Special Issue – AI-Based Environmental Monitoring with UAV Systems

PE&RS

February 2023 Volume 89, Number 2

Photogrammetric Engineering & Remote Sensing The official journal for imaging and geospatial information science and technology

PUBLISHING OPEN-ACCESS IN *PE&RS* IS NOW EASIER!

ASPRS is changing the subscription model of our monthly journal, *PE&RS*. ASPRS is waiving open-access fees for primary authors from subscribing institutions. Additionally, primary authors who are Individual Members of ASPRS will be able to publish one open-access article per year at no cost and will receive a 50% discount on open-access fees for additional articles.

- **Open Access matters!** By providing unrestricted access to research we can advance the geospatial industry and provide research that is available to everyone.
- **Institutions and authors receive more recognition!** Giving permission to everyone to read, share, reuse the research without asking for permission, as long as the author is credited.
- **Reputation matters!** Known for its high standards, *PE&RS* is the industry leading peer-review journal. Adding open access increases authors' visibility and reputation for quality research.
- **Fostering the geospatial industry!** Open access allows for sharing without restriction. Research is freely available to everyone without an embargo period.

Under the previous subscription model, authors and institutions paid $1500 or more in open-access fees per article. This will represent a significant cost savings. Open-access publications benefit authors through greater visibility of their work and conformance with open science mandates of funding agencies.

Subscriptions asprs.org/subscribe
Membership asprs.org/membership

asprs THE IMAGING & GEOSPATIAL INFORMATION SOCIETY

ANNOUNCEMENTS

Pickett is pleased to share the opening of a new office in Hendersonville, North Carolina! This is especially exciting as it is the first space that Pickett has shared with ESP, the parent company of Pickett and Associates. Hendersonville is officially open and shared by geospatial, survey and engineering team members from both Pickett and ESP. The space was opened due to a large demand from the clients of both companies in the Mid-Atlantic region, and offers room for growth of additional team members.

Our new office is located at 1027 Fleming Street, Suite D, Hendersonville, NC, 28791. We look forward to our continued expansion of offices and team members! For more information visit https://www.pickettusa.com/.

Woolpert was contracted by the U.S. Army Corps of Engineers and Pro-ROV Services for the Browns Valley Irrigation District under separate contracts to collect and integrate lidar and multibeam sonar data. These data will be used to determine the accurate capacity of two California reservoirs and to support the management of water resources.

For the Browns Valley Irrigation District project, Woolpert collected lidar data via uncrewed aircraft system and acquired multibeam sonar data with a survey vessel. These data will be used to map and model Collins Lake, a reservoir located northeast of Sacramento in the foothills of the Sierra Nevada Mountains.

Under a separate contract, the Corps selected Woolpert to collect multibeam sonar data for capacity surveys conducted at Lake Isabella, east of Bakersfield. At the same time, Woolpert was under contract to fly topographic lidar over the region for the U.S. Geological Survey 3D Elevation Program. By combining the hydrographic data collected at Lake Isabella and integrating it with USGS lidar data, Woolpert developed a high-resolution digital elevation model of the reservoir and watershed to accurately determine the capacity of Lake Isabella. This opportunity to leverage two surveys produced a valuable dataset that will be used for many years.

"With the ongoing drought, the accuracy of these data is essential to the effective management of water resources," Woolpert Certified Hydrographer David Neff said. "Collecting and integrating lidar and multibeam sonar data ties together two technologies and creates a precise and seamless elevation model. That model is then hosted within a point cloud for state and local officials to access, so they can effectively plan and manage these critical assets."

Neff said UAS was employed for Collins Lake and fixed-wing aircraft was used for Lake Isabella due to the specific needs of each region and each project. He added that the sonar data was collected by eTrac, a Woolpert Company, which is a vessel-based hydrographic survey firm acquired by Woolpert in February 2022 with headquarters in San Rafael.

"We are fortunate to have the staff and equipment to collect data as each project demands," Neff said. "With three offices in California, we understand all too well the impact that the drought has on the state. Advanced technologies such as these hold the key to the state's ability to overcome the water crisis. We are here and ready to help." For more, visit woolpert.com.

Three **Halff Associates, Inc**. (Halff) projects collected Engineering Excellence Awards from the Texas chapter of the American Council of Engineering Companies (ACEC).

Two projects received gold medals, while one earned silver. Halff's French Creek Natural Channel Restoration and Flood Control project was awarded a gold medal in the Water Resources category, while the East Austin Emergency Utility Repair effort received gold in Special Projects. The City of Sugar Land Integrated Stormwater Management Model (ISWMM) earned a silver medal distinction in Studies, Research and Consulting Engineering Services.

The French Creek Natural Channel Restoration and Flood Control project included 2,611 linear feet of channel improvements to mitigate flood conditions. Reconstruction of a half mile of an existing two-lane collector roadway and the addition of one new span bridge and two new bridge-class culverts provided a safe all-weather access facility for this community. The project included natural channel design and stormwater filtration features for water quality, construction of retaining walls and relocation of seven major franchise utilities. Key projects elements were the construction of three new bridge structures and downstream channel improvements to alleviate flooding of the roadway during low- and high-frequency storm events (including the 100-year intensity).

Meanwhile, East Austin water main breaks along a failing box culvert prompted the City of Austin to declare an emergency and initiate a forensic data collection effort. In November 2020, the City contracted with Halff to begin designing repairs to the box culvert joints and failing infrastructure. Halff designed the repairs within an accelerated schedule to ensure construction was not delayed and to reduce the length of delays within the neighborhood. The emergency declaration suspended Austin's normal permit approval process. Halff was responsible for coordinating with

all City departments for informal approvals as the design progressed. Construction began in January 2021. The project was substantially completed in December 2021 below the construction budget determined by Halff.

Sugar Land's ISWMM connects to 27 rain/stream gauges across the city that report real-time information, which creates instantaneous ponding maps that include street ponding depths. The system gives the public reliable information throughout Sugar Land and allows the City to send notifications about intensifying street ponding, flooding and potential street closures. Other benefits include assisting in evacuation routes, coordinating hazard mitigation plans and evaluating drainage infrastructure to better target improvement areas.

Halff has won a firm-record 21 projects in 2022. For more information about Halff, visit halff.com.

SkyWatch Space Applications Inc. ("SkyWatch"), a Canadian space tech company democratizing remote sensing data, announced the launch of EarthCache-X (EC-X) and its partnerships with Capella Space, ICEYE, Satellogic (NYSE: SATL), and Vexcel Imaging, to deliver emerging data types to the commercial Earth observation (EO) market.

EarthCache-X (EC-X) provides a new means for remote sensing partners to go to market with flexible ways to handle data in custom formats that customers might require. With the addition of these industry-leading partners to SkyWatch's repertoire, customers will now have access to SAR, DEMs, aerial imagery and stereo/tri-stereo imagery.

EC-X presents a flexible platform for ingesting and distributing new data types to the EarthCache customer base. The way a customer searches for and receives data remains the same; EC-X still eliminates the friction of multiple contracts by providing a single business relationship with SkyWatch. With these new data providers, SkyWatch is able to create bespoke solutions for customers, defining the right product offerings and price points on a custom basis.

SkyWatch launches the EC-X program with an exceptional set of partners, each providing unique sensing technology that is underutilized by the commercial market today. Customers can now ensure that they are able to rapidly adopt and innovate around new sensing technologies, while our data partners grow their presence in the commercial market and learn how to optimize their solutions for those end markets. Over time, SkyWatch will migrate these in-demand sensing technologies into fully-integrated product offerings, with standardized pricing and output types.

Each of the inaugural EC-X partners was selected to bring strategic value to SkyWatch customers, in response to the demand SkyWatch is seeing in the market and on its Earth observation distribution platform, EarthCache.

Capella Space has the most responsive and agile SAR imaging services in the industry on the leading edge of rapid data acquisition for emergencies or tip and cue, with the added benefit of fully-automated scheduling, downlink, and processing architecture. Owning the world's largest SAR satellite constellation, ICEYE delivers proven and reliable Earth Observation solutions. ICEYE's high revisit and daily coherent change detection capabilities enable users to gain robust and comprehensive situational awareness. Satellogic is a vertically integrated EO satellite and data company that makes EO data more accessible, more reliable, and more affordable, offering the best price across the industry with the largest daily data collection capacity. Satellogic aims to deliver daily global remaps with the end goal of 200+ satellites in orbit. Vexcel runs the world's largest aerial imagery and geospatial data program, collecting multiple types of aerial content in 25+ countries. Vexcel delivers products such as Oblique, True Ortho, and Digital Surface Model data at up to 7.5cm resolution. Imagery and data are also optimized for ML/AI capabilities.

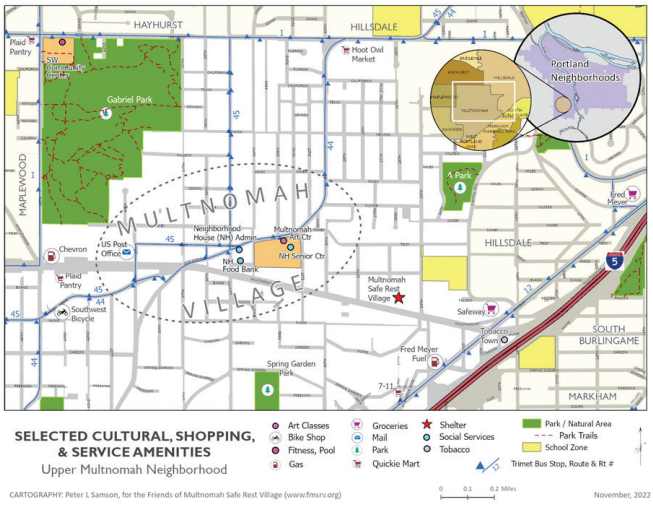
The addition of SAR, DEMs, and aerial data enriches the SkyWatch aggregation model, which provides Earth observation data customers access to the largest network of industry-leading data sets within a single solution with a flexible, pay-per-use pricing model.

To learn more about EarthCache, visit getstarted.skywatch.com/ec-x.

CALENDAR

- 15-17 February, **ASPRS Annual Conference at Geo Week**, Denver, Colorado. For more information, visit https://my.asprs.org/2023conference.
- 5 May, **ASPRS GeoByte — SeaSketch 2.0: A New, Free and Open Source software Service for Map-based Surveys and Collaborative Geodesign.** For more information, visit https://www.asprs.org/geobytes.html.

PE&RS
PHOTOGRAMMETRIC ENGINEERING & REMOTE SENSING
The official journal for imaging and geospatial information science and technology
February 2023 Volume 89 Number 2



73 ASPRS SAC GIS Day Map Contest Winners

COLUMNS

ANNOUNCEMENTS

DEPARTMENTS

77 Special Issue Introduction — AI-Based Environmental Monitoring with UAV Systems

Tolga Bakırman

See the Cover Description on Page 64

COVER DESCRIPTION

Exploding white dwarf stars are thought to be the ultimate source of much of our solar system's lithium. But on Earth, there are certain environments where the soft, light, silvery-white metal is most concentrated and easily mined—notably in briny groundwater aquifers found beneath desert salt flats.

An ideal climate for mining lithium is generally arid, punctuated by seasonal rains or melting. This allows water to pool in shallow, salty lakes and then evaporate during the summer—a cycle that helps concentrate the lithium. Thermal hot springs, volcanic activity, and a subsiding landscape also typically accompany major lithium reserves.

These conditions abound in the Atacama Desert in South America, where some of the world's most largest lithium deposits and mining operations are located. But they can also be found in Nevada's Clayton Valley—the site of the only active lithium mine in the United States. The town of Silver Peak was established in the 1860s around gold and silver mines, but since the 1960s the town has focused on harvesting lithium, which is concentrated under the Clayton Valley in an area where an extinct volcano left lithium-rich deposits.

The Operational Land Imager-2 (OLI) on Landsat 9 acquired this natural-color image of the lithium mining operation in Silver Peak on December 8, 2022. The mine pumps brine to the surface and shunts it into a series of shallow evaporation ponds. Color variations in the ponds are due to varying concentrations of lithium in the water; lighter blue ponds have higher concentrations of lithium. The valley's frequently dry, sunny, and windy weather evaporates water and leaves an increasingly concentrated lithium brine—a process that can take 18 months or more. The brine is then pumped to a nearby facility where it is processed, dried into a powder, and packaged.

Global demand for lithium has ballooned in recent years and is forecast to continue surging in the coming decade because the metal is used in the rechargeable batteries used to power electric vehicles. Makers of laptops, cell phones, and other products with rechargeable batteries also rely heavily on lithium. Lithium is also used in ceramics, certain types of glass, industrial grease, and some types of medication.

Like many lithium mines around the world, the mine at Silver Peak is adding new pumps and evaporation ponds to try to keep up with the demand. The mine, which has been the only U.S. source of lithium for decades, plans to double its production by 2025. Demand for U.S.-produced lithium is particularly strong due to tax incentives that reward consumers for buying electric vehicles with domestically produced and assembled parts.

The Silver Peak mine generates about 1 percent of the world's annual production of lithium. Brine operations, mostly in Chile and Argentina, generate about 75 percent of worldwide production. The mining of lithium-containing ore, which is especially common in Australia, accounts for the rest.

To view the complete image, visit https://landsat.visibleearth.nasa.gov/view.php?id=150730.

NASA Earth Observatory image by Lauren Dauphin, using Landsat data from the U.S. Geological Survey. Story by Adam Voiland.

asprs
THE IMAGING & GEOSPATIAL INFORMATION SOCIETY

PHOTOGRAMMETRIC ENGINEERING & REMOTE SENSING

JOURNAL STAFF

Publisher ASPRS
Editor-In-Chief Alper Yilmaz
Director of Publications Rae Kelley
Electronic Publications Manager/Graphic Artist
Matthew Austin

Photogrammetric Engineering & Remote Sensing is the official journal of the American Society for Photogrammetry and Remote Sensing. It is devoted to the exchange of ideas and information about the applications of photogrammetry, remote sensing, and geographic information systems. The technical activities of the Society are conducted through the following Technical Divisions: Geographic Information Systems, Photogrammetric Applications, Lidar, Primary Data Acquisition, Professional Practice, Remote Sensing Applications, and Unmanned Autonomous Systems. Additional information on the functioning of the Technical Divisions and the Society can be found in the Yearbook issue of *PE&RS*.

All written correspondence should be directed to the American Society for Photogrammetry and Remote Sensing, PO Box 14713, Baton Rouge, LA 70898, including general inquiries, memberships, subscriptions, business and editorial matters, changes in address, manuscripts for publication, advertising, back issues, and publications. The telephone number of the Society Headquarters is 301-493-0290; the fax number is 225-408-4422; web address is www.asprs.org.

PE&RS. *PE&RS* (ISSN0099-1112) is published monthly by the American Society for Photogrammetry and Remote Sensing, 8550 United Plaza Blvd, Suite 1001, Baton Rouge, Louisiana 70809 . Periodicals postage paid at Bethesda, Maryland and at additional mailing offices.

SUBSCRIPTION. *PE&RS* is available as an e-Subscription (single-site and multi-site licenses) and an e-Subscription with print add-on (single-site license only). PE&RS subscriptions are on a calendar-year, beginning in January and ending in December.

The rate for a single-site e-Subscription for the USA/Non-USA is $1040 USD, for Canadian* is $1092 USD.

The rate for a multi-site e-Subscription for the USA/Non-USA is $1040 USD plus $250 USD for each additional license, for Canadian* is $1092 USD plus $263 for each additional license.

The rate for e-Subscription with print add-on for the USA is $1525 USD, for Canadian* is $1612 USD, and for Non-USA is $1565 USD.

*Note: Subscription prices for Canada includes 5% of the total amount for Canada's Goods and Services Tax (GST #135123065). **PLEASE NOTE: All Subscription Agencies receive a 20.00 USD discount.**

POSTMASTER. Send address changes to PE&RS, ASPRS, PO Box 14713, Baton Rouge, LA 70898. CDN CPM #(40020812).

MEMBERSHIP. Membership is open to any person actively engaged in the practice of photogrammetry, photointerpretation, remote sensing and geographic information systems; or who by means of education or profession is interested in the application or development of these arts and sciences. Membership is for one year, with renewal based on the anniversary date of the month joined. Membership Dues include a 12-month electronic subscription to *PE&RS*. Annual Individual Membership dues are $150.00 USD and Student Membership dues are $50.00 USD. A tax of 5% for Canada's Goods and Service Tax (GST #135123065) is applied to all members residing in Canada.

GIS Tips & Tricks

Al Karlin, Ph.D. CMS-L, GISP

When Time-of-Day Matters; Visualizing DEMs

The acronym, DEM, which stands for Digital Elevation Model, can have different meanings to different users. Some users interpret a DEM to be any digital representation of the elevation, including digital contours (vectors) and triangular irregular networks (TINs). Others, myself included, use DEM to refer to raster-based representations of elevation where the values of each raster cell are derived from elevations (adjusted to some datum) of some surface. The surface referenced in the DEM may be the bare-earth terrain, devoid of any above-surface features, man-made or natural, often referred to as a Digital Terrain Model (DTM) or some other surface, such as the tree canopy or building rooftop surface, often referred to as a Digital Surface Model (DSM). DEM enhancement, such as hydro-flattening and/or hydro-enforcing, is the topic of a future column. Now, using lidar (IfSAR, sonar and/or phodar/photogrammetry) as the source data, the ground (or submerged ground) returns are processed into a bare earth DTM, and the first returns are processed into a specific DSM.

This month's tips focus on visualizing DEMs, specifically IfSAR (Interferometric Synthetic Aperture Radar) and/or lidar-derived DTMs. The illustrations were constructed using ArcGIS Pro 3.0 and an IfSAR DTM, but the workflow to produce the end products are identical in both the previous Pro versions (2.X) and the Esri Desktop products (10.X and earlier). Similar workflows can be used for making these visualizations in QGIS and GlobalMapper. NOTE: The DTM used in the illustrations was obtained from the US National Map (TNM Download v2 [nationalmap.gov]); a topic for a future Tips & Tricks column.

Tip #1 — Building Raster Pyramids

When a raster layer is first added to the Contents Pane, the user is asked to build pyramids and calculate statistics. This is a one-time calculation and by doing so will add a metadata file (.aux) and an overlay (.ovr) file to the directory with the raster DTM (Figure 1). The raster pyramids and statistics decrease the rendering time of the raster, and I recommend building the pyramids/statistics as a routine when working with rasters. Deleting these file will not alter the raster, but will increase the rendering time.

By checking the "Always use this choice" box (Figure 1), Pyramids and Statistics for all future rasters newly added to your map will be calculated and this prompt will not reappear. (NOTE: To re-activate the prompt, or change the settings, use Project | Options | Raster and Imagery.)

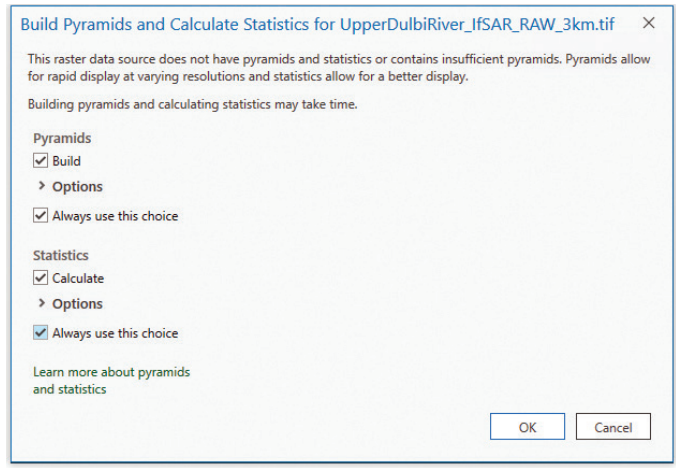


Figure 1. Building Pyramids and Statistics when first adding a raster to ArcGIS Pro3.0. Check the "Always use this choice" to affect all future rasters added to the Contents Pane.

Tip #2 — Adjust the Default Color Scheme (ramp)

The default color scheme ramp for a DTM, or any DEM for that matter, is a the black (low elevation) to-white (higher elevation) color ramp as in Figure 2 below. This default ramp, while representative of the elevation, is neither very informative nor visually pleasing (looks like an X-ray image to me.)

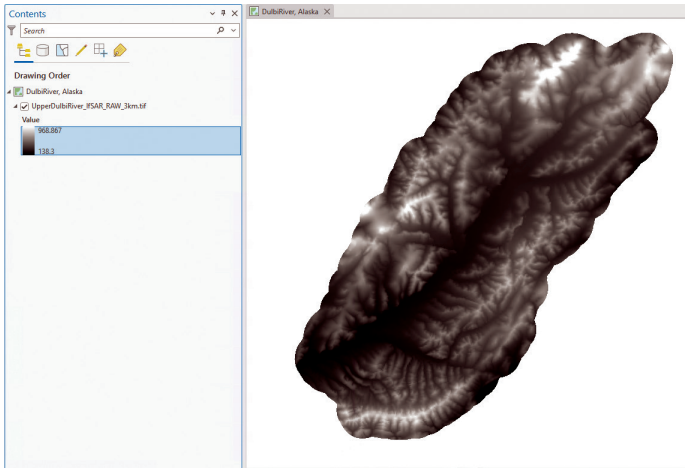


Figure 2. The default color scheme for a DTM newly added to the map. The elevation range is seen in the Contents pane.

Photogrammetric Engineering & Remote Sensing
Vol. 89, No. 2, February 2023, pp. 65-68.
0099-1112/22/65-68

doi: 10.14358/PERS.89.2.65

Of course, it is easy to alter the color scheme and most GIS software programs provide multiple color scheme choices. As the elevation is expressed as a continuous value, choose a continuous, rather than discrete, color scheme. In the ArcGIS Pro, you can either right-click on the color scheme ramp (highlighted in Figure 2) in the Contents Pane, and use the drop-down arrow to reveal the choices, or you can select the DTM in the Contents Pane and open the Symbology Tab (Figure 3) to alter the color scheme.

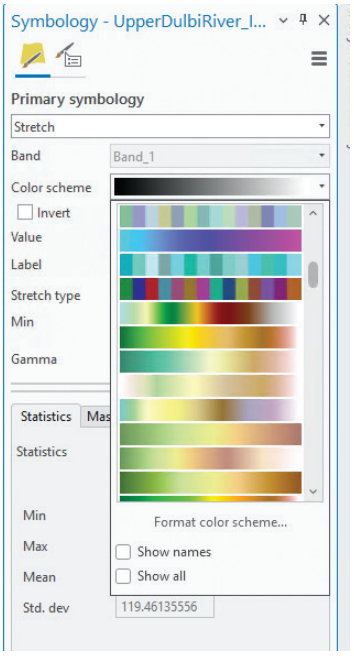


Figure 3. The Symbology Tab in ArcGIS Pro 3.0 shows the Primary Symbology (as Stretch) and the default color scheme. Using the arrow to the right of the Color Scheme reveals user-defined choices for color schemes.

Choosing a continuous color scheme ramp (I chose the Elevation #1 color ramp, fifth color scheme from the top) results in the DTM being rendered as in Figure 4. This U.S. Geological Survey (USGS) standard elevation color scheme runs from Blues (low elevations) to Greens to Red/Browns, and finally to Whites (high elevations) and provides the viewer with additional visual information. Now the user can clearly visualize the river draining from the northeast (higher elevations) to the southwest (lower elevations), as well as the highest elevation (white) mountain peaks. While this rendering is more visually appealing than the standard black-to-white shading, but, at least to me, it looks a little primitive.

Tip #3 — Changing the Primary Symbology of the DTM

To further enhance the rendering of the DTM, notice that the default on the Primary Symbology tab is set to "Stretch" and that there is a dropdown arrow (to the right) on that same line. Using the dropdown (Figure 5) reveals several additional choices for rendering. The choice that I prefer is the "Shaded Relief". This is frequently referred to as a "Hill-shade" in Esri and other GIS software packages, including QGIS and the WhiteBox Toolset (discussed in a previous Tips & Tricks column).

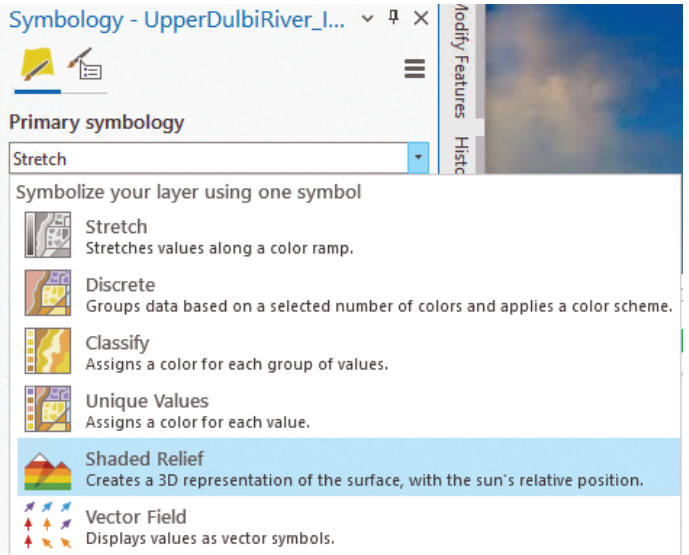


Figure 5. Options for the Primary symbology raster rendering. A simple choice to enhance the rendering is to choose "Shaded Relief".

Deploying the Shaded Relief renderer (Figure 5) opens a new tab (Figure 6) with additional options. The default is to have the sun at 45 degrees (relative to the horizon) and at an Azimuth of 315 degrees (northwestern portion of the sky as in mid-afternoon) and no vertical exaggeration (Z Factor = 1). These parameters result in the rendering of the DTM as in Figure 6.

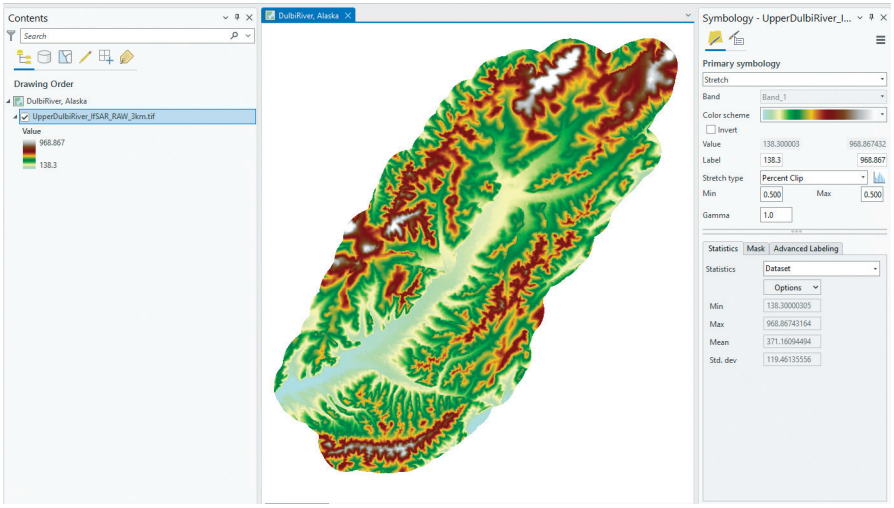


Figure 4. The DTM rendered with the USGS Standard Elevation #1 color scheme. Note that the Primary Symbology is set to "Stretch".

Setting the Z Scale factor to a 5X exaggeration results in a more dramatic and easily interpretable rendition of the DTM as in Figure 7.

With the 5X Z Factor exaggeration, the river is even better defined as are the mountain valleys, ridges and peaks.

Adjusting the sun angle to a position lower in the sky (Altitude = 15 degrees) exaggerates the shadows as in Figure 8 tends to increase shadow lengths, while increasing the sun altitude reduces shadows.

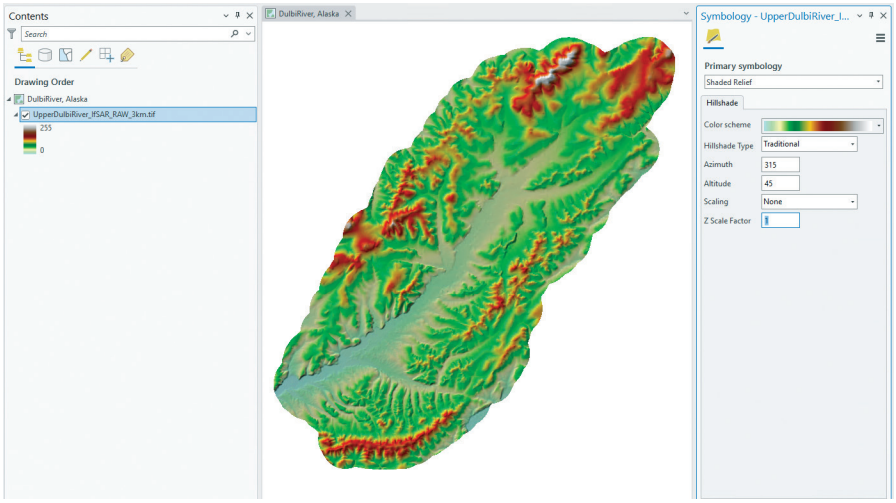


Figure 6. The Shaded Relief options tab. The user can specify the compass bearing (Azimuth) and height on the meridian (Altitude) for solar illumination, and an exaggeration (Z Scale) factor. The default solar parameters simulate a mid-afternoon illumination and shadows.

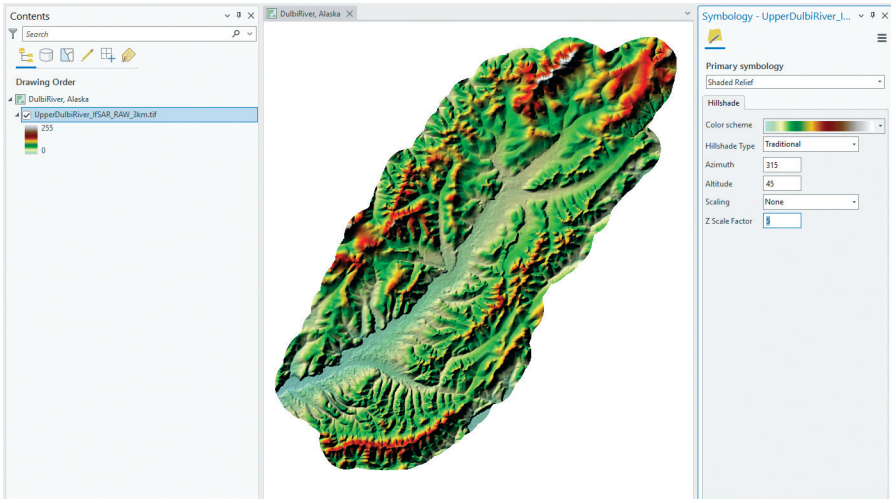


Figure 7. Shaded Relief rendering of the DTM with the default sun Azimuth and Altitude, but with a 5X Z Scale Factor exaggeration.

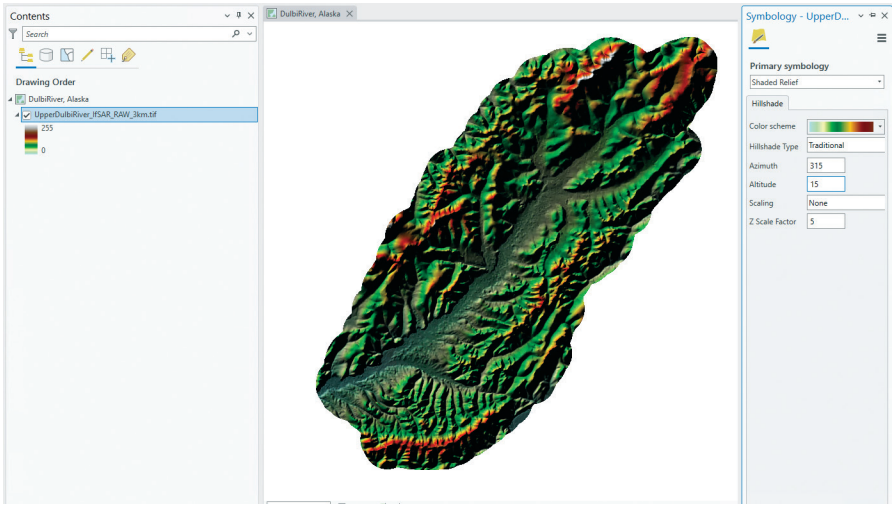


Figure 8. DTM rendered with sun illumination at 15 degrees Altitude showing more shadows than in Figure 7.

Similarly, the direction of the shadows, simulating time of day, can be changed by altering the Azimuth to have the sun in the east (~45 – 115 degrees) simulating a morning view, rather than the west (351 degrees) simulating an afternoon view.

I prefer to maintain the late afternoon sun angle and shadows (Azimuth = 315 and Altitude = 45) but adjust the Z-factor for flatter terrains (Z Factor = 8), normal terrains (Z Factor = 5), and mountainous terrains (Z Factor = 3).

Tip #4 — Create Your Own Hillshade DTM and Customize the Look

To achieve an even more customized look, in ArcGIS Pro open the Geoprocessing Tab and either search for or navigate to the "Hillshade" Geoprocessing Tool (Spatial Analyst | Surface | Hillshade or 3D Analyst Tools | Raster | Surface | Hillshade). You can build this raster in Esri (Hillshade), QGIS (in Raster |Terrain Analysis), in WhiteBox (Geomorphometric Analysis | Hillshade), SAGA (Terrain Analysis- Lighting, Visibility | Analytical Hillshading), and several other GIS programs. In general, you will see the same, or similar, parameters as previously, Sun Azimuth, Sun Altitude and Z Factor (vertical exaggeration) to construct the Hillshade raster.

In the example in Figure 9 (next page), I built the Hillshade raster using the Geoprocessing Tool in Spatial Analyst selected the DTM as input, set the afternoon sun azimuth and altitude (315 degrees and 45 degrees, respectively) and specified a 5X vertical exaggeration (Z Factor).

Then, I placed the Hillshade5X.tif raster below the DTM raster in the Drawing Order on the Contents Pane. Selecting the Original DTM, I returned the Primary Symbology to "Stretch" but retained the USGS Elevation #1 color scheme and made the DTM raster 35% transparent (Figure 10).

This workflow results in a visually appealing rendition (Figure 11) of the DTM similar to the Shaded Relief option in Tip 3, and permits additional customization of the DTM rendering independent of the "Shaded Relief" option.

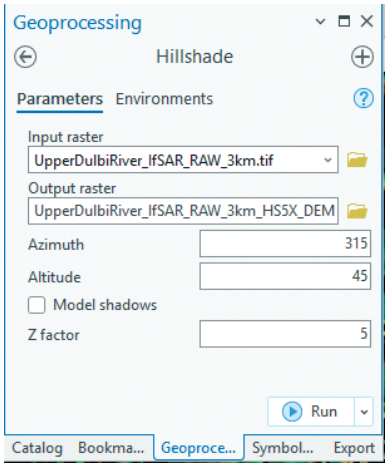


Figure 9. The Spatial Analyst Hillshade Geoprocessing Tool dialog used to construct a 5X exaggerated hillshaded raster.

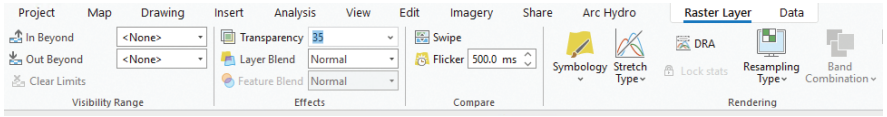


Figure 10. Selecting the original DTM in the Contents Pane, then selecting Raster Layer on the Ribbon and adjusting the Transparency to 35%.

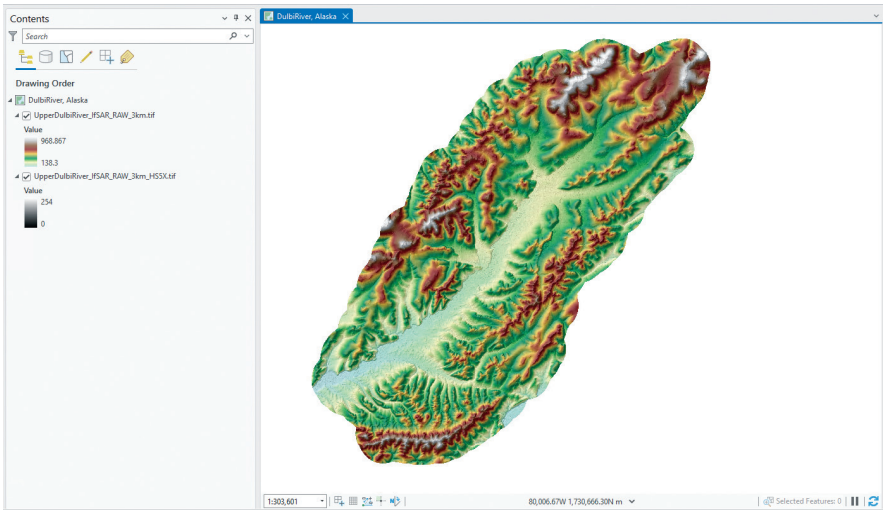
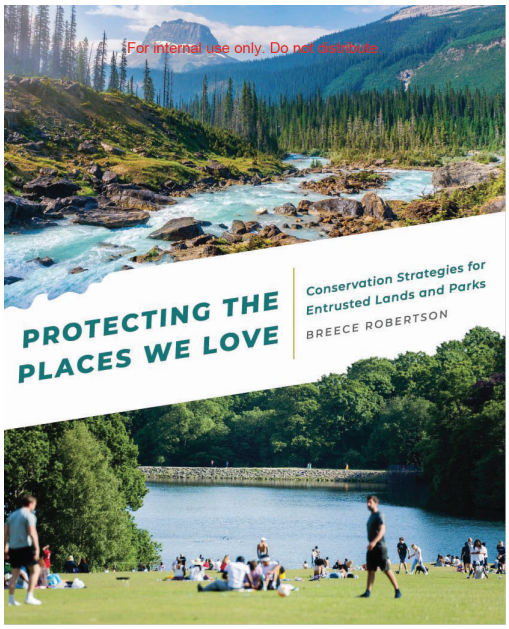


Figure 11. DTM rendered with 35% Transparency and an independent 5X exageration Hillshade raster showing through from beneath.

You can experiment with different color scheme ramps, different lighting conditions and different transparency levels to achieve the precise rendering you desire for your map.

Send your questions, comments, and tips to GISTT@ASPRS.org.

Al Karlin, Ph.D., CMS-L, GISP is with Dewberry's Geospatial and Technology Services group in Tampa, FL. As a senior geospatial scientist, Al works with all aspects of Lidar, remote sensing, photogrammetry, and GIS-related projects. He also teaches beginning map making at the University of Tampa.



Protecting the Places We Love: Conservation Strategies for Entrusted Lands and Parks

By Breece Robertson

ESRI Press: Redlands, California. 2021. 280 pages, including diagrams, maps, photos, images, index. Paperback or eBook. $29.99. ISBN 9781589486164, eISBN 9781589486171.

Reviewed by Matthew E. Ramspott, Ph.D., Professor, Department of Geography, Frostburg State University, Frostburg, Maryland.

According to author Breece Roberston, this book project (her first) was developed to share insights gained from years of experience working at the intersection of the worlds of land conservation advocacy and geospatial science and technology. Breece, currently the Director of Partnerships and Strategy for the Center for Geospatial Solutions at the Lincoln Institute of Land Policy, previously worked at the Trust for Public Lands, leading efforts devoted to land protection, improving park access and utilization, and development of voter-approved public funding for parks and conservation projects. In a 2021 promotional interview[1] with ESRI co-founder and president Jack Dangermond, Breece suggests that her book was written for a broad audience but targeted toward land trusts and other conservation groups that are currently underutilizing GIS. As Jack suggested in the interview, a primary goal of the book is to "catalyze the conservation community with quantitative methods." As access barriers (fiscal, technical, organizational, or otherwise) to geospatial science and tech have become less onerous in recent years, Breece calls attention to this opportunity at a critical time when environmental challenges have arguably reached crisis proportions. The book is organized into ten Chapters, each highlighting opportunities for conservation organizations to improve their effectiveness and efficiency through the enhanced utilization of geospatial data and tools. The book is illustrated with colorful maps, graphics, and photos throughout.

The Foreword sets the tone with a vigorous call to action, making the case that in an environment of complex and competing political and economic priorities, all available tools and methods need to be explored to advance the goals of land conservation. Each subsequent chapter unfolds with reference to resources and ideas addressing various aspects of the subject. Some specific topics include: a broad primer on using maps as a device to focus attention on places and problems; discussion of the importance of enhancing urban parks and greenspace systems with a view toward balancing access, equity, biodiversity, and connectivity; understanding and documenting the relationship between climate resilience & green infrastructure; measuring and understanding landscape connectivity; innovative and effective ways of promoting community engagement; building strategy and impact into the park planning process; and bringing Geographic Information Science into the critical process of evaluating and communicating the impacts and benefits of conservation efforts. The latter chapters highlight the availability of new capabilities that can enhance public interface with parks and conservation efforts and offer a strategic look at ways to improve the integration of GIS capabilities into traditional organizational structures. The cross-referenced Index gives some help to readers wishing to review and revisit specific examples.

[1]Interview: https://www.esri.com/en-us/esri-press/browse/protecting-the-places-we-love-conservation-strategies-for-entrusted-lands-and-parks.

One clear strength of the book is the large amount of practical discussion centered around data sourcing. This is a focus that runs throughout the book, pertaining to a wide variety of issues of interest to those engaged with advocating for and assessing the benefits of wildland conservation, parks, and green infrastructure development. Highlighted are US Census-derived demographic and socioeconomic data, multiscale land use/land cover data, outputs from ecological and habitat models, archives of location-based information on conservation projects, and many more datasets. This fundamental aspect of the book seems effective as an idea

Photogrammetric Engineering & Remote Sensing
Vol. 89, No. 2, February 2023, pp. 69-70.
0099-1112/22/69-70

doi: 10.14358/PERS.89.2.69

generator to help with the potentially challenging process of locating appropriate and readily accessible data. Discussion of data reliability is perhaps a little sparse, but the sources referenced here are credible and authoritative, and issues related to data integrity and accuracy have been covered in depth elsewhere.

The book gives a comprehensive view of the processes and challenges associated with effective management of land conservation efforts, making use of representative examples far too numerous to address completely in this short review. Overall, the tone is grounded in practical organizational considerations such as framing goals and questions, identifying potential stakeholders and collaborators, strategic planning & prioritization, inviting public participation, accessing useful geospatial data and tools, and integrating geographic information into advocacy, outreach, and fundraising efforts.

It should be noted that this is not a highly technical book, nor is it intended to be. The book surveys many methods and structures associated with geospatial analysis but does not venture far into the realm of a technical "how-to" manual. Perhaps unsurprisingly, much of the description of software and apps in the book prominently features ESRI products. The focus instead is on underscoring the potential value of geospatial science and detailing successful examples to point the way forward. The projects described in the book range from the very local scale (e.g. the redevelopment of a local neighborhood park or schoolyard) to intermediate scale (e.g. evaluating park needs in a large metropolitan area with diverse socioeconomic situations) and regional scales (e.g. tracking the locations and connectivity of conservation projects with an eye towards meeting more broadly integrated conservation goals and objectives). The many benefits of community partnerships and collaborations are weaved throughout the narrative.

In conclusion, Protecting the Places We Love seems well-suited to its target audience. With a focused and deep understanding of the operational concerns of organizations devoted to the expansion of land conservation and green space, Robertson thoughtfully explores her subject in a clear, authoritative, and well-organized manner. The underlying message is about the diverse array of opportunities for conservation and parks advocates to enhance their work by leveraging the capabilities of geospatial science, and Robertson successfully presents it in a way that is both inclusive and empowering.

& GRIDS & DATUMS

BY Clifford J. Mugnier, CP, CMS, FASPRS

REPUBLIC OF EL SALVADOR

The Grids & Datums column has completed an exploration of every country on the Earth. For those who did not get to enjoy this world tour the first time, *PE&RS* is reprinting prior articles from the column. This month's article on the Republic of El Salvador was originally printed in 2005 but contains updates to their coordinate system since then.

When the Spanish Conquistadores first entered into Central America in the early 16th century, Indians of the Pipil tribe occupied the area now known as El Salvador. The Pipil were a subgroup of a nomadic people known as the Nahua, who had migrated into Central America around 3,000 B.C. They eventually fell under the Maya Empire, which dominated Central America until about the 9th century A.D. According to the Library of Congress *Country Studies*, "Pipil culture did not reach the advanced level achieved by the Maya; it has been compared, albeit on a smaller scale, to that of the Aztecs in Mexico. The Pipil nation, believed to have been founded in the 11th century, was organized into two major federated states subdivided into smaller principalities." The Spaniards were initially defeated in 1524 when they attempted to enter the area, and it took until 1528 with two more expeditions to finally subdue the Pipil nation. The fierce warrior Atlacatl is revered to this day to the exclusion of Alvarado who finally overcame the natives. "In this sense, the Salvadoran ambivalence toward the conquest bears a resemblance to the prevailing opinion in Mexico, where Cortes is more reviled than celebrated."

El Salvador (The Savior) is the smallest Spanish-speaking nation in the Western Hemisphere. It is located on the western side of the Central American isthmus. The country is slightly smaller than Massachusetts, and its land boundaries are with Guatemala (203 km), and Honduras (342 km) (*PE&RS*, July 1999). El Salvador's coastline is 307 km on the Pacific Ocean and its terrain is mostly mountainous with a narrow coastal belt and central plateau. The lowest point

is the Pacific Ocean (0 m), and the highest point is Cerro El Pital (2,730 m). El Salvador is known as the "Land of Volcanoes," and is one of the most seismologically active regions on Earth, situated atop three large tectonic plates.

Topographic Mapping of the Americas, Australia and New Zealand states, "To 1930, the only detailed accurate surveying done in El Salvador – a country that became independent in 1821 – was the surveying done for the Intercontinental Railroad Commission and a few surveys related to possible road routes." However, Nicaragua and Honduras had been squabbling regarding their border for more than 30 years, and an arbitration agreement was signed in 1930 and soon thereafter surveyed by the U.S. Coast & Geodetic Survey. El Salvador, Guatemala, and Honduras signed a protocol on 26 March 1936 accepting Cerro Monte Cristo as the tripoint of the boundaries of the three states. It was during this era that the first geodetic surveys were performed on the border between El Salvador and Guatemala from 1937 to 1940. The Dirección General de Cartografía was established in 1946 to

Photogrammetric Engineering & Remote Sensing
Vol. 89, No. 2, February 2023, pp. 71-72.
0099-1112/22/71-72

doi: 10.14358/PERS.89.2.71

continue the geodetic surveys and to implement a topographic mapping of the country. The end of the following year 1947, the Inter American Geodetic Survey (part of the U.S. Army Map Service) signed a cooperative agreement with El Salvador. By 1958, all classical geodetic triangulation had been completed in the country, and topographic mapping was well underway. The national mapping agency is now known as the Instituto Geográfico Nacional "Ing. Pablo Arnoldo Guzmán."

The original horizontal datum of El Salvador is the Ocotepeque Datum of 1935 which was established by the U.S.C.&G.S. at Base Norte (in Honduras) where φ_o = 14° 26′ 20.168″ North, λ_o = 89° 11′ 33.964″ West of Greenwich, and H_o = 806.99 meters above mean sea level. The defining geodetic azimuth to Base Sur is: α_o = 235° 54′ 21.790″, and the ellipsoid of reference is the Clarke 1866 where a = 6,378,206.4 meters and $^1/_f$ = 294.9786982. The corresponding astronomic observations at that mountainous location are: Φ_o = 14° 26′ 13.73″ North (±0.07"), Λ_o = 89° 11′ 39.67″ West (±0.045"), and the defining astronomic azimuth to Base Sur is: α_o = 235° 54′ 20.37″ (±0.28"). The difference between these two sets of coordinates is due to the local gravimetric deflection of the vertical. The grid system used in El Salvador is based on the Lambert Conformal Conic projection where the Central Meridian, λ_o = 89º 00′ 00″ West, the Latitude of Origin, φ_o = 13º 47′ 00″ North, the Scale Factor at Origin, m_o = 0.999967040, the False Easting = 500,000 m, and the False Northing = 295,809.184 m. Sometime after that, the National Geospatial-Intelligence Agency (nee NIMA) computed all the classical triangulation of Central America on the North American Datum of 1927. The 1987 published datum shift parameters for Central America based on 19 stations (TR 8350.2) from NAD27 to WGS84 are: ΔX = 0 m ± 8 m, ΔY = +125 m ± 3 m, ΔZ = +194 m ± 5 m.

After the disaster when Hurricane Mitch hit Central America, the U.S. National Geodetic Survey established a number of GPS Continuously Operating Reference Stations in the region. They observed a number of stations in El Salvador in order to establish and publish a High Accuracy Reference Network on the North American Datum of 1983. The National Geodetic Survey (NGS) has posted all of the data, including various other sources of geodetic coordinate data in El Salvador and the region on their website. The documentation of their geodetic observations is a superb resource for the researcher, and it is freely available. Thanks, NGS!

Update on the Republic of El Salvador

"The reference frame of El Salvador is a GPS base network of 38 stations. It was measured by the *Gerencia de Geodesia, Instituto Geográfico y del Catastro Nacional* (GG-IGCN using differential GPS positioning in partial sub-networks and in 36 daily sessions between October and December 2007. Most of the reference stations (28) were determined together with a secondary control point in order to ensure the long-term stability of the network. The GPS data were processed by *Deutsches Geodätisches Forschungsinstitut* (DGFI) within the SIRGASWGII (Geocentric Datum) activities. The analysis strategy is based on the double difference approach (Bernese Software V5.0, Dach *et al.* 2007)." As is typical of Latin American nations that relate their local coordinate systems to the SIRGAS, they do not publish the transformation parameters from their classical native coordinates to SIRGAS. https://www.sirgas.org/fileadmin/docs/Boletines/Bol14/35_Figueroa_SIRGAS-ES2007_8.pdf

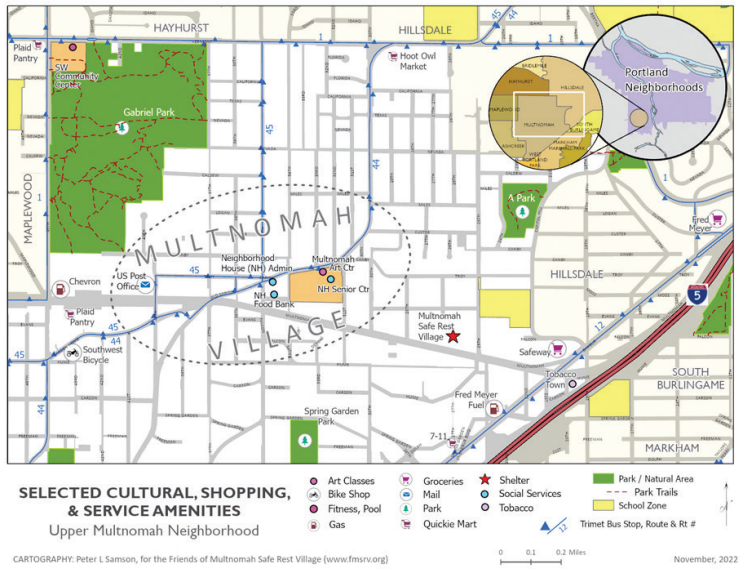
The contents of this column reflect the views of the author, who is responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the American Society for Photogrammetry and Remote Sensing and/or the Louisiana State University Center for GeoInformatics (C^4G).

This column was previously published in *PE&RS*.

ASPRS SAC GIS DAY MAP CONTEST WINNERS!

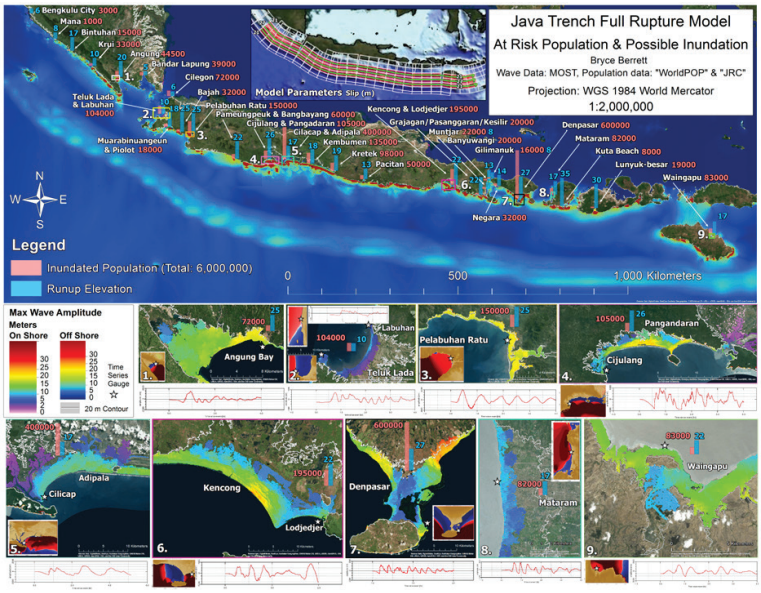
To celebrate GIS Day, the ASPRS Student Advisory Council (SAC) hosted an online Cartography Map Contest. Maps were welcomed in any format including hand drawn submitted maps. The SAC received numerous entries showcasing the application of cartography in multiple disciplines.

Prior to voting, participants were encouraged to share the story behind their maps. All of the contest submissions represented different stories. Four winners were selected. When asked what motivated their participation, the winners said the GIS Map Contest was an amazing opportunity to display and share their work with the remote sensing and geoscience community and participate in the ASPRS community.



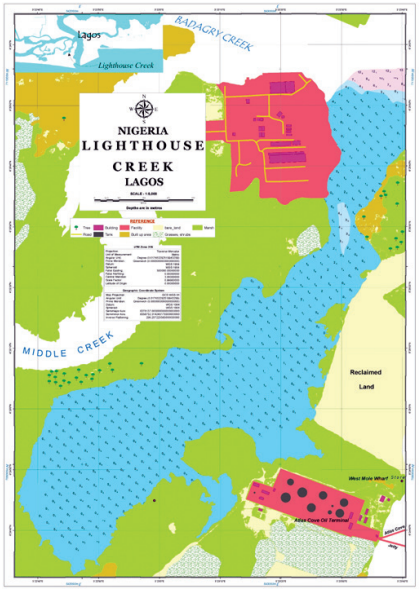
Peter L Samson
Portland State University, Portland OR
Auditing GIS Courses for Fun in Retirement

This map aims to be a welcoming resource for formerly houseless residents recently offered shelter by the City of Portland in a "Safe Rest Village (SRV)" in my neighborhood of Multnomah.. The map includes features, businesses, and facilities most likely to be useful to our new neighbors as they learn skills to navigate their way through the SRV and into more permanent housing. I was requested to make this map by a group of supportive housed neighbors calling itself the *Friends of the Multnomah Safe Rest Village*.



Bryce Berrett
Oregon State University
PhD student in Geomatics

This map shows tsunami inundation and the population at risk from an impending megathrust earthquake that will eventually occur along the Java trench fault in Indonesia. This map was created using dozens of individual tsunami models that were combined to give a comprehensive map of the region. While it is improbable that the whole fault will rupture together, or fail with such high displacement, this map is meant to represent a worse-case scenario for any location along the coast.

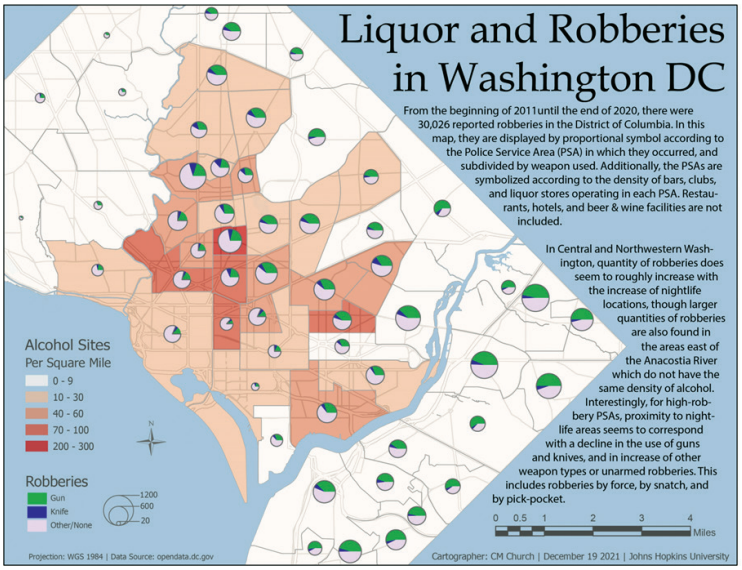


Chukwuma John Okolie
Newcastle University, United Kingdom
Doctoral Candidate in Remote Sensing

This is a bathymetric map of Lighthouse Creek produced from Landsat imagery using Stumpf's ratio transform algorithm. The map shows the variation in depths of the creek. Lighthouse Creek is in the vicinity of Lagos harbour, Nigeria. The creek is relatively shallow with depths ranging from about 0.1 – 6m.

"This contest provides an opportunity to inform readers about an effective and humane model for addressing the houselessness crisis that plagues Portland and so many other American cities".

~ Peter Samson,
2021 GIS Day Map Contest Winner



Conor MacBain Church
Johns Hopkins University
MS, Geospatial Intelligence

This multivariate map uses a classed choropleth to represent the density of hard-liquor serving locations (bars, nightclubs, and liquor stores) in Washington DC, broken down by Police Service Area (PSA, also known as a police beat). At the same time, scaled pie charts are used to portray quantity of reported robberies as well as the proportion in which each type of weapon was used.

GIS Map Day Winners received:

1. Complimentary one-year license of ArcGIS for Personal
2. Map displayed at the ASPRS Annual Conference at Geo Week in February 2023
3. Student Volunteer Registration to the ASPRS Annual Conference at Geo Week

All attendees of the SAC meeting who participated in the live voting received a complimentary one-year premium subscription to Esri's ArcGIS StoryMaps.

2023 ASPRS MEMBERSHIP RENEWALS

Members may notice a few changes in the online renewal process in 2023.

1. As part of their 2023 budget approval, the ASPRS Board of Directors voted to increase annual dues for Individual Members by $25 (from $150 to $175) effective May 1. This is the first dues increase in nearly a decade and is necessitated by rising costs. Sustaining Member and Student Member dues will remain unchanged in 2023.
2. Student Member renewal invoices will be issued at the Individual Member rate of $175, with an option to submit new proof of enrolment to obtain the lower Student Member dues rate of $50.
3. Voluntary tax-deductible donations of $25 to the ASPRS Foundation and $10 to ASPRS Initiatives will be included in the renewal invoice. These contributions are voluntary and can be removed from the cart before proceeding to payment.

If you have any questions about these changes or need help during the renewal process, please contact office@asprs.org or call (301) 493-0290.

NEW ASPRS MEMBERS

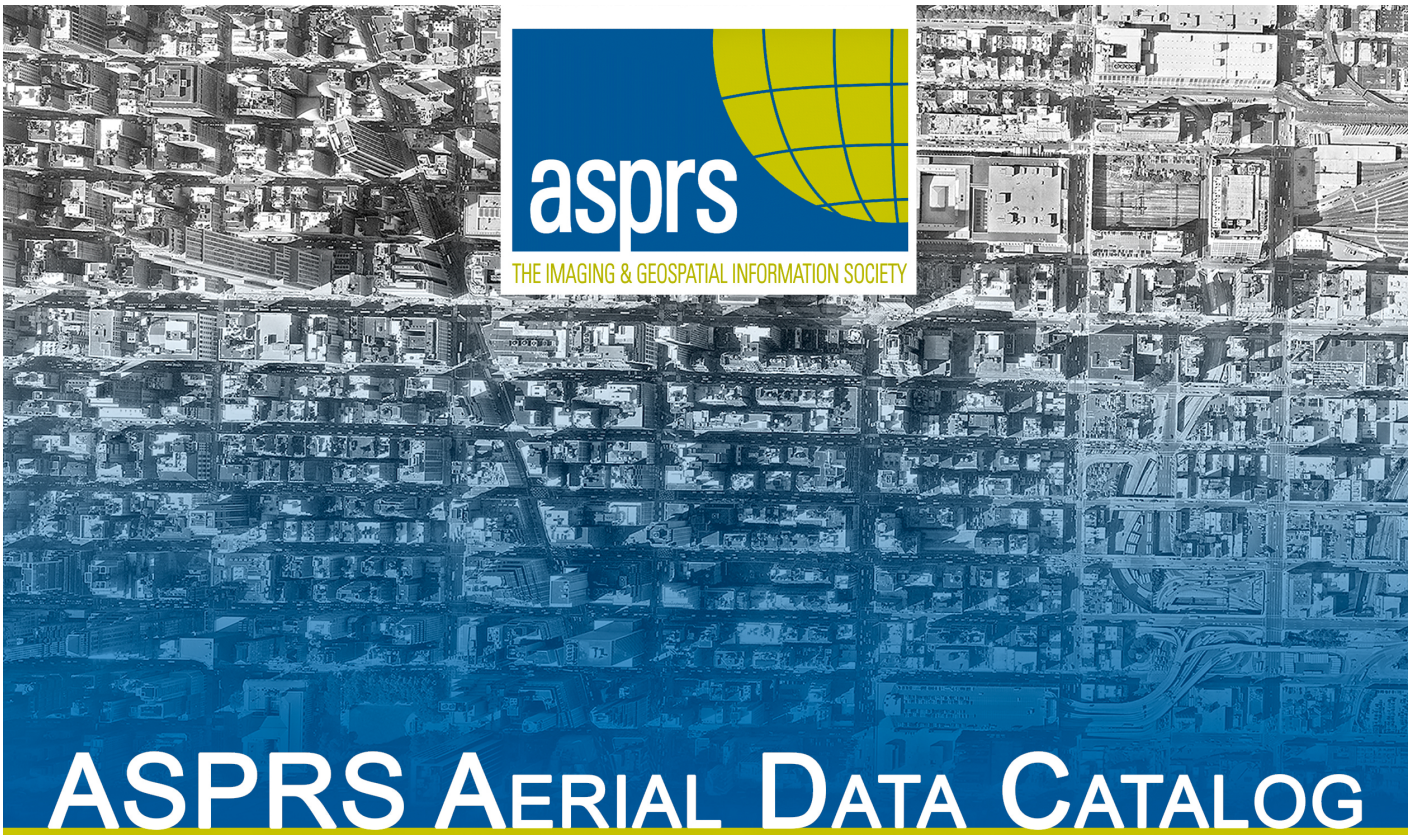
ASPRS would like to welcome the following new members!

Tolga Bakirman
Nishan Bhattarai, PhD
Daniel Bollich
Andre Doria, PhD
Firat Erdem
Michael James Fink
Feng Gao, PhD
Christopher Allen Gray
Seth Gulich
Rob Harrap
Kris Kleiner
Danial Mariampillai
Emma Menio
Andrew Molchan
Andrew Moody
Melanie A. Olson
Robert Patten
Mark Que
Richard Renton
John Tobeck
Victor Valdivia
Michael Winfield

FOR MORE INFORMATION ON ASPRS MEMBERSHIP, VISIT
HTTP://WWW.ASPRS.ORG/JOIN-NOW

Interested in ASPRS apparel? Visit

https://stores.inksoft.com/asprs/shop/home



asprs
THE IMAGING & GEOSPATIAL INFORMATION SOCIETY

ASPRS Aerial Data Catalog

"The Source for Finding Aerial Collections"

http://dpac.asprs.org

The ASPRS Aerial Data Catalog is a tool allowing owners of aerial photography to list details and contact information about individual collections.

By providing this free and open metadata catalog with no commercial intrests, the Data Preservation and Archiving Committee (DPAC) aims to provide a definitive metadata resource for all users in the geospatial community to locate previously unkown imagery.

DPAC hopes this Catalog will contribute to the protection and preservation of aerial photography around the world!

ASPRS Members: We Need Your Help!
There are three ways to get involved

1 USE — Use the catalog to browse over 5,000 entries from all 50 states and many countries. Millions of frames from as early as 1924!

2 SUPPLY — Caretakers of collections, with or without metadata, should contact DPAC to add their datasets to the catolog free of charge!

3 TELL — Spread the word about the catalog! New users and data collections are key to making this a useful tool for the community!

For More Details or To Get Involved Contact:

David Ruiz • druiz@quantumspatial.com • 510-834-2001 OR David Day • dday@kasurveys.com • 215-677-3119

Special Issue Introduction — AI-Based Environmental Monitoring with UAV Systems

Tolga Bakırman, Yildiz Technical University

Global warming and climate change have become the most important factor threatening the world. Climate change results in dramatical environmental hazards and threatens the planet and human life. A wide variety of policies have been proposed to decrease the effects of global warming and climate change. The most important one is the Paris Agreement which aims to limit global warming to well below two degrees Celsius. Many countries have formulated long term low greenhouse gas emission development strategies related to the Paris Agreement which aimed to meet the essential strategies addressing issues with climate change, environmental protection and low carbon.

The astonishing developments on unmanned aerial vehicle (UAV) systems and artificial intelligence (AI) technologies enables a great opportunity to monitor the environment and propose reliable solutions to restore and preserve the planet and human health.

Data acquisition and processing paradigm has been changed as a result of technological developments. It is obvious that new solutions, innovative approaches will make significant contributions to solve the problems which our planet is facing. UAV data can be collected by various platforms (planes or helicopters, fixed wing systems, drones) and sensors for earth observation and sustainable environmental monitoring which are also utilized by the United Nations to support the delivery of its mandates, resolutions, and activities.

UAV based earth observation data and AI techniques have a wide range of applications such as risk management, disaster monitoring and assessment, environmental impact evaluation and restoration, monitoring agriculture and food cycles, urban analysis, digital twin and smart city applications and providing increased situation awareness. This growth of widely available UAV data associated with the exponential increase in digital computing power, machine learning and artificial intelligence plays a key role in the environmental monitoring and solution generation of geospatial information for the benefit of humans and the planet.

This special issue in *Photogrammetric Engineering and Remote Sensing* (*PE&RS*) gathers peer-reviewed papers that advance state-of-the-art in AI-Based Environmental Monitoring with UAV Systems. The papers focus on the use of AI-based method for investigation of the relationship between environmental monitoring, and apricot trees segmentation, wheat leaf nitrogen prediction, energy infrastructure damage assessment, urban building segmentation and car detection.

Forest management and inventory, plant monitoring in urban areas, vegetation distribution, change monitoring, and the establishment of sustainable agricultural systems all require extraction and monitoring of trees. In this regard, Erdem et al. focuses on detecting, counting, and mapping the apricot trees from UAV orthophoto imagery using U-Net and Mask R-CNN deep learning methods.

Nitrogen has a key role in plant development and directly affects the crops' quality. The secret to site-specific management for improved nitrogen usage efficiency and a sustainable crop production system is quantitative measurement of crop nitrogen. Sahoo et al. investigates exploitation of spectroscopy using field portable spectroradiometer on ground in the crop field and imaging spectrometer from UAV to predict leaf nitrogen of wheat crop with artificial neural network, extreme learning machine, least absolute shrinkage and selection operator and support vector machine regression.

In any environmental or natural disaster, effective emergency management is the major key to save lives. The management authorities need rapid and accurate information regarding energy infrastructure's damage assessment in order to guide critical allocations of lifesaving resources within 72 hours. Hughes et al. demonstrates a scalable end-to-end solution using UAV that provides AI-based edge detection, classification and geolocation of utility poles.

Similarly, building segmentation also has numerous application areas including disaster management and urban planning. Sariturk et al. performed a comprehensive performance comparison of state-of-the-art deep learning architectures for building segmentation using aerial imagery.

Environmental monitoring is also related to the urban development and hence it is important to determine the car density in the parking lots, especially in hospitals, large enterprises, and residential areas, which are used intensively, in terms of executing existing management systems and making precise planning for the future. Kaya et al. tested the performance of the two deep learning approaches namely YOLOv3 and Mask R-CNN in order to determine the number of cars in the parking lot using high resolution UAV imagery within the deep learning tools of ESRI ArcGIS Pro.

I believe the topics that the papers have focused on in this Special Issue will have a wider impact in the Environmental Monitoring community. I would like to express my gratitude to the authors, who contributed to this Special Issue of the *PE&RS* for sharing their scientific results. I would also like to acknowledge the voluntary reviewers, whose professional reviews contributed to the further improvement of the submitted papers. I'm also thankful to Prof. Dr. Bülent Bayram from Yildiz Technical University, Istanbul, Türkiye for organizing the International Symposium on Geoinformatics 2021 (https://www.isagsymposium.org/isag2021). Finally, I'm grateful to *PE&RS* Editor Prof. Dr. Alper Yılmaz for his support and guidance in assembling this Special Issue.

Tolga Bakırman is with the Yildiz Technical University, Istanbul, Turkey. bakirman@yildiz.edu.tr

Call for *PE&RS* Special Issue Submissions

Innovative Methods for Geospatial Data using Remote Sensing and GIS

Internationally comparable data is a global need for managing resources, monitoring current trends and taking actions for sustainable living. Even though there has been a significant progress on geospatial data availability, extensive data gaps are still a major problem for general assessment and supervise the progress through the years. According to United Nations 2022 The Sustainable Development Goals Report, while health and energy sectors have the highest data available, limited data available for climate action.

The COVID-19 crisis has also shown that there are innovative data collection methods utilizing information and computer technologies. However, only 5% of the countries have benefit from remote sensing technologies to measure the impact of COVID-19. Additionally, novel approaches such as artificial intelligence should be used in conjunction with assessments to make sure they are put to use for critical situations.

The recent developments in remote sensing, geographic information systems and ICT have provided a wide accessibility to create geospatial data for various purposes. The proposed special issue focuses on *"Innovative Methods for Geospatial Data using Remote Sensing and GIS" for wide range of applications. This special issue aims to bring researchers to share knowledge and their expertise about innovative methods to contribute to fill data gaps around the world for a better future.*

The proposed special issue aims to contributes ASPRS's key mission on 'Simplify and promote the use of image-based geospatial technologies for the end-user', 'Promote collaboration between end users and geospatial experts to match data and technology to applications and solutions' and 'promote the transfer of geospatial data and information technology to developing nations' by providing innovative methods to create geospatial data using remote sensing and geographic information systems utilizing state-of-the-art developments and solutions.

Deadline for Manuscript Submission—July 1, 2023

Submit your Manuscript to http://asprs-pers.edmgr.com

Guest Editors

Dr. Tolga Bakirman, bakirman@yildiz.edu.tr , *Yildiz Technical University, Department of Geomatic Engineering*, Davutpasa Campus, 34220 Esenler-Istanbul/Turkey

Dr. George Arampatzis, garampatzis@pem.tuc.gr, *Technical University Crete, School of Production Engineering & Management*, 73100 Chania – Crete/Greece

UAS Edge Computing of Energy Infrastructure Damage Assessment

Jordan Bowman, Lexie Yang, Orrin Thomas, Jerry Kirk, Andrew Duncan, David Hughes, and Shannon Meade

Abstract

Energy infrastructure assessments are needed within 72 hours of natural disasters, and previous data collection methods have proven too slow. We demonstrate a scalable end-to-end solution using a prototype unmanned aerial system that performs on-the-edge detection, classification (i.e., damaged or undamaged), and geo-location of utility poles. The prototype is suitable for disaster response because it requires no local communication infrastructure and is capable of autonomous missions. Collections before, during, and after Hurricane Ida in 2021 were used to test the system. The system delivered an F1 score of 0.65 operating with a 2.7 s/frame processing speed with the YOLOv5 large model and an F1 score of 0.55 with a 0.48 s/frame with the YOLOv5 small model. Geo-location uncertainty in the bottom half of the frame was ~8 m, mostly driven by error in camera pointing measurement. With additional training data to improve performance and detect additional types of features, a fleet of similar drones could autonomously collect actionable post-disaster data.

Introduction

In the wake of disaster, timely, effective emergency management saves lives (Kerle and Oppenheimer 2002). Emergency management officials such as the US Department of Energy and the Federal Emergency Management Agency (FEMA), and others across the globe, need actionable information for damage assessment to guide critical allocations of lifesaving resources within 72 hours (FEMA 2020; Hodgson *et al.* 2010). Traditionally, damage assessments have been performed by local disaster response crews deployed to the affected area. Any imagery-based data collection (e.g., from unmanned aerial systems (UAS)) suffered from download, processing, and/or upload delays. Collections also frequently relied on local communication infrastructures that may have sustained damage (Boccardo and Tonolo 2015). Methods that rely on social media (e.g., Twitter messages) to infer damage require communications systems in the disaster-damaged area to be operable often experience data gaps and suffer from self-reporting bias in favor of wealthier and more tech savvy individuals (Alam *et al.* 2017). The lack of precise geo-location in social media-based information also makes accurate damage assessment challenging. Efforts using satellite imagery have had limited impact and effectiveness in operation because of challenges related to acquisition lag; low resolution; inconsistent image quality (e.g., clouds, haze); and slow dissemination (Laituri and Kodrich 2008).

Data collection and processing following Hurricane Ida in 2021 illustrated the latency and scope of the problem (ArcGIS Online 2022). Within the 72-hour window of landfall, scattered social media images were collected, ICEYE synthetic aperture radar satellite images were collected for the whole area (for flood detection), National Oceanic and Atmospheric Administration captured crewed aerial imagery along major transportation corridors, the US Civil Air Patrol collected UAS imagery over some areas, and nearly comprehensive crewed aerial imagery was provided by the National Insurance Crime Bureau. The US Civil Air Patrol implemented an effective data management program. However, the data had to be off-loaded from aircraft and satellites and processed to provide actionable information. The crowdsourced analysis began a full four days after landfall. FEMA's ArcGIS damage assessments layer were first published 53 days after landfall. At the time of writing, the most recent update was 219 days after landfall.

This paper describes a scalable method to deliver actionable intelligence (not imagery) within FEMA's designated 72-hour post-disaster window. Our prototype uses a UAS to obtain sufficient resolution; deploys rapidly and autonomously; operates without any need for local communication infrastructure; and has enough onboard processing power to: (1) autonomously assess damage of energy infrastructure such as utility poles and substations using machine learning (ML), (2) geolocate detected features of interest, and (3) transmit relevant damage assessment results over constrained communications. Initial testing, based on collected imagery from before and after Hurricane Ida, attempted to detect utility poles and classify them as damaged or undamaged.

In this paper, the term UAS refers to a remote-controlled vehicle and its operator; the term unmanned aerial vehicle (UAV) refers to the vehicle only.

Related Work

UAS onboard processing is commonly used for flight control and autopilot. Onboard obstacle detection and avoidance (i.e., sense and avoid) research is ongoing (Yu and Zhang 2015; Fasano *et al.* 2016; Liu et at. 2019; Pedro *et al.* 2021; Rave 2021). Onboard simultaneous location and mapping to augment direct geo-positioning, operation in Global Navigation Satellite System (GNSS) denied environments, and avoidance of obstacles has also been a focus (Bershadsky and Johnson 2013; Bender *et al.* 2016, 2017; Rojas-Perez and Martinez-Carranza 2017). Edge processing has also been used to classify image sections for artificial intelligence (AI)–enhanced views aiming to increase operator comprehension (Speranza 2021).

Research has been moving toward onboard feature detection for years. Several feature detection algorithms were proposed, but edge implementation was left for future work (Gerke and Seibold 2014; Kuchhold *et al.* 2018; Nguyen *et al.* 2019; Castellano *et al.* 2020; Balamuralidhar *et al.* 2021; Hernandez *et al.* 2022; Koay *et al.* 2021). Real-time detection has been implemented on offboard devices (Redmon and Farhadi 2018; Wang *et al.* 2018; Harter *et al.* 2019; Jiang and Zhang 2019). Chen *et al.* (2022) proposed using large crewed aerial vehicles as local edge servers and gateways to cloud computing for off-loading edge computing tasks in drone swarms.

There are at least two published reports of UAS onboard feature detection. Jiang *et al.* (2022) reported UAS onboard feature detection of bridge cracks, spalling, and corrosion. Kraft *et al.* (2021) built a custom drone with an onboard global positioning unit (GPU) to automatically detect and geo-position features of interest (e.g., litter). Our solution extends these by addressing operation in constrained communication environments and beyond visual line of sight.

Geospatial Sciences and Human Security Division, Oak Ridge National Laboratory, 1 Bethel Valley Road, Oak Ridge, TN 37830 (hughesdc@ornl.gov).

Contributed by Tolga Bakirman, May 9, 2022 (sent for review July 15, 2022; reviewed by Jianyu Gu).

Photogrammetric Engineering & Remote Sensing
Vol. 89, No. 2, February 2023, pp. 79–87.
0099-1112/22/79–87

doi: 10.14358/PERS.22-00087R2

Nguyen *et al.* (2019) performed work relevant to energy infrastructure. They trained ML models to detect missing utility pole caps, cracked poles, cracked cross arms, rot-damaged cross arms, and woodpecker-damaged poles. Their work differs from ours in two important ways. First, it was not implemented on-the-edge, in restrained communication, or beyond visual line of sight. Second, the image resolutions differed by orders of magnitude because Nguyen *et al.* were looking for small features (i.e., cracks and rot), while we were looking for broken and leaning poles. Hence, direct comparisons of the network performance are not useful. Other efforts to use UAS in disaster response (e.g., floods) stop short of actual edge computing (Koay *et al.* 2021; Hernandez *et al.* 2022).

Methods

We performed a market survey of commercial UAS market in search of a vehicle with suitable onboard processing power to conduct real-time imagery analysis, sufficient flight time to collect data in a real-world scenario, and an open-software architecture conducive to the research environment. The Association of Unmanned Vehicle Systems' International comprehensive database of over 2000 commercial-off-the-shelf (COTS) aircraft was consulted, and we found listed no aircraft that satisfied all three requirements. Therefore, a modified aircraft was required. We chose the TurboAce Matrix quadcopter (Figure 1) because it can be easily modified to meet the specific needs of this project.

Airframe

The prototype aircraft (Figure 2) was named "Cornelia" after the first wife of the Roman emperor Julius Caesar. Modifications include the addition of the PixC4 compute module, camera, and the Multimodal Autonomous Vehicle Network (MAVNet) cellular command and control software (see *Command and Control Link* for additional details). The resulting aircraft has the following performance specifications:

- Weight: 3895 g (with battery and payload)
- Wingspan: ~44 inches
- Battery: 6S (24 V); 10 000 *mAh* lithium polymer
- Flight time: 25 minutes
- Range: 4–5 miles

Figure 1. TurboAce Matrix, a commercial-off-the-shelf unmanned aerial vehicle. (Source: TurboAce.com)

The UAV market has begun signaling increased interest in onboard computer power and advanced communications, which may be applicable for future iterations of this technology. For example, the Parrot Anafi AI has an onboard GPU and cellular communications and has recently become commercially available. The Anafi AI may closely replicate the capabilities of our prototype, but there are still unanswered questions regarding the computational power available for onboard processing.

Onboard Processors

We used the ORNL-developed PixC4-Jetson, which is a flight controller and co-processor carrier board to host the popular Nvidia Jetson NX series of companion computers. Figure 3 shows a schematic of the associated hardware. The PixC4-Jetson provides a lightweight, compact, and robust means of integrating the companion computer with the flight controller. This integration provides the co-processor with the position and orientation data required to geo-locate acquired imagery. The companion computer hosts the Docker ML algorithms used in this project while simultaneously performing real-time onboard processing. Figure 4 shows the PixC4 module with Nvidia Jetson co-processor.

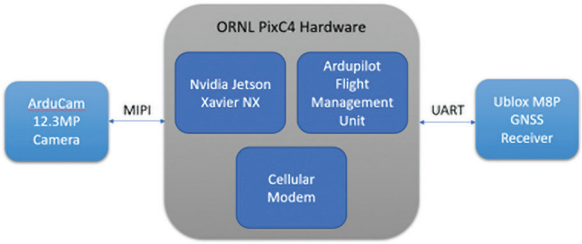


Figure 3. PixC4-Jetson carrier board hardware architecture. GNSS = Global Navigation Satellite System.

Figure 4. PixC4 with Nvidia Jetson Xavier NX.

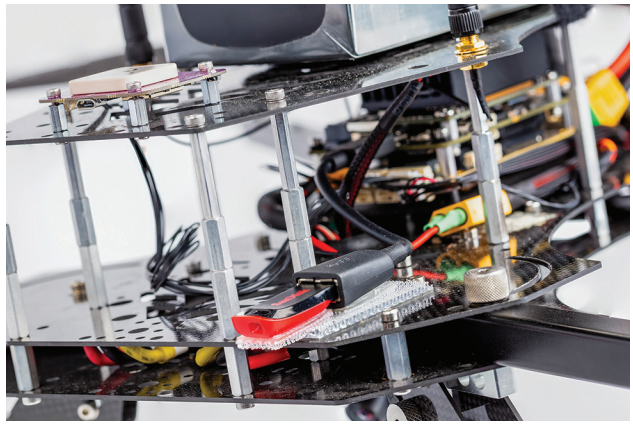


Figure 2. Modified aircraft "Cornelia" prior to test flight. (Source: Carlos Jones, ORNL)

Command and Control Link

We used the MAVNet system for global command and control of the UAV. MAVNet is a suite of hardware and software technologies developed for enabling next-generation unmanned systems research. The MAVNet communication system maintains communication links by allowing mesh networks, cellular/5G, and Iridium satellite communication modems to coexist, automatically falling forwards or backwards among them. This communication paradigm allows real-time operation in remote, austere environments such as those requiring disaster response (Harter *et al.* 2019).

MAVNet includes a web-based ground control system (GCS) that serves as the operator's primary interface with the aircraft. It displays real-time data sent from the UAV and enables complete control of the UAV and payloads from a single interface. For the purposes of this research, the GCS interface was modified to support hosting imagery collected by the Cornelia aircraft in real time. As detections are made, annotations are displayed in real time on the operator's interface, allowing the pilot immediate access to the data.

The GCS interface, shown in Figure 5, was customized to support displaying ML-detected features as icons. These icons have attributes for detection confidence, feature type, and optional image thumbnails. The interface can display up to 2000 icons simultaneously. For ease, an annotation manager was created to find, filter, and delete the icons. As shown in the figure, the aircraft's location is represented as the blue "delta" symbol. Telemetry data is shown in real time in the panel on the left side of the screen. The green icons display the detected location of undamaged utility poles collected during a test flight of the system. The operator can click on the icons to access more information (e.g., detection confidence) and full-resolution images collected by the aircraft. In the screen capture, the quadcopter is returning to a landing spot ("D") after having flown a short mission over some known utility poles in a test location. The black dots indicate where pictures were taken, and the green pole icons indicate where each pole was detected and its condition.

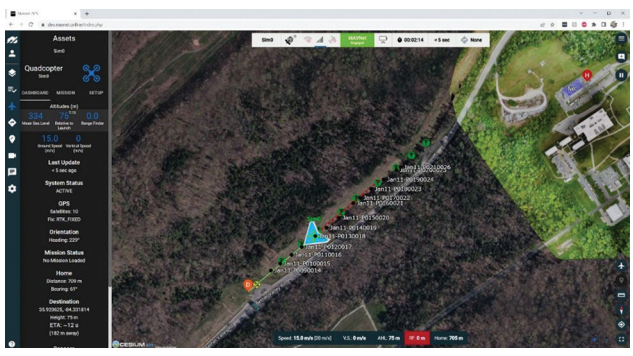


Figure 5. MAVNet GCS with sample detections.

Compute

The software running on the Nvidia Jetson Xavier NX companion computer is a version of Linux with everything not essential for the project removed. The primary applications are:

- MavnetProxy, a C#-based application developed in house that interfaces between the MAVNet server and the flight management unit. Specifically, it transfers raw images to ML Docker and then retrieves, stores, and uploads processed images and results.
- NVGSTCapture/GStreamer for camera control and image acquisition.
- Docker (docker-ce version 19.03), the host for the machine learning algorithms.
- Ardupilot (branch COPTER-4.1.2), the flight management software.
- Jetson Xavier Jetpack version 4.6 for the NVIDIA GPU.

Geo-Positioning Algorithm Development

We implemented algorithms for Cornelia and Anafi AI to translate pixel position to geo-location and upload results to MAVNet. The image-to-ground (I2G) operations is an inversion of the ground-to-image (G2I) operation. G2I is closed form, but it results in the loss of dimensionality from a three-dimensional (3D) ground point to a two-dimensional (2D) image point. The inversion, I2G, cannot recover the dimension lost in the projection. Instead of recovering the 3D point input, I2G returns the line that contains the point. This can be described as range ambiguity, meaning we know the direction toward the ground point but not how far away it is. Range ambiguity can be resolved if two (or more) images are used to observe the same 3D ground point. Barring degenerate geometry, their I2G operations will result in multiple lines whose intersection estimates the 3D ground point.

In this work, the simplest way to recover the lost range was to assume a ground plane. To describe the operation more formally, we define three coordinate systems: pixel space, image space, and ground space.

Pixel space is a 2D coordinate system, (u, v), defined with respect to an image in default landscape orientation. The origin is the top-left corner of the top-left pixel. The positive x-direction is from image left to image right. The positive y-direction is from image top to image bottom. The units are in pixels.

The image coordinate system is defined with respect to the image perspective center and focal length. The origin is at the image principal point (i.e., the intersection of the image's optical axis with the image plane). The positive x-direction is from image left to image right. The positive y-direction is from image bottom to image top. The negative z-direction (to keep with a right-handed coordinate system) is the image look direction from the perspective center out through the principal point. Image coordinates, $x = x \;\; y \;\; z^T$, have only two degrees of freedom because $z = -f$ is constant, where f is the focal length of the camera. The units are the same as the focal length (typically mm or pixels). The GCS can be any right-handed 3D Cartesian system.

The transformation from pixel space to image space can be complicated. The camera matrix (Mugneir *et al.* 2013) captures all linear transformations including scaling to focal length units, correcting for non-square pixels and skew between the image axes. The transformation is commonly extended to include nonlinear distortions (Fryer and Fraser 1986; Claus and Fitzgibbon 2005; Remondino and Fraser 2006). Camera calibration, the process of computing these parameters, is not required. This is acceptable because the accuracy requirements are loose, and image space errors are an insignificant contributor to overall error (as will be shown). It is also unavoidable because of the intended mass deployment to nontechnical users. Hence, it is assumed that pixels are square, axes are perpendicular, the principal point is in the center of the image, and the manufacturer's published focal length is reliable and can be converted to pixel units. This simplifies the transformation to $x = u - w/2$, $v = -(v - h/2)$, and $z = -f$, where w and h are the respective width and height in pixels and f is also in pixels.

The image position, $O = X_c \;\; Y_c \;\; Z_c^{\mathrm{T}}$ and orientation define the transformation from ground space to image space. Orientation has three degrees of freedom and is expressed in a variety of ways. Roll, pitch, and heading; Euler angles; quaternions; and 3 × 3 orthogonal rotation matrices are common examples. The rotation matrix, M, is the form used in the formulas below. M rotates from the GCS to the image coordinate system.

The G2I transformation can be written as $x = \lambda MX - O$, where $X = X \;\; Y \;\; Z^T$ is a ground point and $\lambda = -f/(M_{2,0}(X - X_c) + M_{2,1}(Y - Y_c) + M_{2,2}(Z - Z_c))$ is the scale that moves x into the focal plane. The inverse, I2G, can thus be written as $O + \lambda^{-1}M^T x = X$. Note that λ is the range along the line defined by the camera center O and look vector $M^T x$ to the ground point, X. λ is unrecoverable in the inversion because it is a function of the unknown X. However, the ray $r = O + \rho M^T x$ with $\rho > 0$ contains X. Our I2G operation resolves ρ by intersecting r with the approximate ground plane, $Z = Z_g$:

$$O + M^T x(Z_g - 0_z)/(M^T x)_z = Z, \tag{1}$$

where 0_z is the z-component of the camera position and $(M^T x)_z$ is the z-component of the look vector $M^T x$. This is reasonable for the first intended applications in flat, coastal, hurricane-damaged areas.

Figure 6. Four example views from Cornelia's circular calibration flight.

Cornelia's real-time telemetry data includes airframe position, yaw (α), roll (β), and pitch (γ). The constant rotation between the airframe and camera had to be calibrated separately. This is modeled as $x = \lambda M_c M_A X - O$. $M_A = f(\alpha, \beta, \gamma)$ is the time-variant rotation from the ground to airframe systems provided by the real-time telemetry. M_c is the constant rotation from airframe to camera systems; $M = M_c M_A$, and λ is the same scale factor.

The calibration flight was roughly circular with convergent look vectors (Figure 6). The circular flight sampled the yaw domain. The domains of pitch and roll were not widely sampled, but accelerations from the circular flight and buffeting from winds caused some variation. Pitch was in the range [–5°, 11°], and roll in the range [–3°, 7°].

The dirt and short cut weeds of the field feature matched well in the open-source structure-from-motion software COLMAP (Figure 6). The 18 images had an average of 190.5 feature matches. Each ground point was measured in an average of 3.6 images. The root-mean-square error of the image measurements was 0.82 pixels. The rotation matrix of each image, M_i, was an output of the photo triangulation. M_c was computed to minimize the residual errors ($v_{\alpha,i}$, $v_{\beta,i}$, $v_{\alpha,\gamma}$) in the telemetry, $M_c = M^T_{A,i} M_i$, where $M^T_{A,i} = f(\alpha_i + v_{\alpha,i}, \beta_i + v_{\beta,i}, \gamma_i + v_{\gamma,i})$. The residuals were weighted equally. In this model, M_i was considered errorless. This is a reasonable approach given the triangulated rotations were orders of magnitude more precise than the telemetry.

The I2G transformation (Equation 1) inputs are the six telemetry parameters, the image point, and the ground elevation estimate. Uncertainty estimates for each of these inputs is required to model the uncertainty of the computed ground location. Covariance of the telemetry parameters was estimated from their residuals in the calibration flight. The image location of the features is measured at the bottom of the ML bounding box, and the standard error was estimated at five pixels (for both x- and y-axes). The standard error of the ground elevation was modeled as 3 m. All estimated errors were reported at the one σ level.

The uncertainty of the computed ground location varies across the image (Figure 6). For ease, Figure 7 uses normalized fiducial coordinates ($\hat{x} \in [-1,1]$ and $\hat{y} \in [-1,1]$). The dominant signal is an exponential increase in uncertainty from the bottom ($\hat{y} = -1$) to the top of the image ($\hat{y} = 1$) . The left ($\hat{x} = -1$) and right ($\hat{x} = 1$) sides of the image are not symmetrical because the cameras were mounted with a small roll that tilts the right side of the image up.

Figure 8 shows the contribution of different sources of error. The curves are from the center vertical of the image. The top curve (solid black) is the full error propagation model. The remaining curves show the predicted accuracy when different input categories are modeled as errorless. Modeling the pixel measurements as errorless had a negligible effect on the estimated error. Modeling the camera position as errorless reduced the estimated uncertainty by ~50% in the bottom two-thirds of the field of view (FOV). The improvement falls off to 4% at the top of the image. Modeling the rotations as errorless reduced the error estimate by 45% at the bottom half of the FOV. This increased to 95% at the top of the image.

I2G accuracy for the Parrot Anafi AI used to collect data before, during, and after Hurricane Ida was analyzed similarly. Figure 9 compares the I2G accuracy of the Parrot Anafi and the Cornelia. The Parrot complete model is more accurate in the bottom tenth of the image, but it degraded faster through the bottom two-thirds of the FOV. This was driven mostly by larger systematic errors in the heading. In the top-third of the FOV, the Parrot Anafi's relative accuracy improves because the Cornelia was pitched 10.6° further from nadir.

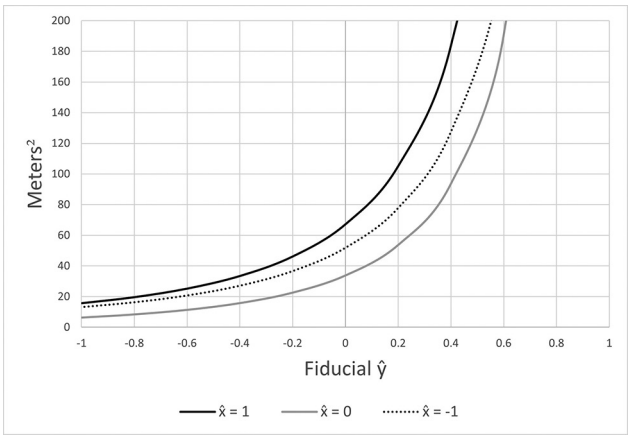


Figure 7. Area of the Cornelia I2G error ellipses for three vertical image lines (one σ).

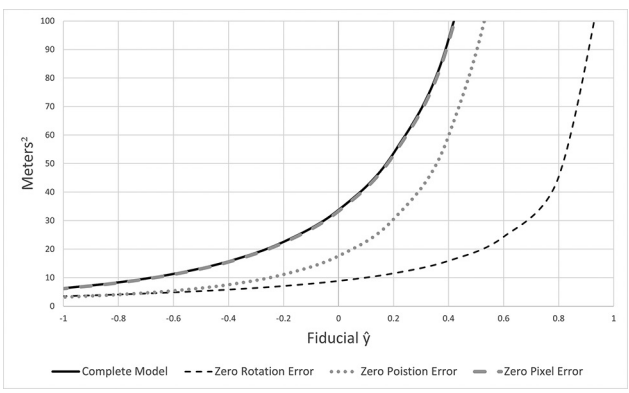


Figure 8. Analysis of error sources by comparison of the complete I2G uncertainty model to models assuming zero error in different input categories.

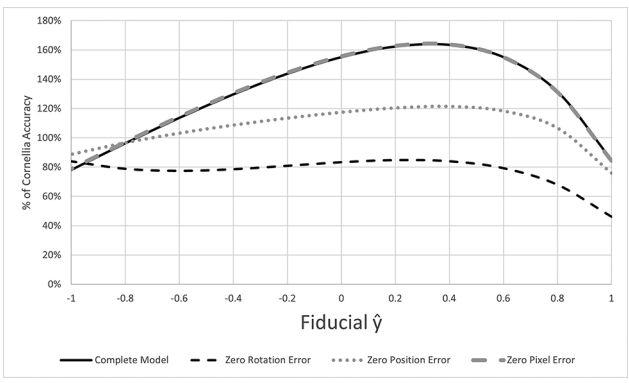


Figure 9. Comparison of uncertainty in the Parrot I2G and Cornelia models.

Machine Learning Algorithm Development

YOLOv5 (Jocher *et al.* 2022) single-stage object detection architecture was selected for utility pole detection and damage assessment. YOLOv5 had superior performance and flexibility to scale the model for real-time results on the UAS edge device.

YOLOv5 can be trained with different model sizes, which vary in their number of layers and parameters. Larger models offer superior object detection performance at the cost of requiring more computational and storage resources, which are both limited when dealing with edge devices. We selected three model sizes (labeled YOLOv5 Small, YOLOv5 Medium, and YOLOv5 Large) to give a sense of how the trade-offs between model size and performance impact our application. Generally, YOLOv5 adapts the architecture of CSPDarknet53 (Bochkovskiy *et al.* 2020) with a Spatial Pyramid Pooling layer as a backbone, PANet as a Neck, and a YOLO detection head (Redmon *et al.* 2016). To further optimize the overall performance, a bag of freebies and specials (Bochkovskiy *et al.* 2020) were incorporated. Because YOLOv5 is the most notable and convenient one-stage detector, we selected it as our base object detection module.

Summary of the Image Collection and Processing Framework

We illustrate the main components of the proposed framework in Figure 10, where the critical steps of MAVNet, georeferencing, and ML processing are provided in detail.

1. At even increments of ground movement (nominally 25 m), the Flight Management Unit (FMU) sends a message to the MavnetProxy application.
2. MavnetProxy recognizes the message and retrieves the most recent inertial measurement unit (IMU)/global positioning system (GPS)/Orientation messages. These messages are continually sent by the FMU to the application at ~5 Hz.
3. MavnetProxy then commands the camera driver to capture the current image buffer and saves the JPEG file with current date/time. The camera is always "on", so this is practically instantaneous.
4. MavnetProxy next creates a new "job" entry in the Workflow database with the image filename, and all relative information from the FMU (acquired in Step 2).
5. The workflow engine serializes each image as a multipart job.

Part 1—Process the Image

a. Copy file into shared filesystem area.
b. Send command to Docker to start the ML application and pass in the filename. (*Machine Learning Algorithm Development*)
c. Wait for ML application to return a status code (OK or error).
d. Copy (from the shared filesystem) the "labels" and processed image files. The "labels" file is a CSV text file with each infrastructure item described by a sperate line. Each line contains X/Y pixel location, status, and confidence of each item located in the image. The processed image file is the input file with bounding rectangles around the objects. The image file returned is the same size as the input file.
e. Clean up the "shared" filesystem in preparation for the next job.

Part 2—Process the Labels File

f. For each detected feature, compute a ground location (*Geo-Positioning Algorithm Development*)
g. Upload the ground location to MAVNet with the feature type, damage rating, and confidence. These are also displayed as icons in the MAVNet user interface.

Part 3—Apply Quadcopters GPS/IMU Data as EXIF Metadata to the Image Files

h. Add EXIF data to the raw and processed images with the GPS, IMU, and Orientation.

Part 4—Store all Data on a Removable USB Drive for further/after-flight analysis.

Part 5—Upload Imagery to MAVNet if a high-speed network is available.

i. Send either image (usually the processed) to MAVNet over the cellular network. This step is queued for later if the quadcopter is operating only on an Iridium satellite connection.

Results and Discussion

This section discusses results of this work as pertains to hardware, geo-positioning, and ML.

Hardware

Hurricane Ida was the only hurricane in the 2021 season that was available for collection of damaged infrastructure imagery needed for this work. We collected images with the Anafi Parrot USA quadcopter and post-processed them because the Cornelia platform was not operational at the time of the hurricane. A total of 334 images were curated and used for ML training and validation and algorithm testing. Imagery was collected from the Oak Ridge Reservation and two locations in neighboring Knox County, Tennessee, due to ease of data collection and curation.

The edge processing UAS prototype was developed to understand the future needs for state and local governments, who typically assess damage following natural disasters, to implement the ML and photogrammetry onboard hardware and analysis required. MAVNet is designed to enable next-generation applications for small autonomous systems. MAVNet's hardware component, PixC4-Jetson, is mated to a COTS quadcopter. A specially designed microcontroller-based vehicle controller with an integrated GPU-accelerated companion computer is onboard the platform. This hardware is compatible with a wide variety

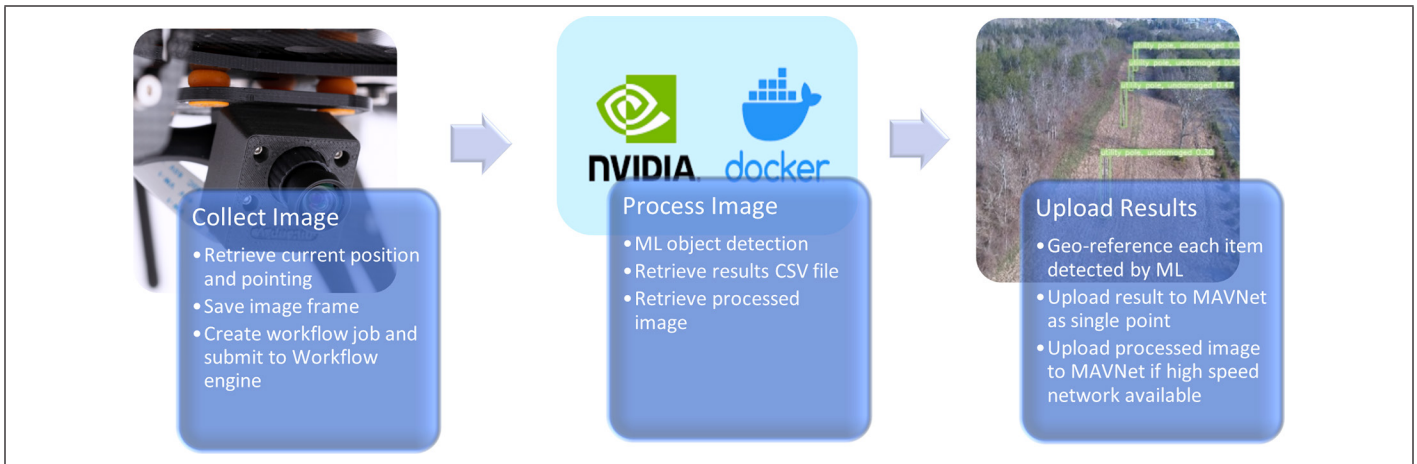


Figure 10. Image collection and processing flowchart. GPS = global positioning system; IMU = inertial measurement unit; ML = machine learning.

of commercial and hobby-level platforms. The hardware collects imagery from an onboard camera and immediately begins processing the data onboard, seeking to identify damaged utility poles and other infrastructure. In just seconds, the resulting data products are uploaded to the MAVNet cloud server and distributed to users anywhere in the world. In contrast, the stages of current systems involve manually creating flight.

Geo-Positioning

The I2G error estimate increases exponentially toward the top of the image. The error propagation model predicts that limiting the detected features to the bottom 60% of the image will bound the predicted semimajor error ellipsoid axis length at 8 m (one σ level). The model predicts that better direct georeferencing hardware (e.g., real-time kinematic GNSS) would reduce the estimated uncertainty by ~50% in the bottom of the image. In the top portion of the image, the rotation uncertainty becomes the dominant source of error. Higher-accuracy direct orientation measurements will make more of the top of the image usable. However, using the whole FOV will likely require a more nadir-looking pitch to manage the exponentially increasing error.

Machine Learning

Training an effective model requires high-quality imagery that captures the variety of object types, scenery, and lighting conditions a deployed system could be expected to encounter. Images of undamaged utility poles can be readily collected. However, we were able to collect only one set of post-disaster images (from Hurricane Ida). Table 1 and Table 2 provides the number of images and labels collected and the split between training and testing.

The Ida images proved inadequate to train a robust detector. To compensate for the scarcity of post-disaster images, we used two-stage model training and a combination of pre- and post-disaster imagery. First, the model was trained to detect utility poles. Second, the model was fine-tuned to distinguish between damaged and undamaged utility poles. Stage 2 used the fine-tune strategy (Li and Hoiem 2017) to update the extracted features' damaged status. This transfer learning approach has been shown to be useful to increase model generalizability under data-scarcity scenarios.

For comparisons, we repeated the same two-stage process with three different model sizes (various numbers of model parameters and layers in YOLOv5) using the combined data set, as well a single-stage model using only the damaged pole images from Hurricane Ida (see Table 3).

In Table 3 we used precision, recall, F-score, and mean Average Precision (mAP) metrics to evaluate utility pole damage assessment detection results. We framed this detection problem as a binary classification problem, for which all detected bounding boxes were either damaged pole or undamaged pole.

The imbalanced sample distribution and fewer number of labels in the Ida data set, as shown in the Table 1 and Table 2, may have contributed to the lower precision, recall and mAP. With the two-stage model training (training and transfer learning), we can observe the consistent improvements over the single-stage model (Table 3). For example, the YOLOv5 two-stage approach improved the F1-score from 0.603 to 0.652 and the mAP from 0.537 to 0.581.

Assuming a nominal flying speed of 5 m/s and image capture separation of 25m, there is ~5 seconds of available processing time between images. YOLOv5 processing time varies from 1.03 to 2.7 seconds per image for the small to large models (Table 4). Hence, there is flexibility to maintain real-time processing using the larger model, flying faster, and/or capturing image more frequently.

We showed an example of predictions from three YOLOv5 models in Figure 11, where the predictions are consistent across three models for this given image. In Figure 12, examples of common issues and incorrect detections were listed, where one utility pole might have been captured by more than one predicted bounding box predictions with conflicting assigned labels (lower middle in Figure 12c) and the utility pole on the ground was classified as undamaged because of its orientation, making it visually similar to those undamaged utility poles nearby (lower right in Figure 12b and 12c).

We also compared the results from the three YOLOv5 models to those obtained from YOLOv3 (Redmon and Farhadi 2018) and older but popular object detection frameworks. YOLOv5 models yielded better F-scores and mAP for one- and two-stage models. The only exception was a worse mAP (0.371 versus 0.403) for the one-stage YOLOv5 Small model. However, YOLOv5 Small model was twice as fast as the YOLOv3 analog (Table 4), which may be a reasonable trade-off for edge-computing applications.

We saw improvements in model performance as new data was incorporated into the training set, and we anticipate performance gains as more post disaster imagery is acquired. Therefore, the feature detection will improve as we iterate on our object detection model. We are also investigating few-shot learning (Bowman and Yang 2021) and model optimization methods to improve overall model performance with limited quantities of training data.

Table 1. Number of images from the three data collection locations, training/validation/testing splits, and contribution to the combined data set.

Location	Total Images	Training	Validation	Testing
Middlebrook Pike	257	182	50	25
Northshore	484	340	96	48
Hurricane Ida	334	235	66	33
Combined	1075	757	212	106

Table 2. Number of labeled utility poles in the Ida data set by condition. Undamaged utility poles account for 78% of the objects in the data set, and damaged utility poles account for 22%.

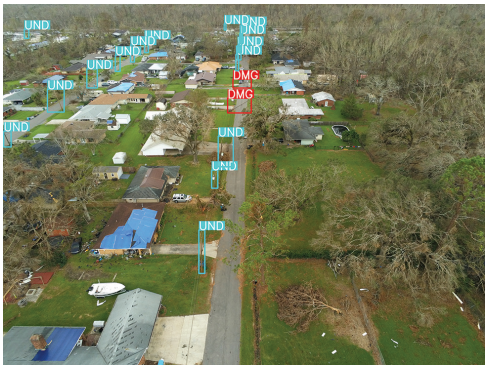
Feature of Interest	Ida Training	Ida Validation	Ida Testing	Ida Total
Undamaged pole	2063	458	329	2850
Damaged pole	504	168	112	784

Table 3. Testing results for one-stage YOLOv3 and YOLOv5 training on Hurricane Ida data and two-stage training on the combined data set (Middlebrook, Northshore, Ida). The bolded numbers indicate the best performing metrics.

Data	Model size	Precision	Recall	F-score	mAP
YOLOv5					
One-stage	Large	0.811	0.48	0.603	0.537
	Medium	0.852	0.423	0.565	0.475
	Small	0.778	0.333	0.466	0.371
Two-stage	Large	0.885	0.516	0.652	0.581
	Medium	0.881	0.488	0.628	0.552
	Small	0.855	0.411	0.555	0.480
YOLOv3					
One-stage	Standard	0.303	0.439	0.358	0.403
Two-stage	Standard	0.422	0.456	0.471	0.438

Table 4. Object detection model size and inferencing speed.

Framework	Number of Parameters	Inferencing Speed (Seconds per Image)
YOLOv5 Large	46 149 064	2.7
YOLOv5 Medium	20 879 400	1.35
YOLOv5 Small	7 027 720	0.48
YOLOv3	61 508 200	1.03



(a) Input image and labels

(b) Predictions from YOLOv5 Large

(c) Predictions from YOLOv5 Medium

(d) Predictions from YOLOv5 Small

Figure 11. Examples of the predictions from three YOLOv5 models. DMG in red boxes: damaged utility pole; UND in cyan boxes: undamaged utility pole.

(a) Input image and labels

(b) Predictions from YOLOv5 Large

(c) Predictions from YOLOv5 Medium

(d) Predictions from YOLOv5 Small

Figure 12. Examples of incorrect predictions and commonly seen failures. DMG in red boxes: damaged utility pole; UND in cyan boxes: undamaged utility pole.

The current processing workflow operates on still-image captures and functions within the capabilities of our current UAS prototype. We see advantages in moving toward processing the video stream. Edge processing would identify and track possible features of interest, providing multiple views. The feature detection performance offsets required for faster processing would ideally be offset by the redundancy. Additionally, multiple observations through a time series would allow I2G operations by photogrammetric intersection independent of ground surface priors. Perhaps more importantly, we could photogrammetrically estimate pole heights and lean to increase robustness of detection and classification.

We also plan improve the Cornelia quadcopter, geo-positioning, and ML implementation. Tasks include investigating more precise GNSS, exploring the addition of an altitude sensor, investigating the possibility of changing the camera to a more nadir viewpoint, and investigating few-shot learning strategies to improve the performance of the utility pole damage classification model. We will also work to identify COTS UAVs with capabilities similar to the Cornelia's so that the work can be more easily scaled.

Conclusion

In this work, we demonstrated our end-to-end deployable framework including prototype to detect, classify, and geo-locate utility poles on the edge. We have advanced the state of the art for this topic in two areas: (1) deployed, onboard UAS analysis that allows ML processing within size, weight, and power (SWaP) parameters and communications constraints and (2) an object detection and classification module that delivers satisfactory performance with a small set of labeled data. The proposed framework can be scaled to solve the post disaster energy infrastructure damage assessment problem. The developed onboard processing system and hardware are available for adoption and emulation by the US Civil Air Patrol and other organizations for real-time disaster assessment. We anticipate data collected from commercial, local, state, and federal entities to improve training, validation, and testing.

Acknowledgments

This material is based upon work supported by the US Department of Energy Office of Cybersecurity, Energy Security, and Emergency Response. The authors would also like to acknowledge the technical editing and document preparation assistance provided by Laurie Varma (ORNL). Notice: This manuscript has been authored by UT-Battelle LLC under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

References

Alam, F., M. Imran, and F. Ofli. 2017. Image4act: Online social media image processing for disaster response. *Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 601–604.

ArcGIS Online. 2022, April 6. Hurricane Ida final damage assessment public view. Accessed April 6, 2022. https://www.arcgis.com/home/item.html?id=48b4348dc6864bddaa15b233781bfea0.

Balamuralidhar, N., et al. 2021. MultEYE: Monitoring system for real-time vehicle detection, tracking and speed estimation from UAV imagery on edge-computing platforms. *Remote Sensing* 13(4): 573.

Bender, D., W. Koch, and D. Cremers. 2017. SLAM-Based Return to Take-Off Point for UAS. In *Multisensor Fusion and Integration for Intelligent Systems*. Springer, Cham, 2017.

Bender, D., F. Rouatbi, M. Schikora, D. Cremersy, and W. Koch, W. 2016, July. Scaling the world of monocular SLAM with INS-measurements for UAS navigation. In *2016 19th International Conference on Information Fusion (FUSION)* (pp. 1493–1500). IEEE.

Bershadsky, D., and Johnson, E. N. 2013. Indoor GPS-denied context based SLAM aided guidance for autonomous unmanned aerial systems. In *AIAA Guidance, Navigation, and Control (GNC) Conference* (p. 4989).

Boccardo, P., and F. G. Tonolo. 2015. Remote sensing role in emergency mapping for disaster response. In *Engineering Geology for Society and Territory*, vol. 5, Springer, Cham, pp. 17–24.

Bochkovskiy, A., C.-Y. Wang, and H.-Y. M. Liao. 2020. *YOLOv4: Optimal Speed and Accuracy of Object Detection*. ArXiv. Accessed June 3, 2022. https://arxiv.org/abs/2004.10934?sid=NDAqzT.

Bowman, J., and L. Yang. 2021. Few-shot learning for post-disaster structure damage assessment. In *Proceedings of the 4th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery (GEOAI '21)*. (pp. 27–32). Association for Computing Machinery.

Castellano, G., C. Castiello, C. Mencar, and G. Vessio. 2020. Crowd counting from unmanned aerial vehicles with fully-convolutional neural networks. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN 2020) (pp. 1–8). IEEE.

Chen, B., H. Zhou, J. Yao and H. Guan. 2022, March/April. RESERVE: An energy-efficient edge cloud architecture for intelligent multi-UAV. *IEEE Transactions on Services Computing* 15(2): 819–832.

Claus D., and A. W. Fitzgibbon. 2005, June. A rational function lens distribution model for general cameras. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (pp. 213–219).

G. Fasano, D. Accado, A. Moccia, and D. Moroney. 2016, November. Sense and avoid for unmanned aircraft systems. *IEEE Aerospace and Electronic Systems Magazine*, vol. 31, no. 11,31(11): 82–110.

Federal Emergency Management Agency. 2020, May. *Preliminary Damage Assessment Guide*. Accessed July 5, 2022. https://www.fema.gov/disaster/how-declared/preliminary-damage-assessments/guide.

Fryer J. G., and C. S. Fraser. 1986. On the calibration of underwater cameras. *The Photogrammetric Record* 12(67): 73–85.

Gerke, M., and P. Seibold, P. 2014. Visual inspection of power lines by U.A.S. In *Proceedings of the 2014 International Conference and Exposition on Electrical and Power Engineering.*

Harter, A., B. Vacaliuc, B. Stinson, A. Duncan, and C. Roberts. 2019, May. MAVNet: Design of a reliable beyond visual line of sight communication system for unmanned vehicles. In *AUVSI XPONENTIAL Conference Proceedings*, vol. 2019, no. 1.

Hernandez, D., et al. 2022. Flood detection using real-time image segmentation from unmanned aerial vehicles on edge-computing platform. *Remote Sensing* 14: 223.

Hodgson, M. E., B. A. Davis, Y. Cheng, and J. Miller. 2010. Modeling remote sensing satellite collection opportunity likelihood for hurricane disaster response. *Cartography and Geographic Information Science* 37(1): 7–15.

Jiang, S., and Zhang, J. 2019. Real-time crack assessment using deep neural networks with wall-climbing unmanned aerial system. *Computer-Aided Civil and Infrastructure Engineering* 35(6): 549–564.

Jiang, S., Y. Cheng, and J. Zhang. 2022, May. Vision-guided unmanned aerial system for rapid multiple-type damage detection and localization. *Structural Health Monitoring*. Jocher, G., et al. 2022, August. *YOLOv5 Classification Models, Apple M1, Reproducibility, ClearML and Deci.ai Integrations*. Zenodo.

Kerle, N., and C. Oppenheimer. 2002. Satellite remote sensing as a tool in lahar disaster management. *Disasters* 26(2): 140–160.

Koay, H. V., et al. 2021. YOLO-RTUAV: Towards real-time vehicle detection through aerial images with low-cost edge devices. *Remote Sensing* 13(21): 4196.

Kraft, M., et al. 2021. Autonomous, onboard vision-based trash and litter detection in low altitude aerial images collected by an unmanned aerial vehicle. *Remote Sensing* 13(5): 965.

Kuchhold, M., M. Simon, V. Eiselein, and T. Sikora. 2018. Scale-adaptive real-time crowd detection and counting for drone images. In *Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP 2018)* (pp. 943–947).

Laituri, M., and K. Kodrich. 2008. On line disaster response community: People as sensors of high magnitude disasters using internet GIS. Sensors 8.5: 3037–3055.

Li, Z., and D. Hoiem. 2017. Learning without forgetting. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40(12): 2935–2947.

K. Liu, S. Chauhan, R. Devaraj, S. Shahi, and U. Sreekumar. 2019. Enabling autonomous unmanned aerial systems via edge computing. In *2019 IEEE International Conference on Service-Oriented System Engineering (SOSE)* (pp. 374–3745).

Mugneir, C. J., et al. 2013. The mathematics of photogrammetry. In *The Manual of Photogrammetry* (6th ed.) (C. McGlone, ed.). American Society for Photogrammetry and Remote Sensing, p. 277.

Nguyen, V. N., R. Jenssen, and D. Roverso. 2019. Intelligent monitoring and inspection of power line components powered by UAVs and deep learning. In *IEEE Power and Energy Technology Systems Journal* 6(1): pp. 11–21. doi: 10.1109/JPETS.2018.2881429.

Pedro, D., et al. 2021. Collision avoidance on unmanned aerial vehicles using neural network pipelines and flow clustering techniques. *Remote Sensing* 13(13): 2643.

Rave, C. J. 2021. "Edge Processing of Image for UAS Sense and Avoidance." PhD Thesis, Wright State University.

Redmon, J, S. Divvala, R. Girshick, A. Farhadi. 2016, May. *You Only Look Once: Unified, real-time object detection*. Accessed September 6, 2022. https://arxiv.org/pdf/1506.02640.pdf.

Redmon, J., and A. Farhadi. 2018. *YOLOv3: An incremental improvement*. arXiv e-prints. Accessed September 6, 2022. https://arxiv.org/abs/1804.02767.

Remondino F., and C. Fraser. 2006. Digital camera calibration methods: Considerations and comparisons. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences* 36: 266–272.

Rojas-Perez, L. O., and J. Martinez-Carranza. 2017. "Metric Monocular SLAM and Colour Segmentation for Multiple Obstacle Avoidance in Autonomous Flight." Presented at *2017 Workshop on Research, Education and Development of Unmanned Aerial Systems (RED-UAS)*. IEEE.

Speranza, N. A. 2021. "Adaptive Two-Stage Edge-Centric Architecture for Deeply-Learned Embedded Real-Time Target Classification in Aerospace Sense-and-Avoidance Applications." PhD Thesis, Wright State University.

Wang, B., J. Xie, S. Li, Y. Wan, S. Fu, and K. Lu. 2018. Enabling high-performance onboard computing with virtualization for unmanned aerial systems. In *Proceedings of the 2018 International Conference on Unmanned Aircraft Systems*.

Yu, X., and Y. Zhang. Sense and avoid technologies with applications to unmanned aircraft systems: Review and prospects. *Progress in Aerospace Sciences* 74 (2015): 152–166.

IN-PRESS ARTICLES

Shuanggen Jin, Ayman M.Elameen, Daniel Olago. Identification of drought events in major basins of Africa based on weighted water storage deficit index from GRACE measurements.

Xin Jia, Qing Zhu, Xuming Ge, Ruifeng Ma, Daiwei Zhang, and Tao Liu. Robust Guardrail Instantiation and Trajectory Optimization of Complex Highways Based on Mobile Laser Scanning Point Clouds.

Zhikang Lin, Wei Liu, Yulong Wang, Yan Xu, Chaoyang Niu. Change Detection in SAR Images through Clustering Fusion Algorithm and Deep Neural Networks.

Dan Li, Hanjie Wu, Yujian Wang, Xiaojun Li, Fanqiang Kong, and Qiang Wang. Lightweight Parallel Octave Convolutional Neural Network for Hyperspectral Image Classification.

Tolga Bakirman, Bahadir Kulavuz, and Bulent Bayram. Use of Artificial Intelligence Toward Climate-Neutral Cultural Heritage.

Jian Wu, Shifeng Fu, Peng Chen, Qinghui Chen, Xiang Pan. Validation of Island 3D-mapping based on UAV spatial point cloud optimization: A case study in Dongluo Island of China.

Bo Xu, Daiwei Zhang, Han Hu, Qing Zhu, Qiang Wang, Xuming Ge, Min Chen, Yan Zhou. Spherical Hough Transform for Robust Line Detection toward a 2D-3D Integrated Mobile Mapping System.

Ruifeng Ma, Xuming Ge, Qing Zhu, Xin Jia, Min Chen, Liu Tao. Model-driven precise degradation analysis method of highway marking using Mobile Laser Scanning point clouds.

Yao Zhou, Ling Tan, Yue Ding. MCAFNet: Multi Channel Attention Fusion Network Based CNN for Remote Sensing Scene Classification.

Xinyan Pang, Na Ren, Changqing Zhu, Shuitao Guo, Ying Xiong. Blind and Robust Watermarking Algorithm for Remote Sensing Images Resistant to Geometric Attacks.

WHO'S WHO IN ASPRS

Founded in 1934, the American Society for Photogrammetry and Remote Sensing (ASPRS) is a scientific association serving thousands of professional members around the world. Our mission is to advance knowledge and improve understanding of mapping sciences to promote the responsible applications of photogrammetry, remote sensing, geographic information systems (GIS) and supporting technologies.

BOARD OF DIRECTORS

BOARD OFFICERS

President
Christopher Parrish, Ph.D
Oregon State University

President-Elect
Lorraine B. Amenda, PLS, CP
Towill, Inc.

Vice President
Bandana Kar
Oak Ridge National Lab

Past President
Jason M. Stoker, Ph.D,
U.S. Geological Survey

Treasurer
John McCombs
NOAA

Secretary
Harold Rempel
ESP Associates, Inc.

COUNCIL OFFICERS

ASPRS has six councils. To learn more, visit https://www.asprs.org/Councils.html.

Sustaining Members Council
Chair: Ryan Bowe
Deputy Chair: Melissa Martin

Technical Division Directors Council
Chair: Bill Swope
Deputy Chair: Hope Morgan

Standing Committee Chairs Council
Chair: David Stolarz
Deputy Chair: TBA

Early-Career Professionals Council
Chair: Madeline Stewart
Deputy Chair: Kyle Knapp

Region Officers Council
Chair: Demetrio Zourarakis
Deputy Chair: Jason Krueger

Student Advisory Council
Chair: Lauren McKinney-Wise
Deputy Chair: Oscar Duran

TECHNICAL DIVISION OFFICERS

ASPRS has seven professional divisions. To learn more, visit https://www.asprs.org/Divisions.html.

Geographic Information Systems Division
Director: Denise Theunissen
Assistant Director: Jin Lee

Lidar Division
Director: Ajit Sampath
Assistant Director: Mat Bethel

Photogrammetric Applications Division
Director: Ben Wilkinson
Assistant Director: Hank Theiss

Primary Data Acquisition Division
Director: Greg Stensaas
Assistant Director: Srini Dharmapuri

Professional Practice Division
Director: Bill Swope
Assistant Director: Hope Morgan

Remote Sensing Applications Division
Director: Amr Abd-Ehrahman
Assistant Director: Tao Liu

Unmanned Autonomous Systems (UAS)
Director: Jacob Lopez
Assistant Director: Bahram Salehi

REGION PRESIDENTS

ASPRS has 13 regions to serve the United States. To learn more, visit https://www.asprs.org/regions.html.

Alaska Region

Cascadia Region
Robert Hariston-Porter

Eastern Great Lakes Region
Michael Joos, CP, GISP

Florida Region
Xan Fredericks

Heartland Region
Whit Lynn

Intermountain Region
Robert T. Pack

Mid-South Region
David Hughes

Northeast Region

North Atlantic Region

Pacific Southwest Region
John Erickson, PLS, CP

Potomac Region
Dave Lasko

Rocky Mountain Region

Western Great Lakes Region
Adam Smith

Apricot Tree Detection from UAV-Images Using Mask R-CNN and U-Net

Firat Erdem, Nuri Erkin Ocer, Dilek Kucuk Matci, Gordana Kaplan, and Ugur Avdan

Abstract

Monitoring trees is necessary to manage and take inventory of forests, monitor plants in urban areas, distribute vegetation, monitor change, and establish sensitive and renewable agricultural systems. This study aims to automatically detect, count, and map apricot trees in an orthophoto, covering an area of approximately 48 ha on the ground surface using two different algorithms based on deep learning. Here, Mask region-based convolutional neural network (Mask R-CNN) and U-Net models were run together with a dilation operator to detect apricot trees in UAV images, and the performances of the models were compared. Results show that Mask R-CNN operated in this way performs better in tree detection, counting, and mapping tasks compared to U-Net. Mask R-CNN with the dilation operator achieved a precision of 98.7%, recall of 99.7%, F1 score of 99.1%, and intersection over union (IoU) of 74.8% for the test orthophoto. U-Net, on the other hand, has achieved a recall of 93.3%, precision of 97.2%, F1 score of 95.2%, and IoU of 58.3% when run with the dilation operator. Mask R-CNN was able to produce successful results in challenging areas. U-Net, on the other hand, showed a tendency to overlook existing trees rather than generate false alarms.

Introduction

Accurate detection and counting of trees are challenging tasks faced by agricultural producers, urban planners, foresters, ecologists, and insurers to make correct assessments on issues such as irrigation, maintenance, planning, value, and risk determination. By their nature, the working areas in these fields can be hundreds of hectares, and performing these tasks in such large areas with traditional methods is very time consuming, expensive, and prone to errors. Therefore, more advanced, economical, and accurate methods are needed.

Tree detection and counting through imagery have been carried out with different methods for decades. For example, to detect and count palm trees from high-resolution UAV imagery, Shafri *et al.* (2011) proposed a method consisting of spectral and texture analysis, edge extraction, segmentation, morphological analysis, and blob analysis stages that results in average detection accuracy of more than 95%. In another study, Santoso *et al.* (2016) used spectroscopic measurements to extract palm trees using QuickBird images. Traditional digital image processing techniques, such as local maximum (Mohan *et al.* 2017), marker-controlled watershed (Huang *et al.* 2018), Hough transform (Koc-San *et al.* 2018), histogram of oriented gradient features, and support vector machine classifier (Wang *et al.* 2019), were also used to detect and count trees in high-resolution images.

In recent years, as an attempt to simulate the learning behavior of the human brain for computers, deep learning has been used in many remote sensing applications, including tree extraction and counting using remote sensing imagery (Safonova *et al.* 2019; Santos *et al.* 2019). Especially, the convolutional neural network (CNN) has been recognized as one of the most successful and widely used deep learning approaches and has found a place for itself in tree detection applications (Chen *et al.* 2017; Li *et al.* 2018). Having a hierarchical architecture that can be trained using a large amount of data, CNN is able to perform object detection tasks. When CNNs are used together with region proposal networks in methods such as Faster R-CNN or Mask R-CNN, they produce fast, efficient, and highly accurate results in the field of object detection (Girshick 2015; He *et al.* 2017). The Mask region-based convolutional neural network (Mask R-CNN) was recently used to detect trees in urban, campus, and rural areas (Ocer *et al.* 2020; Lumnitz *et al.* 2021). Introduced in 2017 as an extension to the Faster R-CNN model, Mask R-CNN is capable of detecting objects in images and creating masks for them, which is named instance segmentation. Quoc *et al.* (2020) compared the performance of this model with U-Net for segmenting agricultural areas in satellite images, obtaining a mean average precision of 95.21% for Mask R-CNN and 92.69% for U-Net. First proposed for segmenting biomedical images, U-Net (Ronneberger *et al.* 2015) has been used for many different areas in the literature, such as cell counting (Falk *et al.* 2019), brain tumor detection (Dong *et al.* 2017), road detection (Yang *et al.* 2019), and tree detection (Zhao *et al.* 2018; Korznikov *et al.* 2021).

The speed performances and capacity limits of deep learning algorithms, almost all of which are run better on a graphics processing unit (GPU), are directly related to the memory and speed values of this hardware. There is a certain upper limit to the image size the GPU can process, depending on the model used. For example, when Matterport's Mask R-CNN implementation (Abdulla 2017) is used as the model, the maximum RGB image size that an Nvidia GTX 1080 Ti GPU with 11 GB of internal memory can handle is 4896×4896 pixels. For this reason, when object detection is to be done larger than the upper limit, it is necessary to split the image into parts or reduce the image size to sufficient dimensions (hence compromising on detail). In studies dealing with such cases (Ma *et al.* 2018; Gao *et al.* 2019; Zhang *et al.* 2019), the image is usually divided into small pieces that overlap to a certain extent, and then additional processes are applied to prevent repetitive object counting in the overlap regions and to determine the object locations correctly. However, as these studies are concerned with only object detection, the problem of the existence of multiple outputs for the same ground truth in the overlap regions is not addressed in terms of mapping tasks.

Encountered in cases where object detection is performed with deep learning–based instance segmentation methods, such as Mask R-CNN or U-Net, the generated masks do not fit correctly to the ground truth, as some artificial gaps between the masks and the output image edges are formed for the objects close to the edges, eventually leading to mask localization inaccuracies (Cheng *et al.* 2020; Kirillov *et al.* 2020). In cases where the dimensions of the image to be examined are larger than the capacity of the GPU used, splitting the image into parts for processing increases the total number of edges, eventually boosting incorrect segmentation and detection results in the regions close to the edges.

In this study, the detection, counting, and canopy mapping of apricot trees in a very high-resolution large orthophoto were conducted with the help of two instance segmentation models: Mask R-CNN and

Firat Erdem, Nuri Erkin Ocer, Dilek Kucuk Matci, Gordana Kaplan, and Ugur Avdan are with Eskisehir Technical University, Institute of Earth and Space Sciences, Eskisehir, Turkey (firaterdem@eskisehir.edu.tr).

Contributed by Tolga Bakirman, May 9, 2022 (sent for review July 4, 2022; reviewed by Weixun Zhou, Murat Uysal).

Photogrammetric Engineering & Remote Sensing
Vol. 89, No. 2, February 2023, pp. 89–96.
0099-1112/22/89–96

doi: 10.14358/PERS.22-00086R2

U-Net. The main objectives of this study are (1) to compare the results produced by these models and (2) to improve these results by alleviating the problem of mask localization inaccuracies in generated tree masks, leading to a vast amount of false alarms encountered in object detection and segmentation. As a solution to the latter, this article proposes to apply a dilation operator to the detection outputs. To make comparisons between model results and to prove the effectiveness of the suggested method, a very large apricot tree orchard in Malatya province was selected as the study area. Results are compared in terms of precision, recall, F1 score, and intersection over union (IoU) metrics.

Study Area

Four different study sites located in Malatya/Turkey were chosen to detect apricot trees (Figure 1). The geographic coordinates of the study sites were 38.2332° E–38.3986° N, 38.2441° E–38.4113° N, 37.6981° E–38.8596° N, and 38.3248° E–38.5074° N for Site 1 (12 ha), Site 2 (13 ha), Site 3 (64 ha), and Site 4 (48 ha), respectively. With an average annual precipitation of 574 mm, Malatya city has a warm and temperate climate. All study sites were orchards that contain apricot (*Prunus armeniaca*) trees. To increase the model's generalization capability, study sites with different characteristics were selected. The crowns of the trees in Site 1 were wider than in other study sites. The trees in Site 3, on the other hand, had a small number of leaves because they belong to the drying period. For this reason, trees in Site 3 were more difficult to distinguish from the background.

Materials and Methods

Data Acquisition

In all study sites, a flight was performed with a DJI Phantom 4 Pro UAV with RGB-integrated 20-megapixel camera, and aerial images were obtained. For all study sites, the flight altitude was chosen as 75 m and the front and side overlap as 80% and 70%, respectively. After the UAV flights, orthophoto production was performed using Agisoft Metashape software. As a result of this process, four orthophotos were obtained. Orthophotos of Sites 1–3 were used for model training, while the orthophoto of Site 4 was used for testing the model. The spatial resolutions of the orthophotos used for training were 2.07 cm for Sites 1 and 2, 1.84 cm for Sites 3 and 4.

To obtain label data, trees on the orthophoto images were manually annotated using ArcMap software, and polygons describing tree boundaries were produced (Figure 2). As a result, a total number of 1692 trees were annotated in orthophotos and used for training. In the orthophoto used for the test, a total of 957 trees were annotated. Orthophoto images and label images were cropped into pieces of 768×768 pixels to make the training of the deep learning model more efficient. Image subsets without trees were removed from the training data set.

Instance Segmentation Models Based on Deep Learning

Being a subset of machine learning, deep learning is a neural network with multiple layers. Similar to the learning behavior of the human brain, these networks use large amounts of data for a specific purpose

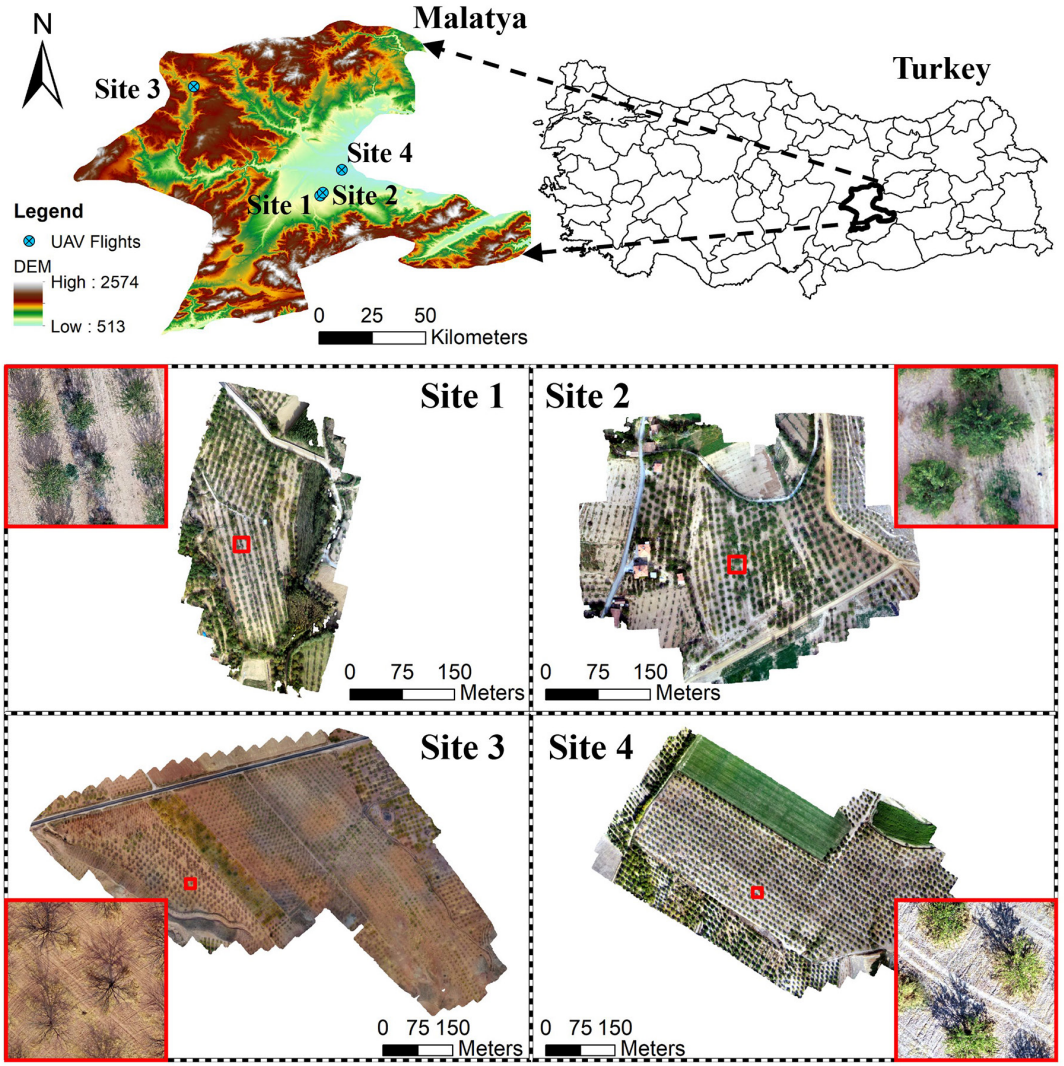


Figure 1. Location of the study sites and generated orthophotos.

Figure 2. Manually annotation of tree boundaries.

in a process to minimize the error. Among the key problems deep neural networks can solve is detecting objects in images. The computer vision task for detecting and localizing an object in an image is defined as instance segmentation. In this study, we use two different types of instance segmentation algorithms—Mask R-CNN and U-Net—for apricot tree detection. In the following parts, details about both methods will be given.

Mask R-CNN

Mask R-CNN is conceptually simple and flexible object detection and instance segmentation model (He *et al.* 2017). In addition to the existing branch for bounding box recognition, the approach extends the Faster R-CNN model by adding a branch to predict an object mask in parallel. Thus, the Mask R-CNN model produces high-quality segmentation masks for detected objects and classifies them. As given in the schematic diagram in Figure 3, Mask R-CNN consists of three main components: (1) a feature pyramid network module, which generates multi-scale feature maps with better-quality information than regular feature extractors by constructing a bottom-up pathway and a top-down pathway in order to extract the hidden features in the input image; (2) a region proposal network module, which uses the appropriate feature output map of the feature pyramid network to propose candidate object bounding boxes or regions of interest; and (3) a prediction and regression module, which predicts the class, the offset values for the bounding box, and the mask of the object located in the proposed region. In this study, in the test phase, RGB images of 4800×4800 pixels were input to the Mask R-CNN model, and binary masks of the same size were produced as output, where trees were represented with white and other regions with black.

U-Net

The U-Net architecture is a fully convolutional network, first designed for image segmentation in biomedicine. The results of the ISBI 2012 challenge for segmentation of neural structures in electron microscopic stacks showed that U-Net architecture can be trained using a small number of photos and exceeded the previous best approach (Ronneberger *et al.* 2015). It is divided into two parts: the contracting path and the expansive path. The contracting path is made up of recurrent 3×3 convolutions, rectified linear unit, and 2×2 max-pooling layers with stride 2. The number of feature channels is doubled after down-sampling. The expansive path is made up of up-sampling the feature map, 2×2 up-convolution, and concatenation with the equivalent feature map from the contracting path. To map each feature vector in the final layer, a 1×1 convolution is employed. The schematic diagram of the U-Net architecture is given in Figure 4. In this study, images of 768×768 pixels were input to the U-Net, and binary masks of the same size were produced as output, where trees were represented with white and other regions with black.

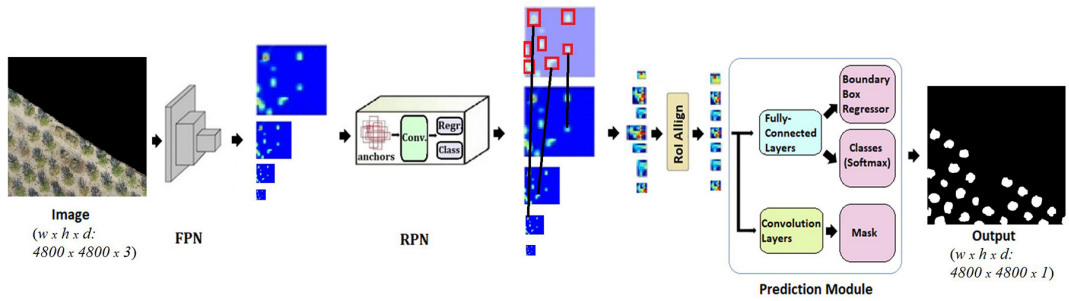


Figure 3. Architecture for Mask region-based convolutional neural network.

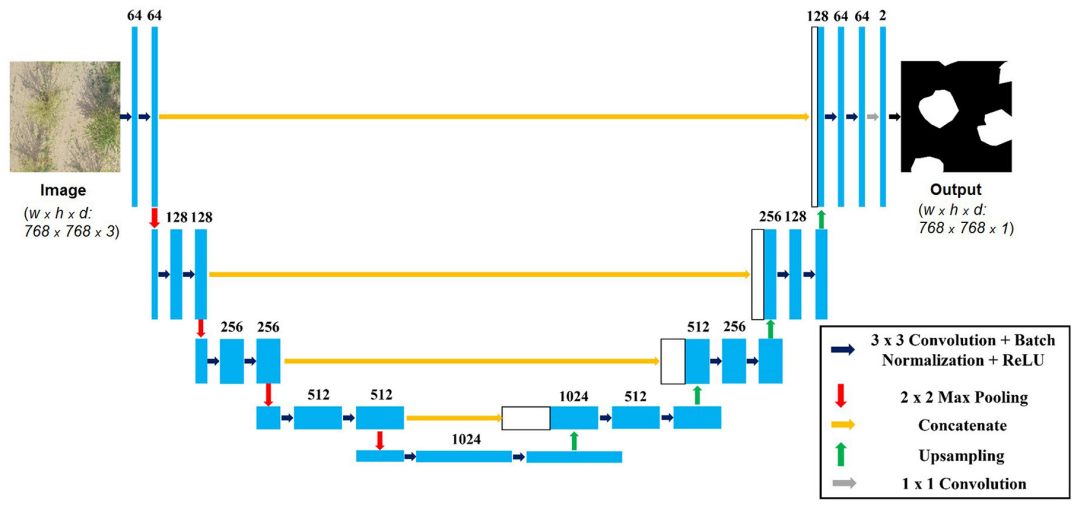


Figure 4. U-Net architecture.

Postprocessing: Dilation

The output map for the entire test orthophoto is produced by combining the binary mask outputs for each image subset. However, due to mask localization inaccuracies, some artificial gaps between the masks and the output image edges are formed, clearly visible at the intersections of small images (Figure 5). To avoid this problem, the morphologic dilation process was applied to the binary image. Erosion and dilation are the fundamental operators in mathematical morphology. When mathematical morphology is utilized in image processing, these operations are applied to an image with a collection of known shapes, called a structuring element. When the dilation operator is applied to an image, it produces an output image that illustrates where the structuring element hits the objects in the image (Benediktsson *et al.* 2003). It enlarges bright regions and shrinks dark regions. In this study, the shape of the structuring element was selected as a disk, and the radius of the disk-shaped footprint was set to 11 for outputs of Mask R-CNN and 7 for outputs of U-Net. These parameters were selected empirically. As a result of dilation operation, tree crowns are dilated, and edge mistakes were mitigated.

Work Flow for Apricot Tree Detection in Large-Scale UAV Images

The detection and segmentation of apricot trees in large-scale images are difficult. Due to hardware limitations, to perform these tasks with deep learning methods, the image size must be brought to a level that the GPU can handle by either cropping the image into parts or shrinking its dimensions. However, resizing the image will reduce details, making it hard to detect small targets. Therefore, it is necessary to examine the image by cropping in order not to lose detail and to analyze the image in its original resolution. So, as in our study (see the flowchart in Figure 6), the orthophoto is divided into patches that do not overlap each other, the Mask R-CNN and U-Net models are fed with these patches, and mask outputs for the trees are produced. However, due to mask localization inaccuracies near the output edges, it is not possible to make accurate detections for the entire orthophoto simply by putting these patches together. Mask localization inaccuracies near the image edges are encountered in deep learning instance segmentation models, where the masks of the objects close to the edges of the output image do not fit the ground truth, and artificial gaps are formed between the mask border and the edge. As in this study, splitting the image into many parts increases the number of edges, further exacerbating the edge problem. In order to alleviate this problem, this article proposes applying a morphology operator, such as dilation, to the detection outputs before putting them together. The binary output detection results for test orthophoto were obtained in four steps:

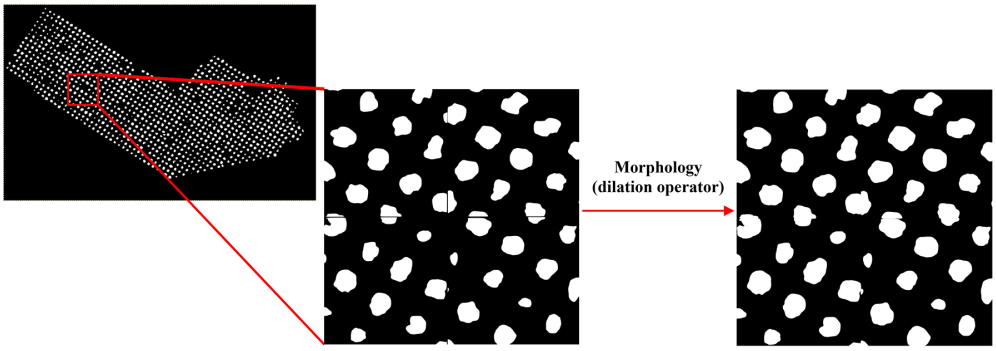


Figure 5. Postprocessing work flow.

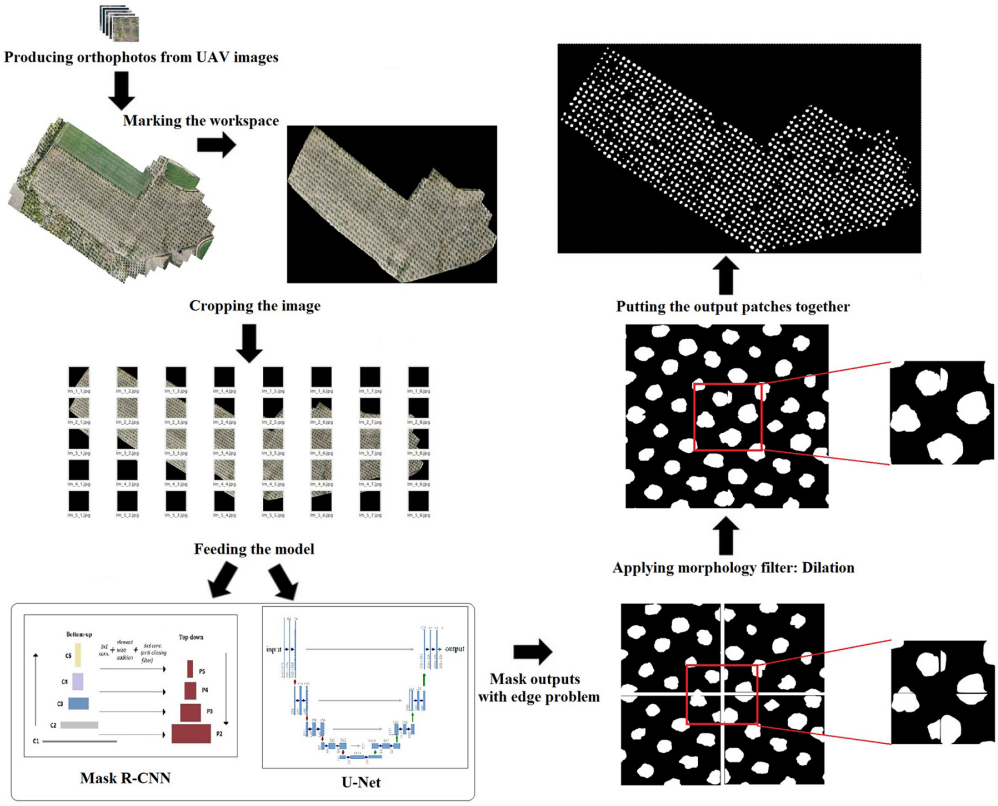


Figure 6. Flowchart of the study.

1. Cropping the orthophoto into patches that do not overlap each other (4800×4800 pixels for Mask R-CNN, 768×768 for U-Net)
2. Feeding the cropped images into the models
3. Applying dilation operator to the model outputs
4. Putting the filtered outputs together

Accuracy Assessment Method

The accuracy assessments of the detection results in this study were performed using the F1 score, precision, recall, and IoU metrics. Manual annotation was performed in order to use it as reference data in accuracy assessment. F1 score, precision, and recall were calculated in an object-based manner. That is, a generated mask is approved as a true positive if it overlaps more than half of the ground truth; otherwise, it is marked as a false positive. Precision is the ratio of the number of correctly predicted positive results to the number of all positive predictions, and recall is the ratio of number of correctly predicted positive results to the number of total positive ground truths. The F1 score is the weighted average of precision and recall. F1 score, precision, and recall are calculated using steps 1 to 3 listed below, respectively (Sokolova *et al.* 2006):

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

$$\text{F1 score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

where TP, FP, TN, and FN are the total numbers of true positives, false positives, true negatives, and false negatives, respectively.

The IoU score has been calculated on the basis of pixels to evaluate segmentation performances. IoU is the crossover rate between the prediction and the ground-truth boundaries, thus the ratio of their intersection to union (Nowozin 2014). It can be calculated using

$$\text{IoU} = \frac{\text{Prediction} \cap \text{Ground Truth}}{\text{Prediction} \cup \text{Ground Truth}} \tag{4}$$

Experimental Setup

The experiments related to Mask R-CNN were carried out using the framework introduced by Abdulla (2017), which uses the Keras and Tensorflow libraries. The U-Net model was also created using the Keras library. Transfer learning was carried out using the weights produced by Ocer *et al.* (2020) as initial weights for both models. Data augmentation techniques (random scaling, rotation, vertical/horizontal flipping, contrast/brightness alteration, and Gaussian blurring) were applied to increase the amount of data. Stochastic gradient descent with momentum is employed to optimize the Mask R-CNN model, with a weight decay of 0.0001 and a momentum of 0.9. The initial learning rate was determined as 0.002 and gradually increased to 0.01. On the other hand, to optimize the U-Net model, the Adam optimization algorithm was used, and the initial learning rate was set to 0.0005. The number of epochs was determined as 100 for both models. The hardware configurations were an Intel i7 6850 K central processing unit with an Nvidia GTX 1080 Ti GPU for training and testing of the models. Training of both models was carried out with images of 768×768 pixels. The Mask R-CNN framework produced instance segmentation results in run length encoding format, the results of which were converted to binary image format using a Python script.

Mask R-CNN and U-Net models were also compared with the machine learning–based classification and regression tree (CART) method (Lewis 2000). CART models use a collection of if-then (split) conditions called tree-building algorithms for classification (Razi and Athappilly 2005). The algorithm was implemented in eCognition Developer software. Tree, shadow, and background classes were defined for classification. After the classification, shadow and background classes were merged, and a binary image that represents tree and nontree classes was obtained.

Results and Discussion

Binary mask outputs produced for the test orthophoto with different models are given in Figure 7. There are 957 trees in total in the orthophoto. When run with the dilation operator, the Mask R-CNN model detected 954 trees correctly, missed three trees, and generated 13 masks that are not actually trees. On the other hand, the U-Net model, run with the dilation operator, detected 893 trees correctly, missed 64, and generated 26 false alarms. On the other hand, when the dilation operator is not applied, Mask R-CNN could detect 936 trees correctly, missed 21 trees, and generated 161 masks that are not actually trees, while U-Net detected 883 trees correctly, missed 74, and generated 55 false alarms. The CART algorithm detected 917 trees correctly, but it missed 40 trees and generated 660 masks that are not actually trees. Accuracy assessment results are given in Table 1.

Comparison of the Models in Tree Detection

When run with the dilation operator, the Mask R-CNN model achieved a precision of 98.7%, a recall of 99.7%, an F1 score of 99.1%, and an IoU of 74.8% for the test orthophoto. The experiment showed that, with the help of dilation operation, the Mask R-CNN model achieved a high success in the detection of apricot trees with a tendency to produce false alarms. A measure of segmentation success, its IoU score was 74.8% due to false alarms and, in some cases, less accurate tree boundary segmentation. U-Net, on the other hand, achieved a recall

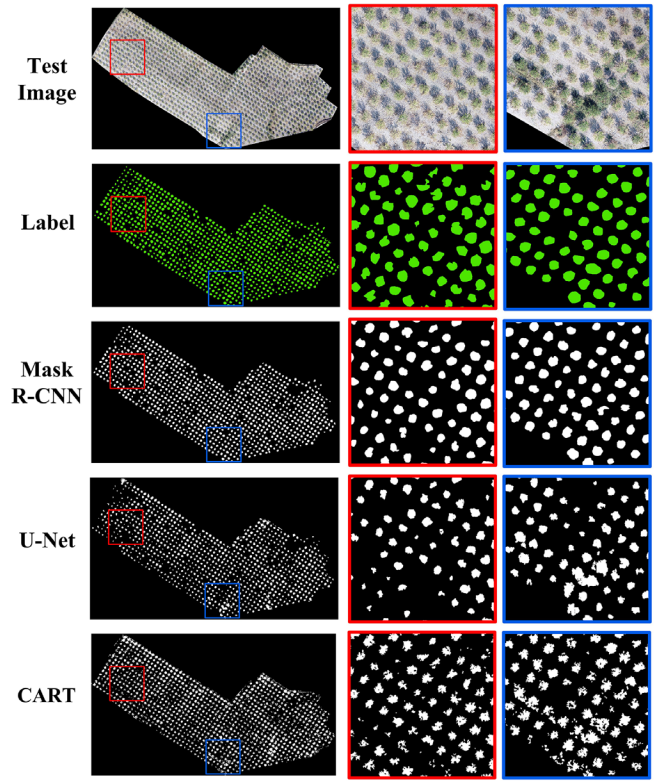


Figure 7. Detection results.

Table 1. Accuracy assessment results.

	Mask R-CNN (%)		U-Net (%)		
Metric	With Dilation	Without Dilation	With Dilation	Without Dilation	CART (%)
F1 score	99.1	91.1	95.2	93.2	73.4
Precision	98.7	85.3	97.2	94.1	58.2
Recall	99.7	97.8	93.3	92.3	95.8
IoU	74.8	69.6	58.3	52.4	57.3

Mask R-CNN = Mask region-based convolutional neural network; CART = classification and regression tree; IoU = intersection over union.

of 93.3%, a precision of 97.2%, an F1 score of 95.2%, and an IoU of 58.3% when run with the dilation operator. Compared to Mask R-CNN, U-Net showed a greater tendency to overlook existing targets. The CART algorithm achieved a recall of 95.8%, a precision of 58.2%, an F1 score of 73.4%, and an IoU of 57.3%. Although the CART algorithm's recall value is not low, it produced detection results with lower success compared to deep learning–based Mask R-CNN and U-Net models.

As shown in Figure 8, the detection capability of Mask R-CNN is outstanding even when it comes to difficult samples. While the model demonstrated the ability to distinguish leafless, dry trees from live trees targeted for detection, U-Net, on the other hand, missed some trees in the scene.

The results support similar studies in the literature. For example, Zhang *et al.* (2022) compared several models for crown delineation of different tree types in a mixed forest, and Mask R-CNN outperformed both U-Net and YOLO. Mask R-CNN holds the edge in the other compared parameters. Zhao *et al.* (2018) conducted similar research on pomegranate trees where Mask R-CNN outperformed U-Net in the accuracy analysis. Different from our study, in their study area, there were multiple connected trees, while in our study area, multiple connected apricot trees are rare. In their study, the biggest disadvantage of Mask R-CNN is the longer training time.

Effectiveness of the Dilation Operator

When the dilation operator did not engage, all detection evaluation metrics for the outputs generated by both models deteriorated. Especially, the precision metric suffers much more for Mask R-CNN, as many false alarms are generated compared to the case where the dilation applied. However, the results produced by U-Net, when run without the dilation operator, were not as badly affected as those of Mask R-CNN. It is noteworthy that the Mask R-CNN result is adversely affected, although IoU, which is a measure of segmentation success, is almost 70% for Mask R-CNN and 52% for U-Net when dilation is not applied. This is due mainly to the inaccurate segmentations produced for objects in the edge regions, which is a characteristic of the Mask R-CNN model. In the U-Net model, this problem is much milder. When the large image is cropped into patches so that it can be processed in models, the new edges of the created patches split some trees into two pieces. When these patches are processed in the models and reassembled, if there are large gaps between the segmentation output of the object and the edge, as in the case of Mask R-CNN output, which is shown on the top row of Figure 9, the tree, which is actually one piece, is detected wrongly as two separate trees. However, as can be seen in the case of U-Net (the image at right in the top row of Figure 9), the gap is much narrower, and two detections from the same object are less likely. When a dilation operator is applied to the outputs, tree masks are enlarged, and dark regions are shrunk (Figure 9, bottom row). Thus, especially as seen in Mask R-CNN output, two masks belonging to the same object are combined and counted as a single object as it should be. Thus, 148 false alarms produced by the Mask R-CNN model on the test orthophoto were eliminated thanks to the applied dilation operator.

Conclusion

This study aims to automatically detect, count, and map apricot trees in a large-scale orthophoto, using Mask R-CNN and U-Net instance

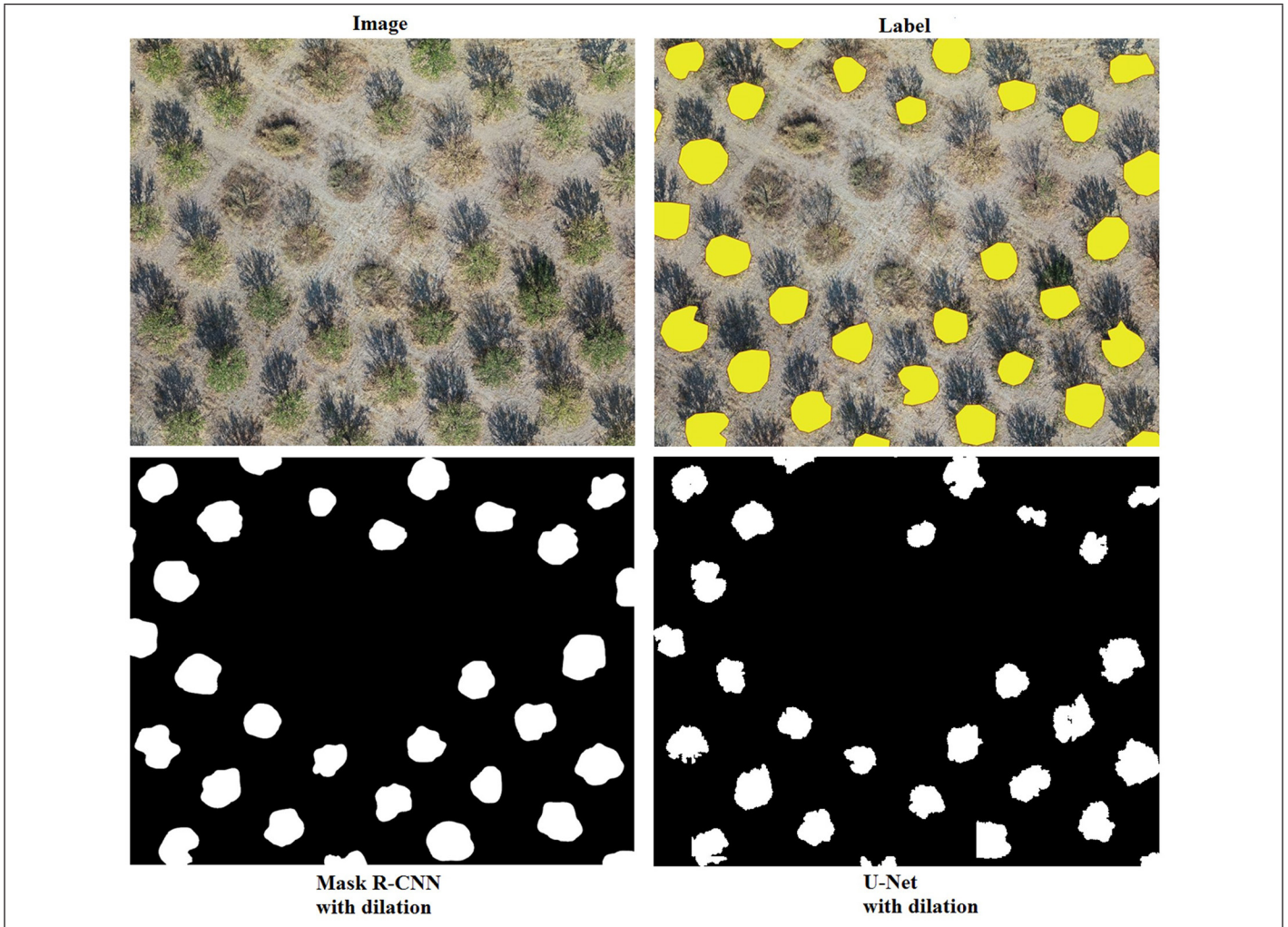


Figure 8. Comparison of the detection performance of the models on hard samples.

segmentation models, and to improve the results by alleviating the object detection problem encountered around image edges. For this purpose, first, training and testing of the models were conducted using the images of apricot trees with distinct characteristics from four different sites to make models robust against variations in diverse tree views. Then a dilation filter was applied on the patches of mask outputs to eliminate gaps between two masks of the same tree in the junction areas of the patches.

The results indicate that, when run with the dilation operator, Mask R-CNN produces better outputs in each of the tree detection, counting, and canopy mapping tasks compared to the U-Net. According to all evaluation metrics used in this study, Mask R-CNN outperforms U-Net. Mask R-CNN was able to produce successful results in areas where accurate detection is difficult even for the trained human eye. U-Net, on the other hand, showed a tendency to overlook existing trees rather than generate false alarms. However, Mask R-CNN seems to suffer more from the edge problem. The results show that when it comes to object detection and mapping tasks for large areas that have to be examined in parts, a dilation operator to be applied on the outputs of instance segmentation models helps to improve detection accuracies greatly.

These results add to the rapidly expanding field of deep learning in the remote sensing area. For future studies, we recommend comparing different models for extracting different tree types. This way, using similar data, the extraction of various tree types with different content in the images can be evaluated.

Acknowledgment

We would like to thank the Turkish Agricultural Insurance Pool (TARSIM), which contributed to our work by sharing UAV data.

References

Abdulla, W.. 2017. Mask R-CNN for object detection and instance segmentation on keras and tensorflow. https://github.com/matterport/Mask_RCNN

Benediktsson, J. A., M. Pesaresi, and K. Amason. 2003. Classification and feature extraction for remote sensing images from urban areas based on morphological transformations. *IEEE Transactions on Geoscience and Remote Sensing* 41(9):1940–1949.

Chen, S. W., S. S. Shivakumar, S. Dcunha, J. Das, E. Okon, C. Qu, C. J. Taylor, and V. Kumar. 2017. Counting apples and oranges with deep learning: A data-driven approach. *IEEE Robotics and Automation Letters* 2(2):781–788.

Cheng, T., X. Wang, L. Huang, and W. Liu. 2020. Boundary-preserving Mask R-CNN. Pages 660–676 in *Proceedings of the European Conference on Computer Vision*, held in Glasgow, United Kingdom, 23-28 August 2020. Berlin: Springer-Verlag.

Dong, H., G. Yang, F. Liu, Y. Mo, and Y. Guo. 2017. Automatic brain tumor detection and segmentation using U-Net based fully convolutional networks. Pages 506–517 in *Proceedings of the Annual Conference on Medical Image Understanding and Analysis*, held in Edinburgh, Scotland, 11-13 Jully 2017. Berlin: Springer-Verlag.

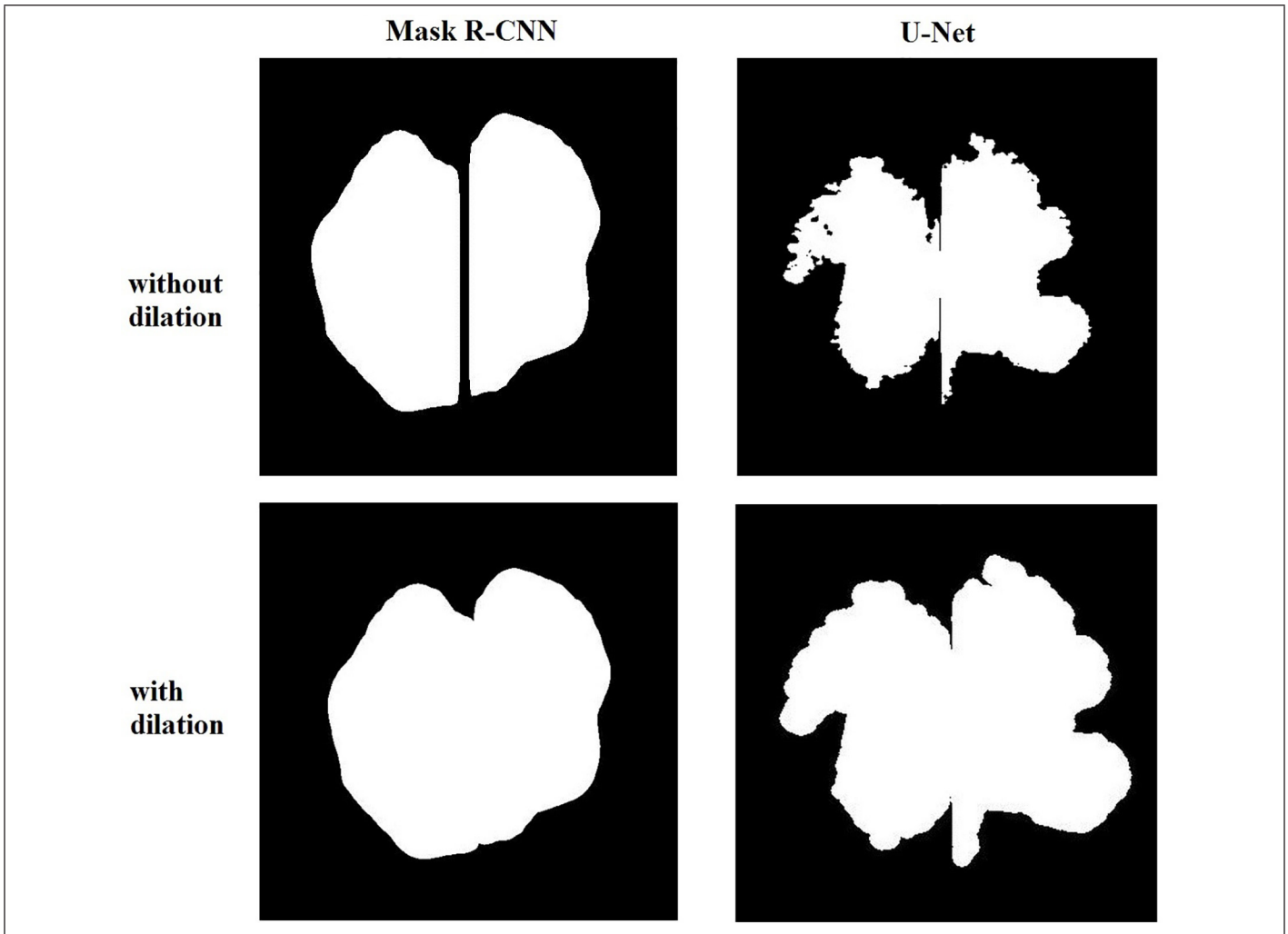


Figure 9. Effect of dilation filter on the mask outputs generated by Mask region-based convolutional neural network and U-Net for the same ground truth.

Falk, T., D. Mai, R. Bensch, Ö. Çiçek, A. Abdulkadir, Y. Marrakchi, A. Böhm, J. Deubner, Z. Jäckel, and K. Seiwald. 2019. U-Net: Deep learning for cell counting, detection, and morphometry. *Nature Methods*, 16(1):67–70.

Gao, L., Y. He, X. Sun, X. Jia, and B. Zhang. 2019. Incorporating negative sample training for ship detection based on deep learning. *Sensors* 19(3):684.

Girshick, R. 2015. Fast R-CNN. Pages 1440–1448 in *Proceedings of the IEEE International Conference on Computer Vision*, held in Santiago, Chile, 7-13 December 2015. Washington, D.C.: IEEE Computer Society.

He, K., G. Gkioxari, P. Dollár, and R. Girshick, R. 2017. Mask R-CNN. Pages 2961–2969 in *Proceedings of the IEEE International Conference on Computer Vision*, held in Venice, Italy, 22-29 October 2017. Washington, D.C.: IEEE Computer Society.

Huang, H., X. Li, and C. Chen. 2018. Individual tree crown detection and delineation from very-high-resolution UAV images based on bias field and marker-controlled watershed segmentation algorithms. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 11(7):2253–2262.

Kirillov, A., Y. Wu, K. He, and P. Girshick. 2020. Pointrend: Image segmentation as rendering. Pages 9799–9808 in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, held in , 13-19 June 2020. Washington, D.C.: IEEE Computer Society.

Koc-San, D., S. Selim, N. Aslan, and B. T. San. 2018. Automatic citrus tree extraction from UAV images and digital surface models using circular Hough transform. *Computers and Electronics in Agriculture* 150:289–301.

Korznikov, K. A., D. E. Kislov, J. Altman, J. Doležal, A. S. Vozmishcheva, and P. V. Krestov. 2021. Using U-Net-like deep convolutional neural networks for precise tree recognition in very high resolution RGB (red, green, blue) satellite images. *Forests* 12(1):66.

Lewis, R. J. 2000. An introduction to classification and regression tree (CART) analysis. In *Proceedings of the Annual Meeting of the Society for Academic Emergency Medicine*, held in San Francisco, Calif., 22-25 May 2000. Citeseer.

Li, W., R. Dong, H. Fu, and L. Yu. 2018. Large-scale oil palm tree detection from high-resolution satellite images using two-stage convolutional neural networks. *Remote Sensing* 11(1):11.

Lumnitz, S., T. Devisscher, J. R. Mayaud, V. Radic, N. C. Coops, and V. C. Griess. 2021. Mapping trees along urban street networks with deep learning and street-level imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* 175:144–157.

Ma, M., J. Chen, W. Liu, and W. Yang. 2018. Ship classification and detection based on CNN using GF-3 SAR images. *Remote Sensing* 10(12):2043.

Mohan, M., C. A. Silva, C. Klauberg, P. Jat, G. Catts, A. Cardil, A. T. Hudak, and M. Dia. 2017. Individual tree detection from unmanned aerial vehicle (UAV) derived canopy height model in an open canopy mixed conifer forest. *Forests* 8(9):340.

Nowozin, S. Optimal decisions from probabilistic models: the intersection-over-union case. *Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 548–555.

Ocer, N. E., G. Kaplan, F. Erdem, D. Kucuk Matci, and U. Avdan. 2020. Tree extraction from multi-scale UAV images using Mask R-CNN with FPN. *Remote Sensing Letters* 11(9):847–856.

Quoc, T. T. P., T. T. Linh, and T. N. T. Minh. 2020. Comparing U-Net convolutional network with mask R-CNN in agricultural area segmentation on satellite images. Pages 124–129 in *Proceedings of the 2020 7th NAFOSTED Conference on Information and Computer Science (NICS)*, held in Ho Chi Minh city, Vietnam, 26-27 November 2020. New York: IEEE).

Razi, M. A. and K. Athappilly. 2005. A comparative predictive analysis of neural networks (NNs), nonlinear regression and classification and regression tree (CART) models. *Expert Systems with Applications* 29(1):65–74.

Ronneberger, O., P. Fischer, and T. Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. Pages 234–241 in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, held in Munich, Germany, 5-9 October 2015. Berlin: Springer-Verlag.

Safonova, A., S. Tabik, D. Alcaraz-Segura, A. Rubtsov, Y. Maglinets, and F. Herrera. 2019. Detection of fir trees (*Abies sibirica*) damaged by the bark beetle in unmanned aerial vehicle images with deep learning. *Remote Sensing* 11(6):643.

Santos, A. A. D., J. Marcato Junior, M. S. Araújo, D. R. Di Martini, E. C. Tetila, H. L. Siqueira, C. Aoki, A. Eltner, E. T. Matsubara, and H. Pistori. 2019. Assessment of CNN-based methods for individual tree detection on images captured by RGB cameras attached to UAVs. *Sensors* 19(16):3595.

Santoso, H., H. Tani, and X. Wang. 2016. A simple method for detection and counting of oil palm trees using high-resolution multispectral satellite imagery. *International Journal of Remote Sensing* 37(21):5122–5134.

Shafri, H. Z., N. Hamdan, and M. I. Saripan. 2011. Semi-automatic detection and counting of oil palm trees from high spatial resolution airborne imagery. *International Journal of Remote Sensing* 32(8):2095–2115.

Sokolova, M., N. Japkowicz, and S. Szpakowicz. 2006. Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation. Pages 1015–1021 in *Proceedings of the Australasian Joint Conference on Artificial Intelligence*, held in Hobart, Australia, 4-8 December 2006. Berlin: Springer-Verlag.

Wang, Y., X. Zhu, and B. Wu. 2019. Automatic detection of individual oil palm trees from UAV images using HOG features and an SVM classifier. *International Journal of Remote Sensing* 40(19):7356–7370.

Yang, X., X. Li, Y. Ye, R. Y. K. Lau, X. Zhang, and X. Huang. 2019. Road detection and centerline extraction via deep recurrent convolutional neural network U-Net. *IEEE Transactions on Geoscience and Remote Sensing* 57(9):7209–7220.

Zhang, C., J. Zhou, H. Wang, T. Tan, M. Cui, Z. Huang, P. Wang, and L. Zhang. 2022. Multi-species individual tree segmentation and identification based on improved Mask R-CNN and UAV imagery in mixed forests. *Remote Sensing* 14(4):874.

Zhang, S., R. Wu, K. Xu, J. Wang, and W. Sun. 2019. R-CNN-based ship detection from high resolution remote sensing imagery. *Remote Sensing* 11(6):631.

Zhao, T., Y. Yang, H. Niu, D. Wang, and Y. Chen. 2018. Comparing U-Net convolutional network with Mask R-CNN in the performances of pomegranate tree canopy segmentation. Pages 210–218. in *SPIE Asia-Pacific Remote Sensing*, held in Honolulu, Hawaii, 24-26 September 2018. Bellingham, WA: SPIE

Comparative Analysis of Different CNN Models for Building Segmentation from Satellite and UAV Images

Batuhan Sariturk, Damla Kumbasar, and Dursun Zafer Seker

Abstract

Building segmentation has numerous application areas such as urban planning and disaster management. In this study, 12 CNN models (U-Net, FPN, and LinkNet using EfficientNet-B5 backbone, U-Net, SegNet, FCN, and six Residual U-Net models) were generated and used for building segmentation. Inria Aerial Image Labeling Data Set was used to train models, and three data sets (Inria Aerial Image Labeling Data Set, Massachusetts Buildings Data Set, and Syedra Archaeological Site Data Set) were used to evaluate trained models. On the Inria test set, Residual-2 U-Net has the highest F1 and Intersection over Union (IoU) scores with 0.824 and 0.722, respectively. On the Syedra test set, LinkNet-EfficientNet-B5 has F1 and IoU scores of 0.336 and 0.246. On the Massachusetts test set, Residual-4 U-Net has F1 and IoU scores of 0.394 and 0.259. It has been observed that, for all sets, at least two of the top three models used residual connections. Therefore, for this study, residual connections are more successful than conventional convolutional layers.

Introduction

In the last few years, particularly with the developments in urbanized areas, building detection and segmentation from aerial images have become challenging and important research topics within the remote sensing and computer vision community (Ye *et al.* 2021). Assigning each pixel into a building or non-building class is challenging because there are big differences in sizes and shapes of buildings, strong similarities between buildings and non-buildings, and small between class and large within-class variance in pixel values in aerial images (Chen *et al.* 2021). Building map generation from aerial images is a costly, laborious, and mostly a manual process (Zhu *et al.* 2021). However, tasks such as detection of illegal buildings (Liu *et al.* 2021), change detection (Sun *et al.* 2020), urban planning (Rathore *et al.* 2016), and disaster management (Liu *et al.* 2020a) all require precise building segmentation.

With continuous developments in current technologies, such as unmanned aerial vehicles (UAVs), sensors, and satellites, a large number of high-resolution aerial images have become more accessible and preferred data for building segmentation applications (Zhu *et al.* 2021), which opened new paradigms for the community. With these high-resolution aerial images, the texture, structure, and spectral information of the buildings have become more refined (Wang and Miao 2022), and these images facilitate the automatic building segmentation greatly (Hu *et al.* 2021). But this brings more challenges that arise from diversity in building characteristics to complex backgrounds (Liu *et al.* 2020b; Tian *et al.* 2021). Therefore, improving the accuracy and efficiency of building segmentation is still a challenging task and the focus of many studies (Ye *et al.* 2021).

Technologies used in building segmentation studies have also progressed over the years (Ye *et al.* 2021) and various techniques have been proposed. In earlier studies, most of these methods were traditional image processing methods that focus on pixels (Sirmacek and Unsalan 2008), lengths, edges (Ferraioli 2010), shapes (Dunaeva and Kornilov 2017), spectrum (Zhong *et al.* 2008), textures (Awrangjeb *et al.* 2013) and shadows (Chen *et al.* 2014; Sirmacek and Unsalan 2008) as input features. Then, later on, conventional machine learning methods such as Support Vector Machines (Inglada 2007), K-Means (Celik 2009), Random Forests (Dong *et al.* 2015), Conditional Random Fields (CRF) (Li *et al.* 2015) and AdaBoost have been started to be used. However, the building structure complexities, similarities with other classes (such as roads and cement backgrounds), dependency on prior knowledge of the topic (Liu *et al.* 2020b), and the need for manual feature selection led to poor generalization capabilities and bias (Liu *et al.* 2021). Thus, they are not well suited for building segmentation from high-resolution aerial images.

Recently, deep learning approaches, most popularly convolutional neural networks (CNN) have come forth as successful methods for computer vision applications (Qin *et al.* 2018), due to the availability of large data sets and increased computation power. CNN, which is a subclass of deep neural networks (DNN), is proposed in (LeCun *et al.* 1989). Differently, from previously mentioned conventional methods, deep learning methods can learn the features and semantic information directly from the input (Liu *et al.* 2020b), perform detailed feature mapping via CNNs, and make classifications and segmentations through sequential convolutions with fully connected layers (Wang and Miao 2022). Therefore, CNN can be regarded as a method that combines feature extraction and segmentation/classification into a single model.

CNN models, which are mostly patch-based, have made notable achievements in building extraction tasks. However, they rely on small patches around targets to predict and ignore the relations between patches (Wang and Miao 2022). In addition, these patch-based models are time-consuming. To overcome these problems, fully convolutional networks (FCN) were proposed by (Long *et al.* 2015). In FCNs, fully connected layers that are used in conventional CNNs are replaced with convolutional layers, and the up-sampling layers are incorporated into the model. This allows the models to predict dense segmentation feature maps. This approach has become very popular and has achieved considerable performance on semantic segmentation tasks (Erdem *et al.* 2021; Ozturk *et al.* 2022; Sariturk *et al.* 2020). In addition, numerous variants based on FCN have been proposed such as SegNet (Badrinarayanan *et al.* 2017), DeepLab (Chen *et al.* 2018), DeconvNet (Noh *et al.* 2015), and U-Net (Ronneberger *et al.* 2015).

Despite FCN's success, there are some limitations. Firstly, fully convolutional layers in FCNs lead to a large number of parameters and computation complexity. Secondly, FCNs fail to provide detailed information about complex structures and lose information about small objects (Khan *et al.* 2022). To overcome these disadvantages, (Ronneberger *et al.* 2015) proposed U-Net, which achieved significant

Istanbul Technical University, Faculty of Civil Engineering, Department of Geomatics Engineering, 34469, Istanbul, Turkey (sariturkb@itu.edu.tr).

Contributed by Tolga Bakirman, May 9, 2022 (sent for review July 28, 2022; reviewed by Ozan Arslan, Weixun Zhou).

Photogrammetric Engineering & Remote Sensing
Vol. 89, No. 2, February 2023, pp. 97–105.
0099-1112/22/97–105

doi: 10.14358/PERS.22-00084R2

success in medical image segmentation studies. U-Net, which adopts a U-shape architecture, consists of two parts. The first part is the encoder, which extracts features from the input image. The second part is the decoder, which performs the dense predictions. Nowadays, most CNNs used in segmentation studies from aerial images are based on FCNs.

In this study, 12 CNN models were generated, trained, evaluated, and comparisons were realized. U-Net, Feature Pyramid Networks (FPN), and LinkNet architectures using EfficientNet-B5 backbone, Original U-Net, SegNet, FCNs, and six different Residual U-Net approaches that use six different residual block designs were used for building segmentation from aerial images. The main objectives of the study were:

(1) Comparison of performance and generalization capabilities of different state-of-the-art CNN models for building segmentation
(2) Comparison of different residual block designs inspired by the ResNet architecture, which has been popular in recent years and achieved successful results, with each other and other used approaches
(3) Evaluating the usability of models trained with publicly available image data covering urban and rural areas, for archaeological sites, and which approaches might be more appropriate

These models were trained using the training set prepared using the images from the Inria Aerial Image Labeling Data Set. To test the trained models, three different test sets were prepared, and evaluations were realized. These test sets were prepared using the Inria Aerial Image Labeling Data Set, the Massachusetts Buildings Data Set, and our Syedra Archaeological Site Data Set. This study aims to realize comparisons between used approaches for building segmentation tasks and test the usability of the models trained with the data from urban and rural areas for the detection of buildings and building remains in archaeological sites.

Methodology

In this study, U-Net, FPN, and LinkNet were used as base architectures to generate the models. In addition to the original U-Net model, EfficientNet-B5 is used as a backbone for all three base architectures. Moreover, six different residual block designs (Naranjo-Alcazar *et al.* 2019) were applied to the original U-Net architecture. Consequently, a total of 12 CNN models were obtained.

Architectures

U-Net

U-Net is a CNN that was developed originally for biomedical image segmentation (Ronneberger *et al.* 2015). Due to its effective GPU usage and good representation ability, it has become a popular architecture for image segmentation tasks (Lei *et al.* 2021). U-Net is an architecture with a U-shaped, symmetrical structure. It consists of encoder, decoder, and bottleneck parts that can extract and then concatenate feature maps (Figure 1). The encoder path extracts feature maps, and these feature maps propagate via skip connections to the decoder path. Afterward, the decoder path uses the learned features to reconstruct the images into wanted dimensions using up-sampling (Punn and Agarwal 2021). The bottleneck, which connects encoder and decoder paths, includes two 3 × 3 convolutions. The encoder consists of four blocks; each one includes two sets of 3 × 3 convolutions, Rectified Linear Unit (ReLU) activations, and 2 × 2 max-pooling. The decoder also consists of four blocks and each block includes a 2 × 2 transposed convolution layer, skip connection to the related feature map, and two 3 × 3 convolutions (Ronneberger *et al.* 2015).

Feature Pyramid Networks (FPN)

Lin *et al.* (2017) introduced FPN in 2017. The authors used the inherent pyramidal and multi-scale hierarchy of deep CNNs to construct feature pyramids. The aim was to naturally use the pyramidal shape of the feature hierarchy of CNN and at the same time create a semantically strong feature pyramid (Lin *et al.* 2017). To achieve this, Lin *et al.* (2017) used an architectural design that combines semantically strong and low-resolution features with semantically weak and high-resolution features using lateral connections and top-down path. Consequently, a feature pyramid that has strong and rich semantic information was obtained.

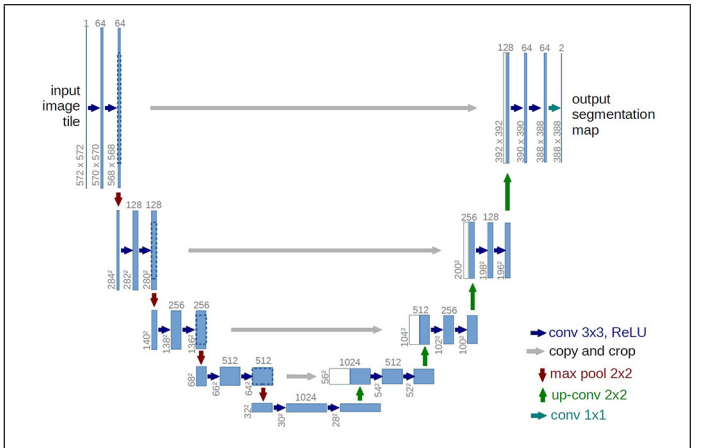


Figure 1. U-Net architecture (Ronneberger *et al.* 2015).

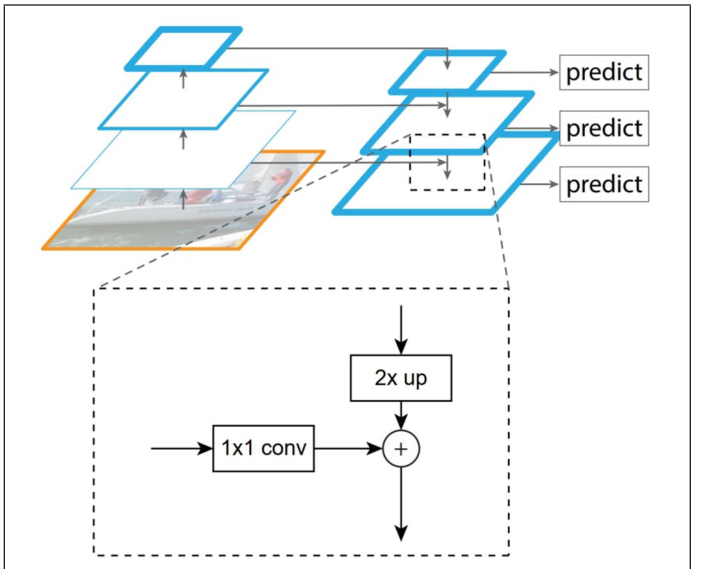


Figure 2. Feature pyramid network architecture (Lin *et al.* 2017).

FPN consists of two paths: a top-down and a bottom-up path (Figure 2). The bottom-up path is a traditional CNN used in image segmentation and classification tasks. The top-down path generates the pyramidal features to be used for object detection. This path takes the last layer of the bottom-up path, up-sample the output of the previous layer, and adds the output of the bottom-up path with the same dimension.

LinkNet

In 2017, Chaurasia and Culurciello proposed a deep neural network called LinkNet, which allows the model to learn without any important increase in the number of parameters (Chaurasia and Culurciello 2017). Unlike many CNN architectures, the novelty of LinkNet lies in the way that they link every encoder with a decoder (Figure 3). Due to the usage of multiple down-sampling in the encoder, some spatial information is lost, and therefore it is quite hard to recover the information by using just these down-sampled outputs. In LinkNet, inputs of each encoder are bypassed to the output of their corresponding decoders (Chaurasia and Culurciello 2017). With this operation, lost spatial information is recovered and can be used by decoders and in up-sampling operations. Additionally, since the encoder and decoder are sharing knowledge learned by the encoder at each layer, decoders can use fewer parameters. This results in a more efficient network design compared to other existing networks (Chaurasia and Culurciello 2017).

SegNet

Badrinarayanan *et al.* (2017) presented a deep fully convolutional network for semantic segmentation, called SegNet. The proposed architecture consists of an encoder, a decoder, and a classification

layer (Figure 4). The topology of the encoder is identical to the first 13 convolutional layers used in the VGG16 network (Badrinarayanan *et al.* 2017). Every convolution layer is followed by a Batch Normalization and a ReLU. There is a corresponding decoder layer for each encoder layer; hence the network has 13 decoder layers. The decoder maps the low-resolution feature maps coming from the encoder to input resolution for pixel-wise classification. The final decoder output is a classifier to produce final class probabilities.

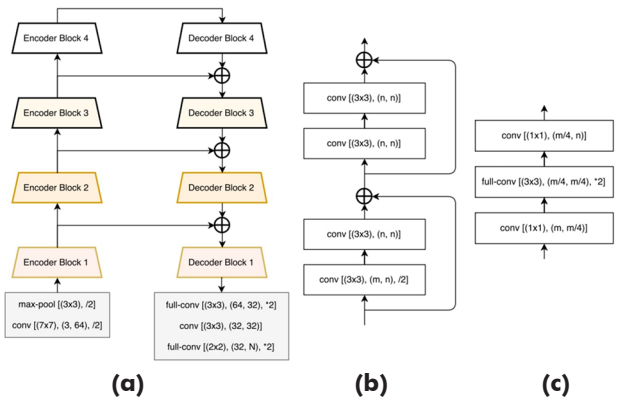


Figure 3. (a) LinkNet architecture, (b) encoder block, (c) decoder block (Chaurasia and Culurciello 2017).

Fully Convolutional Networks (FCN)

In 2015, Long *et al.* (2015) adapted modern networks used for classification (VGGNet, AlexNet, GoogleNet) into fully convolutional networks (FCN) and transferred their representations to image segmentation tasks by fine-tuning. Then, they defined an architecture (Figure 5) that combines the deep, semantic information with the shallow, appearance information by using skip connections to produce detailed and accurate segmentations (Long *et al.* 2015).

Backbone (EfficientNet)

In the study, EfficientNet is used as a backbone for architectures U-Net, FPN, and LinkNet. Tan and Le (Tan and Le 2019) studied the model scaling and found that balancing the width, depth, and resolution of a network can lead to better performance. They proposed a scaling method that scales all dimensions uniformly, using a compound coefficient (Figure 6). As a result of these studies, they designed a baseline network and scaled the network up to obtain a model family called EfficientNet. The EfficientNet family consists of eight models, named from B0 to B7 (Tan and Le 2019). As the number in the model

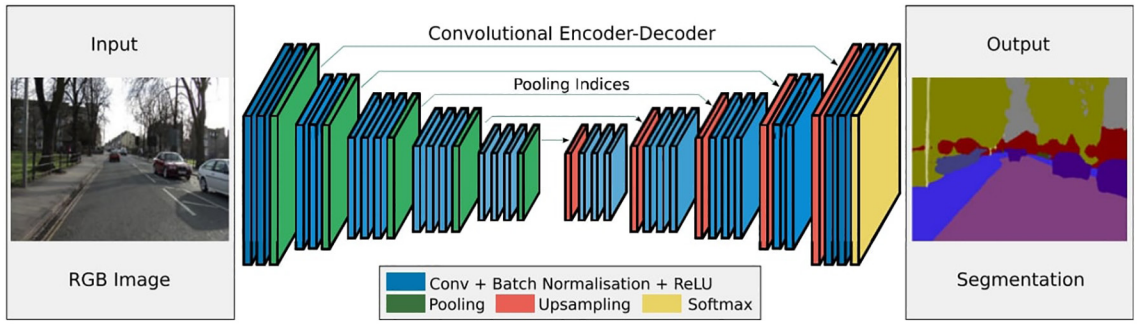


Figure 4. SegNet architecture (Badrinarayanan *et al.* 2017).

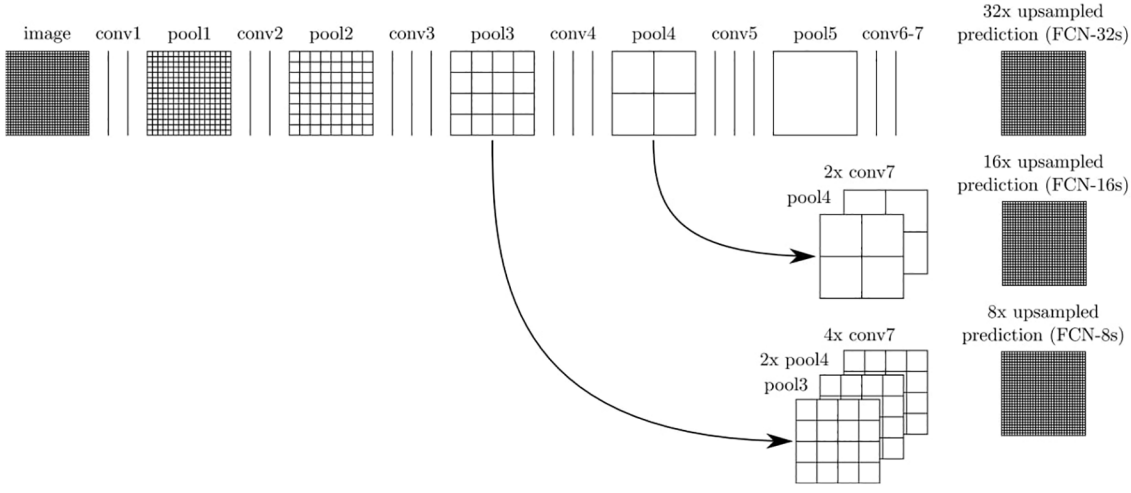


Figure 5. Fully convolutional network architectures (Long *et al.* 2015).

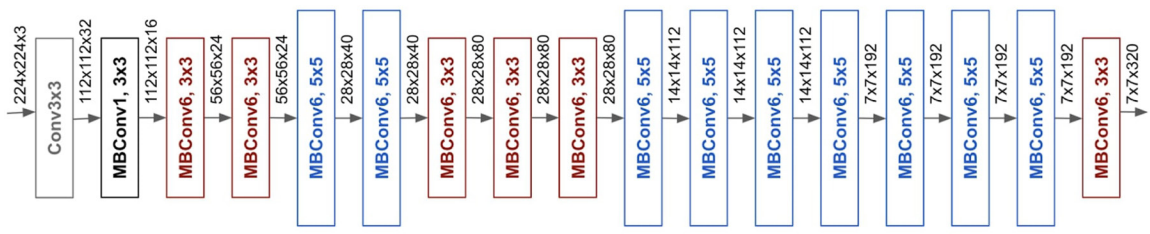


Figure 6. EfficientNet-B0 architecture (Tan and Le 2019).

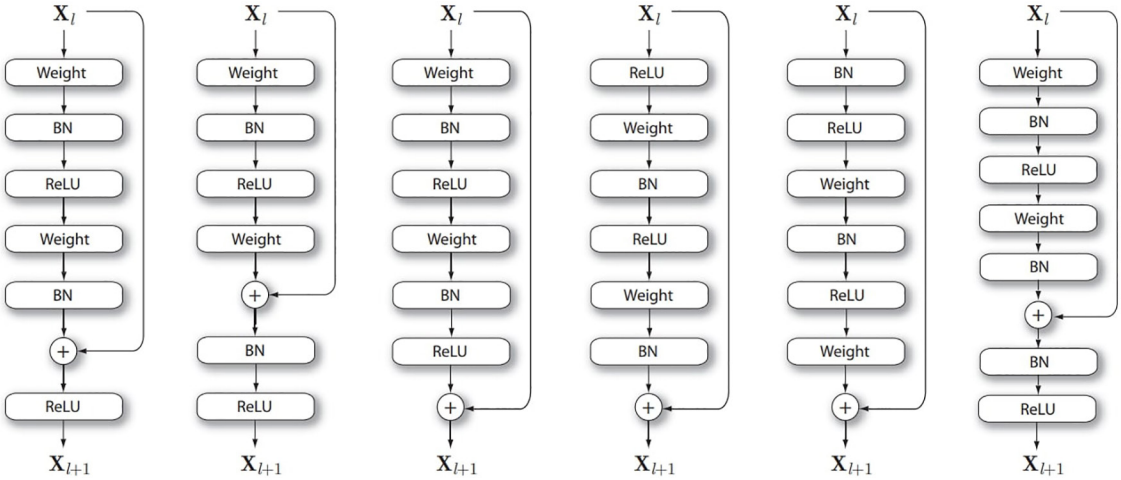


Figure 7. Residual block designs were used in the study (Naranjo-Alcazar *et al.* 2019).

name increases, it is stated that the model has more parameters. For this study, the EfficientNet-B2 variant is selected as the backbone.

Residual Blocks and Residual U-Net Approaches

Residual blocks were originally proposed via ResNet architecture (He *et al.* 2016a), to overcome problems of deep CNNs (He *et al.* 2016a). When the number of layers in a network continues to increase, the "vanishing gradient problem" occurs. During the training of deep models, gradients often get very small or zero, and eventually leave the weights nearly unchanged. Using residual connections, information can be directly passed many layers down (He *et al.* 2016a). In this study, additionally to the residual block (RB1) used in ResNet, four different block designs (RB2-RB5) by (He *et al.* 2016b), and one (RB6) introduced by (Dai *et al.* 2017) were used to generate Residual U-Net approaches (Figure 7). Conventional convolution layers and convolution blocks are replaced with residually connected layers and residual blocks. As a result, six different Residual U-Net approaches have been generated.

Data Sets

In the study, a total of three different data sets were used. Inria Aerial Image Labeling Data Set was selected to train and test the models, and two additional data sets were also used to test the trained models. These additional test sets were prepared using Massachusetts Buildings Data Set and our Syedra Archaeological Site Data Set.

Inria Aerial Image Labeling Data Set

The Inria Aerial Image Labeling Data Set is a benchmark data set provided by Inria for use in building segmentation studies (Maggiori *et al.* 2017). This data set features RGB images with 30 cm spatial resolution. The data set is divided into two sets as training and testing (Maggiori *et al.* 2017). Every image in the training set has a corresponding binary mask, with building and not building classes. Since the images in the provided test set don't have masks, only images and corresponding masks from the training set were used. The regions where the images are from differ from densely populated areas (e.g., Chicago and Vienna) to rural areas (e.g., Austrian Tyrol). The usage of this diverse data set provides a high generalization ability. The training set provided by Inria consists of a total of 180 images and their masks from five cities with different characteristics (Chicago, Ill., USA; Austin, Tex., USA; Kitsap County, Wash., USA; Vienna and Western Tyrol, Austria). These images are in the size of 5000 × 5000 pixels and each covers an area of 1500 × 1500 m².

In order to augment the size of the data set and cut the computational cost, images and masks were cropped into smaller patches in size of 512 × 512 pixels (Figure 8). Images and masks with a low number of buildings and no buildings were eliminated from the set, and as a result, 11 210 images and corresponding masks were obtained. From this data set, 7848 images and masks (70%) were selected to be used as the training set, 1681 images and masks (15%) were selected to be the validation set, and the remaining 1681 images and masks (15%) were selected as the test set. In these selections, attention has been given to the homogeneous distribution of the rural and urban areas.

Figure 8. Sample 512 × 512 image and mask from Inria data set.

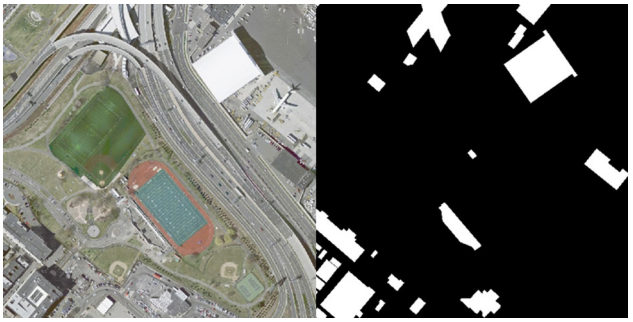


Figure 9. Sample 512 × 512 image and mask from Massachusetts data set.

Massachusetts Buildings Data Set

One of two data sets used to generate additional test sets is Massachusetts Buildings Data Set. This data set was introduced by Volodymyr Mnih in 2013 (Mnih 2013) and includes 151 RGB images and corresponding masks from Boston, Mass., USA. Images in the data set is the size of 5000 × 5000 pixels and covers an area of 2.25 km². Selected images and masks from the set were cropped into 512 × 512 pixel size patches (Figure 9), similar to the preparation of the Inria set, and the images with little to no amount of building classes were

Figure 10. Sample 256 × 256 image and mask from Syedra data set.

eliminated. As a result of all the steps, a test set that includes 225 images and corresponding masks was obtained.

Syedra Archaeological Site Data Set

Syedra Archaeological Site Data Set was acquired over the Syedra Ancient City in November 2020. The ancient city of Syedra is located approximately 20 km southeast of Alanya, Antalya, Turkey. While the central structures of Syedra, which are spread over a very large area, are concentrated on the summit of Asar Tepe (approximately 400 m above sea level) and its south-facing high parts, there are also city-related building groups in the harbor area on the coastline and other smaller hills around Asar Tepe (Can 2017). In the study, the images were collected from the hilly area (approximately 10 ha), which is the old city center, where the buildings that have been preserved until today are located.

The images have been acquired using a DJI Mavic 2 Pro UAV with Hasselblad L1D-20c, 20 MP aerial camera. They were obtained with two flights with fixed take-off heights of 46 m and 58 m in the south region of the city. The collected images were cropped into 256 × 256 pixel size patches (Figure 10), the images with little to no amount of building classes were eliminated, and as a result, a total of 487 images and their corresponding masks were obtained to be used as the second test set.

Training and Testing

In the study, U-Net, FPN, and LinkNet architectures with EfficientNet-B5 backbone, original U-Net architecture, SegNet, FCN, and six Residual U-Net approaches with different residual block designs were used to generate models and comparisons between them were realized. To train these models, the prepared Inria data set was used. To test the trained models, in addition to the Inria data set, Massachusetts and Syedra data sets were used. Tensorflow framework was used to build models, and all models (except Residual U-Net approaches) are available in the "Segmentation Models" repository on GitHub (Yakubovskiy 2019).

As mentioned earlier, 70% of the Inria data set was used for training, and 15% was used for validation. To cut the computational cost, images and their corresponding masks were resized from 512 × 512 pixels to 256 × 256 pixels and normalized. "Adam" optimizer (Kingma and Ba 2015) with a 0.0001 initial learning rate was used during the training. To incrementally reduce the learning rate, with respect to training performance, the "ReduceLROnPlateau" callback was used. This callback monitored the validation loss during training, and when the value does not decrease for five consecutive epochs, the learning rate value was reduced by a 0.1 factor. All models are trained with a batch size of 2, and to calculate loss values, two loss functions were used together. These are Binary Cross-Entropy Loss and Dice Loss. Previous studies on building detection have shown that using joint loss functions is effective (Iglovikov *et al.* 2018). Binary Cross-Entropy is a loss function that is used in studies involving binary decisions (Equation 1) (Rosebrock 2017). The dice coefficient is a metric widely used in computer vision studies to calculate similarities between two images (Jadon 2020). In 2016, it has been adapted as Dice Loss (Equation 2) (Sudre *et al.* 2017).

$$L(y, \hat{y}) = -\frac{1}{N} + \sum_{i=0}^{N} \left(y^{*}\log(\hat{y}_i) + (1-y)^{*}\log(1-\hat{y}_i) \right) \quad (1)$$

$$DL(y, \hat{p}) = 1 - \frac{2y\hat{p} + 1}{y + \hat{p} + 1} \quad (2)$$

To determine the duration of the training, another callback called "EarlyStopping" was used. During the training, this callback monitored the validation loss and if the value doesn't decrease for 10 consecutive epochs, the training stopped. In addition to mentioned callbacks, the "ModelCheckpoint" callback was also used to save the model whenever the validation loss value decreased. With this callback function, the best models were saved during the training, and later on, were used for testing and segmentation. The training epochs for all models, along with the number of trainable parameters, were shown in Table 1.

Table 1. Number of trainable parameters and number of epochs.

Architectures	Number of Trainable Parameters	Epoch of the Best Model	Total Number of Epochs
U-Net	31M	12	22
SegNet	37.3M	20	30
FCN	134.3M	49	50
U-Net (EfficientNet-B5 Backbone)	37.4M	42	52
FPN (EfficientNet-B5 Backbone)	31.8M	21	31
LinkNet (EfficientNet-B5 Backbone)	33.7M	12	22
Residual-1 U-Net	31.7M	15	25
Residual-2 U-Net	31.7M	27	37
Residual-3 U-Net	31.7M	18	28
Residual-4 U-Net	31.7M	17	27
Residual-5 U-Net	31.7M	23	33
Residual-6 U-Net	31.7M	18	28

After training, the best models for each architecture were picked and evaluated on the test sets. To evaluate the models, several metrics were used, and image segmentation was performed with 0.5 threshold applied to the class probabilities. The metrics used for the evaluation process are Precision, Recall, F1 Score, and Intersection over Union (IoU). The precision metric is the ratio of True Positive (TP) values to total positive (TP + False Positive (FP)) (Equation 3) (Patterson and Gibson 2017). The recall is the ratio of TP values to all real positive values (TP + False Negative (FN)) (Equation 3) (Patterson and Gibson 2017). F1 Score is a metric used to measure the general performance of a model in segmentation and classification studies, and it is the harmonic mean of Precision and Recall (Equation 4). IoU gives the ratio of the overlapped area between the mask and the prediction, of the union of these areas (Equation 5) (Géron 2019). Besides these metrics, test loss and test accuracy values were calculated for all models, using each test set.

$$\text{Precision} = \frac{TP}{TP+FP} \qquad \text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

$$\text{F1 Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

$$\text{IoU(A, B)} = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cup B|} \quad (5)$$

Table 2. Evaluation results of the models.

	Inria			Syedra			Massachusetts		
	F1	IoU	Test Acc. (%)	F1	IoU	Test Acc (%)	F1	IoU	Test Acc (%)
U-Net	0.819	0.712	94.1	0.188	0.134	82.6	0.312	0.200	79.3
SegNet	0.674	0.558	92.1	0.073	0.053	83.5	0.165	0.101	81.5
FCN	0.693	0.567	91.6	0.187	0.136	83.7	0.254	0.158	73.9
U-Net EfficientNet-B5	0.701	0.568	89.9	0.179	0.115	77.1	0.275	0.173	80.3
FPN EfficientNet-B5	0.661	0.527	89.8	0.140	0.086	80.0	0.299	0.187	79.6
LinkNet EfficientNet-B5	0.690	0.552	89.8	**0.336**	**0.246**	82.6	0.305	0.190	78.1
Residual-1 U-Net	0.811	0.703	94.1	0.218	0.151	83.7	0.368	0.243	83.0
Residual-2 U-Net	**0.824**	**0.722**	**94.7**	0.199	0.140	83.8	0.311	0.202	82.3
Residual-3 U-Net	0.824	0.705	94.2	0.249	0.178	83.7	0.336	0.220	82.4
Residual-4 U-Net	0.803	0.693	93.9	0.252	0.172	78.9	**0.394**	**0.259**	82.0
Residual-5 U-Net	0.797	0.686	93.9	0.141	0.096	83.2	0.328	0.211	81.7
Residual-6 U-Net	0.817	0.712	94.4	0.151	0.107	**84.0**	0.313	0.204	**83.1**

IoU = Intersection over Union; acc. = Accuracy.

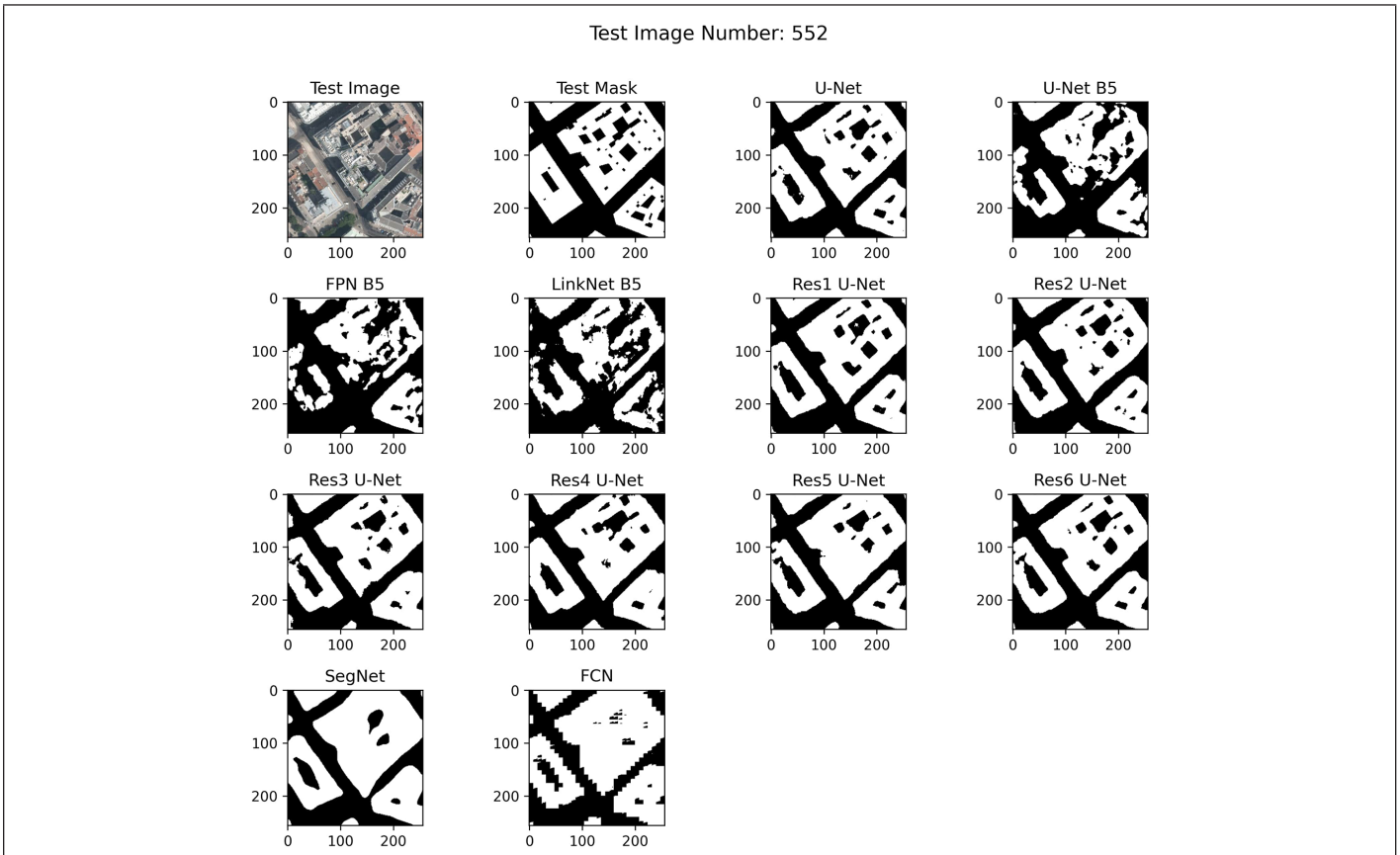


Figure 11. Segmentation results of image no. 552 from the Inria test set.

Results and Discussion

A summary of evaluation metric results is shown in Table 2. It has been observed that on the Inria test set, the Residual-2 U-Net model has the highest IoU score, F1 score, and test accuracy with 0.824, 0.722, and 94.7%, respectively. For the F1 score, Residual-3 U-Net and U-Net follow Residual-2 U-Net with 0.824 and 0.819. For the IoU score, U-Net and Residual-6 U-Net follow with 0.712. On the Syedra test set, the LinkNet model with the EfficientNet-B5 backbone has the highest F1 and IoU scores with 0.336 and 0.246. Residual-4 U-Net and Residual-3 U-Net follow with 0.252 and 0.249 F1 scores, and 0.172 and 0.178 IoU scores, respectively.

On the Massachusetts test set, the Residual-4 U-Net model has the highest F1 and IoU scores with 0.394 and 0.259. Residual-1 U-Net and Residual-3 U-Net models follow with 0.368 and 0.336 F1 scores and 0.243 and 0.220 IoU scores. Residual-6 U-Net model has the highest test accuracy on both Massachusetts and Syedra test sets with 83.1% and 84%, respectively.

Test images, their corresponding masks, and example segmentation images are shown in Figures 11, 12, and 13. Predictions were performed with 0.5 threshold value. In the segmented images presented, white pixels are building classes and black pixels are the background.

According to the results obtained on prepared test sets, the Residual-3 U-Net architecture, using the RB-3 residual block design, is in the top three for all test sets according to the evaluations. Therefore, it can be stated that it is the most inclusive model among all used approaches. When evaluated in general, it has been observed that

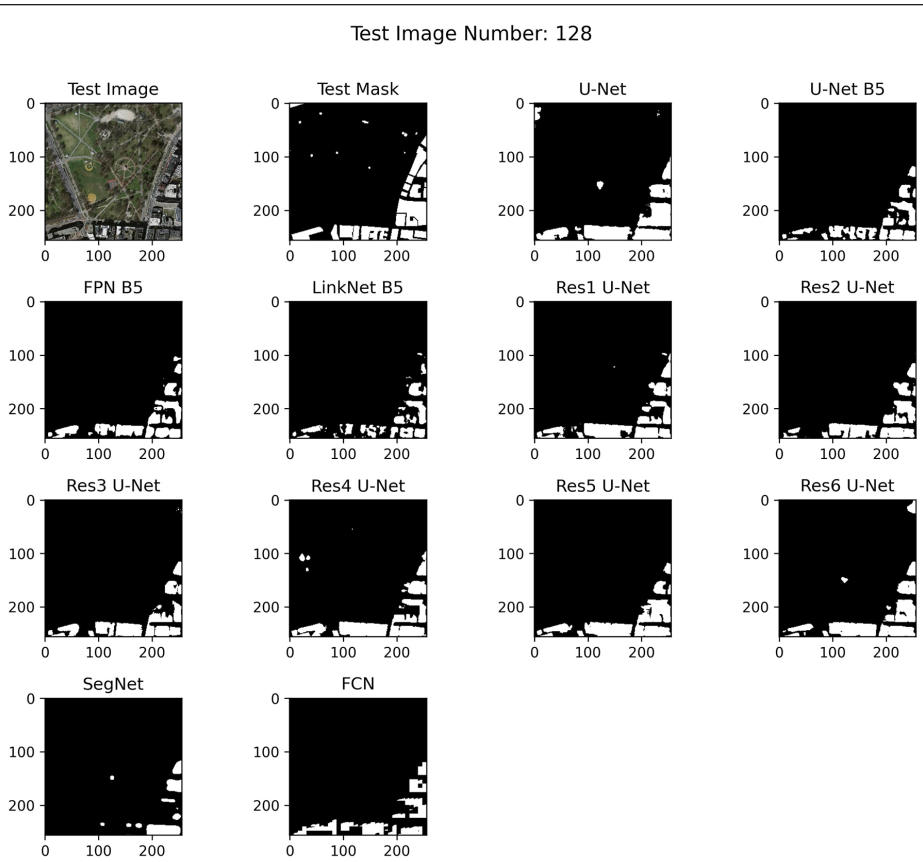


Figure 12. Segmentation results of image no. 128 from the Massachusetts test set.

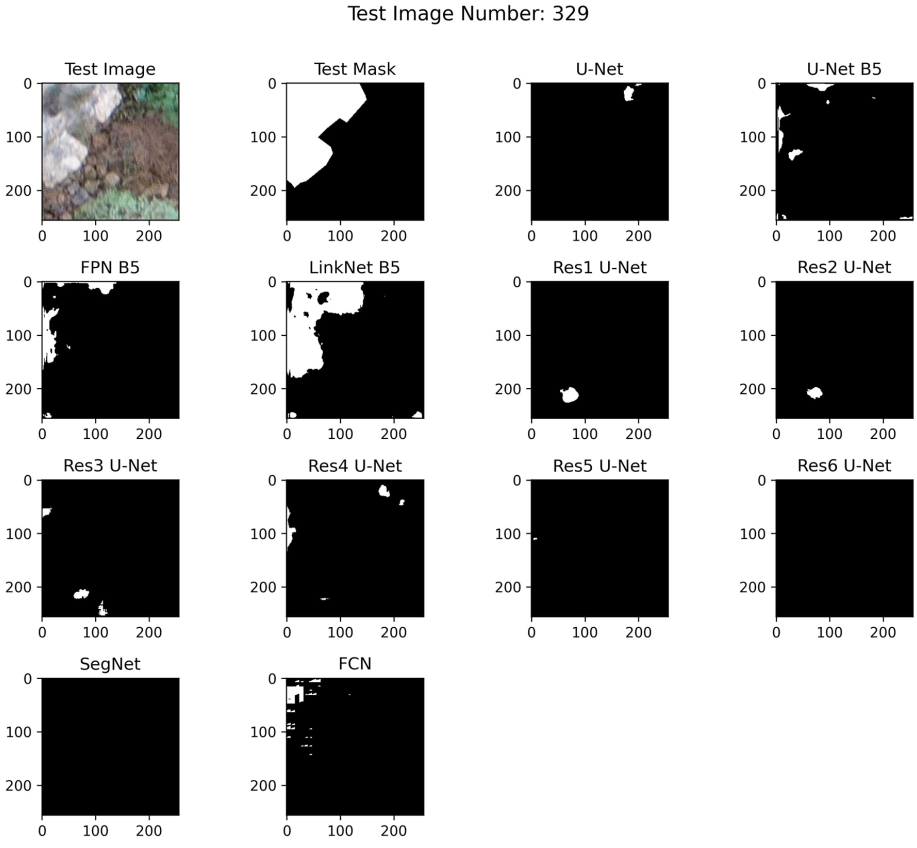


Figure 13. Segmentation results of image no. 329 from the Syedra test set.

models using residual connections are more successful than models using conventional convolution structures. On all three test sets, at least two of the top three most successful models, according to the evaluation metric results, use architectures with residual connections. This can be interpreted as residual connections being more successful due to their ability to carry spatial information from earlier layers. The model uses LinkNet architecture with EfficientNet-B5 backbone achieved the most successful results on the Syedra test set, which is the test set that has the most different feature characteristics, which are archaeological buildings and building remains. Since this architecture achieves the most promising results compared to other models, it could be said that this approach is best suited for different feature characteristics among all models and could be a better choice to be used for future studies on archaeological sites. On the other hand, FPN architecture with EfficientNet-B5 backbone and SegNet were the least successful models for all three test sets. According to the evaluation metric results, SegNet was in the bottom three for all three test sets and provided the lowest F1 and IoU scores for Syedra and Massachusetts test sets. Considering these results on test sets that have different types of feature characteristics than the training set, it could be said that SegNet is the approach that has the least generalization ability among the others used in the study. Therefore, it can be stated that the architectural structures of FPN and SegNet are the least suitable model approaches for building detection and segmentation study compared to other approaches used.

Conclusions

Over the years, automatic building segmentation from aerial images has become a challenging and important research topic with the rapid developments in especially urban areas. Within the study, building segmentation from satellite and UAV images was performed using the Inria Aerial Image Labeling Data Set as the training set, and the Inria Aerial Image Labeling Data Set, the Massachusetts Buildings Data Set, and the Syedra Archaeological Site Data Set as test sets. A total of 12 models were generated, and comparisons were realized.

Results show that the Residual-2 U-Net model, which uses a different residual block from ResNet, performs best in all evaluation metrics on the Inria test set. According to the evaluation metrics, the LinkNet EfficientNet-B5 model on the Syedra Archaeological Site Test Set and the Residual-4 U-Net model on the Massachusetts Test Set showed the best results. Since the features in Syedra Archaeological Site Test Set have different characteristics from the Inria Training Set, it may be said that the LinkNet EfficientNet-B5 model has high generalization ability. Nevertheless, the features from Massachusetts Test Set are similar to the Inria Training Set but from a different kind of sensor, and because of that it can be mentioned that the Residual-4 U-Net model has a generalization ability across similar features and different sensors.

In further studies, performance improvements will be realized on the proposed residual models, with the implementation of approaches such as attention gates and transformers. Additionally, studies to extract buildings and building remains on the archaeological sites will be continued using the Syedra Archaeological Site Data Set and different CNN approaches.

References

Awrangjeb, M., C. Zhang and C. S. Fraser. 2013. Improved building detection using texture information. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XXXVIII-3:143–148. https://doi.org/10.5194/isprsarchives-xxxviii-3-w22-143-2011.

Badrinarayanan, V., A. Kendall and R. Cipolla. 2017. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39:2481–2495.

Can, B. 2017. Syedra ve Teritoryumu. In *Barış Salman Anı Kitabı,* edited by I. A. Adıbelli, E. Baştürk, G. İ. Bertram and K. Matsumura. Istanbul: Ege Yayınları, pp. 263–267.

Celik, T. 2009. Unsupervised change detection in satellite images using principal component analysis and k-means clustering. *IEEE Geoscience and Remote Sensing Letters* 6:772–776.

Chaurasia, A. and E. Culurciello. 2017. LinkNet: Exploiting encoder representations for efficient semantic segmentation. Pages 1–4 in *2017 IEEE Visual Communications and Image Processing (VCIP)*. https://doi.org/10.1109/VCIP.2017.8305148.

Chen, D., S. Shang and C. Wu. 2014. Shadow-based building detection and segmentation in high-resolution remote sensing image. *Journal of Multimedia* 9:181–188. https://doi.org/10.4304/jmm.9.1.

Chen, L. C., G. Papandreou, I. Kokkinos, K. Murphy and A. L. Yuille. 2018. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* https://doi.org/10.1109/TPAMI.2017.2699184.

Chen, M., J. Wu, L. Liu, W. Zhao, F. Tian, Q. Shen and B. Zhao. 2021. DR-Net: An improved network for building extraction from high resolution remote sensing image. *Remote Sensing* 13:294.

Dai, W., C. Dai, S. Qu, J. Li and S. Das. 2017. Very deep convolutional neural networks for raw waveforms. Pages 421–425 in *Proceedings ICASSP, International Conference on Acoustics, Speech, and Signal Processing*. https://doi.org/10.1109/ICASSP.2017.7952190.

Dong, Y., B. Du and L. Zhang. 2015. Target detection based on random forest metric learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8:1830–1838. https://doi.org/10.1109/JSTARS.2015.2416255.

Dunaeva, A. V. and F. A. Kornilov. 2017. Specific shape building detection from aerial imagery in infrared range. *Bulletin of South Ural State University Series, Computational Mathematics and Software Engineering* 6:84–100. https://doi.org/10.14529/cmse170306.

Erdem, F., B. Bayram, T. Bakirman, O. C. Bayrak and B. Akpinar. 2021. An ensemble deep learning based shoreline segmentation approach (WaterNet) from Landsat 8 OLI images. *Advances in Space Research* 67:964–974. https://doi.org/10.1016/j.asr.2020.10.043.

Ferraioli, G. 2010. Multichannel InSAR building edge detection. *IEEE Transactions on Geoscience and Remote Sensing* 48:1224–1231. https://doi.org/10.1109/TGRS.2009.2029338.

Géron, A. 2019. *Hands-On Machine Learning with Scikit-Learn and TensorFlow*, 2d ed. Sebastopol, Calif.: O'Reilly Media.

He, K., X. Zhang, S. Ren and J. Sun. 2016a. Deep residual learning for image recognition. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR.2016.90.

He, K., X. Zhang, S. Ren and J. Sun. 2016b. Identity mappings in deep residual networks. Pages 630–645 in *Lecture Notes in Computer Science*, vol. 9908. Cham: Springer. https://doi.org/10.1007/978-3-319-46493-0_38.

Hu, Q., L. Zhen, Y. Mao, X. Zhou and G. Zhou. 2021. Automated building extraction using satellite remote sensing imagery. *Automation in Construction* 123:103509. https://doi.org/10.1016/j.autcon.2020.103509.

Iglovikov, V., S. Seferbekov, A. Buslaev and A. Shvets. 2018. TernausNetV2: Fully convolutional network for instance segmentation. Pages 228–232 in *IEEE Conference on Computer Vision and Pattern Recognition Workshops*. https://doi.org/10.1109/CVPRW.2018.00042.

Inglada, J. 2007. Automatic recognition of man-made objects in high resolution optical remote sensing images by SVM classification of geometric image features. *ISPRS Journal of Photogrammetry and Remote Sensing* 62:236–248. https://doi.org/10.1016/j.isprsjprs.2007.05.011.

Jadon, S. 2020. A survey of loss functions for semantic segmentation. *IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*. https://doi.org/10.1109/CIBCB48159.2020.9277638.

Khan, S. D., L. Alarabi and S. Basalamah. 2022. An encoder–decoder deep learning framework for building footprints extraction from aerial imagery. *Arabian Journal for Science and Engineering*. https://doi.org/10.1007/s13369-022-06768-8.

Kingma, D. P. and J. L. Ba. 2015. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015*. https://doi.org/10.48550/arXiv.1412.6980.

LeCun, Y., B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard and L. D. Jackel. 1989. Backpropagation applied to handwritten zip code recognition. *Neural Computation*. https://doi.org/10.1162/neco.1989.1.4.541.

Lei, M., Z. Rao, H. Wang, Y. Chen, L. Zou and H. Yu. 2021. Maceral groups analysis of coal based on semantic segmentation of photomicrographs via the improved U-net. *Fuel* 294:120475. https://doi.org/10.1016/j.fuel.2021.120475.

Li, E., J. Femiani, S. Xu, X. Zhang and P. Wonka. 2015. Robust rooftop extraction from visible band images using higher order CRF. *IEEE Transactions on Geoscience and Remote Sensing* 53:4483–4495.

Lin, T.-Y., P. Dollár, R. Girshick, K. He, B. Hariharan and S. Belongie. 2017. Feature pyramid networks for object detection. Pages 2117–2125 in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, held in Honolulu, Hawaii, 21–26 July 2016.

Liu, S., H. Ye, K. Jin and H. Cheng. 2021. CT-UNet: Context-transfer-UNet for building segmentation in remote sensing images. *Neural Processing Letters*. https://doi.org/10.1007/s11063-021-10592-w.

Liu, Y., E. So, Z. Li, G. Su, L. Gross, X. Li, W. Qi, F. Yang, B. Fu, A. Yalikun and L. Wu. 2020a. Scenario-based seismic vulnerability and hazard analyses to help direct disaster risk reduction in rural Weinan, China. *International Journal of Disaster Risk Reduction* 48:101577. https://doi.org/10.1016/j.ijdrr.2020.101577.

Liu, Y., J. Zhou, W. Qi, X. Li, L. Gross, Q. Shao, Z. Zhao, L. Ni, X. Fan and Z. Li. 2020b. ARC-Net: An efficient network for building extraction from high-resolution aerial images. *IEEE Access* 8:154997–155010. https://doi.org/10.1109/ACCESS.2020.3015701.

Long, J., E. Shelhamer and T. Darrell. 2015. Fully convolutional networks for semantic segmentation. Pages 3431–3440 in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, held in Boston, Mass., 7–12 June 2015.

Maggiori, E., Y. Tarabalka, G. Charpiat and P. Alliez. 2017. Can semantic labeling methods generalize to any city? The Inria aerial image labeling benchmark. *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. <https://project.inria.fr/aerialimagelabeling/> Accessed 13 August 2021.

Mnih, V. 2013. Machine *Learning for Aerial Image Labeling*. Ph.D. dissertation. University of Toronto.

Naranjo-Alcazar, J., S. Perez-Castanos, I. Martin-Morato, P. Zuccarello and M. Cobos. 2019. On the performance of residual block design alternatives in convolutional neural networks for end-to-end audio classification. arXiv:1906.10891.

Noh, H., S. Hong and B. Han. 2015. Learning deconvolution network for semantic segmentation. Pages 1520–1528 in *Proceedings IEEE International Conference on Computer Vision*. https://doi.org/10.1109/ICCV.2015.178.

Ozturk, O., M. S. Isik, B. Sariturk and D. Z. Seker. 2022. Generation of Istanbul road data set using Google map API for deep learning-based segmentation. *International Journal of Remote Sensing*. https://doi.org/10.1080/01431161.2022.2068989.

Patterson, J. and A. Gibson. 2017. *Deep learning—A Practitioner's Approach*, 1st ed. Sebastopol, Calif.: O'Reilly Media.

Punn, N. S. and S. Agarwal. 2021. Modality specific U-Net variants for biomedical image segmentation: A survey. arXiv:2107.04537.

Qin, X., S. He, X. Yang, M. Dehghan, Q. Qin and J. Martin. 2018. Accurate outline extraction of individual building from very high-resolution optical images. *IEEE Geoscience and Remote Sensing Letters* 15:1775–1779. https://doi.org/10.1109/LGRS.2018.2857719.

Rathore, M. M., A. Ahmad, A. Paul and S. Rho. 2016. Urban planning and building smart cities based on the Internet of Things using Big Data analytics. *Computer Networks* 101:63–80. https://doi.org/10.1016/j.comnet.2015.12.023.

Ronneberger, O., P. Fischer and T. Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. Pages 234–241 in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. https://doi.org/10.1007/978-3-319-24574-4_28.

Rosebrock, A., 2017. Deep Learning for Computer Vision with Python. PyImageSearch.

Sariturk, B., B. Bayram, Z. Duran and D. Z. Seker. 2020. Feature extraction from satellite images using SegNet and fully convolutional networks (FCN). *International Journal of Engineering and Geosciences* 5:138–143. https://doi.org/10.26833/ijeg.645426.

Sirmacek, B. and C. Unsalan. 2008. Building detection from aerial images using invariant color features and shadow information. In *23rd International Symposium on Computer and Information Sciences*. https://doi.org/10.1109/ISCIS.2008.4717854.

Sudre, C. H., W. Li, T. Vercauteren, S. Ourselin and M. Jorge Cardoso. 2017. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. *Lecture Notes in Computer Science* 10553:240–248. https://doi.org/10.1007/978-3-319-67558-9_28.

Sun, Y., X. Zhang, J. Huang, H. Wang and Q. Xin. 2020. Fine-grained building change detection from very high-spatial-resolution remote sensing images based on deep multitask learning. *IEEE Geoscience and Remote Sensing Letters* 1–5. https://doi.org/10.1109/lgrs.2020.3018858.

Tan, M. and Q. V. Le. 2019. EfficientNet: Rethinking model scaling for convolutional neural networks. Pages 10691–10700 in *Proceedings 36th International Conference on Machine Learning*, held in Long Beach, Calif., 9–15 June 2019.

Tian, Q., Y. Zhao, K. Qin, Y. Li and X. Chen. 2021. Dense feature pyramid fusion deep network for building segmentation in remote sensing image. Page 1176351 in *Proceedings Seventh Symposium on Novel Photoelectronic Detection Technology and Applications*. https://doi.org/10.1117/12.2587144.

Wang, H. and F. Miao. 2022. Building extraction from remote sensing images using deep residual U-Net. *European Journal of Remote Sensing* 55:71–85. https://doi.org/10.1080/22797254.2021.2018944.

Yakubovskiy, P., 2019. Segmentation Models, GitHub Repository. <https://github.com/qubvel/segmentation_models> Accessed 13 August 2021.

Ye, H., S. Liu, K. Jin and H. Cheng. 2021. CT-UNET: An improved neural network based on U-Net for building segmentation in remote sensing images. Pages 166–172 in *Proceedings 2020 25th International Conference on Pattern Recognition (ICPR)*. https://doi.org/10.1109/ICPR48806.2021.9412355.

Zhong, S.-H., J.-J. Huang and W.-X. Xie. 2008. A new method of building detection from a single aerial photograph. Pages 1219–1222 in *Proceedings International Conference on Signal Processing*. https://doi.org/10.1109/ICOSP.2008.4697350.

Zhu, Y., Z. Liang, J. Yan, G. Chen and X. Wang. 2021. E-D-Net: Automatic building extraction from high-resolution aerial images with boundary information. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14:4595–4606. https://doi.org/10.1109/JSTARS.2021.3073994.

SUSTAININGMEMBERS

ACI USA Inc.
Weston, Florida
https://acicorporation.com/
Member Since: 2/2018

Aerial Services, Inc.
Cedar Falls, Iowa
www.AerialServicesInc.com
Member Since: 5/2001

Airworks Solutions Inc.
Boston, Massachusetts
Member Since: 3/2022

Applanix
Richmond Hill, Ontario, Canada
http://www.applanix.com
Member Since: 7/1997

Ayres Associates
Madison, Wisconsin
www.AyresAssociates.com
Member Since: 1/1953

CT Consultants
Mentor, Ohio
Member Since: 3/2022

Dewberry
Fairfax, Virginia
www.dewberry.com
Member Since: 1/1985

Esri
Redlands, California
www.esri.com
Member Since: 1/1987

GeoCue Group
Madison, Alabama
http://www.geocue.com
Member Since: 10/2003

Geographic Imperatives LLC
Centennial, Colorado
Member Since: 12/2020

GeoWing Mapping, Inc.
Richmond, California
www.geowingmapping.com
Member Since: 12/2016

Halff Associates, Inc.
Richardson, Texas
www.halff.com
Member Since: 8/2021

Keystone Aerial Surveys, Inc.
Philadelphia, Pennsylvania
www.kasurveys.com
Member Since: 1/1985

Kucera International
Willoughby, Ohio
www.kucerainternational.com
Member Since: 1/1992

L3Harris Technologies
Broomfield, Colorado
www.l3harris.com
Member Since: 6/2008

Merrick & Company
Greenwood Village, Colorado
www.merrick.com
Member Since: 4/1995

Nearmap
South Jordan, Utah
www.nearmap.com
Member Since: 6/2023

NV5 Geospatial
Sheboygan Falls, Wisconsin
www.quantumspatial.com
Member Since: 1/1974

Pickett and Associates, Inc.
Bartow, Florida
www.pickettusa.com
Member Since: 4/2007

Riegl USA, Inc.
Orlando, Florida
www.rieglusa.com
Member Since: 11/2004

Robinson Aerial Surveys, Inc.(RAS)
Hackettstown, New Jersey
www.robinsonaerial.com
Member Since: 1/1954

Sanborn Map Company
Colorado Springs, Colorado
www.sanborn.com
Member Since: 10/1984

Surdex Corporation
Chesterfield, Missouri
www.surdex.com
Member Since: 12/2011

Surveying And Mapping, LLC (SAM)
Austin, Texas
www.sam.biz
Member Since: 12/2005

T3 Global Strategies, Inc.
Bridgeville, Pennsylvania
https://t3gs.com/
Member Since: 6/2020

Towill, Inc.
San Francisco, California
www.towill.com
Member Since: 1/1952

Woolpert LLP
Dayton, Ohio
www.woolpert.com
Member Since: 1/1985

SUSTAININGMEMBER**BENEFITS**

Membership

- ✓ Provides a means for dissemination of new information
- ✓ Encourages an exchange of ideas and communication
- ✓ Offers prime exposure for companies

Benefits of an ASPRS Membership

- Complimentary and discounted Employee Membership*
- E-mail blast to full ASPRS membership*
- Professional Certification Application fee discount for any employee
- Member price for ASPRS publications
- Discount on group registration to ASPRS virtual conferences
- Sustaining Member company listing in ASPRS directory/website
- Hot link to company website from Sustaining Member company listing page on ASPRS website
- Press Release Priority Listing in *PE&RS* Industry News
- Priority publishing of Highlight Articles in *PE&RS* plus, 20% discount off cover fee
- Discount on *PE&RS* advertising
- Exhibit discounts at ASPRS sponsored conferences (exception ASPRS/ILMF)
- Free training webinar registrations per year*
- Discount on additional training webinar registrations for employees
- Discount for each new SMC member brought on board (Discount for first year only)

*quantity depends on membership level

Unmanned Aerial Vehicle (UAV)–Based Imaging Spectroscopy for Predicting Wheat Leaf Nitrogen

Rabi N. Sahoo, Shalini Gakhar, R.G. Rejith, Rajeev Ranjan, Mahesh C. Meena, Abir Dey, Joydeep Mukherjee, Rajkumar Dhakar, Sunny Arya, Anchal Daas, Subhash Babu, Pravin K. Upadhyay, Kapila Sekhawat, Sudhir Kumar, Mahesh Kumar, Viswanathan Chinnusamy, and Manoj Khanna

Abstract

Quantitative estimation of crop nitrogen is the key to site-specific management for enhanced nitrogen (N) use efficiency and a sustainable crop production system. As an alternate to the conventional approach through wet chemistry, sensor-based noninvasive, rapid, and near-real-time assessment of crop N at the field scale has been the need for precision agriculture. The present study attempts to predict leaf N of wheat crop through spectroscopy using a field portable spectroradiometer (spectral range of 400–2500 nm) on the ground in the crop field and an imaging spectrometer (spectral range of 400–1000 nm) from an unmanned aerial vehicle (UAV) with the objectives to evaluate (1) four multivariate spectral models (i.e., artificial neural network, extreme learning machine [ELM], least absolute shrinkage and selection operator, and support vector machine regression) and (2) two sets of hyperspectral data collected from two platforms and two different sensors. In the former part of the study, ELM outperforms the other methods with maximum calibration and validation R^2 of 0.99 and 0.96, respectively. Furthermore, the image data set acquired from UAV gives higher performance compared to field spectral data. Also, significant bands are identified using stepwise multiple linear regression and used for modeling to generate a wheat leaf N map of the experimental field.

Introduction

Plant growth is highly dependent on nitrogen (N) and has a direct influence on the quality of crops. Leaf N (LNC) content is one of the major factors governing leaf nutrition and is a decision-making factor for accurate and effective fertilizer application. Wheat (*Triticum aestivum* L.) is the second most important cereal crop in India and influences nutritional security of the country (Anuj *et al.* 2014). Therefore, agri-production that increases yield and sustainability requires satisfying the demand and supply trade-off for N use efficiency. For this purpose, management of N is complex and necessitates problem-oriented, interdisciplinary research (Spiertz 2009). Destructive methods for computation of LNC are not only time consuming but also highly expensive and labor intensive at the same time. Remote sensing is a fast, nondestructive, dependable method for accurate assessment of LNC and capturing the LNC variability in the field for site-specific fertilizer application for enhanced N use efficiency and precision agriculture (Barzin *et al.* 2021).

Remote sensing in general has been a very important supporting tool for capturing spatiotemporal variability of different biophysical and biochemical parameters for better managing the required inputs across the field, which encourages sustainable practices through precision farming (Barrientos *et al.* 2011; Shanmugapriya *et al.* 2019). Ground- or field-based spectral data acquisition using a nonimaging sensor like a spectroradiometer is more accurate but is restricted to the collection of limited samples and inefficient for large-scale coverage. Whereas air- or spaceborne imaging sensors suffer from the limitations of a long revisit time and low spatial and spectral resolution and necessitate good light and atmospheric conditions, high cost, and so on (Liu *et al.* 2017), unmanned aerial vehicles (UAVs), or drones, have the extensive capability to address the previously mentioned gaps and acquire agricultural data. They have a multitude of advantages, such as lightweight, low altitude, speed control, high-resolution imagery, quick and iterative missions, and so on (Xiang and Tian 2011). Current satellites providing coarse to medium spatial resolution are not sufficient due to their nonavailability during cloud cover, limited revisit possibility during the crop season, and expensive data set. UAVs have been identified as a viable substitute and/or complement to remote sensing platforms for agricultural applications and mainly for better visualization and quantitative assessment of crop conditions at very high resolution. At present, the sensors that are mounted on UAVs are mostly red, green, blue and multispectral cameras with fewer spectral bands. Therefore, the acquired spectral information is limited to few applications. However, hyperspectral sensors, with hundreds of narrow and contiguous bands, have the capability of capturing subtle variations in plant growth and parameters, thereby having many potential applications in agriculture. The development of hyperspectral sensors has enabled quantitative assessment of plant parameters, encouraging researchers to implement precision agriculture. However, hyperspectral remote sensing comes with its own set of challenges, such as spectral variability, mixed pixel problems, accuracy requirements, selection of an appropriate model for information extraction that would provide better accuracy, and so on (Das *et al.* 2015; Hank *et al.* 2019; Pandey *et al.* 2020).

Civil aviation and other relevant authorities elsewhere in the world are now allowing UAVs to be used for limited and specific trials and, in some cases, commercial operations in agriculture (Ahirwar *et al.* 2019), horticulture (Tu *et al.* 2020), viticulture (Spachos and Stefano 2019), and forestry (Tang and Shao 2015). Spraying for controlling diseases, weeds, and pests; spreading microgranular pesticides and fertilizers and even beneficial insects; and planting are among the diverse uses now being done using UAVs (Rao Mogili and Deepak 2018; Chen *et al.* 2021). Plant parameters obtained from UAV-based sensors can be used not only for monitoring crop growth conditions but also for crop response to differential treatment of nutrients and water (Yang *et al.* 2020). N status monitoring is important for farmers, as it leads

Rabi N. Sahoo, Shalini Gakhar, R.G. Rejith, Rajeev Ranjan, Joydeep Mukherjee, Rajkumar Dhakar and Sunny Arya are with the Division of Agricultural Physics, ICAR–Indian Agricultural Research Institute (IARI), New Delhi 110012, India (rabi.sahoo@icar.gov.in).

Mahesh C. Meena and Abir Dey are with the Division of Soil Science and Agricultural Chemistry, ICAR-IARI, New Delhi 110012, India.

Anchal Daas, Subhash Babu, Pravin K. Upadhyay, and Kapila Sekhawat are with the Division of Agronomy, ICAR-IARI, New Delhi 110012, India.

Sudhir Kumar, Mahesh Kumar, and Viswanathan Chinnusamy are with the Division of Plant Physiology, ICAR-IARI, New Delhi 110012, India.

Manoj Khanna is with the Water Technology Centre, ICAR-IARI, New Delhi 110012, India.

Contributed by Tolga Bakirman, May 12, 2022 (sent for review September 7, 2022; reviewed by Itiya Aneece, Fusun Balik Sanli).

Photogrammetric Engineering & Remote Sensing
Vol. 89, No. 2, February 2023, pp. 107–116.
0099-1112/22/107–116

doi: 10.14358/PERS.22-00089R2

to an efficient fertilizer application (Lee *et al.* 2020a). Various studies have attempted to capture spectral reflectance data of crops at different growth stages and also synchronized measurements of plant parameters, such as leaf area index, leaf dry matter, plant dry matter, and three plant N indicators, that is, leaf N accumulation, plant N accumulation, and N nutrition index, and developed their relationships through modeling in the recent past.

Most of the studies presented in literature are based on either ground-based or airborne image–derived spectral data for assessing plant N content. Spectroradiometric study of plant N may not help one scale up to the field level , as this is mainly single-pixel based regardless of field of view coverage. There is a need to scale up the same at the farm scale using airborne imaging spectrometry. However, intercomparison of the sensor data at different platforms needs to be studied. To assess N content, many studies use univariate and multivariate statistical analyses. Univariate regression models involve the association of the quantitative correlation of crop parameters with spectrally derived indices, such as the normalized difference vegetative index (NDVI). Nevertheless, numerous recent studies involve nonlinear multivariate models to accommodate more spectral data for better prediction with high accuracy. These models include partial least square regression (PLSR) (Santos-Rufo *et al.* 2020), stepwise multiple linear regression (MLR) (Zheng *et al.* 2018), random forest (RF) (Yang *et al.* 2021), multivariate adaptive regression splines (Mahajan *et al.* 2021), artificial neural networks (ANN), support vector machine (SVM) (Yuan *et al.* 2017), extreme learning machine (ELM) (Yu *et al.* 2020), least absolute shrinkage and selection operator (LASSO) (Shafiee *et al.* 2021), and many more.

The present study exploits two sets of spectral data sets collected from imaging sensors on UAV platforms and nonimaging sensors, namely, a spectroradiometer at ground level in the field, to predict leaf the N content of wheat crops. The approach performs an exhaustive comparative assessment of machine learning multivariate models in order to find the suitable predictive model(s) for N estimation. Furthermore, the work done here aims to promote the application of UAV-based hyperspectral remote sensing for the same field scale for site-specific nutrient management.

Materials and Methods

Field Experimentation

An experiment was conducted at the Research Farm of ICAR–Indian Agricultural Research Institute, New Delhi, India, during *rabi* season (from December to April 2021–2022). The experimental field was located at 28°38′28.314″N latitude and 77°9′3.106″E longitude with an average elevation of 230 m above mean sea level. The wheat crop (*T. aestivum* L.) variety HD 3059 was sown on 13 December 2021 under three irrigation treatments—irrigation based on soil moisture sensor (I_1); crop water stress index, CWSI (I_2); and conventional (I_3)—and five N levels—0 kg N ha^{-1} (N_0), 50 kg N ha^{-1} (N_1), 100 kg N ha^{-1} (N_2), 150 kg N ha^{-1} (N_3), and 200 kg N ha^{-1} (N_4). The experiment was conducted in a split-plot design with three replications (R_1, R_2, and R_3) in which irrigation was the main plot treatment and N were the subplot treatment in a plot size of 13×7.2 m. Thus, the total number of plots was 45 (3×5×3). One-third of the recommended amount of N and the full recommended amount of phosphorus (P) and potassium (K) were applied as a basal dose in the form of urea (46% N), diammonium phosphate (18% N and 46% P_2O_5), and muriate of potash (60% K_2O), respectively. The remaining N was applied at the time of first and second irrigation in two equal split doses. The leaf samples from 45 plots were collected, and their total N concentration was determined by the Micro–Kjeldahl method (Guebel *et al.* 1991; Ranjan *et al.* 2012). The plot layout of the research farm is shown in Figure 1 along with irrigation, and N treatment details are tabulated in Table 1.

Collection of Spectral Data Using Ground-Held Spectroradiometer

The spectral reflectance of leaf samples in the range of 350–2500-nm wavelength has been collected at a 1-nm interval using a contact probe with an ASD field spec-3 spectroradiometer (Analytical Spectral Devices Inc., Boulder, CO, USA). After acquiring the spectral data,

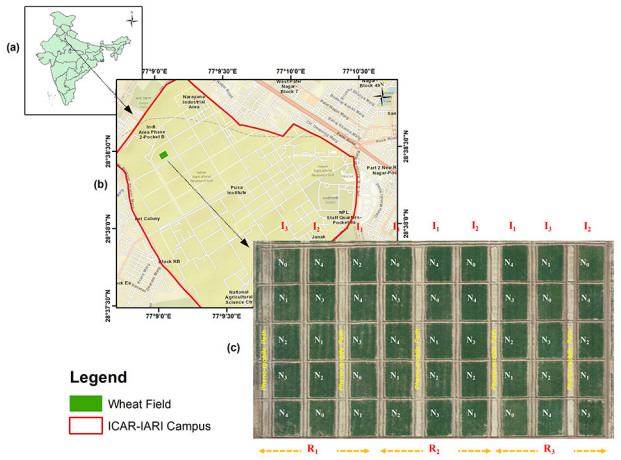


Figure 1. Geographic location of the experimental plot with different fertilization treatments. (a and b) Location of ICAR-IARI campus and experimental wheat plots (source: ESRI base map, India Street Map). (c) Wheat field layout collected on 7 February 2022.

Table 1. Irrigation and fertilization treatments in the field experiment on wheat.

	Irrigation		Fertilization	
Crop	Main Plot	Basis of Irrigation	Subplot	Graded Nitrogen (N) Levels (kg ha^{-1})
Wheat (*Rabi*, 2021–2022)	I_1	Soil moisture sensor	N_0	0
	I_2	CWSI	N_1	50
	I_3	Conventional	N_2	100
			N_3	150
			N_4	200

a series of preprocessing techniques, such as splice correction, noise removal, and smoothening of spectra, have been applied. The spectral drifts at 1001 and 1831 nm are rectified using the splice correction function of ASD Viewspec Pro™. The noise-dominant bands that exist at the spectral range of 350–359 nm and 2451–2500 nm have been removed for further analysis. Finally, the preprocessed spectral data have been used for generating the spectral library of the collected field samples. Since a three-point sampling method has been adopted for the present study, 135 spectral reflectance curves are obtained from each sampling point of 45 plots. The spectral reflectance acquired for the wheat crop is shown in Figure 2.

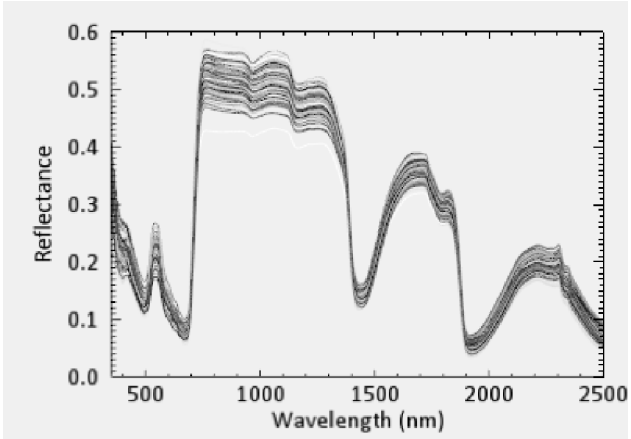


Figure 2. Spectral reflectance curves of the wheat crop using a ground-held spectroradiometer.

Table 2. Specifications of field and airborne hyperspectral sensor used for the study.

	Nano-Hyperspec Hyperspectral Sensor	ASD Field Spectroradiometer
No. of bands	269	2151
Spectral range	400–1000 nm	350–2500 nm
Spectral resolution	2.2 nm	3 nm at 700 nm, 6 nm at 1400 nm, 6 nm at 2100 nm
Frame rate	300 Hz	—
Spatial resolution	0.03 m	—
Signal to noise	>15:1 (1000 nm) < 140:1 (550 nm)	—
Total storage limit	480 GB	—
Focal length	4.8 m	—
Weight	1.2 kg	5.44 kg

UAV Imaging Using a Hyperspectral Sensor

Synchronized with leaf sample collection, spectral data collection has been prearranged, and the images of wheat fields are captured using a visual and near-infrared push-broom style hyperspectral UAV imaging camera with a lens of 4.8 mm. The hyperspectral imagery has been collected with a Headwall Nano-Hyperspec (Headwall Photonics Inc., Bolton, MA, USA) hyperspectral sensor mounted on a DJI Matrice 600 Pro UAV platform with Ronin gimbal. It captures data in 269 spectral bands ranging from 400 to 1000 nm with ~2.2 dispersion/pixel spectral interval (FWHM with 6-nm slit). The flight height has been adjusted to 21 m for getting a spatial resolution of 3.4 cm. The acquisitions of the hyperspectral images are performed between 11:00 AM and 12:00 PM in the daytime in sunny or lightly cloudy weather (cloud cover less than 20%) with relatively stable solar light. The detailed specifications of the hyperspectral sensor used for ground-based and airborne remote sensing are tabulated in Table 2.

Processing of UAV Captured Hyperspectral Images

The radiometric correction, reflectance conversion from digital number values, geometric correction, and mosaicking of the individual hyperspectral data cubes have been done using the designated software named SpectralView. The data was acquired 57 days after sowing of the wheat crop when it was in tillering stage. The orthorectification of reflectance hypercubes has been carried out using the data from Global Navigation Satellite System receivers, the inertial measurement unit, and the Shuttle Radar Topography Mission digital elevation model (Van Zyl 2001; Santos-Rufo *et al.* 2020).

A combination of NDVI threshold and spectral angle mapper (SAM) binary masks has been used for segmenting the wheat field canopy pixels from neighboring soil, shadow, and grass pixels (Oshigami *et al.* 2013). A primary mask image was generated by multiplying the NDVI threshold image with the SAM binary mask followed by application of the morphological erosion operator to remove the mixed pixels at the edges (Moghimi *et al.* 2018). The threshold for NDVI was chosen at the center between the vegetation peak of the wheat canopy and the neighboring trough in the NDVI histogram (Chancia *et al.* 2021). The SAM classification has been carried out using the endmember spectra of the wheat crop, soil, and grass extracted manually from the image. The generated spectral signatures have been smoothened for enhancing the spectral features using the Savitzky–Golay (SG filter) algorithm with a second-order polynomial and a window size of 5 (Savitzky and Golay 1964) (Ge *et al.* 2019). The endmember spectra corresponding to different leaf N values have been generated using the sequential maximum angle convex cone (SMACC) (Wei *et al.* 2019). The SMACC is a sequential algorithm based on a convex cone model (Gruninger *et al.* 2004). The process starts with a single endmember and shows an incremental rise in dimension until a specified number of endmembers is achieved. A convex cone was developed with extreme points and the existing (first) endmember, and an oblique projection was applied to it for deriving the next endmember. It also simultaneously generated abundance maps and was updated at each iterative step. The hyperspectral image with the wheat crop is shown in Figure 3a, and the corresponding extracted endmember spectral library is illustrated in Figure 3b.

Spectral Modeling

More recently, machine learning algorithms have shown remarkable results in the very same field, having exceptional capabilities to automatically learn the relationship between the data (here the spectra captured) and predict accordingly. After an exhaustive literature survey, five models were identified to calibrate between the reflectance spectra (acquired from the field and extracted from the hyperspectral UAV image) and leaf N concentration values. Namely, ANN, ELM, support vector machine regression (SVR), LASSO, and MLR are implemented and compared in order to identify a suitable model for leaf N prediction and mapping. Each data set is split into calibration (two-thirds of all observations) and validation (one-third of all observations) with corresponding N values. A brief description of the implemented models is given in the following sections.

ANN

This network mimics the capability of learning like the human brain, containing neurons as the most fundamental and functional unit. Its architecture contains input hidden and output layers that are fully connected. The input neurons act as computational unit and receive raw information; furthermore, the data pass to a hidden layer that transforms the data into a usable form to generate the output.

For every input x_i, the corresponding weight value w_i is multiplied. The one with higher influence on the output value is triggered:

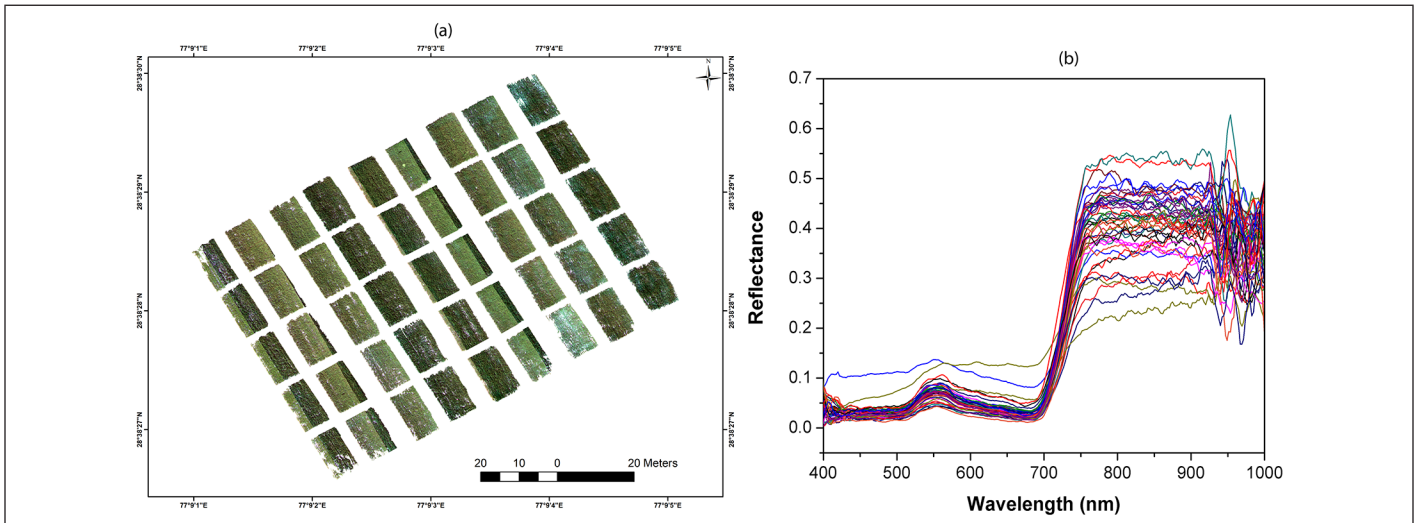


Figure 3. (a) Hyperspectral image of the wheat experimental field. (b) Spectral signatures of wheat crop derived from the image.

$$\Sigma = (x_1 \times w_1) + (x_2 \times w_2) + \ldots (x_n \times w_n) \quad (1)$$

where $x = [x_1, x_2, x_3, \ldots x_n]$ and $w = [w_1, w_2, w_3, \ldots w_n]$ are the row vectors belonging to input and the corresponding weights. Therefore, the dot product is given by

$$x \cdot w = (x_1 \times w_1) + (x_2 \times w_2) + \ldots (x_n \times w_n) \quad (2)$$

which turns out to be

$$\Sigma = x \cdot w \quad (3)$$

Bias b plays a major role in balancing the complete architecture; it often acts as an offset to move the activation function and produce the output value:

$$z = x \cdot w + b \quad (4)$$

This intermediate value that is generated is passed to an activation function that is nonlinear and influences the learning speed of the network. Multiple activation functions are used, depending on the application. The training mechanism includes back-propagating the error by computing the gradient values concerning the weight. The mean square error is calculated by the difference between the actual values and the predicted values of error. Later, the weights can be optimized, and hyperparameters, such as the minimum error, number of epochs, and learning rates, can be fixed. Tao *et al.* (2020) evaluated the accuracies of winter-wheat yield based on ground-measured plant height, which is extracted from UAV-based hyperspectral images using PLSR, ANN, and RF regression techniques (Yue *et al.* 2017; Tao *et al.* 2020). Another study aimed to estimate traits of rice, such as plant N content, leaf N accumulation, and plant N accumulation, using UAV-based hyperspectral data. Univariate regression models on vegetation indices and ANN, RF, and SVM were evaluated over different rice stages (Wang *et al.* 2021).

ELM

In 1994, G. B. Huang proposed the ELM with a single hidden layer that can overcome some major drawbacks, such as slow learning, iterative parameter tuning, and so on, of traditional gradient methods (Guang-Bin *et al.* 2006). The output function and hidden layer output function are given by Equation 5 and Equation 6, respectively,

$$f_L(x) = \sum_{i=1}^{L} \beta_i G(a_i, b_i, x) \quad (5)$$

$$h(x) = [G(a_1, b_1, x), \ldots, G(a_L, b_L, x)] \quad (6)$$

Here, x is the input vector and G is the activation function.

The training of the ELM happens by

$$f_L(x) = \sum_{i=1}^{L} \beta_i G(a_i, b_i, x) = H\beta \quad (7)$$

where H (hidden layer output matrix) is a column vector given by

$$H = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix} = \begin{bmatrix} h_1(x_1) & \cdots & h_L(x_1) \\ \vdots & \vdots & \vdots \\ h_1(x_N) & \cdots & h_L(x_N) \end{bmatrix}$$

Here, $\{(x_i, t_i) | x_i \in R^d, t_i \in R^m, i = 1, \ldots, N\}$ is the training set, $G(a, b, x)$ is the hidden node output function, N is the number of training samples, and L is the number of hidden nodes. The value of β is computed as

$$\beta = H^T T \quad (8)$$

$$H = (w_i, \ldots, w_N, b_i, \ldots, b_N, x_i, \ldots, x_N)$$

$$= \begin{bmatrix} g(w_i . x_i + b_i) & \cdots & g(w_N . x_i + b_N) \\ \vdots & \vdots & \vdots \\ g(w_i . x_N + b_i) & \cdots & g(w_N . x_N + b_N) \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_N^T \end{bmatrix} T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}$$

where T is training data target matrix.

Researchers have tried to exploit the advantage of fast learning process and higher training accuracy with low error percentage in order to estimate N content using ELM for rice crops (Yu *et al.* 2019, 2020). But there is much less evidence in the literature that similar studies are done with respect to wheat crops. As hyperspectral data set is voluminous, and the traditional gradient-based methods take an additional computation effort to perform regression.

SVM

SVM (Corinna and Vapnik 1995) is an extensively used method for hyperspectral data analysis. It maximizes the margin and generates a decision boundary with extreme separation between the data points of multiple classes. The decision boundary may be linear or nonlinear, depending on the separability analysis of the data used. It is a machine learning algorithm that utilizes labeled data, trains the associated model, and predicts unlabeled data.

The mathematical formulation of SVM includes formulas of lines in two dimensions and hyperplanes in multi-dimensions represented by

$$y = mx + b \quad (9)$$

$$\vec{w} \cdot \vec{x} + b = 0 \quad (10)$$

where x is sample data set for which SVM finds weights w such that the data points in the data set are separated using the most optimal hyperplane. The width is calculated by taking the dot product of the distance vector and the perpendicular vector w and then dividing by the magnitude of w:

$$\text{width} = \frac{1}{2} \cdot \frac{w^2}{\|w\|} \quad (11)$$

Since the width needs to be maximized, LaGrange multipliers are used (Shevade *et al.* 2000). Furthermore, L is differentiated with respect to w:

$$L = \frac{1}{2} \cdot \frac{w^2}{\|w\|} - \sum_i^n \left[y_i (\vec{w} \cdot \vec{x}_i + 1) - 1 \right] \quad (12)$$

$$\vec{w} = \sum_i^n a_i \cdot y_i \cdot \vec{x}_i \quad (13)$$

Differentiating L with respect to b gives

$$b = \sum_i^n a_i \cdot y_i = 0 \quad (14)$$

Substituting the value of w from Equation 13 into Equation 12 gives

$$L = \frac{1}{2} \left(\sum_i^n a_i \cdot y_i \cdot \vec{x}_i \right) \left(\sum_j^n a_j \cdot y_j \cdot \vec{x}_j \right) - \left(\sum_i^n a_i \cdot y_i \cdot \vec{x}_i \right) \left(\sum_j^n a_j \cdot j \cdot \vec{x}_j \right) - b \left(\sum_i^n a_i \cdot y_i \right) + \sum_i^n a_i \quad (15)$$

$$L = \sum_i^n a_i - \frac{1}{2} \sum_i^n \sum_j^n a_i a_j y_i y_j \vec{x_i} \cdot \vec{x_j} \quad (16)$$

Since the above derivation deals with linearly separable data points but the practical problems involve nonlinear boundaries to establish separation, a kernel trick is required, and the general equation is

$$k(x, y) = x^T y + c \tag{17}$$

SVM has already proven its efficacy for hyperspectral data, which is also evident from the literature. Various studies estimated the leaf chlorophyll content of maize and wheat and comprehensively examined the effects of the spectral and spatial information of UAV imagery and the effects of phenotype and phenology on N estimation using SVM (Zhu *et al.* 2020; Wang *et al.* 2021). Lee *et al.* (2020b) used linear regression RF and SVM with UAV multispectral images to predict canopy N weight in corn. (see also Zha *et al.* 2020).

LASSO

The LASSO regression is a nonparametric model that shows an ability to eliminate nonsignificant weak predicting variables and thereby improves prediction accuracy (Chen *et al.* 2012). It manages the significant variables that control the dependent parameter and shrinks regression coefficients to zero if no improvement is perceived in model performance (Tibshirani 1996). Those variables with a regression coefficient of zero after shrinkage are excluded from the model (Ranstam and Cook 2018). LASSO regression undergoes L1 regularization, adding a penalty equivalent to absolute values of the magnitude of coefficients. In order to produce simple models, larger penalties generate values close to zero, and some coefficients were eliminated from the model. The mathematical equation of LASSO regression is

$$\sum_{i=1}^{n}\left(y_i - \sum_j x_{ij}\beta_j\right)^2 + \lambda\sum_{j=1}^{p}\left|\beta_j\right| \tag{18}$$

where λ denotes the amount of shrinkage, n is the number of instances, p is the number of features, x is the covariate matrix, and β is the coefficient matrix.

Wheat grain yield was predicted using multiple vegetative indices with LASSO as a feature selection technique for multispectral UAV data (Li *et al.* 2019; Shafiee *et al.* 2021). Comparative assessment of 12 different models was done for LandSat-derived total N with UAV-based image data to do water quality monitoring.

MLR

In MLR, a linear relationship between multiple dependent and independent variables is used to generate appropriate equations for predicting the variable of interest (Besalatpour *et al.* 2012). The global regression model can be represented by

$$Y = c + a_1x_1 + a_2x_2 + a_3x_3 + \ldots + a_nx_n \tag{19}$$

where y is the dependent variable, c is a constant, a_1, a_2, a_3, $\ldots a_n$ are coefficients, and x_1, x_2,$\ldots x_n$ are independent variables.

Model Evaluation

The performance of the predictive models has been assessed using the coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and the ratio of performance deviation (RPD). Higher R^2 and lower RMSE and MAE indicate the best performance for prediction models. Such a model will be applied to the whole data set for generating the N prediction map. Higher RPD values, such as 1.4–2 and greater than 2, indicate good and very good predictions (Chang *et al.* 2001). All these statistical indices are generated using R software (version 4.2, R Development Core Team, 2018). Equations 20–23 are used for calculating R^2, RMSE, MAE, and RPD:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(x_i - y_i\right)^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}_i\right)^2} \tag{20}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - y_i\right)^2} \tag{21}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|x_i - y_i\right| \tag{22}$$

$$\text{RPD} = \frac{\text{SD}}{\text{RMSE}} \tag{23}$$

Here, n is the total sample size, x_i is the ith measured leaf N value, y_i is the ith predicted leaf N value using a suitable model, and $\bar{x}_i$ is the ith mean measured leaf N value. Generally, higher R^2 values denote better performance for the predictive models. But the model having higher R^2 values does not always possess low RMSE or higher RPD. To avoid such circumstances, each model is assigned a rank based on the multi-criteria decision-making rule (Das *et al.* 2020; Mahajan *et al.* 2021). In the present study, each model is assigned a rank based on each statistical index following "more is better" and "less is better" rules. Statistical parameters such as R^2, RPD, and the "more is better" rule were applied, and for parameters such as RMSE, the "less is better" rule was applied. Finally, the ranks were added to get a composite rank suggesting the best-performing model. A flowchart of the proposed approach is shown in Figure 4.

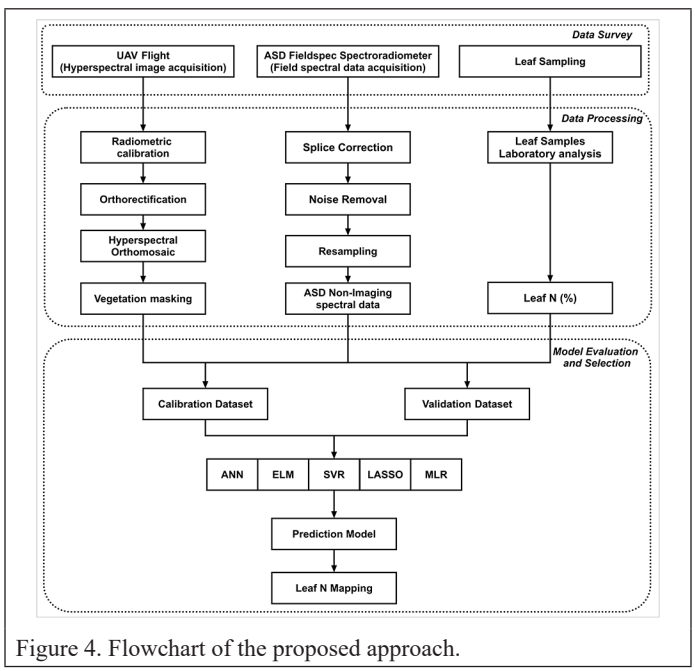


Figure 4. Flowchart of the proposed approach.

Results and Discussion

Descriptive Statistics of Leaf N

A descriptive statistical analysis of leaf N for the complete, calibration, and validation data set is given in Table 3. The leaf N values varied from 3.17% to 5.72% with an average of 4.78% for the complete data set. The values for the calibration and validation data sets varied from 3.17% to 5.69% with an average of 4.80% and from 4.06% to 5.72% with an average of 4.75%, respectively. In terms of the coefficient of variation (CV), all three data sets show less variability. However, the calibration data set shows the highest variability (CV = 12.54%). The complete data set and calibration data set show negative skewness, while the validation data set seems to be positively skewed. The positive skewness of the validation data set implies that the observations are more concentrated toward the lower values and consist of very few higher values. The overall statistical analysis confirmed a good variability among the three data sets, suggesting the usage of these data sets for calibrating and validating the prediction models.

Spectral Characteristics of Leaves with Varying N Content

Leaf pigmentation, water content, leaf anatomy, and nutritional status are the major factors affecting the spectral reflectance of leaves.

Table 3. Descriptive statistical parameters of leaf nitrogen (%) in the study area.

	Sample Size	Minimum	Maximum	Mean	Median	SD	% CV	Skewness	Kurtosis
Whole data set	45	3.17	5.72	4.78	4.85	0.56	11.62	−0.49	0.19
Calibration data set	30	3.17	5.69	4.80	4.94	0.60	12.54	−0.77	0.17
Validation data set	15	4.06	5.72	4.75	4.69	0.46	9.60	0.91	0.69

CV = coefficient of variation; SD = standard deviation.

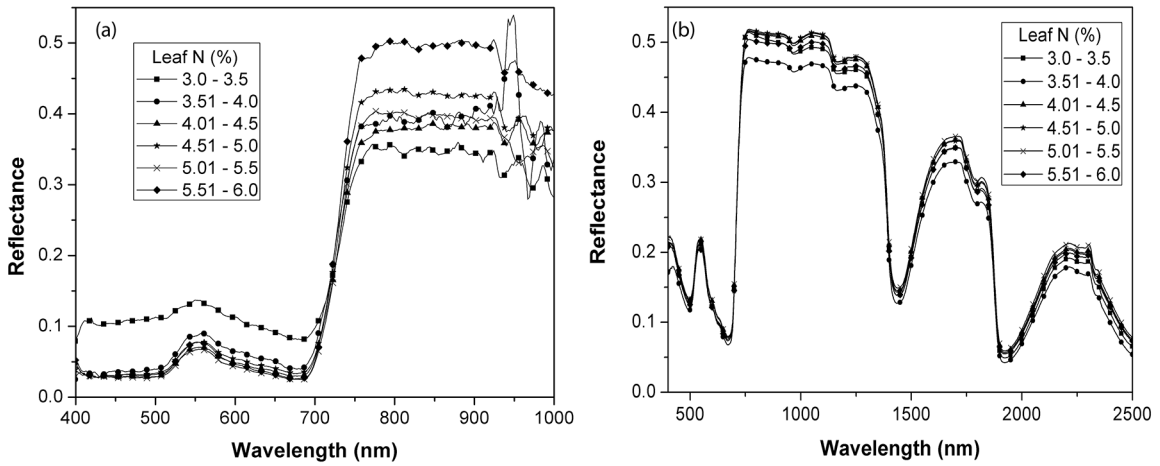


Figure 5. Spectral characteristics of leaf nitrogen (N). (a) Imaging. (b) Non-imaging.

All mature, healthy leaves show high reflectance in the NIR region (700 nm < λ <1500 nm), reduced reflectance in the green region, and very low reflectance in the ultraviolet, blue, and red regions. In low-reflectance regions, the major leaf pigments, such as chlorophyll, xanthophyll, and carotenoids, absorb the energy completely for photosynthetic activity. Thus, the spectrally active constituents, such as leaf pigment concentrations, depend on the leaf N concentration and were thereby responsible for getting typical spectral responses. Low N content in the leaf caused a decrease in chlorophyll and consequently reduced the absorption of radiation, which in turn caused high spectral reflectance in that region. The spectral characteristics of leaves with varying N content generated using the image-derived endmember spectra and spectroradiometer-based spectra data are shown in Figure 5a and 5b. With an increase in N deficiency, reflectance increases from 500 to 700 nm which is due to the absorption in this region is highly affected by pigment concentration which depends on N concentration. Roy (1989) reported an increase in reflectance from 500 to 700 nm in Coleus leaves for decrease in plant N. Al-Abbas *et al.* (1974) also reported high reflectance in maize leaves in the range of 530–750 nm under N-deficient treatments.

Spectral Modeling of Leaf N

As mentioned, four machine learning models (ANN, ELM, LASSO, and SVR) were analyzed and implemented for the prediction and estimation of leaf N content in the considered experimental field. The study is threefold: (1) endmembers derived from the hyperspectral image are used, and (2) spectral signatures acquired in the field were used for modeling followed by (3) selection of a suitable model for prediction and analysis. Furthermore, MLR is used for the modeling equation to generate an in-field N map using hyperspectral images acquired by UAV.

Modeling of UAV Image Data for Leaf N

After the endmember spectra corresponding to different leaf N values have been generated using SMACC and background pixels are removed from the image, calibration and validation data sets (70:30) are given as input to the regression models. The results of the models have been contrastively analyzed and are tabulated in Table 4 followed by the corresponding regression plots, which were shown in Figure 6 (see next page). RMSE, MAE, R^2, and RPD are calculated for all considered models and calibration and validation data sets.

ANN is a widely used regression method for field spectral data, but it is yet to be exploited for UAV data. For the present work, it performs well with R^2 values of 0.99 and 0.97 for the calibration and validation sets, respectively, as tabulated in Table 4 and Figure 6a and 6b. The number of neurons is taken as 10, with the back-propagation algorithm, the sigmoid activation function, and learning rate set as 0.01. The number of neurons in the hidden layer is set to be same, as the training data are small. An early stopping criterion, at 20 epochs, is considered in case there is no remarkable improvement in accuracy. Other activation functions, such as Bayesian regularization, were also applied, yielding less accuracy. The learning rate is selected by random initialization, which helps control the step size to reach minimum loss, further reducing its value to 0.01, and gives a better chance to find the minimum error value. Although analogous studies are carried out with different types of crops and crop assessment parameters, there is still scope to assess leaf N using airborne hyperspectral data (Wang *et al.* 2021). Tao *et al.* (2020) have proposed to estimate yield and plant height using spectral indices, ground-measured plant height, and UAV hyperspectral data with maximum $R^2 = 0.77$ among PLSR, ANN, and RF at different growth stages.

The hyperparameters for the models are set according to the hit-and-trial method to attain maximum accuracy. According to the rank-based multi-criteria decision-making rule, ELM has performed remarkably well in the case of endmembers extracted from the image. The number of hidden neurons is set to be 1000 with the "purelin" activation function and distribution, from which the input weights and the bias should be initialized in the range of [−1, 1]. The calibration $R^2 = 0.998$, whereas the validation $R^2 = 0.968$ for the current study (refer to Table 4 and Figure 6c and 6d). A similar approach is followed for rice crops to measure canopy chlorophyll content using UAV hyperspectral imaging using particle swarm optimization (PSO) as a feature selector and ELM as a regression model with $R^2 = 0.791$ (Cao *et al.* 2020). Furthermore, Genetic Algorithm-ELM (GA-ELM) and Whale Optimization Algorithm-ELM (WOA-ELM) are explored on a hybrid basis with $R^2 = 0.887$. Yu *et al.* (2019) used UAV hyperspectral inversion modeling of rice N content based on WOA-ELM. Similarly, rice canopy N content is estimated using UAV-based hyperspectral image data by PSO-ELM and Beetle Antennae Search (BAS-ELM) with an R^2 value of 0.86 for calibration data set. The high performance of ELM is evident

Table 4. Accuracy assessment of regression models using unmanned aerial vehicle–based hyperspectral data.

Sequence No.	Model	Calibration				Validation			
		RMSE	MAE	R^2	RPD	RMSE	MAE	R^2	RPD
1	ANN	0.029	0.022	0.998	21.059	0.074	0.059	0.978	5.595
2	ELM	0.026	0.019	0.998	22.694	0.080	0.066	0.968	5.525
3	LASSO	0.059	0.043	0.994	9.592	0.160	0.130	0.892	2.247
4	SVR	0.054	0.051	0.994	10.524	0.122	0.103	0.938	3.237

ANN = artificial neural network; ELM = extreme learning machine; LSSO = least absolute shrinkage and selection operator; MAE = mean absolute error; RMSE = root mean square error; RPD = ratio of performance deviation; SVR = support vector machine regression.

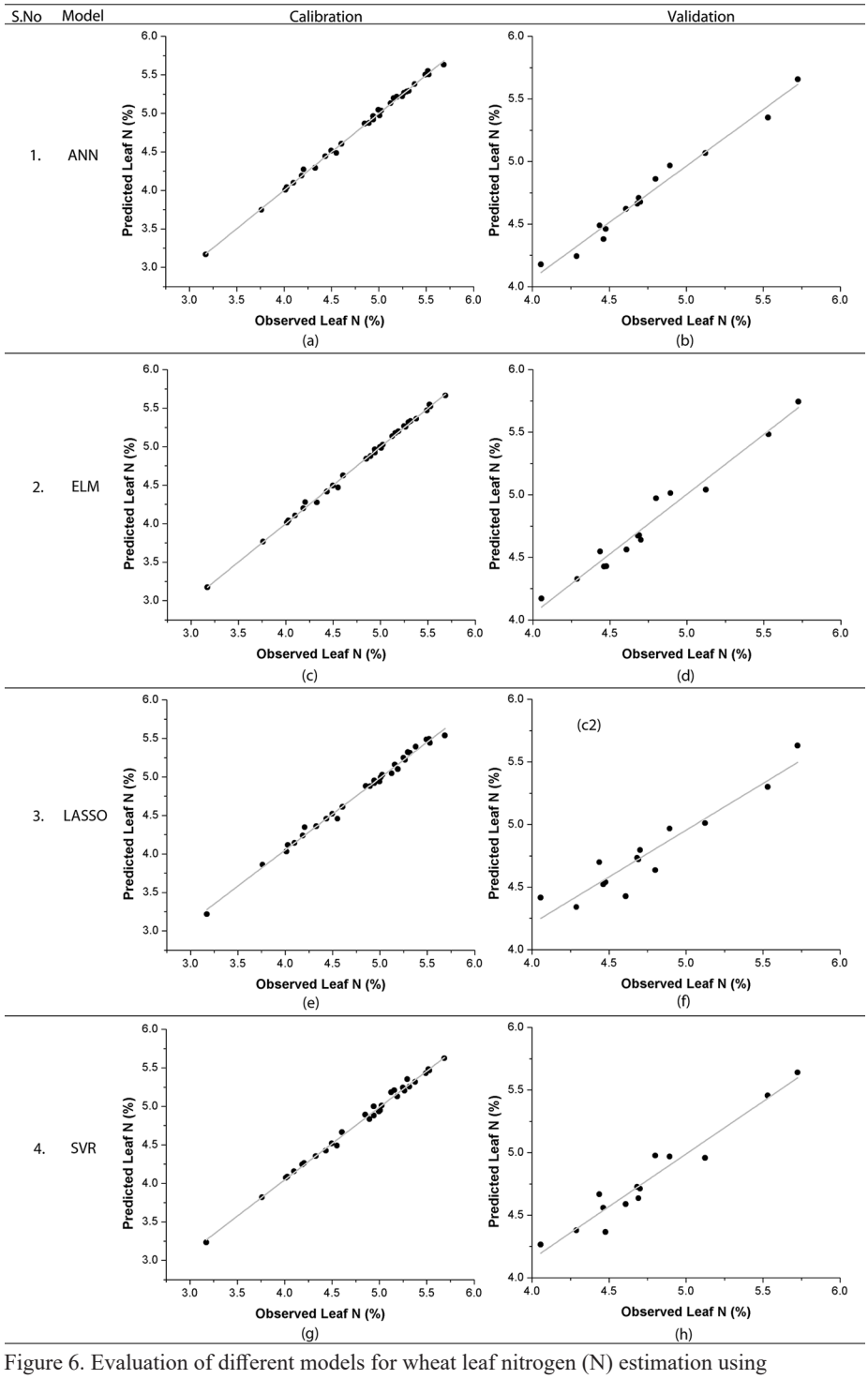


Figure 6. Evaluation of different models for wheat leaf nitrogen (N) estimation using hyperspectral images from an unmanned aerial vehicle.

from the literature as well due to its fast and efficient learning speed, fast convergence, good generalization ability, and ease of implementation (Chen *et al.* 2017).

The LASSO technique is extensively used as a feature extraction method for selecting the significant bands in the literature (Omidi *et al.* 2020) and with multispectral UAV imagery (Barzin *et al.* 2021). In the proposed approach, it is used as a regression method where the coefficients and the parameter of the model are selected through the ordinary least squares method. This method works by minimizing the sum of squares of residuals. It gives R^2 as 0.99 for validation and 0.89 for calibration, as shown in Table 4 and Figure 6e and 6f.

SVR considers the observations within the decision boundary with the best-fit hyperplane. The "radial" kernel function is used with a cost of 1000 and a gamma of 0.0001 taken as hyperparameters for the study. The main parameter to tune in case of SVR is the kernel, as it maps the observations into a suitable feature space to make them separable. Furthermore, Hsu *et al.* (2003) referred to choosing cost and gamma values. R^2 values of 0.99 and 0.89 are achieved for the calibration and validation sets, respectively (Table 4 and Figure 6g and 6h).

Modeling of Ground Spectra Using Spectroradiometer for Leaf N

Four regression techniques for estimating leaf N content in wheat were compared by spectral features derived from field data acquisition. The results are given in Table 5, and the simultaneous regression plots are shown in Figure 7 (see next page).

The calibration R^2 lies between 0.9818 and 0.9894, and the validation R^2 ranges from 0.7353 to 0.9349 for all the four models. It is evident from the results that ANN is performing better than other three models considered. Although the calibration R^2 is comparable for all models, the validation R^2 is low in the case of LASSO.

Also, as compared to the endmembers derived from the UAV hyperspectral image data, the R^2 values are less in the case of field spectral signatures. The reason articulated for the same may be that the ground data are influenced by multiple atmospheric hindrances, such as weather, temperature, humidity, and so on. Therefore, the spectral signatures acquired in field conditions might not be similar to the spectral signatures derived from the image itself (Yadav *et al.* 2018).

MLR for Leaf N Mapping

Stepwise MLR model has been implemented based on the following sensitive wavelengths: 399, 520, 668, 691,767, 774, 803, 827, 830, 848, 904, and 922 nm. The R^2 value of the model is 0.967, and the adjusted R^2 is 0.943. The model equation is

Table 5. Accuracy assessment of regression models using field hyperspectral data.

Sequence No.	Model	Calibration				Validation			
		RMSE	MAE	R^2	RPD	RMSE	MAE	R^2	RPD
1	ANN	0.061	0.044	0.989	9.761	0.117	0.084	0.935	3.521
2	ELM	0.074	0.057	0.984	8.082	0.160	0.110	0.874	2.656
3	LASSO	0.088	0.067	0.982	6.374	0.228	0.157	0.735	1.580
4	SVR	0.081	0.066	0.982	7.105	0.143	0.099	0.906	2.783

ANN = artificial neural network; ELM = extreme learning machine; LSSO = least absolute shrinkage and selection operator; MAE = mean absolute error; RMSE = root mean square error; RPD = ratio of performance deviation; SVR = support vector machine regression.

$$y = 4.4 + (5.9*B_{399}) - (34.17*B_{520}) + (139.3*B_{668}) - (123.7*B_{691}) + (56.11*B_{767}) - (62.81*B_{774}) + (51.37*B_{803}) + (40.44*B_{827}) - (72.7*B_{830}) - (19.96*B_{848}) + (6.94*B_{904}) + (4.81*B_{922})$$

The above equation is used for prediction of leaf N content in the field using the hyperspectral image acquired by UAV. Figure 8 shows the N variability in the field according to the experimentation carried out. The values range from 0.9% to 6.69% with the maximum of them lying between 3.51% and 5.02%.

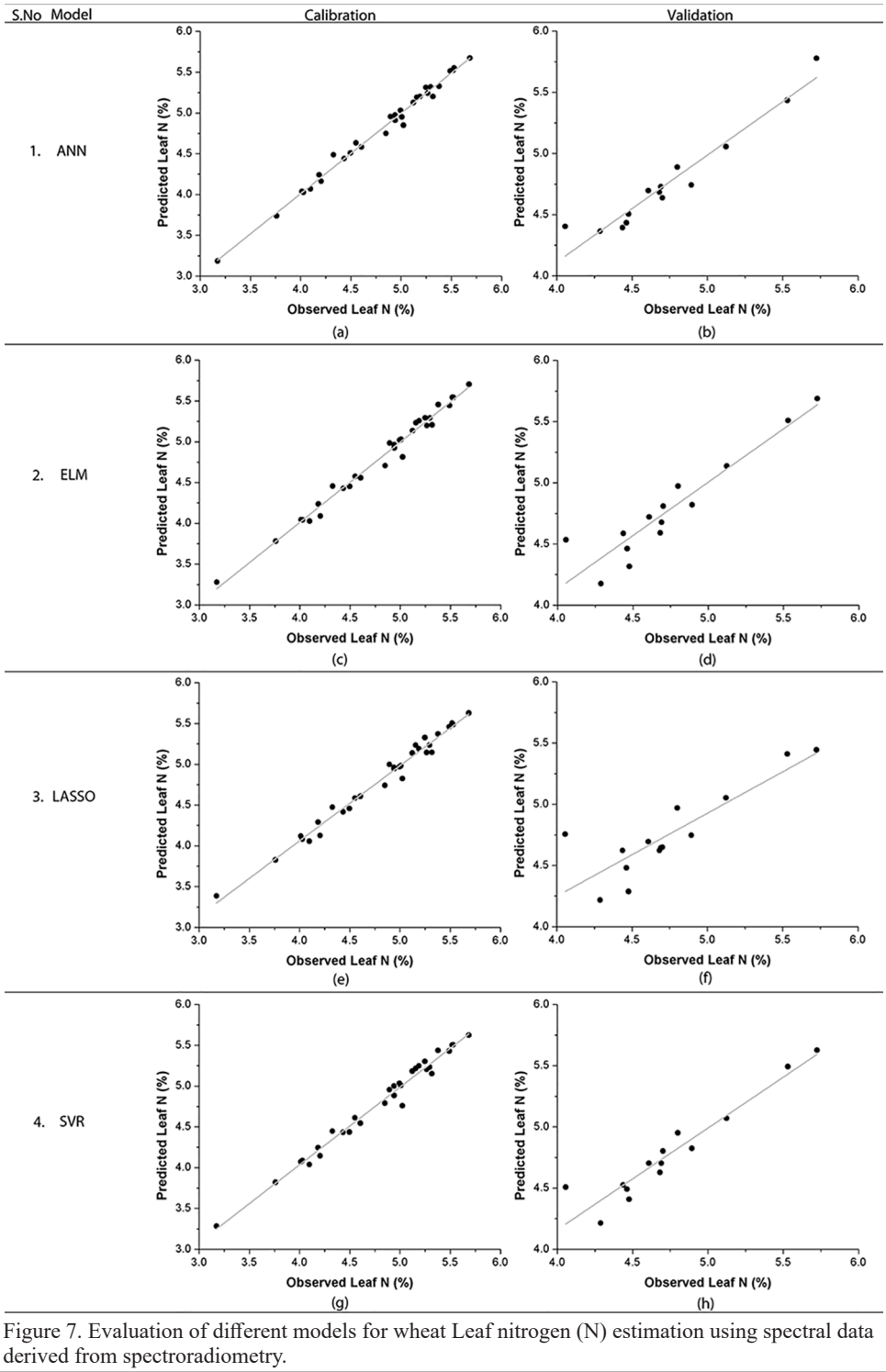


Figure 7. Evaluation of different models for wheat Leaf nitrogen (N) estimation using spectral data derived from spectroradiometry.

Conclusions

Wheat crop is studied for the estimation of leaf N through imaging and nonimaging spectroscopy. The development of hyperspectral remote sensing has enabled quantitative assessment of N by the extraction of a treasure of information associated with a particular spectral signature. The proposed study is twofold. First, among the four models considered, ELM has outdone with calibration performance of RMSE = 0.02, MAE = 0.019, RPD = 22.69, and R^2 = 0.99 and validation performance of RMSE = 0.08, MAE = 0.06, RPD = 5.52, and R^2 = 0.96 in the case of imaging data from the UAV. wherein ANN has performed well for nonimaging data from field spectroradiometry with calibration performance of RMSE = 0.06, MAE = 0.04, RPD = 9.76, and R^2 =0.98 and validation performance of RMSE = 0.117, MAE = 0.08, RPD = 3.52, R^2 = 0.93. Second, the machine learning measures in this stage make use of reference signatures derived from the image and field. The comprehensive analysis shows that the image-based spectral signatures have produced higher performance, whereas the field-based spectral data set, due to intervening effects of background illumination, has yielded low results. The other reason articulated for this inference is that the image-derived spectral signatures are tightly related to the imagery acquired from the UAV, and the challenge of insufficient spectral signatures is not there to give an optimal endmember spectrum. Furthermore, stepwise MLR is also used to derive significant spectral bands of 399, 520, 668, 691,767, 774, 803, 827, 830, 848, 904, and 922 nm with an adjusted R^2 of 0.943. The present study encourages monitoring small to medium-sized fields using UAVs as a viable solution for crop health management.

Acknowledgments

The research activity was conducted under the financial support of Indian Council of Agricultural Research (ICAR) and is hereby duly acknowledged under the project titled, "Network Program on Precision Agriculture (NePPA)".

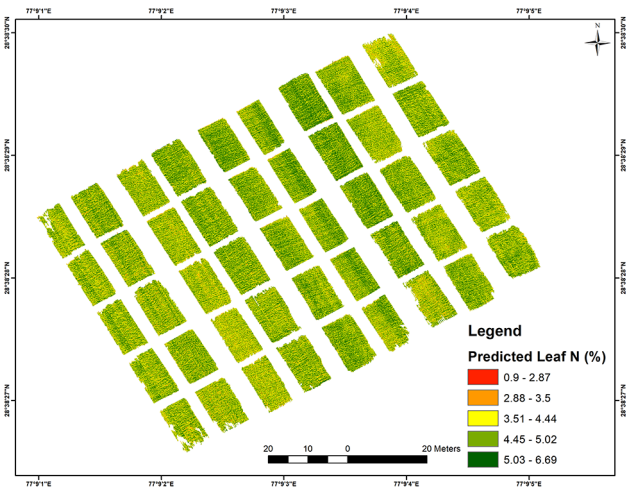


Figure 8. Wheat leaf nitrogen (N) variability map using multiple linear regression.

References

Ahirwar, S., R. Swarnkar S. Bhukya, and G. Namwade G. 2019. Application of drone in agriculture. *International Journal of Current Microbiology and Applied Sciences* 8(1):2500–2505.

Al-Abbas, A. H., R. Barr, J. D. Hall, F. L. Crane and M. F. Baumgardner. 1974. Spectra of normal and nutrient-deficient maize leaves. *Agronomy Journal* 66(1):16–20.

Anuj, K. R. Singh, G. Singh, R. K. Sharma, M. S. Saharan, R. S. Chhokar, B. S. Tyagi, R. Sendhill, R., Chand and I. Sharma. 2014. *Wheat Cultivation in India (Pocket Guide)*. Karnal, India: ICAR-Directorate of Wheat Research.[[comp: Please stet "Sendhill, R." in the author line.]]

Barrientos, A., J. Colorado, J. del Cerro, A. Martinez, C. Rossi, D. Sanz and J. Valente. 2011. Aerial remote sensing in agriculture: A practical approach to area coverage and path planning for fleets of mini aerial robots. *Journal of Field Robotics* 28(5):667–689.

Barzin, R., H. Kamangir and G. C. Bora. 2021. Comparison of machine learning methods for leaf nitrogen estimation in corn using multispectral UAV images. *Transactions of the ASABE* 64(6):2089—2101.

Besalatpour, A., M. A. Hajabbasi, S. Ayoubi, M. Afyuni, A. Jalalian and R.J.S.S. Schulin. 2012. Soil shear strength prediction using intelligent systems: Artificial neural networks and an adaptive neuro-fuzzy inference system. *Soil Science and Plant Nutrition* 58(2):149–160.

Cao, Y. J. Kailun, W Jingxian, Y. Fenghua, D. Wen and X. Tongyu. 2020. Inversion modeling of japonica rice canopy chlorophyll content with UAV hyperspectral remote sensing. *PLoS One* 15(9):e0238530.

Chancia, R. T. Bates, J. Vanden Heuvel and J. van Aardt. 2021. Assessing grapevine nutrient status from unmanned aerial system (UAS) hyperspectral imagery. *Remote Sensing* 13(2):4489.

Chang, C. W., D. A. Laird, M. J. Mausbach and C. R Hurburgh. 2001. Near-infrared reflectance spectroscopy–principal components regression analyses of soil properties. *Soil Science Society of America Journal* 65(2):480–490.

Chen, C. L. Kenli, D. Mingxing and L. Keqin. 2017. Extreme learning machine and its applications in big data processing. In *Big Data Analytics for Sensor-Network Collected Intelligence*, 117–150. New York: Academic Press.

Chen, H., Y. Lan, B. K. Fritz, W. C. Hoffmann and S. Liu. 2021. Review of agricultural spraying technologies for plant protection using unmanned aerial vehicle (UAV). *International Journal of Agricultural and Biological Engineering* 14(1):38–49.

Chen, Y., W. Ma, X. Wang and C. Zhao. 2012. Relationship between soil nutrient and wheat yield based on PLS. *Nongye Jixie Xuebao*. *[Transactions of the Chinese Society for Agricultural Machinery]* 43(2):159–164.

Corinna, C. and V. Vapnik. 1995. Support-vector networks. *Machine Learning* 20(3):273–297.

Das, B. S., M. C. Sarathjith, P. Santra, R. N. Sahoo, R. Srivastava, A. Routray and S. S. Ray. 2015. Hyperspectral remote sensing: Opportunities, status and challenges for rapid soil assessment in India. *Current Science* 108(5):860–868.

Das, B., K. K. Manohara, G. R. Mahajan and R. N. Sahoo. 2020. "pectroscopy based novel spectral indices: PCA-and PLSR-coupled machine learning models for salinity stress phenotyping of rice. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 229:117983.

Ge, X. W. Jingzhe, D. Jianli, C. Xiaoyi, Z. Zipeng, L. Jie and L. Xiaohang. 2019. Combining UAV-based hyperspectral imagery and machine learning algorithms for soil moisture content monitoring. *PeerJ* 7:6926.

Gruninger, J. H., A. J. Ratkowski and M. L. Hoke. 2004. The sequential maximum angle convex cone (SMACC) endmember model. In Algorithms and technologies for multispectral, hyperspectral, and ultraspectral imagery." *SPIE* 525:1–14.

Guang-Bin, H., Z. Qin-Yu and S. Chee-Kheong. 2006. Extreme learning machine: Theory and applications. *Neurocomputing* 70(1–3):489–501.

Guebel, D. V., B. C. Nudel and A. M. Giulietti. 1991. A simple and rapid micro-Kjeldahl method for total nitrogen analysis. *Biotechnology Techniques* 5(6):427–430.

Hank, T. B., K. Berger, H. Bach, J.G.P.W. Clevers, A. Gitelson, P. Zarco-Tejada and W. Mauser. 2019. Spaceborne imaging spectroscopy for sustainable agriculture: Contributions and challenges. *Surveys in Geophysics* 40(3):515–551.

Hsu, Chih-Wei, Chih-Chung Chang, and Chih-Jen Lin. "A practical guide to support vector classification." (2003): 1396-1400.

Lee, H., J. Wang and B. Leblon. 2020a. Intra-field canopy nitrogen retrieval from unmanned aerial vehicle imagery for wheat and corn fields. *Canadian Journal of Remote Sensing* 46(4):454–472.

Lee, H, J. Wang and B. Leblon. 2020b. Using linear regression Random Forests and Support Vector Machine with unmanned aerial vehicle multispectral images to predict canopy nitrogen weight in corn. *Remote Sensing* 12(13):2071.

Li, J., A.-N. Veeranampalayam-Sivakumar, M. Bhatta, N. D. Garst, H. Stoll, P. S. Baenziger, V. Belamkar, R. Howard, Y. Ge and Y. Shi. 2019. Principal variable selection to explain grain yield variation in winter wheat from features extracted from UAV imagery. *Plant Methods* 15(1):1–13.

Liu, H., Z. Hongchun and W. Ping. 2017. Quantitative modelling for leaf nitrogen content of winter wheat using UAV-based hyperspectral data. *International Journal of Remote Sensing* 38(8–10):2117–2134.

Mahajan, G. R., D. Bappa, G. Bhaskar, M. Dayesh, D. Ashwini, M. Shaiesh, P. K. Puna and K. R. Mukund. 2021. Monitoring properties of the salt-affected soils by multivariate analysis of the visible and near-infrared hyperspectral data. *Catena* 198:105041.

Moghimi, A., C. Yang, M. E. Miller, S. F. Kianian and P. M. Marchetto. 2018. A novel approach to assess salt stress tolerance in wheat using hyperspectral imaging. *Frontiers in Plant Science*. https://doi.org/10.3389/fpls.2018.01182.

Omidi, R., A. Moghimi, A. Pourreza, M. El-Hadedy and A. S. Eddin. 2020. Ensemble hyperspectral band selection for detecting nitrogen status in grape leaves. Pages 286–293 in 19th *IEEE International Conference on Machine Learning and Applications (ICMLA)*, held in. 14-17 December 2020 Edited by Wani, A., F. Luo, X (Andy). Li, D. Dou, and F. Bonchi. Virtual event: IEEE Computer Society Conference Publishing Services (CPS).

Oshigami, S., Y. Yasushi, U. Tatsumi, M. Atsushi, A. Yessy, K. Yuu, Y. Taro, M. Shuichi and N. Anna. 2013. Mineralogical mapping of southern Namibia by application of continuum-removal MSAM method to the HyMap data. *International Journal of Remote Sensing* 34(15):5282–5295.

Pandey, P. C., H. Balzter, P. K. Srivastava, G. P. Petropoulos and B. Bhattacharya. 2020. Future perspectives and challenges in hyperspectral remote sensing. In *Hyperspectral Remote Sensing: Theory and Applications*, 429–439. Amsterdam: Elsevier.

Ranjan, R., U. K. Chopra, R. N. Sahoo, A. K. Singh and S Pradhan. 2012. Assessment of plant nitrogen stress in wheat (*Triticum aestivum* L.) through hyperspectral indices. *International Journal of Remote Sensing* 33(20):6342–6360.

Ranstam, J. and J. A. Cook. 2018. LASSO regression. *Journal of British Surgery* 105(10):1348–1348.

Rao Mogili, U.M. and B.B.V.L. Deepak. 2018. Review on application of drone systems in precision agriculture. *Procedia Computer Science* 133:502–509.

Roy, P. S. 1989. Spectral reflectance characteristics of vegetation and their use in estimating productive potential. *Proceedings/Indian Academy of Sciences* 99:59–81.

Santos-Rufo, A., F.-J. Mesas-Carrascosa, A. García-Ferrer and J. Emilio Meroño-Larriva. 2020. Wavelength selection method based on partial least square from hyperspectral unmanned aerial vehicle orthomosaic of irrigated olive orchards. *Remote Sensing* 12(20):3426.

Savitzky, A. and M. J. Golay. 1964. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry* 36(8):1627–1639.

Shafiee, S., L. M. Lied, I. Burud, J. A. Dieseth, M. Alsheikh and M. Lillemo. 2021. Sequential forward selection and support vector regression in comparison to LASSO regression for spring wheat yield prediction based on UAV imagery. *Computers and Electronics in Agriculture* 183:106036.

Shanmugapriya, P., S. Rathika, T. Ramesh and P. Janaki. 2019. Applications of remote sensing in agriculture—A review. *International Journal of Current Microbiology and Applied Sciences* 8(1):2270–2283.

Shevade, S. K., S. S. Keerthi, C. Bhattacharyya and K.R.K. Murthy. 2000. Improvements to the SMO algorithm for SVM regression. *IEEE Transactions on Neural Networks* 11(5):1188–1193.

Spachos, P. and S. Gregori. 2019. Integration of wireless sensor networks and smart UAVs for precision viticulture. *IEEE Internet Computing*23(3):8–16.

Spiertz, J.H.J. 2009. Nitrogen, sustainable agriculture and food security: A review. In *Sustainable Agriculture*, edited by E. Lichtfouse, M. Navarrete, P. Debaeke, S. Véronique and C. Alberola, 535–651. Dordrecht: Springer.

Tang, L. and G. Shao. 2015. Drone remote sensing for forestry research and practices. *Journal of Forestry Research* 26(4):791–797.

Tao, Huilin, F. Haikuan, X. Liangji, M. Mengke, Y. Guijun, Y. Xiaodong and F. Lingling. 2020. Estimation of the yield and plant height of winter wheat using UAV-based hyperspectral images. *Sensors* 20(4):1231.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society* 58(1):267–288.

Tu, Y.-H., S. Phinn, K. Johansen, A. Robson and D. Wu. 2020. Optimising drone flight planning for measuring horticultural tree crop structure. *ISPRS Journal of Photogrammetry and Remote Sensing* 160:83–96.

Van Zyl, J. J. 2001. The Shuttle Radar Topography Mission (SRTM): Abreakthrough in remote sensing of topography. *Acta Astronautica* 48(5–12):559–565.

Wang, L., S. Chen, D. Li, C. Wang, H. Jiang, Q. Zheng and Z. Peng. 2021. Estimation of paddy rice nitrogen content and accumulation both at leaf and plant levels from UAV hyperspectral imagery. *Remote Sensing* 13(15):2956.

Wei, L., M. Yu, Y. Liang, Z. Yuan, C. Huang, R. Li and Y. Yu. 2019. Precise crop classification using spectral-spatial-location fusion based on conditional random fields for UAV-borne hyperspectral remote sensing imagery. *Remote Sensing* 11(17):2011.

Xiang, H. and L. Tian. 2011. Development of a low-cost agricultural remote sensing system based on an autonomous unmanned aerial vehicle (UAV). *Biosystems Engineering* 108(2):174–190.

Yadav, D., M. K. Arora, K. C. Tiwari and J. K. Ghosh. 2018. Parameters affecting target detection in VNIR and SWIR range. *Egyptian Journal of Remote Sensing and Space Science* 21(3):325–333.

Yang, M., M. A. Hassan, K. Xu, C. Zheng, A. Rasheed, Y. Zhang, X. Jin, X. Xia, Y. Xiao and Z. He. 2020. Assessment of water and nitrogen use efficiencies through UAV-based multispectral phenotyping in winter wheat. *Frontiers in Plant Science*. https://doi.org/10.3389/fpls.2020.00927.

Yang, S., L. Hu, H. Wu, H. Ren, H. Qiao, P. Li and W. Fan. 2021. Integration of crop growth model and random forest for winter wheat yield estimation from UAV hyperspectral imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14:6253–6269.

Yu, F., W. Du, Z.Guo, C. Zhou, D. Wang and T. Xu. 2019. UAV hyperspectral inversion modeling of rice nitrogen content based on WOA-ELM. *International Journal of Precision Agricultural Aviation* 2(2):43–48.

Yu, F., S. Feng, W. Yao, D. Wang, S. Xing and T. Xu. 2020. BAS-ELM based UAV hyperspectral remote sensing inversion modeling of rice canopy nitrogen content. *International Journal of Precision Agricultural Aviation* 3(3):59–64.

Yuan, H., G. Yang G. Li, Y. Wang, J. Liu, H. Yu, H. Feng, B. Xu , X. Zhao and X. Yang. 2017. Retrieving soybean leaf area index from unmanned aerial vehicle hyperspectral remote sensing: Analysis of RF, ANN, and SVM regression models. *Remote Sensing* 9(4):309.

Yue, J., G. Yang, C. Li, Z. Li, Y. Wang, H. Feng and B. Xu. 2017. Estimation of winter wheat above-ground biomass using unmanned aerial vehicle-based snapshot hyperspectral sensor and crop height improved models. *Remote Sensing* 9(7):708.

Zha, H., Y. Miao, T. Wang, Y. Li, J. Zhang, W. Sun, Z. Feng and K. Kusnierek. 2020. Improving unmanned aerial vehicle remote sensing-based rice nitrogen nutrition index prediction with machine learning. *Remote Sensing* 12(2):215.

Zheng, H., T. Cheng, D. Li, X. Yao, Y. Tian, W. Cao and Y. Zhu. 2018. Combining unmanned aerial vehicle (UAV)-based multispectral imagery and ground-based hyperspectral data for plant nitrogen concentration estimation in rice." *Frontiers in Plant Science* 9:936.

Zhu, W., Z. Sun, T. Yang, J. Li, J. Peng, K. Zhu, L. Shiji, H. Gong, Y. Lyu, B. Li and X. Liao. 2020. Estimating leaf chlorophyll content of crops via optimal unmanned aerial vehicle hyperspectral data at multi-scales. *Computers and Electronics in Agriculture* 178:105786.

Car Detection from Very High-Resolution UAV Images Using Deep Learning Algorithms

Yunus Kaya, Halil İbrahim Şenol, Abdurahman Yasin Yiğit, and Murat Yakar

Abstract

It is important to determine car density in parking lots, especially in hospitals, large enterprises, and residential areas, which are used intensively, in terms of executing existing management systems and making precise plans for the future. In this study, cars in parking lots were detected using high-resolution unmanned aerial vehicle (UAV) images with deep learning methods. We tested the performance of the two approaches by determining the number of cars in a parking lot using the You Only Look Once (YOLOv3) and Mask Region–Based Convolutional Neural Networks (Mask R-CNN) approaches as deep learning methods and the deep learning tool of Esri ArcGIS Pro. High-resolution UAV images were processed by photogrammetry and used as input products for the R-CNN and YOLOv3 algorithm. Recall, F1 score, precision ratio/uncertainty accuracy, and average producer accuracy of products automatically extracted with the algorithm were determined as 0.862/0.941, 0.874/0.946, 0.885/0.951, and 0.776/0.897 for R-CNN and YOLOv3, respectively.

Introduction

Environmental monitoring is directly related to the urban development. Especially with technological developments, the use of unmanned aerial vehicles (UAVs) for environmental monitoring and ground observation has become widespread (Eskandari *et al.* 2020). The use of high-resolution UAV data, together with the deep learning algorithms that have developed in recent years, facilitates environmental monitoring.

Object detection has been one of the important topics explored by remote sensing and photogrammetric techniques for years. It is almost impossible to detect objects manually, especially due to the growth and crowding of urban areas. However, with the developing image capture technologies and increasing resolution in recent years, detecting objects has started to become an important research topic. In addition, it has been possible to automatically detect these objects with the deep learning methods that have been around for years but have recently been used more frequently.

Literature Review

Studies for object detection are very large scale. The differentiation of data and the methods used and the fact that deep learning techniques offer a wide range of solutions are the results of this. While the method used is important, the characteristics of the selected data are also important. With the development of high-resolution aerial images and light detection and ranging (lidar) applications in recent years, the diversity of data is also increasing. Although high-resolution UAV images (Moranduzzo *et al.* 2013; Moranduzzo and Melgani 2014; Ulvi and Toprak 2016; Perko *et al.* 2021; Bakirman *et al.* 2022) are used in most of the studies, very high-resolution satellite images (Perko *et al.* 2021; Deng *et al.* 2022; Liu *et al.* 2022; Shamsolmoali *et al.* 2022), high-resolution video images (Jiao *et al.* 2021), and point cloud data obtained with lidar (Fiorucci *et al.* 2022) are also used.

Alternative methods based on digital aerial photogrammetry have been proposed as a low-cost, high-resolution, three-dimensional (3D) information source in obtaining high-resolution images (Iglhaut *et al.* 2019; Puliti *et al.* 2020). Structure from motion (SfM) uses multiple images collected from different locations to create a 3D model of the visible surface that can generate point cloud data similar to lidar. SfM is a photogrammetric algorithm that automatically solves the geometry of the scene, camera positions, and orientation without requiring pre-definition of a target mesh with known 3D positions (Vasuki *et al.* 2014; Hastaoğlu *et al.* 2022) . The SfM can be collected from an airborne platform, such as a handheld camera, aircraft, balloon, kite, or unmanned aerial system. The ability to generate high-resolution orthomosaic images from UAV aerial images covering the entire work area means that image-based methods are an attractive alternative to the more costly lidar data collection.

Apart from the data used for object detection, the deep learning algorithms used also differ. Although algorithms for convolutional neural networks (CNN) (Liu *et al.* 2021; Shivappriya *et al.* 2021; de Arruda *et al.* 2022) are frequently used, the Part-Based Convolutional Neural Network (Sun *et al.* 2021), the Critical Feature Capturing Network (Ming *et al.* 2021), Attention Scale Pyramid Deformable (Gao *et al.* 2021), the Stress Concentration Discrimination Model (Duan *et al.* 2022), and the Frequency Extraction Network (Cheng *et al.* 2021) are also used for object detection.

Ammour *et al.* (2017) combined very high-spatial resolution UAV images with the CNN deep learning method and support vector machine classification method to perform car extraction with up to 93.6% accuracy. Benjdira *et al.* (2019) compared the performance of Faster R-CNN and You Only Look Once (YOLOv3) algorithms on very large-car data sets based on a car extraction application. Although YOLOv3 and Faster R-CNN can be compared in terms of accuracy, it has been found that YOLOv3 has higher performance in terms of speed. Amato *et al.* (2019) proposed an approach for real-time parked car detection from UAV videos. They determined the accuracy as 95% in their YOLOv3-based solutions.

Kilic and Ozturk (2021) analyzed high-resolution aerial images for car counting with the heatmap learner convolutional neural networks tool and performed car extraction by uploading a test data. Srivastava *et al.* (2021) used the current methods for tool extraction and compared them in their studies. In the study, it has been determined that the use of different detection methods (UAV, lidar, sonar, and radar) together can produce better results. Zhang *et al.* (2021) made an application for car detection using high-resolution satellite images and CNN algorithms.

In this study, parked cars on high-resolution UAV images were extracted with YOLOv3 and a deep learning application provided by Esri Analytics and based on the Mask R-CNN algorithm. The main motivation of the study is to investigate the accuracy of Esri, which offers a commercial approach to developing deep learning object detection algorithms, and to question its use in projects that require high precision. Unlike previous studies, the usability of the results of a

Yunus Kaya and Halil İbrahim Şenol are with Harran University, Geomatics Engineering Department, Sanliurfa, Turkey.

Abdurahman Yasin Yiğit and Murat Yakar are with Mersin University, Geomatics Engineering Department, Mersin, Turkey.

Contributed by Tolga Bakirman, July 16, 2022 (sent for review September 12, 2022; reviewed by Timea Bezdan, Ibrahim Yilmaz, Peijuan Wang).

Photogrammetric Engineering & Remote Sensing
Vol. 89, No. 2, February 2023, pp. 117–123.
0099-1112/22/117–123

doi: 10.14358/PERS.22-00101R2

commercial solution proposal was investigated. In the following parts of the study, detailed information is given about the data, algorithm, and results used.

Materials and Methods

Study Area

The study was carried out using high-resolution satellite images taken in a parking lot located at Mersin University, Turkey (Figure 1). The reason for choosing this area is to investigate whether it is possible to detect cars parked outside the parking lot as well as in the regular parking structure.

Data Acquisition

A Sensefly Ebee Plus UAV with a built-in 20-MP camera was used for photogrammetric research. It is integrated with a remote flight control and real-time kinematic (RTK) system. The characteristics of the UAV system and the onboard camera are given in Table 1.

Table 1. UAV system features and onboard camera parameters.

Specialty	Value
Weight/size	1100 g/1100 mm
Cruising speed	40–110 m/s
Maximum of flight time	About 50 minutes
PPK/RTK	+
Radiolink distance	3 km
Satellite positioning systems	GPS/GLONASS
Camera model	S.O.D.A.
Sensor	1-inch CMOS effective pixels: 20 M
Lens	10.6 mm (35-mm equivalent: 29 mm)
Image size	5472 × 3648 pixels
f-number	f/2.8–f/11

The main field activities were carried out in three steps: (1) flight mission, (2) placement and acquisition of ground control points (GCPs), and (3) flight operation and aerial image collection.

Terrain Awareness flight planning was preferred because a constant ground sample distance (GSD) value was desired at the end of the study. Terrain Awareness helps pilots achieve improved map quality in a variety of terrains with equal resolution and less probability of holes at higher altitudes. In Terrain Awareness flight planning, the equivalent of a GSD of ~5 cm/pixel is ~210 m in the S.O.D.A camera. In addition, the UAV was equipped with satellite positioning systems (GPS/GLONASS), and all images collected were geo-located in a WGS84/UTM36N metric coordinate system.

The data used in this study include a total of four missions conducted during 1 day in Mersin, Turkey, in July 2021. Four hundred and twelve digital images were taken for the study area, and lens calibration was applied automatically. These images from the region provide approximately 75% overlap in both forward and lateral (overlap) directions.

Method

In this section, detailed information about the processing of the data obtained is given, as shown in Figure 2.

Image Processing and the SfM Algorithm

The acquired images were transferred to a computer for photogrammetric processing. Flight trajectories were reconstructed using Global Navigation Satellite Systems/Inertial Measurement Unit (GNSS/IMU) measurements and adjusted using differentially corrected base station data in Emotion 3.0 software. Locations of cameras were further refined using GCPs, resulting in a derived point cloud with a geometric accuracy of 4.2 cm.

The processing of the images obtained by the UAV was done in Agisoft Metashape software, a commercial software based on image matching and model generation using the SfM algorithm (Şasi and Yakar 2018). In photogrammetric software using the SfM algorithm, bundle block adjustment algorithms are used to optimize the projection errors between the images and the calculated point positions (Duran and Atik 2021; Senkal *et al.* 2021). The software using these algorithms and methods first aligns the images, performs a preliminary

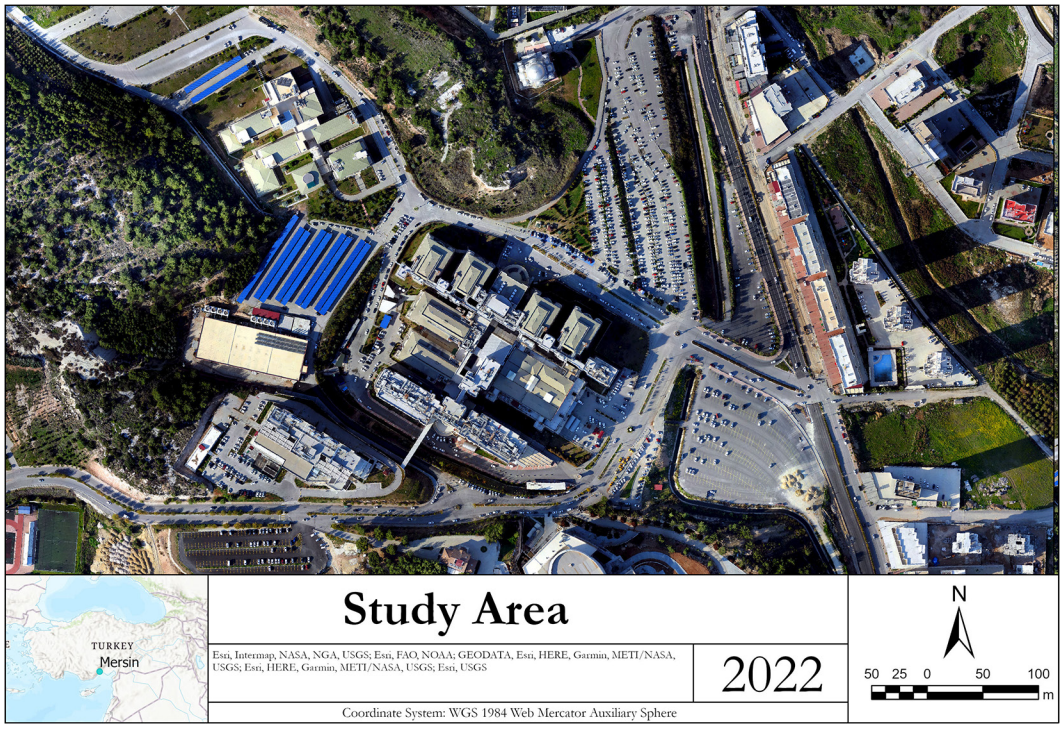


Figure 1. Study area.

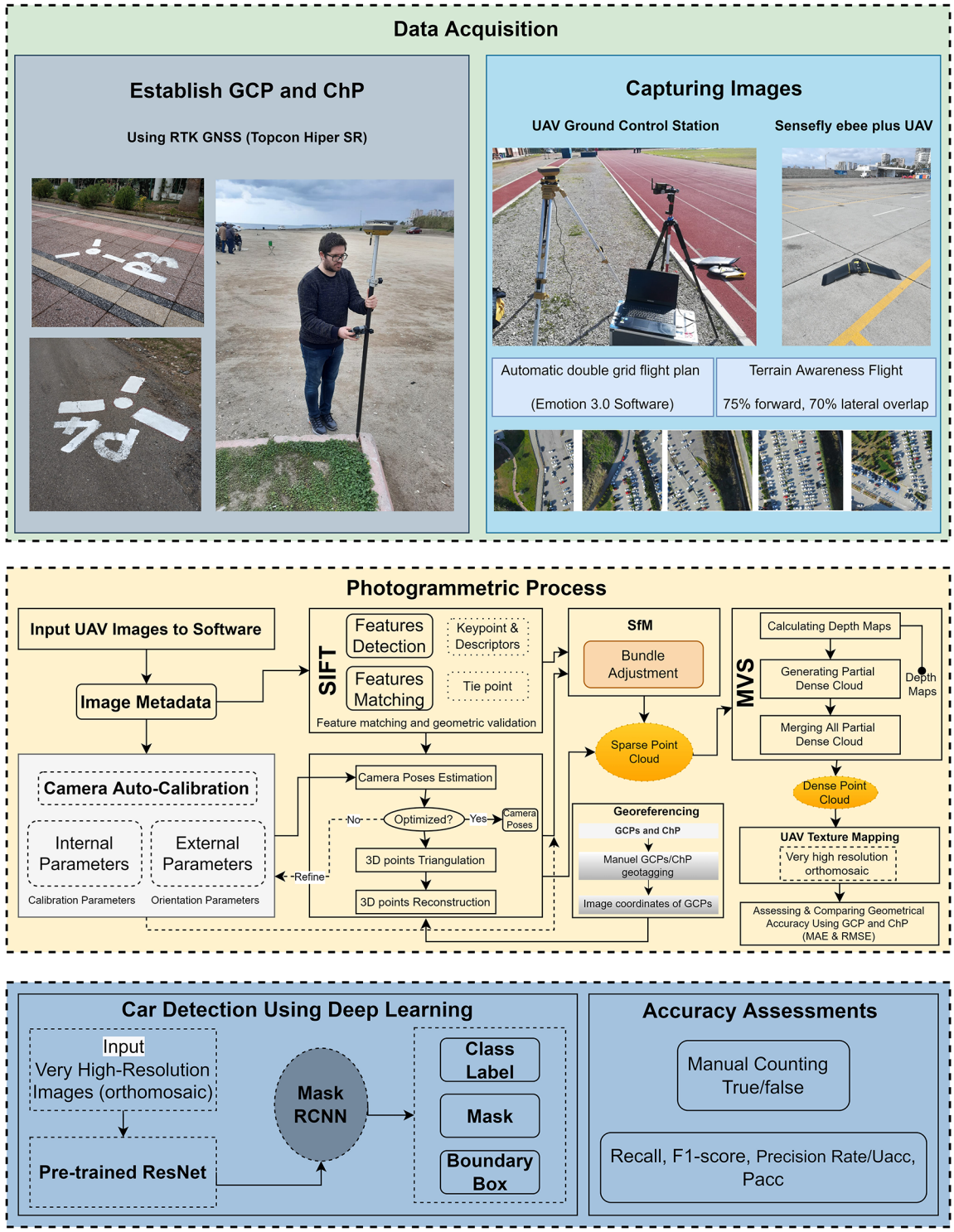


Figure 2. Work flow.

adjusting, and creates a sparse point cloud. For this, algorithms are used to automatically position the matching points between the images (Hamal *et al.* 2020).

Feature Detection and Matching

In this step, Agisoft Metashape initiates the image alignment process by executing the scale-invariant feature transform (SIFT) algorithm (Moranduzzo and Melgani 2014; Ulvi 2020) to generate a starting point cloud consisting of matching key points in different images (Moranduzzo *et al.* 2015; Hamal 2022) .

The SIFT algorithm improves the initial camera parameters using the corrected positions and orientations of each image as a starting point. A set of key points is determined in each image through feature matching algorithms such as SIFT, and these are stored as key points in a database (Hossein-Nejad *et al.* 2021). When processing a new image, the algorithm recognizes the image features and compares them with the features in the database to identify the tie points. The SIFT algorithm automatically detects and matches key points between multiple images determined by local pixel variances (Nesbit *et al.* 2018; Ulvi 2021). In

this step, a low-density point cloud is created with the 3D coordinates of the key points (Westoby *et al.* 2012). We set the limit of points to detect per image to 80k and the limit of matching points to 20k. We used "high" alignment accuracy for all data sets. GCPs were manually defined to refine camera positions and marked on some images to aid optimization. We processed the height data as ellipsoid height (meters). We used the WGS84 UTM 36N coordinate system in the study.

Multi-View Stereo

Another step in SfM is the process of creating a dense point cloud by condensing the created sparse point cloud for an accurate geometrical detail display. The algorithm used at this stage is the dense multi-view stereo algorithm. This algorithm divided the overlapping images into subsets and reconstructed the 3D point data independently of these subsets. Point condensation was performed by a patch-based multi-view stereo algorithm (Lowe 2004; Ulvi 2018). In this step, binary depth maps are calculated for the overlapping image pairs using the GNSS stereo matching algorithm, taking into calculation the relative camera parameters previously calculated. The created binary depth maps were combined into partial dense point clouds. We set the quality of creating depth maps to "high quality" with an aggressive filtering mode.

In the final step, a photo-realistic 3D model and high-quality 2D orthomosaic image and a 2.5D digital elevation model can also be created in this step, alternatively using images to cover a real and natural surface.

Table 2. Accuracy evaluation of study area for the georeferencing of UAV-SfM photogrammetry.

Positioning Ground Control Points (as GCPs)			Check Control Points (as ChPs)		
No.	Horizontal Errors (m)	Vertical Errors (m)	No.	Horizontal Errors (m)	Vertical Errors (m)
1	0.008	0.012	2	0.007	0.010
3	0.014	0.010	5	0.014	0.009
4	0.015	0.007	6	0.011	0.008
7	0.009	0.014	9	0.004	0.011
8	0.010	0.014	10	0.009	0.009
11	0.012	0.009	12	0.011	0.012
…	…	…	…	…	…
MAE	0.014	0.015	MAE	0.013	0.010
RMSE	0.012	0.014	RMSE	0.011	0.08

A final 3D digital model of ~83 million points was created for the study area. The average surface density is about 14k points per meter. The mean absolute error (MAE) and the root mean square error (RMSE) of the GCPs and check control points (ChPs) used in the study were calculated (Table 2). These results show satisfactory accuracy for automatic detail extraction analysis from images produced with high quality and accuracy. MAE (Equation 1) and RMSE (Equation 2) are two metrics that are employed in object detection algorithm evaluation. MAE represents the average absolute difference between true values and predicted values . Similarly, RMSE represents the average of the square root of the squared difference between the true and predicted values:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i}^{n}\left|f_i - y_i\right| \tag{1}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i}^{n}\left(f_i - y_i\right)^2} \tag{2}$$

Mask R-CNN

For object instance segmentation, we used Mask R-CNN (He *et al.* 2017), which is conceptually simple and frequently used in the literature (Ahmed *et al.* 2020; Mahmoud *et al.* 2020; Sun *et al.* 2022). The approach automatically detects objects with specific shapes on the high-resolution image while also creating a high-quality segmentation mask for each sample. The CNN approach consists of several convolutional and pooling layers and ends with more than one fully connected layer. There are three steps in each convolution layer: convolution, nonlinear activation, and pooling. After each convolutional layer, a feature map is created and passed to the next layer. Mask R-CNN extends Faster R-CNN by adding a branch for predicting an object mask in parallel with the existing branch for bounding box recognition. Faster R-CNN (Ren *et al.* 2015) has two outputs for each candidate object. The first step, called the region proposal network (RPN), proposes candidate object bounding boxes. The second step, which is Fast R-CNN (Girshick 2015), extracts features using RoIPool from each candidate box and performs classification and bounding box regression. Mask R-CNN has a third branch that extracts this object mask (Ammirato and Berg 2019). Mask R-CNN is easy to train and can easily be generalized to other tasks, such as estimating human poses, cars, or houses.

Mask R-CNN consists of two steps. In the first step (RPN), it scans feature maps and generates a region of interest (RoI) (Ren *et al.* 2015). RPN uses ResNet as its backbone architecture. The high-level and low-level fusion of different feature layers using a residual convolution network (He *et al.* 2016) and a four-layer feature pyramid takes full advantage of shallow image location information and meaningful semantic information. In the second step, RoI pool processing is applied for each RoI to downsample the feature map using the nearest neighbor approach (Girshick 2015). In the pooling phase, important features are selected from the feature map. The RoI pool can cause misalignment between extracted features. To create more accurate RoIs, RoI alignment created using bilinear interpolation is applied for each RoI. Mask R-CNN generates masks for each RoI as well as predicting the class and bounding boxes for each object (Zhang *et al.* 2020).

YOLOv3

Object detection is used in many areas to facilitate human life. Single-stage models also need to be improved to meet the requirements, as their use is increasing rapidly (Narejo *et al.* 2021). YOLOv3 (Redmon and Farhadi 2018) is one of the state-of-the-art object detection methods evolving from YOLO and YOLOv2. The YOLOv3 object detection process consists of five steps (Figure 2). Unlike darknet19 used by YOLO9000, YOLOv3 uses a new backbone network. It is called darknet53 because it is a 53 convolutional network (Kou *et al.* 2021). Compared with the two-stage object detection network, the YOLOv3 has an obvious speed advantage. Compared with the single-stage network, YOLOv3 can achieve average sensitivity with higher detection speed. YOLOv3 uses the idea of feature pyramid networks to make predictions at three different scales: 13 × 13, 26 × 26, and 52 × 52 (Ge *et al.* 2021). Three scales perform information extraction and merging for feature maps of different sizes and output detection. This improves detection accuracy while maintaining high-speed operation.

Assessment Method

After car detection algorithms are developed with deep learning, the performance of the final product should be evaluated. Currently, various measurement techniques are used to evaluate the quality of object detection algorithms, including recall, F1 score, precision ratio/uncertainty accuracy (UAcc), producer accuracy (Pacc), and manual counting (true, false). In order to evaluate the correct identification capacity of our mentioned accuracy assessment methodology, the following definitions need to be made.

The true positive (TP) value represents the number of cars that are actually cars and correctly identified by the algorithm. False positive (FP) means the number of cars misidentified by the algorithm, not actually cars. False negative (FN) means the undetected cars by the algorithm. The real number of cars in the images is represented as N.

Recall rate (sensitivity) is an important metric for evaluating algorithm performance. It represents how many of the predicted positive cars are actually cars as follows:

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{3}$$

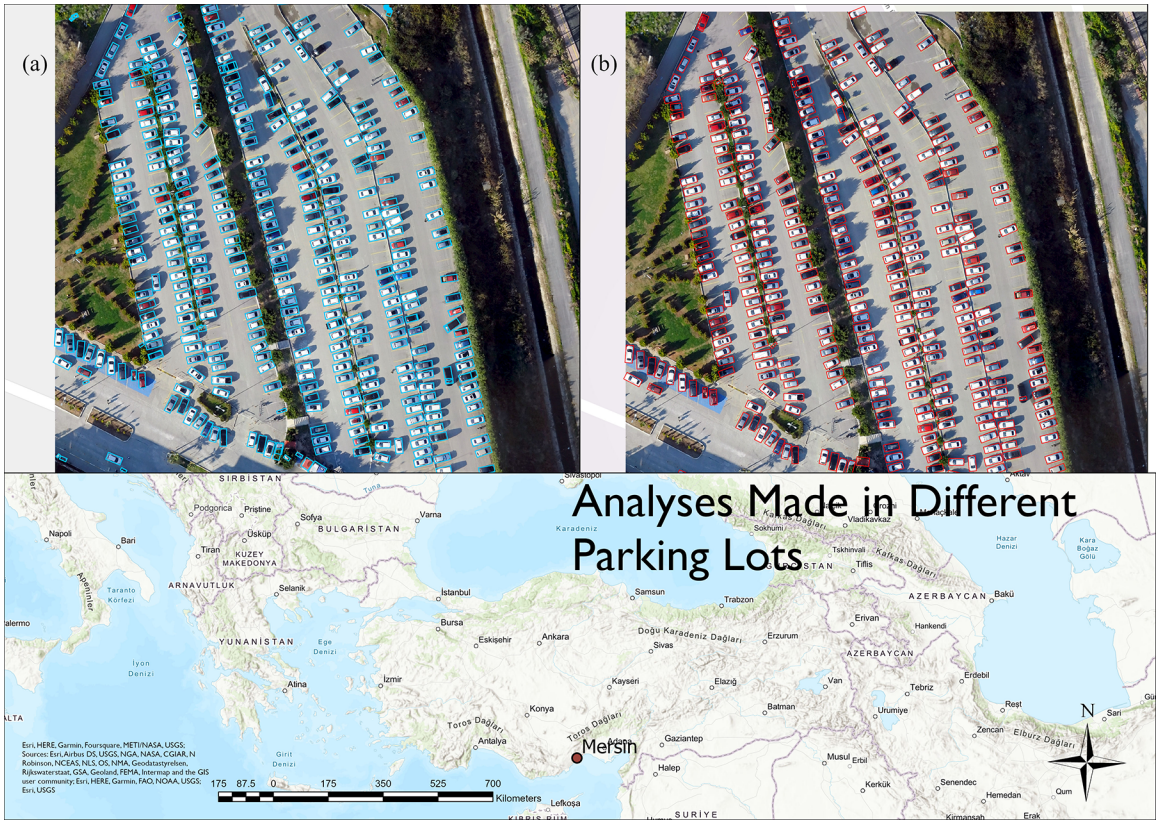


Figure 3. Cars detected by the Mask R-CNN (a) and YOLOv3 (b) algorithms.

The F1 score combines precision and recall rates into single metric; this is the harmonic mean of these combined metrics:

$$\text{F1 score} = \frac{2*TP}{2*TP + FN + FP} \tag{4}$$

UAcc is another important evaluation metric that determines the number of cars predicted to be truly positive across all detections:

$$\text{Precision}(\text{UAcc}) = \frac{TP}{TP + FP} \tag{5}$$

$$\text{Pacc} = \frac{TP}{N} \tag{6}$$

Results and Discussion

Today, various studies continue to be carried out for smart city and urban development. Identifying cars and detecting the use of parking lots can solve important urban problems, such as detecting car density in urban areas and identifying needed parking spaces.

The aim of the study is to distinguish and count the parked cars in a parking area using the Mask R-CNN and YOLO v3 deep learning methods. In this direction, a very high-resolution (5 cm) UAV image was used in the study.

In this study, car extraction was performed using the object detection method from the image in the deep learning tool, a tool that Esri ArcGIS Pro software has been developing recently. Toolbox uses the Mask R-CNN algorithm and allows data anlysis using a pretrained training package. Detailed information on the use of this tool can be found in Esri (2022). In addition, the YOLO v3 deep learning algorithm used for object detection was used through Anaconda. Open parking lots were selected from the very high-resolution UAV image used for the study. The hospital parking lot, where there are always too many cars, was preferred for the distinguishability of car detection. Figure 3 shows the areas used for car detection and the resulting products.

As can be seen in Figure 3, although the algorithms worked successfully and detected cars, they also detected some unrelated objects that looked like cars. In this direction, to reveal the results of car detection, the detected polygons were examined, and an accuracy analysis was performed (Table 3). For the analysis, the accuracy between the ground-truth feature objects and the objects detected by the deep learning algorithm was calculated. Figure 3 shows the final car detection map for test parking, and Table 3 reports the quantitative detection results.

Table 3. Car detection results for the very high-resolution UAV images using deep learning methods (Mask R-CNN and YOLOv3).

	Total	TP	FP	FN	Recall	F1 Score	Precision (UAcc)	Pacc
Mask R-CNN	388	301	39	48	0.862	0.874	0.885	0.776
YOLOv3	388	348	18	22	0.941	0.946	0.951	0.897

FN = false negative; FP = false positive; Pacc = producer accuracy; TP = true positive; UAcc = user accuracy.

When the values in Table 3 and Figure 3 are examined, it is seen that both methods are successful, but the YOLOv3 approach gives better results than the R-CNN approach. It has been noticed that the R-CNN algorithm has difficulty detecting especially in places where shadows fall on the image due to the angle of the sun. However, although the algorithm was successful in detecting cars, it also detected other objects similar to cars as cars. The reason for this may be the light reflected from the car windows from other detected objects and the algorithm making an error for this reason. However, as you can see in Table 3, high accuracy rates were obtained, and the car detection was successful. Although YOLOv3 can be preferred because it gives more accurate results, programming ability is needed for the application of this method. However, the deep learning tool developed by Esri can

solve an engineering problem without the need for programming skills. Therefore, since Mask R-CNN can be used with commercial software, this deep learning approach is easy to use, although its accuracy is low.

Conclusion

The aim of this study is to investigate and compare the performances of the Mask R-CNN and YOLOv3 algorithms for car detection on high-resolution UAV images. The object detection technique automatically detects cars with the recommended approach. The study consists of three stages: acquisition and processing of high-resolution UAV data and car detection using R-CNN and YOLOv3 architecture. First, high-resolution (5-cm) images obtained with Sensefly Ebee Plus UAV with a built-in 20-MP camera were processed with the SfM algorithm. Then automatic detection of cars in the parking area was made using the Mask R-CNN and YOLOv3 architecture. As a result of the study, the recall, F1 score, UAcc, and Pacc average accuracy of products automatically extracted with the algorithm were determined as 0.862/0.941, 0.874/0.946, 0.885/0.951, and 0.776/0.897 for R-CNN and YOLOv3, respectively. Therefore, the proposed methods provide an automatic car counting technique over a processed UAV image. As a challenge of our work, serious planning is required to acquire UAV data automatically and at regular intervals. Therefore, in future studies, we aim to determine the automatic car counting approach through data obtained from security cameras. However, comparisons can be made by adding more object detection deep learning algorithms in future studies. By making these comparisons with different data, a result can be obtained by choosing the right algorithms for the different data used.

References

Ahmed, B., T. A. Gulliver and S. alZahir. 2020. Image splicing detection using mask-RCNN. *Signal, Image and Video Processing* 14(5):1035–1042.

Amato, G., L. Ciampi, F. Falchi and C. Gennaro. 2019. Counting vehicles with deep learning in onboard UAV imagery. Pages 1–6 in *2019 IEEE Symposium on Computers and Communications (ISCC)*, held in Barcelona, Spain 29 June to 3 July 2019. Edited by Sartaj Sahni. Los Alamitos, CA, USA: IEEE Computer Society.

Ammirato, P. and A. C. Berg. 2019. A Mask-RCNN baseline for probabilistic object detection. Preprint, arXiv:1908.03621.

Ammour, N., H. Alhichri, Y. Bazi, B. Benjdira, N. Alajlan and M. Zuair. 2017. Deep learning approach for car detection in UAV imagery. *Remote Sensing* 9(4):312.

Bakirman, T., I. Komurcu and E. Sertel. 2022. Comparative analysis of deep learning based building extraction methods with the new VHR Istanbul dataset. *Expert Systems with Applications* 202:117346.

Benjdira, B., T. Khursheed, A. Koubaa, A. Ammar and K. and Ouni. 2019. Car detection using unmanned aerial vehicles: Comparison between Faster R-CNN and YOLOv3. Pages 1–6 in *2019 1st International Conference on Unmanned Vehicle Systems-Oman (UVS)*, held in Muscat, Oman 5-7 February 2019. Edited by Hadj Bourdoucen. Piscataway, New Jersey: IEEE.

Cheng, G., C. Lang, M. Wu, X. Xie, X. Yao and J. and Han. 2021. Feature enhancement network for object detection in optical remote sensing images. *Journal of Remote Sensing* 2021:9805389.

de Arruda, M.D.S., L. P. Osco, P. R. Acosta, D. N. Gonçalves, J. M. Junior, A. P. M. Ramos and W. N. Gonçalves. 2022. Counting and locating high-density objects using convolutional neural network. *Expert Systems with Applications* 195:116555.

Deng, J., X. Li and Y. Fang. 2022. Few-shot object detection on remote sensing images. Preprint, arXiv:2006.07826.

Duan, Z., S. Wang, H. Di and J. Deng. 2022. Distillation remote sensing object counting via multi-scale context feature aggregation. *IEEE Transactions on Geoscience and Remote Sensing* 60:5613012.

Duran, Z. and M. E. Atik. 2021. Accuracy comparison of interior orientation parameters from different photogrammetric software and direct linear transformation method. *International Journal of Engineering and Geosciences* 6(2):74–80.

Eskandari, R., M. Mahdianpari, F. Mohammadimanesh, B. Salehi, B. Brisco and S. Homayouni. 2020. Meta-analysis of unmanned aerial vehicle (UAV) imagery for agro-environmental monitoring using machine learning and statistical models. *Remote Sensing* 12(21):3511.

Esri. 2022. Deep Learning Model to Detect Cars in High Resolution Imagery. <https://www.arcgis.com/home/item.html?id=cfc57b507f914d1593f5871bf0d52999> Accessed 12 September 2022.

Fiorucci, M., W. B. Verschoof-van der Vaart, P. Soleni, B. Le Saux and A. Traviglia. 2022. Deep learning for archaeological object detection on LiDAR: New evaluation measures and insights. *Remote Sensing* 14(7):1694.

Gao, G., Q. Liu and Y. Wang. 2021. Counting from sky: A large-scale data set for remote sensing object counting and a benchmark method. *IEEE Transactions on Geoscience and Remote Sensing* 59(5):3642–3655.

Ge, Z., S. Liu, F. Wang, Z. Li and J. Sun. 2021. Yolox: Exceeding yolo series in 2021. Preprint, arXiv:2107.08430.

Girshick, R. 2015. Fast R-CNN. Pages 1440–1448 in *IEEE International Conference on Computer Vision (ICCV)*, held in Santiago, Chile, 7-13 December 2015. Edited by Lisa O'Conner. NW Washington, DC, United States: IEEE Computer Society.

Hamal, S.N.G. 2022. Accuracy of digital maps produced from UAV images in rural areas. *Advanced UAV* 2(1):29–34.

Hamal, S.N.G., B. Sarı and A. Ulvi. 2020. Using of hybrid data acquisition techniques for cultural heritage: A case study of Pompeiopolis. *Türkiye İnsansız Hava Araçları Dergisi* 2(2):55–60.

Hastaoğlu, K. Ö., S. Göğsu and Y. Gül. 2022. Determining the relationship between the slope and directional distribution of the UAV point cloud and the accuracy of various IDW interpolation. *International Journal of Engineering and Geosciences* 7(2):161–173.

He, K., G. Gkioxari, P. Dollár and R. Girshick. 2017. Mask R-CNN. Pages 2961–2969 in *Proceedings of the IEEE International Conference on Computer Vision*, held in Venice, Italy, 22-29 October 2017. Edited by Lisa O'Conner. NW Washington, DC, United States: IEEE Computer Society.

He, K., X. Zhang, S. Ren and J. Sun. 2016. Deep residual learning for image recognition. Pages 770–778 in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, held in Las Vegas, Nevada, 26 June – 1 July 2016. Edited by Lisa O'Conner. NW Washington, DC, United States: IEEE Computer Society.

Hossein-Nejad, Z., H. Agahi and A. Mahmoodzadeh. 2021. Image matching based on the adaptive redundant keypoint elimination method in the SIFT algorithm. *Pattern Analysis and Applications* 24(2):669–683.

Iglhaut, J., C. Cabo, S. Puliti, L. Piermattei, J. O'Connor and J. Rosette. 2019. Structure from motion photogrammetry in forestry: A review. *Current Forestry Reports* 5:155–168.

Jiao, L., R. Zhang, F. Liu, S. Yang, B. Hou, L. Li and X. Tang. 2021. New generation deep learning for video object detection: A survey. *IEEE Transactions on Neural Networks and Learning Systems* 33(8):3195–3215.

Kilic, E. and S. Ozturk. 2021. An accurate car counting in aerial images based on convolutional neural networks. *Journal of Ambient Intelligence and Humanized Computing*. https://doi.org/10.1007/s12652-021-03377-5.

Kou, X., S. Liu, K. Cheng and Y. Qian. 2021. Development of a YOLO-V3-based model for detecting defects on steel strip surface. *Measurement* 182:109454.

Liu, S., L. Zhang, H. Lu and Y. He. 2022. Center-boundary dual attention for oriented object detection in remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing* 60:1–14.

Liu, Y., P. Sun, N. Wergeles and Y. Shang. 2021. A survey and performance evaluation of deep learning methods for small object detection. *Expert Systems with Applications* 172:114602.

Lowe, D. G. 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60(2):91–110.

Mahmoud, A., S. Mohamed, R. El-Khoribi and H. Abdelsalam. 2020. Object detection using adaptive mask RCNN in optical remote sensing images. *International Journal of Intelligent Engineering and Systems* 13(1):65–76.

Ming, Q., L. Miao, Z. Zhou and Y. Dong. 2021. CFC-Net: A critical feature capturing network for arbitrary-oriented object detection in remote-sensing images. *IEEE Transactions on Geoscience and Remote Sensing* 60:1–14.

Moranduzzo, T. and F. Melgani. 2014. Automatic car counting method for unmanned aerial vehicle images. *IEEE Transactions on Geoscience and Remote Sensing* 52:1635–1647.

Moranduzzo, T., F. Melgani, Y. Bazi and N. Alajlan. 2015. A fast object detector based on high-order gradients and Gaussian process regression for UAV images. *International Journal of Remote Sensing* 36(10):2713–2733.

Moranduzzo, T., F. Melgani and A. Daamouche. 2013. An object detection technique for very high resolution remote sensing images. Pages 79–83 in *2013 8th International Workshop on Systems, Signal Processing and Their Applications (WoSSPA)*, held in Algiers, Algeria, 12-15 May 2013. Edited by Boualem Boashash. Piscataway, NJ: IEEE.Narejo, S., B. Pandey, C. Rodriguez and M. R. Anjum. 2021. Weapon detection using YOLO V3 for smart surveillance system. *Mathematical Problems in Engineering* 2021:9975700.

Nesbit, P. R., P. R. Durkin, C. H. Hugenholtz, S. M. Hubbard and M. Kucharczyk. 2018. 3-D stratigraphic mapping using a digital outcrop model derived from UAV images and structure-from-motion photogrammetry. *Geosphere* 14(6):2469–2486.

Perko, R., A. Almer, M. Theuermann, M. Klopschitz, T. Schnsbel and P. M. Roth. 2021. Protocol design issues for object density estimation and counting in remote sensing. Pages 2771–2774 in *2021 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, held in Brussels, Belgium, 12 – 16 July, 2021. Edited by Ramon Hanssen. NW Washington, DC, United States: IEEE International.Puliti, S., J. Breidenbach and R. Astrup. 2020. Estimation of forest growing stock volume with UAV laser scanning data: Can it be done without field data? *Remote Sensing* 12(8):1245.

Redmon, J. and A. Farhadi. 2018. Yolov3: An incremental improvement. Preprint, arXiv:1804.02767.Ren, S., K. He, R. Girshick and J. Sun. 2015. Faster R-CNN: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems*, vol. 28, edited by C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama and R. Garnett, 91–99. Red Hook, NY: Curran Associates, Inc.

Şasi, A. and M. Yakar. 2018. Photogrammetric modelling of Hasbey Dar'ülhuffaz (Masjid) using an unmanned aerial vehicle. *International Journal of Engineering and Geosciences* 3 (1):6–11.

Senkal, E., G. Kaplan and U. Avdan. 2021. Accuracy assessment of digital surface models from unmanned aerial vehicles' imagery on archaeological sites. *International Journal of Engineering and Geosciences*, 6 (2), 81-89.

Shamsolmoali, P., J. Chanussot, M. Zareapoor, H. Zhou and J. Yang. 2022. Multipatch feature pyramid network for weakly supervised object detection in optical remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing* 60:3106442.

Shivappriya, S. N., M.J.P. Priyadarsini, A. Stateczny, C. Puttamadappa and B. D. Parameshachari. 2021. Cascade object detection and remote sensing object detection method based on trainable activation function. *Remote Sensing* 13(2):200.

Srivastava, S., S. Narayan and S. Mittal. 2021. A survey of deep learning techniques for vehicle detection from UAV images. *Journal of Systems Architecture* 117:102152.

Sun, X., P. Wang, C. Wang, Y. Liu and K. Fu. 2021. PBNet: Part-based convolutional neural network for complex composite object detection in remote sensing imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* 173:50–65.

Sun, Y., L. Su, Y. Luo, H. Meng, W. Li, Z. Zhang and W. Zhang. 2022. Global Mask R-CNN for marine ship instance segmentation. *Neurocomputing* 480:257–270.

Ulvi, A. 2018. Analysis of the utility of the unmanned aerial vehicle (UAV) in volume calculation by using photogrammetric techniques. *International Journal of Engineering and Geosciences* 3(2):43–49.

Ulvi, A. 2020. Importance of unmanned aerial vehicles (UAVs) in the documentation of cultural heritage. *Turkish Journal of Engineering* 4(3):104–112.

Ulvi, A. 2021. The effect of the distribution and numbers of ground control points on the precision of producing orthophoto maps with an unmanned aerial vehicle. *Journal of Asian Architecture and Building Engineering* 20(6):806–817.

Ulvi, A. and A. S. Toprak. 2016. Investigation of three-dimensional modelling availability taken photograph of the unmanned aerial vehicle: Sample of Kanlıdıvane Church. *International Journal of Engineering and Geosciences* 1(1):1–7.

Vasuki, Y., E. J. Holden, P. Kovesi and S. Micklethwaite. 2014. Semi-automatic mapping of geological structures using UAV-based photogrammetric data: An image analysis approach. *Computers & Geosciences* 69:22–32.

Westoby, M. J., J. Brasington, N. F. Glasser, M. J. Hambrey and J. M. Reynolds. 2012. "Structure-from-Motion" photogrammetry: A low-cost, effective tool for geoscience applications. *Geomorphology* 179:300–314.

Zhang, J., Q. Hu, J. Li and M. Ai. 2021. Learning from GPS trajectories of floating car for CNN-based urban road extraction with high-resolution satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing* 59(3):1836–1847.

Zhang, Q., X. Chang and S. B. Bian. 2020. Vehicle-damage-detection segmentation algorithm based on improved Mask RCNN. *IEEE Access* 8:6997–7004.

Digital Elevation Model Technologies and Applications: The DEM Users Manual, 3rd Edition

Edited by David F. Maune, PhD, CP and Amar Nayegandhi, CP, CMS

To order, visit https://www.asprs.org/dem

The 3rd edition of the DEM Users Manual includes 15 chapters and three appendices. References in the eBook version are hyperlinked. Chapter and appendix titles include:

1. Introduction to DEMs
   David F. Maune, Hans Karl Heidemann, Stephen M. Kopp, and Clayton A. Crawford
2. Vertical Datums
   Dru Smith
3. Standards, Guidelines & Specifications
   David F. Maune
4. The National Elevation Dataset (NED)
   Dean B. Gesch, Gayla A. Evans, Michael J. Oimoen, and Samantha T. Arundel
5. The 3D Elevation Program (3DEP)
   Jason M. Stoker, Vicki Lukas, Allyson L. Jason, Diane F. Eldridge, and Larry J. Sugarbaker
6. Photogrammetry
   J. Chris McGlone and Scott Arko
7. IfSAR
   Scott Hensley and Lorraine Tighe
8. Airborne Topographic Lidar
   Amar Nayegandhi and Joshua Nimetz
9. Lidar Data Processing
   Joshua M. Novac
10. Airborne Lidar Bathymetry
    Jennifer Wozencraft and Amar Nayegandhi
11. Sonar
    Guy T. Noll and Douglas Lockhart
12. Enabling Technologies
    Bruno M. Scherzinger, Joseph J. Hutton, and Mohamed M.R. Mostafa
13. DEM User Applications
    David F. Maune
14. DEM User Requirements & Benefits
    David F. Maune
15. Quality Assessment of Elevation Data
    Jennifer Novac

Appendix A. Acronyms

Appendix B. Definitions

Appendix C. Sample Datasets

This book is your guide to 3D elevation technologies, products and applications. It will guide you through the inception and implementation of the U.S. Geological Survey's (USGS) 3D Elevation Program (3DEP) to provide not just bare earth DEMs, but a full suite of 3D elevation products using Quality Levels (QLs) that are standardized and consistent across the U.S. and territories. The 3DEP is based on the National Enhanced Elevation Assessment (NEEA) which evaluated 602 different mission-critical requirements for and benefits from enhanced elevation data of various QLs for 34 Federal agencies, all 50 states (with local and Tribal input), and 13 non-governmental organizations.

The NEEA documented the highest Return on Investment from QL2 lidar for the conterminous states, Hawaii and U.S. territories, and QL5 IfSAR for Alaska.

Chapters 3, 5, 8, 9, 13, 14, and 15 are "must-read" chapters for users and providers of topographic lidar data. Chapter 8 addresses linear mode, single photon and Geiger mode lidar technologies, and Chapter 10 addresses the latest in topo-bathymetric lidar. The remaining chapters are either relevant to all DEM technologies or address alternative technologies including photogrammetry, IfSAR, and sonar.

As demonstrated by the figures selected for the front cover of this manual, readers will recognize the editors' vision for the future – a 3D Nation that seamlessly merges topographic and bathymetric data from the tops of the mountains, beneath rivers and lakes, to the depths of the sea.

Co-Editors

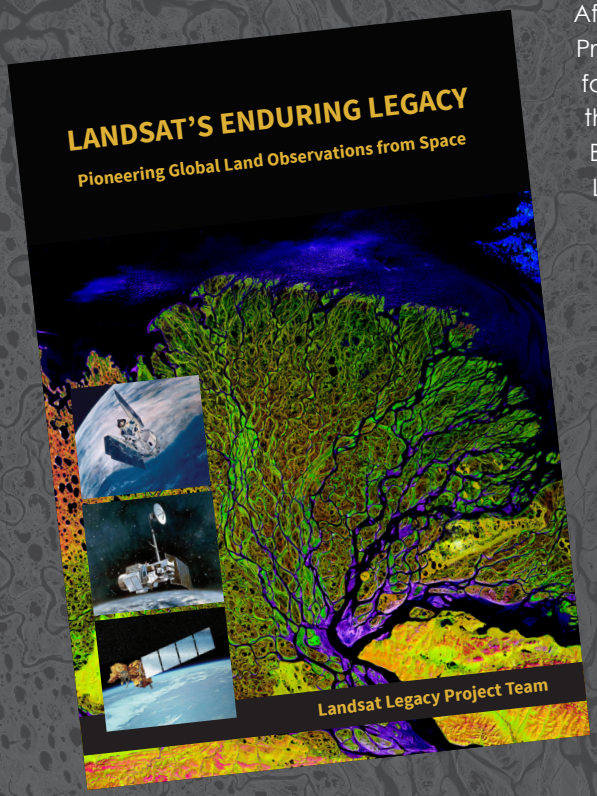
David F. Maune, PhD, CP and
Amar Nayegandhi, CP, CMS

PRICING

Student (must submit copy of Student ID)	$50 +S&H
ASPRS Member	$80 +S&H
Non-member	$100 +S&H
E-Book (only available in the Amazon Kindle store)	$85

Landsat's Enduring Legacy

Pioneering Global Land Observations from Space



After more than 15 years of research and writing, the Landsat Legacy Project Team published, in collaboration with the American Society for Photogrammetry and Remote Sensing (ASPRS), a seminal work on the nearly half-century of monitoring the Earth's lands with Landsat. Born of technologies that evolved from the Second World War, Landsat not only pioneered global land monitoring but in the process drove innovation in digital imaging technologies and encouraged development of global imagery archives. Access to this imagery led to early breakthroughs in natural resources assessments, particularly for agriculture, forestry, and geology. The technical Landsat remote sensing revolution was not simple or straightforward. Early conflicts between civilian and defense satellite remote sensing users gave way to disagreements over whether the Landsat system should be a public service or a private enterprise. The failed attempts to privatize Landsat nearly led to its demise. Only the combined engagement of civilian and defense organizations ultimately saved this pioneer satellite land monitoring program. With the emergence of 21st century Earth system science research, the full value of the Landsat concept and its continuous 45-year global archive has been recognized and embraced. Discussion of Landsat's future continues but its heritage will not be forgotten.

The pioneering satellite system's vital history is captured in this notable volume on Landsat's Enduring Legacy.

Landsat Legacy Project Team

Samuel N. Goward
Darrel L. Williams
Terry Arvidson
Laura E. P. Rocchio
James R. Irons
Carol A. Russell
Shaida S. Johnston

Landsat's Enduring Legacy

Hardback. 2017, ISBN 1-57083-101-7

Member/Non-member	$48*
Student Member	$36*

* Plus shipping

Order online at

www.asprs.org/landsat

asprs THE IMAGING & GEOSPATIAL INFORMATION SOCIETY