PE&RS

August 2023 Volume 89, Number 8

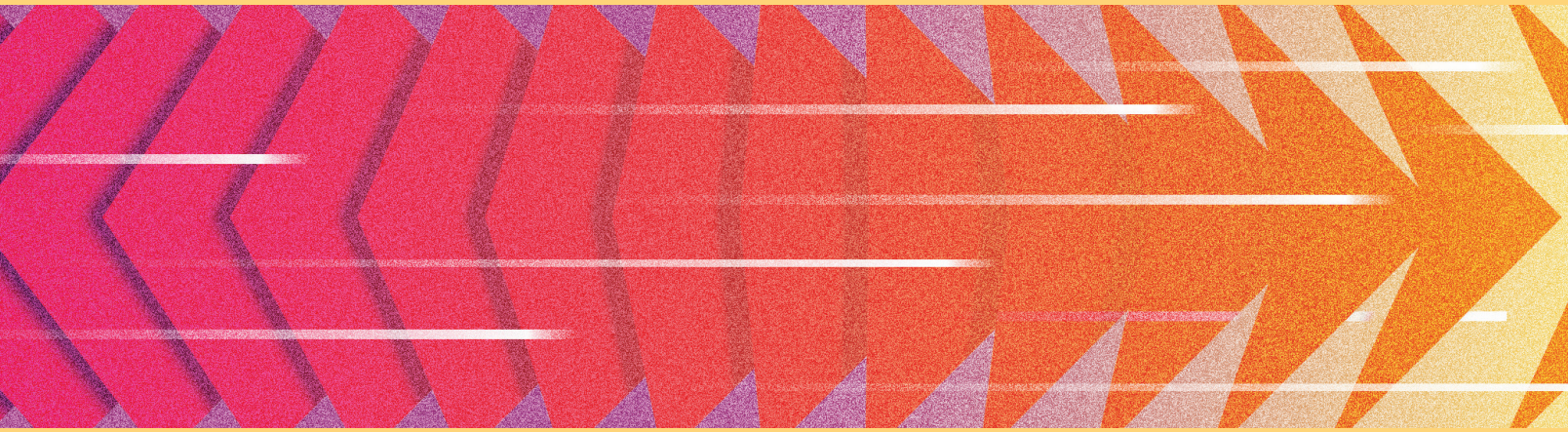
Photogrammetric Engineering & Remote Sensing The official journal for imaging and geospatial information science and technology

JOIN ASPRS TODAY!

asprs THE IMAGING & GEOSPATIAL INFORMATION SOCIETY

ACCELERATE YOUR CAREER!

PHOTOGRAMMETRY · REMOTE SENSING · GIS · LIDAR · UAS …and more!



LEARN

- Read our journal, *PE&RS*
- Attend professional development workshops, GeoBytes, and online courses through the ASPRS ProLearn platform
- Earn professional development hours (PDH)
- Attend our national & regional meetings and conferences

DO

- Write for *PE&RS*
- Innovate to create new geospatial technologies
- Present at our national & regional meetings and conferences
- Engage & network

GIVE

- Participate in the development of standards & best practices
- Influence state licensure through our NCEES affiliation
- Mentor colleagues & support students
- Educate others about geospatial science & technology

BELONG

- Establish yourself as a geospatial expert
- Grow business relationships
- Brand yourself and your company as geospatial leaders
- Connect to the world via our affiliation with ISPRS

Don’t delay, join today at **asprs.org**

ANNOUNCEMENTS

New Partnership Agreement Enables Data Scientists Greater Flexibility to Scale Esri's ArcGIS Capabilities

- Data scientists must often transform gigabytes or terabytes of data—collected from trillions of sources—into valuable insights by finding new ways to make their data work harder. Spatial analysis has become a critical piece of the data science toolkit, enriching all forms of data with geographic context. This allows analysts to uncover patterns, make predictions, and optimize workflows to create operational efficiencies. To support the needs of these professionals, **Esri** the global leader in location intelligence, has partnered with Databricks, the data and AI company. The collaboration will provide users with the advanced spatial analytics capabilities of Esri's ArcGIS software easily accessible in Databricks' big data platform, the Databricks Lakehouse Platform.

"The Databricks Lakehouse Platform merges data engineering, data science, machine learning, and analytics within a single platform and, in many cases, customers also need a powerful spatial component," said Roger Murff, VP of product partnerships at Databricks. "By having geoanalytics tools run directly in Databricks, it empowers those users to perform spatial analysis at a scale they have come to expect with Databricks."

Esri offerings that integrate with the Databricks Lakehouse Platform include ArcGIS GeoAnalytics Engine and Big Data Toolkit, which are built specifically to enable users to perform spatial analysis on big datasets.

"Organizations and data practitioners using Databricks need access to spatial tools where they do their day-to-day work," said Richard Cooke, Esri director of global business development. "The solutions that Esri is providing allow data scientists to access the most comprehensive set of spatial analytics tools available natively within the Databricks environment."

To learn more visit https://www.esri.com/about/newsroom/announcements/.

AlphaGeo UK, a reputable reseller of UAV's, UAV sensors, surveying equipment, and software solutions, is pleased to announce its partnership with GeoCue as the latest distributor in the **GeoCue** global network. AlphaGeo UK will provide cutting-edge TrueView 3D Imaging Sensors and LP360 LiDAR Processing Software to customers throughout the United Kingdom.

Since its establishment in 2015, AlphaGeo UK has served a diverse range of customers in various sectors, including Hazmat, Mining, AECO (Architecture, Engineering, Construction, and Owner-operated), Forestry, Heritage, Disaster management, utilities, and Marine (Off and onshore). As a highly experienced and reputable reseller, AlphaGeo UK offers expertise in drones/UAVs, payloads, ROVs (underwater drones), survey equipment (both land and hydrographic), and software solutions.

"We see our partnership with GeoCue as a great opportunity to showcase our expertise and provide our customers with even more value," said Adrian Fowler, Technical Sales Manager at AlphaGeo UK. "The TrueView and LP360 products impressed us with their innovative drone lidar and photogrammetry solutions integrated into lightweight payloads. They enable fast, easily automated generation of true 3D colorized point clouds, oblique imagery, and orthophotos from drone flights. This revolutionizes the way our customers work and allows them to digitize assets like never before."

The TrueView LiDAR products and LP360 software offer numerous benefits to AlphaGeo UK's customers, enabling the creation of highly accurate 3D models and facilitating the digitization of inaccessible environments or features from the ground. This is especially valuable for surveyors in need of fast, repeatable surveys of either hard-to-reach or vast areas. The GeoCue products and software will primarily cater to B2B clients, including commercial drone users, surveyors, and surveying companies, seeking comprehensive 3D mapping solutions.

Fowler further expressed his excitement about the partnership, stating, "I'm especially thrilled by the lidar rental option offered by GeoCue, which gives customers the opportunity to try before purchasing. It demonstrates GeoCue's commitment to customer satisfaction and aligns with our own philosophy of providing exceptional service and value."

AlphaGeo UK aims to support its customers by offering attractive finance options, allowing companies to acquire the necessary equipment for their growth and success while spreading the costs over affordable monthly payments. This approach enables businesses to conserve cash and invest wisely in their operations.

Abdel Diani, the Director of Global Distribution for GeoCue, expressed his enthusiasm for the partnership with AlphaGeo UK. "We are delighted to welcome AlphaGeo UK to our global network," said Diani. "Their expertise and reputation in the UK market make them an ideal partner for introducing TrueView 3D Imaging Sensors and LP360 LiDAR Processing Software to a broader customer base. We look forward to a successful partnership and the growth it will bring."

For more information, visit https://geocue.com/contact/general-inquiries.

NV5 Global, Inc. announced that its newly acquired subsidiary, Axim Geospatial, LLC ("Axim"), was awarded a prime contract with the National Geospatial-Intelligence Agency (NGA) under the GEOINT Enterprise Operations Service and Solutions Program with Industry, Core Mission Operations (GEO-SPI B).

GEO-SPI B is a seven-year, multi-award contract vehicle valued up to $900 million and focused on delivering trusted geospatial intelligence (GEOINT) to NGA's Analysis and Source Directorates. The GEO-SPI B award is Axim Geospatial's largest contract win to date. GEO-SPI B is the follow-on contract to NGA's Multi-Intelligence Analytics and Collection Support Services (MACSS) contract, the largest geospatial staff augmentation program in the agency.

Under the GEO-SPI B contract, Axim will apply domain expertise, automation, and innovative solutions to improve outcomes and enhance NGA performance across multiple mission sets. Axim's automation will include process automation, computer vision capabilities, and algorithms to address mission specific challenges, capabilities that are now coupled with NV5's broader capabilities to deliver high-end machine learning analytics and maritime domain analysis.

"We are pleased to have been selected by NGA to deliver classified and unclassified GEOINT collection and discovery, exploitation production, and enhancement in support of the GEO-SPI B contract," said Dickerson Wright, PE, Chairman and CEO of NV5. "Axim has built a reputation over the past 20 years supporting NGA's Janus, Economic Indicator Monitoring, and GeoPhysical Data Purchasing programs, and we look forward to expanding our support of NGA core mission operations through the GEO-SPI B program."

For more information, contact ir@nv5.com.

RIEGL Laser Measurement Systems GmbH and Schiebel have successfully completed the integration of a high-end laser scanning system, the RIEGL VQ-840-G topo-bathymetric lidar sensor, on the Schiebel CAMCOPTER® S-100 Unmanned Air System (UAS).

Operating a high-end laser scanning system remotely on an Unmanned Air Vehicle (UAV) requires a tailored solution going beyond what is currently available off-the-shelf. In order to maintain the broad operating range of the UAS, it is imperative to keep the weight of the sensor payload low. In addition, the effective execution of the survey mission requires full remote control of the payload instruments and real-time feedback to the operator via a data link.

The compact topo-bathymetric laser scanner was designed for use in a variety of maritime and hydrographic environments. The lidar sensor payload system is controlled remotely via a data link, which is a crucial for the integration into the S-100 system.

The scanner is controlled by using the onboard software "RiACQUIRE-Embedded" via the available data link; data acquisition and laser safety are also monitored. Once the survey is completed, the raw data seamlessly integrates into the RIEGL data processing workflow.

The RIEGL VQ-840-G, combined with the outstanding technical specifications and performance of the CAMCOPTER® S-100 UAS, enables an efficient and secure way for surveying shallow waters, where monitoring from boats becomes a challenge. The applications of Airborne Lidar Bathymetry (ALB) include the mapping of coastlines and river banks as well as the monitoring of natural habitats, water reservoirs and hydraulic engineering applications (e.g., canals, dams, bridges). In a single data acquisition mission, data below and above the water surface are covered.

Additionally, the topographic laser scanners RIEGL VUX-1UAV/-LR and VUX-12023 can be integrated in the front payload bay of the CAMCOPTER® S-100.

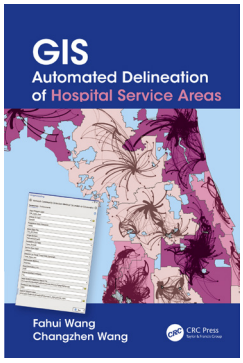
For more information, visit www.rieglusa.com.

CALENDAR

- 5-7 September, **Commercial UAV Expo**; Las Vegas, Nevada; www.expouav.com.
- 16-19 October, **GIS-Pro 2023**, Columbus, Ohio; www.urisa.org/gis-pro.
- 6-8 November, **GoGeomatics Expo**, Calgary, Alberta, Canada; https://gogeomaticsexpo.com.
- 8-10 November, **Smart GEO Expo 2023**, Seoul, South Korea; https://smartgeoexpo.kr.
- 27 November - 1 December, **URISA GIS Leadership Academy**, Denver, Colorado; https://urisa-portal.org/page/URISA_GLA.
- 11-13, February 2024, **Geo Week**, Denver, Colorado; https://www.geo-week.com.
- 13-16 May 2024, **Geospatial World Forum**, Rotterdam, The Netherlands; https://geospatialworldforum.org.

PE&RS
PHOTOGRAMMETRIC ENGINEERING & REMOTE SENSING
The official journal for imaging and geospatial information science and technology
August 2023 Volume 89 Number 8

461 **GIS Tips & Tricks — When Labels and Text Matter**
By Al Karlin, Ph.D. CMS-L, GISP



465 **Book Review — *GIS Automated Delineation of Hospital Service Areas***
By Demetrio P. Zourarakis, PhD, GISP, CMS

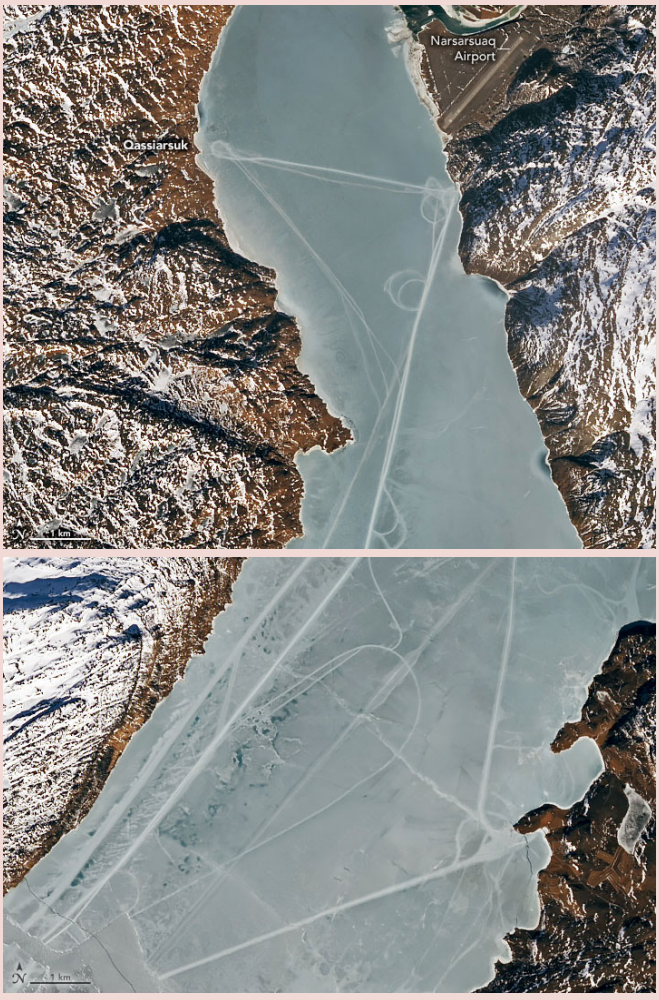
COLUMNS

ANNOUNCEMENTS

DEPARTMENTS

See the Cover Description on Page 460

COVER DESCRIPTION



Greenland lacks an obvious human fingerprint when viewed from space. Instead of sprawling cities or geometrically organized agriculture, an enormous ice sheet spans much of the island. But there is a fjord along the island's southern perimeter where seasonal ice has temporarily revealed the presence of people.

On March 13, 2023, the Operational Land Imager-2 (OLI-2) on Landsat 9 acquired these natural-color images of Tunulliarfik Fjord. The images have been pan-sharpened to bring out more detail.

At the time, the fjord's waters were capped with a layer of sea ice. Atop the ice there are several long, straight lines connecting the towns of Narsarsuaq and Qassiarsuk, and running the length of the fjord. Though cracks and ridges can form naturally in the ice, there are indications that people were likely involved with producing some of the tracks pictured here.

Nathan Kurtz and several other scientists in the Cryospheric Sciences Laboratory at NASA's Goddard Space Flight Center reviewed the images. Their consensus was that many of the tracks, especially those that include loops, were likely produced by vehicles, such as snowmobiles. "It seems likely that residents from the town or nearby were out on the ice, possibly hunting or traveling," Kurtz said.

Traversing the frozen fjord with a vehicle would make sense given its size. The track that connects the small towns of Qassiarsuk and Narsarsuaq, visible in the first detailed image above, spans about 4 kilometers (2.5 miles). The track that runs the length of the fjord is at least 20 kilometers (12 miles) long.

Mixed amid the geometric shapes, some of the "lines" show more variation. They are particularly notable in the second detailed image above. "These look like natural ice deformation features, most likely cracks or ridging due to tidal motion," Kurtz said.

As spring turns to summer, the snow and ice will loosen its grip on the region. The fjord's waters are actually ice-free much of the year, except for floating icebergs that have broken off from the glacier to the north. And as snow melts from the land, white will be replaced with hints of green. Qassiarsuk is within the Kujataa World Heritage Site, which is recognized as the earliest Arctic landscape to have been farmed.

NASA Earth Observatory images by Lauren Dauphin, using Landsat data from the U.S. Geological Survey. Story by Kathryn Hansen. https://landsat.visibleearth.nasa.gov/view.php?id=151243.

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 $175.00 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

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

When Labels and Text Matter

Introduction

As readers of this column know, one of my constant mantras to my GIS and map making students is "never accept the defaults", and I have written several columns to point out the "defaults" that we so often accept without much thought. Lettering on maps falls into that category; we tend to use the "default" font (Tahoma), maybe change the point size (default is 10 points), maybe adjust the style (italics, bold, etc.) and the more advanced students will put a 1- (default) or 2-point white (default) halo behind black letters and call it a day (or a map). ArcGIS Pro offers a lot more control over the lettering than most users realize. The following Tips and Tricks assume that you know how to engage with the Esri labeling engine in ArcGIS Pro to label features of interest. There are several documents available from Esri (https://pro.arcgis.com/en/pro-app/latest/help/mapping/text/labeling-basics.htm) and videos on YouTube (https://www.youtube.com/watch?v=ca8SD6bBS8Y) to help you become familiar with the label engine. So, here are a few ideas and tips for lettering and labeling on your next map. And, of course, many of these options are available in other GIS software packages.

Tip #1 — Choosing Different Fonts

Although not an absolute rule (nothing ever is absolute), but cartographic convention is to label water bodies (rivers, lakes, the ocean) with a Times New Roman/italics font. Also, it is generally a good practice to limit the number of different fonts on a map to two, one serif and one sanserif, but it is acceptable to use multiple styles, i.e. bold, italics, smallcaps, etc. of those two fonts. For the illustrations below, I made a "Graphics Layer" and added the text "Gulf of Mexico" and "Tampa Bay" to my map in Times New Roman:Bold Italics in Apatite blue (I'll discuss custom colors in next month's column).

Tip #2 — Choosing Font Colors and Halos

Also, a "general" suggestion is to render the text label in a color of the same color family, but darker (not black) or lighter (not white) than the surrounding area being labeled (Figure 1). So, in the case of water, the text would be a lighter (or darker) blue than the surrounding water depending on how much contrast you prefer between the lettering and background.

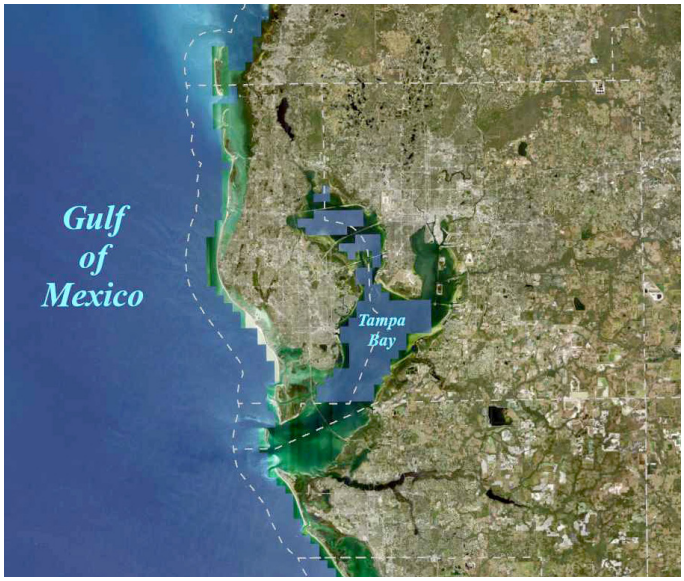


Figure 1. Map of Tampa Bay, Florida area with labels for the Gulf of Mexico and Tampa Bay.

Tip #3

When adding a halo behind the lettering, in general, try different halo sizes (1 point is the default in ArcGIS Pro), colors (white is the default) and transparency levels (default is 0%) to determine which works best for your map. In Figure 2A, I labeled the counties with a 2-point light gray halo and a dark green (Tip #2) Verdana (remember Tahoma is the default font) bold font. In Figure 2A, the county names really stand out. In Figure 2B, I used the same halo and font but with a 50% transparency, to make the county names less prominent.

Tip #4 — Choosing a Different Letter Spacing (Kerning):

Letter spacing, technically called kerning, is a feature that most beginning map makers totally ignore, but adjusting the spacing between letters on labels and in legends, can make a world of difference in legibility. Letter spacing is set on the Label Class | Symbol | Formatting window (Figure 3). The default letter spacing is "0%".

Photogrammetric Engineering & Remote Sensing
Vol. 89, No. 8, August 2023, pp. 461-463.
0099-1112/22/461-463

doi: 10.14358/PERS.89.8.461

Figure 2A. County labels stand out with a 2-point light gray halo surrounding the dark green lettering.

Figure 2B. County labels are less intrusive when the 2-point light gray halo is adjusted to 50% transparent.

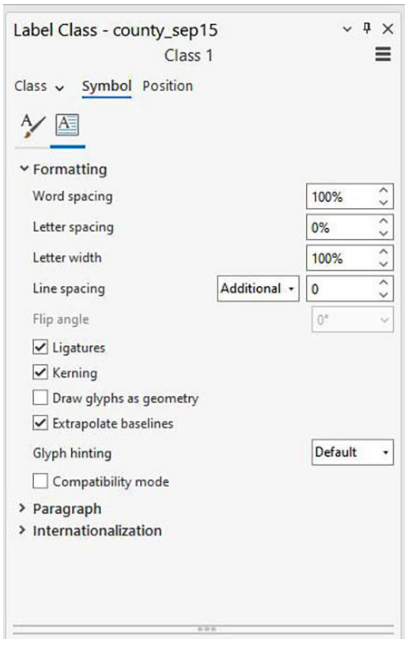


Figure 3. The Label Class | Symbol| Formatting window in ArcGIS Pro provides multiple letter customizations.

In Figure 4, I increased the letter spacing to 35% to make the county names more readable.

Figure 4. County labels are less intrusive when the 2-point light gray halo is adjusted to 50% transparent and more legible when the letter spacing is increased to 35%.

Tip #5

For curved lines (rivers) Esri provides the River Placement style (https://pro.arcgis.com/en/pro-app/latest/help/mapping/text/label-using-the-river-placement-style-for-line-features.htm?rsource=https%3A%2F%2Flinks.esri.com%2FRiverLabels).

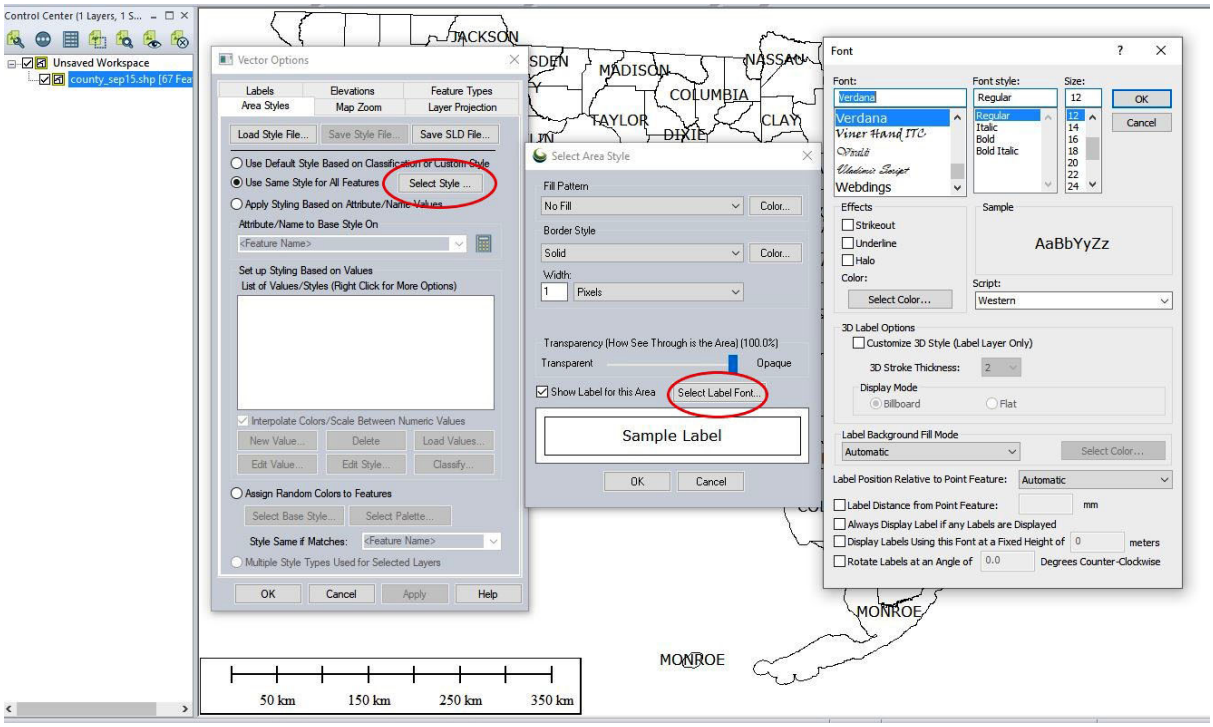


Figure 5. Global Mapper™ menu sequence to access text controls.

For Global Mapper™ Users

To get to many of these same features and controls over text labeling use the Options (right-click) on the layer, then the Select Style… Then Select Label Font… to access the controls as in Figure 5.

For QGIS Users

Label text control is found on the Labels Tab on the Layer Properties menu (Figure 6). Note that to fully enable these features, a label field should be specified.

These are just a few of the of the multitude of labeling features that you can control on your maps. And remember, any/all of these tips also can be applied to your legend text.

Tip #6

Finally, (and this one is directly stolen from well-known Esri cartographers - John Nelson and Ken Field) avoid using the Comic Sans font unless you are making a map for little children.

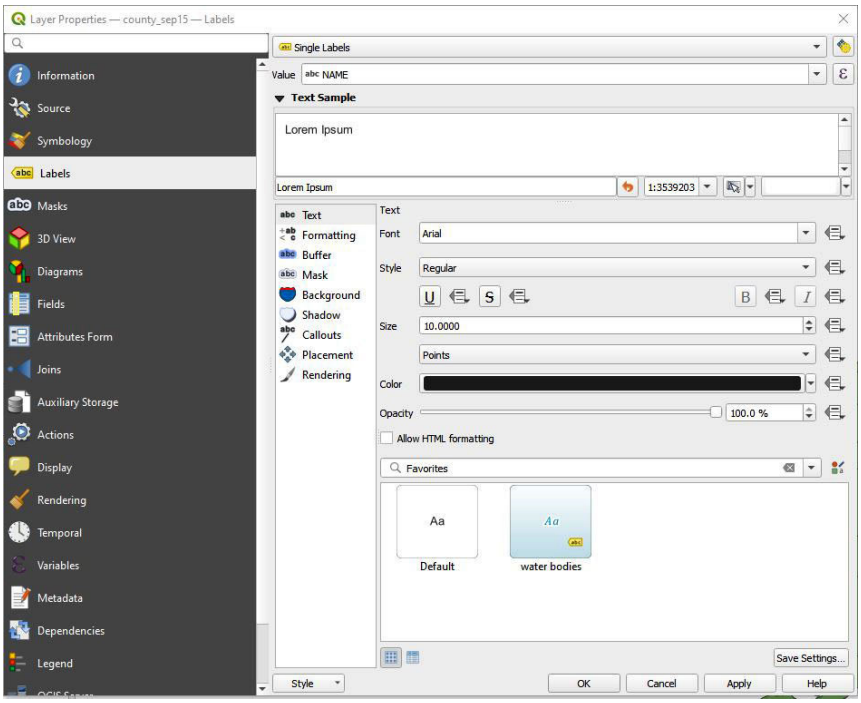


Figure 6. The Layer Properties | Labels tab from QGIS showing text labeling options.

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

Al Karlin, Ph.D., CMS-L, GISP *is a senior geospatial scientist with Dewberry's Geospatial and Technology Services group in Tampa, FL. Al works with all aspects of lidar, remote sensing, photogrammetry, and GIS-related projects. Al also teaches Map-making for the Social Sciences at the University of Tampa.*

The layman's perspective on technical theory and practical applications of mapping and GIS

MAPPING MATTERS

YOUR QUESTIONS ANSWERED

by Qassim Abdullah, Ph.D., PLS, CP
Woolpert Vice President and Chief Scientist

- **Have you ever wondered about what can and can't be achieved with geospatial technologies and processes?**
- **Would you like to understand the geospatial industry in layman's terms?**
- **Have you been intimidated by formulas or equations in scientific journal articles and published reports?**
- **Do you have a challenging technical question that no one you know can answer?**

If you answered "YES" to any of these questions, then you need to read Dr. Qassim Abdullah's column, Mapping Matters.

In it, he answers all geospatial questions—no matter how challenging—and offers accessible solutions.

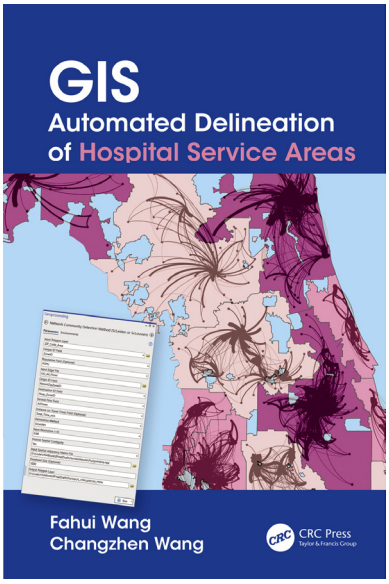
Send your questions to Mapping_Matters@asprs.org

To browse previous articles of Mapping Matters, visit http://www.asprs.org/Mapping-Matters.html

"Your mapping matters publications have helped us a lot in refining our knowledge on the world of Photogrammetry. I always admire what you are doing to the science of Photogrammetry. Thank You Very much! the world wants more of enthusiast scientists like you."

"I read through your comments and calculations twice. It is very clear understandable. I am Honored there are experienced professionals like you, willing to help fellow members and promote knowledge in the Geo-Spatial Sciences."

YOUR COMPANION TO SUCCESS



GIS Automated Delineation of Hospital Service Areas, 1st Edition.

Fahui Wang and Changzhen Wang

CRC Press. Taylor and Francis Group, LLC, Boca Raton, FL. 2022. xvi+208 pages ISBN: 9780367202286 (hardcover). ISBN: 9781032079493 (paperback). ISBN: 9780429260285 (electronic version).

Reviewed by Demetrio P. Zourarakis, PhD, GISP, CMS (GIS, RS, Lidar). Adjunct Assistant Professor, University of Kentucky, CAFE. Visiting Lecturer, Kentucky State University, CACS.

In this first edition of their excellent work, Drs. Wang and Wang provide a roadmap for arriving at the delineation and visualization of hospital service areas (HSAs) by using a variety of geocomputational methods. The wealth of tools and automated geoprocessing workflows presented in this specialized but comprehensive book illustrate how GIS shines at its brightest when it brings to bear its ability to help solve a matter of critical and current importance. The central problem addressed by the authors is how hospitals deliver health services to patients based on both parties' geographic location. All in all, given the depths it explores, the book is elegant and relatively concise with chapters never exceeding thirty-two pages in length. In its seven chapters, it provides the reader with a thoughtfully constructed architecture and precisely written text geared to demonstrate in detail the diverse ways of answering questions based on topological relationships between hospitals and the patients they serve. To do so, the authors first introduce the reader – with clarity and detail – to the necessary mathematical formulations, which are central to modeling the spatial relations of interest, and later use them in software programs tailored to processing the data from case studies.

Occasionally, chapters start out with a discussion connecting them to other chapters, they typically contain five sections, and all have a last section being a rich and articulate summary. The first chapter *Why Hospital Service Areas?* is unique in that it provides an overview of the rest of the chapters, and in that it describes in detail the contents of the GIS datasets and program files, serving as a source of metadata. In the words of the authors, Chapter 2—*Estimating Distance and Travel Time Matrices in GIS*, examines a fundamental task in spatial analysis: estimating distance and travel time. Defining HSAs often requires a distance (or travel time) matrix between patient (resident) locations and hospitals. This chapter and subsequent ones contain screenshots as the authors lead the reader step-by-step through the geoprocessing tasks. The software applications used to calculate origin-destination (OD) travel time matrices are ArcGIS and Google Maps API. Chapter 3—*Analysis of Spatial Behavior of Health Care Utilization in Distance Decay*, examines patterns in health care utilization using the concept of distance decay, by analyzing data from the state of Florida. Chapter 4—*Delineating Hospital Service Areas by the Darmouth Method* starts out with a narrative detailing the background and history of the Darmouth Atlas of Health Care Project and the methods used in that work to delineate HSAs and Hospital Referral Regions (HRRs). Again, the authors show in a tutorial-like way how to create the delineations using the geoprocessing tools provided and applying them to the Florida case study data. Chapter 5—*Delineating Hospital Service Areas by the Huff Model* introduces two different methods to determining HSAs: the Proximal Area Method and the Huff Model, together with mathematical derivations of the formulae used. Implementation of this model in ArcGIS Pro is demonstrated later in the chapter. Chapter 6—*Delineating Hospital Service Areas by Network Community Detection Methods* explains how several types of network community

Photogrammetric Engineering & Remote Sensing
Vol. 89, No. 8, August 2023, pp. 465-466.
0099-1112/22/465-466

doi: 10.14358/PERS.89.8.465

detection methods can help delineate HSAs, by relying on patient service flow datasets between patient residences and hospital locations. The final chapter, Chapter 7—*Delineating Cancer Service Areas in the Northeast Region of the USA* presents the data utilized in the cancer care case study, while applying the concepts introduced and developed in prior chapters – such as distance decay. In closing, in the section entitled *List of Major GIS Datasets and Program Files* the authors provide the reader with online links for accessing datasets, tools and scripts for automated workflows, and their output, for use with a diversity of software. In essence, the authors provide the readership with a toolkit containing multiple tools for delineating HSAs: the Dartmouth method, the Huff model, the two network community detection methods, and several data processing tools that support these HSA delineation methods.

The book also sports two appendices meant as user guides for the estimation of large OD drive time matrices and network flows: Appendix A—*User Guide: Estimating a Large OD Drive Time Matrix* and Appendix B—*User Guide: How to Create Curved-Line and Straight-Line Network Flow Maps*. As in the case of previous chapters, both appendices contain numerous screenshots illustrating the processing results step-by-step.

With this book, exclusively in electronic format, the pleasure of paging through a physical book is absent, and unlike with past reviews, there is also no need to carry voluminous tomes around. To the non-specialist, this is a critical work for understanding the state-of-the-practice of GIS applied to this very real-world problem, particularly in light of the pandemic of recent years, and the current focus and efforts to address service accessibility-related aspects of diversity, equity, and inclusiveness (DEI). At the very least, GIS professionals supporting the mission of agencies and teams in the domains of urban planning and even day-to-day emergency preparedness and response will benefit from the use of this resource.

In closing, this rare jewel of a volume and the methodologies developed, demonstrated, and showcased in it should be the cornerstone for one or more intermediate or advanced courses in GIS curricula.

Journal Staff

NEW ASPRS MEMBERS

ASPRS would like to welcome the following new members!

Dennis A. Bechtel
Anna Burkoff
Jeng Hann Chong
Nicholas Fishburn
Clare Gaffey
Kipper L. Giles
Michael Gribbons
Christopher Heliotes
Tina M. Jackson
Nicholas Jones
Dominic Lee
Amanda J. Miner
Eman Nassar, PhD
Sun Ho Ro
Audrey Roberts
David Alexander Schwab
Melissa Seitzinger
Thomas (Tom) B. Sing, Sr.
Jason Vande Hey
Carson Yinger

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

ASPRS Staff Directory

Membership/*PE&RS* Subscription/ Conferences
Yuki Day
office@asprs.org

Advertising/Exhibit Sales
Bill Spilman
bill@innovativemediasolutions.com

Peer-Review Article Submission
Alper Yilmaz
PERSeditor@asprs.org

Highlight Article Submission
Jie Shan
jshan@ecn.purdue.edu

Feature Article Submission
Featureeditor@asprs.org

Certification
applications@asprs.org

Calendar
calendar@asprs.org

ASPRS Bookstore
office@asprs.org

ASPRS Foundation
foundation@asprs.org

Mailing Address
PO Box 14713
Baton Rouge, LA 70898
301-493-0290, 225-408-4422 (fax)
www.asprs.org

ASPRS Workshop Series

It's not too late to earn Professional Development Hours

Miss one of our Live Online Workshops? You can purchase the workshops now and watch when you are ready!

Check out the workshops offered by visiting
https://asprs.prolearn.io/catalog

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

Edge Detection Method for High-Resolution Remote Sensing Imagery by Combining Superpixels with Dual-Threshold Edge Tracking

Yanxiong Liu, Zhipeng Dong, Yikai Feng, Yilan Chen, and Long Yang

Abstract

Edge detection in high-spatial-resolution remote sensing images (HSRIs) is a key technology for automatic extraction, analysis, and understanding of image information. With respect to the problem of fake edges in image edge detection caused by image noise and the phenomenon of the same class objects reflecting different spectra, this article proposes a novel edge detection method for HSRIs by combining superpixels with dual-threshold edge tracking. First, the image is smoothed using the simple linear iterative clustering algorithm to eliminate the influence of image noise and the phenomenon of the same class objects reflecting different spectra on image edge detection. Second, initial edge detection results of the image are obtained using the dual-threshold edge tracking algorithm. Finally, the initial image edge detection results are post-processed by removing the burrs and extracting skeleton lines to obtain accurate edge detection results. The experimental results confirm that the proposed method outperforms the others and can obtain smooth, continuous, and single-pixel response edge detection results for HSRIs.

Introduction

With the development of earth observation satellite technology, high-spatial-resolution remote sensing images (HSRIS) have been applied to urban planning, military reconnaissance, precision strike, crop monitoring, and so on (Sirmacek and Unsalan 2011; Li *et al.* 2017; Wang *et al.* 2019; Dong *et al.* 2020). Edge detection in HSRIS is a key technology for the automatic extraction, analysis, and understanding of HSRI information and plays an important role in object detection, image segmentation, and ground object extraction of HSRIS (Wang *et al.* 2009; Jin *et al.* 2012; Hong *et al.* 2021). Image edge detection refers to finding the region edge where the gray level in the image changes sharply. Due to the great significance of HSRI edge detection, edge detection in HSRIS has attracted different scholars' research attention (Tan *et al.* 2002; Li *et al.* 2008; Huang *et al.* 2017).

With respect to edge detection for HSRIS, some research has been carried out. Typical first-order differential operators (Sobel [Sobel 1970]; Robert [Duan *et al.* 2005]; Prewitt [Prewitt 1970]; and Kirsch [Kirsch 1971]), second-order different operators (Gauss-Laplace [Marr and Hildreth 1980]), and the Canny algorithm (Canny 1986) are widely used to detect edges in remote sensing images. For example, Ali and Clausi (2001) used the Canny algorithm to extract image edges for the feature extraction and enhancement of remote sensing images. Xi and Zhang (2012) combined the Canny algorithm and the Hough transform to detect the real edges in HSRIS. Shi and Suo (2016) proposed the improved Canny algorithm to detect the edges of HSRIS. The improved Canny algorithm used the Otsu method to specify the threshold adaptively to obtain more continuous edge-detected results. Ye *et al.* (2018) used a fast-guided filter instead of a Gauss filter to reduce the loss of image edge information. Then the improved Sobel operator was used to calculate the gradient and direction of the gradient. Finally, the edges in HSRIS were detected using the Canny algorithm. Liu *et al.* (2014) integrated the Canny algorithm and the locally adaptive thresholding method to extract coastline from remote sensing images. However, the edge detection operators (Sobel, Robert, Prewitt, and so on) are sensitive to image noises (Tan *et al.* 2002; Jin *et al.* 2012). There are a large number of fake edges in edge detection results using the previously mentioned edge detection operators.

Other edge detection methods have also been proposed. For example, Jing *et al.* (2011) proposed an edge detection algorithm based on a global minimization active contour model to obtain continuous and closed edges of an oil slick in an infrared aerial image. Jubai *et al.* (2006) made some improvements to the Pal-King fuzzy edge detection algorithm and proposed an algorithm combing improved fuzzy theory and a genetic algorithm for the edge detection of oil spilled into the sea by remote sensing images. Han *et al.* (2016) took the space-computing capacity of cellular automata and the data pattern search ability of extreme learning machines into account and put forward an extreme learning machine based on cellular automata to detect edges in remote sensing images. Kiani and Sahebi (2015) proposed an edge detection method based on the features of remote sensing images. Frist, spectral thresholds of different regions in the image were determined and extracted in a piecewise manner. Then the edge between these regions was extracted using Shannon entropy. Li *et al.* (2010) put forward a revised parallel beam Radon transform (RPRT) method. An algorithm suitable for straight-edge detection of roads in HSRIS was designed based on the ridgelet transform with the RPRT. Zhao *et al.* (2017) presented an optimal Gabor-based edge detection method for HSRIS. As we know, HSRIS have the characteristics of more detailed spatial and structural information about landscapes (Segl and Kaufmann 2001; Aksoy *et al.* 2010). With spatial resolution refinement, the internal variability within homogeneous land cover units increase. There are the phenomena of the same class objects reflecting different spectra in the HSRIS (Blaschke 2010; Yi *et al.* 2012). Furthermore, there are salt-and-pepper noises in the HSRIS (Yan *et al.* 2006; Myint *et al.* 2011). Due to image noise and the phenomenon of the same class objects reflecting different spectra, there are a large number of fake edges in the image edge detection results, as shown in Figure 1. In Figure 1, the green lines are fake edges, and the yellow lines are real edges. However, the problem of fake edges in image edge detection results caused by image noise

Yanxiong Liu is with the First Institute of Oceanography, Ministry of Natural Resources, Qingdao 266061, China; and the National Engineering Laboratory for Integrated Aero-Space-Ground-Ocean Big Data Application Technology, Xian 710072, China (yxliu@fio.org.cn).

Zhipeng Dong, Yikai Feng, Yilan Chen, and Long Yang are with the First Institute of Oceanography, Ministry of Natural Resources, Qingdao 266061, China..

Corresponding author: Zhipeng Dong (zhipengdong@foxmail.com)

Contributed by Prasad S. Thenkabail, January 15, 2023 (sent for review February 28, 2023; reviewed by Itiya Aneece, Shrinidhi Ambinakudige).

Photogrammetric Engineering & Remote Sensing
Vol. 89, No. 8, August 2023, pp. 469–476.
0099-1112/22/469–476

doi: 10.14358/PERS.23-00003R2

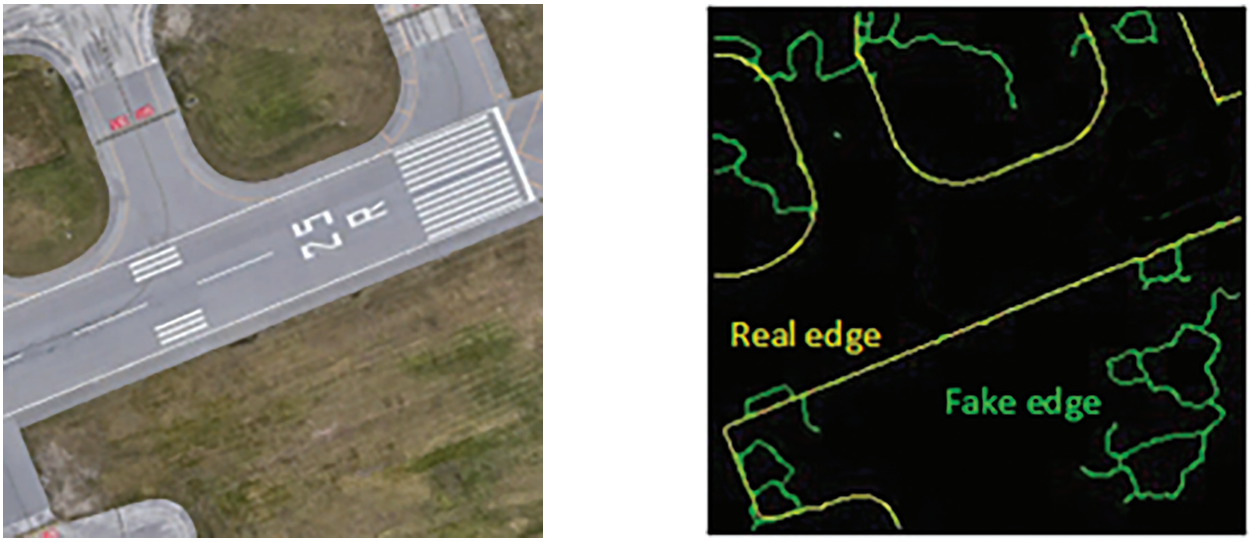


Figure 1. (a) High-spatial-resolution remote sensing images. (b) Edge detection result.

and the phenomenon of the same class objects reflecting different spectrum has not been effectively solved in the literature.

With respect to how to solve the problem of fake edges in image edge detection results caused by image noise and the phenomenon of the same class objects reflecting different spectra, this article proposes a novel edge detection method for HSRIS by combining superpixels with dual-threshold edge tracking. The main contributions of this paper are summarized as follows:

1. The influence of image noise and the phenomenon of the same class objects reflecting different spectrum on image edge detection is eliminated using the simple linear iterative clustering (SLIC) algorithm.
2. A novel edge detection method for HSRIS by combining superpixels with dual-threshold edge tracking is proposed.
3. The smooth, continuous, and single-pixel response edge detection results of HSRIS can be obtained.

The remainder of this article is organized as follows. Next, we demonstrate the performance and characteristics of the key steps to achieve edge detection results for HSRIS. Then we describe the test data and present the results and discussion. Finally, we summarize the results and present conclusions.

Methodology

The proposed edge detection method for HSRIS consists of three steps. First, the image is smoothed using SLIC superpixels. Second, the initial image edge detection results are obtained using the dual-threshold edge tracking algorithm. Third, the initial image edge detection results are post-processed. An overview of the methods used for implementing the three steps is displayed in Figure 2.

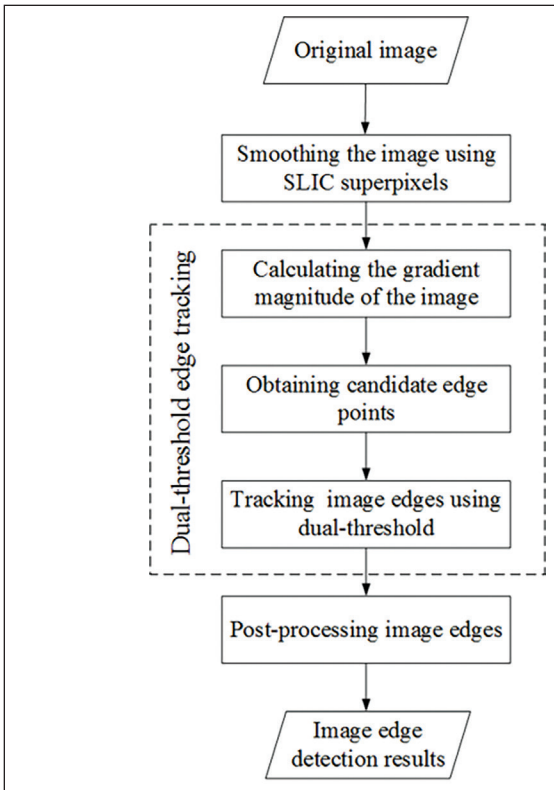
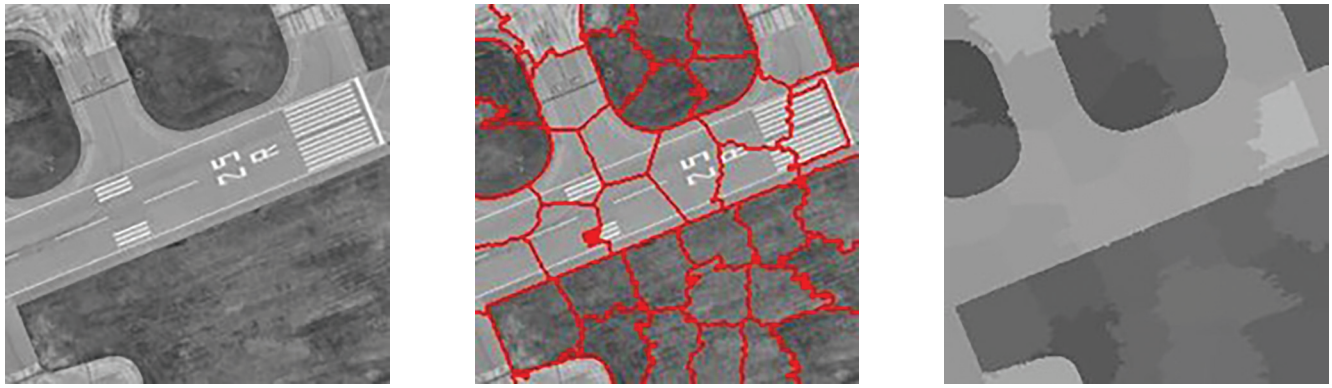


Figure 2. Workflow of the proposed method.

Smoothing Image

Due to image noise and the phenomenon of the same class objects reflecting different spectra, there are a large number of fake edges in the image edge detection results, as shown in Figure 1b. The image needs to be smoothed to eliminate the influence of image noise and the phenomenon of the same class objects reflecting different spectra on edge detection results. An object-based image analysis has become the first choice for HSRI application recently (Yan *et al.* 2006; Myint *et al.* 2011). Furthermore, object-based image processing and analysis is insensitive to spectral variation, which aids in resolving the problem of salt-and-pepper noise effects (Wang *et al.* 2018; Dong *et al.* 2019). Superpixels, which are the image blocks composed of adjacent pixels with similar texture, color, and brightness, were first proposed by Ren and Malik (2003). Achanta *et al.* (2012) proposed the SLIC algorithm for generating superpixels. The SLIC algorithm is simple to use and understand. The color image is converted from an RGB color space to a CIELAB color space. In CIELAB color space, L stands for lightness, a for saturation of red and green colors, and b for saturation of yellow and blue colors. Compared to RGB, CIELAB color space is designed to approximate human vision. It aspires to perceptual uniformity, and its L component closely matches human perception of lightness. It can thus be used to make accurate color balance corrections by modifying output curves in the a and b components or to adjust the lightness contrast using the L component. Another advantage of CIELAB is that the distance that can be calculated between two colors is directly proportional to the difference between the two colors as perceived by the human eye (Ganesan *et al.* 2011).



(a) (b) (c)

Figure 3. (a) Gray image. (b) Simple linear iterative clustering superpixel segmentation result. (c) Image smoothing result.

In CIELAB color space, a pixel color is represented as $[l, a, b]^T$, and $[x, y]^T$ denotes the feature vector of a pixel in the XY coordinates. Each pixel has a five-dimensional feature vector $[l, a, b, x, y]^T$. Image pixels are clustered to generated superpixels using their five-dimensional feature vector. Contrary to the superpixels generated using other image segmentation algorithms, such as the mean-shift algorithm (Comaniciu and Meer 2002), the watershed algorithm (Vincent and Soille 1991), the region growing algorithm, and the graph-based image segmentation algorithm (Felzenszwalb and Huttenlocher 2004), SLIC superpixels can adhere well to image boundaries, and the shape of superpixels is similar and compact. Furthermore, the SLIC algorithm has good noise resistance (Wang *et al.* 2018; Dong *et al.* 2019). Therefore, the image is smoothed using SLIC superpixels in this article.

Image edge detection results are often obtained based on gray image. The gray image of color image is generated using Equation 1 (Dong *et al.* 2019). Figure 3a is the gray image of Figure 1a. The gray image is segmented to generate SLIC superpixels. Figure 3b is the SLIC superpixel segmentation result. Based on the SLIC superpixel segmentation result, image smoothing result is obtained using Equation 2. Figure 3c is the image smoothing result. In Figure 3c, image noises are eliminated, and the phenomenon of the same class objects reflecting different spectra is effectively suppressed. Furthermore, the gradient magnitude of the pixels at real edges is increased:

$$\text{Gray} = R\times 0.299 + G\times 0.587 + B\times 0.114 \tag{1}$$

$$\text{Gray}_{ij} = \frac{\sum_{j=1}^{n_i} \text{Gray}_{ij}}{n_i} \tag{2}$$

Here (R, G, B) are the spectral attributes of the pixel in the color image, Gray is the gray value of the pixel, n_i is the number of pixels included in the label i superpixel, and Gray_{ij} is the gray value of the label j pixel in the label i superpixel.

Obtaining Initial Image Edges

In this article, based on the image smoothing results, initial image edge detection results are obtained using the dual-threshold edge tracking method (Canny 1986). The dual-threshold edge tracking method provides the following three benefits:(1) it accurately locates detected edge pixels, (2) it detects edges only once, and (3) it detects edges while suppressing non-edge pixels as much as feasible. The dual-threshold edge tracking method is expressed as follows.

Calculating Image Gradient

The gradient image G of the image smoothing result is obtained using following equations. First, the gradients of a pixel in the X and Y coordinate directions are calculated using Equations 3 and 4, respectively. Then the gradient magnitude and direction of a pixel are calculated using Equations 5 and 6, respectively:

$$G_x(i,j) = S(i+1,j) - S(i-1,j) \tag{3}$$

$$G_y(i,j) = S(i,j+1) - S(i,j-1) \tag{4}$$

$$G(i,j) = \sqrt{G_x^2(i,j) + G_y^2(i,j)} \tag{5}$$

$$\theta(i,j) = \arctan\left[\frac{G_y(i,j)}{G_x(i,j)}\right] \tag{6}$$

Here (i,j) denotes the location of a pixel in the XY coordinate; $G_x(i,j)$ and $G_y(i,j)$ are the gradient magnitudes of the pixel (i,j) in the X and Y coordinate direction, respectively; $S(i,j)$ is the gray value of the pixel (i,j) in the image smoothing result; and $G(i,j)$ and $\theta(i,j)$ are the gradient magnitude and direction of the pixel (i,j) in the gradient image, respectively.

Obtaining Candidate Edge Pixels

The gradient of edge pixels usually has large magnitude. The local maximum values in the gradient image are obtained using the non-maximum suppression algorithm. The pixels corresponding to local maximum values are as the candidate edge pixels. The process of the non-maximum suppression is as follows.

In the neighborhood of 3 × 3 with pixel (i,j) as the center, the interpolations are calculated along the gradient direction $\theta(i,j)$. If the gradient magnitude of pixel (i,j) is greater than the two interpolations adjacent to it in the gradient direction , the pixel is a candidate edge pixel; otherwise, it is as a non-edge pixel.

Tracking Image Edges

Based on candidate edge pixels, the initial image edges are detected and connected using the dual-threshold edge tracking algorithm. In the dual-threshold edge tracking algorithm, high gradient magnitude threshold T_h and low gradient magnitude threshold T_l are set. Among candidate edge pixels, if the gradient magnitude of the pixel is greater than T_h, the pixel is an edge pixel. If the gradient magnitude of the pixel is less than T_l, the pixel is a non-edge pixel. The pixels which their gradient magnitude is between two thresholds are as suspected edge pixels. For a suspected edge pixel, if there is an edge pixel in the adjacent pixel of the pixel, it is as an edge pixel; otherwise, it is a non-edge pixel.

Post-Processing Image Edges

Initial image edge detection results are obtained using the dual-threshold edge tracking method. Figure 4a shows the initial image edge detection results of Figure 1a. In Figure 4a, there are burrs and cracks in the edge detection results. Therefore, the initial edge detection results need to be post-processed to eliminate burrs and cracks. The post-processing edge detection results are shown in Figure 4b. In Figure 4b, smooth and continuous edge detection results are obtained. The post-processing consists of two steps. First, eight neighborhood pixels centered on the edge pixels of the initial edge detection results are edge pixels. Then the skeleton lines of expanded initial edge detection results are extracted to obtain smooth and continuous edge detection

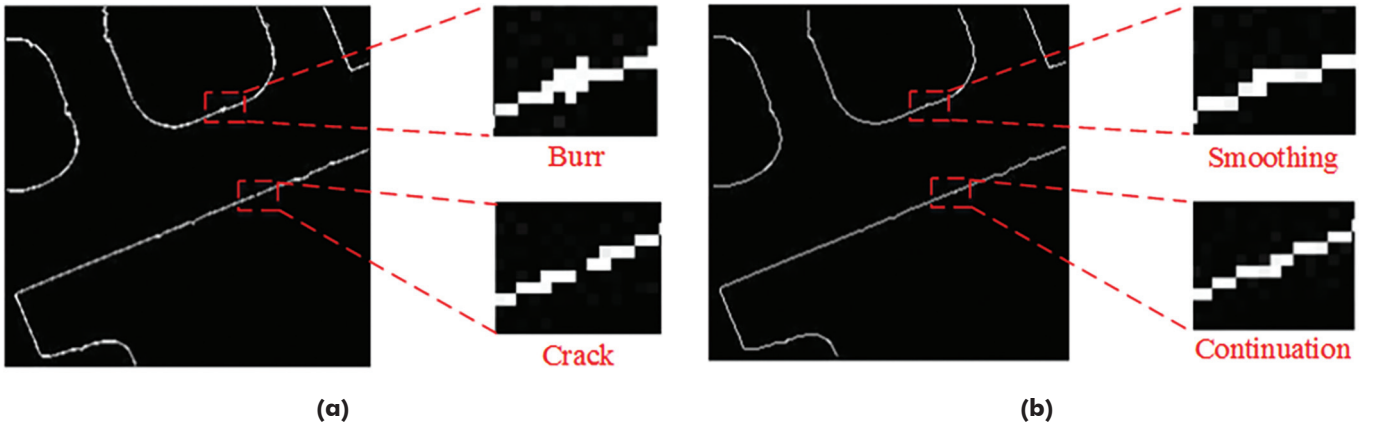


Figure 4. (a) Initial edge detection result. (b) Post-processing edge detection result.

results. The extracting skeleton lines algorithm (He 2002) is expressed as follows.

The *P*-value of the edge pixel in expanded initial edge detection results is set to 0, and the *P*-value of the non-edge pixel is set to 1. The neighborhood of 3 × 3 with the edge pixel as the center is shown in Figure 5, and edge pixel is at P_1. When the edge pixel satisfies the following four conditions at the same time, the edge pixel is changed into a non-edge pixel, and the *P*-value of edge pixel is set to 1. Repeat the judgment for each pixel in expanded initial edge detection results according to the following four conditions until no pixel changes:

1. The sum of the *P*-value of the eight neighboring pixels of the edge pixel is greater than or equal to 2 and less than or equal to 6.
2. Among the eight neighborhood pixels of the edge pixel, only one pair of pixels has a *P*-value of 0–1 in the clockwise direction.
3. $P\text{-value}_2 \times P\text{-value}_4 \times P\text{-value}_8 = 0$, or, among the eight neighborhood pixels, the pixel at P_2, the number of pixel pairs with a *P*-value of 0–1, is not equal to 1 in the clockwise direction.
4. $P\text{-value}_2 \times P\text{-value}_4 \times P\text{-value}_6 = 0$, or, among the eight neighborhood pixels, the pixel at P_4, the number of pixel pairs with a *P*-value of 0–1, is not equal to 1 in the clockwise direction, where $P\text{-value}_2$, $P\text{-value}_4$, $P\text{-value}_6$, and $P\text{-value}_8$ are the *P*-values of the pixels at P_2, P_4, P_6, and P_8, respectively.

Experimental Results

To examine the effectiveness of the proposed method, QuickBird multispectral images and GaoFen-2 (GF-2) panchromatic images are used to compare the proposed method with four commonly used image edge detection algorithms (Sobel, Robert, Gauss-Laplace, and Canny). Experiments are implemented using C++ and performed on a computer with an Intel Core i5-4210U CPU running at 2.40 GHZ with 12 GB RAM using Windows 7.

Edge Detection Results for QuickBird Images

Two 0.61-m-resolution QuickBird multispectral images are obtained from Google Earth, as shown in Figure 6a1 and 6b1. The images contain red, green, and blue spectral bands, which are composed of true color. The size of the images is 600 × 600 pixels (366 × 366 m). Figure 6a1 shows a factory area with buildings, lakes, trees, grassland, roads, and storage tanks. Figure 6a2–a6 show the image edge detection results of Figure 6a1 using the Sobel, Robert, Gauss-Laplace, and Canny algorithms and the proposed method, respectively. In Figure 6a2, there are a lot of burrs in the edge detection results of the Sobel algorithm, as shown in the area in the red rectangle. In Figure 6a3, there are a lot of cracks in the edge detection results of the Robert algorithm, as shown in the area in the red rectangle. In Figure 6a4 and 6a5, there are many fake edges in the edge detection results of the Gauss-Laplace and Canny algorithms. For example, for the ground objects in the rectangle, due to the image noise and different spectral characteristics of the grassland area, there are a lot of fake edges in the edge detection results. In Figure 6a6, the proposed method can obtain smooth, continuous, and single-pixel response image edge detection results.

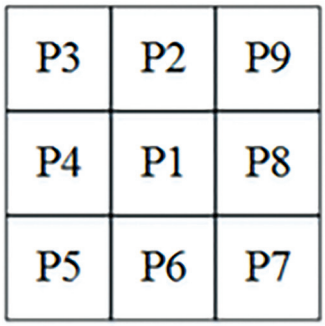


Figure 5. The neighborhood of 3 × 3 with edge pixel as the center.

Moreover, the proposed method can effectively suppress the fake edges and obtain realistic object edge extraction results, as shown in the area in the red rectangle.

Figure 6b1 shows a road network area with roads, lakes, rivers, and trees. Figure 6a2–a6 shows the image edge detection results of Figure 6a1 using the Sobel, Robert, Gauss-Laplace, and Canny algorithms and the proposed method, respectively. In Figure 6b2, 6b4, and 6b5, as the grassland within the red rectangle shows different spectral features, a large number of fake edges and burrs are generated in the edge detection results. In Figure 6b3, there are a lot of cracks and burrs in the edge detection result of the Robert algorithm, as shown in the area in the red rectangle. In Figure 6a6, the proposed method can well suppress the fake edges generated by the phenomenon of the same class objects reflecting different spectra and can accurately extract the real edge of ground objects.

The experimental results show that the edge detection results of the Sobel, Robert, Gauss-Laplace, and Canny algorithms contain many fake edges due to image noise and the phenomenon of the same class objects reflecting different spectra. In addition, the edge detection results contain burrs and fissures and are tough to get single-pixel response on the edge. The proposed method can effectively overcome the influence of the image noise and the phenomenon of the same class objects reflecting different spectra to obtain good image edge detection results. The edge detection results are smooth, continuous, and a single-pixel response on the edge.

Figure 6a1 and 6b1 are QuickBird images. Figure 6a2 and 6b2, 6c2 shows experimental results using the Sobel algorithm. Figure 6a3 and 6b3 shows experimental results using the Robert algorithm. Figure 6a4 and 6b4 shows experimental results using the Gauss-Laplace algorithm. Figure 6a5 and 6b5 shows experimental results using the Canny algorithm. Figure 6a6 and 6b6 shows experimental results using the proposed method.

Edge Detection Results for GF-2 Panchromatic Images

In this section, two GF-2 panchromatic images are used to compare the proposed method with other image edge detection algorithms, as shown

in Figure 7(a1 and 7b1. The spatial resolution of the GF-2 panchromatic images is 0.8 m, and their size is 600 × 600 pixels (480 × 480 m). Unlike the QuickBird image, GF-2 panchromatic images contain only one band information. Through different types of data, the effectiveness and universality of the proposed method are further verified.

Figure 7a1 shows a residential area with buildings, roads, lakes, rivers, and trees. In Figure 7a2 and 7a4, because different areas of the building show different spectral characteristics, fake edges are generated in the edge detection results of buildings within the rectangle by the Sobel and Gauss-Laplace algorithms. In Figure 7a3, a lot of cracks appear in the edge detection results of the Robert algorithm for residential features. For example, the edges of buildings in the rectangle are difficult to detect. In Figure 7a5, there are fake edges and burrs in the edge detection result of the Canny algorithm. In Figure 7a6, the proposed method can effectively suppress the phenomenon that different areas of buildings within the rectangle show different spectral characteristics and accurately extract the true building edges.

Figure 7b1 shows a marine fishery area. In Figure 7b2, 7b4, and 7b5, the Sobel, Gauss-Laplace, and Canny algorithms all generate a lot of burrs and fake edges for the marine fishery areas with good spectral homogeneity in the rectangle. In Figure 7b3, it is difficult to detect the edge of the marine fishery areas in the rectangle by the Robert algorithm. In Figure 7b6, for the marine fishery areas within the rectangle, the proposed method can extract the true feature edges accurately. Moreover, the edge detection results are smooth, continuous, and a single-pixel response.

The above experimental results show that for single-band GF-2 panchromatic images, there are a lot of burrs and fake edges in the edge detection results of the Sobel, Robert, Gauss-Laplace, and Canny algorithms, which makes it difficult to accurately detect the real edge of ground objects. However, the proposed method can effectively overcome burr, crack, and fake edge problems to obtain smooth, continuous, and single-pixel response real edge extraction results for objects in

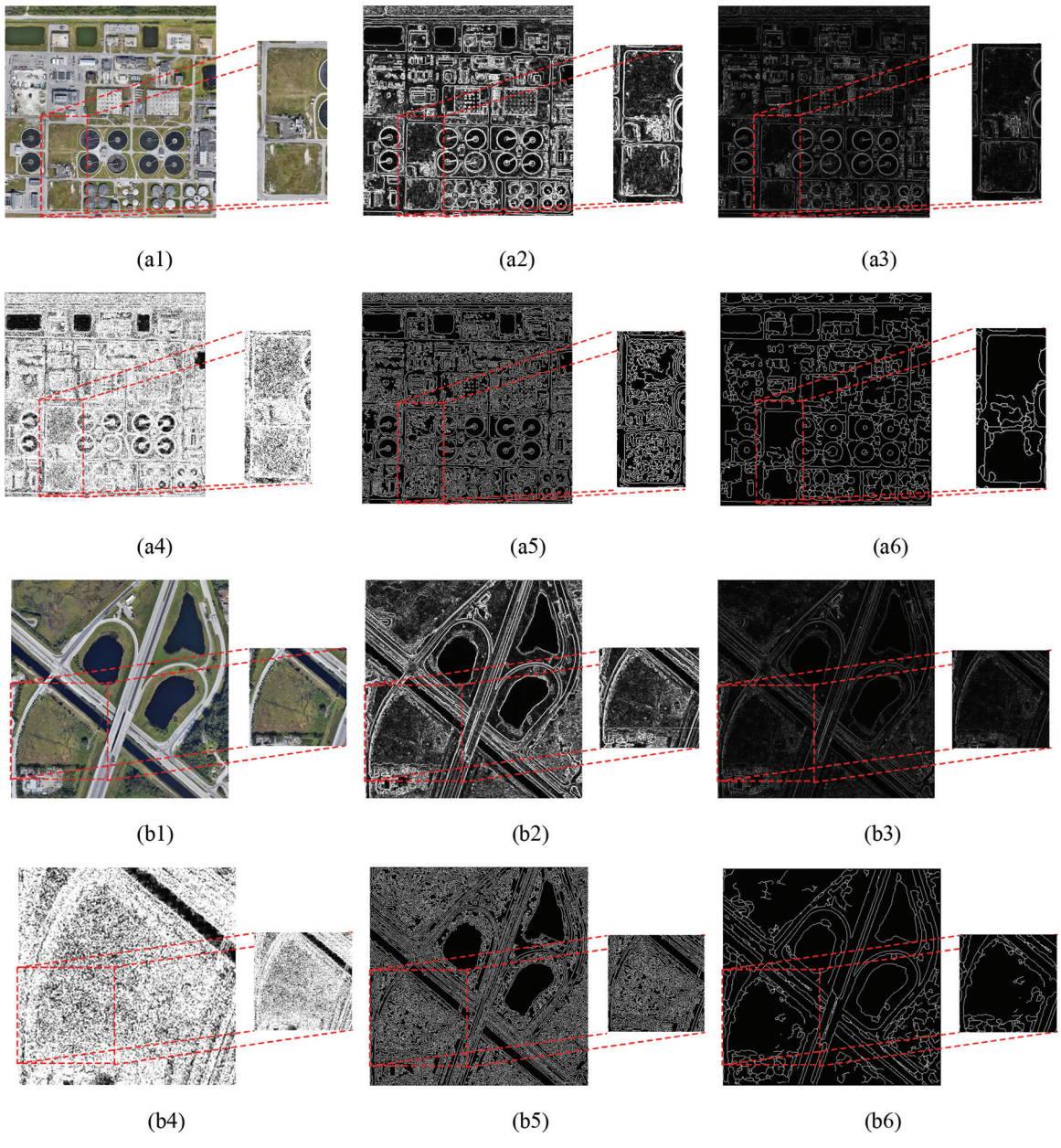


Figure 6. (a1) and (b1) QuickBird images. (a2), (b2), and (c2) Experimental results using the Sobel algorithm. (a3) and (b3) Experimental results using the Robert algorithm. (a4) and (b4) Experimental results using the Gauss-Laplace algorithm. (a5) and (b5) Experimental results using the Canny algorithm. (a6) and (b6) Experimental results using the proposed method.

GF-2 panchromatic images. The proposed method has good robustness and generalizability for different types of image edge detection.

Figure 7a1 and 7b1 shows GF-2 panchromatic images. Figure 7a2 and 7b2 shows experimental results using the Sobel algorithm. Figure 7a3 and 7b3 shows experimental results using the Robert algorithm. Figure 7a4 and 7b4 shows experimental results using the Gauss-Laplace algorithm. Figure 7a5 and 7b5 shows experimental results using the Canny algorithm. Figure 7a6 and 7b6 shows experimental results using the proposed method.

Comparison of the Time Consumption of Five Edge Detection Algorithms

Table 1 shows the average consuming time of edge detection for per image of QuickBird and GF-2 images using five edge detection algorithms. In Table 1, the average consuming times of five edge detection algorithms are 0.17, 0.07, 0.18, 0.38, and 0.36 seconds, respectively. The consuming times of five edge algorithms are less than 1 second. The experimental results show that it is efficient for HSRI edge detection using the proposed method.

Discussion

In this section, the edge detection results of the above QuickBird multispectral images and GaoFen-2 (GF-2) panchromatic images are quantitatively evaluated and discussed. It is proved by mathematical induction that the better the connectivity of edge detection results, the better edge detection results of image (Kitchen and Rosenfeld 1981). The connectivity of edge detection results is quantitatively evaluated by using the ratio of the number of eight-connected components to the number of four-connected components (Lin *et al.* 2003). A four-connected component is a connected edge that is obtained using a four-neighborhood to connect the edge pixel, as shown in Figure 8a. An eight-connected component is a connected edge that is obtained using an eight-neighborhood to connect the edge pixel, as shown in Figure 8b. In this article, B and C represent the number of four-connected components and the number of eight-connected components in the edge detection results, respectively. The value range of C/B is (0, 1]. The smaller the value of C/B is, the better the connectivity of edge detection results is, and vice versa.

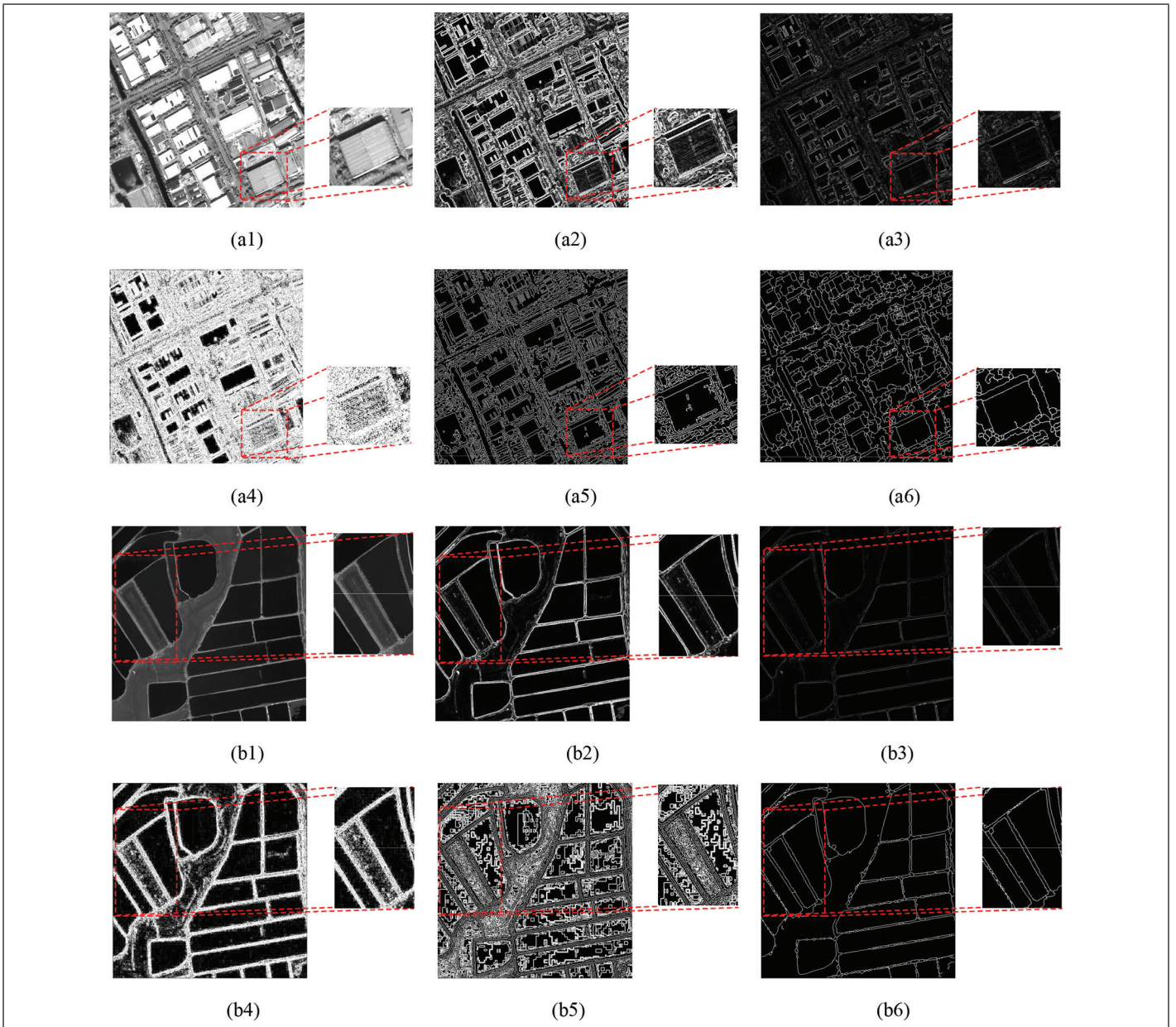


Figure 7. (a1) and (b1) Panchromatic images. (a2) and (b2) Experimental results using the Sobel algorithm. (a3) and (b3) Experimental results using the Robert algorithm. (a4) and (b4) Experimental results using the Gauss-Laplace algorithm. (a5) and (b5) Experimental results using the Canny algorithm. (a6) and (b6) Experimental results using the proposed method.

Edge Connectivity Evaluation for Edge Detection Results of QuickBird Images

The quantitative evaluation results of the edge connectivity of the edge detection results of QuickBird multispectral images are presented in Table 2. Since the proposed method uses the SLIC algorithm to smooth the image, it effectively suppresses the image fake edge caused by image noise and the phenomenon of the same class objects reflecting different spectra. Moreover, the initial edge detection results of the proposed method are post-processed, which effectively reduces burrs and cracks in the edge detection results and improves the connectivity. Therefore, the C value of the proposed method is small. There are a lot of fake edges, burrs, and cracks in the edge detection results of the Sobel, Robert, Gauss-Laplace, and Canny algorithms, and their C values are larger. In Table 2, the C/B values of the proposed method for the edge detection results in Figure 6a1 and 6b1 are 0.07 and 0.04, respectively, which are smaller than other edge detection algorithms. The quantitative evaluation results show that the proposed method has better connectivity than other edge detection results and can obtain better edge detection results for QuickBird images.

Edge Connectivity Evaluation for Edge Detection Results of GF-2 Panchromatic Images

The quantitative evaluation results of the edge connectivity of the edge detection results of GF-2 panchromatic images are presented in Table 3. Since the edge detection results of the proposed method are smooth, continuous, and a single-pixel response for GF-2 panchromatic images, the C value of the proposed method is small. However, there are a lot of fake edges, burrs, and cracks in the edge detection results of the other edge detection algorithms, resulting in their larger C value. Therefore, in Table 3, the C/B values of the proposed method are 0.002 and 0.004, respectively, which is the smallest among the five edge detection algorithms. The experimental results show that the proposed method outperforms the other edge detection algorithms for the edge detection of Figure 7a1 and 7b1.

Conclusion

In this article, an edge detection method for HSRIS by combining superpixels with dual-threshold edge tracking is proposed. The proposed method can solve the problems of the influence of image noise and the phenomenon of the same class objects reflecting different spectra on HSRI edge detection. To verify the effectiveness of the proposed method, QuickBird and GF-2 images containing different ground objects are used to qualitatively and quantitatively compare the proposed method with some image edge detection algorithms. The experimental results show that the proposed method outperforms other image edge detection algorithms and can obtain smooth, continuous, and single-pixel response edge detection results. Furthermore, compared with other image edge detection algorithms, the proposed method can suppress fake edges to the greatest extent and obtain realistic object edge extraction results for HSRIS. In terms of algorithm efficiency, the efficiency of the proposed method is lower than the Sobel, Robert, and Gauss-Laplace algorithms. Going forward, we will optimize the proposed method to improve its operational efficiency.

Acknowledgments

The authors would like to thank anonymous reviewers for their valuable comments, which helped improve this article. This work was supported by the Shandong Natural Science Foundation (grant no. ZR2022QD104), the Shandong Postdoctoral Innovation Project (grant no. SDCX-ZG-202202041), the China Postdoctoral Science Foundation (grant no. 2022M722444), and the National Engineering Laboratory for Integrated Aero-Space-Ground-Ocean Big Data Application Technology Open Fund.

Table 1. The consuming time of five edge detection algorithms.

	Time / Per image (second)
Sobel	0.17
Robert	0.07
Gauss-Laplace	0.18
Canny	0.38
Our method	0.36

Table 2. Quantitative evaluation results of edge detection resutls of Quickbird images.

	Figure 6 (a1)			Figure 6 (b1)		
	B	C	C/B	B	C	C/B
Sobel	3501	1879	0.537	6204	2772	0.447
Robert	3718	2425	0.652	2425	1661	0.685
Gauss-Laplace	996	242	0.243	1026	231	0.225
Canny	19212	3100	0.161	29371	3255	0.111
Our method	4399	29	**0.007**	5176	23	**0.004**

Table 3. Quantitative evaluation results of edge detection resutls of GF-2 panchromatic images.

	Figure 7 (a1)			Figure 7 (b1)		
	B	C	C/B	B	C	C/B
Sobel	2467	1070	0.434	647	359	0.555
Robert	2514	1325	0.527	102	88	0.863
Gauss-Laplace	1520	565	0.372	2580	1522	0.59
Canny	22616	2836	0.125	9082	1289	0.142
Our method	6670	12	**0.002**	2393	10	**0.004**

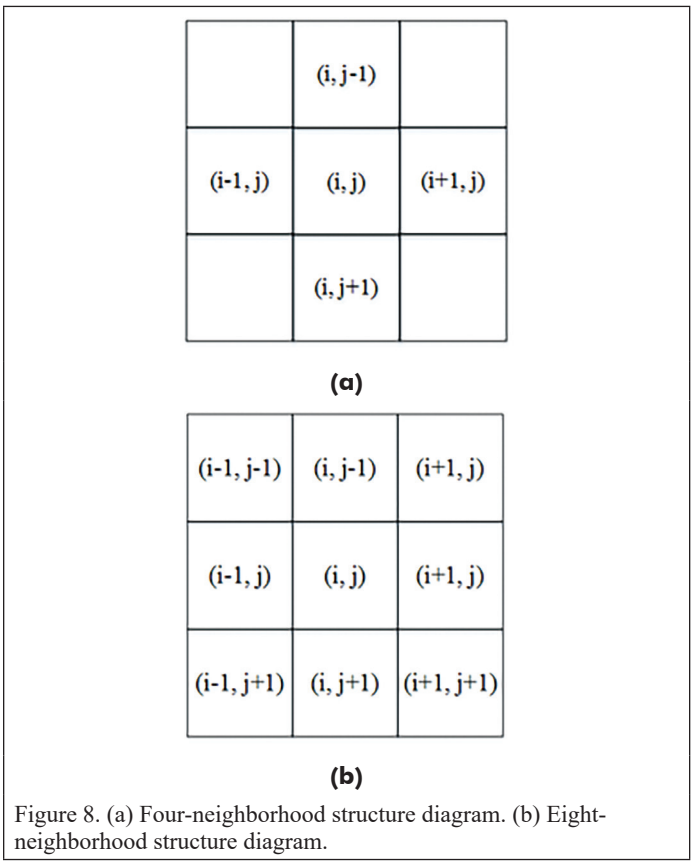


Figure 8. (a) Four-neighborhood structure diagram. (b) Eight-neighborhood structure diagram.

References

Achanta, R., A. Shaji, K. Smith, A. Lucchi, P. Fua and S. Süsstrunk. 2012. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34(11):2274–2282.

Ali, M. and D. Clausi. 2001. Using the Canny edge detector for feature extraction and enhancement of remote sensing image. Pages 2298–2300 in *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, July 2001.

Aksoy, S., H. G. Akcay and T. Wassenaar. 2010. Automatic mapping of linear woody vegetation features in agricultural landscapes using very high resolution imagery. *IEEE Transactions on Geoscience and Remote Sensing* 48(1):511–522.

Blaschke, T. 2010. Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing* 65(1):2–16.

Canny, J. 1986. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8(6):679–698.

Comaniciu, D. and P. Meer. 2002. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(5):603–619.

Dong, Z., M. Wang, D. Li, Y. Wang and Z. Zhang. 2019. Cloud detection method for high resolution remote sensing imagery based on the spectrum and texture of superpixels. *Photogrammetric Engineering and Remote Sensing* 85(4):257–268.

Dong, Z., M. Wang, Y. Wang, Y. Zhu and Z. Zhang. 2020. Object detection in high resolution remote sensing imagery based on convolutional neural networks with suitable object scale features. *IEEE Transactions on Geoscience and Remote Sensing* 58(3):2104–2114.

Duan, L, Q. Li and Y. Li. 2005. Summary of image edge detection. *Optical Technique* 31(3):415–419.

Felzenszwalb, P. F. and D. P. Huttenlocher. 2004. Efficient graph-based image segmentation. *International Journal of Computer Vision* 59(2):167–181.

Ganesan, P., V. Rajini and R. I. Rajkumar. 2011. Segmentation and edge detection of color images using CIELAB color space and edge detectors, in *Proceedings of IEEE International Conference on Emerging Trends in Robotics and Communication Technologies*, December 2010.

Han, M., X. Yang and E. Jiang. 2016. An extreme learning machine based on cellular automata of edge detection for remote sensing images. *Neurocomputing* 198:27–34.

He, B. 2002. *Visual C++ Digital Image Processing*. Beijing: Posts and Telecommunications Press.

Hong, R., J. Park, S. Jang, H. Shin, H. Kim and I. Song. 2021. Development of a parcel-level land boundary extraction algorithm for aerial imagery of regularly arranged agricultural areas. *Remote Sensing* 13:1167.

Huang, L., X. Yu and X. Zuo. 2017. Edge detection in UAV remote sensing images using the method integrating Zernike moments with clustering algorithms. *International Journal of Aerospace Engineering* 5:1–7.

Jin, X., X. Wang, X. Xu and C. Shao. 2012. A robust tensor-based edge detection method for noisy remote sensing images. Pages 229–232 in *Proceedings of IEEE International Conference on Computer Science and Automation Engineering*, May 2012.

Jing, Y., J. An and Z. Liu. 2011. A novel edge detection algorithm based on global minimization active contour model for oil slick infrared aerial image. *IEEE Transactions on Geoscience and Remote Sensing* 49(6):2005–2013.

Jubai, A., B. Jing and J. Yang. 2006. Combining fuzzy theory and a genetic algorithm for satellite image edge detection. *International Journal of Remote Sensing* 27(14):3013–3024.

Kiani, A. and M. R. Sahebi. 2015. Edge detection based on the Shannon entropy by piecewise thresholding on remote sensing images. *IET Computer Vision* 9(5):758–768.

Kirsch, R. A. 1971. Computer determination of the constituent structure of biological images. *Computers in Biomedical Research* 4(3):315–328.

Kitchen, L. and A. Rosenfeld. 1981. Edge evaluation using local edge coherence. *IEEE Transactions on Systems, Man and Cybernetics* 11(9):597–605.

Li, D., M. Wang, Z. Dong, X. Shen and L. Shi. 2017. Earth observation brain (EOB): An intelligent earth observation system. *Geo-Spatial Information Science* 20(2):134–140.

Li, X., S. Zhang, X. Pan, P. Dale and R. Cropp. 2010. Straight road edge detection from high-resolution remote sensing images based on the ridgelet transform with the revised parallel-beam Radon transform. *International Journal of Remote Sensing* 31(19):5041–5059.

Li, Y., Y. Fang and L. Wang. 2008. An algorithm for remote sensing image edge detection based on fuzzy sets. Pages 1087–1090 in *Proceedings of International Symposium on Intelligent Information Technology Application*, December 2008.

Lin, H., N. Shu and C. Zhao. 2003. A new edge evaluation method based on connection components. *Remote Sensing for Land and Resources* 40(3):37–40.

Liu, H. and K. C. Jezek. 2004. Automated extraction of coastline from satellite imagery by integrating Canny edge detection and locally adaptive thresholding methods. *International Journal of Remote Sensing* 25(5):937–958.

Marr, D. and E. Hildreth. 1980. Theory of edge detection. *Proceedings of the Royal Society B: Biological Sciences* 207(1167):187–217.

Myint, S. W., P. Gober, A. Brazel, S. Grossman-Clarke and Q. Weng. 2011. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote Sensing of Environment* 115(5):1145–1161.

Prewitt, J.M.S. 1970. Object enhancement and extraction. In *Picture Processing and Psychopictorics*, edited by B. Lipkin and A. Rosenfeld, 75–149. New York: Academic Press.

Ren, X. and J. Malik. 2003. Learning a classification model for segmentation. Pages 10–17 in *Proceedings of International Conference Computer Vision*, vol. 1, October 2003.

Segl, K. and H. Kaufmann. 2001. Detection of small objects from high-resolution panchromatic satellite imagery based on supervised image segmentation. *IEEE Transactions on Geoscience and Remote Sensing* 39(9):2080–2083.

Shi, G. and J. Suo. 2016. Remote sensing image edge-detection based on improved Canny operator. Pages 652–656 in *Proceedings of 2016 8th IEEE International Conference on Communication Software and Networks (ICCSN)*, June 2016.

Sirmacek, B. and C. Unsalan. 2011. A probabilistic framework to detect buildings in aerial and images. *IEEE Transactions on Geoscience and Remote Sensing* 49(1):211–221.

Sobel, I. E., 1970. *Camera Models and Machine Perception*. Ph.D. dissertation, Stanford University.

Tan, Q., Y. Shao and X. Fan. 2002. A novel edge detection algorithm for remote sensing images based on the self-similarity of fractal character. Pages 2510–2512 in *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, June 2002.

Vincent, L. and P. Soille. 1991. Watersheds in digital spaces: An efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13(6):583–598.

Wang, J., C. Yang and C. Sun. 2009. A novel algorithm for edge detection of remote sensing image based on CNN and PSO. Pages 1–5 in *Proceedings of 2nd International Congress on Image and Signal Processing*, October 2009.

Wang, M., Z. Dong, Y. Cheng and D. Li. 2018. Optimal segmentation of high-resolution remote sensing image by combining superpixels with the minimum spanning tree. *IEEE Transactions on Geoscience and Remote Sensing* 56(1):228–338.

Wang, Y., Z. Dong and Y. Zhu. 2019. Multiscale block fusion object detection method for large-scale high-resolution remote sensing imagery. *IEEE Access* 7(1):99530–99539.

Xi, J. and J. Zhang. 2012. Edge detection from remote sensing images based on Canny operator and Hough transform. In *Advances in Computer Science and Engineering*, vol 141, edited by D. Zeng, 807–814. Berlin: Springer.

Yan, G., J. F. Mas, B.H.P. Maathuis, X. Zhang and P. M. Van Dijk. 2006. Comparison of pixel-based and object-oriented image classification approaches—A case study in a coal fire area, Wuda, Inner Mongolia, China. *International Journal of Remote Sensing* 27(18):4039–4055.

Ye, H., M. Ding and S. Yan. 2018. Improved edge detection algorithm of high-resolution remote sensing images based on fast guided filter. Pages 29–33 in *Proceedings of 2018 IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC)*, December 2018.

Yi, L., G. Zhang and Z. Wu. 2012. A scale-synthesis method for high spatial resolution remote sensing image segmentation. *IEEE Transactions on Geoscience and Remote Sensing* 50(10):4062–4070.

Zhao, H., P. Xiao and X. Feng. 2017. Optimal Gabor filter-based edge detection of high spatial resolution remotely sensed images. *Journal of Applied Remote Sensing* 11(1):015019.

Call for *PE&RS* Special Issue Submissions

Ushering a New Era of Hyperspectral Remote Sensing to Advance Remote Sensing Science in the Twenty-first Century

Great advances are taking place in remote sensing with the advent of new generation of hyperspectral sensors. These include data from, already in orbit sensors such as: 1. Germany's Deutsches Zentrum fur Luftund Raumfahrt (DLR's) Earth Sensing Imaging Spectrometer (DESIS) sensor onboard the International Space Station (ISS), 2. Italian Space Agency's (ASI's) PRISMA (Hyperspectral Precursor of the Application Mission), and 3. Germany's DLR's Environmental Mapping and Analysis Program (EnMAP). Further, Planet Labs PBC recently announced the launch of two hyperspectral sensors called Tanager in 2023. NASA is planning for the hyperspectral sensor Surface Biology and Geology (SBG) to be launched in the coming years. Further, we already have over 70,000 hyperspectral images of the world acquired from NASA's Earth Observing-1 (EO-1) Hyperion that are freely available to anyone from the U.S. Geological Survey's data archives.

These suites of sensors acquire data in 200 plus hyperspectral narrowbands (HNBs) in 2.55 to 12 nm bandwidth, either in 400-1000 or 400-2500 nm spectral range with SBG also acquiring data in the thermal range. In addition, Landsat-NEXT is planning a constellation of 3 satellites each carrying 26 bands in the 400-12,000 nm wavelength range. HNBs provide data as "spectral signatures" in stark contrast to "a few data points along the spectrum" provided by multispectral broadbands (MBBs) such as the Landsat satellite series.

The goal of this special issue is to seek scientific papers that perform research utilizing data from these new generation hyperspectral narrowband (HNB) sensors for a wide array of science applications and compare them with the performance of the multispectral broadband (MBB) sensors such as Landsat, Sentinels, MODIS, IRS, SPOT, and a host of others.

Papers on the following topics are of particular interest:

1. Methods and techniques of understanding, processing, and computing hyperspectral data with specific emphasis on machine learning, deep learning, artificial intelligence (ML/DL/AI), and cloud computing.
2. Issues of hyperspectral data volumes, data redundancy, and overcoming Hughes' phenomenon.
3. Building hyperspectral libraries for purposes of creating reference training, testing, and validation data.
4. Utilizing time-series multispectral data and hyperspectral data over many years to build data cubes and apply advanced computational methods of ML/DL/AI methods and approaches on the cloud.
5. Discussions of hyperspectral data analysis techniques like full spectral analysis versus optimal band analysis.
6. Developing hyperspectral vegetation indices (HVIs) for targeted applications to model and map plant biophysical (e.g., Yield, biomass, leaf area index), biochemical (e.g., Nitrogen, anthocyanins, carotenoids), plant health/stress, and plant structural quantities.
7. Classification of complex vegetation and crop types/species using HNBs and HVIs and comparing them with the performance of multispectral broadband data.

All submissions will be peer-reviewed in line with *PE&RS* policy. Because of page limits, not all submissions recommended for acceptance by the review panel may be included in the special issue. Under this circumstance, the guest editors will select the most relevant papers for inclusion in the special issue. Authors must prepare manuscripts according to the *PE&RS* Instructions to Authors, published in each issue of *PE&RS* and also available on the ASPRS website, https://www.asprs.org/asprs-publications/pers/manuscript-submission.

Special Issue Editors

Dr. Prasad S. Thenkabail, pthenkabail@usgs.gov, thenkabail@gmail.com
Senior Scientist (ST), USGS, Flagstaff, Arizona

Dr. Itiya Aneece, ianeece@usgs.gov
USGS, Flagstaff, Arizona

Dr. Pardhasaradhi Teluguntla, pteluguntla@usgs.gov
USGS, Flagstaff, Arizona

Important Dates

Manuscripts Due — December 15, 2023

Final Papers Due — May 1, 2024

Tentative Publication Date — 2024

Please submit your manuscript — www.editorialmanager.com/asprs-pers/ select "Hyperspectral Remote Sensing"

Expansion of Urban Impervious Surfaces in Lahore (1993–2022) Based on GEE and Remote Sensing Data

Zhenfeng Shao, Muhammad Nasar Ahmad, Akib Javed, Fakhrul Islam, Zahid Jahangir and Israr Ahmad

Abstract

Impervious surfaces are an essential component of our environment and are mainly triggered by human developments. Rapid urbanization and population expansion have increased Lahore's urban impervious surface area. This research is based on estimating the urban impervious surface area (UISA) growth from 1993 to 2022. Therefore, we aimed to generate an accurate urban impervious surfaces area map based on Landsat time series data on Google Earth Engine (GEE). We have used a novel global impervious surface area index (GISAI) for impervious surface area (UISA) extraction. The GISAI accomplished significant results, with an average overall accuracy of 90.93% and an average kappa coefficient of 0.78. We also compared the results of GISAI with Global Human Settlement Layer-Built and harmonized nighttime light (NTL) ISA data products. The accuracy assessment and cross-validation of UISA results were performed using ground truth data on ArcGIS and GEE. Our research findings revealed that the spatial extent of UISA increased by 198.69 km^2 from 1993 to 2022 in Lahore. Additionally, the UISA has increased at an average growth rate of 39.74 km^2. The GISAI index was highly accurate at extracting UISA and can be used for other cities to map impervious surface area growth. This research can help urban planners and policymakers to delineate urban development boundaries. Also, there should be controlled urban expansion policies for sustainable metropolis and should use less impermeable materials for future city developments.

Introduction

The urban impervious surfaces are an essential component of our anthropogenic environment (Tian *et al.* 2018; Vaddiraju and Savitha 2022; Wang *et al.* 2022a) that can be used as a substitute for determining the sustainability of urban development of a city at a regional and global scale. Assessing urbanization and environmental change is a crucial indicator (Wen *et al.* 2021; Wu *et al.* 2022). An increase in urban impervious surface area (UISA) accelerates critical ecological problems such as urban heat islands (Ahmad *et al.* 2022; Manoli *et al.* 2019), urban flooding (Mignot *et al.* 2019; Sattar *et al.* 2020), public health problems, habitat disappearing (Ouyang *et al.* 2022), and the decline of the urban environment (Gao *et al.* 2012).

Generally, urban impervious surfaces are defined as permanent artificial structures that do not allow water to penetrate the ground, such as roads, buildings, and pavements (Sun *et al.* 2015, 2022a; Weng and Hu 2008). It can also include asphalt roads (Chen *et al.* 2022), buildings, stone, concrete, bricks, or unoccupied ground with an impermeable structure (Gong *et al.* 2019). Due to the increase of UISA, most of the vegetation is converted into buildings, societies, and roadways. This conversion of natural green surfaces into urban impervious surfaces caused an increase in air temperature and radiation budget (Kotarba and Aleksandrowicz 2016; Su *et al.* 2022). According to Peroni *et al.* (2022), 392 research articles were published from 2000 to 2020 focused on soil sealing; researchers used several terms including impervious surface, soil consumption, artificial structure, soil sealing, land consumption, and land take, incorporated with multispectral satellite products. But the term impervious surface was used in most articles, and the Landsat satellite data set (Yin *et al.* 2021) was used in 80% of the related research work.

Therefore, mapping urban impervious surfaces in cities is very important to analyze the quality of the urban environment (Liang *et al.* 2022). Among the existing studies, two methods have been applied to map and quantify urban impervious surfaces (1) field survey, ground-based GPS/drone surveys, (2) remote sensing techniques, use of remote sensing technology. Nevertheless, ground-based, in situ measurements, such as a field survey, are costly and time-consuming (Sun *et al.* 2015; Weng 2012). And a conventional survey method covers a small area. But remote sensing data can even cover the globe and use substantial methods, including deep learning, machine learning and artificial intelligence, adding a speedier and more accurate assessment of the UISA.

Remote sensing-based technology provides significant results for urban impervious surface estimation. Remote sensing data from multiscale platforms offer a variety of perspectives for monitoring urban impervious surfaces (Cao *et al.* 2020). Therefore, taking advantage of remotely sensed data to map urban impervious surfaces has become an important topic (Li *et al.* 2022). According to Sun *et al.* (2022b), numerous satellite-based urban impervious surface area products are available. But all the products have different coverage extent, spatial-temporal resolution and data availability.

Among the existing moderate resolution (Wang *et al.* 2022b) satellite data products, Landsat (Kuang *et al.* 2014; Li *et al.* 2016) imagery has proven to be one of the most optimal data sources to generate global products because of its long-term records since 1972, nearly global coverage, and open access at 30 m resolution (Huang *et al.* 2021; Sun *et al.* 2022b). In addition, nighttime light (NTL) imagery (Tang *et al.* 2021) can detect the intensity of human activity and thus can provide valuable prospects for the occurrence of UISA growth (Zhang *et al.* 2020).

Moreover, researchers have used a variety of well-known techniques, including band ratios, supervised segmentation for object-based image analysis, decision tree, threshold-based segmentation, spectral mixture analysis, regression analysis, artificial neural networks, and pixel/sub-pixel-based algorithms (Peroni *et al.* 2022). However, the

Zhenfeng Shao, Muhammad Nasar Ahmad, Akib Javed, Zahid Jahangir and Israr Ahmad are with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, Hubei 43007, China (shaozhenfeng@whu.edu.cn; akibjaved@whu.edu.cn; jahangir93@whu.edu.cn; israr.rs@whu.edu.cn).

Fakhrul Islam is with Khushal Khan Khattak University, Karak, Khyber Pakhtunkhwa, Pakistan (fakhrul.islam@kkkuk.edu.pk).

Corresponding author: Muhammad Nasar Ahmad (mnassarahmad@gmail.com)

Contributed by Desheng Liu, January 8, 2023 (sent for review February 17, 2023; reviewed by Weishu Gong, DongMei Chen).

Photogrammetric Engineering & Remote Sensing
Vol. 89, No. 8, August 2023, pp. 479–486.
0099-1112/22/479–486

doi: 10.14358/PERS.23-00001R2

automatic classification method is most commonly used by researchers for UISA extraction. But still, all the processes have limitations and low accuracy. According to Lu *et al.* (2011), due to the heterogeneous nature of urban environments and the limited spatial resolution, a large number of pixels are mixed with bare soil and frequently result in poor mapping for UISA.

Therefore, any change detection, like urban impervious surfaces, requires tremendous computing power and storage capacity for satellite data. At present, the Google Earth Engine (GEE) is found to be a cloud-based platform with analysis-ready data and high performance (Zhang *et al.* 2020) that can process time series remote sensing data efficiently (Gong *et al.* 2019; Kuang *et al.* 2021). In this research, we used the global urban impervious surface area index (GISAI) developed by (Javed *et al.* 2022), co-author of this research work. The GISAI uses only multispectral bands of satellite imagery and can extract UISA with an overall accuracy of 87.29% (Javed *et al.* 2022).

The GISAI has developed using well-known indices such as modified normalized difference water index (MNDWI), deep clear water body delineation index (DCWDI), and one novel index, modified built-up and bare index (MBBI). Therefore, the GISAI can use remote sensing data to compare cities under specific climatic and environmental conditions. Consequently, we used Landsat time series data on GEE to extract UISA for Lahore, Pakistan. Furthermore, to validate our research outcomes, we used two additional ISA products, including Global Human Settlement Layer (GHSL), available from 1975 to 2030, and Harmonized Nighttime Light (1992 to 2022), as described in Table 2. Two key objectives of this research are (i) identification of UISA change detection from 1993 to 2022 and (ii) cross-validation and comparison of GISAI and other ISA products. This research is noteworthy because it identifies the expansion of UISA in Lahore from 1993 to 2022 for 30 years. It can help researchers to apply this index to other regions and urban planners to make policies to control UISA expansion in future.

Study Area and Data

Study Area

Lahore is the capital city of the Punjab province. The location of the study area on the map is presented in (Figure 1). Lahore city has the highest GDP production and economic activities in the province. Furthermore, this city has different climatic conditions, from low to high temperatures. Lahore is the second-largest city in terms of population among all the megacities of Pakistan. Its population is 12 642 000, and it is located in northeastern Punjab, Pakistan. It has a total area of 1772 square kilometers (Ahmad *et al.* 2022a) with a geographical position of 31.5204° N latitude and 74.3587° E longitude. The city has developed horizontally and consists of a walled city adjacent to urban and commercial areas. Lahore has developed as a technological center with significant commercial, industrial, and trade prospects. It has an average temperature of 33.2 °C in summer and an average of 12.1 °C in winter.

Data

In this study, we have used Landsat time series data for UISA extraction from 1993 to 2022. Furthermore, we have used two ISA products freely available (i) GHSL BUILT-S and (ii) Harmonized Nighttime Light to compare with the results of Landsat data. Table 3 contains information about the data sets used, including their providers, spatial resolution,

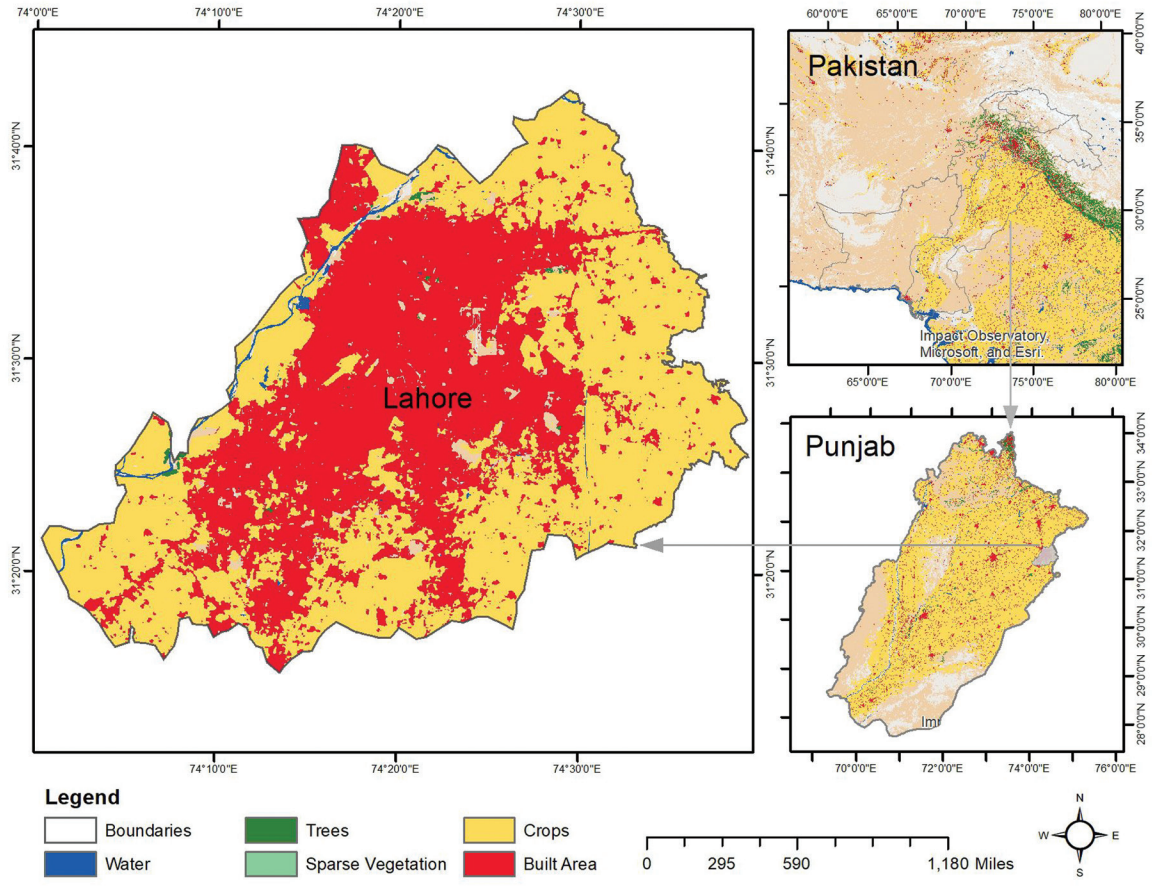


Figure 1. Location of Lahore city on the map.

and temporal availability. Additionally, each data product is further discussed separately.

Landsat

Landsat data provide multiple bands with different wavelengths, but our objective was to focus on UISA. Therefore, we used only selected bands for each Landsat product. Table 1 shows details of Landsat products and bands used to carry out this research.

Landsat products are the longest-running and most continuous space-based satellite imagery that NASA/USGS missions provide. Since 1972, Landsat imagery has been accessible. *Landsat 5* data product has attained a Guinness World Record for the most extended continuously operating earth observation satellite for approximately 29 years (Ahmad *et al.* 2022b). This study used Landsat data from 1993 to 2022 (30 years). Moreover, we have made composite output by combining five years of Landsat images for different year ranges. Table 2 provides the Landsat data acquired through the GEE to make a composite image for each year range.

Global Human Settlement Layer–Built-Up Surface GHSL-BUILT-S

The GHSL was initiated by the Joint Research Centre, the Directorate-General for Regional and Urban Policy, and the European Commission (Liu *et al.* 2020) to map human settlement and built-up surface at a global scale (Schiavina *et al.* 2022). The GHSL creates emerging global spatial data, evidence-based analytics, and knowledge about the existence of humans on Earth. The built-up (BU) surfaces estimated between 1975 and 2030 are available in the Global Human Settlement (GHS)-BUILT-S geographic raster data set together with two practical use components: the total BU surface and the non-residential BU surface. The data created by spatial-temporal interpolation of five observed collections of multiple-sensor, multiple-platform satellite imageries: Sentinel-2 (S2) composite (GHS-composite-S2 R2020A) supports the 2018 epoch, while Landsat (MSS, TM, ETM sensor) supports 1975, 1990, 2000, and 2014 epochs (Schiavina *et al.* 2022; Uhl and Leyk 2022).

Harmonized NTL

Harmonized NTL is calibrated data developed by (Li *et al.* 2020; Li and Zhou 2017). Harmonized NTL data set was created by merging the inter-calibrated NTL observations from the Defense Meteorological Satellite Program (DMSP) data with the simulated DMSP-like NTL observations from the Visible Infrared Imaging Radiometer Suite (VIIRS) data at the global scale. This temporally extended DMSP NTL data set provides helpful assistance for various studies on human activities, including urban impervious surface and urban growth patterns. The data set consists of temporarily calibrated DMSP-OLS NTL time series data from 1992 to 2013 and converted NTL time series made from VIIRS data (2014–2022) with a spatial resolution of 30 arc seconds (1 km). Table 3 provides the details of the harmonized NTL with temporal and spatial resolution.

Methodology

Google Earth Engine and Sentinel Application Platform

Regarding the GEE cloud computing platform, the methodology for this research work includes satellite data fetching and band selection, and performing urban impervious surface extraction based on the GISIA index explained in (Figure 1). The Google Earth Engine provides open access to the Landsat archive, including satellite imagery from *Landsat 1* to *Landsat 9*. This research uses specific bands from *Landsat 4-5* to *Landsat 8* for the last three decades (Table 1). In addition, the open-source platform QGIS was used to validate the image analysis and processing of the satellite data.

Table 1. List of Landsat products and bands acquired on Google Earth Engine for urban impervious surface extraction.

Band Name	*Landsat 9* OLI	*Landsat 8* OLI	*Landsat 7* ETM+	*Landsat 4* and *5* TM
Blue	2	2	1	1
Green	3	3	2	2
Red	4	4	3	3
Near Infrared	5	5	4	4
Short-wave Infrared 1	6	6	5	5
Short-wave Infrared 2	7	7	7	7

OLI = Operational Land Imager; ETM+ = Enhanced Thematic Mapper Plus; TM = Thematic Mapper.

Table 2. Urban impervious surface area (UISA) data sets used in this research.

No	Year (Range)	Composite Image
1	1993–1997	1993
2	1998–2002	1998
3	2003–2007	2003
4	2008–2012	2008
5	2013–2017	2013
6	2018–2022	2018

The authors used a combined Landsat data set consisting of *Landsat 5* TM, *Landsat 7* ETM+, and *Landsat 8* OLI data sets. In the study, only surface reflectance data sets were used and any scan lines can corrector (SLC) failure images from the data set were manually removed. To generate a composite GISAI image, the authors took five yearly composites from the combined Landsat data set. We used a threshold value of 0.1 to separate urban and non-urban areas. It means that any pixel with a value greater than 0.1 was classified as an urban area, while any pixel with a value less than or equal to 0.1 was classified as a non-urban area.

Overall, the methodology involved using Landsat data sets, removing any SLC failure images, generating yearly composites, and classifying urban and non-urban areas using a threshold value. These steps were carried out using GEE, a cloud-based platform for satellite image analysis. GEE is a cloud-based platform that allows studying satellite imagery using various programming languages, including JavaScript and Python.

Preprocessing

This study uses annual image collection from the USGS *Landsat 8* Level 2, Collection 2, and Tier 1 surface reflectance dataset using a data range from January to December each month composite instead of single remote sensing imagery. Landsat's high-quality band Bitmask has been used for atmospheric correction. Table 2 provides information on the data sets used from 1993 to 2022.

We first separate the water bodies using two well-known water indices for urban impervious surface area extraction. An MNDWI was developed (Xu 2006), and a DCWDI was formed by (Yue and Liu 2019). We build a composite free water area (CWFA) using both

Table 3. UISA data sets used in this research.

No.	Name	Download URL	Spatial Resolution	Temporal Resolution
1	GHS-BUILT	https://ghsl.jrc.ec.europa.eu/datasets.php	38 m	1975–2030
2	Landsat	https://earthengine.google.com/	30 m	1993–2022
3	Harmonized NTL	https://samapriya.github.io/awesome-gee-community-datasets/projects/hntl/	30 arc-seconds ~ 1 Km	1992–2022

GHS-BUILT = Global Human Settlement (GHS)-BUILT; NTL = nighttime light.

indexes. Furthermore, the indices and bands used to develop GISAI are presented in Table 4, including MNDWI, DCWDI, CWFA index, Normalized Difference Built-Up Index (NDBI), Normalized Difference Urban Index (NDUI), MBBI, Composite impervious surface area index (CISAI), terraMNDWI, shortwave-infrared (swir)Soil, and GISAI. This research methodology used the minimal MNDWI index composite without water, multiplied by the current ISA composite. Additionally, the water-free MNDWI value distinguishes ISA somewhat from bare soil. Finally, swirSoil eliminates the extremely bright sandy region. Also, the data flow diagram (Figure 2) details indices implementation and sequencing.

Results

This section describes the research outcomes for the urban impervious surface area increase discussed in the methodology section. Furthermore, the results are further explained below separately for UISA extraction.

Urban Impervious Surface Area Change Detection

GISAI demonstrates strong skills for mapping metropolitan regions using the Landsat data set. The study investigated a variety of thresholds and discovered that values between 0.1 and 0.5 can provide substantial results. Additionally, multispectral earth observation satellites and past Landsat missions frequently have multispectral bands. As a result, GISAI can also be applied to numerous satellite imagery products. We have implemented the GISAI to delineate UISA in this research work. The results show an abrupt change in the urban impervious surface area increases from 1993 to 2022 in the selected region.

Furthermore, GISAI, a novel index, was likened to (GHSL) BUILT-S and Harmonized Nighttime Light data to cross-validate our research results. Also, the result shows significant improvement in UISA extraction using GISAI compared to the conventional single index-based mapping techniques used in previous research papers. We can get more valuable results by combining multiple indices to enhance urban impervious surface area segregation from the permeable surface (Javed *et al.* 2022; Zhao and Zhu 2022). Therefore, using the GISAI index, we expressively reduced soil signature and separated bare soil from built-up structures. In this research, we used a threshold value of 0.1, and it can be observed from (Figure 3) that there was an abrupt change in the urban impervious surface area. The increase of urban impervious surface in Lahore has been observed from 211.04 km² to 409.72 km² from 1993 to 2022.

In this research, we used two additional data sets based on NTL observations and GHSL built-up surface to examine the spatiotemporal patterns of ISA over 30 years. Figure 4 show the GISAI, GHSL, and Harmonized NTL data results for an interval of 10 years. Similar to GISAI outcomes, the GHSL-S results are also significant but comparatively not symmetrical. However, the spatial resolution of GHSL data is 38 m, close to Landsat time series data (30 m). It can be observed in Figure 4 that there was no distinct difference observed between 2012 and 2020. So, the result produced by GISAI was more symmetrical at the city level than other ISA products.

The UISA delineated from NTL data with a low spatial resolution provided significant results and showed similar trends like GISAI and GHSL built surface as shown in Figure 4. Because of its extremely low spatial resolution, the Harmonized NTL findings were discovered to be pixel exposed in contrast to GISAI and GHSL results. However, Harmonized NTL works best when covering a larger area, such as a continent or region; it is less effective when mapping at the city scale.

Moreover, a trend analysis was also performed to observe the UISA change in the study area by interpolating the GISAI results from 1993 to 2022 in a single map (Figure 5). From trend analysis, we found that most of the UISA change was observed in 2012 and 2022. It also indicates that UISA increase trends were observed in the east and south-west directions for Lahore city.

Figure 6 shows the overall trend of ISA from 1993 to 2022 based on composite images. Furthermore, the change detection from 1993 to 2022 (Figure 6) shows an overall increase in UISA. The growth rate was the highest for 2013–2017 years.

Table 4. Indices sequencing to achieve global urban impervious surface area index (GISAI) (Javed *et al.* 2022).

$$\text{MNDWI} = \frac{\text{Green} - \text{SWIR1}}{\text{Green} + \text{SWIR1}} \quad (1)$$

$$\text{DCWDI} = \sqrt{\left(\text{Red}^2 + \text{NIR}^2\right)} \quad (2)$$

$$\text{CWFA} = (\text{median MNDWI}_T + \text{median DCWDI}_T) < 1 \quad (3)$$

$$\text{NDBI} = \frac{\text{SWIR1} - \text{NIR}}{\text{SWIR1} + \text{NIR}} \quad (4)$$

$$\text{NDUI} = \frac{\text{SWIR2} - \text{NIR}}{\text{SWIR2} + \text{NIR}} \quad (5)$$

$$\text{MBBI} = \frac{\text{SWIR2} - \text{SWIR1}}{\text{SWIR2} + \text{SWIR1}} \quad (6)$$

$$\text{CISAI} = (\text{min NDBI} + 1) \times (\text{min NDUI} + 1) \times (\text{min MBBI} + 1) \times \text{CWFA} \quad (7)$$

$$\text{terra MNDWI} = (\text{min MNDWI} + 1) \times \text{CWFA} \quad (8)$$

$$\text{SWIR soil} = \text{SWIR1} \times \text{SWIR2} \times 4 \quad (9)$$

$$\text{GISAI} = (\text{CISAI} \times \text{terra MNDWI}) - \text{SWIR soil} \quad (10)$$

MNDWI = modified normalized difference water index; SWIR = shortwave-infrared; DCWDI = deep clear water body delineation index; NIR = near infrared; CWFA = composite free water area; NDBI = Normalized Difference Built-Up Index; NDUI = Normalized Difference Urban Index; MBBI = Modified Bare and Built-up Index; CISAI = Composite Impervious surface area index .

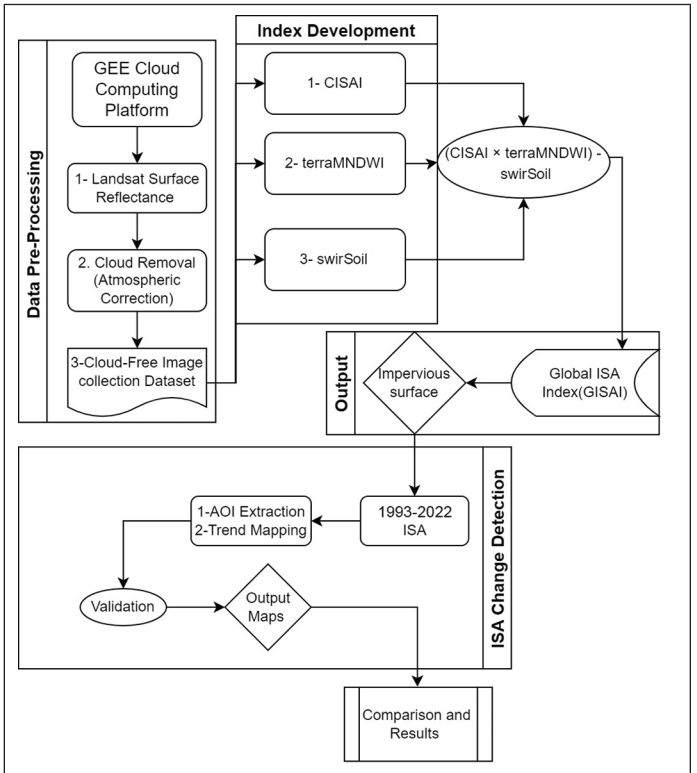


Figure 2. Methodology (data flow diagram). GEE = Google Earth Engine; CISAI = Composite Impervious surface area index ; MNDWI = modified normalized difference water index; AOI = Area of Interest; ISA = impervious surface area.

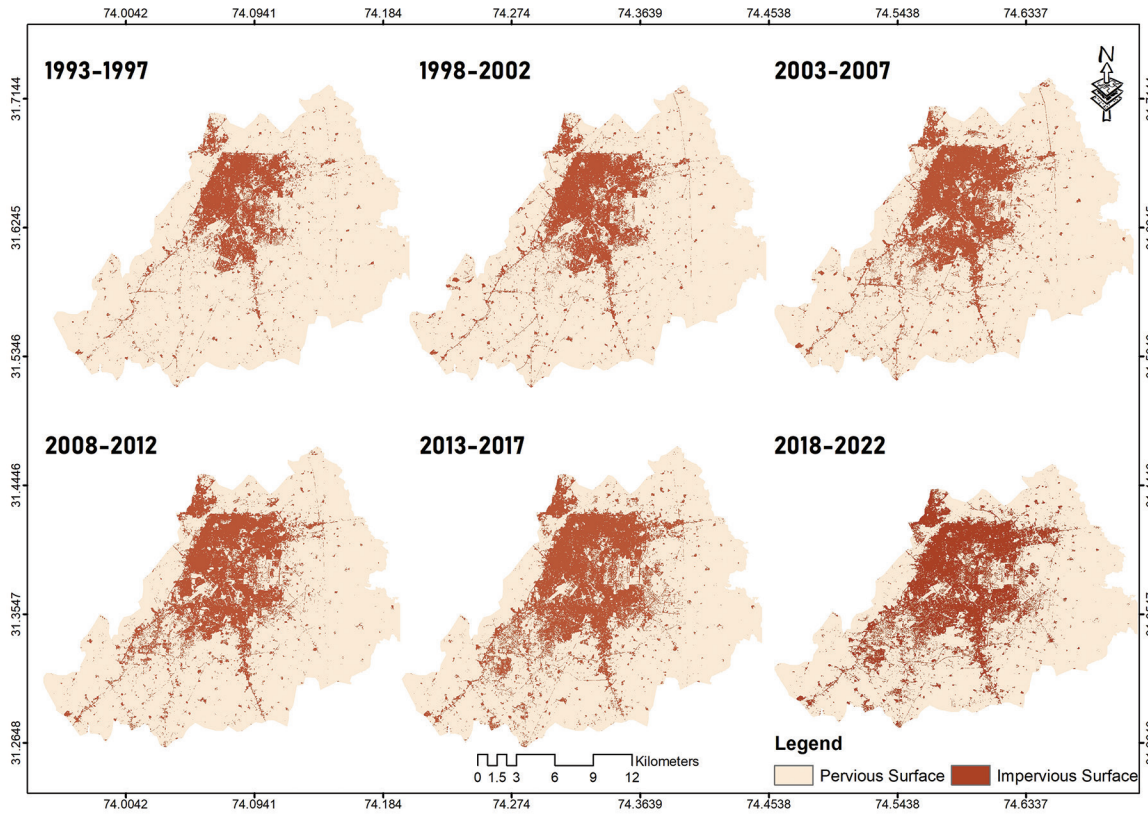


Figure 3. Urban impervious surface area extraction using the global urban impervious surface area index (GISAI) index on the Google Earth Engine.

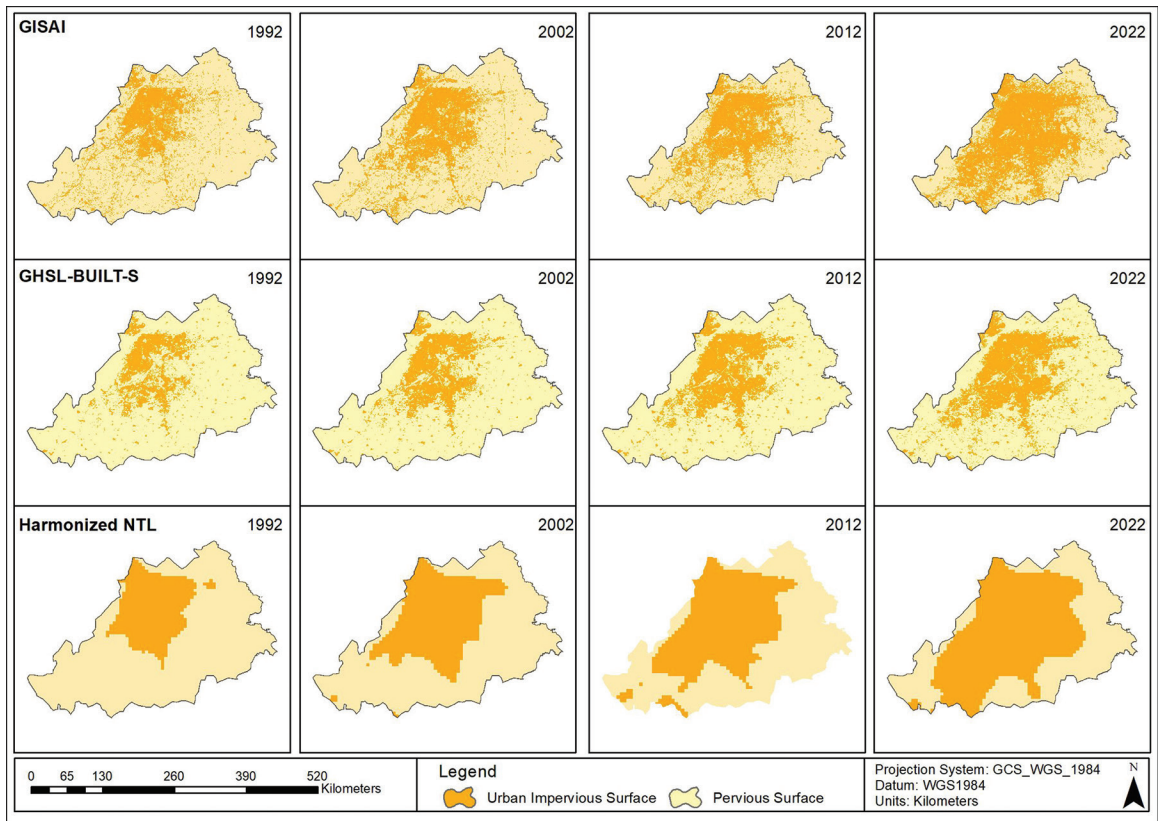


Figure 4. Urban impervious surface area extraction using GHSL-BUILT-S data. GISAI = global urban impervious surface area index; GHSL = Global Human Settlement Layer; NTL = nighttime light.

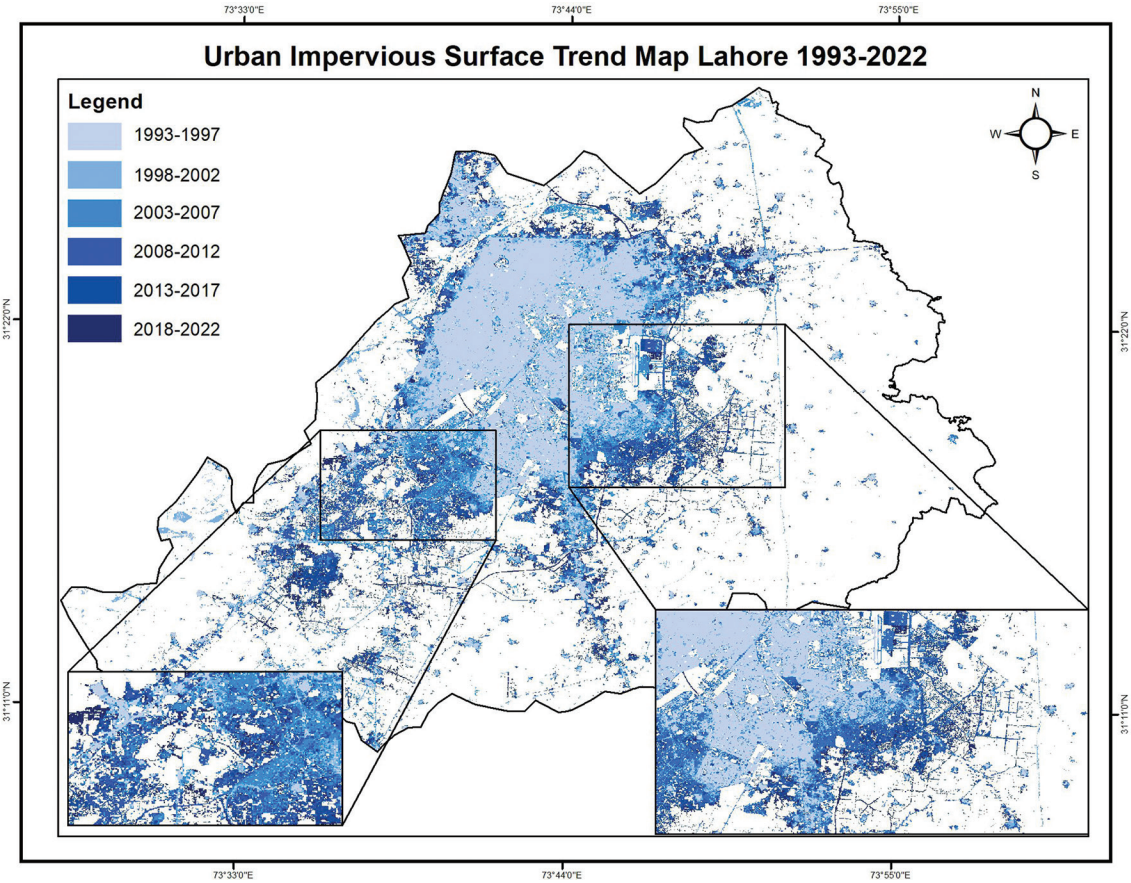


Figure 5. Trend analysis from 1993 to 2022 using the global urban impervious surface area index (GISAI) index.

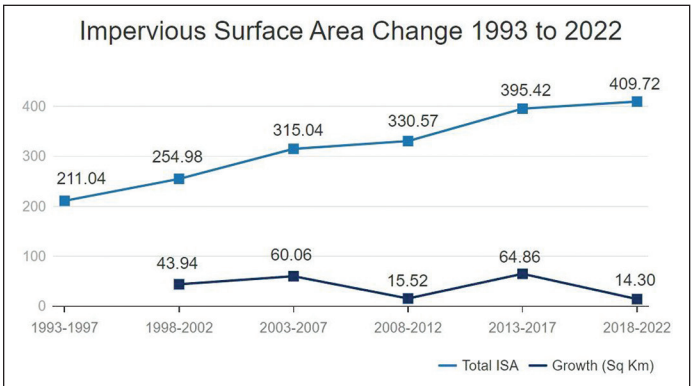


Figure 6. Overall change in the urban impervious surface from 1992 to 2022. ISA = impervious surface area.

Table 5. Binary Confusion matrix.

	Classified: No	Classified: Yes
Reference: No	True Negative (TN)	False Positive (FP)
Reference: Yes	False Negative (FN)	True Positive (TP)

Accuracy Assessment

In this study, we used a conventional approach widely used in earlier research (Story and Congalton 1986) to assess the accuracy of urban impervious surface area increase. The study collected sample points from Google Earth images for 1998, 2003, 2008, 2013, and 2018. The study used ArcGIS to calculate classified point values for ground truth points. There were no images between 1985 and 1998. Therefore, we use 1998 ground truth values for 1993 as well. We compute overall accuracy, sensitivity, omission error, commission error, Kappa coefficient, and F1 score values for accuracy assessment (Table 6). In this study, we took around 100 points for both urban and non-urban areas.

The binary confusion matrix of Table 5 is useful for studies with only two classes, in this case, ISA and other classes. Accuracy assessment metrics are overall accuracy, Cohen's Kappa coefficient (Story and Congalton 1986), and F1 score (Chicco and Jurman 2020) used in this study. The accuracy assessment equations (Equations 11–16) are formulated to deal with only two classes (Chicco *et al.* 2021). All the equations are described below (Equations 11–16).

$$OA = (TP+TN)/(TN+FP+FN+TP) \quad (11)$$

$$Sensitivity = TP/(TP+FN) \quad (12)$$

$$Omission\ error = FN/(FN+TP) \quad (13)$$

$$Comission\ error = FP/(FP+TP) \quad (14)$$

$$Kappa\ coefficient = \frac{2(TP \cdot TN - FP \cdot FN)}{(TP+FP)\cdot(FP+TN)+(TP+FN)\cdot(FN+TN)} \quad (15)$$

$$F1\ score = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (16)$$

Where OA= Overall Accuracy, TP = True Positive, TN = True Negative, FP = False Positive, and FN = False Negative. The accuracy assessment comparison Table 6 shows combined results for six years of images.

Table 6. Accuracy assessment of urban impervious surface extraction for megacities of Pakistan.

Year	Overall Accuracy	Sensitivity	Omission Error	Commission Error	Kappa	F1 Score
2018–2022	0.94444	0.95745	0.04255	0.04255	0.87745	0.95745
2013–2017	0.88462	0.86364	0.13636	0.03797	0.75019	0.91018
2008–2012	0.84270	0.86000	0.14000	0.14000	0.68051	0.86000
2003–2007	0.85000	0.97619	0.02381	0.21154	0.69543	0.87234
1998–2002	0.95050	0.98039	0.01961	0.07407	0.90092	0.95238
1993–1997	0.91050	0.93039	0.01761	0.05407	0.80092	0.87238

Discussions

In recent years, ISA has become a significant indicator of environmental quality and a characteristic of urbanization. Measuring change detection in ISA is essential for city planning, sustainable development, and environmental protection. In recent research articles, some existing approaches were used, such as mixture analysis and classification-based methods, which require multi-part computations and high-quality training data. But this research proposed a novel UISA (GISAI) index based on the GEE cloud platform. In urban remote sensing applications, the impervious area is usually mixed with bare soil because of their spectral similarities.

Additionally, the variability of ISA has an impact on the outcomes of the extraction. Therefore, the current study used a combination of indices explained in the methodology section to reduce the bare soil signature. First, using SWIR1 and SWIR2 bands, bright and highly reflecting sandy regions were eliminated. These sandy areas reflect quite well in both SWIR and other classes. Secondly, the temporal parameter eradicates barren soil regions with vegetation or water. As a result, GISAI provided significant results for comparative research of megacities. Our experimental results showed an overall accuracy of 89% for ISA extraction.

This index also shows how effectively it can map cities with different environmental backgrounds. The study outcomes allow urban planners to organize resources for future sustainable urban development. The results of UISA can be useful in observing the sustainable development of these cities. Last but not least, there were some limitations in this research conducted for UISA extraction. The primary limitation is that the GISAI estimation method is highly computationally intensive and dependent on cloud platforms. Besides increasing the differentiation of bare soil from UISA, some bare soil types are still difficult to distinguish from other land use (Ahmad *et al.* 2022b) types (built-up area, bare agriculture land). But the result can be improved by incorporating NTL data products and build-up surface products. In the future, indices combined with deep learning, image fusion, improved spatial resolution, and enhanced spectral resolution can provide a more dynamic result for UISA extraction.

Conclusions

Urban impervious surface area (UISA) is important to our anthropogenic urban environment. We implemented a novel GISAI index based on *Landsat 8* time-series data on GEE to map ISA in Lahore, Pakistan. The research findings have shown that the GISAI index can provide significant results for mapping urban impervious surfaces at the city level. The percentage increase in the urban impervious surface has been observed from 11%, 13.3%, 16.4%, 17.2%, 20.6%, and 21.4% from 1993 to 2022, respectively. But it showed the maximum increase in 2013–2017 for the last 30 years. The results reveal that GISAI retrieved UISA extraction accurately with medium spatial resolution data of 30 m. The experimental results also show. The development of a trend analysis map (Figure 6) intended to assist the government and decision-makers in the execution of spatial planning strategies and action plans for lowering the UISA in future. So that the population, infrastructure, and environment can be protected. It is also concluded that the method proposed in this study can be implemented in other metropolitan cities worldwide because it is easy to implement with open-access satellite data. Researchers and decision-makers can comprehend the environmental response to urban growth with UISA calculation.

Acknowledgments

The authors express gratitude to the USGS and NASA for making satellite imagery available to researchers. The authors acknowledge the anonymous reviewers for their constructive comments and suggestions for improving the manuscript. All persons who meet authorship criteria are listed as authors. All authors certify that they have participated sufficiently in the work to take public responsibility for the content, including participation in the manuscript's concept, design, analysis, writing, or revision. Furthermore, each author certifies that this or similar material has not been submitted to publication in any other journal. Conceptualization, M.N.A; data preparation, A.J.; methodology, M.N.A. and, A.J.; analysis, M.N.A; supervision, Z.S.; validation, M.N.A; visualization, M.N.A; writing – original draft, M.N.A.; writing – review and editing, M.N.A, A.J; All authors have read and agreed to the published current version of the manuscript. This work was supported in part by the National Natural Science Foundation of China under Grants 42090012, in part by the Guangxi science and technology program (Guangxi key R & D plan, GuiKe 2021AB30019), 03 special research and 5G project of Jiangxi Province in China (20212ABC03A09); Zhuhai industry university research cooperation project of China (ZH22017001210098PWC); Sichuan Science and Technology Program (2022YFN0031); Hubei key R & D plan (2022BAA048), and Zhizhuo Research Fund on Spatial-Temporal Artificial Intelligence (Grant No. ZZJJ202202).

References

Ahmad, M. N., Q. Cheng and F. Luo. 2022a. Dynamic linkage between urbanization, electrical power consumption, and suitability analysis using remote sensing and GIS techniques. *Photogrammetric Engineering & Remote Sensing* 88(3):171–179.

Ahmad, M. N., Z. Shao and A. Javed. 2022b. Modelling land use/land cover (LULC) change dynamics, future prospects, and its environmental impacts based on geospatial data models and remote sensing data. *Environmental Science and Pollution Research*:1–17.

Ahmad, M. N., Z. Shao, A. Javed, F. Islam, H. H. Ahmad and R. W. Aslam. 2023. The cellular automata approach in dynamic modelling of land use change detection and future simulations based on remote sensing data in Lahore Pakistan. *Photogrammetric Engineering & Remote Sensing* 89(1):47–55.

Ahmad, M. N., Z. Shao, R. W. Aslam, I. Ahmad, M. Liao, X. Li and Y. Song. 2022. Landslide hazard, susceptibility and risk assessment (HSRA) based on remote sensing and GIS data models: A case study of Muzaffarabad Pakistan. *Stochastic Environmental Research and Risk Assessment*:1–16.

Cao, X., X. Gao, Z. Shen and R. Li. 2020. Expansion of urban impervious surfaces in Xining city based on GEE and Landsat time series data. *IEEE Access* 8:147097–147111.

Chen, X., K. Yang, J. Wang, Z. Wang, L. Wang and F. Su. 2022. Improving long-term impervious surface percentage mapping in mountainous areas based on multi-source remote sensing data. *Geocarto International* 37(26):1–25.

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC genomics, 21, 1-13.

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The Matthews correlation coefficient (MCC) is more informative than Cohen's Kappa and Brier score in binary classification assessment. IEEE Access, 9, 78368-78381.

Gao, F., E. B. De Colstoun, R. Ma, Q. Weng, J. G. Masek, J. Chen, Y. Pan and C. Song. 2012. Mapping impervious surface expansion using medium-resolution satellite image time series: a case study in the Yangtze River Delta, China. *International Journal of Remote Sensing* 33(24):7609–7628.

Gong, P., X. Li and W. Zhang. 2019. 40-Year (1978–2017) human settlement changes in China reflected by impervious surfaces from satellite remote sensing. *Science Bulletin* 64(11):756–763.

Huang, X., J. Li, J. Yang, Z. Zhang, D. Li and X. Liu. 2021. 30 m global impervious surface area dynamics and urban expansion pattern observed by Landsat satellites: From 1972 to 2019. *Science China Earth Sciences* 64(11):1922–1933.

Javed, A., Z. Shao, I. Ara, M. E. Huq, M. Y. Ali, N. Saleem and M. N. Ahmad. 2022. Development of global impervious surface area index for automatic spatiotemporal urban mapping. *Preprints.org*. https://doi.org/10.20944/preprints202207.0071.v1.

Kotarba, A. Z. and S. Aleksandrowicz. 2016. Impervious surface detection with nighttime photography from the International Space Station. *Remote Sensing of Environment* 176:295–307.

Kuang, W., W. Chi, D. Lu and Y. Dou. 2014. A comparative analysis of megacity expansions in China and the US: Patterns, rates and driving forces. *Landscape and Urban Planning* 132:121–135.

Kuang, W., Y. Hou, Y. Dou, D. Lu and S. Yang. 2021. Mapping global urban impervious surface and green space fractions using Google Earth Engine. *Remote Sensing* 13(20):4187.

Li, H., F. Zhu, X. Zheng, M. Liu and G. Chen. 2022. MSCDUNet: A deep learning framework for built-up area change detection integrating multispectral, SAR and VHR data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 15:5163–5176.

Li, L., D. Lu and W. Kuang. 2016. Examining urban impervious surface distribution and its dynamic change in Hangzhou metropolis. *Remote Sensing* 8(3):265.

Li, X. and Y. Zhou. 2017. A stepwise calibration of global DMSP/OLS stable nighttime light data (1992–2013). *Remote Sensing* 9(6):637.

Li, X., Y. Zhou, M. Zhao and X. Zhao. 2020. A harmonized global nighttime light dataset 1992–2018. *Scientific Data* 7(1):1–9.

Liang, X., Y. Lin and H. Zhang. 2022. Mapping urban impervious surface with an unsupervised approach using interferometric coherence of SAR images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 15:2734–2744.

Liu, F., S. Wang, Y. Xu, Q. Ying, F. Yang and Y. Qin. 2020. Accuracy assessment of global human settlement layer (GHSL) built-up products over China. *Plos One* 15(5):e0233164.

Lu, D., E. Moran and S. Hetrick. 2011. Detection of impervious surface change with multitemporal Landsat images in an urban–rural frontier. *ISPRS Journal of Photogrammetry and Remote Sensing* 66(3):298–306.

Manoli, G., S. Fatichi, M. Schläpfer, K. Yu, T. W. Crowther, N. Meili, P. Burlando, G. G. Katul and E. Bou-Zeid. 2019. Magnitude of urban heat islands largely explained by climate and population. *Nature* 573(7772):55–60.

Mignot, E., X. Li and B. Dewals. 2019. Experimental modelling of urban flooding: A review. *Journal of Hydrology* 568:334–342.

Ouyang, L., C. Wu, J. Li, Y. Liu, M. Wang, J. Han, C. Song, Q. Yu and D. Haase. 2022. Mapping impervious surface using phenology-integrated and fisher transformed linear spectral mixture analysis. *Remote Sensing* 14(7):1673.

Peroni, F., S. E. Pappalardo, F. Facchinelli, E. Crescini, M. Munafò, M. E. Hodgson and M. De Marchi. 2022. How to map soil sealing, land take and impervious surfaces? A systematic review. *Environmental Research Letters* 17(5):53005.

Sattar, A., A. Goswami, A. V. Kulkarni and A. Emmer. 2020. Lake evolution, hydrodynamic outburst flood modeling and sensitivity analysis in the Central Himalaya: A case study. *Water* 12(1):237.

Schiavina, M., M. Melchiorri, M. Pesaresi, P. Politis, S. Freire, L. Maffenini, P. Florio, D. Ehrlich, K. Goch and P. Tommasi. 2022. *GHSL Data Package 2022*. Luxembourg: Publications Office of the European Union. https://doi.org/10.2760/19817.

Story, M. and R. G. Congalton. 1986. Accuracy assessment: A user's perspective. *Photogrammetric Engineering and Remote Sensing* 52(3):397–399.

Su, S., J. Tian, X. Dong, Q. Tian, N. Wang and Y. Xi. 2022. An impervious surface spectral index on multispectral imagery using visible and near-infrared bands. *Remote Sensing* 14(14):3391.

Sun, G., X. Chen, X. Jia, Y. Yao and Z. Wang. 2015. Combinational build-up index (CBI) for effective impervious surface mapping in urban areas. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9(5):2081–2092.

Sun, G., J. Cheng, A. Zhang, X. Jia, Y. Yao and Z. Jiao. 2022a. Hierarchical fusion of optical and dual-polarized SAR on impervious surface mapping at city scale. *ISPRS Journal of Photogrammetry and Remote Sensing* 184:264–278.

Sun, Z., W. Du, H. Jiang, Q. Weng, H. Guo, Y. Han, Q. Xing and Y. Ma. 2022b. Global 10-m impervious surface area mapping: A big earth data based extraction and updating approach. *International Journal of Applied Earth Observation and Geoinformation* 109:102800. https://doi.org/10.1016/j.jag.2022.102800.

Tang, Y., Z. Shao, X. Huang and B. Cai. 2021. Mapping impervious surface areas using time-series nighttime light and MODIS imagery. *Remote Sensing* 13(10):1900.

Tian, Y., H. Chen, Q. Song and K. Zheng. 2018. A novel index for impervious surface area mapping: Development and validation. *Remote Sensing* 10(10):1521.

Uhl, J. H. and S. Leyk. 2022. Uncertainty prediction of built-up areas from global human settlement data in the United States based on landscape metrics. *ArXiv Preprint ArXiv:*2205.09023.

Vaddiraju, S. C. and C. Savitha. 2022. Determination of impervious area of Saroor Nagar Watershed of Telangana using spectral indices, MLC, and machine learning (SVM) techniques. *Environmental Monitoring and Assessment* 194(4):1–13.

Wang, J., Y. Zhao, Y. Fu, L. Xia and J. Chen. 2022a. Improving LSMA for impervious surface estimation in an urban area. *European Journal of Remote Sensing* 55(1):37–51.

Wang, M., W. Ding, F. Wang, Y. Song, X. Chen and Z. Liu. 2022b. A novel bayes approach to impervious surface extraction from high-resolution remote sensing images. *Sensors* 22(10):3924.

Wen, D., X. Huang, F. Bovolo, J. Li, X. Ke, A. Zhang and J. A. Benediktsson. 2021. Change detection from very-high-spatial-resolution optical remote sensing images: Methods, applications, and future directions. *IEEE Geoscience and Remote Sensing Magazine* 9(4):68–101.

Weng, Q. 2012. Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends. *Remote Sensing of Environment* 117:34–49.

Weng, Q. and X. Hu. 2008. Medium spatial resolution satellite imagery for estimating and mapping urban impervious surfaces using LSMA and ANN. *IEEE Transactions on Geoscience and Remote Sensing* 46(8):2397–2406.

Wu, W., Z. Shao, J. Teng, X. Huang, X. Zhao and S. Guo. 2022. Urban impervious surface extraction using seasonal time series SAR images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 15:6033–6045. https://doi.org/10.1109/JSTARS.2022.3193137.

Xu, H. 2006. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing* 27(14):3025–3033.

Yin, R., G. He, G. Wang, T. Long, H. Li, D. Zhou and C. Gong. 2021. Automatic framework of mapping impervious surface growth with long-term landsat imagery based on temporal deep learning model. *IEEE Geoscience and Remote Sensing Letters* 19:1–5.

Yue, H. and Y. Liu. 2019. Method for delineating open water bodies based on the deeply clear waterbody delineation index. *Journal of Applied Remote Sensing* 13(3):38504.

Zhang, X., L. Liu, C. Wu, X. Chen, Y. Gao, S. Xie and B. Zhang. 2020. Development of a global 30 m impervious surface map using multisource and multitemporal remote sensing datasets with the Google Earth Engine platform. *Earth System Science Data* 12(3):1625–1648. https://doi.org/10.5194/essd-12-1625-2020.

Zhao, Y. and Z. Zhu. 2022. ASI: An artificial surface Index for Landsat 8 imagery. *International Journal of Applied Earth Observation and Geoinformation* 107:102703.

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

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

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

Cardinal Systems, LLC
Flagler Beach, Florida
www.cardinalsystems.net
Member Since: 1/2001

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

GPI Geospatial Inc.
Orlando, Florida
www.aca-net.com
Member Since: 1/1994

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

Miller Creek Associates
SeaTac Washington
www.mcamaps.com
Member Since: 12/2014

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

PixElement
Belmont, Michigan
https://pixelement.com
Member Since: 2/2017

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

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

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
Lorraine B. Amenda, PLS, CP
Towill, Inc

President-Elect
Bandana Kar
Oak Ridge National Lab

Vice President
Amr Abd-Elrahman
University of Florida

Past President
Christopher Parrish, Ph.D
Oregon State University

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: Hope Morgan
Deputy Chair: Tao Liu

Early-Career Professionals Council
Chair: Youssef Kaddoura
Deputy Chair: : Greg Stamnes

Region Officers Council
Chair: Demetrio Zourarakis
Deputy Chair: Cody Condron

Student Advisory Council
Chair: Oscar Duran
Deputy Chair: Ali Alruzuq

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: Srini Dharmapuri
Assistant Director: Ravi Soneja

Professional Practice Division
Director: Hope Morgan
Assistant Director: Christian Stallings

Remote Sensing Applications Division
Director: Tao Liu
Assistant Director: Indu Jeyachandran

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
Dave Parret

Cascadia Region
Jimmy Schulz

Eastern Great Lakes Region
Craig Fry

Florida Region
Matt LaLuzerne

Gulf South
Stuart Babin

Heartland Region
Whit Lynn

Mid-South Region
David Hughes

North Atlantic Region
Kurt Lutz

Northeast Region

Pacific Southwest Region
Omar Mora

Potomac Region
Jason Brown

Rocky Mountain Region
Melissa Martin

Western Great Lakes Region
Adam Smith

Small Object Detection in Remote Sensing Images Based on Window Self-Attention Mechanism

Jiaxin Xu, Qiao Zhang, Yu Liu, and Mengting Zheng

Abstract

For remote sensing image object detection tasks in the small object feature, extraction ability is insufficient and difficult to locate, and other problems. This paper proposes an improved algorithm for small object detection in remote sensing images based on a window self-attention mechanism. On the basis of You Only Look Once (YOLO)v5s, a shallow feature extraction layer with four times downsampling is added to the feature fusion pyramid and the window self-attention mechanism is added to the Path Aggregation Network. Experiments show that the improved model obtained the Mean Average Precision (mAP) of 78.3% and 91.8% on the DIOR and Remote Sensing Object Detection public data sets with frames per second of 65 and 51, respectively. Compared with the basal YOLOv5s network, the mAP has improved by 5.8% and 3.3%, respectively. Compared with other object detection methods, the detection accuracy and real-time performance have been improved.

Introduction

With the rapid development of satellite remote sensing technology at home and abroad, the observed earth surface information that can be obtained is becoming more and more abundant. There are numerous important applications for the quick and precise automatic detection of specific targets in remote sensing images, including defense and military response, the development of smart cities, the mitigation and control of geological disasters, the monitoring of the agricultural environment, the updating of geographic information systems, and other fields (Li *et al.* 2010; Castelluccio *et al.* 2015; Chen *et al.* 2017; Kanjir *et al.* 2018; Song *et al.* 2021; Tian *et al.* 2022). Currently, remote sensing images have ever-finer resolutions. The structure, texture, and details of ground objects and landscapes can be seen more clearly in high-resolution remote sensing images, which makes it easier to present the target in the field of view (Li *et al.* 2016). Object detection has met great challenges, since remote sensing images with high resolution and high coverage also offer redundant information, and observed objects are varied, small, and dense (Xie *et al.* 2021). So, fast and accurate automatic detection of small targets is one of the most challenging and valuable tasks in the field of remote sensing.

For some early object detection algorithms, the Haar cascade classifier (Viola and Jones 2001) is based on Haar-like features and used a strong classifier cascade trained by Adaboost (Freund and Schapire 1997) for face detection. The Scale Invariant Feature Transform method extracts invariant features from pictures to achieve reliable matching of objects or scenes under different viewing angles (Lowe 2004). The Histogram of Oriented Gradient (HOG) algorithm calculates statistics on the gradient information (Dalal and Triggs 2005), and then generates the final feature description, and combines support vector machine (SVM) for object classification and detection (Cortes and Vapnik 1995). The Deformable Parts Model detector makes some improvements on the basis of HOG and SVM classifiers (Felzenszwalb *et al.* 2008), which greatly improves the detection performance. The above-mentioned traditional machine learning object detector adopts artificially designed features and sliding window ideas, which have problems such as single features, high time cost, and low detection efficiency.

With the introduction of the convolutional neural network (CNN), a breakthrough has been made in the object detection method based on deep learning. For object detection, there are anchor-based detectors and anchor-free detectors. Depending on whether they must create region proposals, the anchor-based detectors are separated into two-stage detectors and single-stage detectors. The first two-stage algorithm, region-based (R)-CNN (Girshick *et al.* 2014), which is a qualitative leap over conventional techniques, uses CNN as the fundamental network for object detection and a selective search method to extract candidate boxes. The Scalable Sequential Pyramid Networks (SPP-Net) algorithm implements region feature extraction at different scales (He *et al.* 2015). Fast R-CNN uses region of interest pooling layers to extract features from candidate regions of different sizes and shapes (Girshick 2015). By generating candidate regions with Region Proposal Network (RPN) (Ren *et al.* 2015), the detection efficiency of the Faster R-CNN algorithm is significantly the Fast R-CNN. In 2017, the mask branch was added to the Mask R-CNN algorithm to accomplish both object detection and semantic segmentation (He *et al.* 2017). However, the two-stage object detection methods were difficult to achieve real-time detection. In response to this problem, Redmon and others transformed the classification problem into a regression problem (Redmon *et al.* 2015), named the proposed algorithm You Only Look Once (YOLO). On this basis, the YOLOv2 (Redmon and Farhadi 2017) and YOLOv3 (Redmon and Farhadi 2018) algorithms are successively proposed. Among them, YOLOv3 uses Darknet-53 as the backbone network and uses the Feature Pyramid Network (FPN) and multi-scale feature map. The Single Shot Multibox Detector combines the anchor idea in RPN to design dense anchors with multiple aspect ratios at multiple scales (Liu *et al.* 2015). In 2020, Alexey modified the backbone of the YOLOv3 network to Cross Stage Partial Dark Network (CSPDarknet53) and proposed YOLOv4 (Bochkovskiy *et al.* 2020), and adopts the idea of spatial pyramid pooling. In the same year, Glenn Jocher and others proposed YOLOv5 (https://github.com/ultralytics/YOLOv5), which is more lightweight and stable than the YOLOv1~v4 series; and can be easily deployed in engineering applications. There are also some anchor-free detection techniques, such as FCOS (Tian *et al.* 2019), CornerNet (Law and Deng 2020), etc. Although these methods have quick detection times, their accuracy is still a concern.

There are two definitions of small targets in object detection, one based on relative size and one based on absolute size. According to the definition of the international organization, Society of Photographic Instrumentation Engineers, a target area of fewer than 80 pixels in a

Southwest Petroleum University, School of Geoscience and Technology, No. 8, Xindu Avenue, Xindu District, Chengdu, Sichuan, China.

Corresponding author: Qiao Zhang (swpuqzh@swpu.edu.cn)

Contributed by Bo Wu, January 15, 2023 (sent for review April 7, 2023; reviewed by Zeyu Chen, Riccardo La Grassa).

Photogrammetric Engineering & Remote Sensing
Vol. 89, No. 8, August 2023, pp. 489–497.
0099-1112/22/489–497

doi: 10.14358/PERS.23-00004R3

256 × 256 image is a small target (Zhang *et al.* 2021). According to the definition of the Common Objects in Context (COCO) data set (Lin *et al.* 2014), a target with a size of fewer than 32 × 32 pixels can be considered a small target. This paper uses the absolute definitions from the COCO data set. Because of their low SNR, tiny size, and limited structural information, small objects are challenging to detect. Bashir *et al.* (2021) combined recurrent generative adversarial networks and residual feature aggregation to improve the detection performance of small objects. Li *et al.* (2021) proposed a cross-layer attention network that further enhances its representation capability. Han *et al.* (2022) proposed a multi-visual small object detector that detects small objects by using multiple low-level feature maps with multi-branch convolution. Zhang *et al.* (2022) proposed the CotYOLO-v3 target detection algorithm to extract contextual information about small objects and enhance visual representation in Darknet-53. Fang *et al.* (2022) proposed an improved S2ANET-SR model to design a super-resolution enhancement module to enhance the feature extraction of small objects. Shi *et al.* (2022) proposed an algorithm called FE-CenterNet anchorless detector to improve the perception of small objects by mining multi-scale contextual information and combining it with a coordinate attention mechanism. However, the above methods still have the problem of low accuracy or low real-time detection of small objects. This paper proposes the following modifications based on YOLOv5. (1.) To improve the detection ability of small targets in remote sensing images, a shallow feature extraction layer with four times downsampling is added to the feature fusion pyramid and a small target detection layer should be added to the detection head. (2.) To improve the network feature extraction capability for small targets, the convolution layer of the BottleneckCSP in the Path Aggregation Network (PANet) is replaced by two continuous Swin-Transformer blocks.

Introduction of YOLOV5

YOLOv5 differs from previous generations of YOLO in that it is divided into four models, s, m, l, and x, according to the volume size of the model. In various models, the width and depth of the network can be freely adjusted to facilitate the deployment of the model in hardware devices that meet the conditions. In this paper, we adopt the smallest volume YOLOv5s model for improvement. YOLOv5s mainly consists of three parts: backbone, neck, and head, and its network structure is shown in Figure 1.

Backbone

The backbone of YOLOv5s consists of an input and a backbone network for extracting the features of the image. The input is a color image with a size of 640 × 640 and three channels of RGB. Data pre-prossessing is then performed to enrich the image information using Mosaic data enhancement. The k-means clustering algorithm is used to re-match the anchors and adaptively obtain the best anchors. The backbone network is CSPDarknet53, which contains modules such as focus, CSP, and Shirley Proctor Puller Foundation (SPPF). In YOLOv5-v6.0 (this is the sixth official released version of YOLOv5), the focus module is equivalently replaced by a convolutional layer with a convolutional kernel size of 6 × 6, which is more efficient when applied to hardware devices. The CSP block consists of a series of residual blocks that effectively prevent gradient explosion and gradient disappearance. The SPPF block is different from the SPP block in YOLOv4. The SPP block inputs the extracted features into multiple Maxpooling layers of different sizes in parallel and then further fuses them, while the SPPF block makes the input features pass through multiple maxpooling layers of 5 × 5 size serially in turn. Although the final processing results of the two blocks are the same, the latter is twice as fast as the former in terms of computation

Neck

The neck part of YOLOv5s consists of FPN and PANet for fusion of features extracted from the backbone network. The FPN transfers rich semantic information from the deep feature layer to the shallow feature layer, enhances the feature representation capability of small targets, and solves the multi-scale problem well to generate feature maps of different scales for prediction. The PANet adds a bottom-up channel based on FPN and combines the CSP block to better transfer the shallow information to the deep feature layer, fuse features in both directions, and strengthen the fusion ability of features.

Head

The head part of YOLOv5s includes the prediction and output of the network, which uses the detection head to perform object detection on the input image, generate the object bounding box, and predict the category of that object. For the loss function, YOLOv5s calculates three kinds of losses: confidence, bounding box, and classification. In the loss calculation of bounding box, the Generalized Intersection over Union (GIoU) loss function is used (Rezatofighi *et al.* 2019). The GIoU loss calculation formula is as follows:

$$\text{loss}_{\text{GIoU}} = 1 - \left(\text{IoU} - \frac{|C - A \cap B|}{|C|} \right) \tag{1}$$

where A denotes the area of the true box, B denotes the area of the predicted box, and C denotes the area of the smallest rectangular box that can surround A and B. When the ratio of the area except the true box and the predicted box in C to the total area of C is larger, it means that the distance between the true box and the predicted box is farther, and the degree of overlap is smaller. Conversely, the closer the distance, the greater the overlap. The calculation of confidence loss and classification loss is composed of a sigmoid layer and binary cross entropy loss function "Binary Cross Entropy (BCE) loss". Because the input value range of BCE loss is [0, 1], it is necessary to use an activation

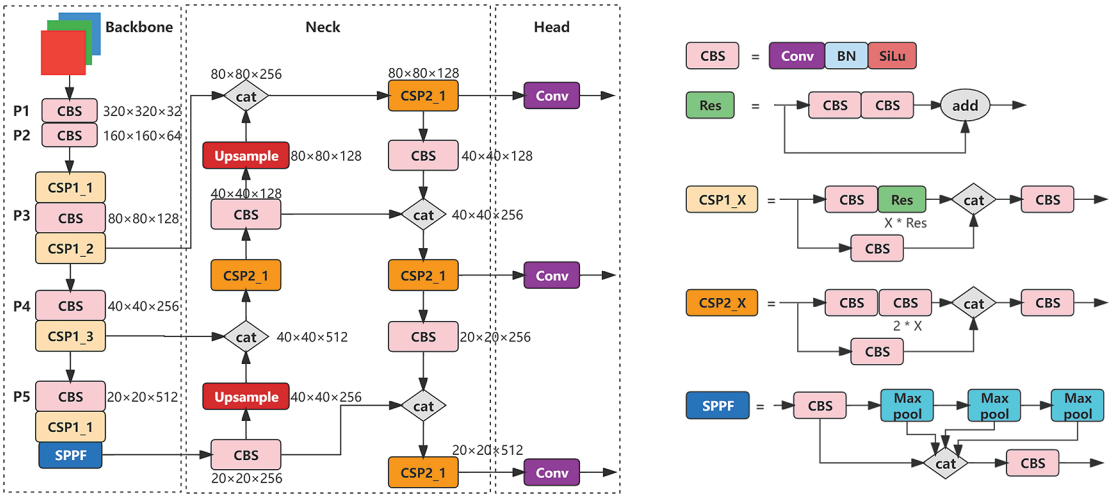


Figure 1. You Only Look Once (YOLO)v5s network structure diagram.

function to ensure that the input is correct. The formula of BCE loss is as follows:

$$\text{loss}_{\text{BCE}} = -\frac{1}{n}\sum_{x}\left[ylna + (1-y)\ln(1-a)\right] \quad (2)$$

where x is the sample, y is the label, a is the predicted output, and n is the total number of samples. The total loss is the sum of confidence loss, bounding box loss, and classification loss. The total loss formula is as follows:

$$\text{loss}_{\text{total}} = \lambda_1\text{loss}_{\text{confidence}} + \lambda_2\text{loss}_{\text{box}} + \lambda_3\text{loss}_{\text{classification}} \quad (3)$$

The coefficient λ in front of each loss represents the respective weight. In this paper, we set the weight coefficient of confidence loss to 1.0 and only count positive samples. The weight coefficient of the classification loss is 0.5, and only positive samples are calculated. The bounding box loss has a weight factor of 0.05.

Improvements of YOLOV5

In this paper, some improvements based on YOLOv5s are proposed. A small target detection layer is added to the original detection layer, and four feature layers of different scales are used for target detection. Furthermore, the window self-attention mechanism in the Swin Transformer block is added to the PANet to enhance the detection capability of small targets. The improved network structure diagram is shown in Figure 2.

Add a Small Object Detection Layer

Three scales of feature maps for detection are created in the basic YOLOv5s network by fusing the feature layers of 8x, 16x, and 32x downsampling. According to the idea of the feature fusion pyramid, the 8x downsampling feature layer belongs to shallow features with smaller perceptual field and more accurate target location information, which is suitable for detecting small targets in images. The 32x downsampling feature layer belongs to deep features containing richer semantic information, which will be interfered with by redundant background information in the process of detecting small targets and is more suitable for detecting large targets. Our research adds a 4x downsampling feature layer to the feature extractor, in which the target has larger pixels and richer position information than the 8x downsampling feature layer. So the network may extract more fine-grained features to improve the performance of tiny targets. Figure 3 displays the schematic diagram of the improved feature extractor.

Add Window Self-Attention Mechanism

Conv layers have been the mainstay existence in convolutional neural networks, which is good at capturing local information, but the locality limits its perceptual field. Transformers can obtain global features to make up for the global capture ability and can learn high-quality intermediate features (Dosovitskiy *et al.* 2020). However, the transformer is computationally intensive in the network, and the shifted windows self-attention mechanism in the swin transformer can improve the computational efficiency (Liu *et al.* 2021). In this paper, two consecutive Swin Transformer blocks (hereafter referred to as STR) are added to the PANet structure to replace one of the convolutional layer structures in BottleneckCSP, and name this structure C3STR. The structure diagram of the C3STR module is shown in Figure 4. The window multi-head attention mechanism in the STR module is used to improve the representational capability of extracted feature, and improve the detection accuracy of small targets.

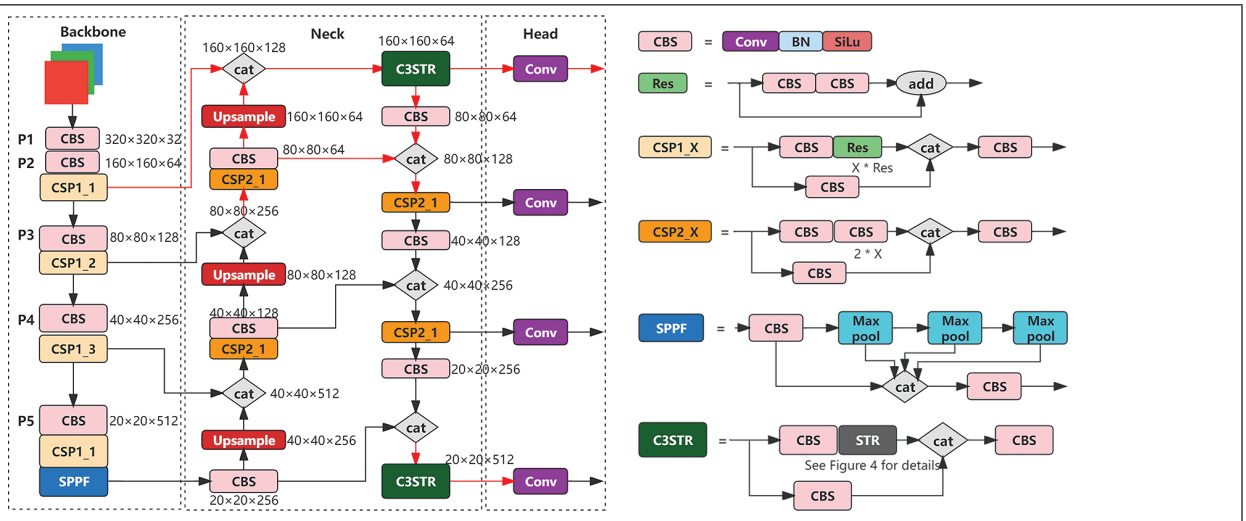


Figure 2. Network structure diagram of improved You Only Look Once (YOLO)v5, where the red line is the improvement part.

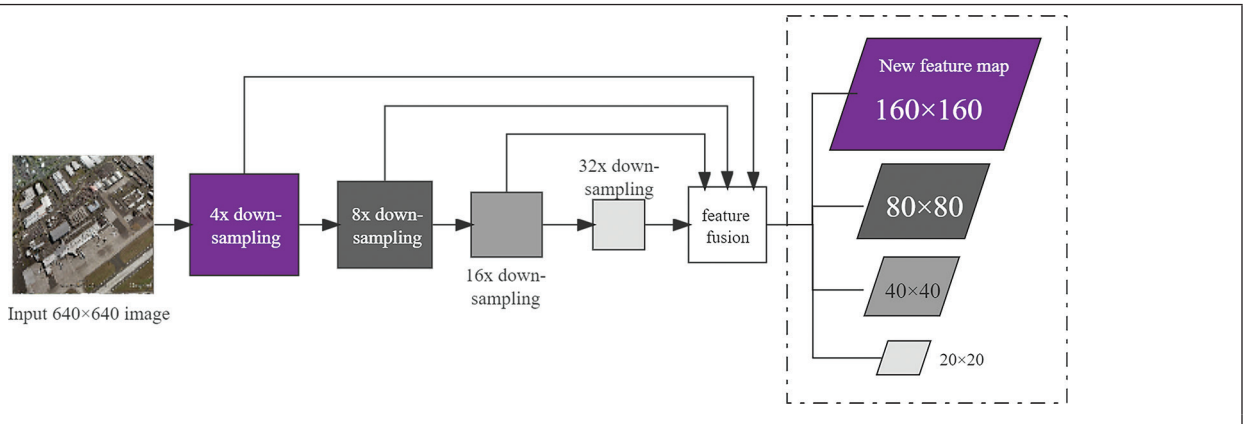


Figure 3. Schematic of the feature extractor after adding a 4x downsampled shallow feature layer.

STR module consists of two consecutive Swin Transformer blocks. Firstly, the input feature map is spreading the dimensions of width and height directions by convolution, taking the deep feature map 20 × 20 × 256 as an example, after convolution it becomes a two-dimensional vector [400,256], then it enters Layer Normal processing layer to do mean and variance calculation with the following formula:

$$y = \frac{x - E[x]}{\sqrt{\mathrm{Var}[x] + \epsilon}} \times \gamma + \beta \tag{4}$$

where x denotes the value of each dimension of the input, y denotes the processed value, ϵ is a very small value, and the default $\epsilon = 10^{-5}$, γ, and β are trainable parameters.

After processing, the features are extracted by a window multi-headed attention mechanism. The features are then fed into the Dropout layer to do stochastic gradient descent to prevent model overfitting. The obtained results are then added with the previous input of the first Layer Normal layer, which is similar to the residual structure. After addition, the features are continued to be input to the next Layer Normal block for processing, and then processed by the Multi-Layer Perceptrons (MLP) block, and then add to the result of the previous addition. The MLP block consists of a fully connected layer, a Gaussian Error Linear Unit activation function, and a Dropout layer. In the first fully connected layer, the number of input channels is multiplied by four times, that is, [400,256] to [400,1024], and in the second fully connected layer, the number of channels is restored. Similarly, the next layer of the Swin Transformer block is similar to the previous layer, but the window multi-head attention mechanism uses the shifted window processing method.

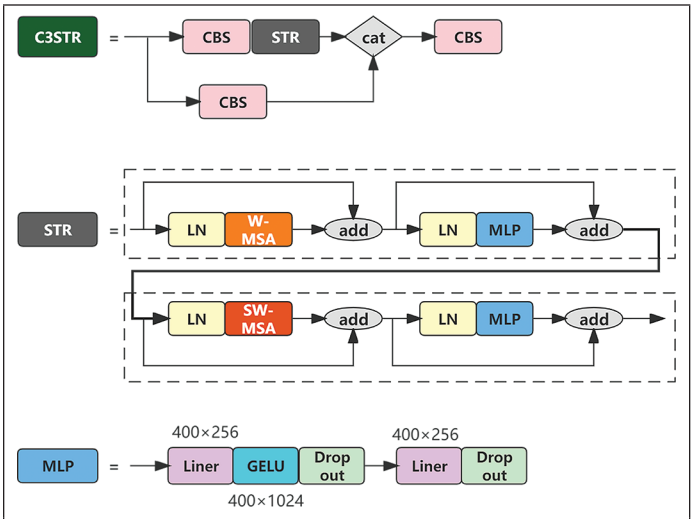


Figure 4. C3STR module structure diagram.

In the general Multi-Headed Self-Attention mechanism (MSA), for each pixel in the feature map, it is necessary to do a computation with all other pixels on the whole map, which is very computationally intensive. However, in the Windows-based Multi-Head Self Attention (W-MSA), the feature map is first divided into small windows according to the size of $M \times M$, and then the self-attention computation is performed inside each window separately. Where M is the size of the window, in this paper's algorithm for example, M is set to 8. W-MSA greatly reduces the computational effort compared with MSA, but due to the limitation of windows, the perceptual field becomes smaller and there is no information interaction between windows. To solve this problem, the Shifted Windows Multi-Head Self-Attention (SW-MSA) is added in the second layer of the STR module, which shifts the window two pixels to the right and two pixels down. As shown in Figure 5, the calculation of self-attention is performed in the shifted window to fuse the information of the neighboring windows in the previous layer to achieve information interaction between different windows.

In W-MSA and SW-MSA, the information from different head parts is learned jointly using a multi-headed attention mechanism. The input node x_i is mapped into vector a_i by means of mapping, and the corresponding k^i, q^i, v^i are obtained by means of three trainable transformation matrices W_k, W_q, W_v. Depending on the number of heads used, k^i, q^i, v^i are equally divided into head parts, and then get Q_i, K_i, V_i corresponding to each head$_i$, and calculate their self-attention (Vaswani *et al.* 2017), the formula is as follows:

$$\mathrm{Attention}(Q,K,V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{5}$$

where Q denotes the matrix of q, q is query, which will be subsequently matched with each k; K denotes the matrix of k, k is key, which will be subsequently matched by each q; V denotes the matrix of v, v is the information extracted from the weight matrix q and k matching process, the greater its relevance corresponds to the greater the weight of v; d represents the length of vector k^i. With the following equation, the self-attention calculation results from each head are stitched together and fused by a trainable parameter W^O to obtain the final multi-head attention results:

$$\mathrm{MultiHead}(Q,K,V) = \mathrm{Concat}\,(\mathrm{head}_1, \mathrm{head}_2, \ldots, \mathrm{head}_h) \tag{6}$$
$$\text{where } \mathrm{head}_i = \mathrm{Attention}\,(QW_i^Q, KW_i^K, VW_i^V)$$

Figure 5. Change of Windows-based Multi-Head Self Attention (W-MSA) to Shifted Windows Multi-Head Self-Attention (SW-MSA).

Experiments and Analysis

Data Preparation and Pre-Processing

We use a large-scale benchmark public data set DIOR (Li *et al.* 2019) for our experiments. The data set consists of 23 463 images and 190 288 target instances with image size of 800 × 800 and resolution ranging from 0.5 m to 30 m. DIOR has twenty categories of targets, namely airplane, airport, baseball-field, basketball-court, bridge chimney, dam, expressway-service-area, expressway-toll-station, golf-field, ground-track-field, harbor, overpass, ship, stadium, storage-tank tennis-court, train-station, vehicle, windmill. The data set is annotated in PASCAL VOC format and converted to YOLO data set format. The training, validation, and test sets are divided in a 1:1:2 ratio. Before training, Mosaic data enhancement was done on the images. Mosaic data enhancement was done by randomly selecting four images and performing operations such as flipping, scaling, and color gamut transformation to obtain the images shown in Figure 6. The anchors are then re-clustered using the k-means clustering algorithm, yielding 12 appropriately sized anchors adaptively.

Model Training

The experiments in this paper make use of the Windows 10 operating system, an NVIDIA RTX A4000 GPU, and the Pytorch open-source framework with Python version 3.8. In this paper, we do not use the official YOLOv5 pre-training weights file provided, and initialize the weights randomly. The batch size for model training is set to 4, the epoch is 300, and the initial learning rate is 0.01. During training, we use an early stop mechanism to track the validation loss and stop training when the validation loss does not decrease for ten consecutive epochs. Figure 7 displays the loss function change curves of the training set and the validation set.

Evaluation Metrics and Analysis of Results

The evaluation metrics in this paper use Average Precision (AP50, AP50:95), Mean Average Precision (mAP@0.5), Recall, the AP for small objects (APS, area < 32^2), the AP for medium objects (APM, 32^2 < area < 96^2), the AP for large objects (APL, area > 96^2), and Frames Per Second (FPS). AP50 is the average accuracy value at an Intersection over Union (IoU) threshold of 0.5. AP50:95 is the average accuracy

Figure 6. Example of images after Mosaic data enhancement process.

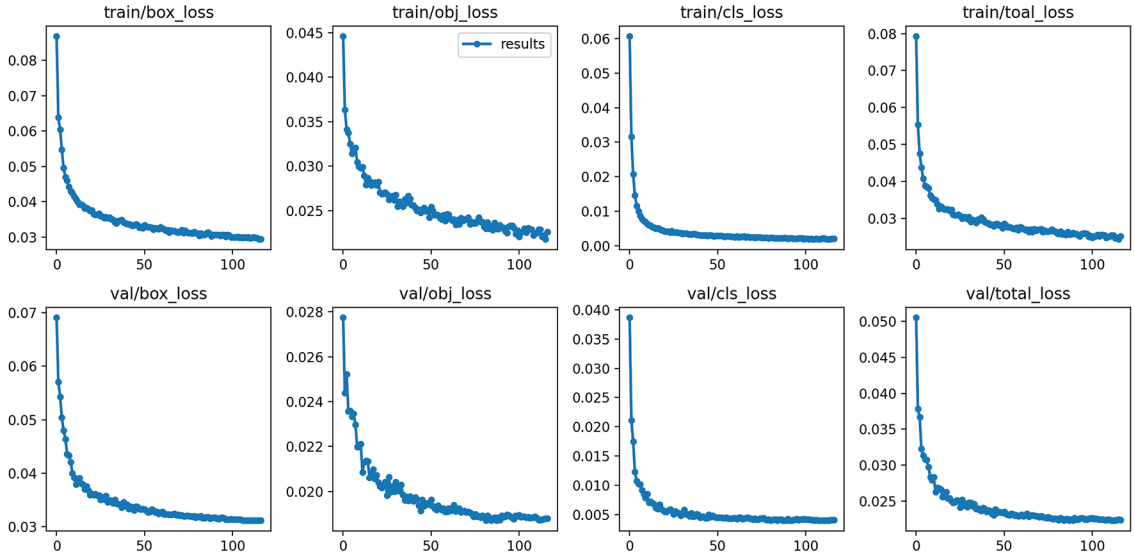


Figure 7. Loss function change curve of training set and validation set, where "train" represents the training set, "val" represents the validation set, "box_loss" represents the localization loss (GIoU), "obj_loss"' represents the confidence loss, and "cls_loss" represents the classification loss. It can be observed that the overall convergence of the model is fast.

value at an IoU threshold of 0.5 to 0.95. The mAP@0.5 is the average value of all categories of AP at an IoU threshold of 0.5. Recall measures the ability of the classifier to find all the positive classes with the following formula. Where TP is true positive and FN is false negative. FPS represents the number of frames transmitted per second, that is, the number of pictures that the network can detect per second.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{7}$$

In this study, ablation experiments and contrast experiments are designed with DIOR as the primary data set to confirm the efficacy and superiority of the improved model. (1) Ablation experiments: Adding only a small target detection layer is referred to as Method A. Adding only three C3STR modules to the detection head is referred to as Method B. Adding a small target detection layer, and replacing all detection heads to C3STR is called Method C. Adding a target detection layer, and replacing all the CSP blocks in the backbone network with C3STR is called Method D. Adding a small target detection layer, and adding C3STR modules to the detection head of P2 and P5 is called Method E. These different improvement schemes are added based on the YOLOv5s network to verify the improvement effect of each improvement scheme. The results in Table 1 demonstrate that adding only Method A increases the AP_S values, indicating that more small target features can be extracted from the shallow feature map. Adding only Method B, the AP_S, AP_M, AP_L, and Mean Average Precision (mAP) both are promoted, indicating that adding the self-attention mechanism in PANet not only helps to extract small target features in shallow features but also helps to extract large target features in deep features. By adding Method E, the mAP value is further improved. However, compared to the basal YOLOv5s method, the speed is slower since a feature fusion layer has been added, and the straightforward convolutional layer has been replaced with a more complicated STR structure, resulting in increased inference time of the network. Despite this, the FPS of 64 can still match the real-time criteria. Comparing Method C with Method E, the mAP and recall are reduced by 0.001 and 0.01, respectively. It is observed that although the mAP is reduced, the AP of small objects such as airplanes, ships, tennis courts, and storage tanks has been improved. We think the possible reason is that other objects in the data set are already in an overfitted state before replacing C3STR due to the small amount of data. The transformer is too concerned with the global and has a large number of parameters, and the replacement increases the complexity of the model, resulting in the performance can be unchanged or decreased. For some small objects with improved AP, they may be under-fitted before adding parameters. Perhaps the overall performance improvement can be obtained by increasing the scale of the data set. Method D improves AP_S, AP_M, and AP_L compared to the base YOLOv5s, but recall decreases and is not as good as the Method E improvement. We think the possible reason is that the self-attention mechanism in Transformer can capture long-distance feature, but ignores local feature details, which reduces the discriminability between background and foreground. In Figure 8, we plot the confusion matrix generated by Method E. In the figure, each row is the predicted category and each column is the true category. The diagonal data are the percentage of true positives. The *TP* values above 0.9 are for airplanes, baseball fields, and ships. Bridges have the highest *FN*, which is due to the small training sample and limited feature extraction for bridges. The *FP* for ships is also higher, which may be attributed to the smaller and denser targets of ships, which are easily misdetected. In the ablation experiments, we did not replace all CSP blocks with C3STR blocks, which we do not recommend. From the convolutional point of view, this simple convolutional kernel allows for efficient template matching in low-level semantic processing. In the self-attention mechanism, they can theoretically represent any convolutional layer when a sufficient number of heads are used, but are computationally expensive and not necessarily more accurate than convolution (Carion *et al.* 2020). From the Transformer's point of view, it has powerful global modeling capability. But CNNs can effectively handle low-level features to enhance the localization of the transformer. Therefore, in order to ensure computational efficiency and accuracy (Wang *et al.* 2021; Chu *et al.* 2021; Yuan *et al.* 2021; Heo *et al.* 2021), it is still mainly convolutional, and the Swin Transformer is used to assist in enhancing the feature extraction capability. (2) Contrast experiments: For

Table 1. The ablation experiment results of the improved algorithm on the DIOR data set.

	Model and AP					
Category	**YOLOv5s**	**A**	**B**	**C**	**D**	**E**
Airplane	0.937	0.937	0.938	0.950	0.925	0.938
Airport	0.572	0.612	0.713	0.764	0.659	0.791
Baseball field	0.943	0.942	0.945	0.942	0.943	0.946
Basketball court	0.850	0.856	0.851	0.862	0.813	0.860
Bridge	0.480	0.489	0.498	0.509	0.453	0.519
Chimney	0.894	0.898	0.901	0.912	0.893	0.903
Dam	0.487	0.582	0.583	0.627	0.532	0.642
Expressway service area	0.660	0.712	0.715	0.717	0.675	0.737
Expressway toll station	0.683	0.695	0.701	0.730	0.649	0.708
Harbor	0.537	0.703	0.738	0.722	0.640	0.737
Golf course	0.791	0.804	0.807	0.812	0.790	0.813
Ground track field	0.610	0.659	0.671	0.664	0.592	0.664
Overpass	0.561	0.593	0.618	0.644	0.614	0.652
Ship	0.923	0.926	0.927	0.940	0.946	0.931
Stadium	0.922	0.905	0.916	0.904	0.892	0.914
Storage tank	0.829	0.836	0.841	0.855	0.845	0.840
Tennis court	0.920	0.921	0.921	0.931	0.933	0.923
Train station	0.372	0.507	0.537	0.567	0.492	0.549
Vehicle	0.734	0.762	0.769	0.768	0.768	0.772
Windmill	0.805	0.809	0.833	0.820	0.819	0.825
AP_S	0.294	0.317	0.357	0.362	0.338	**0.367**
AP_M	0.474	0.483	0.531	0.550	0.499	**0.563**
AP_L	0.521	0.608	0.607	0.601	0.549	**0.613**
$AP_{50:95}$	0.432	0.481	0.496	0.508	0.453	**0.515**
Recall	0.86	0.87	0.88	0.87	0.85	**0.88**
mAP@0.5/%	72.5	75.7	77.1	78.2	74.4	**78.3**
FPS	66	65	64	61	62	**64**

YOLO = You Only Look Once; FPS = frames per second.

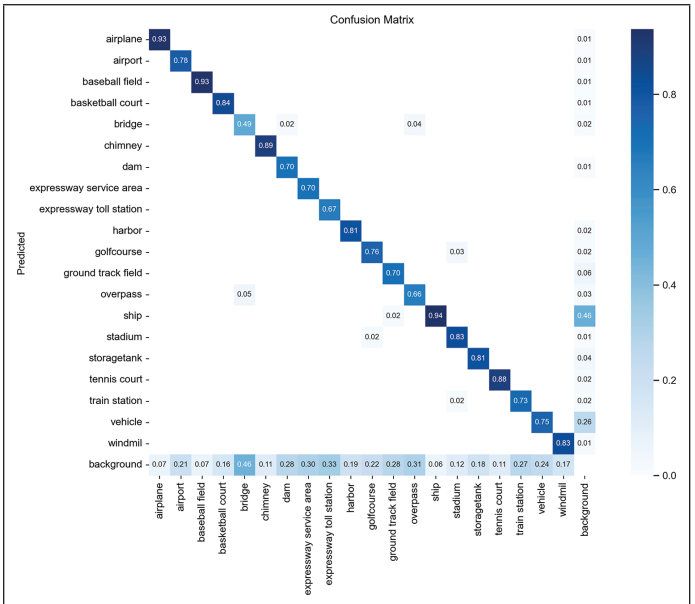


Figure 8. The confusion matrix of the final improved method on the DIOR data set with an Intersection over Union (IoU) threshold of 0.5 and a confidence threshold of 0.5.

testing, the final improved algorithm is put up against the YOLOv4, the YOLOv5s, the YOLOv5m, the Mask R-CNN, and the Swin Transformer algorithms based on the YOLO detection head. The results are given in Table 2, which gives the value of AP_S, AP_M, AP_L, Recall, map@0.5, and FPS. Meanwhile, we use the publicly available data set remote sensing object detection (RSOD) (Long *et al.* 2017) to verify the generalization ability of the model. The data set includes four different types of targets: aircraft, oil tanks, playgrounds, and overpasses. There are 446 aircraft images with 4993 instances, 165 oil tank images with 1586 instances, 189 playground images with 191 instances, and 176 overpass images with 180 instances. The ablation experiments and contrast experiments of RSOD data set are shown in Table 3 and Table 4. Compared to other common object detection methods, the improved algorithm outperforms in the DIOR and the RSOD data set of object detection. Improved in particular is the detection accuracy of small targets. The algorithm proposed in this paper works better when small targets are being detected in remote sensing images. Figure 9 shows the graph of the detection effect of Method E in order to provide a more intuitive assessment of the improved algorithm.

Table 2. The contrast experimental results on the DIOR data set.

Model	AP_S	AP_M	AP_L	Recall	mAP@0.5/%	FPS
YOLOv4	0.161	0.284	0.432	0.79	68.7	36
YOLOv5s	0.294	0.474	0.521	0.86	72.5	66
YOLOv5m	0.265	0.473	0.508	0.83	72.0	59
Mask R-CNN	0.213	0.362	0.487	0.82	69.1	18
Swin Transformer	0.355	0.519	0.556	0.85	73.7	57
YOLOv5s-E (ours)	**0.367**	**0.563**	**0.613**	**0.88**	**78.3**	**64**

YOLO = You Only Look Once; FPS = frames per second; R-CNN = region-based convolutional neural network.

Table 3. The ablation experiment results of the improved algorithm on the remote sensing object detection (RSOD) data set.

	Model and AP					
Category	**YOLOv5s**	**A**	**B**	**C**	**D**	**E**
Aircraft	0.912	0.916	0.935	0.936	0.925	0.937
Oil tank	0.963	0.973	0.974	0.976	0.968	0.974
Overpass	0.678	0.697	0.681	0.696	0.650	0.767
Playground	0.986	0.981	0.995	0.987	0.989	0.993
AP_S	0.362	0.379	0.385	0.406	0.387	**0.486**
AP_M	0.589	0.599	0.653	0.650	0.540	**0.713**
AP_L	0.571	0.571	0.606	0.626	0.542	**0.708**
$AP_{50:95}$	0.556	0.579	0.578	0.580	0.545	**0.633**
Recall	0.92	0.93	0.93	0.93	0.92	**0.95**
mAP@0.5/%	88.5	89.2	89.6	89.9	88.3	**91.8**
FPS	57	53	53	48	52	**51**

YOLO = You Only Look Once; FPS = frames per second.

Conclusion

Given the missed detection and false detection of small targets in remote sensing image target detection with complex backgrounds, this paper proposes some improvements based on YOLOv5s. The first is to add the detection layer of small targets to further integrate shallow features. The second is to add the window self-attention mechanism of the Swin Transformer to enhance the feature extraction ability of

Table 4. The contrast experimental results on the remote sensing object detection (RSOD) data set.

Model	AP_S	AP_M	AP_L	Recall	mAP@0.5/%	FPS
YOLOv4	0.267	0.355	0.439	0.83	75.5	17
YOLOv5s	0.362	0.589	0.571	0.92	88.5	57
YOLOv5m	0.264	0.476	0.511	0.89	84.5	45
Mask R-CNN	0.203	0.426	0.495	0.86	76.5	7
Swin Transformer	0.423	0.691	0.664	0.93	87.9	41
YOLOv5s-E (ours)	**0.486**	**0.713**	**0.708**	**0.95**	**91.8**	**51**

YOLO = You Only Look Once; FPS = frames per second; R-CNN = region-based convolutional neural network.

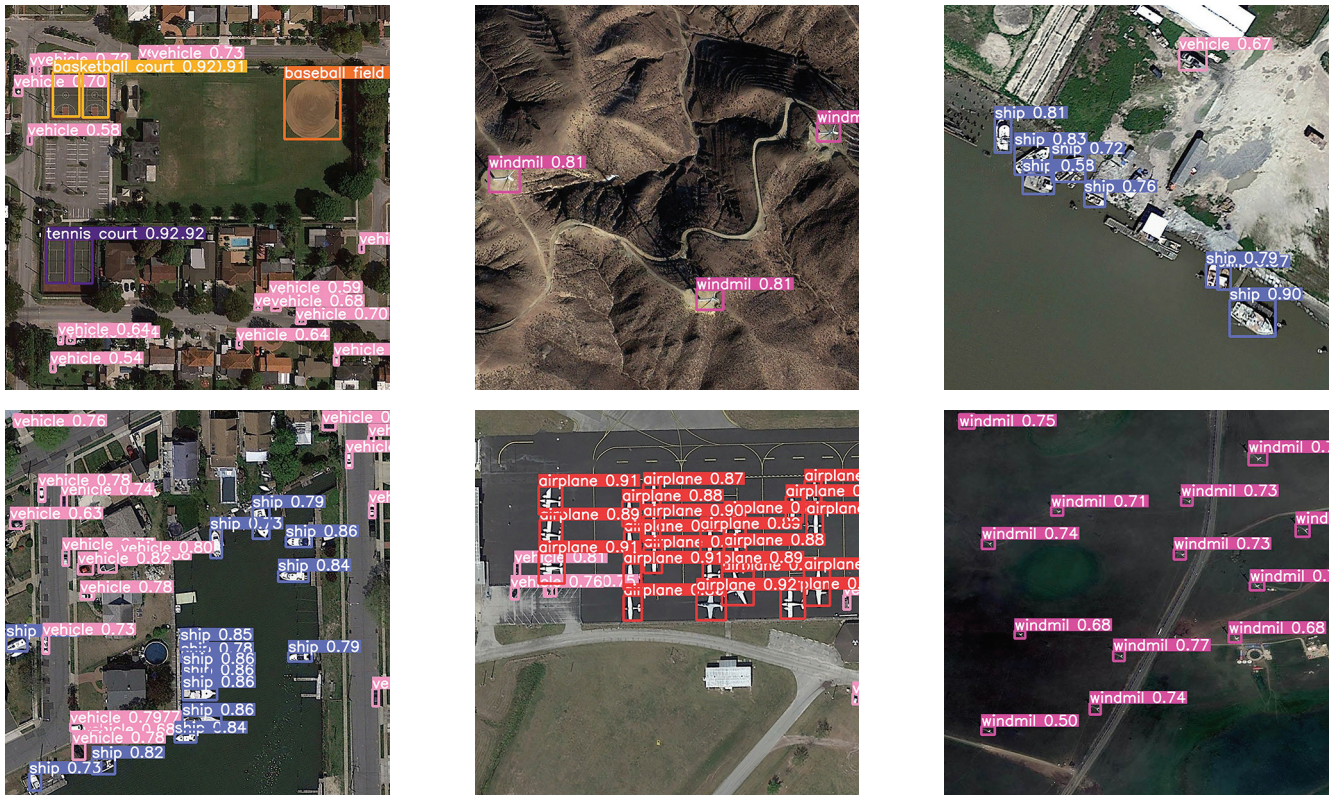


Figure 9. Method E visualization results on DIOR data set vs. remote sensing object detection (RSOD) data set.

targets. Finally, the optimization of the YOLOv5s detection algorithm is realized. The findings demonstrate that in order to enhance the small target feature performance capability, more location information for small targets can be gained by increasing the shallow feature map with 4x downsampling. The capacity to extract features can be improved by incorporating a window self-attention mechanism into the feature fusion pyramid to help the network concentrate more on the area of interest. The final improved algorithm obtained mAP of 0.783 and FPS of 64 on the DIOR data set and mAP of 0.918 and FPS of 51 on the RSOD data set. It is better than YOLOv4, YOLOv5s, YOLOv5m, Mask R-CNN, and other models with higher detection accuracy and real-time performance. It has better application prospects in the field of small object detection for remote sensing images. However, there are still some tiny undetected targets and the algorithm in this paper has not really been implemented into engineering applications. In the next step, we will continue to improve the algorithm, try to incorporate other more advanced attention mechanisms, and make the model lightweight to embed it into mobile devices for realistic scenarios.

Acknowledgements

This study was supported by Sichuan Science and Technology Program(2023YFS0406).

References

Bashir, S.M.A., Y. Wang, P. Melo-Pinto and M. E. Paoletti. 2021. Small object detection in remote sensing images with residual feature aggregation-based super-resolution and object detector network. *Remote Sensing* 13(9):1854.

Bochkovskiy, A., C. Y. Wang and H. M. Liao. 2020. YOLOv4: Optimal speed and accuracy of object detection. https://doi.org/10.48550/arXiv.2004.10934.

Carion, N., F. Massa, G. Synnaeve, N. Usunier, A. Kirillov and S. Zagoruyko. 2020. End-to-end object detection with transformers. *Lecture Notes in Computer Science* 12346:213–229.

https://doi.org/10.48550/arXiv.2005.12872.

Castelluccio, M., G. Poggi, C. Sansone and L. Verdoliva, 2015. Land use classification in remote sensing images by convolutional neural networks. *Acta Ecologica Sinica* 28(2):627–635.

Chen, Y., X. Chen, J. Zhou, Y. Ji and W. Shen 2017. Camouflage target detection via hyperspectral imaging plus information divergence measurement. In *Proceedings SPIE 10244, International Conference on Optoelectronics and Microelectronics Technology and Application*, 5 January 2017. https://doi.org/10.1117/12.2257938.

Chu, X. X., Z. Tian, Y. Q. Wang, B. Zhang, H. B. Ren, X. L. Wei, H. Xia and C. Shen. 2021. Twins: Revisiting the design of spatial attention in vision transformers. *Advances in Neural Information Processing Systems* 34:9355–9366.

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

Dalal, N. and B. Triggs. 2005. Histograms of oriented gradients for human detection. Pages 886–893 in *Proceedings 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, held in San Diego, Calif., 20–26 June 2005.

Dosovitskiy, A., L. Beyer, A. Kolesnikov, D. Weissenborn, X. H. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit and N. Houlsby. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. https://doi.org/10.48550/arXiv.2010.11929.

Fang, X. L., F. Hu, M. Yang, T. X. Zhu, R. Bi, Z. H. Zhang and Z. Gao. 2022. Small object detection in remote sensing images based on super-resolution. *Pattern Recognition Letters* 153(c):107–112.

Felzenszwalb, P., D. McAllester and D. Ramanan. 2008. A discriminatively trained, multiscale, deformable part model. Pages 1–8 in *Proceedings 2008 IEEE Conference on Computer Vision and Pattern Recognition*, held in Anchorage, Alsk., 24–26 June 2008.

Freund, Y. and R. E. Schapire. 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences* 55(1):119–139.

Girshick, R. 2015. Fast R-CNN. https://doi.org/10.48550/arXiv.1504.08083.

Girshick, R., J. Donahue, T. Darrell and J. Malik. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. https://doi.org/10.48550/arXiv.1311.2524.

Han, W. X., A. Kuerban, Y. C. Yang, Z. T. Huang, B. H. Liu and J. Gao. 2022. Multi-vision network for accurate and real-time small object detection in optical remote sensing images. *IEEE Geoscience and Remote Sensing Letters* 19:1–5.

He, K., G. Gkioxari, P. Dollár and R. Girshick. 2017. Mask R-CNN. Pages 2980–2988 in *Proceedings 2017 IEEE International Conference on Computer Vision (ICCV)*, held in Venice, Italy, 22–29 October 2017.

He, K., X. Zhang, S. Ren and J. Sun. 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37(9):1904–1916.

Heo, B., S. Yun, D. Han, S. Chun, J. Choe and S. J. Oh. 2021. Rethinking spatial dimensions of vision transformers. Pages: 11916-11925 in *Proceedings 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. https://doi.org/10.1109/ICCV48922.2021.01172.

Kanjir, U., H. Greidanus and K. Oštir. 2018. Vessel detection and classification from spaceborne optical images: A literature survey. *Remote Sensing of Environment* 207:1–26.

Law, H. and J. Deng. 2020. CornerNet: Detecting objects as paired keypoints. *International Journal of Computer Vision* 128(3):642–656.

Li, K., G. Wan, G. Cheng, L. Meng and J. Han. 2019. Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS Journal of Photogrammetry and Remote Sensing* 159:296–307.

Li, X., S. Zhang, X. Pan, P. Dale and R. Cropp. 2010. Straight road edge detection from high-resolution remote sensing images based on the ridgelet transform with the revised parallel-beam radon transform. *International Journal of Remote Sensing* 31(19):5041–5059.

Li, X. Z., C. L. Wang, N. Li and H. Sun. 2016. Window fusion feature contrast for optical remote sensing target detection. *Journal of Optical Precision Engineering* 24(08):2067–2077.

Li, Y. Y., Q. Huang, X. Pei, Y. Q. Chen, L. C. Jiao and R. H. Shang. 2021. Cross-layer attention network for small object detection in remote sensing imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14(0):2148–2161.

Lin, T. Y., M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick and P. Dollár. 2014. Microsoft COCO: Common objects in context. *Lecture Notes in Computer Science* 8693(1):740–755.

Liu, W., D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A. C. Berg. 2015. SSD: Single shot multibox detector. https://doi.org/10.48550/arXiv.1512.02325

Liu, Z., Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, et al. 2021. Swin transformer: Hierarchical vision transformer using shifted windows. Pages 9992–10002 in *Proceedings 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, held in Montreal, QC, Canada, 10–17 October 2021. https://doi.org/10.1109/ICCV48922.2021.00986.

Long, Y., Y. Gong, Z. Xiao and Q. Liu. 2017. Accurate object localization in remote sensing images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing* 55(5):2486–2498.

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

Redmon, J. and A. Farhadi. 2017. YOLO9000: Better, faster, stronger. Pages 6517–6525 in *Proceedings 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, held in Honolulu, Hawaii, 21–26 July 2017.

Redmon, J. and A. Farhadi. 2018. YOLOv3: An Incremental Improvement. https://doi.org/10.48550/arXiv.1804.02767.

Redmon, J., S. Divvala, R. Girshick and A. Farhadi. 2015. You only look once: Unified, real-time object detection. https://doi.org/10.48550/arXiv.1506.02640.

Ren, S., K. He, R. Girshick and J. Sun. 2015. Faster R-CNN: Towards real-time object detection with region proposal networks. https://doi.org/10.48550/arXiv.1506.01497.

Rezatofighi, H., N. Tsoi, J. Y. Gwak, A. Sadeghian, I. Reid and S. Savarese. 2019. Generalized intersection over union: A metric and a loss for bounding box regression. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.48550/arXiv.1902.09630.

Shi, T. J., J. N. Gong, J. M. Hu, X. Y. Zhi, W. Zhang, Y. Zhang, P. Zhang and G. Bao. 2022. Feature-enhanced CenterNet for small object detection in remote sensing images. *Remote Sensing* 14(5488):5488.

Song, Z. N., H. G. Sui and Y. C. Li. 2021. A review of ship target detection by high-resolution visible remote sensing images. *Journal of Wuhan University (Information Science Edition)* 46(11):1703–1715.

Tian, S., L. Kang, X. Xing, J. Tian, C. Fan and Y. Zhang. 2022. A relation-augmented embedded graph attention network for remote sensing object detection. *IEEE Transactions on Geoscience and Remote Sensing* 60:1–18.

Tian, Z., C. H. Shen, H. Chen and T. He. 2019. FCOS: Fully convolutional one-stage object detection. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. https://doi.org/10.48550/arXiv.1904.01355.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser and I. Polosukhin. 2017. Attention is all you need. https://doi.org/10.48550/arXiv.1706.03762.

Viola, P. and M. Jones. 2001. Rapid object detection using a boosted cascade of simple features. Pages I-I in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, held in Kauai, Hawaii, 2001, 8–14 December 2001. https://doi.org/10.1109/CVPR.2001.990517.

Wang, W. H., E. Xie, X. Li, D. P. Fan, K. T. Song, D. Liang, T. Lu, P. Luo and L. Shao. 2021. Pyramid vision transformer: A versatile backbone for dense prediction without convolutions. Pages 548–558 in *Proceedings 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, held in Montreal, QC, Canada. https://doi.org/10.1109/ICCV48922.2021.00061.

Xie, J. Z., H. Peng, J. F. Tang, Y. C. Hou and Q. X. Zeng. 2021. Improving YOLOv4 for dense remote sensing target detection. *Journal of Computer Engineering and Applications* 57(22):247–256.

Yuan, L., Y. P. Chen, T. Wang, W. H. Yu, Y. J. Shi, Z. H. Jiang, F.E.H. Tay, J. Feng and S. Yan. 2021. Tokens-to-token ViT: Training vision transformers from scratch on ImageNet. Pages 558–567 in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. https://doi.org/10.48550/arXiv.2101.11986.

Zhang, K., S. Y. Ni, D. Yan and A. D. Zhang. 2021. Review of dim small target detection algorithms in single-frame infrared images. Pages 2115–2120 in *Proceedings 2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, held in Chongqing, China. https://doi.org/10.1109/IMCEC51613.2021.9482137.

Zhang, L. G., L. Wang, M. Jin, X. S. Geng and Q. Shen. 2022. Small object detection in remote sensing images based on attention mechanism and multi-scale feature fusion. *International Journal of Remote Sensing* 43(9):3280–3297.

In Press

Leveraging NAIP Imagery for Accurate Large-Area Land Use/Land Cover Mapping: A Case Study in Central Texas. Mukti Ram Subedi, Carlos Portillo-Quintero, Samanthaa S. Kahl, Nancy E. McIntyre, Robert D. Cox, and Gad Perry.

Effect of Latitude as a Significant Element on the Results of Direct UTM Coordinates Transformation Method. Mohammed Anwer Jassim and Darin Mohammed Tofiq Mohammed.

CFAR Edge Detection Using Hysteresis Thresholding for Polarimetric SAR Imagery. Chaoyang Niu, Wanjie Lu, Wei Liu, Tao Hu, Shiju Wang, and Yajie Wu.

DORSL-FIN: A Self-Supervised Neural Network for Recovering Missing Bathymetry from ICESat-2. Forrest Corcoran and Christopher E. Parrish.

ASPRS
CODE OF ETHICS

Honesty, justice, and courtesy form a moral philosophy which associated with mutual interest among people should be the principles on which ethics are founded.

Each person who is engaged in the use development and improvement of the mapping sciences (Photogrammetry Remote Sensing Geographic Information Systems and related disciplines) should accept those principles as a set of dynamic guides for conduct and a way of life rather than merely for passive observance. It is an inherent obligation to apply oneself to one's profession with all diligence and in so doing to be guided by this Code of Ethics.

Accordingly, each person in the mapping sciences profession shall have full regard for achieving excellence in the practice of the profession and the essentiality of maintaining the highest standards of ethical conduct in responsibilities and work for an employer all clients colleagues and associates and society at large and shall…

asprs THE IMAGING & GEOSPATIAL INFORMATION SOCIETY

www.asprs.org

1. Be guided in all professional activities by the highest standards and be a faithful trustee or agent in all matters for each client or employer.
2. At all times, function in such a manner as will bring credit and dignity to the mapping sciences profession.
3. Not compete unfairly with anyone who is engaged in the mapping sciences profession by:
   a. Advertising in a self-laudatory manner;
   b. Monetarily exploiting one's own or another's employment position;
   c. Publicly criticizing other persons working in or having an interest in the mapping sciences;
   d. Exercising undue influence or pressure or soliciting favors through offering monetary inducements.
4. Work to strengthen the profession of mapping sciences by:
   a. Personal effort directed toward improving personal skills and knowledge;
   b. Interchange of information and experience with other persons interested in and using a mapping science with other professions and with students and the public;
   c. Seeking to provide opportunities for professional development and advancement of persons working under his or her supervision;
   d. Promoting the principle of appropriate compensation for work done by person in their employ..
5. Undertake only such assignments in the use of mapping sciences for which one is qualified by education training and experience and employ or advise the employment of experts and specialists when and whenever clients' or employers' interests will be best served thereby.
6. Give appropriate credit to other persons and/or firms for their professional contributions.
7. Recognize the proprietary privacy legal and ethical interests and rights of others. This not only refers to the adoption of these principles in the general conduct of business and professional activities but also as they relate specifically to the appropriate and honest application of photogrammetry remote sensing geographic information systems and related spatial technologies. Subscribers to this code shall not condone promote advocate or tolerate any organization's or individual's use of these technologies in a manner that knowingly contributes to:
   a. deception through data alteration;
   b. circumvention of the law;
   c. transgression of reasonable and legitimate expectation of privacy.
8. Promote equity, inclusion and intellectual diversity in the mapping sciences. Encourage participation without regard to race, religion, gender, disability, age, national origin, political affiliation, sexual orientation, gender identity, or gender expression.

Application of Improved YOLO V5s Model for Regional Poverty Assessment Using Remote Sensing Image Target Detection

Zhang Chenguang and Teng Guifa

Abstract

This study aims at applying the improved You Only Look Once V5s model for the assessment of regional poverty using remote sensing image target detection. The model was improved from structure, algorithm, and components. Objects in the remote sensing images were used to identify poverty, and the poverty alleviation situation could be predicted according to the existing detection results. The results showed that the values of Precision, Recall, mean Average Precision (mAP)@0.5, and mAP@0.5:0.95 of the model increased 7.3%, 0.7%, 1%, and 7.2%, respectively on the Common Objects in Context data set in the detection stage; the four values increased 3.1%, 2.2%, 1.3%, and 5.7%, respectively on the custom remote sensing image data set in the verification stage. The loss values decreased 2.6% and 37.4%, respectively, on the two data sets. Hence, the application of the improved model led to the more accurate detection of the targets. Compared with the other papers, the improved model in this paper proved to be better. Artificial poverty alleviation can be replaced by remote sensing image processing because it is inexpensive, efficient, accurate, objective, does not require data, and has the same evaluation effect. The proposed model can be considered as a promising approach in the assessment of regional poverty.

Introduction

As an important application of hyper spectral remote sensing image, target detection belongs to the high dimensional data (Bitar *et al.* 2019; Peng *et al.* 2020). Moreover, it has higher coupling and correlation in spectral as well as spatial resolutions (Liu and Qiao 2020; Mei *et al.* 2020). While acquiring poverty data is time consuming and costly and greatly influenced by human factors, the use of remote sensing images to detect representation objects can objectively reflect the regional economic development with the use of no data.

Recently, deep learning has played an important role in remote sensing image target detection, and consequently, has broadened the knowledge in this area (Li *et al.* 2018). It is based on convolutional neural network, the performance and design of which are constantly improving (Yuan *et al.* 2020). Target detection based on deep learning includes candidate region-based and regression-based approaches (Zheng *et al.* 2022). In the former approach, also known as the two-stage method, candidate regions are firstly extracted, and subsequently the detection results are obtained, which includes Region-based Convolutional Neural Networks (R-CNN), Fast R-CNN, Faster R-CNN, and Mask R-CNN (Yan *et al.* 2019; Girshick 2015; Ji *et al.* 2019; Zhang *et al.* 2020). On the other hand, in the latter approach, also known as the one-stage method, the extraction process is removed for the candidate regions, and the results are directly obtained through regression. This method includes You Only Look Once (YOLO) and Single Shot MultiBox Detector (SSD) (Xu and Wu 2020; Qu *et al.* 2020).

YOLO V5 is a model with good detection effect among the one-stage methods, enabling real-time detection with high accuracy and low computation. At present, YOLO V5 model has been recognized as one of the best detection tools, and has been widely used in remote sensing image target detection process. In Tang *et al.* (2021), an improved N-YOLO model was proposed to solve the problem of noise in synthetic aperture radar (SAR) images. Firstly, the classification component noise level classifier was used to derive and classify the noise, and then the target extraction module SAR target potential area extraction was used to obtain the complete region of the target. Finally, the results of the two parts were fused. By comparing with CNN algorithm, it was proved that the model had better performance and could realize the identification of ship target. In Luo *et al.* (2022), an improved YOLO V5-aircraft model was proposed to solve the problem of insufficient speed and accuracy of aircraft recognition in remote sensing images. Firstly, the centering and scaling operations were added in the batch process to increase the feature extraction capability of the model. Then, the cross entropy loss function was replaced by the smooth Karhunen-Loève algorithm. Finally, the CSandGlass module was used to replace the residual module, and the low-resolution feature layer was removed. By comparing with the original model, it was proved that the improved model had better speed, accuracy, and convergence. In Zhou *et al.* (2022), a multi-scale ship detection method based on the improved YOLO V5s model was proposed. Firstly, combined with the module-collaborative attention mechanism (CAM), the improved residual module Res2Net was introduced into the new module CSPMRes2 to make it have the feature extraction ability in both depth and scale dimensions. Then, a new feature pyramid network with fusion coefficients module was created to provide coefficients for the optimal feature fusion scheme. Finally, multi-scale detection of ship objects was realized.

In the present study, the improved YOLO V5s model was used for the assessment of regional poverty using remote sensing image target detection. In view of the limitations of the existing research, the following solutions were proposed. First, we used extensive data sources for the analysis and presentation of results. Previously, regional poverty assessment was based on statistical methods or big data technology, influenced by human factors. Now, the remote sensing images target detection results were used for auxiliary verification to achieve an objective decision process. Second, the original model has been improved comprehensively. Previous improvement methods were based on components in a simple form with limited improvement in the target detection ability. However, in this paper, model structure, loss function, and component design were all considered, and the

Zhang Chenguang is with the College of Information Science and Technology, Hebei Agricultural University, Baoding Hebei China 071000; School of Computer and Information Technology, Cangzhou Jiaotong College, Huanghua Hebei China 061100 (cgzhang@czjtu.edu.cn).

Teng Guifa is with the College of Information Science and Technology, Hebei Agricultural University, Baoding Hebei China 071000.

Corresponding authors: Teng Guifa (tguifa@126.com) and Zhang Chenguang (cgzhang@czjtu.edu.cn)

Contributed by Dongdong Wang, January 17, 2023 (sent for review January 29, 2023; reviewed by Mohammad Hajeb).

Photogrammetric Engineering & Remote Sensing
Vol. 89, No. 8, August 2023, pp. 499–513.
0099-1112/22/499–513

doi: 10.14358/PERS.23-00005R3

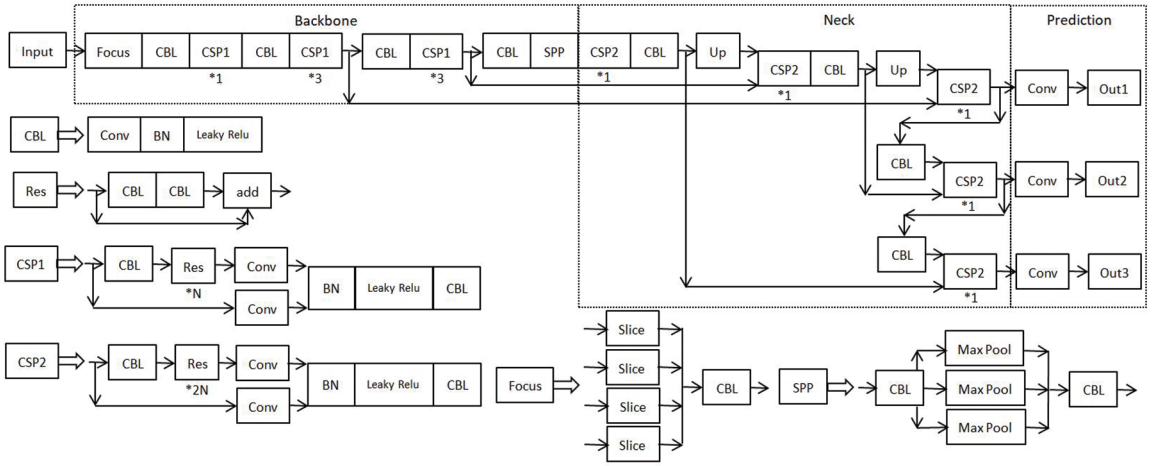


Figure 1. YOLO V5s model structure diagram. CBL = Conv + BN + Leaky relu; Conv = convolution; CSP = Cross Stage Partial; BN = BatchNorm2d function; Up = up-sampling; SPP = Spatial Pyramid Pooling.

feature localization capability was improved significantly by increasing the depth of the pyramid attention network (PAN). The angle relation replaced the position relation between the ground truth (GT) box and the prediction (Pre) box in the complete intersection over union (CIoU) loss function and the detection performance was improved. The attention mechanism was introduced and improved, and the multi-channel convolution technique was combined to enhance extraction capability for different types of features. Third, the detection results were more comprehensive. Performance indicators to evaluate the YOLO V5s model are P (Precision), R (Recall), mAP@0.5 (mean Average Precision @ IoU = 0.5), mAP@0.5:0.95, and Loss. Most studies that focus on the detection result parameters do not provide a comprehensive reference for readers. However, this paper detected all parameters, so as to provide a more authoritative reference basis.

Materials and Methods

YOLO V5s Model

Figure 1 shows the structure of YOLO V5s model as a whole and its parts. Up indicates up-sampling, Conv means convolution, and BN stands for BatchNorm2d function. Focus, CBL (Conv + BN + Leaky relu), CSP (Cross Stage Partial), and SPP (Spatial Pyramid Pooling) structures are shown as the components. Compared with the previous versions, new functions such as adaptive anchor box calculation, adaptive image scaling, and improved Neck structure are introduced.

Data Set Preparation

Remote sensing images were derived from the data integration of satellite images and aerial photography, among which: satellite images were derived from DigitalGlobe's QuickBird (USA) commercial satellite and EarthSat's *LANDSAT-7* (USA) remote sensing data; the aerial images were derived from BlueSky (UK), Sanborn (USA), *IKONOS* (USA), and *SPOT5* (France). Parameters of the image are shown in Table 1.

BIGEMAP[1] software, with a resolution of 0.5 m, was used to obtain the remote sensing images. A total number of 842 304 images were gathered. The images in the determined years were downloaded after selecting the region and adjusting the time node, as shown in Figure 2.

In this study, we used Google tiles in JPG format, as shown in Table 2.

The LabelImg tool[2], developed by Python, was used for labeling the images. Figure 3 shows the annotation process. The results of the annotations were files in XML format that must be converted to TXT format (Appendix A) before they can be used in the YOLO V5s model. File contents were location informations of the objects in remote sensing images, that was, the production of the GT box data.

Table 1. Common satellite bands and resolutions.

Satellite	Bands (nm)	Resolution (m)
QuickBird	blue: 450–520	2.44
	green: 520–660	2.44
	red: 630–690	2.44
	near-infrared: 760–900	2.44
Landsat-7	blue-green: 450–520	30
	green: 520–600	30
	red: 630–690	30
	near-infrared: 760–900	30
	intermediate infrared: 1550–1750	30
	thermal infrared: 10 400–12 500	60
	intermediate infrared: 2080–2350	30
	panchromatic: 520–900	15
SPOT5	green: 500–590	10
	red: 610–680	10
	near-infrared: 780–890	10
	intermediate infrared: 1580–1750	10
	panchromatic: 480–710	5
	super panchromatic: 480–710	2.5

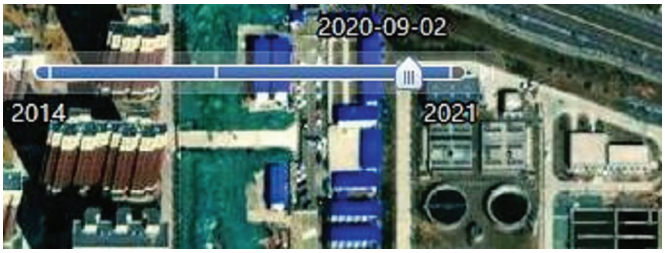


Figure 2. Selection of the image from the determined years.

Table 2. Selection of the download format.

Level	Scale	Resolution	Format	Size
18	1:1757	0.5 m	Google Tile (*.jpg)	256 × 256

1. Chengdu Bigemap Data Processing Co., Ltd. 2023. Bigemap. V25.5.0.1. Windows All. Chengdu: Chengdu Bigemap Data Processing Co., Ltd.
2. Python community. 2021. LabelImg. V1.8.6. Windows All. https://pypi.org/project/labelImg/: Python community.

Poverty Factors and Representational Relationship

Objects in remote sensing images indicates the regional economic development factors. Table 3 shows the poverty factors and representational relationships. According to the poverty alleviation policy, the poverty factors include food, housing, transportation, and education. Contents that cannot be directly observed through remote sensing images need to be calculated through traditional statistical data, which has been done before this study.

YOLO V5s Model Improvement

The following steps were used to improve the YOLO V5s model:

Step 1: Enhanced pan Structure

Table 4 shows the PAN + FPN + PAN structure of the original model. The FPN is an up-sampling process, through which feature extraction results are enlarged layer by layer for semantic recognition. However, the PAN is a down-sampling process, through which the positioning information is preciser layer by layer and is used for positioning recognition. Each value in the table has a uniform format: serial number + name + parameter set.

(1) serial number: L(N), namely the layer N network. In YOLO V5s model's Yaml file, the order in which the network layers are written is the order in which they are executed.
(2) name: the name of the network layer being executed. These names are the same as in Figure 1 and in the same order.
(3) parameter set: in order they are: the number of batches, the number of channels for input data, and the width and height of output data.

As shown in Table 5, the structure of the improved model has the following changes, compared with the original model: first, layers were expanded from 25 to 28. Second, the FPN process was shortened. In layer 16, the image size increased from 32 to 128, which was the end of the FPN process and the beginning of the second PAN process, making the second PAN have greater depth. Third, layers 17–19 were introduced, and the intermediate features extracted were fused with layer 3 in the channel dimension. Thus, both PAN processes implemented down-sampling from 128 to 8. The improved model indirectly increased the overall depth by increasing the PAN depth. This process enhanced the feature extraction ability and improved the feature localization accuracy.

The improved portions of the model are shown in Figure 4.

Step 2: RIoU_Loss (Round IoU) Regression Loss Function

Equations 1, 2, and 3 are used to express the CIoU regression loss function (Zheng *et al.* 2021).

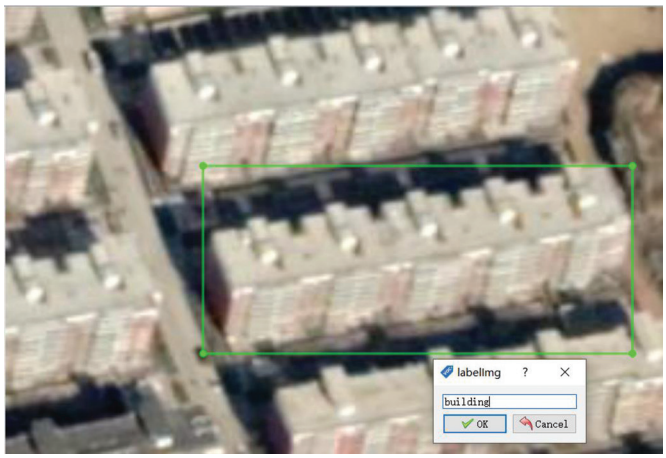


Figure 3. Labeling of the representation objects in the remote sensing image.

Table 3. Label type and representational relationships.

Poverty Factors	Representations	Statistical Data
Food	Greenhouses	Water resources
	Cultivated area	Sewage/household refuse
		Food expenditure
		Per capita income
		Consumption expenditure
Housing	Bungalows	Urban/rural population
	Buildings	Population growth/mortality rate
		Housing expenditure
		Housing sales
		Household appliances
Transportation	Highway	Transportation expenditure
		Travel
		Electricity/post
Education	Playgrounds	Educational expenditure
		Science and technology input/output
		Number of students/teachers
		Sports and entertainment
		Employment situation

Table 4. PAN + FPN + PAN structure of the original YOLO V5s model.

PAN	FPN	PAN
L0:input:[1,32,128,128]	—	—
L1:Focus:[1,64,64,64]	—	—
L2:CBL:[1,64,64,64]	—	—
L3:CSP1:[1,64,64,64]	L18:CSP2:[1,128,32,32]	L19:CBL:[1,128,16,16]
L4:CBL:[1,128,32,32]	L17:Concat L5:[1,256,32,32]	L20:Concat L15:[1,256,16,16]
L5:CSP1:[1,128,32,32]	L16:Up:[1,128,32,32]	L21:CSP2:[1,256,16,16]
L6:CBL:[1,256,16,16]	L15:CBL:[1,128,16,16]	L22:CBL:[1,256,8,8]
L7:CSP1:[1,256,16,16]	L14:CSP2:[1,256,16,16]	L23:Concat L11:[1,512,8,8]
L8:CBL:[1,512,8,8]	L13:Concat L7:[1,512,16,16]	L24:CSP2:[1,512,8,8]
L9:SPP:[1,512,8,8]	L12:Up:[1,256,16,16]	L25:Detect:(L18,L21,L24)
L10:CSP2:[1,512,8,8]	L11:CBL:[1,256,8,8]	—

CBL = Conv + BN + Leaky relu; CSP = Cross Stage Partial; FPN = feature pyramid network; L = layer; PAN = pyramid attention network.

Table 5. Enhanced PAN structure of the improved YOLO V5s model.

PAN	FPN	PAN
L0:input:[1,32,128,128]	—	—
L1:Focus:[1,64,64,64]	—	—
L2:CBL:[1,64,64,64]	—	—
L3:CSP1:[1,64,64,64]	—	—
L4:CBL:[1,128,32,32]	—	—
L5:CSP1:[1,128,32,32]	L16:Up:[1,128,128,128]	L17:CBL:[1,64,64,64]
L6:CBL:[1,256,16,16]	L15:CBL:[1,128,16,16]	L18:Concat L3:[1,128,64,64]
L7:CSP1:[1,256,16,16]	L14:CSP2:[1,256,16,16]	L19:CBL:[1,128,32,32]
L8:CBL:[1,512,8,8]	L13:Concat L7:[1,512,16,16]	L20:Concat L5:[1,256,32,32]
L9:SPP:[1,512,8,8]	L12:Up:[1,256,16,16]	L21:CSP2:[1,128,32,32]
L10:CSP2:[1,512,8,8]	L11:CBL:[1,256,8,8]	L22:CBL:[1,128,16,16]
—	—	L23:Concat L15:[1,256,16,16]
—	—	L24:CSP2:[1,256,16,16]
—	—	L25:CBL:[1,256,8,8]
—	—	L26:Concat L11:[1,512,8,8]
—	—	L27:CSP2:[1,512,8,8]
—	—	L28:Detect:(L21,L24,L27)

CBL = Conv + BN + Leaky relu; CSP = Cross Stage Partial; FPN = feature pyramid network; L = layer; PAN = pyramid attention network.

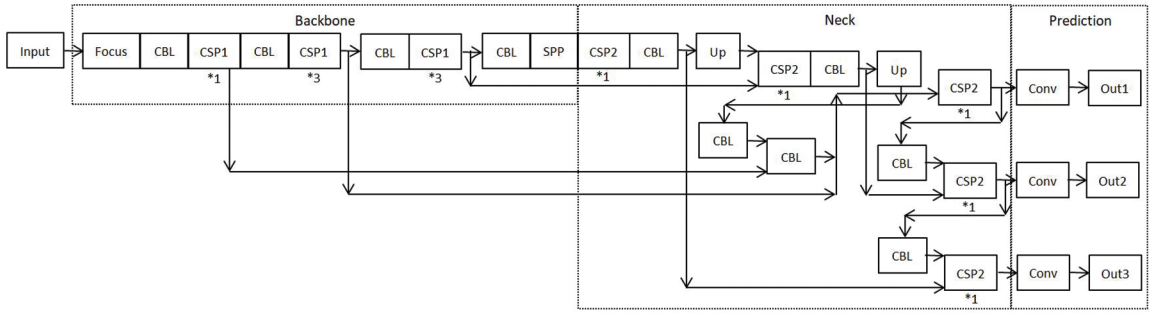


Figure 4. Comparison of enhanced pyramid attention network (PAN) structure with original model. CBL = Conv + BN + Leaky relu; Conv = convolution; CSP = Cross Stage Partial; Up = up-sampling; SPP = Spatial Pyramid Pooling.

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w^{p}}{h^{p}}\right)^2 \quad (1)$$

$$\text{alpha} = \frac{v}{1-\text{iou}+v} \quad (2)$$

$$\text{CIoU} = 1-\left(\text{iou}-\frac{D\_2^2}{D\_C^2}-v*\text{alpha}\right) \quad (3)$$

where w^{gt} denotes the width of the GT box; h^{gt} stands for the height of the GT box; w^{p} represents the width of the Pre box; h^{p} is the height of the Pre box; D_2 signifies the distance between the two centers, and D_C is the diagonal distance of the two enclosing rectangles.

Thus, the factors of overlap area, central point distance, and aspect ratio have been considered. However, location difference is disregarded. For example, P_1 and P_2 are embedded in GT (Figure 5); all have the same aspect ratio. Hence, $v = 0$ is always true in Equation 1. Accordingly $v \cdot \text{alpha} = 0$ is always true in Equation 3. While the areas enclosed by P_1 and P_2 are unequal, when

$$\text{iou}_1 - \frac{D\_2_1^2}{D\_C^2} = \text{iou}_2 - \frac{D\_2_2^2}{D\_C^2} \quad (4)$$

is true, the CIoU values of P_1 and P_2 are equal and cannot be distinguished. Since P_1 is far from the GT center and P_2 is close, D_2_1 is larger than D_2_2. If iou_1 is larger than iou_2 simultaneously, then Equation 4 holds some probability.

As shown in Figure 6, point O is the center of GT; point A denotes the top left of GT; point B stands for the center of Pre. Circle R is formed, taking O as the center, and D_2 is the radius. Then, R is the trajectory of B. However, IoU, D_C, and D_2 are all decidable, while the position of B is random, and the paths of B and R coincide. In the triangle OAB, the lengths of OA and OB are fixed, whereas the length of AB is variable as B moves, and the internal angles change accordingly. By the law of sine, when

$$\text{BO} / \sin\angle\text{BAO} = \text{AO} / \sin\angle\text{ABO} \quad (5)$$

is true,

$$\sin\angle\text{BAO} / \sin\angle\text{ABO} = \text{BO}/\text{AO} \quad (6)$$

is established. Moreover, the position relationship between Pre and GT is transformed into the angle relationship. The angle ∠BAO reflects the movement of B on R, which is the position of Pre concerning GT. B moves from 0° to 360°, covering every possible scenario. When B moves along R to E, the value of does not change. was used for auxiliary verification. Referring to Equation 5 and Equation 6, when

$$\text{AB} / \sin\angle\text{AOB} = \text{AO} \sin\angle\text{ABO} \quad (7)$$

is true,

$$\sin\angle\text{AOB} / \sin\angle\text{ABO} = \text{AB}/\text{AO} \quad (8)$$

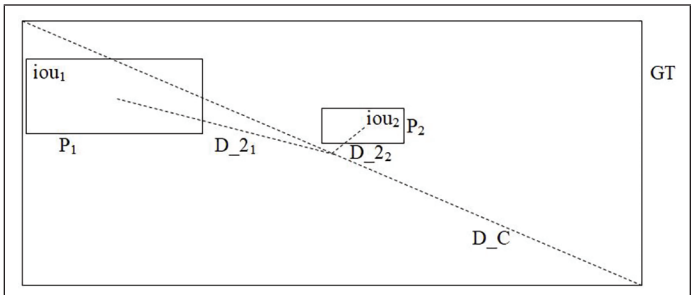


Figure 5. Invalid RIoU_Loss regression loss function. IoU = intersection over union; RIoU =round IoU.

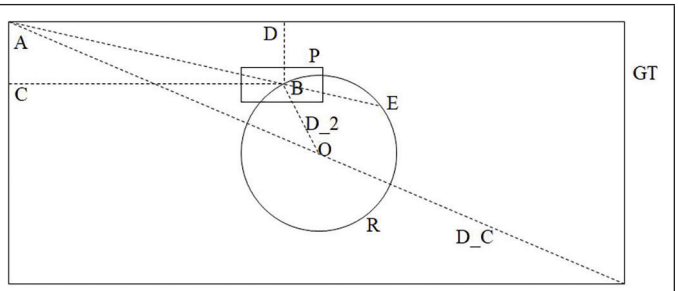


Figure 6. RIoU_Loss regression loss function schematic diagram. RIoU = round intersection over union.

is established. Moreover, BC and BD are the vertical distances from B to GT, and the following equations

$$\text{BC} = |x_B - x_A| \quad (9)$$

$$\text{BD} = |y_B - y_A| \quad (10)$$

$$\text{AB} = \sqrt{\text{BC}^2+\text{BD}^2} \quad (11)$$

are established. In summary, the RIoU_Loss regression loss function is defined as

$$\text{iou} - (\text{BO}/\text{AO})\times(\text{AB}/\text{A}O) = \text{iou} - (4\times D\_2\ \text{AB}/D\_C^2) \quad (12)$$

Step 3: Collaborative Attention Mechanism

Simply combining the attention mechanism with the YOLO V5s model cannot improve the performance of detection results effectively. Table 6 lists common attention mechanisms (Hu *et al.* 2018; Woo *et al.* 2018; Wang *et al.* 2020; Hou *et al.* 2021), compared with Step 2. The meanings of the parameters are as follows:

(1) Memory(G): the memory space required to run the model, the unit of measurement is gigabyte.
(2) Model Layers: the number of layers of the model, including convolution, batch normalization, activation, and other functions.
(3) Yaml Layers: the number of layers of Yaml file, which reflects the order in which the model is run and the related functions.

Table 6. Detection results of YOLO V5s model combined with a common attention mechanism.

	Memory (G)	Model Layers	Yaml Layers	Time/ Image(ms)	Weight	P	R	mAP @0.5	mAP @.5:.95	GFLOPS (B)
Step 2 (Baseline)	4.69	310	28	91.927	15.3M / 7518909	0.7143	0.98	0.9746	0.8425	23.9
+SE	3.69	316	28	105.525	15.4M / 7551677	0.6544	0.9714	0.9649	0.817	24
+CBAM	3.69	321	28	105.173	15.4M / 7551775	0.6691	0.9715	0.9655	0.8041	24
+ECA	3.69	314	28	106.383	15.3M / 7518912	0.6672	0.9687	0.9631	0.8114	23.9
+CA	3.73	320	28	105.891	15.4M / 7544557	0.6611	0.9758	0.9671	0.8003	24

CA = Coordinate Attention; CBAM = Convolutional Block Attention Module; ECA = Efficient Channel Attention; GFLOPS = giga floating-point operations per second; mAP = mean Average Precision; P = Precision; R = Recall; SE = Squeeze-and-Excitation; YOLO = You Only Look Once.

(4) Time/Image(ms): the time taken to execute a single image, the unit is millisecond.
(5) Weight: there are two ways to measure it. The first is the size of weight file, the unit is megabyte. The second is the total number of parameters during model running.
(6) P: Precision.
(7) R: Recall.
(8) mAP@0.5: mean Average Precision @ IoU = 0.5.
(9) mAP@0.5:0.95: mean Average Precision @ IoU = 0.5:0.95.
(10) GFLOPS(B): giga floating-point operations per second, the unit is byte.

The main performance indicators of the model (P, R, mAP@0.5, and mAP@0.5:0.95, obtained based on the COCO data set), does not improve but decrease. There are two possible reasons: first, the traditional attention mechanisms must conduct pooling operations in the dimensions of scale or channel, making the results inconsistent with the overall original data dimensions and causing data loss. Second, the sigmoid function is used to generate coefficients for data enhancement, transforming from a linear to nonlinear distribution. Figure 7 shows that the conversion results in the interval $[a,b]$ are better than the initial linear function ($y = nx + m$), whereas in the intervals $[-\infty,a]$ and $[b,+\infty]$ it approaches the lower and upper limits indefinitely, making the data distribution not properly reflected. Experimental data showed large fluctuations in the early stage of the training process. The model performance was difficult to continuously improve at the later stage, because its ability to extract features was limited and challenging.

The new CAM consisted of two parts. First, the reverse filling attention mechanism (RFAM) component was proposed. Unlike the traditional combinations, this component did not plug into the existing model but replaced layer 11. In other words, the RFAM module replaced the Conv + BN + Leaky Relu (CBL) module, based on the connection between the Backbone and the Neck. Second, a multi-channel convolution (MCC) component was added after layer 21, layer 25, and layer 29 for extracting different features based on Step 2, as the connection between the Neck and the Prediction. Meanwhile, the CBL modules in layer 23 and layer 27 were replaced by dilated convolution (DC) modules to alleviate data loss in the down-sampling process, as shown in Table 7 (*see next page*).

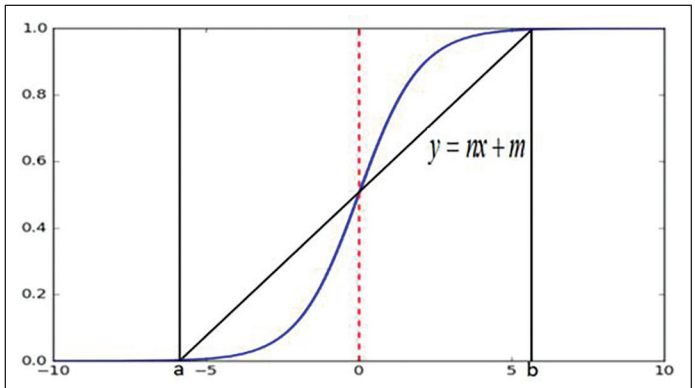


Figure 7. Sigmoid function versus linear function.

Figure 8 shows the RFAM component, which is located at the junction of the Backbone and the Neck. The meanings of portions are as follows:

(1) AP: nn.AdaptiveAvgPool2d(h, w) function. Apply a 2D adaptive average pooling over an input signal composed of several input planes. The output is of size $h \times w$, for any input size. The number of output features is equal to the number of input planes.
(2) Re: Torch.Tensor.Repeat(size) function. Repeat a tensor along the specified dimensions. Parameter "size" is the number of times to repeat.
(3) M: Torch.Mean(input) function. Return the mean value of all elements in the input tensor.
(4) Sub: Torch.Subtract(input) function.

Three sequences were executed simultaneously. In the first sequence, the input data were completely preserved without any processing, similar to the input data in the residual network. In the second sequence, channel calculation and pixel calculation were performed in two steps. In the channel calculation, AP was first used to convert the input data size from $h \times w$ to 1×1. Subsequently, Re was used to reverse fill and restore the size to $h \times w$. Thus, the value of each pixel was the same for any channel. Finally, the current result was used to calculate a difference in the input data by Sub. The difference between each pixel

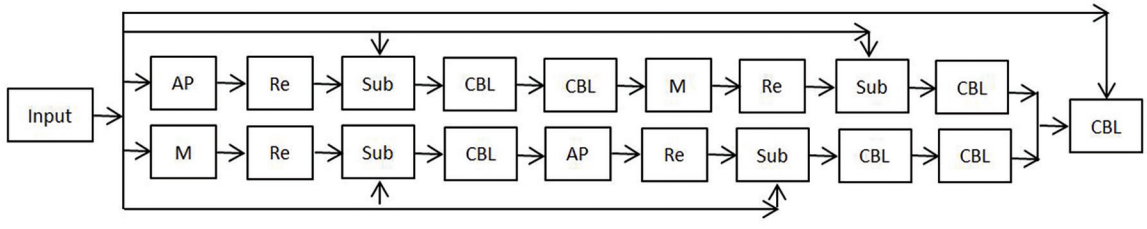


Figure 8. Reverse filling attention mechanism (RFAM) structure. AP = nn.AdaptiveAvgPool2d(h, w) function; CBL = Conv + BN + Leaky relu; M = Torch.Mean(input) function; Re = Torch.Tensor.Repeat(size) function; Sub = Torch.Subtract(input) function.

Table 7. Collaborative attention mechanism structure of the improved YOLO V5s model.

PAN	FPN	PAN
L0:input:[1,32,128,128]	—	—
L1:Focus:[1,64,64,64]	—	—
L2:CBL:[1,64,64,64]	—	—
L3:CSP1:[1,64,64,64]	—	—
L4:CBL:[1,128,32,32]	—	—
L5:CSP1:[1,128,32,32]	L16:Up:[1,128,128,128]	L17:CBL:[1,64,64,64]
L6:CBL:[1,256,16,16]	L15:CBL:[1,128,16,16]	L18:Concat L3:[1,128,64,64]
L7:CSP1:[1,256,16,16]	L14:CSP2:[1,256,16,16]	L19:CBL:[1,128,32,32]
L8:CBL:[1,512,8,8]	L13:Concat L7:[1,512,16,16]	L20:Concat L5:[1,256,32,32]
L9:SPP:[1,512,8,8]	L12:Up:[1,256,16,16]	L21:CSP2:[1,128,32,32]
L10:CSP2:[1,512,8,8]	L11:RFAM:[1,256,8,8]	L22:MCC:[1,128,32,32]
—	—	L23: DC:[1,128,16,16]
—	—	L24:Concat L15:[1,256,16,16]
—	—	L25:CSP2:[1,256,16,16]
—	—	L26:MCC:[1,256,16,16]
—	—	L27: DC:[1,128,8,8]
—	—	L28:Concat L11:[1,512,8,8]
—	—	L29:CSP2:[1,512,8,8]
—	—	L30:MCC:[1,512,8,8]
—	—	L31:Detect:(L22,L26,L30)

CBL = Conv + BN + Leaky relu; CSP = Cross Stage Partial; FPN = feature pyramid network; L = layer; PAN = pyramid attention network; YOLO: You Only Look Once.

of the input data was noted, which simultaneously meets the requirement of increasing the target difference in the feature extraction process. The CBL modules were executed sequentially to further extract features. Moreover, the number of channels was compressed (c/16) and decompressed (c). In the subsequent pixel calculation process, M was used to convert the number of channels from c to 1. Then, Re was used to reverse fill to restore the number to c, and the difference in the input data was calculated based on this result by Sub. Similarly, the CBL module with a convolution kernel size of 7 was used for further feature extraction. The third sequence was like the second, except that the channel calculation was swapped with the order of the pixel calculation. Finally, the results of the three sequences were combined and the CBL model was obtained. The RFAM component performed channel calculation and pixel calculation for the input data at different angles. Unlike the traditional attention mechanisms, the sigmoid function for coefficient was replaced by reverse filling, which reduced data loss and created feature extraction optimization. The parameter set is shown in Table 8.

The performance of RFAM is shown in Table 9.

T- hypothesis test was used to compare the performances of step 2, RFAM, and the attention mechanisms in Table 6. The implementation process is as follows:

In Table 10, method1 and method2 are the two objects to be compared, d is the difference between them, and n is the number of values in each method. Suppose $d_i = x_i - y_i$, (i = 1, 2, …, n) is the sample from normal population $N(\mu_D, \mu_D^2)$, μ_D, and μ_D^2 are unknown, and set significance level $\alpha = 0.1$, then hypothesis test can be expressed as:

$$H_0 : \mu_D \geq 0, H_1 : \mu_D < 0,$$

namely, compare

$$\frac{\overline{x_d}}{s_d / \sqrt{n}}$$

and $-t_{0.1}(n-1)$, where s_d is the standard deviation. The performance ordering is shown in Table 11.

Table 8. Parameter set generated during Reverse filling attention mechanism (RFAM) running.

Sequence	Name	Batch	Channel	Height	Width
input	Input	b	c	h	w
	AP	b	c	1	1
	Re	b	c	h	w
	Sub	b	c	h	w
	CBL	b	c/16	h	w
2	CBL	b	c	h	w
	M	b	1	h	w
	Re	b	c	h	w
	Sub	b	c	h	w
	CBL	b	c	h	w
	M	b	1	h	w
	Re	b	c	h	w
	Sub	b	c	h	w
	CBL	b	c	h	w
3	AP	b	c	1	1
	Re	b	c	h	w
	Sub	b	c	h	w
	CBL	b	c/16	h	w
	CBL	b	c	h	w
output	CBL	b	c	h	w

AP = nn.AdaptiveAvgPool2d(h, w) function; CBL = Conv + BN + Leaky relu; Conv = convolution; BN = BatchNorm2d function; M = Torch.Mean(input) function; Re = Torch.Tensor.Repeat(size) function; Sub = Torch.Subtract(input) function.

Table 9. Detection results of You Only Look Once (YOLO) V5s model combined with RFAM component.

	P	R	mAP@0.5	mAP@.5:.95
+RFAM	0.6894	0.9758	0.9692	0.8317

mAP = mean Average Precision; P = Precision; R = Recall; RFAM = reverse filling attention mechanism.

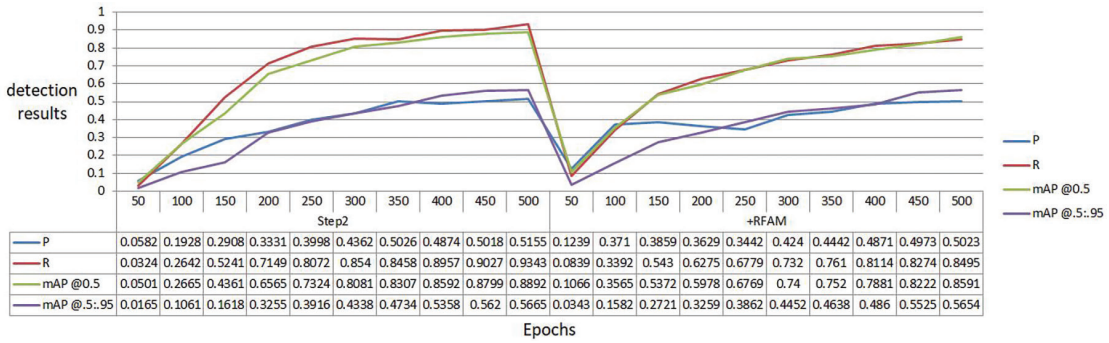
Table 10. T-hypothesis test analog data.

Objects	Values			
method1 (x)	value1_1	value1_2	···	value1_n
method2 (y)	value2_1	value2_2	···	value2_n
$D = x - y$	d1	d2	···	dn

Table 11. Performance ordering among Step 2, reverse filling attention mechanism (RFAM) and attention mechanisms.

Objects	Ordering
Step 2	1
RFAM	2
ECA	3
CBAM	4
SE	5
CA	6

CA = Coordinate Attention; CBAM = Convolutional Block Attention Module; ECA = Efficient Channel Attention; RFAM = reverse filling attention mechanism; SE = Squeeze-and-Excitation.



	50	100	150	200	250	300	350	400	450	500	50	100	150	200	250	300	350	400	450	500
P	0.0582	0.1928	0.2908	0.3331	0.3998	0.4362	0.5026	0.4874	0.5018	0.5155	0.1239	0.371	0.3859	0.3629	0.3442	0.424	0.4442	0.4871	0.4973	0.5023
R	0.0324	0.2642	0.5241	0.7149	0.8072	0.854	0.8458	0.8957	0.9027	0.9343	0.0839	0.3392	0.543	0.6275	0.6779	0.732	0.761	0.8114	0.8274	0.8495
mAP @0.5	0.0501	0.2665	0.4361	0.6565	0.7324	0.8081	0.8307	0.8592	0.8799	0.8892	0.1066	0.3565	0.5372	0.5978	0.6769	0.74	0.752	0.7881	0.8222	0.8591
mAP @.5:.95	0.0165	0.1061	0.1618	0.3255	0.3916	0.4338	0.4734	0.5358	0.562	0.5665	0.0343	0.1582	0.2721	0.3259	0.3862	0.4452	0.4638	0.486	0.5525	0.5654

Figure 9. Comparison of reverse filling attention mechanism (RFAM) and Step 2 detection results. mAP = mean Average Precision; P = Precision; R = Recall.

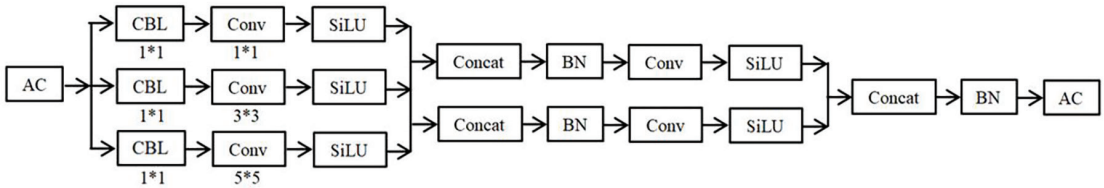


Figure 10. Multi-channel convolution (MCC) component structure. AC = Asymmetric Convolution; BN = BatchNorm2d function; CBL = Conv + BN + Leaky relu; Conv = convolution; SiLU = Sigmoid Linear Unit.

The performance of the current model is not optimal. As shown in Table 11, while RFAM performs better (P, R, mAP@0.5, and mAP@0.5:0.95) than other attention mechanisms in Table 6 (SE, CBAM, ECA, and CA), this does not exceed the Step 2 Baseline.

Figure 9 shows the comparison of the detection results between RFAM and Step 2 at different time nodes. According to the data, the detection results are counted at every 50 epochs, 10 times in total. The performance of RFAM shows a fast growth trend from 0 to 150 epochs, and all indicators are better than those in Step 2. However, this is gradually surpassed. A possible reason is that RFAM takes into account both channel and image scale. Therefore, compared with the single feature extraction method, multi-dimensional feature extraction can significantly accelerate the convergence speed, which is why RFAM is better than Step 2 in the early stage of detection process. However, in the YOLO V5s model, PAN and FPN play a decisive role in object detection performance. Because RFAM does not essentially optimize the model structure, it lacks the ability to extract features continuously in the later stage of the detection process, so that various indicators reach a bottleneck in poor performance.

Experimental data showed that using different convolution kernels to disassemble features can effectively improve the feature extraction ability. The MCC component with multiple convolution kernels can meet the requirements, proposed to connect the Neck and the Prediction. The Asymmetric Convolution Network (ACNet) was used to divide the input data into three channels and perform 1×1 CBL (the size of the convolution kernel in the CBL is 1×1), as shown in Figure 10. Further, 1×1, 3×3, and 5×5 convolution and Sigmoid Linear Unit (SiLU) activation were performed. Subsequently, the results of channels 1 and 2 were combined, with the same operation performed on channel 2 and channel 3. Moreover, BN + Conv + SiLU operations were conducted. This was followed by merging the results of the two channels and performing the BN + ACNet operations.

The meanings of portions are as follows:

(1) ACNet (Ding *et al.* 2019): As shown in Figure 11, the operation is based on the following principles: in the first case, the input data is convolved with two different kernels and the results are added together. In the second case, the two kernels are added before the input data is convolved. The result is the same in both cases. It can be expressed as Equation 13.

$$I * K^{(1)} + I * K^{(2)} = I * (K^{(1)} \oplus K^{(2)}) \quad (13)$$

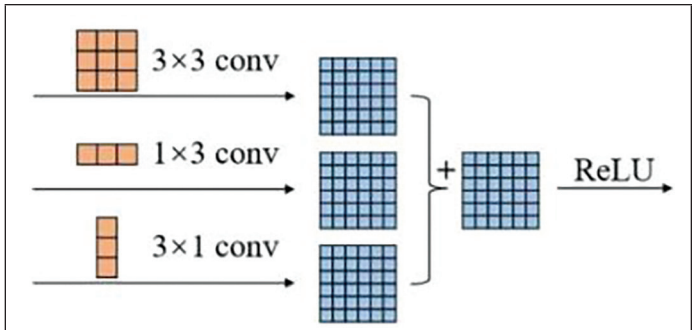


Figure 11. Asymmetric Convolution Network structure. ReLu = ReLu activation function.

The CBL component was replaced by DL (kernel = 3, stride = 2, padding = 2, dilation = 2) to reduce data loss in the down-sampling process. The convolution process is expressed as follows:

$$f = k + (k-1)(d-1) \quad (14)$$

$$\text{out} = (\text{in} - f + 2p + 1)/s \quad (15)$$

(2) SiLU: Torch.nn.SiLU() function. Apply the SiLU function. The SiLU function is also known as the swish function. , where is the logistic sigmoid.

Results and Discussion

Three improvements were achieved with the model. The running results and the performance enhancements of each part are as follows: COCO data set was used in the detection phase, and a custom remote sensing image data set was used in the verification phase.

Detection Phase

Enhanced PAN Structure

Table 12 presents the performance comparison of the enhanced PAN structure and the original model. The results show that with the increase of the depth of the improved model, the amount of computation and memory demand also increase, but the object detection ability of the model has been significantly improved. Increase in P value indicates increase in accuracy, and increase in mAP values indicate increase in object classification ability.

Table 12. Performance comparison of enhanced pyramid attention network (PAN) structure and original model.

	Memory (G)	Model Layers	Yaml Layers	Time/Image (ms)	Weight	P	R	mAP@0.5	mAP@.5:.95	GFLOPS (B)
Baseline	3.96	283	25	59.063	14.1M / 7 276 605	0.6852	0.9729	0.9664	0.797	17.1
+EPAN	4.69	292	28	83.602	15.3M / 7 518 909	0.7039	0.979	0.9716	0.8348	23.9

EPAN = enhanced PAN; GFLOPS = giga floating-point operations per second;mAP = mean Average Precision; P = Precision; R = Recall.

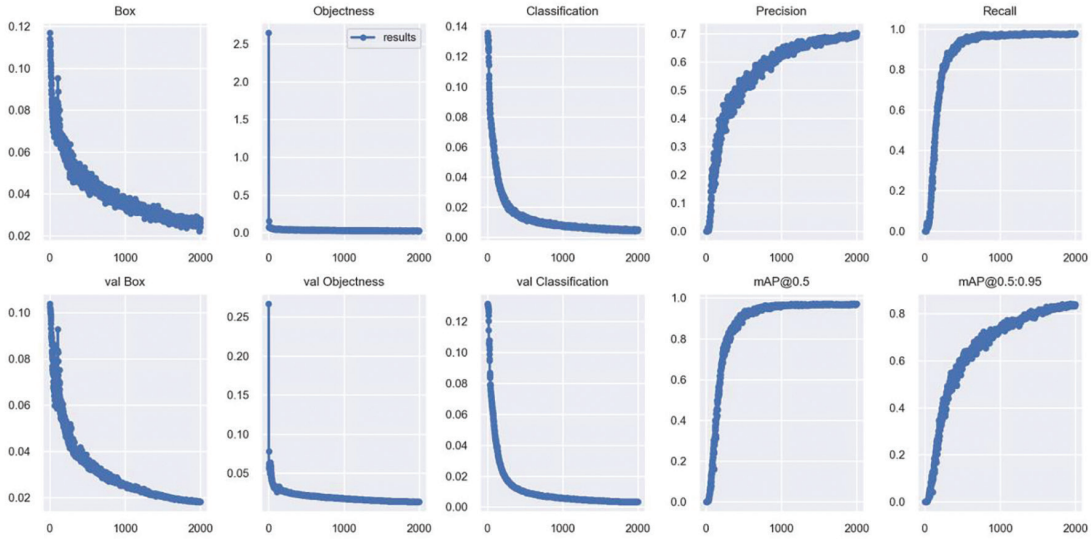


Figure 12. Detection results of enhanced pyramid attention network (PAN) structure. mAP = mean Average Precision.

Figure 12 shows the obtained results, where graphic title "Box" represents the bounding box loss; graphic title "Objectness" stands for the target detection loss, and graphic title "Classification" suggests the Classification loss. These three constitute loss value. Graphic title "val" represents validation set results. (X-axis represents the number of executions, and Y-axis represents the detection values. Since the figure is automatically generated by YOLO V5s model, the two axes cannot be named).

Table 13 shows the comparison of the loss function values between the improved model and the original model. The loss function of the model comprises three parts. Bounding box, represented by "Box", suggests the size and exact position of the target. Confidence, represented by "Objectness", is used to indicate the reliability of the bounding box. Moreover, Classification, represented by "Classification" variable, is used to signify the classification of the targets. The results show that the accuracy and confidence of the improved model are improved, but the classification ability is decreased, and the final loss value is decreased.

Table 13. Comparison of the loss function values between enhanced pyramid attention network (PAN) and the original model.

	Box	Objectness	Classification	Loss
Baseline	0.02941	0.03177	0.00495	0.06613
+EPAN	0.02651	0.02738	0.005304	0.05919

EPAN = enhanced PAN.

Results of the RIoU_Loss Regression Loss Function

Table 14 shows the comparison of the performance of the RIoU_Loss regression loss function and the original model. The results show that in addition to the increase of P value, the value of mAP@0.5:0.95 also increases significantly, which indicates that the improved model has better classification ability when multiple objects are superimposed in a region.

Detection results are shown in Figure 13.

Table 15 shows the comparison of the loss function values between the improved model and the original model. The results show that the feature localization error of the improved model is increased, but the object classification ability is improved.

Results of the CAM Component

Table 16 shows the comparison of the performance between the CAM component and the original model. The results show that all indicators of the model have been improved, and the improvement of accuracy

Table 14. Performance comparison of the RIoU_Loss regression loss function and original model.

	Memory (G)	Model Layers	Yaml Layers	Time/ Image(ms)	Weight	P	R	mAP@0.5	mAP@.5:.95	GFLOPS (B)
Baseline	3.96	283	25	59.063	14.1M / 7 276 605	0.6852	0.9729	0.9664	0.797	17.1
+EPAN	4.69	292	28	83.602	15.3M / 7 518 909	0.7039	0.979	0.9716	0.8348	23.9
+RIoU_Loss	4.69	292	28	91.927	15.3M / 7 518 909	0.7143	0.98	0.9746	0.8425	23.9

EPAN = enhanced PAN; GFLOPS = giga floating-point operations per second; mAP = mean Average Precision; P = Precision; R = Recall; RIoU = round intersection over union.

is obvious, but the cost is the surge of computation. On the other hand, the processing time of a single image increases significantly. This indicates that the structure of the model needs to be further improved. If the overall depth is not easy to simplify, the convolution kernel structure could be considered to be improved to make it more efficient in image processing.

Detection results are shown in Figure 14.

Table 15. Comparison of the loss function values between the round intersection over union (RIoU) and the original model.

	Box	Objectness	Classification	Loss
Baseline	0.02941	0.03177	0.00495	0.06613
+EPAN	0.02651	0.02738	0.005304	0.05919
+RIoU_Loss	0.03305	0.02782	0.005284	0.06616

EPAN = enhanced PAN; RIoU = round intersection over union.

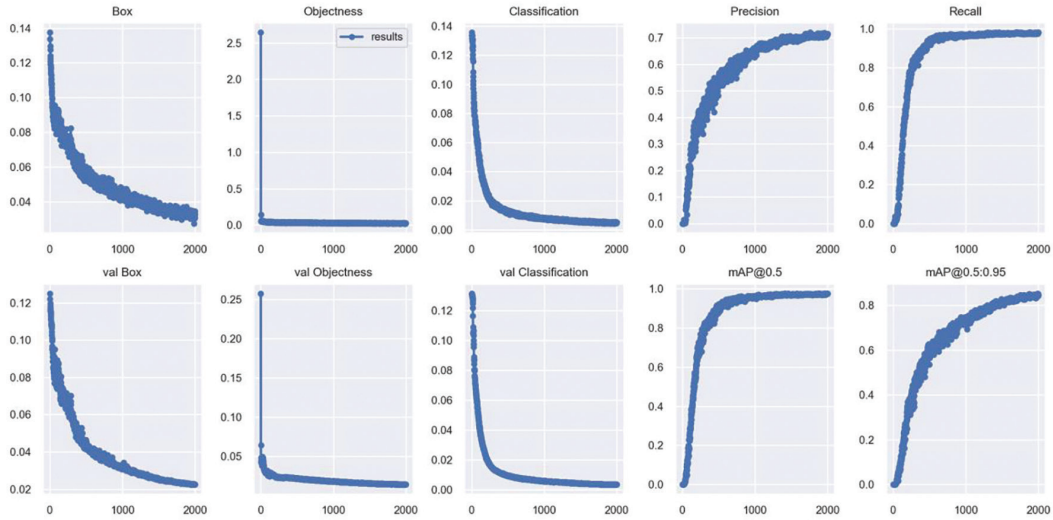


Figure 13. Detection results of the RIoU_Loss regression loss function. mAP = mean Average Precision; RIoU = round intersection over union.

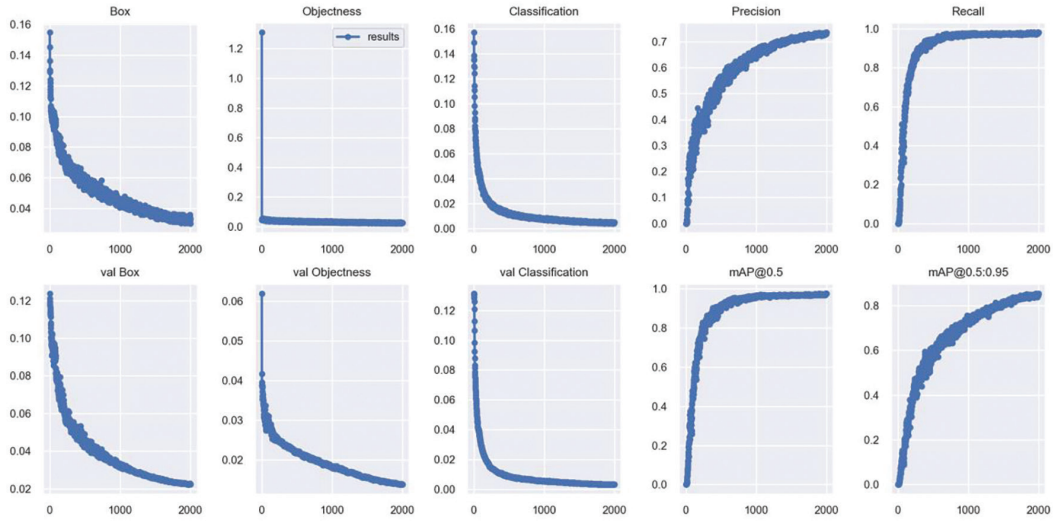


Figure 14. Detection results of the collaborative attention mechanism (CAM) component. mAP = mean Average Precision.

Table 16. Performance comparison of the collaborative attention mechanism (CAM) component and original model.

	Memory (G)	Model Layers	Yaml Layers	Time/Image(ms)	Weight	P	R	mAP@0.5	mAP@.5:.95	GFLOPS (B)
Baseline	3.96	283	25	59.063	14.1M / 7 276 605	0.6852	0.9729	0.9664	0.797	17.1
+EPAN	4.69	292	28	83.602	15.3M / 7 518 909	0.7039	0.979	0.9716	0.8348	23.9
+RIoU_Loss	4.69	292	28	91.927	15.3M / 7 518 909	0.7143	0.98	0.9746	0.8425	23.9
+CAM	4.68	461	31	218.756	76.7M / 38 139 645	0.7346	0.9802	0.9763	0.8541	77.1

CAM = collaborative attention mechanism; EPAN = enhanced PAN; GFLOPS = giga floating-point operations per second; mAP = mean Average Precision; P = Precision; R = Recall; RIoU = round intersection over union.

Table 17 shows the comparison of the loss function values between the improved model and the original model. The results show that the confidence and classification ability of the improved model are further improved, and the loss value decreases on the basis of the original model.

Verification Phase

Table 18 shows the comparison of the performance between the improved model and the original model using the custom remote sensing image data set. The results show that the performance of the improved model on the custom data set is better than that on the COCO data set. There are two possible reasons. First, the number of images in the custom data set is larger, so that the model can recognize more features in the training stage. Second, there is only one kind of remote sensing image in the customized data set, relatively speaking, within the limited training times, the model can identify features more intensively.

Verification results are shown in Figure 15.

Table 17. Comparison of the loss function values between the collaborative attention mechanism (CAM) component and original model.

	Box	Objectness	Classification	Loss
Baseline	0.02941	0.03177	0.00495	0.06613
+EPAN	0.02651	0.02738	0.005304	0.05919
+RIoU_Loss	0.03305	0.02782	0.005284	0.06616
+CAM	0.03334	0.02618	0.004939	0.06445

CAM = collaborative attention mechanism; EPAN = enhanced PAN; RIoU = round intersection over union.

Table 18. Performance comparison of the improved model and original model.

	Memory (G)	Model Layers	Yaml Layers	Time/Image(ms)	Weight	P	R	mAP@0.5	mAP@.5:.95	GFLOPS (B)
Baseline	4.1	283	25	67.612	14.4M / 7 085 118	0.9217	0.9753	0.9795	0.8433	16.5
improved	4.75	461	31	209.845	76.3M / 37 948 158	0.9677	0.9958	0.9927	0.951	76.5

GFLOPS = giga floating-point operations per second; mAP = mean Average Precision; P = Precision; R = Recall.

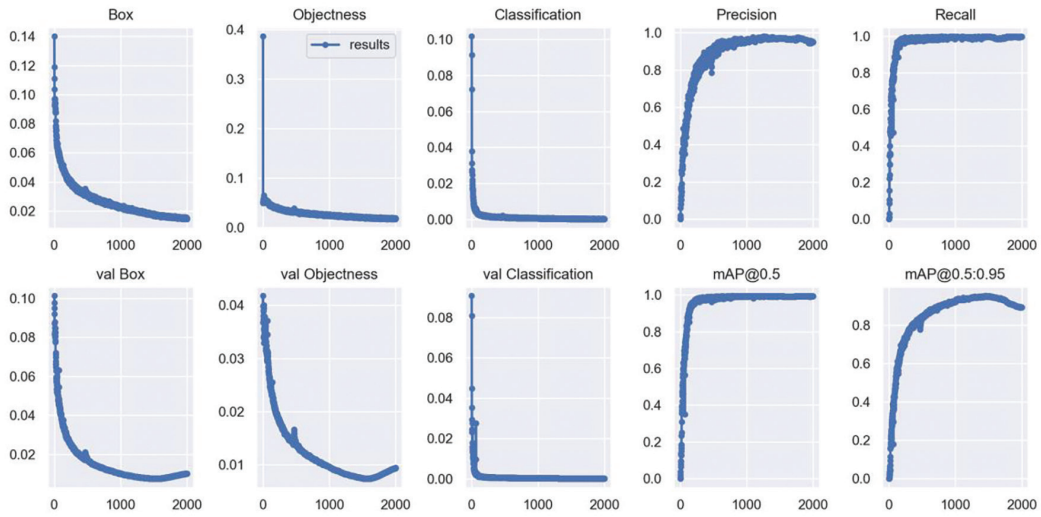


Figure 15. Detection results of the improved model based custom data set. mAP = mean Average Precision.

Table 19 shows the comparison of the loss function values between the improved model and the original model. The results show that the loss value of the improved model is smaller on the custom data set, and the overall loss value is better due to the obvious decrease of the Box value. This indicates that the feature localization ability of the model is better in the verification stage.

Table 19. Comparison of the loss function values between the improved model and original model.

	Box	Objectness	Classification	Loss
Baseline	0.02008	0.03071	0.000738	0.05153
improved	0.01835	0.0209	0.000381	0.03963

Detection Results

Table 20 shows the detection results of each representation object from 2014 to 2019, the poverty rates were obtained by consulting local government.

Table 21 shows the results based on a 95% regression analysis of each detection result.

The results of the analysis corroborate the effectiveness of the target detection results obtained by the improved YOLO V5s model. The

Table 20. Detection results for each representation object.

Year	Highway (m²)	Bungalow1 (pieces)	Bungalow2 (pieces)	Building (pieces)	Sports Ground (pieces)	Green House (pieces)	Terrace1 (m²)	Field (m²)	Terrace2 (m²)	Poverty Rate
2014	4717	11 569	3964	353	12	1148	1773	290	220	0.5437
2015	8171	17 948	4656	457	21	1377	2616	327	311	0.3004
2016	9717	16 463	5521	743	22	1659	3823	478	519	0.186
2017	15 074	18 241	5156	1122	25	2565	4587	529	673	0.1378
2018	16 812	19 462	5087	1192	25	2827	4614	555	731	0.0693
2019	16 881	25 890	4657	1988	30	3080	5291	550	855	0.0045

Table 21(a). Regression analysis of remote sensing image target detection.

Covariance Items	Values
Multiple R	0.999676
R Square	0.999352
Adjusted R Square	0.998272
Standard Error	368.8552

Table 21(b). Variance analysis of remote sensing image target detection.

Variance Items	df	SS	MS	F	Significance F
Regression Analysis	5	6.29E + 08	1.26E + 08	925.0926	5.6E-05
Residual	3	408 162.5	136 054.2		
Total	8	6.3E + 08			

df = degree freedom; F = F-hypothesis test; MS = mean square; SS = stdev square.

Table 22. Detection results for each representation object and poverty rate reference.

Year	Highway (m^2)	Bungalow1 (pieces)	Bungalow2 (pieces)	Building (pieces)	Sports Ground (pieces)	Green House (pieces)	Terrace1 (m^2)	Field (m^2)	Terrace2 (m^2)	Poverty Rate
2020	21 105	26 055	5279	2052	32	3601	6219	550	879	–0.1367

Significant F is much less than 0.05, which proves that the regression process is valid. Consequently, changes in representative objects in the region can be obtained over the years, as shown in Figure 16a–f (the labels BGY–XZ on the X-axis stand for the abbreviations of the county names; *see next page*).

The rate of poverty could reflects regional economic development. As shown in Table 22, the poverty rate of 2020 is unknown, which can be predicted by regression compared with the data in Table 20. According to the prediction result, the poverty rate in 2020 is less than 0, indicating that there is no poverty in this region, that is, the state of out of poverty has been achieved.

As shown in Table 23, the traditional statistical data are used to analyze the economic development of the region as an ancillary validation. According to the regional policy, these indicators are the standard for measuring poverty. The data from 2014 to 2019 are known and reflect the level of poverty in a certain field. Using regression to predict the data in 2020, it can be seen that the poverty degree in all fields are negative, indicating that the region has been completely lifted out of poverty, which is consistent with the prediction result in Table 22. Thus, in the area of regional poverty assessment, remote sensing image target detection has the same effect as traditional statistical methods.

Part of target detection results are shown in Figure 17 (where the "conf" and "IoU" parameters were set to 0.5 during model running), and they are: "highway", "flat-roofed bungalow", "peaked bungalow", "building", "greenhouse", "sports ground", "field", and "terrace". (All the detection results in this paper are based on google tile format with image size of 256, as a result, compared to the remote sensing images that cover the entire region, these images look smaller.)

Table 23. Poverty rate analysis of the traditional statistical data.

Year	Food	Education	Houses	Transportation
2014	2.7953	2.2208	2.8920	2.1978
2015	1.5799	1.2442	1.5229	1.2422
2016	1.1272	0.9437	1.2436	1.0093
2017	0.8692	0.7763	0.8194	0.8608
2018	0.4920	0.4587	0.4796	0.4720
2019	0.0346	0.0327	0.0321	0.0317
2020	–1.1821	–0.2223	–0.5615	–0.2696

Table 24. Comparison of the experimental results.

Papers	P	R	mAP @0.5	mAP @.5:.95	Loss
Shao *et al.* 2021	0.952	0.931	0.949	0.509	/
Xie *et al.* 2022	0.956	0.956	0.978	0.658	/
Kim *et al.* 2021	0.729	0.493	0.476	0.176	/
This paper	0.9677	0.9958	0.9927	0.951	0.0396

mAP = mean Average Precision; P = Precision; R = Recall.

Detection Capability

COCO data set was used in detection stage. Improvements to the model increased 7.3% in the value of P from 0.6852 to 0.7346, 0.7% in the value of R from 0.9729 to 0.9802, 1% in the value of mAP@0.5 from 0.9664 to 0.9763, and 7.2% in the value of mAP@0.5:0.95 from 0.797 to 0.8541, respectively. In the verification stage, the custom data set (842 304 pieces of remote sensing images were downloaded and labeled) was used. The corresponding four values increased 3.1% from 0.9217 to 0.9507, 2.2% from 0.9753 to 0.9958, 1.3% from 0.9795 to 0.9931, and 5.7% from 0.8433 to 0.8909, respectively. The loss value decreased 37.4% from 0.0515 to 0.0322.

Comparison of the Experimental Results

Experimental results were compared with similar models to verify the performance of the improved model. A study (Shao *et al.* 2021) proposed a Tiny Adaptively Spatial Feature Fusion YOLOv5 algorithm using fishing boats in night remote sensing images as a custom data set. Moreover, a small target recognition layer and an adaptive spatial feature fusion network were added. Another study (Xie *et al.* 2022) proposed a lightweight end-to-end target detection framework by introducing an attention mechanism into the YOLO V5 model, using SAR images as a custom data set. In Kim *et al.* (2021), aerial vehicle remote sensing images were used as custom data set, and the Efficient Channel Attention Pyramid YOLO model was proposed by improving the channel attention module of YOLO V5 model. The new detection layer was used to replace the large target detection module, and the transposed convolution was used to replace up-sampling. The comparison results are shown in Table 24. According to the data, the improved model in this paper can obtain comprehensive evaluation indicators and offer better performance.

Discussion

(1) Enhanced PAN Structure: As shown in Figure 1 and Table 4, the original YOLO V5s model consists of PAN + FPN + PAN structure. The first PAN is a down-sampling process where the image size converts from 128 to 8, whereas the number of channels increases from 32 to 512. The FPN is an up-sampling process where the image size converts from 8 to 32, but the number of channels reduces from 512 to 128. The second PAN is also a down-sampling process, where the image size converts from 32 to 8; however, the number of channels increases from 128 to 512. The principle is as follows: when the image size reduces, it is helpful to feature localization, whereas part of the informations lost, thus the number of channels needs to be increased as compensation; on the contrary, when the image size enlarges, the informations increase, however the computation will also be increased accordingly, thus the number of channels should be reduced as mitigation. The following methods have been tried to improve the structure of the model: (i) The CSPs were used as much as possible in both PANs, since each CSP contains three convolution operations, which can improve the feature extraction ability by increasing the depth of the model. However, the experimental data showed that stacking a large number of CSPs not only fail to extract features, but also increase the running time. (ii) The depth of FPN was decreased,

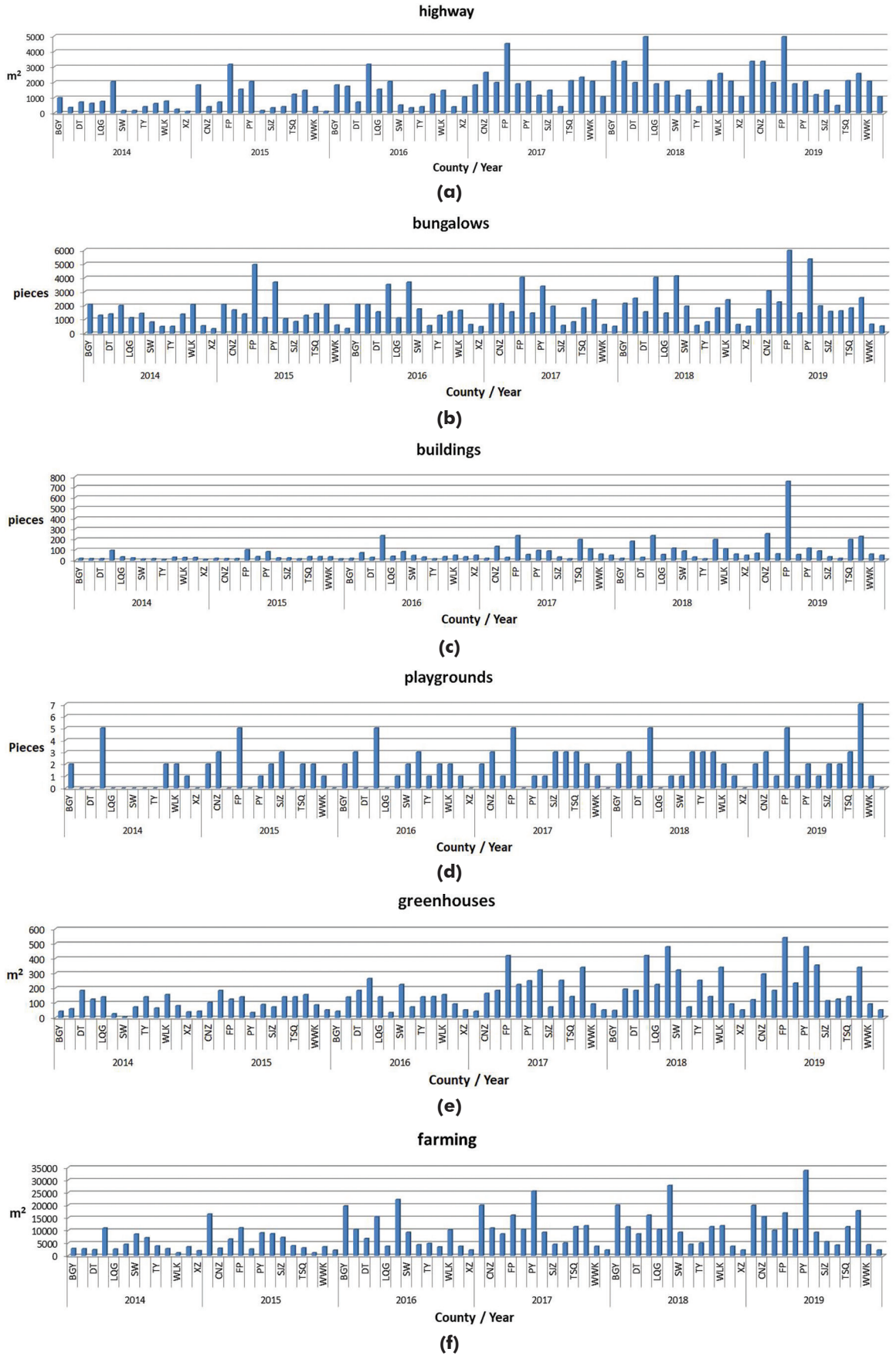


Figure 16. (a) Changes in the number of highways in the region. (b) Changes in the number of bungalows in the region. (c) Changes in the number of buildings in the region. (d) Changes in the number of playgrounds in the region. (e) Changes in the number of greenhouses in the region. (f) Changes in the extent of farming in the region. The labels BGY–XZ on the X-axis stand for the abbreviations of the county names.

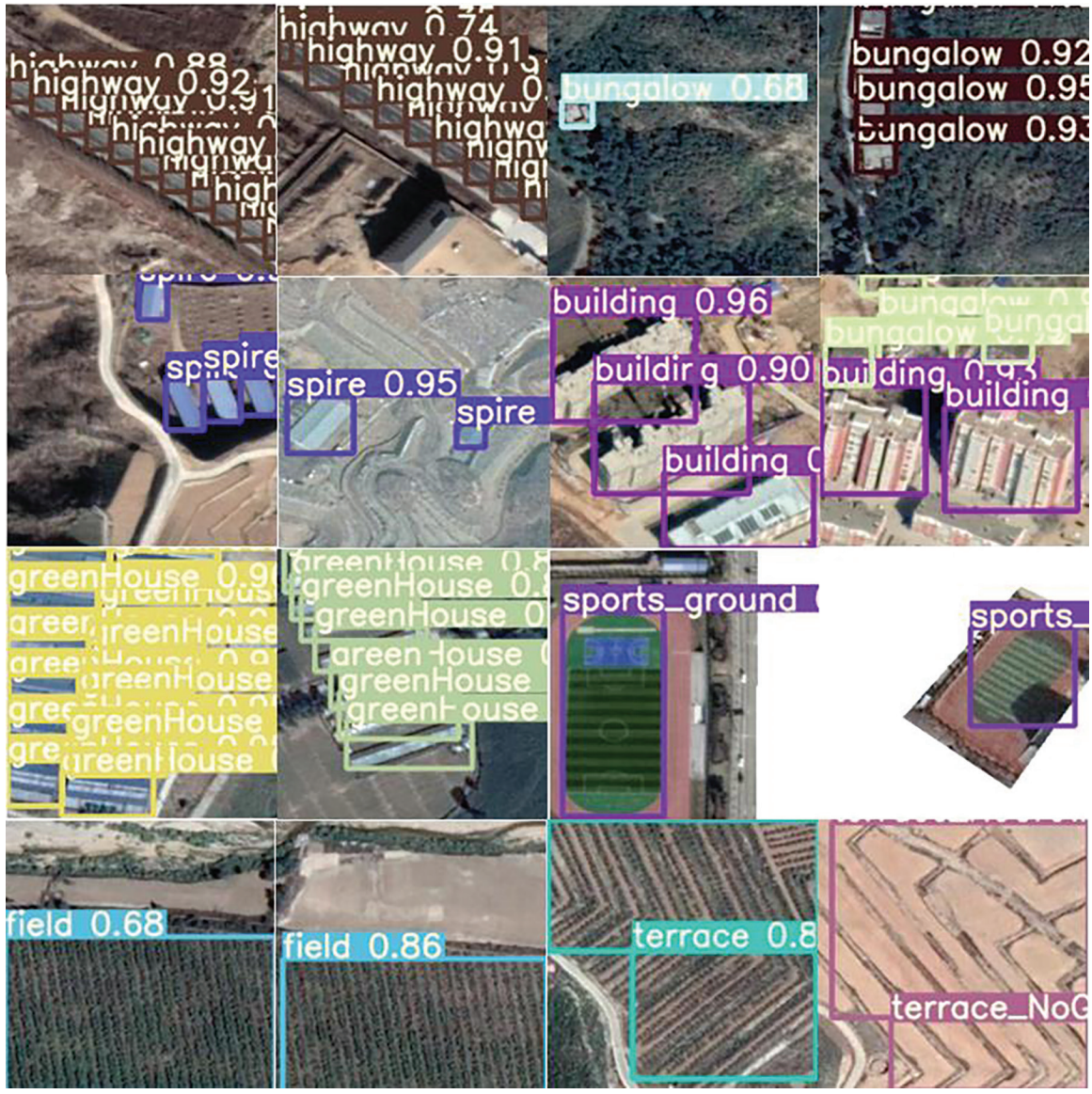


Figure 17. Remote sensing image target detection results.

thereby reducing the computation and parameters. However, the loss value was high, which indicated that there was a large error in the calculation process. According to the definition, PAN is used for feature localization, while FPN is used for semantic recognition. Therefore, the improved scheme was proposed as shown in Figure 4 and Table 5. (i) The depth of FPN was appropriately reduced, removing one CSP. (ii) The image size at the end of FPN was increased from 32 to 128, as the beginning of the second PAN. (iii) Part of CSPs were replaced by CBLs in the second PAN, although both have the same effect, the latter required much less computation than the former. (iv) The new network layer (L18) in the second PAN was fused with the first PAN (L3), taking into account the original characteristics of the data. Thus, the model formed two symmetric PANs, all of them converted the image size from 128 to 8. It was not recommended that the image size be smaller than 8, because too small size would lead to the loss of informations and the anomalous extraction of features. According to Table 12 and Table 13, the detection performance improved, and the loss value decreased, indicating that the improved structure was effective.

(2) RIoU_Loss Regression Loss Function: The role of IoU is to determine the differences between the GT box and the Pre box, so as to obtain the model error. Where GT is the true location of the object and Pre is the predicted location. Up to now, there are Generalized IoU (GIoU), Distance IoU (DIoU), and CIoU, and they're all based on the positional relationship between the two boxes. Take CIoU as an example, which is currently the latest and most commonly used way. It considers the factors of overlap area, central point distance, and aspect ratio; however, when the aspect ratio of two boxes is equal, there would be the phenomenon in Equation 4; that is, the values of CIoU are equal and cannot be differentiated. The new RIoU_Loss regression loss function transforms the positional problem into an angle problem, which is equivalent to fixing the position of GT firstly, and then treating the position of Pre as a variable, then the moving trajectory of Pre is round (Figure 6). Therefore, RIoU is more comprehensive than CIoU, because any kind of positional relationship between the two boxes can be represented by an angle on the circle.

(3) CAM component: Simply increasing the number of convolutions or the depth of the model cannot improve the feature extraction ability, thus the connections between PAN and FPN can be considered. Because the two structures have opposite functions such as down-sampling and up-sampling, a more appropriate join can optimize features and reduce error. The new CAM component consists of the RFAM and MCC components. RFAM connects the first PAN with FPN, which combines attention mechanism for feature

extraction from two dimensions: channel and image size. The advantage of attention mechanism is that it can make feature extraction targeted, whereas the disadvantage is that it needs to calculate all channels and pixels on each image. The traditional calculation methods use Sigmoid function to generate coefficients, since the function is nonlinear, the changes of the output results are not obvious when they approach the extreme values. Thus, the new reverse filling mechanism was proposed, which replaced the Sigmoid function and identified features by increasing differences between the input data and the results in the channel and scale dimensions. The MCC component makes use of the multi-channel technique and connects the FPN with the second PAN. It can extract different types of features through different convolution kernels. Experimental data showed that the number of parallel channels should not exceed three, otherwise memory overflow would occur. ACNet can maintain the robustness of the model; therefore, the first one (Figure 10) is used to separate different features into respective channels, while the last one is used to transmit the extracted features to subsequent components stably.

(4) Limitation of the Improved Model: Based on the original model, the improved model introduced the attention mechanism and the multi-channel technology, which lead to a surge in the number of parameters and computation time. In the next research phase, using the principle of wavelet transform is planned for feature extraction and to implement it by convolution, and the attention mechanism would be replaced. On the other hand, the convolution kernel may be improved to solve the problems of parameter number, computation and loss value caused by model depth.

(5) Ablation Experiment: The detection performance of each component and combinations are shown in Table 25. The experimental data showed that adding three modules respectively on the basis of Baseline can improve the values of P, mAP@0.5, and mAP@0.5:0.95 to different degrees, and the loss values decrease. By arranging and combining modules, the R values of the model can be significantly improved. Finally, the combination of all modules made the performance of the model reach equilibrium. The results showed that the single performance of RIoU_Loss is good, indicating that the design of the structures of the other two modules need to be further optimized, and the conflict between the model depth and the loss value also need to be further alleviated.

Conclusion

To verify the effectiveness of China's poverty alleviation efforts, an improved YOLO V5s model was proposed for regional poverty assessment using remote sensing image target detection. The previous poverty alleviation work was done through manpower, which was easily affected by human factors. In this paper, target detection results of remote sensing images were used for auxiliary verification so as to make the poverty alleviation process more objective. Objects in images were detected to extract features, and their changes were used to evaluate the poverty rate in successive years; subsequently, the poverty alleviation situation could be predicted according to the existing detection results.

The improvements of the model included three aspects. First, the enhanced PAN structure was proposed. The depth of FPN was shorted whereas the second PAN was increased, thereby increasing the overall depth of the model and improving the feature extraction ability. Second, the RIoU_Loss regression loss function was proposed. The position relationship between the GT box and the Pre box was transformed into the angle relationship, so that the decision factors were more comprehensive. Third, new component of the collaborative attention mechanism was proposed. Firstly, at the connection between the Backbone and the Neck part, a new RFAM component was proposed by combining the attention mechanism, and the original CBL component was replaced. This component replaced the Sigmoid function in the attention mechanism with a new reverse filling technique, which increased the difference between each pixel in the image and made the features easier to identify. Secondly, at the junction of the Neck and the Prediction part, the new MCC component was proposed by combining the multi-channel convolution technique. This component used different convolution kernels to split different features. ACNet and DC technology were introduced to reduce the data loss on the basis of ensuring the robustness of the model.

Experimental data showed that compared with the original model and similar papers, the improved model proposed in this paper has better performances on the key evaluation indicators (P, R, mAP@0.5, mAP@0.5:0.95, and Loss), which proved that it had better target detection ability. The detection results were applied to regional poverty assessment and compared with the traditional statistical methods to prove that they have the same effect. It can be seen that remote sensing image target detection has advantages of cheap, efficient, accurate and objective, which provides a promising approach for regional poverty assessment.

Table 25. Performance comparison of the improved model and original model.

Baseline	EPAN	RIoU_Loss	CAM	P	R	mAP@0.5	mAP@.5:.95	Loss
√				0.6852	0.9729	0.9664	0.797	0.0661
	√			0.7039	0.979	0.9716	0.8348	0.0592
		√		0.7632	0.9774	0.9738	0.8922	0.0504
			√	0.7409	0.9774	0.9719	0.8672	0.05
	√	√		0.7143	0.98	0.9746	0.8425	0.0662
	√		√	0.715	0.981	0.9729	0.8418	0.0582
		√	√	0.7434	0.978	0.9722	0.8563	0.0612
	√	√	√	0.7346	0.9802	0.9763	0.8541	0.0645

CAM = collaborative attention mechanism; EPAN = enhanced PAN; mAP = mean Average Precision; P = Precision; R = Recall; RIoU = round intersection over union.

Acknowledgments

This research was funded by Cangzhou Science and Technology Bureau project "research planning of regional economic development based on big data recommendation algorithm and remote sensing", grant number 213102003; This research was funded by Cangzhou Science and Technology Bureau project "Research on air quality prediction based on big data analysis", grant number 213102001; This research was funded by Human Resources and Social Security Department of Hebei Province project "Research on the training of application-oriented big data talents in the construction of new engineering for the job market", grant number JRSHZ-2022-02037.

References

Bitar, A. W., L. F. Cheong and J. P. Ovarlez. 2019. Sparse and low-rank matrix decomposition for automatic target detection in hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing* 57(8):5239–5251.

Ding, X., Y. Guo, G. Ding and J. Han. 2019. ACNet: Strengthening the kernel skeletons for powerful CNN via asymmetric convolution blocks. Pages 1911–1920 in *Proceedings 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, held in Seoul, South Korea, 27 October–2 November 2019. https://doi.org/10.1109/ICCV.2019.00200.

Girshick, R. 2015. Fast R-CNN. Pages 1440–1448 in *Proceedings 2015 IEEE International Conference on Computer Vision (ICCV)*, held in Santiago, Chile, 7–13 December 2015. https://doi.org/10.1109/ICCV.2015.169.

Hou, Q., D. Zhou and J. Feng. 2021. Coordinate attention for efficient mobile network design. Pages 13708–13717 in *Proceedings IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, held in Nashville, Tenn., 20–25 June 2021. https://doi.org/10.1109/CVPR46437.2021.01350.

Hu, J., L. Shen and G. Sun. 2018. Squeeze-and-excitation networks. Pages 7132–7141 in *Proceedings 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, held in Salt Lake City, Utah. https://doi.org/10.1109/CVPR.2018.00745.

Ji, H., Z. Gao, T. Mei and Y. Li. 2019. Improved faster R-CNN with multiscale feature fusion and homography augmentation for vehicle detection in remote sensing images. *IEEE Geoscience and Remote Sensing Letters* 16(11):1761–1765.

Kim, M., J. Jeong and S. Kim. 2021. ECAP-YOLO: Efficient channel attention pyramid YOLO for small object detection in aerial image. *Remote Sensing* 13(23):4851.

Li, Y., H. Zhang, X. Xue, Y. Jiang and Q. Shen. 2018. Deep learning for remote sensing image classification: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 8(6):e1264.

Liu, J. and Y. Qiao. 2020. Mahalanobis distance–based kernel supervised machine learning in spectral dimensionality reduction for hyperspectral imaging remote sensing. *International Journal of Distributed Sensor Networks* 16(11):1550147720968467.

Luo, S., J. Yu, Y. Xi and X. Liao. 2022. Aircraft target detection in remote sensing images based on improved YOLOv5. *IEEE Access* 10:5184–5192.

Mei, S., R. Jiang, X. Li and Q. Du. 2020. Spatial and spectral joint super-resolution using convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing* 58(7):4590–4603.

Peng, M., L. Zhang, X. Sun, Y. Cen and X. Zhao. 2020. A fast three-dimensional convolutional neural network-based spatiotemporal fusion method (STF3DCNN) using a spatial-temporal-spectral dataset. *Remote Sensing* 12(23):3888.

Qu, J., C. Su, Z. Zhang and A. Razi. 2020. Dilated convolution and feature fusion SSD network for small object detection in remote sensing images. *IEEE Access* 8:82832–82843.

Shao, J., Q. Yang, C. Luo, R. Li, Y. Zhou and F. Zhang. 2021. Vessel detection from nighttime remote sensing imagery based on deep learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14:12536–12544.

Tang, G., Y. Zhuge, C. Claramunt and S. Men. 2021. N-Yolo: A SAR ship detection using noise-classifying and complete-target extraction. *Remote Sensing* 13(5):871.

Wang, Q., B. Wu, P. Zhu, P. Li, W. Zuo and Q. Hu. 2020. ECA-Net: Efficient channel attention for deep convolutional neural networks. Pages 11531–11539 in *Proceedings 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, held in Seattle, Wash.

https://doi.org/10.1109/CVPR42600.2020.01155.

Woo, S., J. Park, J. Y. Lee and I. S. Kweon. 2018. Cbam: Convolutional block attention module. *Lecture Notes in Computer Science*. https://doi.org/10.1007/978-3-030-01234-2_1.

Xie, F., B. Lin and Y. Liu. 2022. Research on the coordinate attention mechanism fuse in a YOLOv5 deep learning detector for the SAR ship detection task. *Sensors* 22(9):3370.

Xu, D. and Y. Wu. 2020. Improved YOLO-V3 with DenseNet for multi-scale remote sensing target detection. *Sensors* 20(15):4276.

Yan, J., H. Wang, M. Yan, W. Diao, X. Sun and H. Li. 2019. IoU-adaptive deformable R-CNN: Make full use of IoU for multi-class object detection in remote sensing imagery. *Remote Sensing* 11(3):286.

Yuan, Q., H. Shen, T. Li, Z. Li, S. Li, Y. Jiang, H. Xu, W. Tan, Q. Yang, J. Wang, J. Gao and L. Zhang. 2020. Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sensing of Environment* 241:111716.

Zhang, L., J. Wu, Y. Fan, H. Gao and Y. Shao. 2020. An efficient building extraction method from high spatial resolution remote sensing images based on improved mask R-CNN. *Sensors* 20(5):1465.

Zheng, Y., C. Dong-Ye, P. Wang and M. Yasir. 2022. Improved YOLOv5 network method for remote sensing image based ground objects recognition. *Soft Computing* 26(20):10879–10889.

Zheng, Z., P. Wang, D. Ren, W. Liu, R. Ye, Q. Hu and W. Zuo. 2021. Enhancing geometric factors in model learning and inference for object detection and instance segmentation. *IEEE Transactions on Cybernetics* 52(8):8574–8586.

Zhou, K., M. Zhang, H. Wang and J. Tan. 2022. Ship detection in SAR images based on multi-scale feature extraction and adaptive feature fusion. *Remote Sensing* 14(3):755.

Appendix A

```
# -*- coding: utf-8 -*-
import os.path
import xml.etree.ElementTree as ET
class_names = [‘class names’]
xmlpath = 'path of XML file'
txtpath = 'path of TXT file'
files=[]
for root, dirs, files in os.walk(xmlpath):
    None
number = len(files)
i = 0
while i < number:
        name = files[i][0:-4]
        xml_name = name + ".xml"
        txt_name = name + ".txt"
        xml_file_name = xmlpath + xml_name
        txt_file_name = txtpath + txt_name
        xml_file = open(xml_file_name)
        tree = ET.parse(xml_file)
        root = tree.getroot()
        filename = root.find('filename').text
        image_name = root.find('filename').text
        w = int(root.find('size').find('width').text)
        h = int(root.find('size').find('height').text)
        f_txt = open(txt_file_name, 'w+')
        content = ""
        first = True
        for obj in root.iter('object'):
        name = obj.find('name').text
        class_num = class_names.index(name)
        xmlbox = obj.find('bndbox')
        x1 = int(xmlbox.find('xmin').text)
        x2 = int(xmlbox.find('xmax').text)
        y1 = int(xmlbox.find('ymin').text)
        y2 = int(xmlbox.find('ymax').text)
        if first:
            content += str(class_num) + " " + \
                str((x1+x2)/2/w) + " " + str((y1+y2)/2/h) + " " + \
                str((x2-x1)/w) + " " + str((y2-y1)/h)
            first=False
        else:
            content += "\n" + \
                str(class_num) + " " + \
                str((x1 + x2) / 2 / w) + " " + str((y1 + y2) / 2 / h) + " " + \
                str((x2 - x1) / w) + " " + str((y2 - y1) / h)
    print(str(i/(number - 1) * 100) + "%\n")
    f_txt.write(content)
    f_txt.close()
    xml_file.close()
    i += 1
print("done!")
```

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

Special Advertising Opportunities

FRONT COVER SPONSORSHIP

A *PE&RS* cover sponsorship is a unique opportunity to capture the undivided attention of your target market through three premium points of contact.

1— *PE&RS* FRONT COVER

(Only twelve available, first-come, first-served)
PE&RS is world-renowned for the outstanding imagery displayed monthly on its front cover—and readers have told us they eagerly anticipate every issue. This is a premium opportunity for any company, government agency, university or non-profit organization to provide a strong image that demonstrates their expertise in the geospatial information industry.

2— FREE ACCOMPANYING "HIGHLIGHT" ARTICLE

A detailed article to enhance your cover image is welcome but not a condition of placing an image. Many readers have asked for more information about the covers and your article is a highly visible way to tell your story in more depth for an audience keenly interested in your products and services. No article is guaranteed publication, as it must pass ASPRS editorial review. For more information, contact Rae Kelley at rkelley@asprs.org.

3— FREE TABLE OF CONTENTS COVER DESCRIPTION

Use this highly visible position to showcase your organization by featuring highlights of the technology used in capturing the front cover imagery. Limit 200-word description.

Terms: Fifty percent nonrefundable deposit with space reservation and payment of balance on or before materials closing deadline.

Cover Specifications: Bleed size: 8 5/8" × 11 1/4", Trim: 8 3/8" × 10 7/8"

PRICING

	Sustaining Member Exhibiting at a 2023 ASPRS Conference	Sustaining Member	Exhibitor	Non Member
Cover 1	$1,850	$2,000	$2,350	$2,500

Belly Bands, Inserts, Outserts & More!

Make your material the first impression readers have when they get their copy of *PE&RS*. Contact Bill Spilman at bill@innovativemediasolutions.com

VENDOR SEMINARS

ASPRS Sustaining Members now have the opportunity to hold a 1-hour informational session as a Virtual Vendor Seminar that will be free to all ASPRS Members wishing to attend. There will be one opportunity per month to reach out to all ASPRS Members with a demonstration of a new product, service, or other information. ASPRS will promote the Seminar through a blast email to all members, a notice on the ASPRS web site home page, and ads in the print and digital editions of *PE&RS*.

The Virtual Seminar will be hosted by ASPRS through its Zoom capability and has the capacity to accommodate 500 attendees.

Vendor Seminars	
Fee	$2,500 (no discounts)

DIGITAL ADVERTISING OPPORTUNITIES

EMPLOYMENT PROMOTION

When you need to fill a position right away, use this direct, right-to-the-desktop approach to announce your employment opportunity. The employment opportunity will be sent once to all ASPRS members in our regular Wednesday email newsletter to members, and will be posted on the ASPRS Web site for one month. This type of advertising gets results when you provide a web link with your text.

Employment Opportunity	Net Rate
30-Day Web + 1 email	$500/opportunity
Web-only (no email)	$300/opportunity

Do you have multiple vacancies that need to be filled? Contact us for pricing details for multiple listings.

NEWSLETTER DISPLAY ADVERTISING

Your vertical ad will show up in the right hand column of our weekly newsletter, which is sent to more than 3,000 people, including our membership and interested parties. **Open Rate: 32.9%**

Newsletter vertical banner ad	Net Rate
180 pixels x 240 pixels max	$500/opportunity

DEDICATED CONTENT EMAIL BLAST

Send a dedicated email blast to the ASPRS email list. Advertiser supplies HTML (including images). Lead time: 14 days.

Materials	Net Rate
Advertiser supplies HTML, including images.	$3000/ opportunity

PE&RS Digital Edition

Digital Edition Announcement E-Mail: 5,800+

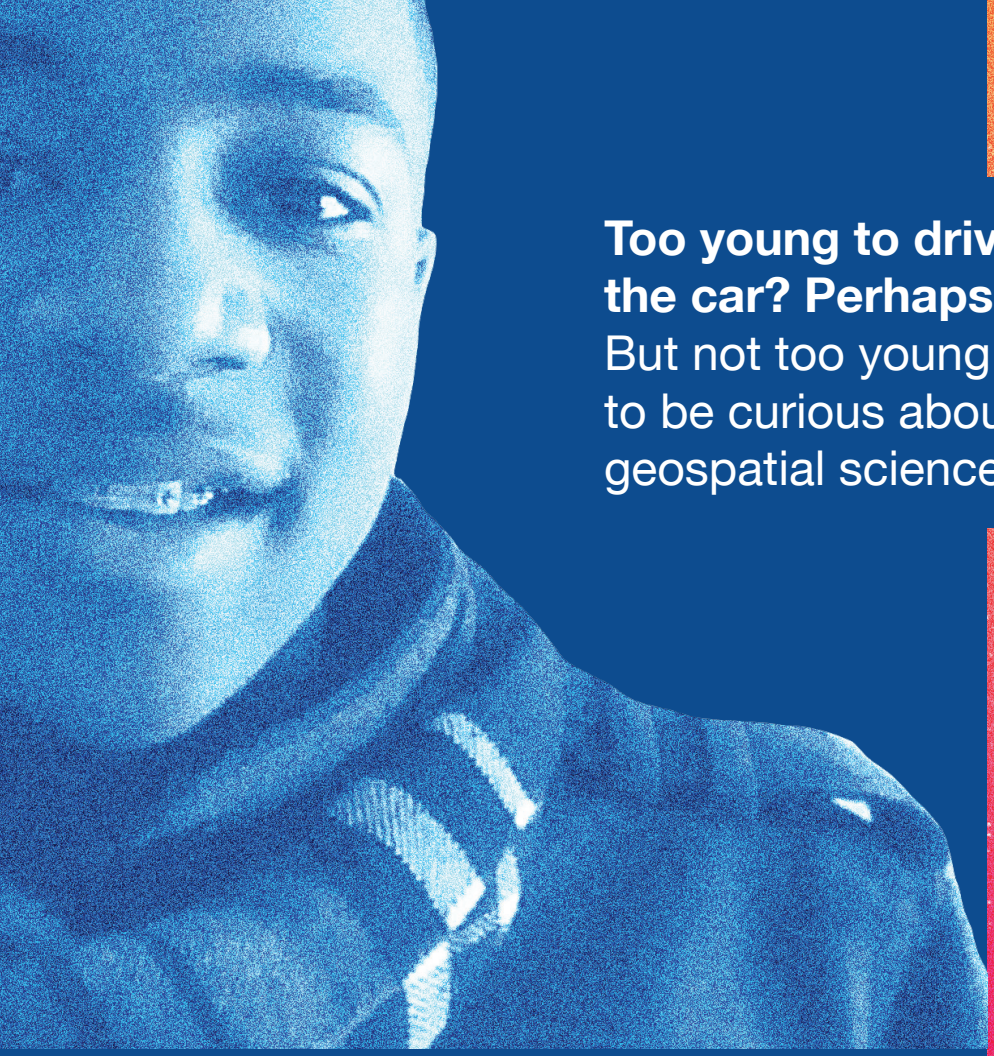
PE&RS is available online in both a public version that is available to anyone but does not include the peer-reviewed articles, and a full version that is available to ASPRS members only upon login.

The enhanced version of *PE&RS* contains hot links for all ASPRS Sustaining Member Companies, as well as hot links on advertisements, ASPRS Who's Who, and internet references.

Become a sponsor today!

The e-mail blast sponsorship opportunity includes a **180 x 240 pixel ad** in the email announcement that goes out to our membership announcing the availability of the electronic issue.

Digital Edition Opportunities	Net Rate
E-mail Blast Sponsorship*	$1,000



Too young to drive the car? Perhaps!
But not too young to be curious about geospatial sciences.

asprs FOUNDATION

The ASPRS Foundation was established to advance the understanding and use of spatial data for the betterment of humankind.

The Foundation provides grants, scholarships, loans and other forms of aid to individuals or organizations pursuing knowledge of imaging and geospatial information science and technology, and their applications across the scientific, governmental, and commercial sectors.

Support the foundation, so when they are ready, we are too.

asprsfoundation.org/donate