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.
ANNOUNCEMENTS

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

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

---

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

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

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

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

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

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

---

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

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

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

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

---

To have your press release published in *PE&RS*, contact Rae Kelley, rkelley@asprs.org.
all City departments for informal approvals as the design progressed. Construction began in January 2021. The project was substantially completed in December 2021 below the construction budget determined by Halff.

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

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

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

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

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

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

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

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

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

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

---

**CALENDAR**

- **15-17** February, [ASPRS Annual Conference at Geo Week](https://my.asprs.org/2023conference), Denver, Colorado. For more information, visit https://my.asprs.org/2023conference.

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

**79 UAS Edge Computing of Energy Infrastructure Damage Assessment**  
Jordan Bowman, Lexie Yang, Ornir Thomas, Jerry Kirk, Andrew Duncan, David Hughes, and Shannon Meade

Energy infrastructure assessments are needed within 72 hours of natural disasters, and previous data collection methods have proven too slow. This article demonstrates a scalable end-to-end solution using a prototype unmanned aerial system that performs on-the-edge detection, classification (i.e., damaged or undamaged), and geo-location of utility poles.

**89 Apricot Tree Detection from UAV-Images Using Mask R-CNN and U-Net**  
Firat Erdem, Nuri Erkin Ocer, Dilek Kucuk Matci, Gordana Kaplan, and Ugur Avdan

Monitoring trees is necessary to manage and take inventory of forests, monitor plants in urban areas, distribute vegetation, monitor change, and establish sensitive and renewable agricultural systems. This article aims to automatically detect, count, and map apricot trees in an orthophoto, covering an area of approximately 48 ha on the ground surface using two different algorithms based on deep learning.

**97 Comparative Analysis of Different CNN Models for Building Segmentation from Satellite and UAV Images**  
Batuhan Sariturk, Damla Kumbasar, and Dursun Zafer Seker

Building segmentation has numerous application areas such as urban planning and disaster management. In this article, 12 CNN models (U-Net, FPN, and LinkNet using EfficientNet-B5 backbone, U-Net, SegNet, FCN, and six Residual U-Net models) were generated and used for building segmentation. Inria Aerial Image Labeling Data Set was used to train models, and three data sets (Inria Aerial Image Labeling Data Set, Massachusetts Buildings Data Set, and Syedra Archaeological Site Data Set) were used to evaluate trained models.

**107 Unmanned Aerial Vehicle (UAV)–Based Imaging Spectroscopy for Predicting Wheat Leaf Nitrogen**  
Rabi N. Sahoo, Shalini Gakhar, R.G. Rejith, Rajeev Ranjan, Mahesh C. Meena, Abir Dey, Joydeep Mukherjee, Rajkumar Dhakar, Sunny Arya, Anchal Daas, Subhash Babu, Pravin K. Upadhyay, Kapila Sekhawat, Sudhir Kumar, Mahesh Kumar, Viswanathan Chinnusamy, and Manoj Khanna

Quantitative estimation of crop nitrogen is the key to site-specific management for enhanced nitrogen (N) use efficiency and a sustainable crop production system. This article attempts to predict leaf N of wheat crop through spectroscopy using a field portable spectroradiometer (spectral range of 400–2500 nm) on the ground in the crop field and an imaging spectrometer (spectral range of 400–1000 nm) from an unmanned aerial vehicle (UAV) with the objectives to evaluate four multivariate spectral models and two sets of hyperspectral data collected from two platforms and two different sensors.

**117 Car Detection from Very High-Resolution UAV Images Using Deep Learning Algorithms**  
Yunus Kaya, Halil Ibrahim Şenol, Abdurahman Yasım Yigit, and Murat Yaker

It is important to determine car density in parking lots, especially in hospitals, large enterprises, and residential areas, which are used intensively, in terms of executing existing management systems and making precise plans for the future. In this article, cars in parking lots were detected using high-resolution unmanned aerial vehicle (UAV) images with deep learning methods.
Exploding white dwarf stars are thought to be the ultimate source of much of our solar system’s lithium. But on Earth, there are certain environments where the soft, light, silvery-white metal is most concentrated and easily mined—notably in briny groundwater aquifers found beneath desert salt flats.

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

These conditions abound in the Atacama Desert in South America, where some of the world’s largest lithium deposits and mining operations are located. But they can also be found in Nevada’s Clayton Valley—the site of the only active lithium mine in the United States. The town of Silver Peak was established in the 1860s around gold and silver mines, but since the 1980s the town has focused on harvesting lithium, which is concentrated under the Clayton Valley in an area where an extinct volcano left lithium-rich deposits.

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

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

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

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

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

When Time-of-Day Matters; Visualizing DEMs

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

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

**Tip #1 — Building Raster Pyramids**

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

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

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

The default color scheme ramp for a DTM, or any DEM for that matter, is a the black (low elevation) to-white (higher elevation) color ramp as in Figure 2 below. This default ramp, while representative of the elevation, is neither very informative nor visually pleasing (looks like an X-ray image to me.)
Of course, it is easy to alter the color scheme and most GIS software programs provide multiple color scheme choices. As the elevation is expressed as a continuous value, choose a continuous, rather than discrete, color scheme. In the ArcGIS Pro, you can either right-click on the color scheme ramp (highlighted in Figure 2) in the Contents Pane, and use the drop-down arrow to reveal the choices, or you can select the DTM in the Contents Pane and open the Symbology Tab (Figure 3) to alter the color scheme.

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

**Tip #3 — Changing the Primary Symbology of the DTM**

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

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

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

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

Adjusting the sun angle to a position lower in the sky (Altitude = 15 degrees) exaggerates the shadows as in Figure 8 tends to increase shadow lengths, while increasing the sun altitude reduces shadows.
Similarly, the direction of the shadows, simulating time of day, can be changed by altering the Azimuth to have the sun in the east (~45 – 115 degrees) simulating a morning view, rather than the west (351 degrees) simulating an afternoon view.

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

**Tip #4 — Create your own Hillshade DTM and customize the look**

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

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

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

This workflow results in a visually appealing rendition (Figure 11) of the DTM similar to the Shaded Relief option in Tip 3, and permits additional customization of the DTM rendering independent of the “Shaded Relief” option.
Figure 9. The Spatial Analyst Hillshade Geoprocessing Tool dialog used to construct a 5X exaggerated hillshaded raster.

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

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

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

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

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

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

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, because when he is ready so will we.

asprsfoundation.org/donate

ASPRS WORKSHOP SERIES

It’s not too late to earn Professional Development Hours

Miss an ASPRS Workshop or GeoByte? Don’t worry! Many ASPRS events are available through our online learning catalog.

https://asprs.prolearn.io/catalog
According to author Breece Robertson, this book project (her first) was developed to share insights gained from years of experience working at the intersection of the worlds of land conservation advocacy and geospatial science and technology. Breece, currently the Director of Partnerships and Strategy for the Center for Geospatial Solutions at the Lincoln Institute of Land Policy, previously worked at the Trust for Public Lands, leading efforts devoted to land protection, improving park access and utilization, and development of voter-approved public funding for parks and conservation projects. In a 2021 promotional interview with ESRI co-founder and president Jack Dangermond, Breece suggests that her book was written for a broad audience but targeted toward land trusts and other conservation groups that are currently underutilizing GIS. As Jack suggested in the interview, a primary goal of the book is to “catalyze the conservation community with quantitative methods.” As access barriers (fiscal, technical, organizational, or otherwise) to geospatial science and tech have become less onerous in recent years, Breece calls attention to this opportunity at a critical time when environmental challenges have arguably reached crisis proportions.

The book is organized into ten Chapters, each highlighting opportunities for conservation organizations to improve their effectiveness and efficiency through the enhanced utilization of geospatial data and tools. The book is illustrated with colorful maps, graphics, and photos throughout.

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

One clear strength of the book is the large amount of practical discussion centered around data sourcing. This is a focus that runs throughout the book, pertaining to a wide variety of issues of interest to those engaged with advocating for and assessing the benefits of wildland conservation, parks, and green infrastructure development. Highlighted are US Census-derived demographic and socioeconomic data, multiscale land use/land cover data, outputs from ecological and habitat models, archives of location-based information on conservation projects, and many more datasets. This fundamental aspect of the book seems effective as an idea
generator to help with the potentially challenging process of locating appropriate and readily accessible data. Discussion of data reliability is perhaps a little sparse, but the sources referenced here are credible and authoritative, and issues related to data integrity and accuracy have been covered in depth elsewhere.

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

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

In conclusion, Protecting the Places We Love seems well-suited to its target audience. With a focused and deep understanding of the operational concerns of organizations devoted to the expansion of land conservation and green space, Robertson thoughtfully explores her subject in a clear, authoritative, and well-organized manner. The underlying message is about the diverse array of opportunities for conservation and parks advocates to enhance their work by leveraging the capabilities of geospatial science, and Robertson successfully presents it in a way that is both inclusive and empowering.
The Grids & Datums column has completed an exploration of every country on the Earth. For those who did not get to enjoy this world tour the first time, PE&RS is reprinting prior articles from the column. This month’s article on the Republic of El Salvador was originally printed in 2005 but contains updates to their coordinate system since then.

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

El Salvador (The Savior) is the smallest Spanish-speaking nation in the Western Hemisphere. It is located on the western side of the Central American isthmus. The country is slightly smaller than Massachusetts, and its land boundaries are with Guatemala (203 km), and Honduras (342 km) (PE&RS, July 1999). El Salvador’s coastline is 307 km on the Pacific Ocean and its terrain is mostly mountainous with a narrow coastal belt and central plateau. The lowest point is the Pacific Ocean (0 m), and the highest point is Cerro El Pital (2,730 m). El Salvador is known as the “Land of Volcanoes,” and is one of the most seismologically active regions on Earth, situated atop three large tectonic plates.

Topographic Mapping of the Americas, Australia and New Zealand states, “To 1930, the only detailed accurate surveying done in El Salvador – a country that became independent in 1821 – was the surveying done for the Intercontinental Railroad Commission and a few surveys related to possible road routes.” However, Nicaragua and Honduras had been squabbling regarding their border for more than 30 years, and an arbitration agreement was signed in 1930 and soon thereafter surveyed by the U.S. Coast & Geodetic Survey. El Salvador, Guatemala, and Honduras signed a protocol on 26 March 1936 accepting Cerro Monte Cristo as the tripoint of the boundaries of the three states. It was during this era that the first geodetic surveys were performed on the border between El Salvador and Guatemala from 1937 to 1940. The Dirección General de Cartografía was established in 1946 to...
continue the geodetic surveys and to implement a topographic mapping of the country. The end of the following year 1947, the Inter American Geodetic Survey (part of the U.S. Army Map Service) signed a cooperative agreement with El Salvador. By 1958, all classical geodetic triangulation had been completed in the country, and topographic mapping was well underway. The national mapping agency is now known as the Instituto Geográfico Nacional “Ing. Pablo Arnoldo Guzmán.”

The original horizontal datum of El Salvador is the Ocotepeque Datum of 1935 which was established by the U.S.C.&G.S. at Base Norte (in Honduras) where $\varphi_o = 14^\circ 26’ - 20.168’’$ North, $\lambda_o = 89^\circ 11’ 33.964’’$ West of Greenwich, and $H_o = 806.99$ meters above mean sea level. The defining geodetic azimuth to Base Sur is: $\alpha_o = 235^\circ 54’ 21.790’’$, and the ellipsoid of reference is the Clarke 1866 where $a = 6,378,206.4$ meters and $1/f = 294.9786982$. The corresponding astronomic observations at that mountainous location are: $\Phi_o = 14^\circ 26’ 13.73’’$ North ($\pm 0.07’’$), $\Lambda_o = 89^\circ 11’ 39.67’’$ West ($\pm 0.045’’$), and the defining astronomic azimuth to Base Sur is: $\alpha_o = 235^\circ 54’ 20.37’’$ ($\pm 0.28’’$). The difference between these two sets of coordinates is due to the local gravimetric deflection of the vertical. The grid system used in El Salvador is based on the Lambert Conformal Conic projection where the Central Meridian, $\lambda_o = 89^\circ 00’ 00’’$ West, the Latitude of Origin, $\varphi_o = 13^\circ 47’ 00’’$ North, the Scale Factor at Origin, $m_o = 0.999967040$, the False Easting = 500,000 m, and the False Northing = 295,809.184 m. Sometime after that, the National Geospatial-Intelligence Agency (nee NIMA) computed all the classical triangulation of Central America on the North American Datum of 1927. The 1987 published datum shift parameters for Central America based on 19 stations (TR 8350.2) from NAD27 to WGS84 are: $\Delta X = 0 \text{ m } \pm 8 \text{ m}$, $\Delta Y = +125 \text{ m } \pm 3 \text{ m}$, $\Delta Z = +194 \text{ m } \pm 5 \text{ m}$. After the disaster when Hurricane Mitch hit Central America, the U.S. National Geodetic Survey established a number of GPS Continuously Operating Reference Stations in the region. They observed a number of stations in El Salvador in order to establish and publish a High Accuracy Reference Network on the North American Datum of 1983. The National Geodetic Survey (NGS) has posted all of the data, including various other sources of geodetic coordinate data in El Salvador and the region on their website. The documentation of their geodetic observations is a superb resource for the researcher, and it is freely available. Thanks, NGS!

**Update on the Republic of El Salvador**

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

The contents of this column reflect the views of the author, who is responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the American Society for Photogrammetry and Remote Sensing and/or the Louisiana State University Center for GeoInformatics (C4G).

This column was previously published in PE&RS.
ASPRS SAC GIS DAY MAP CONTEST WINNERS!

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

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

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

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

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

~ Peter Samson, 2021 GIS Day Map Contest Winner
Members may notice a few changes in the online renewal process in 2023.

1. As part of their 2023 budget approval, the ASPRS Board of Directors voted to increase annual dues for Individual Members by $25 (from $150 to $175) effective May 1. This is the first dues increase in nearly a decade and is necessitated by rising costs. Sustaining Member and Student Member dues will remain unchanged in 2023.

2. Student Member renewal invoices will be issued at the Individual Member rate of $175, with an option to submit new proof of enrolment to obtain the lower Student Member dues rate of $50.

3. Voluntary tax-deductible donations of $25 to the ASPRS Foundation and $10 to ASPRS Initiatives will be included in the renewal invoice. These contributions are voluntary and can be removed from the cart before proceeding to payment.

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

NEW ASPRS MEMBERS

ASPRS would like to welcome the following new members!

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

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

Interested in ASPRS apparel? Visit https://stores.inksoft.com/asprs/shop/home
ASPRS Aerial Data Catalog

“THE SOURCE FOR FINDING AERIAL COLLECTIONS”

HTTP://DPAC.ASPRS.ORG

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

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

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

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

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

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

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

For More Details or To Get Involved Contact:

David Ruiz • druz@quantumspatial.com • 510-834-2001 OR David Day • dday@kasurveys.com • 215-677-3119
Global warming and climate change have become the most important factor threatening the world. Climate change results in dramatical environmental hazards and threatens the planet and human life. A wide variety of policies have been proposed to decrease the effects of global warming and climate change. The most important one is the Paris Agreement which aims to limit global warming to well below two degrees Celsius. Many countries have formulated long term low greenhouse gas emission development strategies related to the Paris Agreement which aimed to meet the essential strategies addressing issues with climate change, environmental protection and low carbon.

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

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

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

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

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

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

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

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

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

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

---

Tolga Bakırman is with the Yıldız Technical University, Istanbul, Turkey. bakirman@yildiz.edu.tr
Call for *PE&RS* Special Issue Submissions

**Innovative Methods for Geospatial Data using Remote Sensing and GIS**

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

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

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

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

**Deadline for Manuscript Submission—July 1, 2023**

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

**Guest Editors**

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

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

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

Introduction

In the wake of disaster, timely, effective emergency management saves lives (Kerle and Oppenheimer 2002). Emergency management officials such as the US Department of Energy and the Federal Emergency Management Agency (FEMA), and others across the globe, need actionable information for damage assessment to guide critical allocations of lifesaving resources within 72 hours (FEMA 2020; Hodgson et al. 2010). Traditionally, damage assessments have been performed by local disaster response crews deployed to the affected area. Any imagery-based data collection (e.g., from unmanned aerial systems (UAS)) suffered from download, processing, and/or upload delays. Collections also frequently relied on local communication infrastructures that may have sustained damage (Boccardo and Tonolo 2015). Methods that rely on social media (e.g., Twitter messages) to infer damage require communication environments and beyond visual line of sight. Solution extensions address operation in constrained communication infrastructure; and has enough onboard processing power to: (1) autonomously assess damage of energy infrastructure such as utility poles and substations using machine learning (ML), (2) geo-locate detected features of interest, and (3) transmit relevant damage assessment results over constrained communications. Initial testing, based on collected imagery from before and after Hurricane Ida, attempted to detect utility poles and classify them as damaged or undamaged.

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

Related Work

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

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

There are at least two published reports of UAS onboard feature detection. Jiang et al. (2022) reported UAS onboard feature detection of bridge cracks, spalling, and corrosion. Kraft et al. (2021) built a custom drone with an onboard global positioning unit (GPU) to automatically detect and geo-position features of interest (e.g., litter). Our solution extends these by addressing operation in constrained communication environments and beyond visual line of sight.
Nguyen et al. (2019) performed work relevant to energy infrastructure. They trained ML models to detect missing utility pole caps, cracked poles, cracked cross arms, rot-damaged cross arms, and woodpecker-damaged poles. Their work differs from ours in two important ways. First, it was not implemented on-the-edge, in restrained communication, or beyond visual line of sight. Second, the image resolutions differed by orders of magnitude because Nguyen et al. were looking for small features (i.e., cracks and rot), while we were looking for broken and leaning poles. Hence, direct comparisons of the network performance are not useful. Other efforts to use UAS in disaster response (e.g., floods) stop short of actual edge computing (Koay et al. 2021; Hernandez et al. 2022).

**Methods**

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

**Airframe**

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

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

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

**Onboard Processors**

We used the ORNL-developed PixC4-Jetson, which is a flight controller and co-processor carrier board to host the popular Nvidia Jetson NX series of companion computers. Figure 3 shows a schematic of the associated hardware. The PixC4-Jetson provides a lightweight, compact, and robust means of integrating the companion computer with the flight controller. This integration provides the co-processor with the position and orientation data required to geo-locate acquired imagery. The companion computer hosts the Docker ML algorithms used in this project while simultaneously performing real-time onboard processing. Figure 4 shows the PixC4 module with Nvidia Jetson co-processor.
Command and Control Link
We used the MAVNet system for global command and control of the UAV. MAVNet is a suite of hardware and software technologies developed for enabling next-generation unmanned systems research. The MAVNet communication system maintains communication links by allowing mesh networks, cellular/5G, and Iridium satellite communication modes to coexist, automatically falling forwards or backwards among them. This communication paradigm allows real-time operation in remote, austere environments such as those requiring disaster response (Harter et al. 2019).

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

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

Figure 5. MAVNet GCS with sample detections.

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

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

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

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

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

The image coordinate system is defined with respect to the image perspective center and focal length. The origin is at the image principal point (i.e., the intersection of the image’s optical axis with the image plane). The positive \(x\)-direction is from image left to image right. The positive \(y\)-direction is from image bottom to image top. The negative \(z\)-direction (to keep with a right-handed coordinate system) is the image look direction from the perspective center out through the principal point. Image coordinates, \(x = x' y' z'\), have only two degrees of freedom because \(z' = -f'\) is constant, where \(f'\) is the focal length of the camera. The units are the same as the focal length (typically mm or pixels). The GCS can be any right-handed 3D Cartesian system.

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

The image position, \(O = X' Y' Z'\), and orientation define the transformation from ground space to image space. Orientation has three degrees of freedom and is expressed in a variety of ways. Roll, pitch, and heading; Euler angles; quaternions; and \(3 \times 3\) orthogonal rotation matrices are common examples. The rotation matrix, \(M\), is the form used in the formulas below. \(M\) rotates from the GCS to the image coordinate system.

The G2I transformation can be written as \(x = 2Mx - O\), where \(X = X' Y' Z'\) is a ground point and \(\lambda = -f'(M_{11}x + M_{12}y + M_{13}z + Z)\), where \(Z = (Z_x Z_y Z_z)\) is the scale that moves \(x\) into the focal plane. The inverse, I2G, can thus be written as \(O = \lambda^{-1}M'x = X\).

Note that \(\lambda\) is the range along the line defined by the camera center \(O\) and look vector \(M'x\) to the ground point. \(\lambda\) is unrecoverable in the inversion because it is a function of the unknown \(X\). However, the ray \(r = O + \rho M'x\) with \(\rho > 0\) contains \(X\). Our I2G operation resolves \(\rho\) by intersecting \(r\) with the approximate ground plane, \(Z = Z_\perp\):
Cornelia’s real-time telemetry data includes airframe position, yaw ($\alpha$), roll ($\beta$), and pitch ($\gamma$). The constant rotation between the airframe and camera had to be calibrated separately. This is modeled as $x = \lambda M M_c X - O$. $M_c = f(\alpha, \beta, \gamma)$ is the time-variant rotation from the ground to airframe systems provided by the real-time telemetry. $M$ is the constant rotation from airframe to camera systems; $M = M_c M_a$, and $\lambda$ is the same scale factor.

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

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

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

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

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

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

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

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

**Summary of the Image Collection and Processing Framework**

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

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

Part 1—Process the Image

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

Part 2—Process the Labels File

f. For each detected feature, compute a ground location (Geopositioning Algorithm Development)

g. Upload the ground location to MAVNet with the feature type, damage rating, and confidence. These are also displayed as icons in the MAVNet user interface.

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

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

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

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

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

**Results and Discussion**

This section discusses results of this work as pertains to hardware, geopositioning, and ML.

**Hardware**

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

The edge processing UAS prototype was developed to understand the future needs for state and local governments, who typically assess damage following natural disasters, to implement the ML and photogrammetry onboard hardware and analysis required. MAVNet is designed to enable next-generation applications for small autonomous systems. MAVNet’s hardware component, PixC4-Jetson, is mated to a COTS quadcopter. A specially designed microcontroller-based vehicle controller with an integrated GPU-accelerated companion computer is onboard the platform. This hardware is compatible with a wide variety...
of commercial and hobby-level platforms. The hardware collects imagery from an onboard camera and immediately begins processing the data onboard, seeking to identify damaged utility poles and other infrastructure. In just seconds, the resulting data products are uploaded to the MAVNG cloud server and distributed to users anywhere in the world. In contrast, the stages of current systems involve manually creating flight.

**Geo-Positioning**

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

**Machine Learning**

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

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

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

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

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

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

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

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

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

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

<table>
<thead>
<tr>
<th>Location</th>
<th>Total Images</th>
<th>Training</th>
<th>Validation</th>
<th>Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Middlebrook Pike</td>
<td>257</td>
<td>182</td>
<td>50</td>
<td>25</td>
</tr>
<tr>
<td>Northshore</td>
<td>484</td>
<td>340</td>
<td>96</td>
<td>48</td>
</tr>
<tr>
<td>Hurricane Ida</td>
<td>334</td>
<td>235</td>
<td>66</td>
<td>33</td>
</tr>
<tr>
<td>Combined</td>
<td>1075</td>
<td>757</td>
<td>212</td>
<td>106</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Feature of Interest</th>
<th>Ida</th>
<th>Ida</th>
<th>Ida</th>
<th>Ida</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training</td>
<td>Validation</td>
<td>Testing</td>
<td>Total</td>
</tr>
<tr>
<td>Undamaged pole</td>
<td>2063</td>
<td>458</td>
<td>329</td>
<td>2850</td>
</tr>
<tr>
<td>Damaged pole</td>
<td>504</td>
<td>168</td>
<td>112</td>
<td>784</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Data</th>
<th>Model size</th>
<th>Precision</th>
<th>Recall</th>
<th>F-score</th>
<th>mAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>YOLOv5</td>
<td>One-stage</td>
<td>Large</td>
<td>0.811</td>
<td>0.48</td>
<td>0.603</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Medium</td>
<td>0.852</td>
<td>0.423</td>
<td>0.565</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Small</td>
<td>0.778</td>
<td>0.333</td>
<td>0.466</td>
</tr>
<tr>
<td></td>
<td>Two-stage</td>
<td>Large</td>
<td>0.885</td>
<td>0.516</td>
<td>0.652</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Medium</td>
<td>0.881</td>
<td>0.488</td>
<td>0.628</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Small</td>
<td>0.855</td>
<td>0.411</td>
<td>0.555</td>
</tr>
<tr>
<td>YOLOv3</td>
<td>One-stage</td>
<td>Standard</td>
<td>0.303</td>
<td>0.439</td>
<td>0.358</td>
</tr>
<tr>
<td></td>
<td>Two-stage</td>
<td>Standard</td>
<td>0.422</td>
<td>0.456</td>
<td>0.471</td>
</tr>
</tbody>
</table>

### Table 4. Object detection model size and inferencing speed.

<table>
<thead>
<tr>
<th>Framework</th>
<th>Number of Parameters</th>
<th>Inferencing Speed (Seconds per Image)</th>
</tr>
</thead>
<tbody>
<tr>
<td>YOLOv5 Large</td>
<td>46 149 064</td>
<td>2.7</td>
</tr>
<tr>
<td>YOLOv5 Medium</td>
<td>20 879 400</td>
<td>1.35</td>
</tr>
<tr>
<td>YOLOv5 Small</td>
<td>7 027 720</td>
<td>0.48</td>
</tr>
<tr>
<td>YOLOv3</td>
<td>61 508 200</td>
<td>1.03</td>
</tr>
</tbody>
</table>
Figure 11. Examples of the predictions from three YOLOv5 models. DMG in red boxes: damaged utility pole; UND in cyan boxes: undamaged utility pole.

Figure 12. Examples of incorrect predictions and commonly seen failures. DMG in red boxes: damaged utility pole; UND in cyan boxes: undamaged utility pole.
The current processing workflow operates on still-image captures and functions within the capabilities of our current UAS prototype. We see advantages in moving toward processing the video stream. Edge processing would identify and track possible features of interest, providing multiple views. The feature detection performance offsets required for faster processing would ideally be offset by the redundancy. Additionally, multiple observations through a time series would allow I2G operations by photogrammetric intersection independent of ground surface priors. Perhaps more importantly, we could photogrammetrically estimate pole heights and lean to increase robustness of detection and classification.

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

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

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

References


Kuchhold, M., et al. 2021. MultEYE: Monitoring system for real-time access to these results of federally sponsored research in accordance with the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

References


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

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

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

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

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

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

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

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

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

WHO’S WHO IN ASPRS

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

BOARD OF DIRECTORS

BOARD OFFICERS

President
Christopher Parrish, Ph.D
Oregon State University

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

Vice President
Bandana Kar
Oak Ridge National Lab

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

Treasurer
John McCombs
NOAA

Secretary
Harold Rempel
ESP Associates, Inc.

COUNCIL OFFICERS

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

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

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

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

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

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

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

TECHNICAL DIVISION OFFICERS

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

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

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

Lidar Division
Director: Ajit Sampath
Assistant Director: Mat Bethel

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

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

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

UAS Division
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
Robert Hariston-Porter

Cascadia Region
Robert Hariston-Porter

Eastern Great Lakes Region
Michael Joos, CP, GISP

Florida Region
Xan Fredericks

Heartland Region
Whit Lynn

Intermountain Region
Robert T. Pack

Mid-South Region
David Hughes

Northeast Region

North Atlantic Region

Pacific Southwest Region
John Erickson, PLS, CP

Rocky Mountain Region

Western Great Lakes Region

Potomac Region
Dave Lasko

Western Great Lakes Region
Adam Smith
Apricot Tree Detection from UAV-Images
Using Mask R-CNN and U-Net

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

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

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

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

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

The speed performances and capacity limits of deep learning algorithms, almost all of which are run better on a graphics processing unit (GPU), are directly related to the memory and speed values of this hardware. There is a certain upper limit to the image size the GPU can process, depending on the model used. For example, when Matterport’s Mask R-CNN implementation (Abdulla 2017) is used as the model, the maximum RGB image size that an Nvidia GTX 1080 Ti GPU with 11 GB of internal memory can handle is 4896×4896 pixels. For this reason, when object detection is to be done larger than the upper limit, it is necessary to split the image into parts or reduce the image size to sufficient dimensions (hence compromising on detail). In studies dealing with such cases (Ma et al. 2018; Gao et al. 2019; Zhang et al. 2019), the image is usually divided into small pieces that overlap to a certain extent, and then additional processes are applied to prevent repetitive object counting in the overlap regions and to determine the object locations correctly. However, as these studies are concerned with only object detection, the problem of the existence of multiple outputs for the same ground truth in the overlap regions is not addressed in terms of mapping tasks. Encountered in cases where object detection is performed with deep learning–based instance segmentation methods, such as Mask R-CNN or U-Net, the generated masks do not fit correctly to the ground truth, as some artificial gaps between the masks and the output image edges are formed for the objects close to the edges, eventually leading to mask localization inaccuracies (Cheng et al. 2020; Kirillov et al. 2020). In cases where the dimensions of the image to be examined are larger than the capacity of the GPU used, splitting the image into parts for processing increases the total number of edges, eventually boosting incorrect segmentation and detection results in the regions close to the edges.

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

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

Materials and Methods

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

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

Instance Segmentation Models Based on Deep Learning
Being a subset of machine learning, deep learning is a neural network with multiple layers. Similar to the learning behavior of the human brain, these networks use large amounts of data for a specific purpose.
in a process to minimize the error. Among the key problems deep neural networks can solve is detecting objects in images. The computer vision task for detecting and localizing an object in an image is defined as instance segmentation. In this study, we use two different types of instance segmentation algorithms—Mask R-CNN and U-Net—for apricot tree detection. In the following parts, details about both methods will be given.

**Mask R-CNN**

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

**U-Net**

The U-Net architecture is a fully convolutional network, first designed for image segmentation in biomedicine. The results of the ISBI 2012 challenge for segmentation of neural structures in electron microscopic stacks showed that U-Net architecture can be trained using a small number of photos and exceeded the previous best approach (Ronneberger et al. 2015). It is divided into two parts: the contracting path and the expansive path. The contracting path is made up of recurrent 3×3 convolutions, rectified linear unit, and 2×2 max-pooling layers with stride 2. The number of feature channels is doubled after down-sampling. The expansive path is made up of up-sampling the feature map, 2×2 up-convolution, and concatenation with the equivalent feature map from the contracting path. To map each feature vector in the final layer, a 1×1 convolution is employed. The schematic diagram of the U-Net architecture is given in Figure 4. In this study, images of 768×768 pixels were input to the U-Net, and binary masks of the same size were produced as output, where trees were represented with white and other regions with black.
Postprocessing: Dilation
The output map for the entire test orthophoto is produced by combining the binary mask outputs for each image subset. However, due to mask localization inaccuracies, some artificial gaps between the masks and the output image edges are formed, clearly visible at the intersections of small images (Figure 5). To avoid this problem, the morphologic dilation process was applied to the binary image. Erosion and dilation are the fundamental operators in mathematical morphology. When mathematical morphology is utilized in image processing, these operations are applied to an image with a collection of known shapes, called a structuring element. When the dilation operator is applied to an image, it produces an output image that illustrates where the structuring element hits the objects in the image (Benediktsson et al. 2003). It enlarges bright regions and shrinks dark regions. In this study, the shape of the structuring element was selected as a disk, and the radius of the disk-shaped footprint was set to 11 for outputs of Mask R-CNN and 7 for outputs of U-Net. These parameters were selected empirically. As a result of dilation operation, tree crowns are dilated, and edge mistakes were mitigated.

Work Flow for Apricot Tree Detection in Large-Scale UAV Images
The detection and segmentation of apricot trees in large-scale images are difficult. Due to hardware limitations, to perform these tasks with deep learning methods, the image size must be brought to a level that the GPU can handle by either cropping the image into parts or shrinking its dimensions. However, resizing the image will reduce details, making it hard to detect small targets. Therefore, it is necessary to examine the image by cropping in order not to lose detail and to analyze the image in its original resolution. So, as in our study (see the flowchart in Figure 6), the orthophoto is divided into patches that do not overlap each other, the Mask R-CNN and U-Net models are fed with these patches, and mask outputs for the trees are produced. However, due to mask localization inaccuracies near the output edges, it is not possible to make accurate detections for the entire orthophoto simply by putting these patches together. Mask localization inaccuracies near the image edges are encountered in deep learning instance segmentation models, where the masks of the objects close to the edges of the output image do not fit the ground truth, and artificial gaps are formed between the mask border and the edge. As in this study, splitting the image into many parts increases the number of edges, further exacerbating the edge problem. In order to alleviate this problem, this article proposes applying a morphology operator, such as dilation, to the detection outputs before putting them together. The binary output detection results for test orthophoto were obtained in four steps:
1. Cropping the orthophoto into patches that do not overlap each other (4800×4800 pixels for Mask R-CNN, 768×768 for U-Net)
2. Feeding the cropped images into the models
3. Applying dilation operator to the model outputs
4. Putting the filtered outputs together

**Accuracy Assessment Method**

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

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

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

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

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

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

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

**Experimental Setup**

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

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

**Results and Discussion**

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

**Comparison of the Models in Tree Detection**

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

<table>
<thead>
<tr>
<th>Test Image</th>
<th>Label</th>
<th>Mask R-CNN</th>
<th>U-Net</th>
<th>CART</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Image" /></td>
<td><img src="image" alt="Image" /></td>
<td><img src="image" alt="Image" /></td>
<td><img src="image" alt="Image" /></td>
<td><img src="image" alt="Image" /></td>
</tr>
</tbody>
</table>

Figure 7. Detection results.

**Table 1. Accuracy assessment results.**

<table>
<thead>
<tr>
<th>Metric</th>
<th>Mask R-CNN (%)</th>
<th>U-Net (%)</th>
<th>CART (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>With Dilation</td>
<td>Without Dilation</td>
<td>With Dilation</td>
</tr>
<tr>
<td>F1 score</td>
<td>99.1</td>
<td>91.1</td>
<td>95.2</td>
</tr>
<tr>
<td>Precision</td>
<td>98.7</td>
<td>85.3</td>
<td>97.2</td>
</tr>
<tr>
<td>Recall</td>
<td>99.7</td>
<td>97.8</td>
<td>93.3</td>
</tr>
<tr>
<td>IoU</td>
<td>74.8</td>
<td>69.6</td>
<td>58.3</td>
</tr>
</tbody>
</table>

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

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

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

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

**Conclusion**
This study aims to automatically detect, count, and map apricot trees in a large-scale orthophoto, using Mask R-CNN and U-Net instance
segmentation models, and to improve the results by alleviating the object detection problem encountered around image edges. For this purpose, first, training and testing of the models were conducted using the images of apricot trees with distinct characteristics from four different sites to make models robust against variations in diverse tree views. Then a dilation filter was applied on the patches of mask outputs to eliminate gaps between two masks of the same tree in the junction areas of the patches.

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

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

**Acknowledgment**

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

**References**


---

**Figure 9.** Effect of dilation filter on the mask outputs generated by Mask region-based convolutional neural network and U-Net for the same ground truth.
Comparative Analysis of Different CNN Models for Building Segmentation from Satellite and UAV Images

Batuhan Sarıturk, Dağla Kumbasar, and Dursun Zafer Seker

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

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

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

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

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

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

Despite FCN’s success, there are some limitations. Firstly, fully convolutional layers in FCNs lead to a large number of parameters and computation complexity. Secondly, FCNs fail to provide detailed information about complex structures and lose information about small objects (Khan et al. 2022). To overcome these disadvantages, (Ronneberger et al. 2015) proposed U-Net, which achieved significant...
success in medical image segmentation studies. U-Net, which adopts a U-shape architecture, consists of two parts. The first part is the encoder, which extracts features from the input image. The second part is the decoder, which performs the dense predictions. Nowadays, most CNNs used in segmentation studies from aerial images are based on FCNs.

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

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

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

Methodology

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

Architectures

U-Net

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

Feature Pyramid Networks (FPN)

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

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

LinkNet

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

SegNet

Badrinarayanan et al. (2017) presented a deep fully convolutional network for semantic segmentation, called SegNet. The proposed architecture consists of an encoder, a decoder, and a classification
The topology of the encoder is identical to the first 13 convolutional layers used in the VGG16 network (Badrinarayanan et al. 2017). Every convolution layer is followed by a Batch Normalization and a ReLU. There is a corresponding decoder layer for each encoder layer; hence the network has 13 decoder layers. The decoder maps the low-resolution feature maps coming from the encoder to input resolution for pixel-wise classification. The final decoder output is a classifier to produce final class probabilities.

**Fully Convolutional Networks (FCN)**

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

**Backbone (EfficientNet)**

In the study, EfficientNet is used as a backbone for architectures U-Net, FPN, and LinkNet. Tan and Le (Tan and Le 2019) studied the model scaling and found that balancing the width, depth, and resolution of a network can lead to better performance. They proposed a scaling method that scales all dimensions uniformly, using a compound coefficient (Figure 6). As a result of these studies, they designed a baseline network and scaled the network up to obtain a model family called EfficientNet. The EfficientNet family consists of eight models, named from B0 to B7 (Tan and Le 2019). As the number in the model increases, the network becomes deeper, wider, and reaches a higher resolution.
name increases, it is stated that the model has more parameters. For this study, the EfficientNet-B2 variant is selected as the backbone.

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

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

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

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

Massachusetts Buildings Data Set
One of two data sets used to generate additional test sets is Massachusetts Buildings Data Set. This data set was introduced by Volodymyr Mnih in 2013 (Mnih 2013) and includes 151 RGB images and corresponding masks from Boston, Mass., USA. Images in the data set is the size of 5000 × 5000 pixels and covers an area of 2.25 km². Selected images and masks from the set were cropped into 512 × 512 pixel size patches (Figure 9), similar to the preparation of the Inria set, and the images with little to no amount of building classes were
elimited. As a result of all the steps, a test set that includes 225 images and corresponding masks was obtained.

**Syedra Archaeological Site Data Set**

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

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

**Training and Testing**

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

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

\[
L(y, \hat{y}) = \frac{1}{N} + \sum_{i=0}^{N} \left( y \log(\hat{y}) + (1-y) \log(1-\hat{y}) \right)
\]  
\[
DL(y, \hat{y}) = 1 - \frac{2y\hat{y} + 1}{y + \hat{y} + 1}
\]

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

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

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

\[
\text{Precision} = \frac{TP}{TP + FP}
\]
\[
\text{Recall} = \frac{TP}{TP + FN}
\]
\[
\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}
\]
\[
\text{IoU}(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cup B|}
\]
Results and Discussion

A summary of evaluation metric results is shown in Table 2. It has been observed that on the Inria test set, the Residual-2 U-Net model has the highest IoU score, F1 score, and test accuracy with 0.824, 0.722, and 94.7%, respectively. For the F1 score, Residual-3 U-Net and U-Net follow Residual-2 U-Net with 0.824 and 0.819. For the IoU score, U-Net and Residual-6 U-Net follow with 0.712.

On the Syedra test set, the LinkNet model with the EfficientNet-B5 backbone has the highest F1 and IoU scores with 0.336 and 0.246. Residual-4 U-Net and Residual-3 U-Net follow with 0.252 and 0.249 F1 scores, and 0.172 and 0.178 IoU scores, respectively.

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

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

According to the results obtained on prepared test sets, the Residual-3 U-Net architecture, using the RB-3 residual block design, is in the top three for all test sets according to the evaluations. Therefore, it can be stated that it is the most inclusive model among all used approaches. When evaluated in general, it has been observed that
Figure 12. Segmentation results of image no. 128 from the Massachusetts test set.

Figure 13. Segmentation results of image no. 329 from the Syedra test set.
models using residual connections are more successful than models using conventional convolution structures. On all three test sets, at least two of the top three most successful models, according to the evaluation metric results, use architectures with residual connections. This can be interpreted as residual connections being more successful due to their ability to carry spatial information from earlier layers. The model uses LinkNet architecture with EfficientNet-B5 backbone achieved the most successful results on the Syedra test set, which is the test set that has the most different feature characteristics, which are archaeological buildings and building remains. Since this architecture achieves the most promising results compared to other models, it could be said that this approach is best suited for different feature characteristics among all models and could be a better choice to be used for future studies on archaeological sites. On the other hand, FPN architecture with EfficientNet-B5 backbone and SegNet were the least successful models for all three test sets. According to the evaluation metric results, SegNet was in the bottom three for all three test sets and provided the lowest F1 and IoU scores for Syedra and Massachusetts test sets. Considering these results on test sets that have different types of feature characteristics than the training set, it could be said that SegNet is the approach that has the least generalization ability among the others used in the study. Therefore, it can be stated that the architectural structures of FPN and SegNet are the least suitable model approaches for building detection and segmentation study compared to other approaches used.

Conclusions
Over the years, automatic building segmentation from aerial images has become a challenging and important research topic with the rapid developments in especially urban areas. Within the study, building segmentation from satellite and UAV images was performed using the Inria Aerial Image Labeling Data Set as the training set, and the Inria Aerial Image Labeling Data Set, the Massachusetts Buildings Data Set, and the Syedra Archaeological Site Data Set as test sets. A total of 12 models were generated, and comparisons were realized. Results show that the Residual-2 U-Net model, which uses a different residual block from ResNet, performs best in all evaluation metrics on the Inria test set. According to the evaluation metrics, the LinkNet EfficientNet-B5 model on the Syedra Archaeological Site Test Set and the Residual-4 U-Net model on the Massachusetts Test Set showed the best results. Since the features in Syedra Archaeological Site Test Set have different characteristics from the Inria Training Set, it may be said that the LinkNet EfficientNet-B5 model has high generalization ability. Nevertheless, the features from Massachusetts Test Set are similar to the Inria Training Set but from a different kind of sensor, and because of that it can be mentioned that the Residual-4 U-Net model has a generalization ability across similar features and different sensors. In further studies, performance improvements will be realized on the proposed residual models, with the implementation of approaches such as attention gates and transformers. Additionally, studies to extract buildings and building remains on the archaeological sites will be continued using the Syedra Archaeological Site Data Set and different CNN approaches.

References


SUSTAINING MEMBERS

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

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

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

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

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

CT Consultants
Mentor, Ohio
Member Since: 3/2022

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

Esti
Redlands, California
www.esti.com
Member Since: 1/1987

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

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

Half Associates, Inc.
Richardson, Texas
www.half.com
Member Since: 8/2021

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

SUSTAINING MEMBER 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
Abstract
Quantitative estimation of crop nitrogen is the key to site-specific management for enhanced nitrogen (N) use efficiency and a sustainable crop production system. As an alternate to the conventional approach through wet chemistry, sensor-based noninvasive, rapid, and near-real-time assessment of crop N at the field scale has been the need for precision agriculture. The present study attempts to predict leaf N of wheat crop through spectroscopy using a field portable spectroradiometer (spectral range of 400–2500 nm) on the ground in the crop field and an imaging spectrometer (spectral range of 400–1000 nm) from an unmanned aerial vehicle (UAV) with the objectives to evaluate (1) four multivariate spectral models (i.e., artificial neural network, extreme learning machine [ELM], least absolute shrinkage and selection operator, and support vector machine regression) and (2) two sets of hyperspectral data collected from two platforms and two different sensors. In the former part of the study, ELM outperforms the other methods with maximum calibration and validation R² of 0.99 and 0.96, respectively. Furthermore, the image data set acquired from UAV gives higher performance compared to field spectral data. Also, significant bands are identified using stepwise multiple linear regression and used for modeling to generate a wheat leaf N map of the experimental field.

Introduction
Plant growth is highly dependent on nitrogen (N) and has a direct influence on the quality of crops. Leaf N (LNC) content is one of the major factors governing leaf nutrition and is a decision-making factor for nutritional security of the country (Anuj et al. 2014). Therefore, agriculture production that increases yield and sustainability requires satisfying the demand and supply trade-off for N use efficiency. For this purpose, management of N is complex and necessitates problem-oriented, interdisciplinary research (Spiertz 2009). Destructive methods for crop nutrient assessment of crop conditions at very high resolution. At present, the sensors that are mounted on UAVs are mostly red, green, blue and multispectral cameras with fewer spectral bands. Therefore, the acquired spectral information is limited to few applications. However, hyperspectral sensors, with hundreds of narrow and contiguous bands, have the capability of capturing subtle variations in plant growth and parameters, thereby having many potential applications in agriculture. The development of hyperspectral sensors has enabled quantitative assessment of plant parameters, encouraging researchers to implement precision agriculture. However, hyperspectral remote sensing comes with its own set of challenges, such as spectral variability, mixed pixel problems, accuracy requirements, selection of an appropriate model for information extraction that would provide better accuracy, and so on (Das et al. 2015; Hank et al. 2019; Pandey et al. 2020). Civil aviation and other relevant authorities elsewhere in the world are now allowing UAVs to be used for limited and specific trials and, in some cases, commercial operations in agriculture (Ahirwar et al. 2019), horticulture (Tu et al. 2020), viticulture (Spachos and Stefano 2019), and forestry (Tang and Shao 2015). Spraying for controlling diseases, weeds, and pests; spreading microgranular pesticides and fertilizers and even beneficial insects; and planting are among the diverse uses now being done using UAVs (Rao Mogili and Deepak 2018; Chen et al. 2021). Plant parameters obtained from UAV-based sensors can be used not only for monitoring crop growth conditions but also for crop response to differential treatment of nutrients and water (Yang et al. 2020). N status monitoring is important for farmers, as it leads to higher performance compared to field spectral data. Also, significant bands are identified using stepwise multiple linear regression and used for modeling to generate a wheat leaf N map of the experimental field.

Unmanned Aerial Vehicle (UAV)–Based Imaging Spectroscopy for Predicting Wheat Leaf Nitrogen
Rabi N. Sahoo, Shalini Gakhar, R.G. Rejith, Rajeev Ranjan, Mahesh C. Meena, Abir Dey, Joydeep Mukherjee, Rajkumar Dhakar, Sunny Arya, Anchal Daas, Subhash Babu, Pravin K. Upadhyay, Kapila Sekhawat, Sudhir Kumar, Mahesh Kumar, Viswanathan Chinnusamy, and Manoj Khanna

Abstract
Quantitative estimation of crop nitrogen is the key to site-specific management for enhanced nitrogen (N) use efficiency and a sustainable crop production system. As an alternate to the conventional approach through wet chemistry, sensor-based noninvasive, rapid, and near-real-time assessment of crop N at the field scale has been the need for precision agriculture. The present study attempts to predict leaf N of wheat crop through spectroscopy using a field portable spectroradiometer (spectral range of 400–2500 nm) on the ground in the crop field and an imaging spectrometer (spectral range of 400–1000 nm) from an unmanned aerial vehicle (UAV) with the objectives to evaluate (1) four multivariate spectral models (i.e., artificial neural network, extreme learning machine [ELM], least absolute shrinkage and selection operator, and support vector machine regression) and (2) two sets of hyperspectral data collected from two platforms and two different sensors. In the former part of the study, ELM outperforms the other methods with maximum calibration and validation R² of 0.99 and 0.96, respectively. Furthermore, the image data set acquired from UAV gives higher performance compared to field spectral data. Also, significant bands are identified using stepwise multiple linear regression and used for modeling to generate a wheat leaf N map of the experimental field.

Introduction
Plant growth is highly dependent on nitrogen (N) and has a direct influence on the quality of crops. Leaf N (LNC) content is one of the major factors governing leaf nutrition and is a decision-making factor for nutritional security of the country (Anuj et al. 2014). Therefore, agriculture production that increases yield and sustainability requires satisfying the demand and supply trade-off for N use efficiency. For this purpose, management of N is complex and necessitates problem-oriented, interdisciplinary research (Spiertz 2009). Destructive methods for crop nutrient assessment of plant parameters, thereby having many potential applications in agriculture. The development of hyperspectral sensors has enabled quantitative assessment of plant parameters, encouraging researchers to implement precision agriculture. However, hyperspectral remote sensing comes with its own set of challenges, such as spectral variability, mixed pixel problems, accuracy requirements, selection of an appropriate model for information extraction that would provide better accuracy, and so on (Das et al. 2015; Hank et al. 2019; Pandey et al. 2020). Civil aviation and other relevant authorities elsewhere in the world are now allowing UAVs to be used for limited and specific trials and, in some cases, commercial operations in agriculture (Ahirwar et al. 2019), horticulture (Tu et al. 2020), viticulture (Spachos and Stefano 2019), and forestry (Tang and Shao 2015). Spraying for controlling diseases, weeds, and pests; spreading microgranular pesticides and fertilizers and even beneficial insects; and planting are among the diverse uses now being done using UAVs (Rao Mogili and Deepak 2018; Chen et al. 2021). Plant parameters obtained from UAV-based sensors can be used not only for monitoring crop growth conditions but also for crop response to differential treatment of nutrients and water (Yang et al. 2020). N status monitoring is important for farmers, as it leads to higher performance compared to field spectral data. Also, significant bands are identified using stepwise multiple linear regression and used for modeling to generate a wheat leaf N map of the experimental field.

Unmanned Aerial Vehicle (UAV)–Based Imaging Spectroscopy for Predicting Wheat Leaf Nitrogen
Rabi N. Sahoo, Shalini Gakhar, R.G. Rejith, Rajeev Ranjan, Mahesh C. Meena, Abir Dey, Joydeep Mukherjee, Rajkumar Dhakar, Sunny Arya, Anchal Daas, Subhash Babu, Pravin K. Upadhyay, Kapila Sekhawat, Sudhir Kumar, Mahesh Kumar, Viswanathan Chinnusamy, and Manoj Khanna

Abstract
Quantitative estimation of crop nitrogen is the key to site-specific management for enhanced nitrogen (N) use efficiency and a sustainable crop production system. As an alternate to the conventional approach through wet chemistry, sensor-based noninvasive, rapid, and near-real-time assessment of crop N at the field scale has been the need for precision agriculture. The present study attempts to predict leaf N of wheat crop through spectroscopy using a field portable spectroradiometer (spectral range of 400–2500 nm) on the ground in the crop field and an imaging spectrometer (spectral range of 400–1000 nm) from an unmanned aerial vehicle (UAV) with the objectives to evaluate (1) four multivariate spectral models (i.e., artificial neural network, extreme learning machine [ELM], least absolute shrinkage and selection operator, and support vector machine regression) and (2) two sets of hyperspectral data collected from two platforms and two different sensors. In the former part of the study, ELM outperforms the other methods with maximum calibration and validation R² of 0.99 and 0.96, respectively. Furthermore, the image data set acquired from UAV gives higher performance compared to field spectral data. Also, significant bands are identified using stepwise multiple linear regression and used for modeling to generate a wheat leaf N map of the experimental field.

Introduction
Plant growth is highly dependent on nitrogen (N) and has a direct influence on the quality of crops. Leaf N (LNC) content is one of the major factors governing leaf nutrition and is a decision-making factor for nutritional security of the country (Anuj et al. 2014). Therefore, agriculture production that increases yield and sustainability requires satisfying the demand and supply trade-off for N use efficiency. For this purpose, management of N is complex and necessitates problem-oriented, interdisciplinary research (Spiertz 2009). Destructive methods for crop nutrient assessment of plant parameters, thereby having many potential applications in agriculture. The development of hyperspectral sensors has enabled quantitative assessment of plant parameters, encouraging researchers to implement precision agriculture. However, hyperspectral remote sensing comes with its own set of challenges, such as spectral variability, mixed pixel problems, accuracy requirements, selection of an appropriate model for information extraction that would provide better accuracy, and so on (Das et al. 2015; Hank et al. 2019; Pandey et al. 2020). Civil aviation and other relevant authorities elsewhere in the world are now allowing UAVs to be used for limited and specific trials and, in some cases, commercial operations in agriculture (Ahirwar et al. 2019), horticulture (Tu et al. 2020), viticulture (Spachos and Stefano 2019), and forestry (Tang and Shao 2015). Spraying for controlling diseases, weeds, and pests; spreading microgranular pesticides and fertilizers and even beneficial insects; and planting are among the diverse uses now being done using UAVs (Rao Mogili and Deepak 2018; Chen et al. 2021). Plant parameters obtained from UAV-based sensors can be used not only for monitoring crop growth conditions but also for crop response to differential treatment of nutrients and water (Yang et al. 2020). N status monitoring is important for farmers, as it leads to higher performance compared to field spectral data. Also, significant bands are identified using stepwise multiple linear regression and used for modeling to generate a wheat leaf N map of the experimental field.
PEER-REVIEWED CONTENT IS ONLY AVAILABLE TO ASPRS MEMBERS AND SUBSCRIBERS

FOR MORE INFORMATION VISIT MY.ASPRS.ORG
PEER-REVIEWED CONTENT IS ONLY AVAILABLE TO ASPRS MEMBERS AND SUBSCRIBERS

FOR MORE INFORMATION VISIT MY.ASPRS.ORG
PEER-REVIEWED CONTENT IS ONLY AVAILABLE TO ASPRS MEMBERS AND SUBSCRIBERS

FOR MORE INFORMATION VISIT MY.ASPRS.ORG
PEER-REVIEWED CONTENT IS ONLY AVAILABLE TO ASPRS MEMBERS AND SUBSCRIBERS

FOR MORE INFORMATION VISIT MY.ASPRS.ORG
PEER-REVIEWED CONTENT IS ONLY AVAILABLE TO ASPRS MEMBERS AND SUBSCRIBERS

FOR MORE INFORMATION VISIT MY.ASPRS.ORG
PEER-REVIEWED CONTENT IS ONLY AVAILABLE TO ASPRS MEMBERS AND SUBSCRIBERS

FOR MORE INFORMATION VISIT MY.ASPRS.ORG
PEER-REVIEWED CONTENT IS ONLY AVAILABLE TO ASPRS MEMBERS AND SUBSCRIBERS

FOR MORE INFORMATION VISIT MY.ASPRS.ORG
Car Detection from Very High-Resolution UAV Images Using Deep Learning Algorithms

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

Abstract

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

Introduction

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

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

Literature Review

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

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

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

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

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

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

Materials and Methods

Study Area

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

Data Acquisition

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

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

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

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

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

Method

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

Image Processing and the SFM Algorithm

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

The processing of the images obtained by the UAV was done in Agisoft Metashape software, a commercial software based on image matching and model generation using the SFM algorithm (Şasi and Yakar 2018). In photogrammetric software using the SFM algorithm, bundle block adjustment algorithms are used to optimize the projection errors between the images and the calculated point positions (Duran and Atik 2021; Senkal et al. 2021). The software using these algorithms and methods first aligns the images, performs a preliminary...
adjusting, and creates a sparse point cloud. For this, algorithms are used to automatically position the matching points between the images (Hamal et al. 2020).

**Feature Detection and Matching**

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

The SIFT algorithm improves the initial camera parameters using the corrected positions and orientations of each image as a starting point. A set of key points is determined in each image through feature matching algorithms such as SIFT, and these are stored as key points in a database (Hossein-Nejad et al. 2021). When processing a new image, the algorithm recognizes the image features and compares them with the features in the database to identify the tie points. The SIFT algorithm automatically detects and matches key points between multiple images determined by local pixel variances (Nesbit et al. 2018; Ulvi 2021).
this step, a low-density point cloud is created with the 3D coordinates of the key points (Westoby et al. 2012). We set the limit of points to detect per image to 80k and the limit of matching points to 20k. We used “high” alignment accuracy for all data sets. GCPs were manually defined to refine camera positions and marked on some images to aid optimization. We processed the height data as ellipsoid height (meters). We used the WGS84 UTM 36N coordinate system in the study.

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

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

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

<table>
<thead>
<tr>
<th>Positioning Ground Control Points (as GCPs)</th>
<th>Check Control Points (as ChPs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No.</td>
<td>Horizontal Errors (m)</td>
</tr>
<tr>
<td>-----</td>
<td>-----------------------</td>
</tr>
<tr>
<td>1</td>
<td>0.008</td>
</tr>
<tr>
<td>3</td>
<td>0.014</td>
</tr>
<tr>
<td>4</td>
<td>0.015</td>
</tr>
<tr>
<td>7</td>
<td>0.009</td>
</tr>
<tr>
<td>8</td>
<td>0.010</td>
</tr>
<tr>
<td>11</td>
<td>0.012</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>MAE</td>
<td>0.014</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.012</td>
</tr>
</tbody>
</table>

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

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

RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (f_i - y_i)^2}  

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

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

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

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

The true positive (TP) value represents the number of cars that are actually cars and correctly identified by the algorithm. False positive (FP) means the number of cars misidentified by the algorithm, not actually cars. False negative (FN) means the undetected cars by the algorithm. The real number of cars in the images is represented as N. Recall rate (sensitivity) is an important metric for evaluating algorithm performance. It represents how many of the predicted positive cars are actually cars as follows:

$$ \text{Recall} = \frac{TP}{TP + FN} $$
The F1 score combines precision and recall rates into single metric; this is the harmonic mean of these combined metrics:

$$F1 \text{ score} = \frac{2 \times TP}{2 \times TP + FN + FP}$$  \hspace{1cm} (4)

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

$$\text{Precision (UAcc)} = \frac{TP}{TP + FP}$$ \hspace{1cm} (5)

$$\text{Pacc} = \frac{TP}{N}$$ \hspace{1cm} (6)

Results and Discussion

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

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

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

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

Table 3. Car detection results for the very high-resolution UAV images using deep learning methods (Mask R-CNN and YOLO v3).

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>Recall</th>
<th>F1 Score</th>
<th>Precision (UAcc)</th>
<th>Pacc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mask R-CNN</td>
<td>388</td>
<td>301</td>
<td>39</td>
<td>48</td>
<td>0.862</td>
<td>0.874</td>
<td>0.885</td>
<td>0.776</td>
</tr>
<tr>
<td>YOLO v3</td>
<td>388</td>
<td>348</td>
<td>18</td>
<td>22</td>
<td>0.941</td>
<td>0.946</td>
<td>0.951</td>
<td>0.897</td>
</tr>
</tbody>
</table>

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

When the values in Table 3 and Figure 3 are examined, it is seen that both methods are successful, but the YOLO v3 approach gives better results than the R-CNN approach. It has been noticed that the R-CNN algorithm has difficulty detecting especially in places where shadows fall on the image due to the angle of the sun. However, although the algorithm was successful in detecting cars, it also detected other objects similar to cars as cars. The reason for this may be the light reflected from the car windows from other detected objects and the algorithm making an error for this reason. However, as you can see in Table 3, high accuracy rates were obtained, and the car detection was successful. Although YOLO v3 can be preferred because it gives more accurate results, programming ability is needed for the application of this method. However, the deep learning tool developed by Esri can
solve an engineering problem without the need for programming skills. Therefore, since Mask R-CNN can be used with commercial software, this deep learning approach is easy to use, although its accuracy is low.

Conclusion

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

References


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. IF SAR
   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 IF SAR 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 topobathymetric lidar. The remaining chapters are either relevant to all DEM technologies or address alternative technologies including photogrammetry, IF SAR, 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

<table>
<thead>
<tr>
<th>PRICING</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student (must submit copy of Student ID)</td>
</tr>
<tr>
<td>ASPRS Member</td>
</tr>
<tr>
<td>Non-member</td>
</tr>
<tr>
<td>E-Book (only available in the Amazon Kindle store)</td>
</tr>
</tbody>
</table>
After more than 15 years of research and writing, the Landsat Legacy Project Team published, in collaboration with the American Society for Photogrammetry and Remote Sensing (ASPRS), a seminal work on the nearly half-century of monitoring the Earth's lands with Landsat. Born of technologies that evolved from the Second World War, Landsat not only pioneered global land monitoring but in the process drove innovation in digital imaging technologies and encouraged development of global imagery archives. Access to this imagery led to early breakthroughs in natural resources assessments, particularly for agriculture, forestry, and geology. The technical Landsat remote sensing revolution was not simple or straightforward. Early conflicts between civilian and defense satellite remote sensing users gave way to disagreements over whether the Landsat system should be a public service or a private enterprise. The failed attempts to privatize Landsat nearly led to its demise. Only the combined engagement of civilian and defense organizations ultimately saved this pioneer satellite land monitoring program. With the emergence of 21st century Earth system science research, the full value of the Landsat concept and its continuous 45-year global archive has been recognized and embraced. Discussion of Landsat's future continues but its heritage will not be forgotten.

The pioneering satellite system's vital history is captured in this notable volume on Landsat's Enduring Legacy.

Landsat Legacy Project Team
Samuel N. Goward
Darrel L. Williams
Terry Arvidson
Laura E. P. Rocchio
James R. Irons
Carol A. Russell
Shaida S. Johnston

Landsat's Enduring Legacy
Member/Non-member $48*
Student Member $36*
* Plus shipping

Order online at www.asprs.org/landsat
LEARN
DO
GIVE
BELONG

ASPRS Offers
» Cutting-edge conference programs
» Professional development workshops
» Accredited professional certifications
» Scholarships and awards
» Career advancing mentoring programs
» PE&RS, the scientific journal of ASPRS

asprs.org