Masters Program in Geospatial Technologies

DETECTION AND MAPPING OF SOLAR PANELS BASED ON DEEP LEARNING, INSTANCE SEGMENTATION AND VERY HIGH-RESOLUTION MULTIBAND UNMANNED AERIAL VEHICLE (UAV) PHOTOGRAMMETRIC SURVEY.

Nnadozie Uzoma Onyeukwu.

Dissertation submitted in partial fulfilment of the requirements
for the Degree of *Master of Science in Geospatial Technologies*

NOVA IMS Information Management School
Universitat Jaume I
Westfälische Wilhelms-Universität Münster

TITLE PAGE.

Detection and Mapping of Solar panels based on Deep learning, Instance segmentation and Very High-Resolution Multiband Unmanned Aerial Vehicle (UAV) Photogrammetric Survey.

Dissertation submitted in partial fulfilment of the requirement for the master's degree of science in Geospatial Technologies.
Westfälische Wilhelms-Universität Münster, Germany (WWU),
Universidade Nova de Lisboa, Lisbon, Portugal (UNL),
Universität Jaume I Castellón, Spain (UJI).

By

Nnadozie Uzoma Onyeukwu.

Dissertation Supervised by

Prof Michael Gould Carlson

(Universität Jaume I. Castellón, Spain.) - Principal Supervisor.

Prof Joel Silva
(Universidade Nova de Lisboa. Lisbon, Portugal.) - Co Supervisor.

Prof. Ionut Sandric
(University of Bucharest. Bucharest, Romania.) Co Supervisor

February 2024.

ABSTRACT.

Solar energy is becoming a pivotal resource in the field of electricity generation as it seeks to create clean energy which makes the earth environmentally friendly. Unmanned Aerial Vehicle (UAV) on the other hand has become a low-cost data collection technology used for remote sensing in Geoscience. With 300 days of sunshine in the Kingdom of Spain and 3,321 hours of annual sunshine in Castellon, residents are eager to harness the solar energy at their disposal through the mass installation of solar panels. In light of this, there is an urgent need for studies on the accurate detection of solar panels with minimal edge loss and energy quantification of these solar panels in order to create a spatial data infrastructure that serves as a geospatial tool for urban planning and green policy implementation. A case in point is the spatial identification of solar clusters which can aid in the location of EV charging stations in urban areas. To this end, this thesis addresses these challenges by conducting a UAV photogrammetric survey and developing a deep learning model for the accurate detection, mapping, and quantification of solar panels in Castellon, Spain using ESRI ArcGIS Pro and Drone2Map. To ensure accurate detection and minimal detection edge loss, we make use of instance segmentation which combines the use of object detection and semantic segmentation. This method accurately delineates the boundary of each solar panel in the study area from a very high-resolution UAV multiband photogrammetry survey while minimising false detection through the addition of a normalised digital surface model. The deep learning model was trained on a record high 0.03m spatial resolution RGBnDSM UAV imagery with the state-of-the-art instance segmentation deep learning architecture called Mask RCNN on the ResNet-101 backbone in diverse weather conditions. This research was tested on real world scenarios, and we achieved a mean accuracy of 0.8445, a recall of 0.9162 and an F1 score of 0.8782 where a higher intersection over union was set at 0.75. Following these results, the deep learning model was applied to an urban planning use case to determine spatial distribution of solar clusters. The findings of this research contribute to advancing solar energy integration into urban landscapes with a robust and accurate geospatial framework.

TABLE OF CONTENTS.

Table of Contents

LIST OF TABLES.

LIST OF FIGURES.

DEDICATION.

This thesis is dedicated to the future of global dreamers who are resilient towards the pursuit of their dreams. This finished thesis is the result of resilience. It is also dedicated to the future of global clean energy.

ACKNOWLEDGEMENT.

My journey through this dissertation was met with great challenges which I am grateful to have learnt from. I am thankful to my parents Mr and Mrs Martins Onyeukwu who have supported me in prayers and finance through to this point. Mr Onyeukwu had kept an impressive tab with every step through the way and I am grateful.

My sincere gratitude goes to the principal supervisor of my thesis; Prof Michael Gould who trusted in my ability to go forth with this research. He ensured I had everything I needed for the success of this research while providing a calm and encouraging atmosphere for me to do more. His kind academic mentorship from the first day I attended class at UJI gave me more reasons to press forwards towards success. Thank you so much Prof Michael Gould. I am also thankful to Prof Ionut Sandric and Prof Silva Joel for their relentless time and effort towards this thesis.

Back home in Nigeria, I have left a professional family from whom I continued to learn about geospatial technology. I am grateful to the leadership, Officers, and Men of 203 ISR Group Yola, Nigeria for giving me the opportunity to hone my skillset in GIS. I am also thankful to Victoria Madedor, Prof Alaci Davidson, Prof Efobi Kingsley Dr Igbokwe Onyedika and Prof Chime Ndu for their encouragement to undertake this master's program.

During the first semester of this master's program, the Erasmus Mundus Nigerian community came through in support for me to ensure I completed this master's program, and I am grateful for their massive support. I am also thankful to Gloria Edo who worked tirelessly to ensure we had a great administrative support. While in Germany, Dr Brox Christopher gave me all the support I needed for the international mapathon I held at IFGI, and I am grateful for his kind support. I am thankful to classmates like Sunil Tamang, Babatunde Aremu and Nyi Nyi Nyan who also helped to make it a success.

This dissertation would not be complete without the help of Clare Lennon who proofread this work multiple times and provided emotional support through this research. On a lighter note, she now calls herself a deep learning engineer haven read the work many times. I am grateful Clare, Thank you so much.

I am thankful to my twin brother Ekezie, my baby sister Uloma and my elder sister Akunna. They have been amazing through this time. I am grateful to Fine Ogechi Ezuka, Thelma Bosah, Okoh Micheal, Prasad Madushanka, Ishola Oluwa Funmi, Adefisan Oluwatobi, Chidinma Ogbu, Obidah Joy, Anyadike Chiwendu and Eze Precious for their moral support through this thesis. Thank you all.

I am further grateful to the brethren of House of Refuge Intl Yola, Nigeria as well as its Protocol Team and the brethren of Holy ghost Consulate Castellon Spain for their kind support through prayers.

This thesis was possible with you all. Thank you!!!

INTRODUCTION.

Background.

With increased public attention on climate change and its effect on the environment, countries are beginning to look towards clean and renewable energy. A rise in population and industrialisation over the years has also brought with it exacerbating pressure on the need for increased energy consumption across cities with a shared cost which has accounted for the daunting energy crisis in Europe. The challenges of urban industrialisation have made solar energy gain popularity as an alternative to fossil fuels and a means to reduce the cost of energy consumption (Kannan & Vakeesan, 2016), especially in areas where no grid is set for electricity supply. The cost of energy does not correspond to the growth of the nation's economy as (Nasrallah et al., 2022) state that Germany and Hungary have experienced temporary energy tariff deficits in European countries with vulnerable economies such as Spain. With climate change in mind and cities growing exponentially, and an added city transportation riddled with carbon emissions that have not been solved, it only makes sense to adopt a carbon-free energy solution for cities as put by (Daus & Yudaev, 2017). With the move towards clean energy, there is a further need for cities to keep a global track-record of this adaptation, its database, and its estimated contribution to the national electricity grid as this would foster better urban and economic planning in cities. This data could prove crucial to urban planners, solar panel sales marketers, solar maintenance engineers and government taxation offices, who would like to give tax incentives to citizens who contribute renewable energy to their neighbourhoods. Mapping solar panels can also be effective for energy estimation and its associated financial incentive programs. For example, Fla. Stat. § 196.012(14) allowed citizens in Florida to enjoy a rebate of up to $20,000 for a single household and up to $100,000 for commercial installations for citizens who installed solar energy systems in their buildings. A further example of a tax incentive took place in August 2008, when North Carolina enacted a real estate property tax exemption equal to 80% of the appraised value attributed to the addition of a photovoltaic solar energy system to their buildings. Not only can such data be used to quantify solar energy, but also the location of solar panels in deep rural areas was used as a source of intelligence in 2015 by the defence intelligence community in Africa to determine location of terrorist activities in Gamboru Ngala, Nigeria. With these quantifications, mapping and location finding of solar panels, geoscientists will spend an ample amount of time manually

digitising solar panels in an entire state or country. The communication from the commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions **EU** solar energy strategy hold that solar panels provide 5% of total electricity in the EU and that the EU intends to increase its 2030 target to 45%. They assert that the full potential of solar energy can only be exploited when citizens are provided with the right incentives to be prosumers. Hence the committee encouraged EU member states to establish appropriate incentives to encourage solar installations. This underscores the need for and value of development of effective deep learning algorithms for the detection of solar panels.

Research Workflow and Objectives.

This research seeks to conduct an aerial Photogrammetry Survey over the study area located in and around the UJI campus, in Castellon, Spain. This survey included capturing high-resolution UAV imagery to serve as the foundation for subsequent analysis. Following data acquisition, the project focused on training, testing, and assessing an instance segmentation deep learning model utilizing the Mask R-CNN deep learning architecture. This phase involves systematically developing the model to accurately detect and delineate solar panels within the UAV imagery.

Subsequently, the trained model was applied to perform detection and mapping of solar panels across the study locations, to understand how well an instance segmentation model generalizes accurately. Finally, we determined the spatial distribution and quantification of solar panels present within the study area, providing valuable insights towards the improved time and accuracy performance of an instance segmentation model if trained in diverse atmospheric conditions.

Research Questions.

1. How does low lighting condition in aerial images affect image reconstruction and detection of solar panels for homogenous texture in UAV photogrammetry?
2. To what degree of accuracy can instance segmentation models be employed in the detection and mapping of solar panels?
3. How well can the deep learning model perform in a real-world scenario if trained with data from diverse atmospheric condition and tested in real world scenarios?
4. If solar panels are detected and mapped, how does their spatial distribution inform the installation of city furniture like EV charging stations?

Study Area.

The study area for this research is limited to an area covering and just outside the Universität Jaume I in Castellón Spain. The choice of the study area is unique to its climatic condition where Castellón receives 3,321 hours of annual sunlight which is greater than other communities of Spain. Hence the need for the development of a deep learning model to keep track of emerging solar panels by the inhabitants. The study area was also chosen due to the availability of equipment to carry out the research. Because deep learning models are known to perform better in the locations in which they are trained, the study area is also put forth by the unavailability of an instance segmentation deep learning models for the accurate detection of solar panels in Castellon as this thesis seeks to achieve as a contribution to science in the Kingdom of Spain.

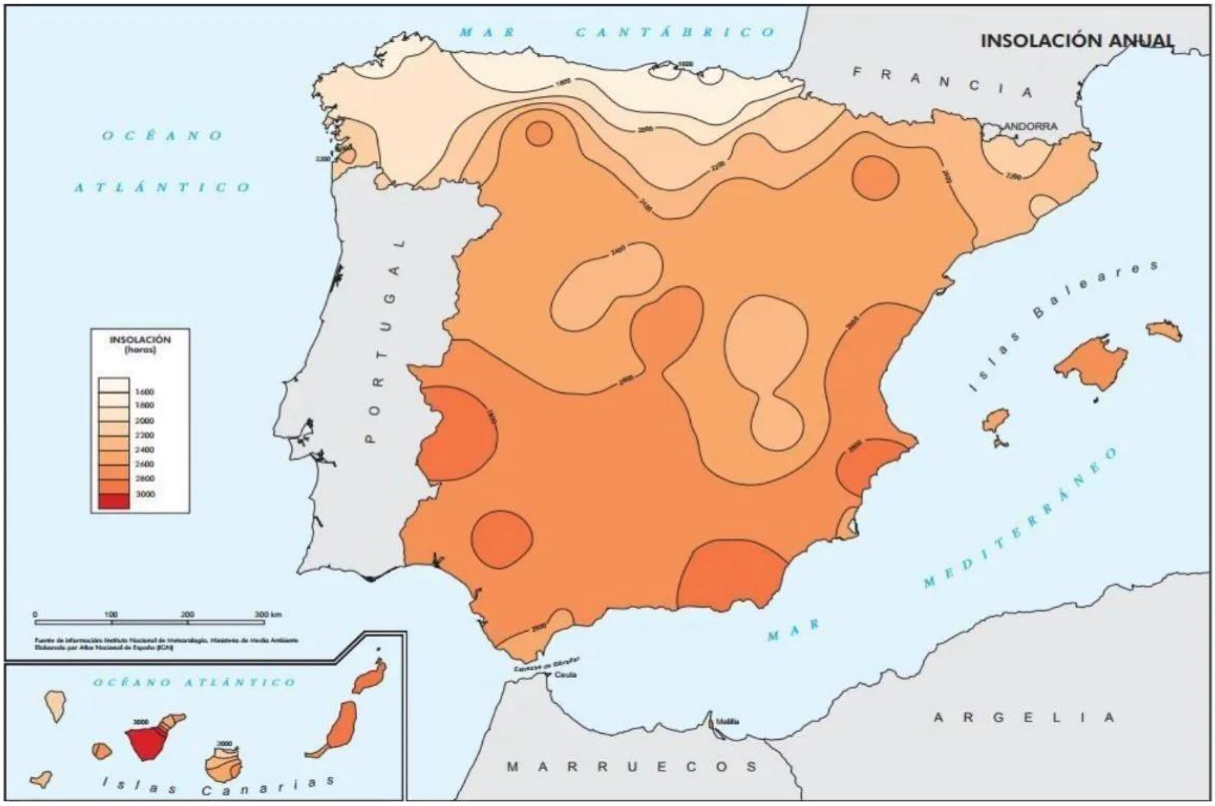


Figure 1 - Map showing areas in Spain with hours of sunshine.
Source – Quiero Sol Advisor 2023.

Significance/Justification.

This research seeks to adopt the use of a more precise deep learning architecture which utilises instance segmentation to accurately delineate the segments of solar panels in aerial imagery. Here, image segmentation uses a combination of object detection and image segmentation to give a more accurate delineation of the solar panels without losing pixels data. This research adopts instance segmentation because instance segmentation combine the principle of object detection and image segmentation without quaternization in the ROI pool layer as a more accurate and faster means for the detection of solar panels with less training dataset and less training time compared to other deep learning architecture which utilises just object detection or just semantic segmentation.

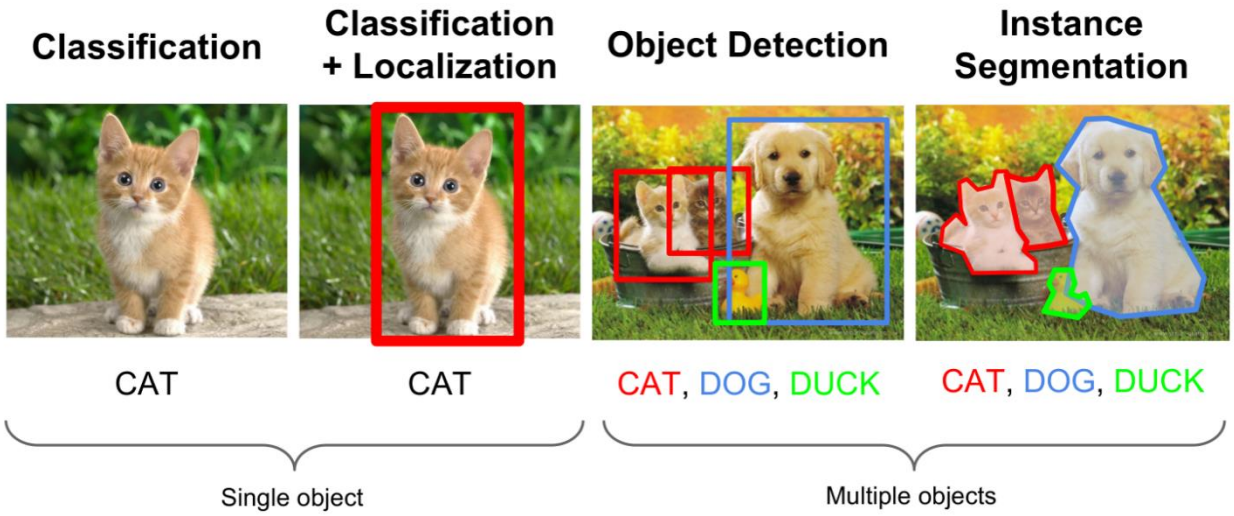


Figure 2 - Figure showing classical explanation of segmentation.
Source - https://imperialcollegelondon.github.io/recode-perceptions/index.html.

In previous research, image segmentation models have been known to detect solar panels but however losses fractions of the panels thereby making it difficult to compute information like electricity generated by each panel. Other research also utilizes object detection algorithm which detects solar panels but do not delineate the segmentation. The result of an instance segmentation model is an accurate bounding box delineating individual solar panels themselves (instance segmentation). In object detection, objects are detected but only from a large region of interest where the solar panels are located (object detection) while in image segmentation solar panels are detected however pixel information are lost resulting in low intersection over union as seen in previous research. The types, variant and associated variant in energy production of the instance of the solar panels can be further distinguished by this research. The diverse weather condition of

the study area for which the deep learning model is trained on, makes it a valuable model in other regions like Germany which has less sunshine and Morocco which can be very sunny. This research would address the need for solar panel maintenance companies who seek to know exactly where to focus their marketing strategy. The project would also be useful for military intelligence community, who would need it to infer terrorist locations. Government agencies would also use this project to identify people who contribute to clean energy if tax exemption needs to be given to such citizens and, most importantly, to quantify how many solar panels are mounted on these roofs and how much energy they contribute to the national grid. The research may be also beneficial to urban planners who could as well determine which parking lots could benefit by proximity to solar panel clusters. As data for training and testing deep learning models result have become scarce, this research contributes very high-resolution UAV dataset to Geoscientist for continuous study in geospatial deep learning.

A further justification of this research is that the deep learning model gives solar marketing firms the localized opportunity to define the scope for which they can channel their marketing efforts, thereby saving time and money with an assured return on investment. If the government decides to help its citizens detect defects in their solar panels, they should know where to look or else they would fly photogrammetric missions aimlessly looking in the wrong places. Consequently, accurate delineating boundaries of solar installations aids accurate flight planning for solar installation inspection projects. The accurate detection of solar panels can help cities progress towards the adoption of green building practices, including a better ability to apply for EU funding. With the localisation of solar panels, policy makers can optimise power distribution and energy storage infrastructural planning. This research might also contribute to an urban green zone map or solar friendly urban map.

LITERATURE REVIEW.

UAV Photogrammetry in Geoscience.

With the advent of technological developments in remote sensing research, (Nex et al., 2022) indicates that over the last two decades, UAVs have become the most relevant emerging technology in the discipline of Geoscience and Remote Sensing. According to (Nex et al., 2022) this technology offers the user the possibility of having economical (Samad et al., 2013), high-defined spatial (Samad et al., 2013) and (Yao et al., 2019) temporal resolution imagery. They all opine its usefulness for the range of remote sensing problems such as infrastructural improvement, image classification (Sunarya et al., 2020) semantic segmentation, and object detection with a focus on instance segmentation, upon which this research is based.

Despite the increase in the usage of UAVs in recent times, the aviation industry and Electronics industry have embraced the application of UAVs even more so than geoscientists (Samad et al., 2013) with (Nex et al., 2022) finding that relatively limited research has been done on the adoption of deep learning for semantic scene analysis using UAVs photogrammetry. Although (Samad et al., 2013) declares UAV technology to be a revolution in the geoscience industry, it still faces complex challenges with regards to its analysis. (Nex et al., 2022) and (Yao et al., 2019) purports that due to the high resolution of GSD obtained from UAV photogrammetry, challenges exist over the extraction of semantic information. For example, ultra-high-resolution imagery would create more unwanted classes during image classification making it difficult to map land use accurately. UAV technology has not yet been able to overcome the challenges of motion blur and its inability to overlay objects in homogenous scenes accurately like seas, trees, vegetation, and beaches as (Samad et al., 2013) asserts.

Previous research has shown that UAV photogrammetry has been used to conduct research for studies in urban development analysis (Iheaturu et al., 2022) quantification of mudflat Morpho dynamics (Chen et al., 2022) production of digital base maps (Sunarya et al., 2020), archaeological surveying (Saleri et al., 2013) aerial terrain mapping, digital terrain model mapping (Udin et al., 2012) and, most recently, to study environmental impact assessment (Olivatto et al., 2023). Novel research in gravimetry (Luo et al., 2022) has also emerged from the principles of UAV photogrammetry. Even given the vast number of studies in UAV photogrammetry, it has thus far not been used to survey solar panels on building rooftops as this thesis seeks to achieve.

(Anuar et al., 2023) presented the potential of UAV technology for UAV photogrammetry in GIS by placing a small format Pentax Option A40 digital camera underneath the wings of the Cropcam UAV to collect photographs over their study area. In their research, the images were collected with a frontal overlap of 60% and a 30% side lap. To ensure accuracy for the GIS study, they established ground control points using the rapid static technique of GPS and check point. The convergent configuration of the UAV camera was employed to strengthen the geometry for successful recovery of the focal length. They calibrated the camera by taking 10 photographs from 5 different locations of the test field. During the flight, they disabled automatic focus and enabled the infinity focus of the camera with two images taken in NIDIR and Oblique per location over a retro reflective target. To avoid oversaturation on the retro reflective target, the flashlight of the camera was covered with a piece of tissue paper to filter the flash such that an adequate amount of light would be transmitted. To process their imagery, the ERDAS Imagine processing software was used where an orthophoto, Digital terrain model and a digitized vector map were generated. The result of their study showed that lens distortion parameters K1, K2 had a significant effect on their photogrammetric result. Also, an orthophoto was produced from images that had 60% frontal overlap and 30% side overlap alongside a digital elevation model and 3D stereo model. Although this research was successful, it experienced certain UAV photogrammetry challenges that are considered worthy to note. Anuar reports that due to crabbing - a condition where an aircraft drifts from its flight plan - the orthophotos were not promising. They strongly advocate that wind plays a vital role in UAV photogrammetry as this could cause crabbing leading to the overall ineffectiveness of the application of UAV photogrammetry product for geographical information system studies. Their study proved that small format digital cameras can be used in UAV photogrammetry for GIS study as it addresses the gap in the use of small format digital cameras in UAV photogrammetry. However, they give that small format digital cameras need to be calibrated to achieve high geometric results especially if the project area does not require high accuracy and is of low budget.

(Iheaturu et al., 2022) conducted a simplified approach for mapping and performing urban analysis by utilizing structures from motion photogrammetry. In their research, they aimed at developing a systematic workflow that could be adopted by geoscientists and urban planners for quantifying urban development. Their means of quantification was to assess the heights of buildings to determine if the land in the study area was being used parsimoniously. Building upon the work of

(Anuar et al., 2023), Iheaturu goes further to justify their reason for the choice of UAV used for UAV photogrammetry. (Iheaturu et al., 2022) states that the choice of UAV for the study is justified by its ability to generate photogrammetric products useful for their study. These products are orthophotos from which digitization can be made to produce a vector map and a digital surface model from which building heights can be computed. To collect data for this study, Iheaturu used a low-cost DJI Mavic 2 Pro UAV which features a focal length of 35mm, a sensor of 20 megapixel (1" CMOS), hovering vertical: ± 0.1 m (when vision positioning is active), ±0.5 m (with GPS positioning) and horizontal: ± 0.3 m (when vision positioning is active) and ±1.5 m (with GPS positioning) which allows the UAV to acquire geolocated images. To produce the model with good accuracy, images were acquired with an adequate overlap of 60% front overlap and 75% side overlap with nine targets evenly distributed over the study area. The targets were made from PVC while the GPS coordinates were measured using a Hi-target V30 GNSS unit in real time kinematic mode. For flight planning, Iheaturo used the Drone Deploy flight planning software to plan waypoints that aircraft would travel to collect data. Their flight parameters include a flight height of 61m which was above that of the highest obstacle of 55m in the study area. At this altitude, the UAV captured a total of 470 images covering 22.5 ha under 25min 41s with an overlap of 60% forward and 75% side lap. Because their research needed to make spatio-temporal comparisons of building growth, they used the ELSHAYAL Smart GIS software to download 2005 imagery of the study area. To process their data from the UAV, they used the PIX4D Mapper software to align the images, and to produce a sparse 3D point cloud reconstruction of the images they used the inbuilt SiftGPU algorithm to identify match feature key points throughout the entire photo. They adopted a different method for imagery accuracy by using the rayCloud Editor tool in PIX4D. The generated quality report shows they achieved a georeferencing accuracy of < 0.1 m which falls within the photogrammetric allowable limit. Since they needed to compare two images, they purport that it was essential for the two images to have the same coordinate system. To achieve this the UAV imagery was referenced to the local datum of the study area while the Google Earth imagery which was collected in WGS 84 was also transformed to the local coordinates of the study area using the transformation techniques in ArcGIS Software. Once all the imagery had been transformed and was ready, they performed an on-screen digitization to extract building footprints, roads, and plot features for urban analysis. They express that due to the size of the study area and the level of precision needed for digitization, the automatic and semi-automatic classification

method would not be ideal for extracting those features. Upon extraction, they classified the extracted features to show development status in terms of tarred and untarred for roads, developed and non-developed for plots. With this, they used the building heights to ascertain parsimoniosity of land alongside building count ratio and plot status to quantify urban analytical changes that occurred between 2005 and 2009. The result of their study shows that all position errors of their ground control points fall between 18.73 mm to 49.84mm which is within the Royal institute of charted survey acceptable limit of less than 80mm. Conversely, they generated an orthophoto and digital surface model with a ground sampling distance of 1.43cm which is largely affected by the flying height of the UAV. They inversely opine that UAVs flying at a lower height would generate better spatial resolution. Their result showed that there were 8 new buildings per year, 3 plots developed per year, an increase of 7m road per year and an overall decrease in underdevelopment of land by 11%. Their research proves that UAV photogrammetry again can be successfully used in urban analysis towards the study of parsimonious use of land.

(Chen et al., 2022) took a different approach in UAV photogrammetry for geoscience but this time without ground control. This is a move away from the approach chosen by (Iheaturu et al., 2022) and (Anuar et al., 2023) who used ground control points for their UAV photogrammetry. The main objective of the study by (Chen et al., 2022) was to evaluate the potential of RTK – assisted UAV photogrammetry without GCP in quantifying intertidal mudflat morpho dynamics. They state that with challenges facing the setup of ground control points in monitoring coastal environments, UAVs with onboard RTK-GNSS can also achieve high level of geodetic accuracy like the establishing ground control points. Prior to this study, the potential of this process was not known. For this research, Chen used a DJI Phantom 4 RTK drone to collect photographs of the study area. The lens of the drone had a focal length of length of 24mm, a CMOS sensor of 1inch and pixel size of 20 mega pixels. They performed a rigorous calibration on the drone lens to measure radial and tangential distortion. Instead of establishing ground control points, they connected the RTK drone to the RTK network base which provided centimetre level positioning accuracy. Although ground control points were not installed, wooden squares of 40 x 40cm were established as checkpoints for accuracy assessment and measured with a differential GPS. For flight parameters, they set their flight height to 120m with 80% frontal overlap, 80% side overlap and a camera angle of NADIR from which they collected 3214 images. Chen expressed the importance of flying on a

cloudy day to avoid the effect of sunlight reflection on mudflat surface. All images in their research were processed using the PIX4Dmapper software and root mean square error was determined from the RTK assisted UAV Photogrammetry without ground control points. To determine the spatial characteristics of annual erosion and accretion that has occurred in the mudflat between the two years under comparison, they used the pixel difference across the pixel location. The result of their study shows that the accuracy of their UAV photogrammetry gives a centimetre level accuracy. They agree that performing UAV photogrammetry without GCPs falls short of accuracy performed by previous studies that achieved millimetre level accuracy, but they express optimism that that an RTK assisted UAV can obtain centimetre level topographic data and would save more time when collecting data for large topographic data. The result of their research shows that the digital elevation of the UAV photogrammetry detected small elevation changes that affect water flow and sediment deposition. They assert that UAV photogrammetry can help ecologist map areas where salt marshes survive for further restoration of the vegetation colonization.

At the 2020 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM), (Sunarya et al., 2020) presented a study on digital maps based on unmanned aerial vehicles to study the performance of photogrammetry software. In their research, they expressed worry over the overpriced photogrammetry software available to perform photogrammetry and tried to compare them with open-source photogrammetry software. They speak firmly of UAV as a medium that has largely helped map making. To begin this research, Sunarya made the flight plan by determining the way points and flight region of the study area. This differs from (Chen et al., 2022) who collected data in a cloudy atmosphere. (Sunarya et al., 2020) took a different approach towards the time of the day for their flight by collecting data between 6am and 8am during sunny weather. They argue that this was done to keep the lighting obtained during flight consistent across the study region. For data collection using UAV, a DJI Phantom 4 pro was used with flight speed between 5 to 7km per hour, camera angle at 90 degrees, an altitude between 50 to 100 meters considering the building height in the study area and an overlap of 70 to 80 percent. This flight parameter resulted in 31 sorties and 3703 aerial photographs taking the team 16 days with 2 missions per day. During their research, they expressed challenges of insufficient batteries, stating that having enough batteries during flight could have limited the number of days spent collecting data. Another challenge they faced was memory overload after

the ortho mosaicking ran for 34 hours on an Intel Core i7 9750H GPU NDVIA GEFORCE GTX 1650 8GB. This made them use only 76 aerial images to perform the photogrammetric process of mosaicking. Upon completion of the image processing, they compared the performance of Open Structure for Motion and Agisoft PhotoScan. They found out that despite the long processing time by Agisoft, the software performed relatively better in terms of image resolution. But in terms of the software ability to match homogeneous images as tie points to generate the orthophoto, Agisoft PhotoScan performed poorly with 24.53% holes detected in the imagery while Open Structure from Motion had 2.0690% detected holes. Again, to compare the distance in error of the coordinates, they used the haversine equation which they argue is commonly used to express the distance between two points in a spherical object. The result of their experiment shows that there was an error of 0.650285m when comparing points taken from the ortho photo and on the field. The result of their research also shows that the geotag results on the map have an average error at the latitude of 0; 00032%, and an average error on longitude 0; 00024%. Although their research gives results, it is not clear what coordinate system was used in making this comparison and what coordinate system was used in aerial data collection and field data collection. Neither was it stated why an error of 0.6m occurred, especially where (Chen et al., 2022; Iheaturu et al., 2022; Anuar et al., 2023) had all established ground control points and achieved centimetre-level accuracy.

(Zerafa & Azzopardi, 2022) have tried for the first time to use UAV photogrammetry in the study of electricity. They argue that with the growing policy of incentives towards PV installations in Malta, it is essential for government and financial institutions to have ample information on the potential energy a PV system installation can generate before allocating funds and incentives to owners of these systems. Their study was aimed at providing a simple method for estimating electricity generation potential based on orthophoto generated using Unmanned Aerial Vehicle, geospatial algorithm, and photovoltaic data. Spatial data was collected for this research using the DJI Matrice V2pro drone and an RGB camera. During multiple flights they collected the data with the camera set to NADIR and oblique at 45 degrees. Each image had a size of 4000 x 3000 pixels, a ground spatial resolution of 3cm, a flight height of 50 meters from the ground surface which was a different approach from (Sunarya et al., 2020) who collected photographs with building height in perspective. Sunarya set their overlap to 80%. To estimate how much energy is generated from the available roof, they calculate the number of buildings and number of units within each building.

Given that they regard each unit as one roof, they used the total collective unit to find the cumulative surface area available for PV system installation. A geo-orthophoto tool from PIXAL Ltd was used to create the Ortho mosaic. Energy data per meter squared was generated from three PV systems in three different locations. This data was logged on to a cloud server daily over a period of two years using a software from UNITENERGY Ltd. The data from this PV system was used to calculate the average energy generation potential for a square meter. The result of their research shows that the UAV photogrammetry demonstrated that the study area had a total available area of 64,703m sq to host PV systems. To obtain their result, they divided the total energy generated from the PV system by the total surface area of the panels, giving a constant of 20.56m2. Their research shows that a total of 7.17 GWh of electricity can be generated from the study if the buildings in the study area were filled with solar panels over their rooftops. The research presented by (Zerafa & Azzopardi, 2022) showed a promising use of UAV photogrammetry for energy potential estimation however their research was not concerned with the geodetic accuracy of the imagery collected from the drone.

(Ali & Abed, 2019) focused their research on the impact of flight parameters on UAV photogrammetry for something they termed mobile topographic mapping. They achieved this by establishing 13 ground control points and observed 4 ground control points while correcting the rest using the OPUS website and TOPCON tool office software. A DGPS type TOPCON GR5 was used to make these observations in static mode. To collect imagery for their experiment, a DJI Phantom 4 Pro with a camera specification of 1 inches CMOS and 20 mega pixels. Utilising the autonomous function of the drone, they planned their flight using the Pix4D capture software. Five different scenarios were established with altitudes of 40, 60, 80, 100 and 120 metres from which only three were selected for the research. Ali and Abed give that PIX4D and Agisoft Photoscan are the two most popular software used in photogrammetry. This is supported by the fact that (Chen et al., 2022; Iheaturu et al., 2022) also used the same software in their research. They give that the software uses the internal meta data of the images to create a reasonable accurate geo rectified orthophoto imagery without ground control points. They also affirm that where accuracy in geodetic accuracy is a priority, ground control points need to be established. The impact of image orientation was also tested in NADIR and 70 degrees oblique.

Object Detection Using Deep Learning.

Research in object detection is a complex computer vision task with evolving science for improvement. (Wang et al., 2022) state that the overall aim of object detection from satellite imagery or UAV image is to localize the object within an image. Such localization makes image detection a core geospatial task to undertake. Various researchers in the field of computer vision and remote sensing have studied the use of deep learning architecture in the detection of features and objects, whether it is a swimming pool, a crop type, a football field, or a terrorist vehicle, etc.

A machine learning model trained on virtual laser scanning data was used by (Zahs et al., 2023) in the detection of structural building damages. In their research, they simulated pre- and post-event point clouds through UAV-borne laser scanning of virtual scenes. Using k-mean clustering, their research identified coarseness in changed and unchanged buildings, which helped in the extraction of robust object-specific change features. Training and classification of point clouds, object-specific change features, multi-class building damage was done using the random forest machine learning model. The result of their classifier when applied to the real-world evaluation dataset yielded high classification accuracies for the target damage grades with overall accuracies of 92.0%–96.8% and an F1 score of 73.2%–94.6%).

In the work of (Bazi & Melgani, 2018), they used a novel convolutional support vector machine network for the detection of vehicles and solar panels from a 2cm UAV imagery. Here, they built the network using the linear support vector machine as the convolutional filter for generating the feature maps. They proposed a configuration with five convolutional layers, three reduction layers, and a classification layer. To efficiently compute the weights of the SVM filters, they used the multicore liblinear software package and a penalty parameter. Instead of backpropagation for learning weight, they used forward supervised learning. The result of this method when compared to a pretrained convolutional neural network achieved an accuracy of 93% in detecting the location on the image where the solar panel is located, but not a bounding box characterizing the panel itself.

Geospatial deep learning for solar panel detection.

The all-encompassing efforts to recreate statistical models that act like the human brain with an ability to make predictions based on historical data have given rise to studies in artificial neural

networks. (Wei & Qingna, 2021) and (Topol, 2019) opine that an important element of artificial neural networks is data. Given that such data can be geographic, studies in artificial intelligence can also be applied to geography. The nexus between artificial intelligence and geospatial technology is based on the phenomenon that spatial is special. (Bookstein, 1991) gives that spatial is special due to its application in geo-statistics and spatial point patterns. Stating further that since every observation or measurement taken on the surface of the earth is done with the knowledge of its location, such observations or measurements are spatial data and can be studied using artificial neural networks to solve geographic problems. (Gao, 2021) further gives that geospatial artificial intelligence is a field of study which concerns itself with mimicking the processes of how the human brain perceives things, how it reasons and make discoveries about geographical phenomena to solve problems of human interaction with its environment through the development of computer programmes.

In the context of data as a driving force for artificial intelligence and neural networks, (Janowicz et al., 2020) affirm that the availability of high-quality imagery as well as software and hardware to process them have given rise to the advances in geospatial artificial intelligence. The availability of high-resolution imagery, high performance computers and deep learning has fostered advances in object detection. This is a core point of study in geospatial science, just as this master's thesis seeks to achieve. The common challenge in geospatial artificial intelligence is a lack of training data as stated by (Li et al., 2021). Another challenge put forward by (Gao, 2021) is the exposure of field conditions and equipment expenses, limiting the training of geospatial professionals as well as the unavailability of high-performance machines for developing, training, and testing of geospatial deep learning models (Hu et al., 2022).

Despite these challenges, previous researchers (Castello et al., 2019; Hu et al., 2022; Kausika et al., 2021; Li et al., 2021; Malof et al., 2016; Yuan et al., 2016) have all employed various techniques in applying geospatial artificial intelligence toward the detection of solar panels as this master's Thesis focuses on. (S. Wang & Li, 2021) have studied geospatial artificial intelligence in terrain analysis for natural feature extraction. (Laxmi 2019) also studies geospatial artificial intelligence for detecting ships, (Yue et al., 2022) for earth observation, (Usery et al., 2022) and for topographic mapping. Despite the significant number of studies in artificial intelligence,

geospatial artificial intelligence and geospatial deep learning has thus far not been vastly explored in the detection of solar panels on building rooftops as this thesis seeks to achieve.

(Malof et al., 2016) carried out a novel study in geospatial artificial intelligence to automatically detect solar photovoltaic arrays from high resolution aerial imagery. In their research, they state that information about the location of solar panels, their power capacity and energy potential are now sought after by government agencies, utility engineers and third-party decision makers. They agree that this information can be acquired by means of survey and interconnection filling. However, they stress that these methods of obtaining information about solar panels are arduous and economically damaging. In their work, they investigated the use of computer algorithms to extract solar panel information from high resolution imagery of 0.3m ground spatial resolution. In this novel research, they utilised a large data set of orthorectified RGB aerial imagery collected within the same month over California with a total coverage of 135 km Sq. The images had a size of 5000 by 5000 pixels each and a total of 600 images. Malof explains that due to the high presence of solar panels and availability of aerial imagery, the study location of California was used. In their research, they manually annotated all the visible solar panels totalling 2,794 solar panels to develop an effective computer algorithm. To avoid a positive bias on the algorithm, Malof splitted the 135km Sq imagery into training and testing at a ratio of 2:1. To carry out the analysis, they performed a feature extraction process that takes the 3-channel RGB image as an input and returns an M-channel feature image using a window size of 3 x 3 pixels because it roughly corresponds to the size of the individual solar panels to extract pixels that characterise the colour, textures and other patterns surrounding the pixel. With the supervised machine learning model, they employed a random forest classifier model that assigned probability or confidence values to each pixel, indicating the likelihood of the presence of a solar panel by training it with roughly 500,000 annotated solar panels and 5 million pixels. They argue that the random forest classifier has the advantage of learning complex nonlinear relations between input and output variables with a high computational speed. Solar panels were detected by identifying groups of adjacent or neighbouring high confidence pixels as detected objects. Although the classifier performed fairly, they post-processed by extracting only pixels with high confidence value.

The results of their research showed that the random forest classifier was effective at discriminating non panel pixels from panel pixels. They show that with 0.07% of the pixel in the

testing data, the detector achieved a P value of 0.0007. The detector detected 90% of the targets with a J value of 0.7 as baseline for future improvements. Their novel work shows that the algorithm was highly effective at detecting solar panels on a per pixel basis and object basis. However, it performed poorly in mapping the precise shape of the solar panel array.

Building upon the work of (Malof et al., 2016) who used a supervised learning random classifier model, (Malof et al., 2017) presented an improved work at the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Their improved research was a deep convolutional neural network, which leverages on pretraining for detection of solar panels array in aerial imagery. Dataset and manual solar panel annotation used in this work were the same as in (Malof et al., 2016). However, to train the convolutional neural network, they extracted a large number image patch that corresponds to each class and designed the image extractor to obtain roughly 2.5 million total training patches. They obtained non-PV training images by sampling every fifth pixel in the (Malof et al., 2016) imagery and retained only locations that did not intersect with a PV annotation. They further sampled every 3rd pixel of the training imagery data set and retained only pixel locations that intersected with solar panel annotation. To reduce computational load, they randomly sampled only 30% of the images containing solar panels and duplicated it 4 times. Their research utilises the convolutional neural network, a deep learning architecture consisting of a visual geometry group module and a fully connected neuron module which is followed by a rectified linear unit (ReLU) activation. This architecture relied on gradient descent where they used a batch size of 64, a learning rate of 0.001, a momentum of 0.9 and a weight decay of 0.0001. They trained their model consisting of 16 epochs and used 10% of the total training imagery for validation. They trained three different CNN weight initialisations.: a random initialisation network and two other networks initialised with pretraining. The result of their experiment was scored using the Jaccard index scoring matrix and they found that all three using the convolutional neural network performed better than (Malof et al., 2016) on the same data used in (Malof et al., 2016) suggesting that the CNN architecture chosen for this problem is effective and superior for image recognition. Of all three CNN initialisation procedures they report the pretrained weight (fixed) network performed the worst while the random initialisation detector achieved a precision score of 0.95.

(Golovko et al., 2017) studied a deep convolutional neural network for detection of solar panels. In their study, they tried to solve the problem of determining the presence of solar panels in the image and their detection by obtaining the exact coordinates of the location. In this research, they used a low-resolution 200 by 200-pixel photo from Google Maps to train their proposed model while the Faster R-CNN deep learning architecture was deployed. They split the research task in two by first defining the images which had solar panels using the convolutional neural network with six layers. A second task was then to detect the solar panels using the faster R-CNN deep neural network based on the ResNet – 50 classifiers. In this research Golovko had a sample of 3347 photographs. For the detection task, 1000 images containing solar panels were used from which 800 were used as training set and 200 images for testing. To determine the images with the presence of solar panels they manually sorted the images by assigning them into two groups: containing and not containing solar panels. For the second task of panel detection, they manually digitised the solar panels rectangularly. They expressed challenges that some images could not be digitised because the solar panels were located at the corners of the images. The neural network was trained with 70 epochs, a K values of 0, a learning speed of 0.001, momentum parameter of 0.9, weight decay of 0.0005, a batch size of 20 and a drop out with probability of 0.5. The training data achieved an accuracy of 87.46% and ROC curve of 0.92. The trained model was then operated on the testing data using the faster R-CNN deep neural network based on the ResNet-50 classifier trained on 5000 iterations. The trained model achieved an AP result of 0.929 on the 200 testing images, an indication of the model's high generalisation ability. Despite the success of this research, (Golovko et al., 2017) still faced the inability to have their model detect the boundaries of the solar panels in the photograph as this thesis seeks to undertake. This is a similar challenge to that faced by (Malof et al., 2016).

In Switzerland, (Castello et al., 2019) carried out research on the automatic detection of rooftop solar panels using convolutional neural networks. They argue that apart from the United States, no country has a comprehensive solar panel database covering the entire country. In a move away from the method employed by (Golovko et al., 2017) who used a faster R-CNN and (Malof et al., 2016, 2017) who used a random forest classifier, (Castello et al., 2019) used the U-Net deep learning architecture and assert that it is the most popular CNN architecture used in image segmentation problems. Their research was aimed at creating a comprehensive database with

location and size of existing solar panel installations in urban areas. To complete this research, they used an ortho rectified 8-bit RGB image of Switzerland provided by the Swiss Federal Office of Topography in TIF format, with a tile size of 17500 by 12000 pixels and spatial resolution of 0.25m. This research work is the first-time imagery in geoscience had been used in the study of the detection of solar panels. They express however that due to the large tile size of the imagery, it became difficult to detect the solar panels hence they further divided the tiles into 250 x 250 pixels in PGN format. They employed the use of an open cv-based tool over 700 images to manually segment and classify the solar panels to create the labelled samples unlike (Golovko et al., 2017; Malof et al., 2016) who manually digitised the solar panels to create labelled samples. In their work, they stress the importance of choosing a good loss function in image segmentation. After experimenting with various loss function pair values ranging from (1, 1.5) up to (1, 9) they discovered that the best convergence was a loss function pair of (1,5). Unlike (Malof et al., 2017) who starts from a pre-trained algorithm, (Castello et al., 2019) trained their algorithm starting from a random set of weights: 80% training data, 20% testing data, a batch size of 32 images, a learning rate of 0.1, a discovery optimal epoch of 75, and a dropout rate of 0.2. Their training model algorithm achieved a precision recall curve of AP = 0.85 and an F1 score of 0.80, model accuracy of 0.93 and model validation of 0.95. Although a few spikes were seen in the graph, they explain that increasing the data set or image batch size could resolve this. Upon successful training of the algorithm, they applied the trained algorithm to random testing images. It was observed that the algorithm localised well and produced boundaries across the solar panels. (Castello et al., 2019) opine that intersection over union is a robust matrix for assessing model performance of object detection algorithms as against the accuracy matrix. The result of their research shows an IOU score of 0.64 and an accuracy of 0.94 when the probability threshold of the prediction array is set to 95%. Their result shows optimism in that when a new imagery of Switzerland is released, it can be used to track the evolution of solar panels towards the establishment of an updated national database.

(Hu et al., 2022) considered that the unsatisfactory solar panel detection results have led them to embark on research where they proposed the combination of residual network and channel attention to improve the faster RCNN framework. Although (Hu et al., 2022) used the same faster RCNN deep learning architecture as used by (Castello et al., 2019; Golovko et al., 2017), they had

a different approach to their research. In their research, they employed the clipping input method to replace the down sampling method in faster RCNN. Secondly, they replaced the VGG 16 backbone for feature extraction with ResNet50 and creatively introduced the channel attention mechanism. Unlike other researchers (Hu et al., 2022) expressed a lack of high-resolution imagery dataset which had made them create a novel solar panel dataset. This research has however not provided information on its novelty in aerial photography. The dataset consists of 150 aerial images with a uniform size of 4000 x 3000 pixels. They split them randomly into 120 images for training and 30 images for testing. They pre-processed the images by clipping areas that were not solar panels out of the images and retaining only images containing solar panels. The resulting operation gave rise to a larger training dataset of 1493. To train the network, they set the number of anchor boxes to 400, an ROI box of 550, an input box of 440, an aspect ratio of (4,8,16,32), a sliding window of 1000 x 1000, learning rate of 0.0001, a batch size of 1, a momentum of 0.09 and a weight decay of 0.0005 implemented in pytorch on an Nvidia GTX1050 Ti. i7-7820X CPU, 32 GB of RAM computer. The algorithm was trained across VGG16, ResNet50 and ResNet101as backbone. The results of this training showed that the ResNet50 achieved an accuracy of 0.8709 which is reported to be 1.09% and 1.67% more than the VGG16 and ResNet101. But when the SE module was introduced, the ResNet50 and ResNet101 achieved a higher accuracy result by 0.44% and 1.05% higher respectively. They assert that upon comparison with the SSD, the faster RCNN algorithm performed better than the SSD while the improved faster RCNN used in the paper outperformed both. For the first time in the detection of solar panels, this research quantified solar panels. The result of their analysis gives a count of 228 panels. The trained algorithm was tested on six random aerial images. The detection probability reached 88% at the lowest and 100% at the highest indicating that their proposed method is superior. With the excellent success of this research, (Hu et al., 2022) did not demonstrate how they performed an accuracy assessment on the test imagery. Neither was the method for data collection addressed in this paper. Again, their research proves a high accuracy rate., It would be interesting to see the spatial distribution of these panels as no other researcher has yet done this.

In (Malof et al., 2019) they proposed a general framework for the mapping of solar panels and the attributes such as size and power generation capacity. In this research they developed a deep learning model packaged in a software called SolarMapper which for the first time focuses on

actual spatial mapping of solar panels geographically. They show that SolarMapper generates a pixel-wise labelling of the solar arrays in overhead imagery alongside its attribute data and they also demonstrate its ability to map solar arrays over the entire US state of Connecticut stretching 14,000km2. In training a deep learning model, (Malof et al., 2019) assert the importance using a large and diverse dataset. In their research, they trained their SolarMapper over 400km2 of imagery across three cities in California with a hand labelled annotation of 16,000 solar arrays. They argue that as at the year of the research, their dataset holds the largest hand labelled solar panel data. To assess the performance of SolarMapper, they employed the use of cross validation and intersection over union. They report that SolarMapper achieved an intersection over union of 0.67, a precision score of 0.76, a Recall of 0.077 and an F1 score of 0.76. In this research, the SolarMapper was deployed in a new environment in Connecticut. The researcher suggests that deploying SolarMapper was not a trivial endeavour. They further explain that such triviality can be caused by the differences in characteristics of the imagery deep learning algorithms are trained in versus the imagery they are deployed on. These characteristics may include a change in landscape, different urban structures, changes in lighting conditions or more often changes in camera perspective. To solve this common supervised machine learning problem, they leverage the use of transfer learning by fine tuning the training parameters of SolarMapper such that it generalises properly in the new environment where it is deployed. Upon transfer learning, they deployed SolarMapper in the new location. Its accuracy assessment was assessed by the human inspectors looking over each prediction one after the other. The researcher gives that this method is faster than the intersection over union. Their results show that SolarMapper gives a very good overall performance with precision score of 0.88, a Recall score of 0.83, and an F1 score of 0.85. From the detected solar panels, this researcher went further to estimate and map the energy generation capacity in Connecticut over 168 cities. They argue the inference that the power generation capacity of a solar array is proportional to its surface area. With this, a linear regression model was formulated to predict the capacity of solar array. In addition, they used the electricity generation constant and surface area. The result of their model upon assessment using the person correlation coefficient give a value of 0.88 and a p-value of 0.01. The estimated energy per city was then produced in a map to show the spatial distribution of solar energy production in Connecticut. For the first time solar panels and their estimations have been mapped spatially across geographical areas denoting energy capacity across the state.

In the works of (Yuan et al., 2016), they trained a convolutional neural network with the aim of mapping solar panels from aerial images. They trained the algorithm on the San Francisco CA and Boston, MA orthorectified RGB aerial imagery with a spatial resolution of 0.3 meters. The imagery data set was reported to have been collected over a period of three years by different sensors at different times, making the images of distinct characteristics. To collect training samples, three analysts manually digitised a total of 4000 solar panels from the imagery and cropped images around labelled solar panels. They realised that creating an image tile size of 500 x 500 made the network convergence very slow. They explain that this has occurred because the solar panel pixel was too small to generate sufficient back propagation error at a certain training period hence the need for a higher resolution. With a higher resolution, they trained their model based on the stochastic gradient descent with a batch size of 5 images, a learning rate of 0.02, a momentum of 0.6 and weight decay of 5 and an epoch of 140 using a single NVIDIA Tesla 12GB GPU. Upon completion, they visually inspected the result. They experienced a lot of false alarms resulting from cropping the images. They explain that cropping images gives the network the inability to learn other patterns that are not solar panels but look like solar panels. For example, the model detected shadows on the roof top and, surprisingly, a zebra crossing. This was improved by retraining the network by adding 110 new images and labelling the false alarms in them with a different label. This yielded an improved result that took 4 days to complete. The result of their analysis showed that when the trained network was applied on the testing data set which covers 18 times the size of the training image, it was reported that the network took 3 seconds to process a 1000 x 1000 tile and one hour to complete the entire detection. The network detected a total of 4500 solar panels in San Francisco and 1500 in Boston. To compute the accuracy of their model, they used the completeness and correctness method which is different from other researchers in this review. Their model achieved a completeness of 0.873 in San Francisco and 0.840 in Boston. For correctness, they achieved 0.855 in San Francisco and 0.812 in Boston. The success of this research proves that convolutional neural networks (CNN) perform well with high resolution imagery. It is important to note that shadows in imagery are a very important factor that could cause false alarms in computer vision research for solar panel detection. Hence, they should be avoided.

In (Kausika et al., 2021), they stress that the misguided information about solar panels can give rise to ineffective policy formulation and monitoring as the Netherland government seeks this information for the realisation of its climate agreement. To this end they carried out a study on GEOAI for the detection of solar photovoltaic installations in the Netherlands. In this research, they utilise a semantic segmentation method that can detect the location of the solar panel and estimate its shape from high resolution imagery. Given that data is an important aspect of GEOAi, they leveraged a repository of aerial photographs from the Netherlands' Cadastre, Land Registry and Mapping Agency captured twice a year. The repository used here contains orthophotos and true orthophotos with a resolution of 10cm in winter months and 25cm in summer months. However, imagery with a spatial resolution of 10cm was used throughout the research. In addition to the imagery, ancillary building footprints were collected to match the aerial imagery for post processing. In this research, a different modified version of convolutional neural networks called TernausNet was used. This deep learning architecture incorporates U-Net architecture with VGG16 as encoder. In this research they utilised a pretrained model with an image size of 1024 x 1024 pixels. They manually collected 5,000 ground truthing data by manually digitising the panels on the images and then using the vector feature class as feature layer input data to mask the raster image tiles. The result of this operation was an imagery set of solar panels alone. A total of 800 image tiles were used to train their model. The data was split into training and validation at a ratio of 80% to 20%. Upon training their model, they operated it in a different scenario. The result of their model shows that TernausNet had false positives on greenhouses and tanks. The model was also reported to have performed better on residential buildings than on industrial buildings. To ensure the algorithm was valid for the entire country, they incorporated more training data to expose the algorithm to diverse panel types, spatial background, and image type from different providers. Although they trained 16 models on different parameters, the final model was trained using a learning rate of 0.0001, batch size of 8 and RGB layer. The result of this research shows that orthophotos had better accuracy in detecting the shapes of the panel correctly while True Orthophoto had poor accuracy due to its image quality but had better location accuracy. They observe that false positives occurred mostly on greenhouses and large commercial buildings with glass domes or roofs. Also, a recurring false positive on rooftop dormers was largely seen as well as wet patches. They give that to access the performance of the model, classification accuracy does not give the right metric to evaluate the robustness of their model. Using the accuracy matrix in

three locations where they tested their model, their model achieved an average precision score of 0.933, an average Recall of 0.92 and an average F1 score of 0.93. At the height of this high performance of the model, they experienced a high discrepancy when comparing their result to the CBD database of the Netherlands, they give that time of acquisition, level of detail, and quality of source data is an attributing factor to the discrepancy. Hence, they call for a regional level investigation of methodology into solar panel detection. They stress that training a deep learning model with varying quality, spatial resolution, different sensors, and processing techniques makes the task of detection of solar panels difficult in deep learning. Again, they admit their difficulty in determining beforehand, imagery quality. Since they used orthophotos and true orthophotos in this research, post processing was indeed important to correct anomalies. This research paper is an excellent work of science that integrates geospatial imagery in deep learning while taking into consideration geodetic features. It is important to note that for the first time in solar panel research, displacement in geodetic solar panels was witnessed because of using non ortho rectified imagery.

To generate statistics for solar panels (Lindahl et al., 2023) mapped decentralised photovoltaic and solar thermal systems using remotely sensed data and deep machine learning. They tried to address the challenge of being able to identify small, decentralised grid-connected and off-grid PV along with ST system by assess the real-world accuracy of identifying small, decentralised PV and ST systems using a convolutional neural network aerial image classification algorithm called Deep solar developed by (Yu et al., 2018). They assert that there is a need to create a database of solar thermal systems differently from solar panel PV systems as this does not exist. Just as (Castello et al., 2019; Kausika et al., 2021; Malof et al., 2017; Yuan et al., 2016) who used orthophotos as geospatial data in their research, (Lindahl et al., 2023) also used orthophoto provided from the Swedish Land Survey. Each image in the repository covers a tile of 2.5 x 2.5 km with a spatial resolution of 0.016m and pixel size of 115 x 115. However, since the deep solar algorithm used in this research was trained using an image tile size of 299 x 299, (Lindahl et al., 2023) up sampled their image tiles to 299 x 299. To carry out their research, they trained the Seep Solar using two different data sets combined. One from the openNRW_train_16 containing1814 positive and 36790 negative image tiles. Secondly was the Swedish set of images for fine tuning the deep solar algorithm to fit to the Swedish conditions. All images were down scaled to 299 x 299 pixels and the deep solar algorithm was trained with 100 epoch and an imbalance rate of 5. They created

ground truth by applying the deep solar algorithm over the test data set. The result was a set of vector feature classes. In their study, Python was used as the programming language while QGIS was used as the program for analysing the gathered and created geodata. They used the accuracy analysis to determine the performance of the trained deep solar algorithm after it had been retained by them. The result of their research showed that the deep solar algorithm achieved a precision of 93.4, recall of 81.3, and an F1 of 86.9 when operated over the openNRW_train_16 data. In the same vein, the deep solar algorithms when operated over Sweden achieved a precision of 63.9%, a recall of 81.8%, and an F1 of 71.1%. This research is a clear methodology of transfer learning and again proves that convolutional neural networks can again be used in the detection of solar panels.

In (Tan et al., 2022) they pose that segmentation edge lose has been a prevailing challenge in the detection of solar panels. Also, they give that due to the poor colour contrast of remote sensing images, it is important to have knowledge of solar panel colour. Their work proposes that; rather than using a single layer manually digitized solar panels, they use different layers of shape target which reduces loss on the edge of the PV plate. From an imagery dataset of low-resolution satellite imagery, they cropped each image to a tile size of 512 * 512 pixel. A total of 19,863 images was used to train their model while they tested on 8,262 image tiles. To this end their work uses pyTorch to build a semantic segmentation model using MMSegmentation. They use a constraint refine module based on the prior knowledge of the colour of the solar panel to increase colour perception. Their research show that colour constrain refinement yields a significant improvement in the colour perception of solar panel. With an intersect over union of 73.46% they achieve a recall of 82.30%.

For the first time in the research of solar panel detection and mapping, (Sizhouhi et al., 2020) uses a mask RCNN deep learning architecture with a modified VGG16 backbone. They use the AMIR Dataset which consist of 3,580 aerial images of large-scale solar PV plants located in 12 countries and six continents. To train their deep learning model, they utilize the sigmoid activation for the last layer of the deep learning architecture, ReLU was used for the inner layer activation function, binary cross entropy for loss entropy, ImageNet was used for the initial weight transfer learning and a batch size of 1 trained on 150 epoch. To achieve highest accuracy and lowest losses, they adopt the deep learning transfer learning method. Here, ImageNet pretrained weight was used and

the model's weight was initialized to prevent random initialization. Their training accuracy reaches a high score of 97.61% while their testing accuracy gives an accuracy of 96.63%.

Deep learning architecture.

The seemingly magical functionality of written text recognition, speech recognition and image recognition has fascinated humans for years. That the computer could tell what animal was contained in a picture or what number is written in a text is amazing. What is more amazing is that the computer could somehow tell what species of animal it had identified in the image by merely looking at it, is even more fascinating. This act mimics closely human capability to reason and accurately identify anything. (Elazami Elhassani et al., 2022) explains that the ability of a deep learning architecture to automatically extrapolate knowledge from a training data set and predict unseen samples in record has made its adoption widely acceptable. The importance and efficacy of deep learning was further brought to light by (Feki et al., 2021) who explained that in the COVID 19 pandemic, deep learning architecture had shown an unprecedented performance towards screening COVID 19 from chest X ray images without undermining privacy of patients. The outburst of deep learning has been characterized by its ability to save time and do more work at a speed unaccomplished by human energy. (Liao et al., 2021) further assert in their study that deep reinforcement learning approach (DRLA) effectively rescheduled train timetables under disturbances, resulting in significant energy savings compared to traditional methods, saving an average of energy by 5.11% and 7.29% with an average reaction time of only 0.0013 and 0.0019 seconds against disturbances, respectively. This again demonstrates that deep learning can better the time taken by humans to perform tasks, give improved accuracy and reduce the cost of achieving this task. This does not support the notion that machines would take over human jobs, however deep learning would improve the way humans achieve everyday results.

Convolutional Neural Network

In the work of (LeCun et al., 2015), they describe deep learning architecture as a multilayer stack of simple modules, which are subject to learning, and many of which compute non-linear input–output mappings. They explain that deep learning methods are representation learning methods with multiple levels of representation transformed from one level to another. They give that this representation learning is a set of methods that allows a machine to accept raw data automatically and find the best representation for detection or classification. This representation is learnt from

the data they take in as input using widely known learning procedures. They also explain that the common form of deep learning is supervised learning. Here the machine is shown an image of an object. It then produces a vector score however the scores are most likely not to be the desired score therefore an objective function measures the error between the output and desired score. It then modifies its internal adjustable parameters to reduce this error. These are called weights or Knobs. To adjust these weights, deep learning algorithm computes gradient vector using a popular procedure called the stochastic gradient descent. They further give that the stochastic gradient descent slope uses a small set of computed examples to adjust the entire weights accordingly across the network. The output of the vector scores is moved to another layer that takes this score as input through the most popular nonlinear function called the rectified linear unit ReLU. They give that this process happens across the layers of the deep learning until the network has the lowest possible error rate between its calculated vector score and desired scored. They affirm that upon training a deep learning network, it is best practice to test deep learning models on a different set of examples called the data set, to ascertain its generalization ability. They also give that, challenges of deep learnings ability to solve images problem gave rise to a type of deep learning architecture called the convolutional neural Network which can process data that come in form of multidimensional arrays like RGB images. They describe the convolutional neural network to have local connections, shared weights, pooling and makes use of multiple layers as the distinct ideas behind the architecture. According to (LeCun et al., 2015) recent convolutional network architectures have 10 to 20 layers of ReLu, hundreds of millions of weights and a billion of connections between units.

Recurrent Neural Network

(Elazami Elhassani et al., 2022) describes deep learning as a subset of machine learning that uses artificial neurons that serve as the building blocks for the different neural network architectures. They give that those artificial neurons which comprise of a single output layer and one output node are the backbone of the deep learning technology as an interconnection of these neurons have the capability to learn more complex pattern. The fact that a neural network has several hidden layers makes it a deep neural network (Elazami Elhassani et al., 2022). They assert that the need for solving problems with sequential data such as genomics or text gave rise to the development of the Recurrent Neural Network and its ability to model spatiotemporal data makes it outstanding.

An example is its outstanding performance speech recognition and handwriting recognition. They give that the drawback of the RNN is it exploding and vanishing gradient however variants of RNN like the ‘‘Long Short-Term Memory’’(LSTM) and ‘‘Gated Recurrent Unit’’(GRU) have been introduced.

Fast Region Based Convolutional Neural Network

With the improvement of object detection by the convolutional neural network, the need arises for localization of object in images. (Girshick, 2015) states that although the region based convolutional neural network excellently achieves object detection, the RCNN has the disadvantage of a multistage pipeline training, expensive training, and slow object detection. (Girshick, 2015) proposed a new algorithm Fast Recurrent Neural Network that solves the draw backs of RCNN. They opine that the Fast RCNN is fast to train and test and has the advantage of a higher detection quality, single stage training and less need disk storage for feature caching. The fast RCNN takes an image as an input the network process the whole image with several convolution and then for each object proposal, a region of interest pooling layer extracts a fixed length feature vector from the feature map. (Girshick, 2015) further explains that each feature vector is fed into a sequence of fully connected (fc) layers that finally branch into two sibling output layers: one that produces softmax probability estimates over K object classes plus a catch-all “background” class and another layer that outputs four real-valued numbers for each of the K object classes. That is a soft probability and per class bounding regression offset. Each set of 4 values encodes refined bounding-box positions for one of the K classes. (Girshick, 2015). According to (Girshick, 2015) the FRCNN has reduced training time and can perform 9 times faster than the RCNN. The fast region based convolutional neural network performs well at object detection and real time object detection in motion pictures.

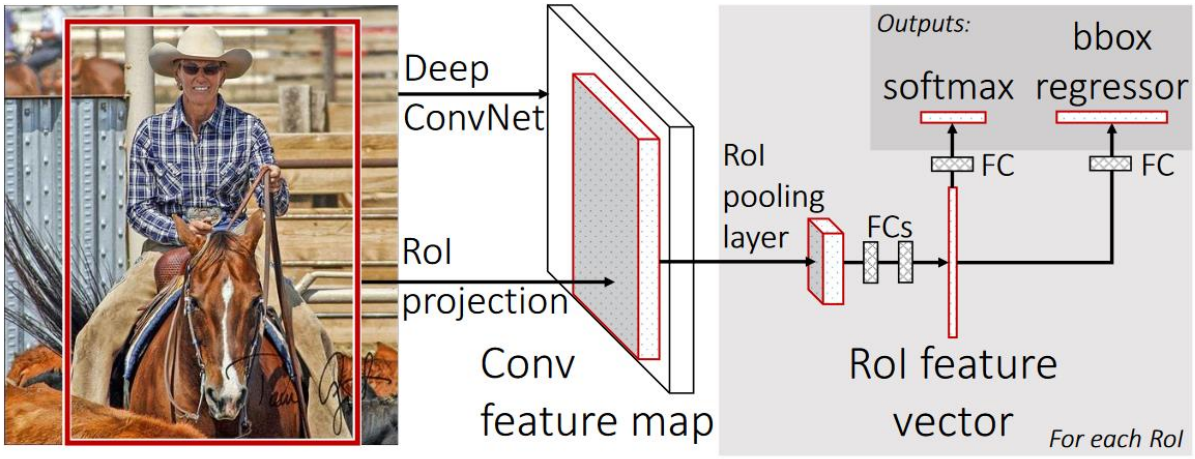


Figure 3 - Illustration of the fast region based convolutional neural network.
Source - (Girshick, 2015)

Mask Region Based Convolutional Neural Network (Proposed)

state

(He et al., 2018) opines that Mask RCNN can train the COCO dataset under one to two days on a single 8 GPU machine and outperforms the winner of the 2016 COCO key point competition. They purport that the Mask RCNN is a flexible framework, for instance level recognition.

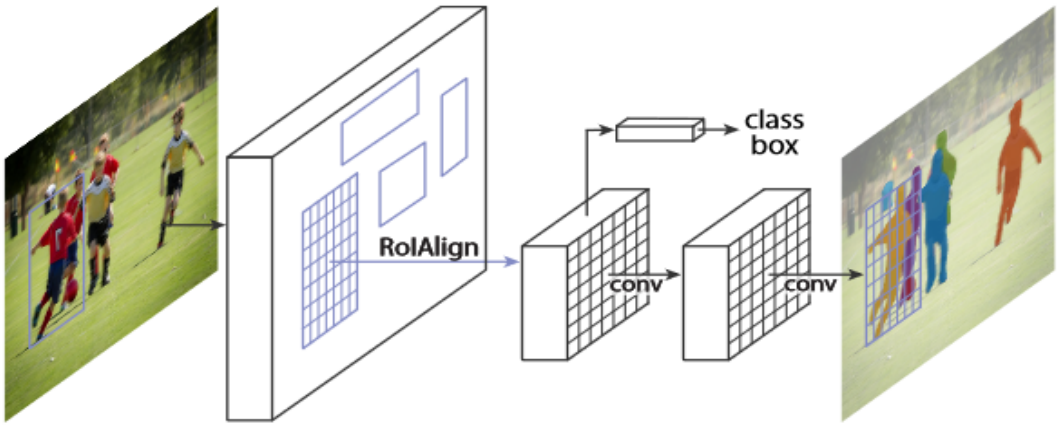


Figure 4 - Figure showing Mask RCNN deep learning architecture.

(He et al., 2018) explains that the Mask RCNN has third output of the object mask in addition to a class label and a bounding box offset which already exist in the Faster RCNN. All of which

makes for a finer spatial layout. This architecture performs great at task dealing with instance segmentation, fine grained object understanding as well as precise object localization.

The architecture starts off with the first two stages of the Faster RCNN. The first stage is the region proposed network which proposes candidate object bounding box. The second stage extracts features using ROIPool from each candidate box and performs classification and bounding box regression. In faster RCNN quantization of images are done in the ROI Pool however this introduces misalignments between the ROI and the extracted feature. (He et al., 2018) assert that this has a negative impact in pixel accurate mask. They solved this by removing the quantization of ROI Pool instead they use binary interpolation to compute the exact values of the features at four regular sampled location in each ROI bin and aggregate the result. They state that when Mask RCNN is used in combination with the ResNet 101 feature pyramid network, it achieves greater results.

Conclusion

This literature review covers various studies for which deep learning has been used for the study of solar panel detection and mapping but to what degree of delineation. From this review, it can be concluded that transfer learning can improve the performance of a deep learning model. As data becomes an important role for geospatial artificial intelligence, true orthophotos are best used for deep learning studies in geosciences. Moreover, it can be noted that various deep learning architectures have been used in previous research but only one of these researchers have used Mask RCNN deep learning architecture towards the identification of solar panels which calls for further studies toward the study of accurate solar panel detection. Again, most challenges of solar panel research have been treated as an image segmentation problem however this thesis employs an instance segmentation method which combine object detection and image segmentation to detect and delineate solar panels. Even though Spain has over 300 days of sun and has an abundance of solar panels, it lacks research on a deep learning algorithm for the detection of solar panels to the best of our knowledge. UAV technology in these literatures showed that photogrammetry has been applied to a range of studies in geoscience however to the best of our knowledge, very high-resolution multiband UAV photogrammetry has not been used in the study of solar panel detection neither have previous researchers leverages the use of very high-resolution imagery. The question

of establishing ground control point to improve geodetic accuracy in UAV photogrammetry is becoming less of a priority as modern UAV now comes with centimetre-grade accuracy in RTK. Again, it can be deduced that geodetic implication of UAV photogrammetry is subject to its use and application in geoscience. Where the geodetic characteristics of the features being detected in a photograph are not important, setting up ground control points might be considered a waste of time and resources. Like in the development of a deep learning model such as this. It is not to say that geodetic accuracy is not important especially when mapping small features such as solar panels. It can be concluded that 60% - 80% overlaps have been reviewed to be successfully used to carry out UAV photogrammetry as this ensures that the Ortho mosaicking algorithm has enough imagery to identify tie point features for reconstruction. Although Ortho mosaicking software is yet to find solutions to homogenous environments and blur correction from moving objects while in flight. Weather conditions like cloud cover and time of the day for collection of data with UAV plays a critical role in forming the characteristics and quality of the imagery. Again, in UAV photogrammetry, geoscientists must ensure that images are calibrated to eradicate inherent distortion inherited during flight. Since deep learning models do not require geodetic attributes to perform computer vision operations, this research seeks to apply UAV photogrammetry without ground control points to develop a deep learning model for the detection of solar panels with the aim of quantifying solar panels in urban areas.

DATA, METHODOLOGY AND RESOURCES.

This chapter describes the study area used where the research was conducted, the type of data that was used to carry out this research and why such data was used as well as its source. This section also describes the various methods applied in the precise collection of data and the various resources used in carrying out the research analysis. This includes the various hardware and software used.

Study Area

In order to collect aerial imagery for the detection of solar panels, Universität Jaume I in Castellón was used as the study area. This study area was chosen due to its flight operational accessibility. The university also has fewer high-rise buildings as compared to other residential areas which could impede flight operation. The study location flown below 100m is not in conflict with manned flight operations within the vicinity. The study gave the researcher the advantage of utilising high radio frequency gain due to less high buildings which could lead to image and aircraft transmission failure. Lastly the study location had no vertical obstacle above 100m in the flight path of the UAV.

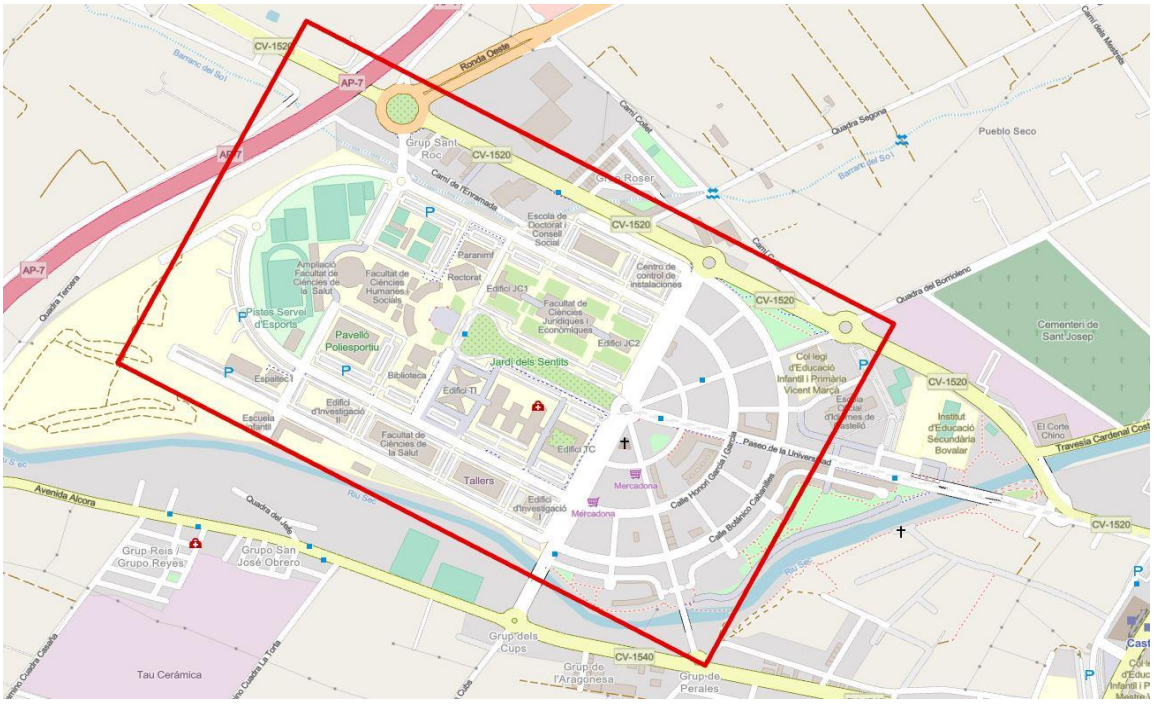


Figure 5 - Image of the study area showing bounding area for flight data collection.

Description of Data

To detect solar panels using deep learning, the deep learning models and computer vision technology utilise digital values from images which are used to detect solar panels. To show the computer vision models the solar panels, those solar panels need to be represented in the form of an image or photograph. To this end, this research makes use of aerial imagery, and the scope of this research is limited to solar panels installed horizontally on rooftops. Therefore, aerial images or photographs would capture the surface area and extent of the solar panels hence the use of aerial imagery. The use of high resolution UAV imagery is in not in consonance with previous studies on solar panel detection such as (Castello et al., 2019; Hu et al., 2022; Lindahl et al., 2023; Malof et al., 2016, 2017; Puttemans et al., n.d.; M. Wang et al., 2018; S. Wang & Li, 2021; Zhuang et al., 2020) all used a mix of satellite imagery and other imagery of a resolution lower than what was used in this research. Again, to the best of our knowledge, no previous research has used very high-resolution imagery from UAV photogrammetry as used in this research. The data set used for this research contains a total of 1,486 aerial images collected in JPEG file format with a horizontal orientation. Each image has an exposure time of 1/1250 and an ISO of 240 with the image having a width of 5280 and a height of 3956. The aerial images are all of barrel distortion, 8 bit per sample, 20 megapixels and sized between 10mb and 12 mb. The aerial imagery collected for this study is a 3 channel RGB image. These images were divided into 1078 aerial images as a training data set and 417 aerial images as a testing data set. Although these individual images were not used singularly, this research used a true orthophoto processed from the training dataset and testing data set. The true orthophoto training dataset has a tile size of 46,000 x 42,000 pixels, a ground sampling distance of 0.03m, an uncompressed size of 28.79gb in 32-bit FGDBR file format and a multiband of 4 channels. Band 1 is red, band 2 is green, band 3 is blue while band 4 is a normalised digital surface model composited. The true orthophoto training dataset has a projected coordinate system of WGS 1984 UTM Zone 30N projected on the universal transverse Mercator, a false easting of 500000, a false Northing of 0.0 and a vertical coordinate system on the EGM96 Geoid. All orthophoto bands have a raster value from 1 – 254 except band 4 (normalised digital surface model) whose raster values are from -41.225 - 44.291 covering 1.7km Sq. The true orthophoto testing dataset covers an expanse of 0.9km Sq with a tile size of 32,000 X 30,000 pixels and a ground sampling distance of 0.03m. The true orthophoto testing dataset has an uncompressed size of 2.68gb in 8-bit TIFF file format and a 3 channel RGB band. All bands have a raster value from 1

– 254, a projected coordinate system of WGS 1984 UTM Zone 30N projected on the universal transverse Mercator, a false easting of 500000, a false Northing of 0.0 and a vertical coordinate system on the EGM96 Geoid.

Resources And Description of Resources Used

Unmanned Aerial Vehicle.

This research utilised a DJI Mavic 3 enterprise unmanned aerial vehicle. The UAV is a multirotor quadcopter fitted with a digital RGB sensor whose sensor is a wide-angle of 4/3 CMOS, 20MP, a focus of 1m to infinity, an aperture value of f/2.8-f/11 without a flash and a field of view of 84 degrees. The sensor features a mechanical shutter that prevents motion blur, a prominent challenge in photogrammetry as expressed by (Samad et al., 2013), supporting rapid shooting up to 0.7 sec and a general flight time of 45 minutes. The sensor also has a telecamera of ½ inch CMO sensor capable of shooting 12MP photos with an aperture of f/4.4 which shoots at 3m to infinity. The aircraft is fitted with GPS, Galileo, BEIDOU and GLONASS global navigation satellite systems. Its GNSS has a vertical geodetic accuracy of 0.1m with vision system, 0.5m with GNSS and 0.1m with real time kinematics. Horizontally, the UAV has a geodetic accuracy of 0.3m with a vision.

Figure 6 - Image of aircraft used for the research – A DJI Mavic 3 Enterprise
Source – Elchapuzas Informatico.

system, 0.5m GNSS and 0.1m when connected to a real time kinematic GNSS receiver. For this research, the researcher relied upon the geodetic accuracy of the UAVs GNSS.

Computer

The computational resources needed for most deep learning research are usually high. ESRI recommends using a computer with 4 cores, 32GB free storage space on the solid-state drive, a RAM of 32GB, a 4GB dedicated graphics card, OpenGL 4.5 with the ARB_shader_draw_parameters, EXT_swap_control, EXT_texture_compression_s3tc, and EXT_texture_filter_anisotropic extensions and a screen resolution of 1080p. This research however used a computer with Intel Core i7, NDVIA GeForce GTX 1660Ti 6GB dedicated graphic card with a 32GB RAM and dual storage of 500GB solid state drive and a 1TB hard disk drive.

Software

This research declares that no programming codes were implemented during the study. However, to perform the various analyses in this research, this research utilised the DJI Map Pilot 2 for flight operation while the DJI flight planner was used for office flight planning. All spatial analysis including deep learning training and testing was done using ESRI ArcGIS PRO version 3.2 while the photogrammetric process was achieved using the ESRI Drone2Map version 2.0.

Data Collection

Unlike other researchers who used available secondary, open geospatial data for the detection of solar panels, the primary data used in this research was originally developed through a UAV photogrammetric process. The data collection process was divided into three data collection phases covering a total study area of 2.6km Sq. During data collection, training data sets were collected separately from the testing data under different atmospheric conditions.

Figure 7 - Researcher on the field setting up Unmanned Aerial Vehicle for data collection.

Training Data

The area used for collecting testing data was divided into two sections. This was done to solve the problem of deep learning models being trained only in restricted atmospheric weather conditions which could have resulted in negative performance by the model when faced with a real-world scenario. To solve this challenge encountered by previous researchers, this thesis proposes the collection of training data in two different weather conditions.

Using the DJI Pilot 2 flight planning application embedded in the flight controller system of the UAV, a feature polygon was created over section 1 of the study area. The flight planning

application automatically produces 11 waypoints with nadir camera angle and one diagonal way point with its camera at 60 degrees oblique across the flight area to improve orthorectification.

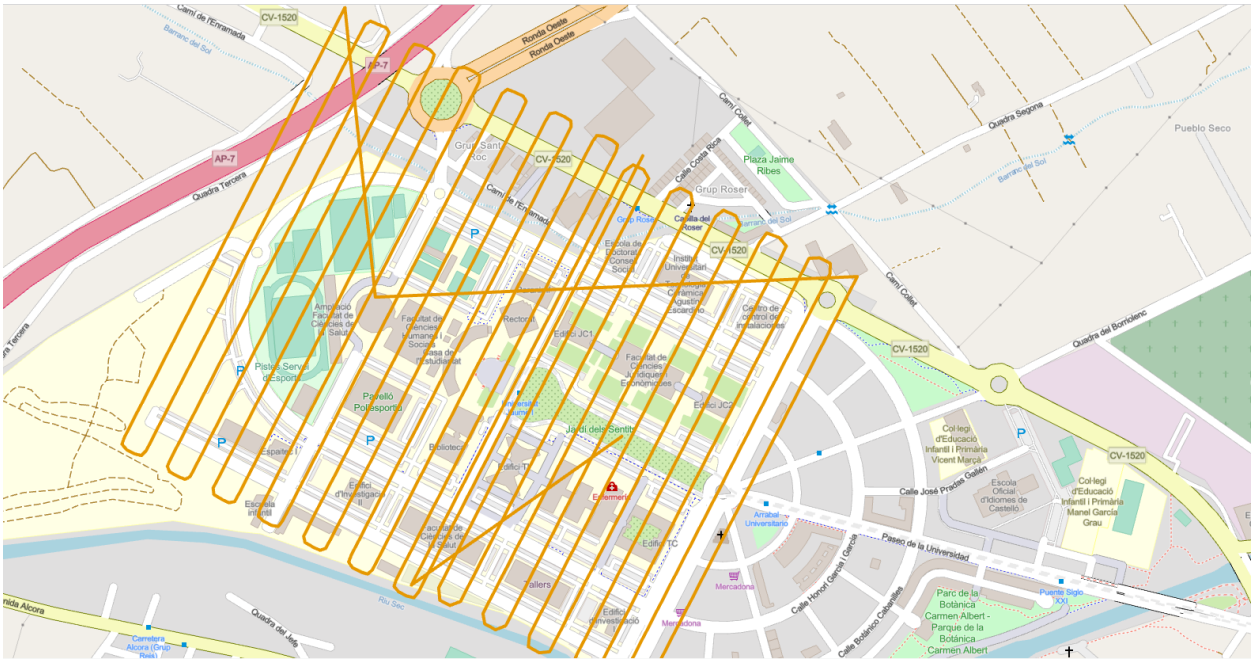


Figure 8 - Flight planning for aircraft way point.

Aerial images were collected using a high overlap of 70% forward overlap while the side overlap was set at 80%. This was done because (Iheaturu et al., 2022) asserts that an overlap of at least 60% front overlap and 75% side overlap is required to achieve a good model for structure from motion. A higher overlap was used to ensure high multiple tie points across successive images. This is drawn from the well-known fact that photogrammetry software is still unable to solve the problem of reconstructing homogeneous image texture. This is a common problem in mapping water bodies as well due to the homogeneous nature of water. Since solar installations are slightly homogenous, a high overlap was used. In the same vein, (Zhang et al., 2023) state that low overlapping images result in large matching error.

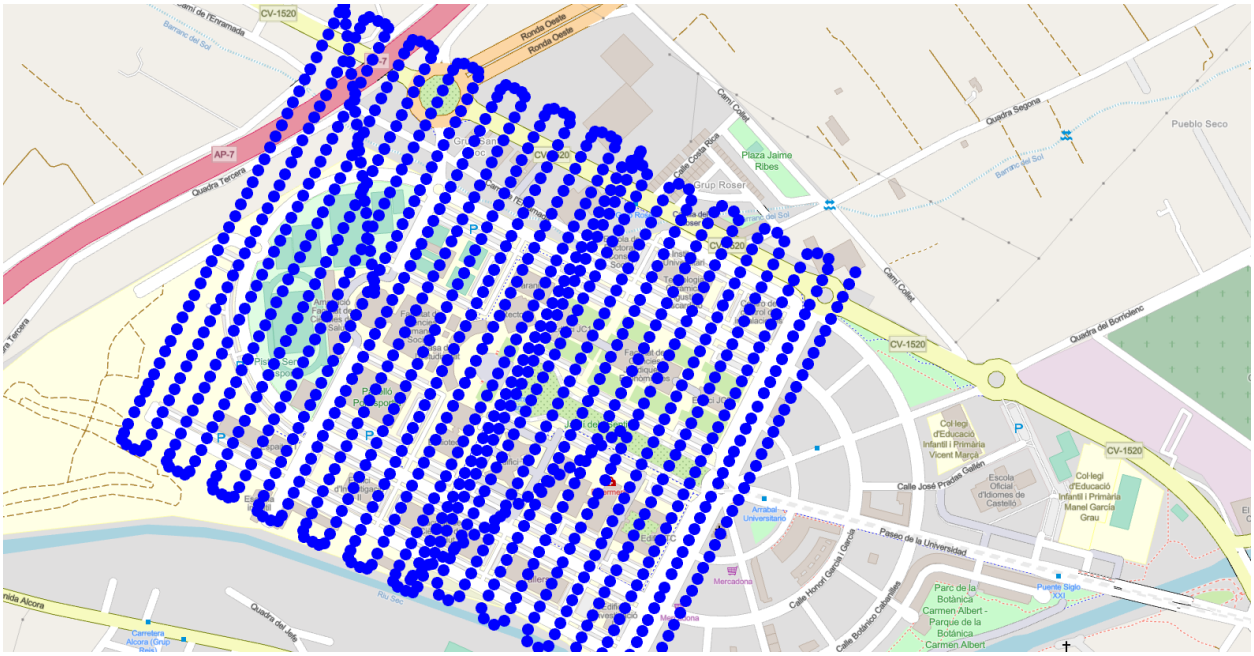


Figure 9 – Point location where Images was taken across study area.

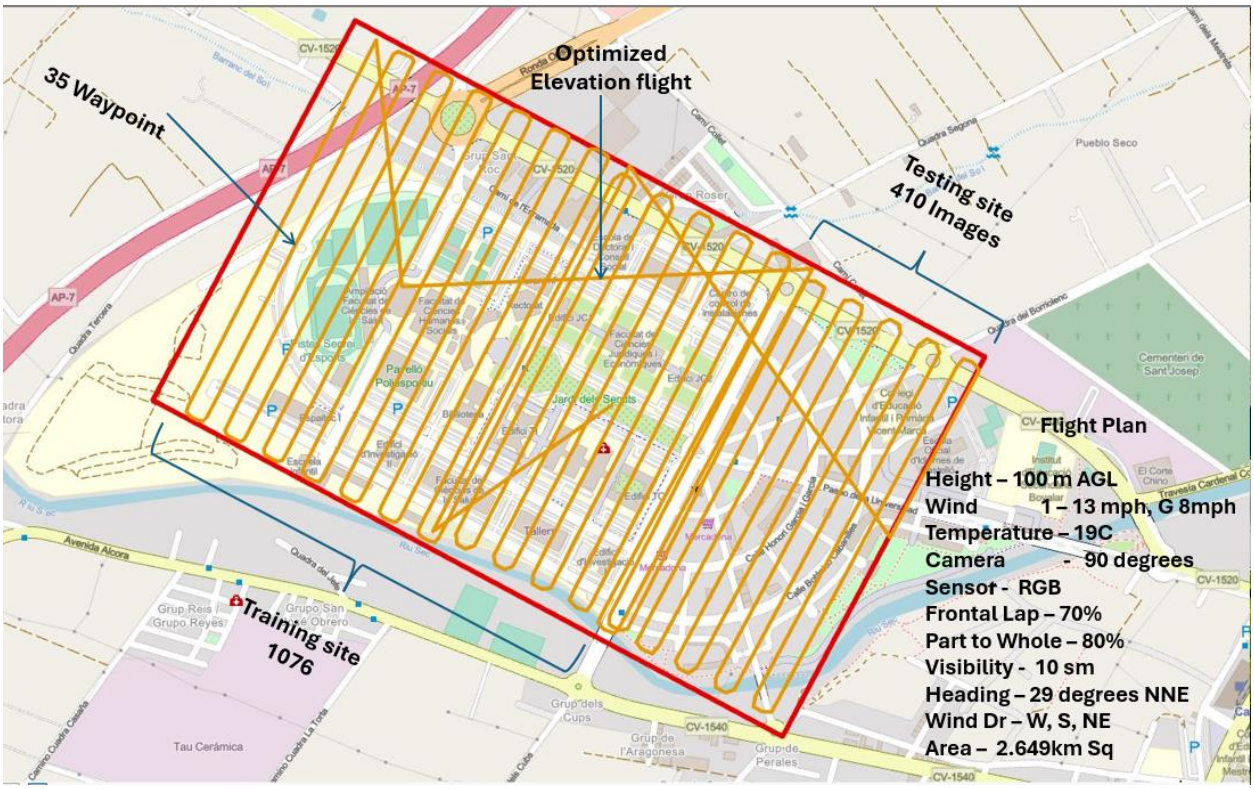


Figure 10 - Way point flight distribution of the study area including optimized flight path.

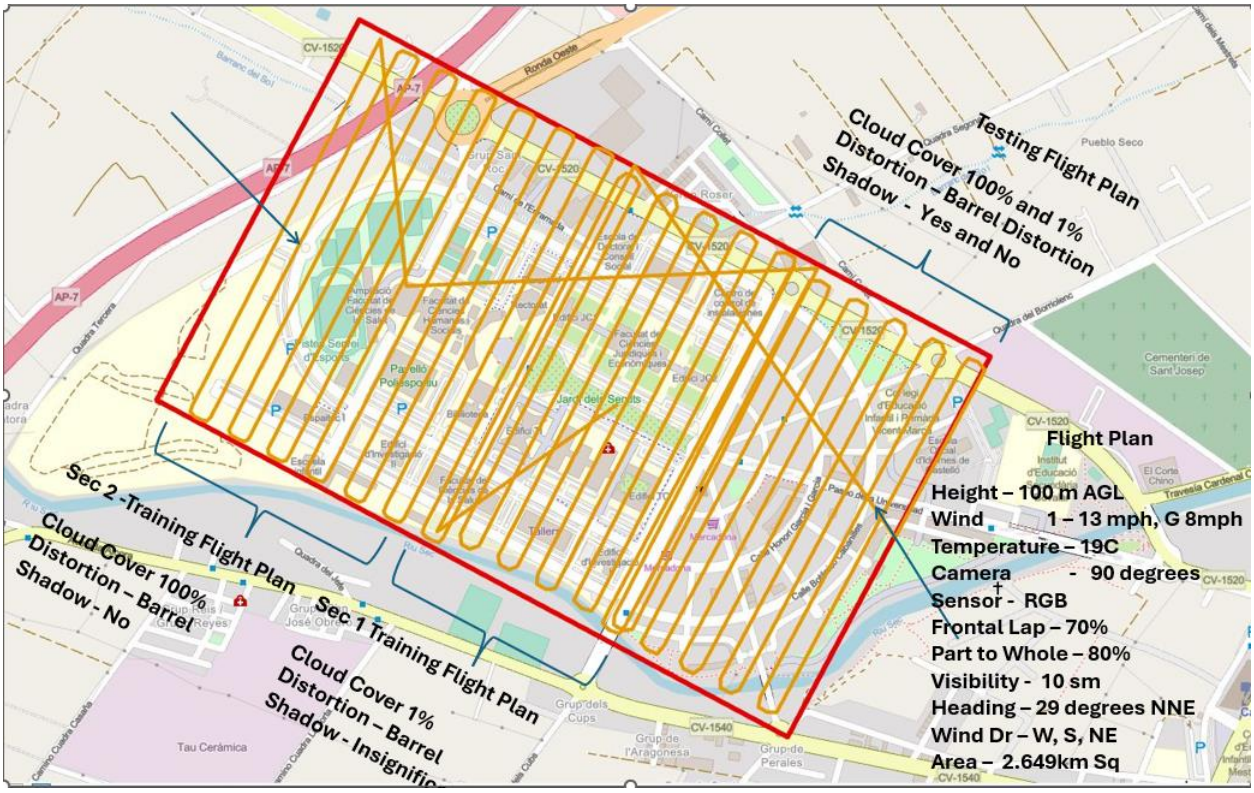


Figure 11 - Flight planning distribution across study area for data collection in diverse weather conditions.

Aerial images were collected with the presence of vignetting because the DJI Mavic 3 has a radial lens distortion from its wide-angle camera. Vignetting according to (Bal & Palus, 2023)b is a phenomenon in which the image brightness is reduced from the optical centre of an image toward its edges. However, according to DJI, the presence of the vignetting provides valuable information about the orientation and distortion of the images which can be used to correct the images for distortion while processing. The UAV was flown at a heading of 117 degrees to and 243 degrees from. The choice of flight heading was obtained from the result of a reconnaissance survey initially carried out to determine the direction the solar panels were facing. This was such that during flight, the UAV camera sees the surface of the solar panels first and not its sides, especially where these solar panels are installed with an inclination.

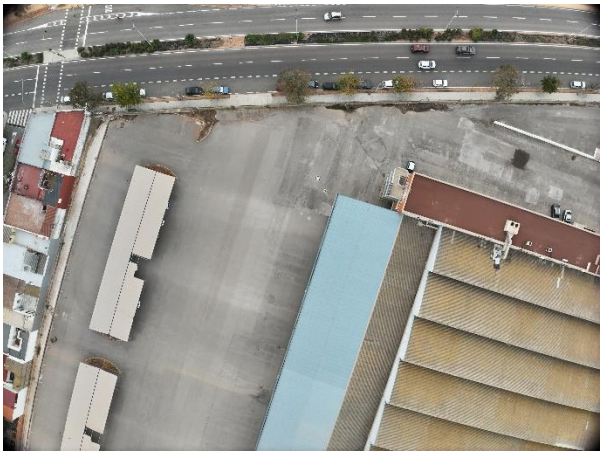
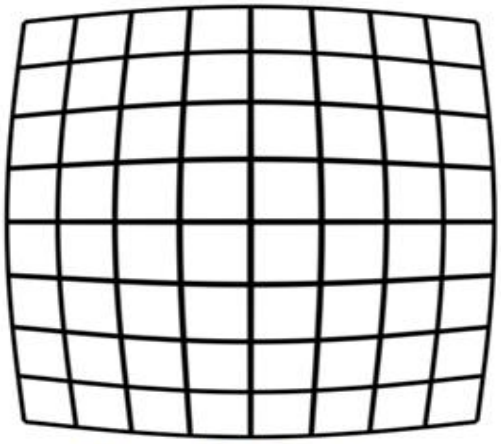
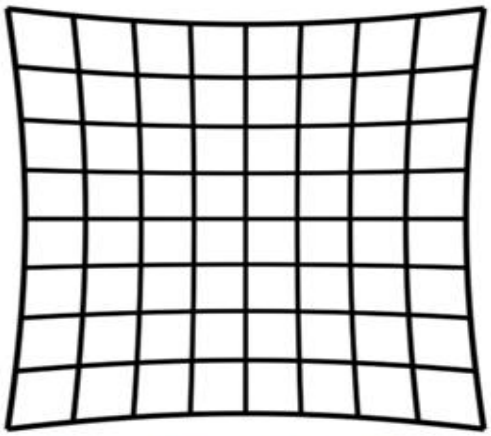


Figure 12 - Barres Distortion in Images

Figure 13 - Non-Barrel Distorted Images



Barrel Distortion



Pincushion Distortion

$$D = \frac{\Delta H}{H} \cdot 100 = \frac{H^* - H}{H} \cdot 100$$

Figure 14 - Lens distortion correction formular.

The UAV speed was set at 10m/s to counter the effect of surface wind of the day which measured at 3mph gusting 11mph which could cause the UAV to drift in flight and in turn resulting in bad aerial imagery as explained by (Zhang et al., 2023). The camera operation mode was set to shutter priority at speed of 1/1250 and sensitivity of ISO at 100. This is in sequence with (Samad et al., 2013) who gives that blurred images error can be caused by shutter speed or UAV flight speed. To solve this, they give that surface wind must be checked before selecting UAV speed.

The UAV was flown at a height of 100m above ground level as (Ali & Abed, 2019) found out that geometric accuracy of aerial images changes significantly at higher altitude. Additionally, in accordance with EU regulation on UAV, UAVs are to be flown below 120m. According to (Mao et al., 2023) strong winds and high temperatures influence UAV flight and the accuracy of its data. They also give that cloud cover could also affect the spectral information of the images. Because this thesis pays attention to training a deep learning model to recognise solar panels in diverse atmospheric conditions, data was collected with an atmospheric temperature of 19 degrees C, 0% precipitation, a cloud cover of 1%, visibility of 10 miles and a wind speed of 3mph west. The UAV was set to allow the camera to take a photograph every 3 seconds even when turning into the next way point.

In section 2 of the training data collection, all other parameters were the same as section 1 except for the atmospheric conditions. Here, data was collected with the temperature at 18 degrees Celsius, cloud cover of 100%, 0% precipitation, visibility of 10 miles and a wind speed of 9mph west gusting 18mph. The UAV was again set to allow the camera to take a photograph every 3 seconds even when turning into the next way point. In this section a total of 14-way points were created and a total of 626 images in JPG file format were collected between 11.40am to 12.09pm Central European Time in December.

Testing data

Testing data sets are important when checking for overfitting, which (Safonova et al., 2023) defines as a situation where a deep learning model performs poorly on the testing dataset but performs well on the training data set. They further give that testing data should be collected independently of the training data set just as this thesis has done.

Since the deep learning model in this thesis will be trained with images collected from diverse weather conditions, it makes sense for the model to be tested on diverse scenarios to ascertain its performance. Therefore, the testing data was collected over different locations using flight parameters of both section 1 and section 2. A total of 11 waypoints were planned and 417 images were collected in JPG file format.

DATA ANALYSIS

Deep Learning Data Preparation

This thesis declares that no programming codes were written to achieve any of the analysis, instead it was accomplished using commercial software. The training of the deep leaning algorithm for this thesis relied on the use of the ArcGIS Pro 'Train deep learning Model' geoprocessing tool.

Upon collection of all aerial data, this research commenced data processing. According to (Ali & Abed, 2019) they identify Pix4D and Agisoft PhotoScan as the most popular photogrammetric software used in orthomosaicking. This thesis used ESRI Drone2Map photogrammetric software for the orientation of images, production of true orthophoto, digital surface models and digital terrain models. Drone2Map is a desktop software with the ability to create contour lines, True Ortho, Multispectral Orthomosaic, Thermal True Ortho, Pansharpened Orthomosaic, Panchromatic Orthomosaic, Digital surface model, Digital terrain model, Shaded relief, Inspection report and NDVI from overlapping nadir images in natural colour, thermal Infrared, or multispectral dataset (ESRI 2024). It is well integrated into the ArcGIS environment which allows interoperability and can open completed projects directly in ArcGIS Pro from Drone2Map as well as sharing projects to the ArcGIS Online or ArcGIS Enterprise. This is different from (Chen et al., 2022; Iheaturu et al., 2022; Ismael & Henari, 2019; Starek et al., 2014; Watanabe & Kawahara, 2016; Zhang et al., 2023) who all used PIX4D photogrammetric software and (Ahmad & Samad, 2010; Tahar et al., 2011) who used ERDAS Imagine and (Cabo et al., 2021; Gonçalves & Henriques, 2015; Jayathunga et al., 2018; Sunarya et al., 2020) who used Agisoft PhotoScan. However, (Șandric et al., 2022) also used ESRI Drone2Map software in their own research.

True Orthophoto Training Data Processing

In the ESRI Drone2Map software, 1076 images of the training data obtained from the flight operations in section and section 2 of the training study were imported into the 2D Orthophoto space of the software and inspected to ensure all images were geotagged and free of problems. To process these images, processing parameters needed to be set. In the software, the point cloud density setting used to derive the level of geometric detail for image reconstruction was set to ultra. The project resolution was set to 1x which allows the selection of the Ultra point cloud setting. Due to the computational load required to process the images, the hardware setting was set to utilize CPU + GPU. In the block adjustment, tie point matching and point cloud generation

process, the ESRI Drone2Map software was set to use the internal orientation data to orient the images. Here, the yaw, pitch and roll are used by the software to adjust the image orientation. For the tie point options, the initial image scale is set to 1(original image size). The refine adjustment setting is checked and set to 1(original size). A tie point residual error threshold of 5,000 was used to such that tie points with a residual greater than the threshold values will not be used in computing the adjustment. To determine the number of images to be used to compute image matches, the large setting was selected. With this, an image is matched to the 12 closest images for each search neighbourhood in the area between each of the four ordinal directions. Because the images have radial distortion in them, camera calibration was performed utilising the focal length, principal point which is the offset between the focal centre and the principal point of auto collimation. The Konrady coefficient was also used to calculate radial distortion while P1 P2, which is the tangential coefficient, was used to calculate the distortion between the lens and image plane. To generate the orthomosaic, the colour balancing setting was not utilised because this thesis proposes to train the deep learning model with data from diverse weather and atmospheric conditions. Seamlines were used to sort overlapping imagery and to produce a smoother-looking mosaic. The software was also prompted to produce the digital surface model as well as the digital elevation model. The images had a coordinate system of GCS WGS 1984 and a z value of EGM96 Geoid and projected to WGS 1984 UTM zone 30N with a vertical projection in the EGM96 Geoid.

The processing took about 6 hours to complete and produced a true ortho photo of 1.744km Sq, a digital elevation model and a digital surface model with spatial resolution of 0.027m. The processing achieved a tie point reprojection RMS error of 0.422. On the external orientation, the processing achieves a minimum standard deviation of 0.009m on the X axis and a maximum of 0.018m on the X axis. However, this standard deviation is a measure from the vertical and horizontal geodetic accuracy of the DJI Mavic 3E GNSS and not from an established ground control point.

Figure 15 - Fully processed true orthophoto of UAV images.

True Orthophoto Testing Data processing.

All parameters used in processing the training data were also used in processing the testing data set except for colour balancing. At the same time, a total area of 0.936km Sq of true orthophoto, digital surface model and digital terrain model was produced with a ground sampling distance of 0.027m. The processing achieved a tie point reprojection root mean square of 0.431. On the external orientation, the processing achieves a minimum standard deviation of 0.015m on the X axis and a maximum of 0.030m on the X axis. Again, this standard deviation is a measure from the vertical and horizontal geodetic accuracy of the DJI Mavic 3E GNSS and not from an established ground control point.

Orthophoto Training Data Preparation for deep learning

In preparation for training the deep learning model, this thesis proposes the addition of the normalized digital surface model as a composite band to the true ortho photo. This process has been suggested by (Kausika et al., 2021) and (Șandric et al., 2022) to improve results of deep learning models. To do this, the normalized digital elevation model had to be obtained using the raster calculator function in ESRI ArcGIS Pro geoprocessing tool. In the toolbox, the digital surface model of the training study area was subtracted from the digital elevation model of the

training study area. The result of this geospatial analysis was a normalised digital surface model in grey scale with a minimum raster value of -41.22 and a maximum raster value of 44.29 in FGDBR file format. The normalised digital surface model was then added to the RGB true orthophoto as a composite using the composite band geoprocessing tool in ArcGIS Pro. This tool creates a raster dataset containing a subset of the original raster bands into a new raster dataset. The result of this analysis is a multiband true orthophoto containing 4 bands and a file format in FGDBR. This was done because of the need for the convolutional neural network to be trained on solar panels in not just RGB but also in grayscale thereby eliminating false positive detections such as shadows upon deployment in the testing data set. False positive detection of solar panels was seen in the work of (Zhuang et al., 2020) whose work expressed mis-segmentation from their deep leaning models.

Annotation

According to (Wang et al., 2018) supervised classification models rely highly on training samples for training deep learning models. In the same vein, (Malof et al., 2016) further stresses its need for training a model and testing its accuracy. Those training samples, according to (Yu et al., 2018), are expensive to construct as convolutional neural networks require training samples to be massive. This thesis collected training samples by manually annotating the solar panels in the entire study area of both training and testing dataset. This was done by creating a feature layer in the ESRI ArcGIS Pro while using the edit digitisation tool to draw polygons or perform on screen digitisation around each solar panel across the entire study area. The digitisation or manual annotation of the solar panels from the multiband true ortho photo resulted in a total of 3,273 manually annotated solar panels with minimum shape area of 0.959 m Sq and 2.485m Sq. Upon completion, a new field was created to give the annotated solar panel a new class label of “1”. This annotation in this research may not be compared to previous research because previous research annotated solar PV installation while this research annotated solar PV and (W. Hu et al., 2022; Malof et al., 2016, 2017; Puttemans et al., n.d.; Yuan et al., 2016) had all used human annotations during their research.

Masking

The processed multiband true orthophoto shows that upon physical inspection, only a few buildings have solar panels on them. It would therefore be computationally expensive and time

consuming to feed the deep learning network with the entire imagery collection. Therefore, a polygon in form of a feature layer was created in ArcGIS Pro around the areas that have solar panels. This polygon has total area of 0.108 km sq. which is 5.7% of the entire study area. This polygon was used as output extent to clip the multiband true orthophoto just as (Li et al., 2021) did. However, in this research, the clip raster geoprocessing tool in ArcGIS Pro was used. The result of the clip raster analysis was a multiband true orthophoto of areas that contain solar panels in the study area alone. There was no need to perform this spatial operation in the test dataset because it would not undergo a convolutional neural network training and real-world data sets do not come clipped. Moreover, it would be interesting to see the generalisation of the model as well as false positive detection.

Figure 16 - Masked multiband true orthophoto of UAV images.

Exporting Training Data

As earlier stated, this research was conducted solely using ArcGIS Pro. Therefore, it follows the analytical process of the software. To train the proposed deep learning model, the clipped

multiband true orthophoto and labelled vector annotation needs to be converted into a training data set. According to ESRI, the deep learning class training samples are based on small sub images called image chips that contain the class of interest. (Castello et al., 2019) used an image chip of 250x250 pixels, (Malof et al., 2017) used a detector which takes an input image of 41x41 pixels. (Golovko et al., 2017) used an image chip of 200x200 pixels. (Kausika et al., 2021) used an image chip of 1024x1024 pixels while (Lindahl et al., 2023) used 299x299. Of all these image chip sizes, (ESRI 2022) gives that image chip size is dependent on the size of the label of interest. In the case of this research, a single solar panel has a geometric area between 0.95-meter sq to 2.48-meter sq and can be fully seen in an image chip of 256x256 pixels. Hence this research exports the clipped multiband true orthophoto to sub images of 256x256 pixels. To do this, the ArcGIS Pro "Export Training Data for Deep Learning (Image Analyst)" tool was used.

The input raster was the clipped multiband true orthophoto, while a designated folder was specified to store the image chips and associated metadata. On the input feature class parameter, the geoprocessing tool takes the hand-annotated polygon of the training data section as input data. TIFF format was specified as output image chip format while a tile size of 256 pixels was specified on the X axis and a tile size of 256 pixel was specified on the Y axis with 50% stride of 128 pixels. Because this thesis proposes the use of a Mask RCNN, RCNN Mask was specified as the output metadata format in the geoprocessing tool. The output chip set according to ESRI will have a mask on the area where the sample exists for each instance of an object in the image chip. According to ESRI 2022, this format is based on Feature Pyramid Network (FPN) and a ResNet101 backbone in the deep learning framework model. (He et al., 2018) explains that the feature pyramid network uses a top-down architecture with literal connections to build an in-network feature pyramid from a single scale input. They further state that using a ResNet-FPN backbone for feature extraction with Mask RCNN gives excellent gain in both accuracy and speed as (Li et al., 2021) recommends that FPN can be instrumental in dealing with robustness in multi-resolution situation image. A class label of "1" from the annotation feature class was selected as class value field for the geoprocessing tool. No buffer radius or mask polygon was used here while rotation angle was left at 0 degrees. The local reference system of the map space was used throughout the analysis for consistency. To give context for the generalisation of the model, the area surrounding the class label of interest was not blackened out. This is such that the model can also learn relatively what

is not a solar panel. During the process of creating the image chipset a minimum overlap ration of 100% was used such that only images chips with 100% of the solar panels visible in them will be exported for training. A total of 5977 images chips of 265x256 were created. Out of the 5977 image chips, 1599 image chip contained solar panels in them with a total of 6400 solar panel features. The result had a minimum of 1 feature per image chip, a mean of 4 features per image chip and a maximum of 13 features per image chip.

Training Deep Learning Model.

The focus of this thesis is on the detection of the individual solar panels and not solar installations. It is important to note that this research seeks to map the individual boundaries of each solar panel. To this end, detecting individual solar panels is therefore a problem of instance segmentation which according to (Ferreira et al., 2020; Șandric et al., 2022) combines object detection and classical segmentation. According to (Șandric et al., 2022) they give that Mask R-CNN is a widely used in solving the problem of instance segmentation and has been validated as efficient in object-based segmentation and classification. Previous studies on detection of solar panels have used various deep learning models such as (Malof et al., 2017) Convolutional Neural Network, (Malof et al., 2016) Random Forest Classifier, (Golovko et al., 2017) Faster R-CNN, (Malof et al., 2019) Solar Mapper, (Castello et al., 2019) U-Net which they assert to be the most popular CNN architecture for fast and precise segmentation of images, TernausNet (Kausika et al., 2021), DeepSolar (Lindahl et al., 2023), faster R-CNN (J. Hu et al., 2022). To the best of my knowledge, no researcher has tried to employ the Mask R-CNN deep learning architecture in the detection of solar panels as this research seeks. It is important to state that where the boundaries of objects are not of interest, Mask RCNN should be avoided as (Ferreira et al., 2020) puts.

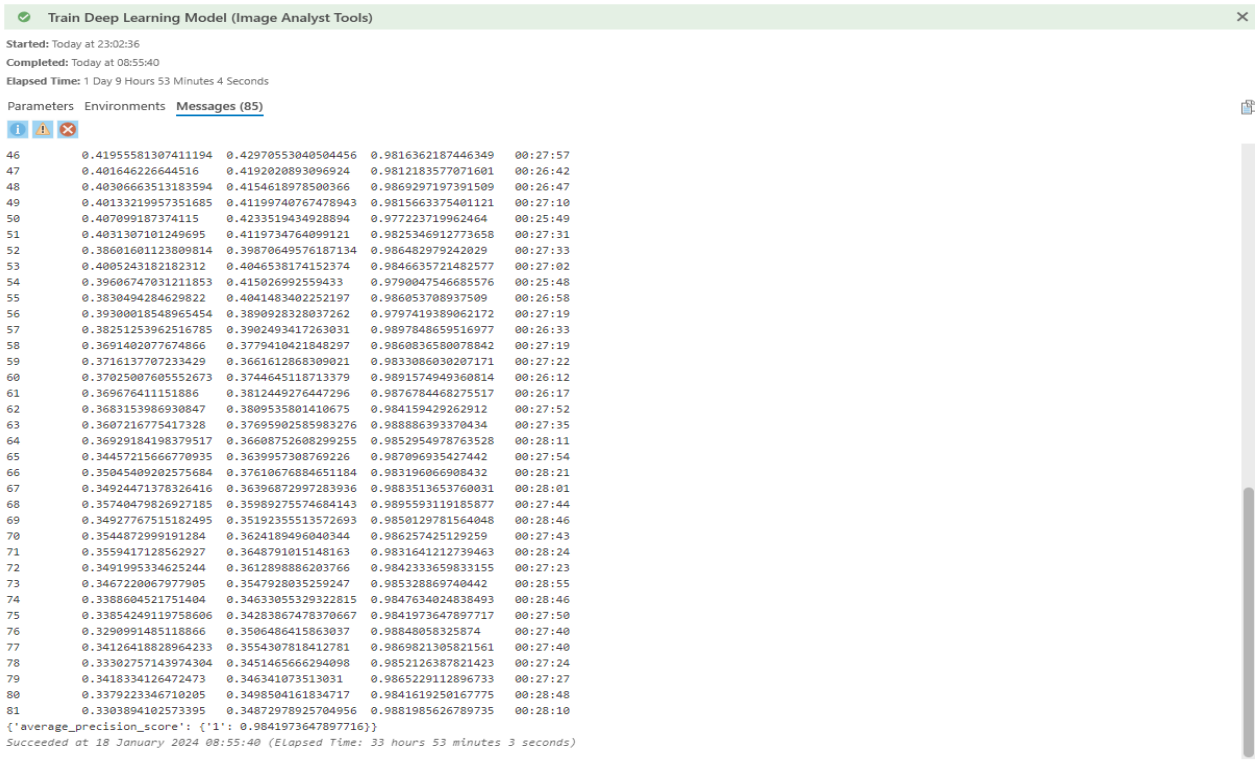


Figure 17 - Deep learning Epoch, Training/Accuracy loss and Epoch Time.

The deep learning model was trained with the proposed Mask R-CNN deep learning architecture in ESRI ArcGIS Pro using the "Train Deep Learning Model (Image Analyst)" geoprocessing tool in a disconnected environment on GPU. In the geoprocessing tool, the folder containing the image chip was loaded as input training data and an output folder was specified. A maximum epoch for training the model was set at 100 however it was programmed to stop training once the model accuracy no longer improved. In the model type, Mask RCNN was specified with the class value as 1. A batch size of 8, chip size of 256 pixels and a learning rate of slice ('1.0965e-05', '1.0965e-04', None). A backbone of ResNet-101 which was trained on the ImageNet dataset was used as backbone. Furthermore, no pretrained model was used in this research as the deep learning model was trained from scratch. The "Train Deep Learning Model (Image Analyst)" geoprocessing tool was set to use 10% of the data set for validation. The weight of the backbone model was not frozen thereby making the weight and biases fit the training samples. The chip size of the images was fixed at 256 pixels. A weight initialisation was performed because the dataset is multispectral data, and the model needs to accommodate the various bands to reinitialise the first layer of the model ESRI 2023. Validation loss was chosen as monitor metric checkpointing and to monitor early

stopping. The training of the deep learning model was done on a computer with Intel Core i7, NDVIA GeForce GTX 1660Ti 6GB dedicated graphic card with a 32GB RAM and dual storage of 500GB solid state drive and 1TB hard disk drive. This training took a total of 33 hours 53 minutes and 3 seconds. Upon completion, it achieved a training accuracy of 98% accuracy of class label "1": solar panels on 81 epochs.

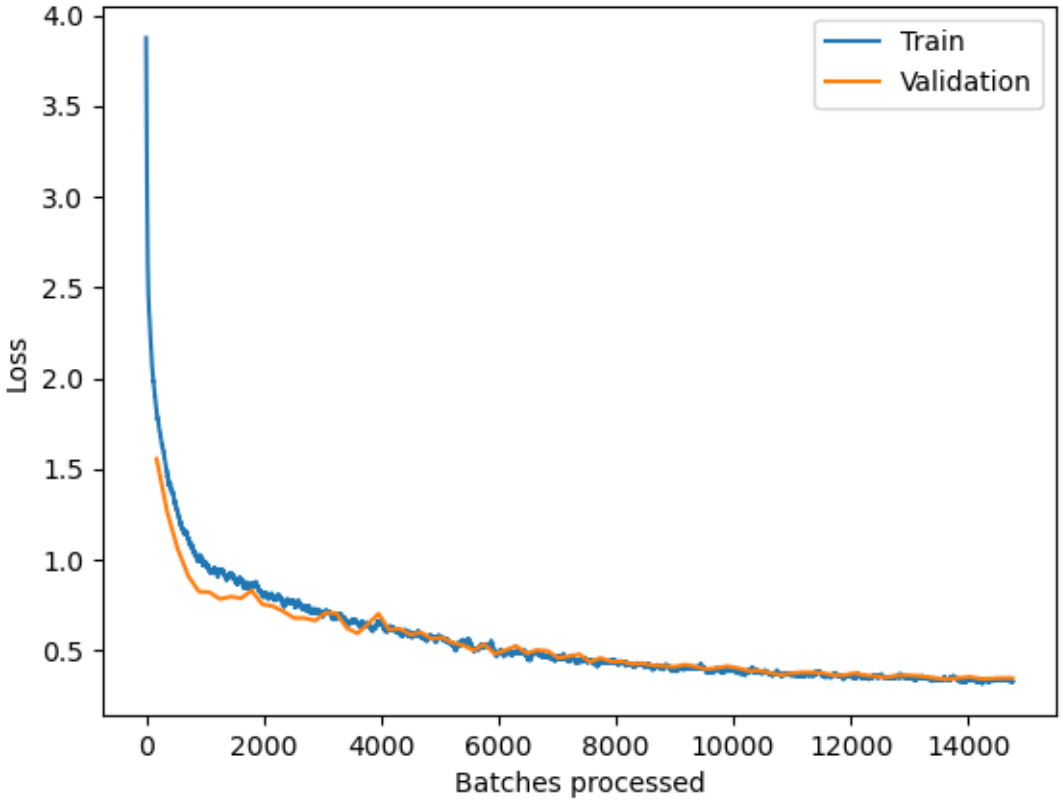


Figure 18 - Training/Validation loss graph of the trained deep learning model using Mask RCNN.

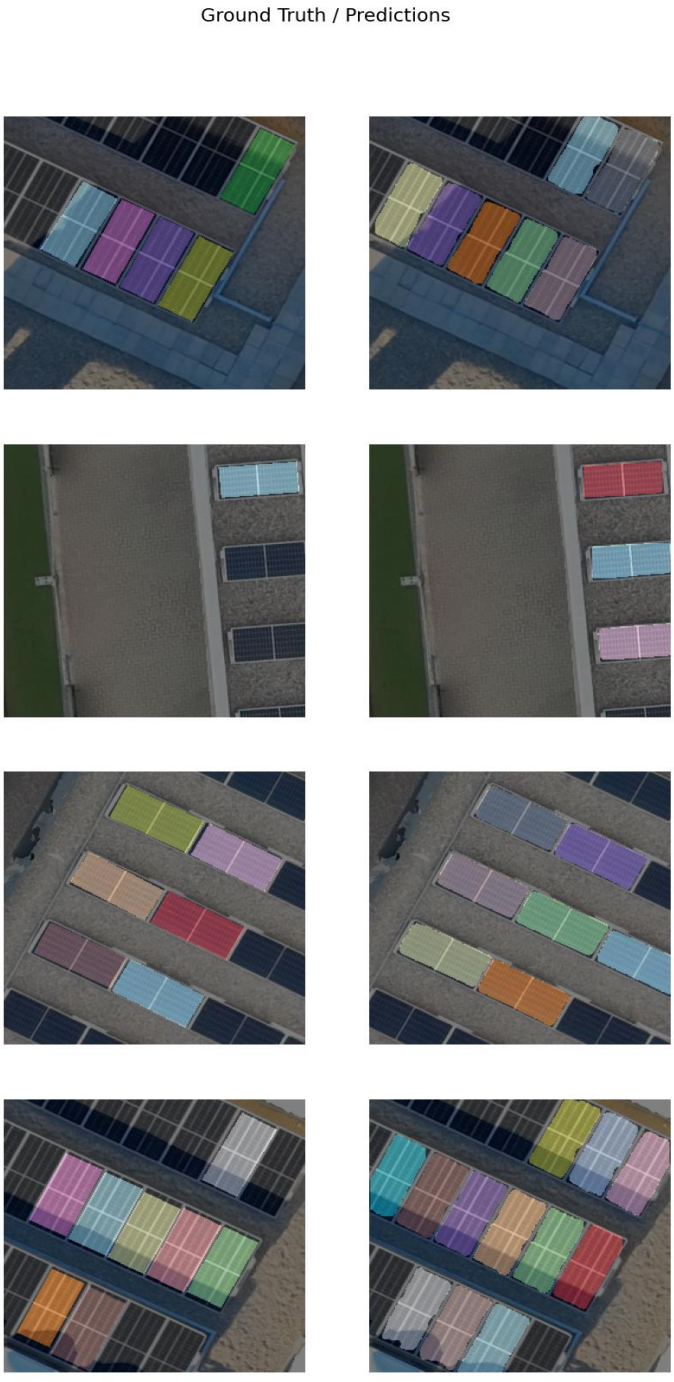


Figure 19 - Image showing deep learning generalisation on the training dataset after training with minimal edge loss.

RESULTS AND DISCUSSION.

Testing Deep Learning Model

It is a common practice as seen in previous literature to test deep learning models on data sets that have not been seen by the deep learning model or else that would amount to bad science. Previous studies by (Feki et al., 2021; Li et al., 2021; Malof et al., 2016, 2017; Wang & Li, 2021) report their accuracy rate on their testing data set hence this research also reports performance of the deep learning model on its testing dataset. Testing data sets are also important for checking for overfitting which (Safonova et al., 2023) defines as a situation where a deep learning model performs poorly on the testing dataset but performs well on the training data set. They further give that test data should be collected independently of the training data set, as this research has done. The trained deep learning model was tested on the testing dataset collected independently of the training dataset using the "Detect Objects Using Deep Learning (Image Analyst)" geoprocessing tool in ArcGIS Pro. This tool produces a bounding box feature class containing the object it finds (ESRI 2023). The multiband true orthophoto of the testing data set was loaded as input raster with a specified output folder. The deep learning model which was trained in this research was defined here as input for model definition. A non-maximum suppression was performed to remove duplicate detected objects. As a result, a new field was created to give a confidence score for detected objects. The cell size used here was the same as the multiband true orthophoto of the testing dataset. This was processed on the computer GPU.

Result

The result of the proposed model shows a promising performance on the test data set. The deep learning model detects nearly all the solar panels in the study area. The model also produced a feature class boundary polygon around each detected solar panel. The model was observed to detect some features that look like solar panels. These false positive detections are dark waffle square shaped objects. Prior to this, the deep learning model was trained without the normalized digital surface model as a fourth raster band. This yielded very poor accuracy with the detection of a lot of false positives and a high degree of false positives. The model was also unable to detect lots of solar panels. But the addition of the normalized digital surface model as the fourth raster band reduced the detection of false positives, thereby improving the precision accuracy of the deep learning model by 10%.

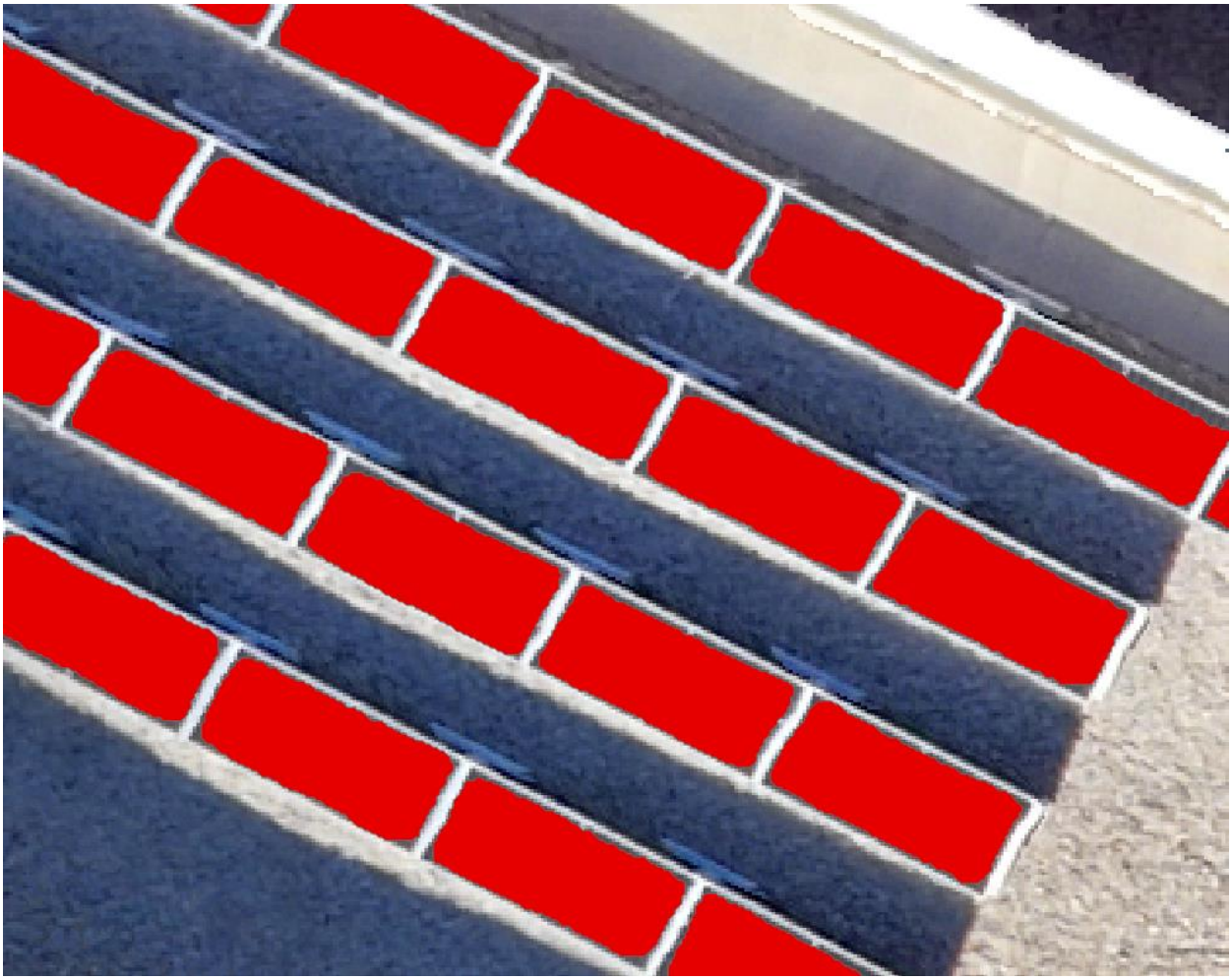
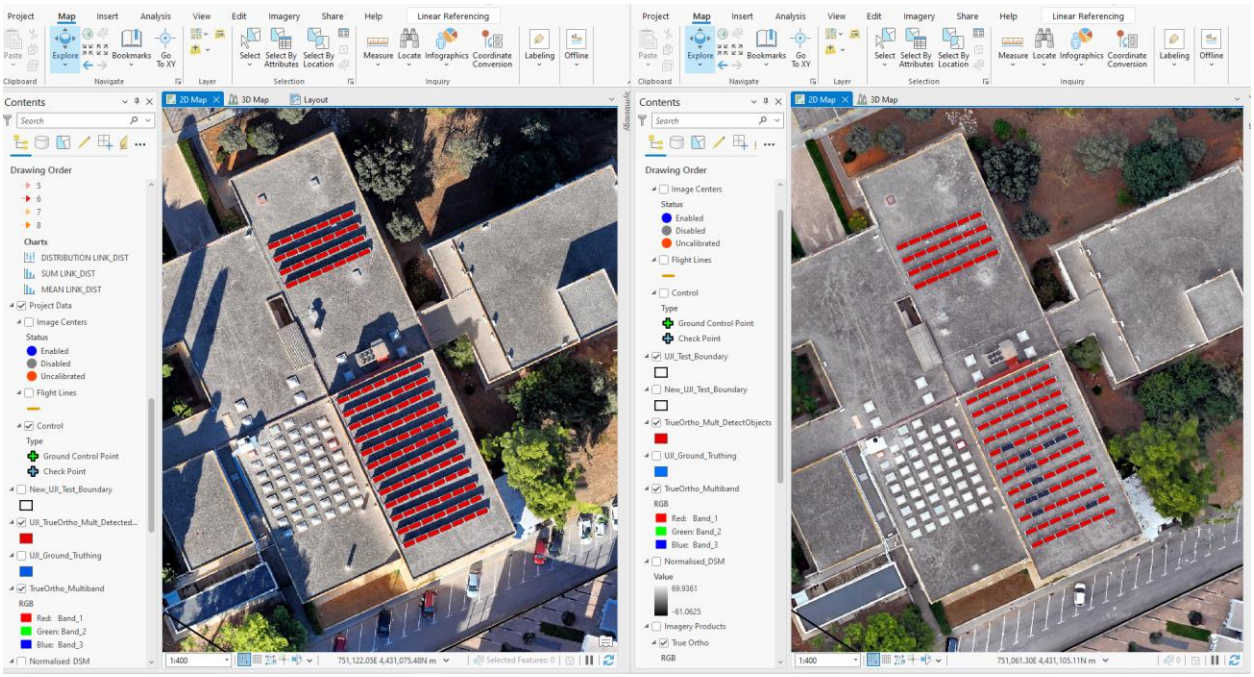


Figure 20 - Deep learning architecture accurately detecting solar panels with minimal edge loss.

This result happens to be different from (Kausika et al., 2021) who affirmed that addition of nDSM showed no significant effect in reducing the detection of false positives in their work.

Table 1 - Accuracy assessment of deep learning model without addition of normalized digital surface model.

IoU >= 0.75	Precision	Recall	F1 Score	AP	True Positive	False Positive	False Negative	Cloud Cover	Clipped Out Homogeneity error
	0.7495	0.7031	0.7256	0.6003	386	129	163	1%	No



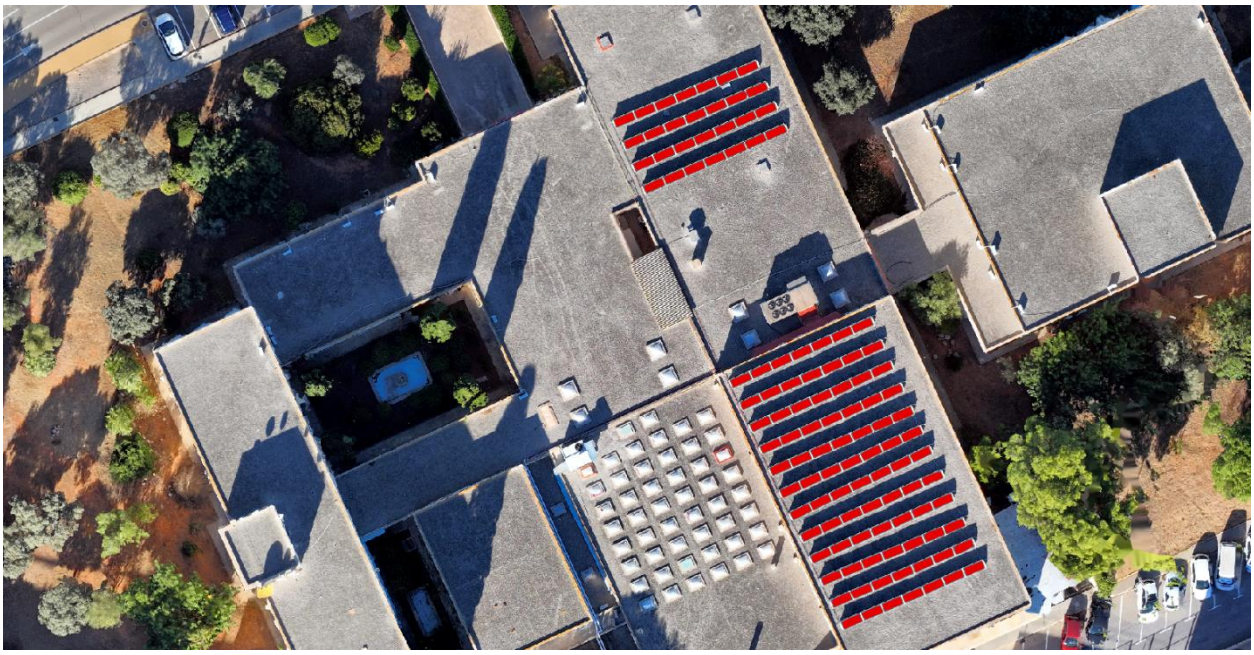


Figure 21 - Image showing performance of the trained deep learning model on the testing dataset with 1% cloud cover in the study area.

Figure 22 - Image showing performance of the trained deep learning model on the testing dataset with 100% cloud cover in the study area.

Accuracy Assessment

Detecting solar panels by physical assessment is not a true test of the performance of the proposed model. For this reason, a standardized deep learning accuracy measurement was used. Previous studies in (Castello et al., 2019; Hou et al., 2019; Li et al., 2021; Sharma et al., 2021; Zhuang et al., 2020) all used the intersection over union to access the performance of their deep learning model. (Castello et al., 2019) particularly puts that the intersect over union is a more suitable metric for computing common pixel area between the prediction and the ground-truth and dividing the area. In the research, the intersect over union was adopted to assess the performance of the model. To do this, the "Compute Accuracy for Object Detection (Image Analyst)" geoprocessing tool in ArcGIS Pro was used. The geoprocessing tool calculates the accuracy of the deep learning model by comparing the detected object to the ground truth objects. ESRI 2024 holds that the accuracy of a model is evaluated using four accuracy metrics: the Average Precision (AP), the F1 score, the COCO mean Average Precision (mAP), and the Precision x Recall curve. The "Compute Accuracy for Object Detection (Image Analyst)" geoprocessing tool takes the feature class of the detected polygon as input for detected features, ground truth hand-annotated feature class as input for

ground truth feature class, class value fields were selected from the feature class attribute field and an intersect over union of 0.2. The result of the computation is given below:

Precision – Precision is the ratio of the number of true positives to the total number of positive predictions (ESRI 2024). Precision is given as (True Positive)/ (True Positive + False Positive)

Recall—Recall is the ratio of the number of true positives to the total number of actual (relevant) objects (ESRI 2024). This given as Recall = (True Positive)/ (True Positive + False Negative)

F1 score— According to (ESRI 2024) the F1 score is a weighted average of the precision and recall. Values range from 0 to 1, where 1 means highest accuracy given as (Precision × Recall)/ [(Precision + Recall)/2]

In the same manner, precision-recall curve is a plot of precision (y-axis) and recall (x-axis), and it serves as an evaluation of the performance of an object detection model. The model is considered a good predictive model if the precision stays high as the recall increases.

In this research, the model with a proposed architecture achieved a precision of 0.8511, a recall of 0.8775 and an F1 score of 0.8641 on an intersection over union of 0.75. This was achieved on the dataset with 1% cloud cover. The precision recall curve was observed to have stayed high as the recall increased. The model was observed to detect 480 true positives, 84 false positives and 67 false negatives. The model was observed to accurately delineate the boundary for each solar panel it detected. A percentage of false positives were seen over waffle-like objects with shadow cover and objects that take the shape and form of a solar panel. It can be said that if the aerial imagery were to be collected off shadow cast, the accuracy of the model would drastically improve. The model was also not able to detect some solar panels resulting in false negative.

To again test the accuracy of the proposed deep learning model in diverse real-world scenarios, the model was operated on the dataset with 100% cloud cover. The proposed architecture achieved a precision of 0.7594, a recall of 0.4864 and an F1 score of 0.593 on an intersection over union of 0.75. This was achieved on the dataset with 100 % cloud cover. The precision recall curve was observed to have stayed high as the recall increased however it dropped midpoint. The model was observed detecting 161 true positives, 51 false positives and 170 false negatives. Here, the model was observed to have also accurately delineated the boundary of each solar panel it detected but

had difficulty recognising 170 solar panels. A detailed assessment was carried out to understand the decline in F1 score as well as an increase in false negatives. It was observed that during the photogrammetric image processing, the multiband true orthophoto for data with 100% cloud cover was unable to properly reconstruct a section of the study area which contained many solar panels.

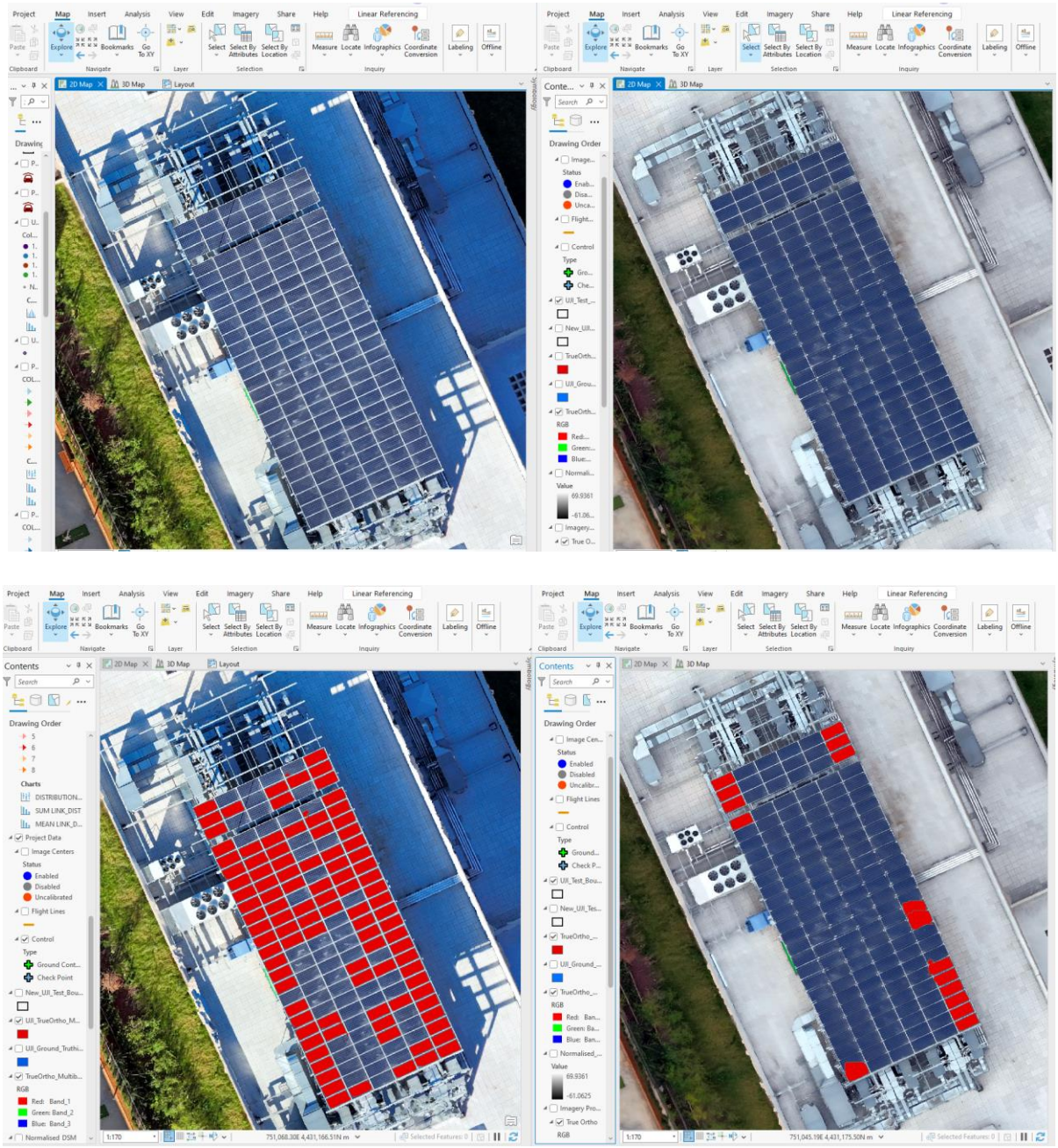


Figure 23 – Image showing reconstruction problem due to homogeneity problem in dataset with 100% cloud cover.

This problem has been described by (Starek et al., 2014) as emanating from the confusion and inability of the SIFT point matching algorithm to detect corresponding key points within overlapping images due to homogeneity of image texture. This is also a typical problem when flying over water. In this case it cannot be safe to assume that the aforementioned accuracy is a true accuracy assessment for the deep learning model.

A proper delineation of the multiband true ortho photo testing dataset was done to clip off area of the imagery with bad image reconstruction as it amounted to poor image quality. The proposed deep learning model was again deployed on the clipped multiband true ortho photo for the testing data set. Having clipped the multiband image, the ground truth was again clipped to contain the new area with good image quality. Upon deployment of the deep learning model with the proposed deep learning architecture on the imagery with 1% cloud cover, a drastic improvement was observed. The proposed deep learning architecture was now observed to achieve a precision of 0.8482, a recall of 0.9719 and an F1 score of 0.8445 on an intersection over union of 0.75. This was achieved on the new clipped dataset with proper image reconstruction having 1% cloud cover. The model was observed to detect 380 true positives, 68 false positives and only 11 false negatives. On the clipped multiband true ortho photo with 100% cloud cover, the proposed deep learning architecture was now observed to achieve a precision of 0.8409, a recall of 0.8605 and an F1 score of 0.8506 on an intersection over union of 0.75. The model was observed to detect 148 true positives, 28 false positives and 24 false negatives. The model was observed to accurately delineate the boundary to each solar panel it detected.

The accuracy assessment reports on multiband true ortho photos with image reconstruction problems and without image reconstruction problems.

Table 2 – Accuracy table of multiband true orthophoto testing dataset with 1% cloud cover.

IoU >= 0.75	Precision	Recall	F1 Score	AP	True Positive	False Positive	False Negative	Cloud Cover	Clipped Out Image reconstruction error
	0.8511	0.8775	0.8641	0.7684	480	84	67	1%	No
	0.8482	0.9719	0.9058	0.8445	380	68	11	1%	Yes
Difference	0.0029	0.1	0.0417	0.0761	-100	-16	-56		

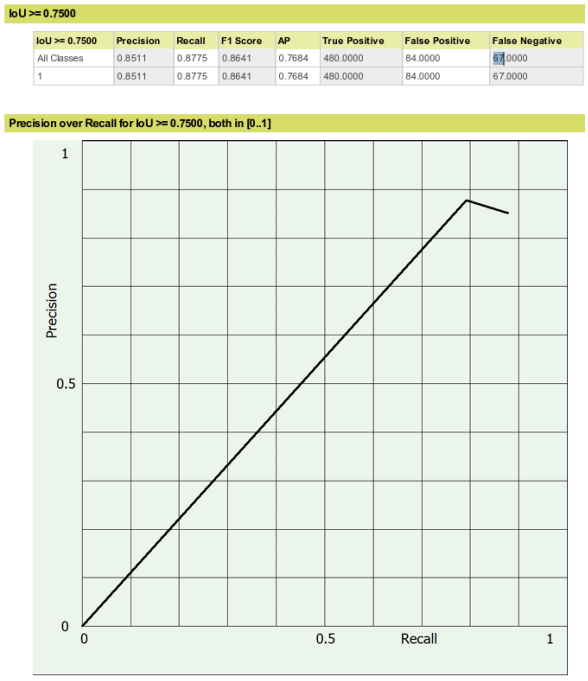


Figure 24 - Precision over recall graph.

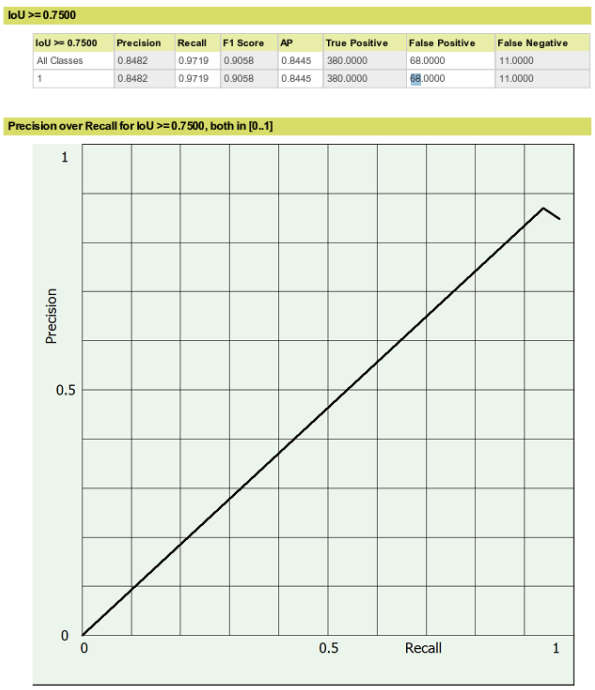
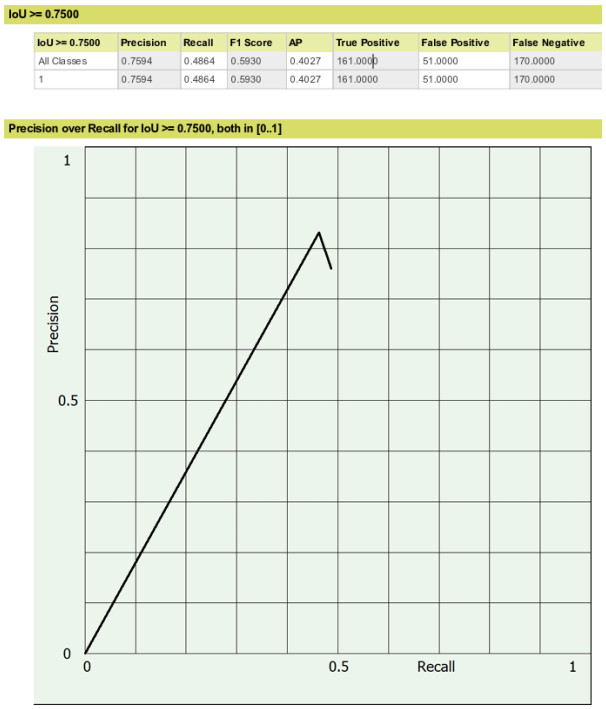


Figure 25 - Precision over recall graph.

From the result seen here when the model was operated on the multiband true ortho training dataset with the clipped-out area, no significant increase was observed in the precision score. However, a great improvement was seen in the Recall score of about 0.1 and a 0.04 increase in F1 score.

Table 3 - Accuracy table of multiband true orthophoto testing dataset with 100% cloud cover.

IoU >= 0.75	Precision	Recall	F1 Score	AP	True Positive	False Positive	False Negative	Cloud Cover	Clipped Out Homogeneity error
	0.7594	0.4864	0.5930	0.4027	161	51	170	100%	No
	0.8409	0.8605	0.8506	0.7730	148	28	24	100%	Yes
Difference	0.0815	0.3741	0.2576	0.3703	-13	-23	-146		



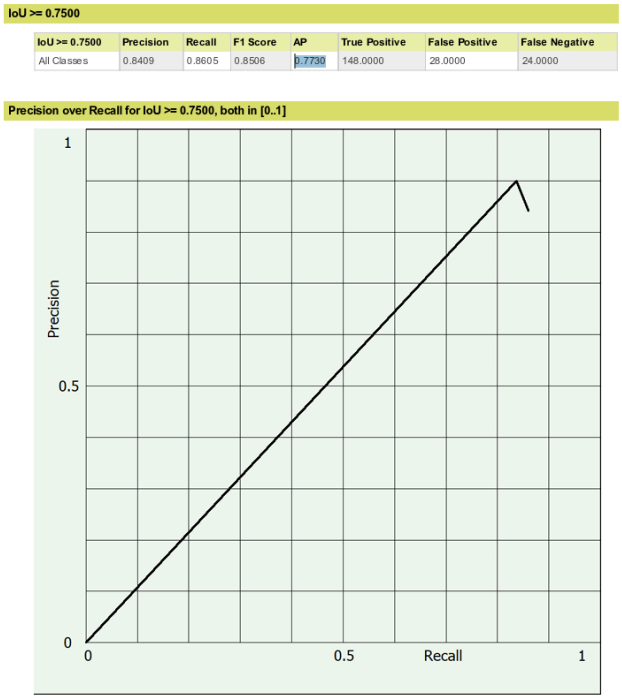


Figure 26 - Precision over recall graph

Figure 27 - Precision over recall graph.

When the area with reconstruction problem was clipped out in the testing dataset with 100% cloud cover, the model was observed to have performed well enough with 0.08 increase in precision and 0.37 increase in Recall and 0.25 increase in F1 score. It can be generally seen from the table above that poor image reconstruction problems from the SIFT algorithm confusion have a big impact on the performance of the deep learning model. This problem could have emanated from low lighting condition when aerial data was collected in 100% cloud cover or the general know reconstruction

problem in the SIFT algorithm. For fairness in research, having stated the problem of image reconstruction from the SIFT algorithm, this research takes the clipped testing dataset with 1% cloud clover and takes the clipped off testing dataset with 100% cloud cover as second testing data set to test the generalization of the deep learning model. This research therefore reports its proposed deep learning model accuracy as follows.

Table 4 - Accuracy table of multiband true orthophoto testing dataset with diverse weather condition.

IoU >= 0.75	Precision	Recall	F1 Score	AP	True Positive	False Positive	False Negative	Cloud Cover	Clipped Out Homogeneity error
	0.8482	0.9719	0.9058	0.8445	380	68	11	1%	Yes
	0.8409	0.8605	0.8506	0.7730	148	28	24	100%	Yes
Mean	0.8445	0.9162	0.8782	0.8087	264	48	17		

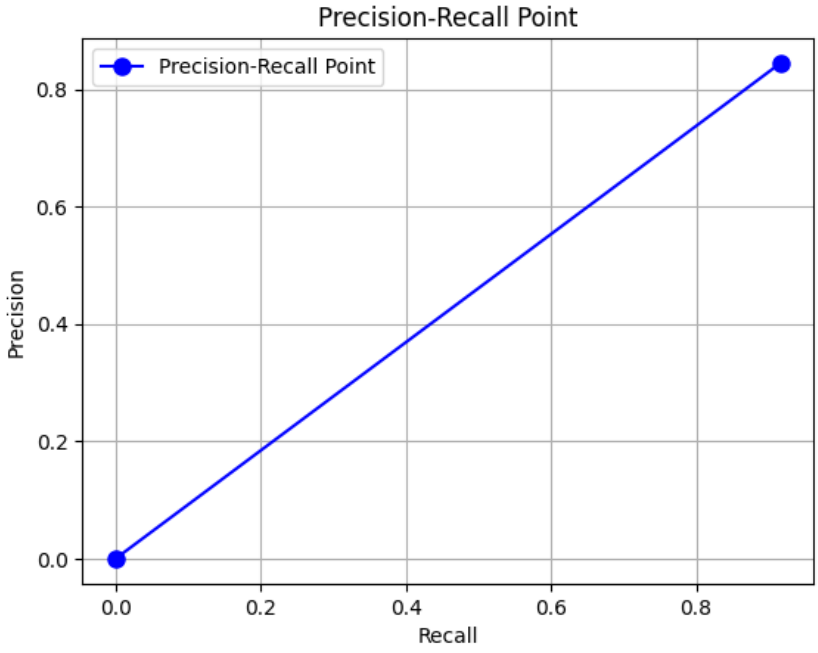


Figure 28 - Precision over recall graph for mean accuracy of the study.

The research findings unveil the robustness of the deep learning model, showcasing its adaptability across various weather conditions and achieving an impressive mean accuracy of over 84% in both test datasets. Despite reconstruction problems from aerial imagery, the proposed deep learning model trained in diverse weather condition had a precision score 0.8445, a Recall of 0.9162, an F1 score of 0.8782 and MaP of 0.8087 on a high intersection over union of 0.75 as against previous research who report intersection over union less than 0.75. The model was observed to have detected an average of 264 true positives, 48 false positive and 17 false negatives. In comparison with the small size of training dataset used in training this model, it is fair to the model to have achieved this impressive feat in comparison with previous research work who have used dataset as large as an entire state like Connecticut, North Rhine Westphalia and even the entire country like Netherlands, Sweeden and Switzerland while reporting detection on 0.5 IOU. This research had only used a university campus the size of only 1.74km square and a high intersection over union of 0.75. This research compares with previous research as follows.

Table 5 – Comparison of previous study on the detection solar panels

Study	Model	Country	Precision [%]	Recall [%]	F1 [%]	Image resolution [m/ pixel]	Number of images evaluated	Share of positive samples of the images [%]
(Yuan et al., 2016)	ConvNet	US	81.2 / 85.5	84.0 / 87.3	82.6 / 86.4	0.30		
(Malof et al., 2019)	SolarMapper	US	76.00	77.00	76.00	0.30		
(Yu et al., 2018)	DeepSolar	US	93.1 / 93.7	88.5 / 90.5	90.7 / 92.1	0.15	93 500	1.31
(Mayer et al., 2020)	DeepSolar	US	91.0	98.1	94.4	0.05	3 798	4.08
(Kausika et al., 2021)	TernausNet	the Netherlands	93.1	90.7	91.9	0.10	2 791 904	5.61
(Rausch et al., 2020)	DeepSolar	Germany	87.3	87.5	87.4	0.10	45 060	0.85
(Lindahl et al., 2023)	DeepSolar — CNN	Germany	93.4	81.3	86.9	0.10	45 060	0.85
(Lindahl et al., 2023)	DeepSolar — CNN	Sweden	63.9	81.8	71.1	0.16	877 142	0.09
This Study	Mask RCNN	Spain	84.45	91.62	87.82	0.03	417	0.1

While this research achieves remarkable results, challenges exist for future studies. It is the first time a deep learning model for the detection of solar panels has been implemented in the Kingdom of Spain, trained with diverse weather conditions. This research is a demonstration that the proposed deep learning model can be used to monitor the growth of solar panels at the university (UJI) in Spain. Moreover, given that this deep learning model was implemented upon instance segmentation, the volume of electricity which solar panels contribute to the campus can also be accurately computed. With this, government agencies, policy makers and financial institutions can

adequately determine homeowners who have installed solar panels on their roof top and as well give incentives adequately to them.

Spatial Distribution of Solar Panels

A spatial distribution of the solar panels was done to map the concentration of solar panels in the study area through point data. A centroid of the solar panels was calculated and interpolated as heat maps to show areas with sparsely populated solar panels and areas with densely populated solar panels. The point data was also aggregated to count the number of solar panels hosted by buildings. The building called Edifici JC 2 on the campus was observed to be hosting 93 solar panels. The Edifici TC building also hosts 196 solar panels, col-legi d Educacio infantile I primaria Vincent Marca host 111 solar panels on its building while Escola official d idiomes de castelló host 112 solar panels. From the spatial distribution of solar panels in the test dataset, urban planners can adequately locate urban furniture that requires the utilization of solar panels. For example, the parking lots that share proximity with densely populated solar panels can be installed with car charging points for electric vehicles in those parking lots.

Figure 29 - Spatial distribution of solar panel clusters in the testing dataset.

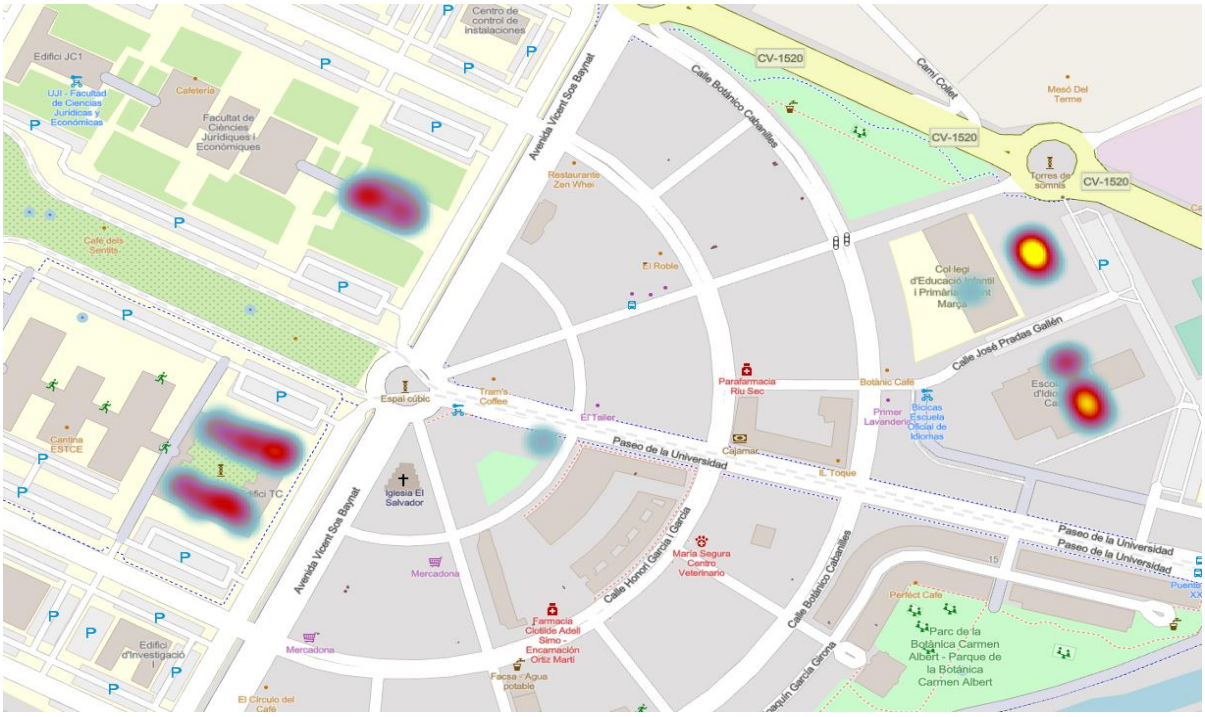


Figure 30 - Heat map showing cluster density of solar panels in the study area.

Figure 31 - Heat map showing cluster quantification of solar panels in the study area.

Conclusion.

The findings from the development of an instance segmentation deep learning model for the detection of solar panels, shows that instance segmentation provides high accuracy for the detection of solar panels with less training time and minimal edge loss using small dataset from very high-resolution UAV imagery. It is important for users to note that detecting solar panels in industrial areas may pose significant problem because of image reconstruction problem of solar panels and homogeneity of texture using VHR imagery but only in low lighting condition. If solar panels of large areas must be detected, the photogrammetry process must be conducted to avoid image reconstruction problems due to low lighting condition and homogeneity of image texture. Instance segmentation deep learning models which combine object detection and image segmentation proves to generalize properly with a small amount of training dataset as opposed to the popular use of large datasets to train deep learning models. Again, despite image reconstruction problems, the deep learning model performs well in real world scenario. This research affirms that training models in diverse weather condition improves generalization of deep learning model to accurately detect solar panel when faced with real world scenarios. Since the deep learning model can detect solar panels of all instances, subclasses of solar panel variant can be created to detect granularity of the instances of the panels. With accurate detection of solar panels and a high intersection over union at 0.75, the deep learning model delineate accurately a bounding box over each solar panel with high confidence. This high confidence of the model shows that the model can be deployed to calculate the amount of electricity generated by each solar panel. The fear of very high-resolution imagery creating unwanted multiclass during segmentation does not exist as this did not happen in this research. This research also proves that false positives can be eliminated greatly by the addition of a nDSM to the training dataset as fourth band to eliminate false positives. Homogeneity of image texture poses a problem to solar panel reconstruction and solar panel detection. Future studies in solar panels detection are encouraged to train deep learning models in diverse weather conditions while incorporating a nDSM to enhance model generalization.

REFERENCES

Ahmad, A., & Samad, A. M. (2010). Aerial mapping using high resolution digital camera and unmanned aerial vehicle for Geographical Information System. 2010 6th International Colloquium on Signal Processing & Its Applications, 1–6. https://doi.org/10.1109/CSPA.2010.5545303

Ali, H. H., & Abed, F. M. (2019). The impact of UAV flight planning parameters on topographic mapping quality control. IOP Conference Series: Materials Science and Engineering, 518(2), 022018. https://doi.org/10.1088/1757-899X/518/2/022018

Anuar A, Khairil A H & Abd. M S (2010). Aerial Mapping using High Resolution Digital Camera and Unmanned Aerial Vehicle for Geographical Information System. 2010 6th International Colloquium on Signal Processing & Its Applications (CSPA)

Bal, A.; Palus, H. (2023). Image Vignetting Correction Using a Deformable Radial Polynomial Model. Sensors 2023, 23, 1157. https://doi.org/10.3390/s23031157

Bazi, Y., & Melgani, F. (2018). Convolutional SVM Networks for Object Detection in UAV Imagery. IEEE Transactions on Geoscience and Remote Sensing, 56(6), 3107–3118. https://doi.org/10.1109/TGRS.2018.2790926

Bookstein, F. L. (1991). Morphometric tools for landmark data: Geometry and biology. Cambridge University Press.

Cabo, C., Sanz-Ablanedo, E., Roca-Pardinas, J., & Ordonez, C. (2021). Influence of the Number and Spatial Distribution of Ground Control Points in the Accuracy of UAV-SfM DEMs: An Approach Based on Generalized Additive Models. IEEE Transactions on Geoscience and Remote Sensing, 59(12), 10618–10627. https://doi.org/10.1109/TGRS.2021.3050693

Castello, R., Roquette, S., Esguerra, M., Guerra, A., & Scartezzini, J.-L. (2019). Deep learning in the built environment: Automatic detection of rooftop solar panels using Convolutional Neural Networks. Journal of Physics: Conference Series, 1343(1), 012034. https://doi.org/10.1088/1742-6596/1343/1/012034

Chen, C., Zhang, C., Wu, W., Jiang, W., Tian, B., & Zhou, Y. (2022). Application of UAV-Based Photogrammetry Without Ground Control Points in Quantifying Intertidal Mudflat Morpho dynamics. IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium, 7767–7770. https://doi.org/10.1109/IGARSS46834.2022.9884346

Daus, Y., & Yudaev, I. (2017). Estimation of Solar Energy Potential under Conditions of Urban Development. Proceedings of the International Conference 'Actual Issues of Mechanical Engineering' 2017 (AIME 2017). International Conference 'Actual Issues of Mechanical Engineering' 2017 (AIME 2017), Tomsk, Russia. https://doi.org/10.2991/aime-17.2017.26

Elazami Elhassani, M., Maisonnasse, L., Olgiati, A., Jerome, R., Rehali, M., Duroux, P., Giudicelli, V., & Kossida, S. (2022). Deep Learning concepts for genomics: An overview. EMBnet.Journal, 27. https://doi.org/10.14806/ej.27.0.990

Feki, I., Ammar, S., Kessentini, Y., & Muhammad, K. (2021). Federated learning for COVID-19 screening from Chest X-ray images. Applied Soft Computing, 106, 107330. https://doi.org/10.1016/j.asoc.2021.107330

Ferreira, M. P., Almeida, D. R. A. D., Papa, D. D. A., Minervino, J. B. S., Veras, H. F. P., Formighieri, A., Santos, C. A. N., Ferreira, M. A. D., Figueiredo, E. O., & Ferreira, E. J. L. (2020). Individual tree detection and species classification of Amazonian palms using UAV images and deep learning. Forest Ecology and Management, 475, 118397. https://doi.org/10.1016/j.foreco.2020.118397

Gao, S. (2021). Geospatial Artificial Intelligence (GeoAI). In S. Gao, Geography. Oxford University Press. https://doi.org/10.1093/obo/9780199874002-0228

Girshick, R. (2015). Fast R-CNN. 2015 IEEE International Conference on Computer Vision (ICCV), 1440–1448. https://doi.org/10.1109/ICCV.2015.169

Golovko, V., Bezobrazov, S., Kroshchanka, A., Sachenko, A., Komar, M., & Karachka, A. (2017). Convolutional neural network based solar photovoltaic panel detection in satellite photos. 2017 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), 14–19. https://doi.org/10.1109/IDAACS.2017.8094501

Gonçalves, J. A., & Henriques, R. (2015). UAV photogrammetry for topographic monitoring of coastal areas. ISPRS Journal of Photogrammetry and Remote Sensing, 104, 101–111. https://doi.org/10.1016/j.isprsjprs.2015.02.009

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2018). Mask R-CNN (arXiv:1703.06870). arXiv. http://arxiv.org/abs/1703.06870

Hongjun Tan, Zhiling Guo, Haoran Zhang, Qi Chen, Zhenjia Lin, Yuntian Chen, Jinyue Yan (2023).Enhancing PV panel segmentation in remote sensing images with constraint refinement modules, Applied Energy, Volume 350, 2023, 121757, ISSN 0306-2619, https://doi.org/10.1016/j.apenergy.2023.121757. https://www.sciencedirect.com/science/article/pii/S0306261923011212)

Hu, J., Wang, Z., Pan, X., Cong, P., Yu, H., & Chu, J. (2022). Solar Panels Detection of High-Resolution Aerial Images Based on Improved Faster-RCNN. IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium, 3544–3547. https://doi.org/10.1109/IGARSS46834.2022.9884399

Hu, W., Bradbury, K., Malof, J. M., Li, B., Huang, B., Streltsov, A., Sydny Fujita, K., & Hoen, B. (2022). What you get is not always what you see—Pitfalls in solar array assessment using overhead imagery. Applied Energy, 327, 120143. https://doi.org/10.1016/j.apenergy.2022.120143

Iheaturu, C., Okolie, C., Ayodele, E., Egogo-Stanley, A., Musa, S., & Ifejika Speranza, C. (2022). A simplified structure-from-motion photogrammetry approach for urban development analysis. Remote Sensing Applications: Society and Environment, 28, 100850. https://doi.org/10.1016/j.rsase.2022.100850

Ismael, R. Q., & Henari, Q. Z. (2019). Accuracy Assessment of UAV photogrammetry for Large Scale Topographic Mapping. 2019 International Engineering Conference (IEC), 1–5. https://doi.org/10.1109/IEC47844.2019.8950607

Janowicz, K., Gao, S., McKenzie, G., Hu, Y., & Bhaduri, B. (2020). GeoAI: Spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond. International Journal of Geographical Information Science, 34(4), 625–636. https://doi.org/10.1080/13658816.2019.1684500

Jayathunga, S., Owari, T., & Tsuyuki, S. (2018). The use of fixed–wing UAV photogrammetry with LiDAR DTM to estimate merchantable volume and carbon stock in living biomass over a mixed conifer–broadleaf forest. International Journal of Applied Earth Observation and Geoinformation, 73, 767–777. https://doi.org/10.1016/j.jag.2018.08.017

Kannan, N., & Vakeesan, D. (2016). Solar energy for future world: - A review. Renewable and Sustainable Energy Reviews, 62, 1092–1105. https://doi.org/10.1016/j.rser.2016.05.022

Kausika, B. B., Nijmeijer, D., Reimerink, I., Brouwer, P., & Liem, V. (2021). GeoAI for detection of solar photovoltaic installations in the Netherlands. Energy and AI, 6, 100111. https://doi.org/10.1016/j.egyai.2021.100111

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539

Li, P., Zhang, H., Guo, Z., Lyu, S., Chen, J., Li, W., Song, X., Shibasaki, R., & Yan, J. (2021). Understanding rooftop PV panel semantic segmentation of satellite and aerial images for better using machine learning. Advances in Applied Energy, 4, 100057. https://doi.org/10.1016/j.adapen.2021.100057

Liao, J., Yang, G., Zhang, S., Zhang, F., & Gong, C. (2021). A Deep Reinforcement Learning Approach for the Energy-Aimed Train Timetable Rescheduling Problem Under Disturbances. IEEE Transactions on Transportation Electrification, 7(4), 3096–3109. https://doi.org/10.1109/TTE.2021.3075462

Lindahl, J., Johansson, R., & Lingfors, D. (2023). Mapping of decentralised photovoltaic and solar thermal systems by remote sensing aerial imagery and deep machine learning for statistic generation. Energy and AI, 14, 100300. https://doi.org/10.1016/j.egyai.2023.100300

Luo, K., Cao, J., Wang, C., Cai, S., Yu, R., Wu, M., Yang, B., & Xiang, W. (2022). First unmanned aerial vehicle airborne gravimetry based on the CH-4 UAV in China. Journal of Applied Geophysics, 206, 104835. https://doi.org/10.1016/j.jappgeo.2022.104835

Malof, J. M., Bradbury, K., Collins, L. M., & Newell, R. G. (2016). Automatic detection of solar photovoltaic arrays in high resolution aerial imagery. Applied Energy, 183, 229–240. https://doi.org/10.1016/j.apenergy.2016.08.191

Malof, J. M., Collins, L. M., & Bradbury, K. (2017). A deep convolutional neural network, with pre-training, for solar photovoltaic array detection in aerial imagery. 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 874–877. https://doi.org/10.1109/IGARSS.2017.8127092

Malof, J. M., Li, B., Huang, B., Bradbury, K., & Stretslov, A. (2019). Mapping solar array location, size, and capacity using deep learning and overhead imagery.

Mao, P., Jiang, B., Shi, Z., He, Y., Shen, T., & Qiu, G. Y. (2023). Effects of UAV flight height on biomass estimation of desert shrub communities. Ecological Indicators, 154, 110698. https://doi.org/10.1016/j.ecolind.2023.110698

Mayer, K., Wang, Z., Arlt, M.-L., Neumann, D., & Rajagopal, R. (2020). DeepSolar for Germany: A deep learning framework for PV system mapping from aerial imagery. 2020 International Conference on Smart Energy Systems and Technologies (SEST), 1–6. https://doi.org/10.1109/SEST48500.2020.9203258

Moradi A. M., Sizkouhi, M. Aghaei, S. M. Esmailifar, M. R. Mohammadi and F. Grimaccia, "Automatic Boundary Extraction of Large-Scale Photovoltaic Plants Using a Fully Convolutional Network on Aerial Imagery," in *IEEE Journal of Photovoltaics*, vol. 10, no. 4, pp. 1061-1067, July 2020, doi: 10.1109/JPHOTOV.2020.2992339

Nadarajah, K., Divagar Vakeesan, Solar energy for future world: - A review, Renewable and Sustainable Energy Reviews, Volume 62, 2016, Pages 1092-1105, ISSN 1364-0321, https://doi.org/10.1016/j.rser.2016.05.022. (https://www.sciencedirect.com/science/article/pii/S1364032116301320)

Nasrallah, H., Samhat, A. E., Shi, Y., Zhu, X. X., Faour, G., & Ghandour, A. J. (2022). Lebanon Solar Rooftop Potential Assessment Using Buildings Segmentation From Aerial Images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 15, 4909–4918. https://doi.org/10.1109/JSTARS.2022.3181446

Nasrallah, H., Samhat, A. E., Shi, Y., Zhu, X., Faour, G., & Ghandour, A. J. (2022). Lebanon Solar Rooftop Potential Assessment using Buildings Segmentation from Aerial Images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 15, 4909 4918. https://doi.org/10.1109/JSTARS.2022.3181446

Nex, F., Armenakis, C., Cramer, M., Cucci, D. A., Gerke, M., Honkavaara, E., Kukko, A., Persello, C., & Skaloud, J. (2022). UAV in the advent of the twenties: Where we stand and what is next. ISPRS Journal of Photogrammetry and Remote Sensing, 184, 215–242. https://doi.org/10.1016/j.isprsjprs.2021.12.006

Olivatto, T. F., Inguaggiato, F. F., & Stanganini, F. N. (2023). Urban mapping and impacts assessment in a Brazilian irregular settlement using UAV-based imaging. Remote Sensing Applications: Society and Environment, 29, 100911. https://doi.org/10.1016/j.rsase.2022.100911

Puttemans, S., Ranst, W. V., & Goedeme, T. (n.d.). Detection Of Photovoltaic Installations in Rgb Aerial Imaging: A Comparative Study.

Rausch, B., Mayer, K., Arlt, M.-L., Gust, G., Staudt, P., Weinhardt, C., Neumann, D., & Rajagopal, R. (2020). An Enriched Automated PV Registry: Combining Image Recognition and 3D Building Data (arXiv:2012.03690). arXiv. http://arxiv.org/abs/2012.03690

Saleri, R., Cappellini, V., Nony, N., De Luca, L., Pierrot-Deseilligny, M., Bardiere, E., & Campi, M. (2013). UAV photogrammetry for archaeological survey: The Theatres area of Pompeii. 2013 Digital Heritage International Congress (Digital Heritage), 497–502. https://doi.org/10.1109/DigitalHeritage.2013.6744818

Samad, Abd. M., Kamarulzaman, N., Hamdani, M. A., Mastor, T. A., & Hashim, K. A. (2013). The potential of Unmanned Aerial Vehicle (UAV) for civilian and mapping application. 2013 IEEE 3rd International Conference on System Engineering and Technology, 313–318. https://doi.org/10.1109/ICSEngT.2013.6650191

Șandric, I., Irimia, R., Petropoulos, G. P., Anand, A., Srivastava, P. K., Pleșoianu, A., Faraslis, I., Stateras, D., & Kalivas, D. (2022). Tree's detection & health's assessment from ultra-high resolution UAV imagery and deep learning. Geocarto International, 37(25), 10459–10479. https://doi.org/10.1080/10106049.2022.2036824

Starek, M. J., Davis, T., Prouty, D., & Berryhill, J. (2014). Small-scale UAS for geoinformatics applications on an island campus. 2014 Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS), 120–127. https://doi.org/10.1109/UPINLBS.2014.7033718

Sunarya, I. M. G., Al Affan, M. R., Kurniawan, A., & Yuniarno, E. M. (2020). Digital Map Based on Unmanned Aerial Vehicle. 2020 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM), 211–216. https://doi.org/10.1109/CENIM51130.2020.9297883

Tahar, K. N., Ahmad, A., & Wan Abdul Aziz Wan Mohd Akib. (2011). UAV-based stereo vision for photogrammetric survey in aerial terrain mapping. 2011 IEEE International Conference on Computer Applications and Industrial Electronics (ICCAIE), 443–447. https://doi.org/10.1109/ICCAIE.2011.6162176

Udin, W. S., Hassan, A. F., Ahmad, A., & Tahar, K. N. (2012). Digital Terrain Model extraction using digital aerial imagery of Unmanned Aerial Vehicle. 2012 IEEE 8th International Colloquium on Signal Processing and Its Applications, 272–275. https://doi.org/10.1109/CSPA.2012.6194732

Usery, E. L., Arundel, S. T., Shavers, E., Stanislawski, L., Thiem, P., & Varanka, D. (2022). GeoAI in the US Geological Survey for topographic mapping. Transactions in GIS, 26(1), 25–40. https://doi.org/10.1111/tgis.12830

Wang, M., Cui, Q., Sun, Y., & Wang, Q. (2018). Photovoltaic panel extraction from very high-resolution aerial imagery using region–line primitive association analysis and template matching. ISPRS Journal of Photogrammetry and Remote Sensing, 141, 100–111. https://doi.org/10.1016/j.isprsjprs.2018.04.010

Wang, S., & Li, W. (2021). GeoAI in terrain analysis: Enabling multi-source deep learning and data fusion for natural feature detection. Computers, Environment and Urban Systems, 90, 101715. https://doi.org/10.1016/j.compenvurbsys.2021.101715

Wang, Y., Bashir, S. M. A., Khan, M., Ullah, Q., Wang, R., Song, Y., Guo, Z., & Niu, Y. (2022). Remote sensing image super-resolution and object detection: Benchmark and state of the art. Expert Systems with Applications, 197, 116793. https://doi.org/10.1016/j.eswa.2022.116793

Watanabe, Y., & Kawahara, Y. (2016). UAV Photogrammetry for Monitoring Changes in River Topography and Vegetation. Procedia Engineering, 154, 317–325. https://doi.org/10.1016/j.proeng.2016.07.482

Wei, Q., & Qingna, L. (2021). Construction of cultural industry development factor model based on factor analysis, artificial intelligence, and big data. Microprocessors and Microsystems, 82, 103880. https://doi.org/10.1016/j.micpro.2021.103880

Yao, H., Qin, R., & Chen, X. (2019). Unmanned Aerial Vehicle for Remote Sensing Applications—A Review. Remote Sensing, 11(12), 1443. https://doi.org/10.3390/rs11121443

Yu, J., Wang, Z., Majumdar, A., & Rajagopal, R. (2018). DeepSolar: A Machine Learning Framework to Efficiently Construct a Solar Deployment Database in the United States. Joule, 2(12), 2605–2617. https://doi.org/10.1016/j.joule.2018.11.021

Yuan, J., Yang, H.-H. L., Omitaomu, O. A., & Bhaduri, B. L. (2016). Large-scale solar panel mapping from aerial images using deep convolutional networks. 2016 IEEE International Conference on Big Data (Big Data), 2703–2708. https://doi.org/10.1109/BigData.2016.7840915

Yuliia Daus and Igor Yudaev (2017) Estimation of Solar Energy Potential under Conditions of Urban Development. Proceedings of the International Conference "Actual Issues of Mechanical Engineering" 2017 (AIME 2017) Atlantis Press 156 - 161, SN - 2352-5401 https://doi.org/10.2991/aime-17.2017.26

Zahs, V., Anders, K., Kohns, J., Stark, A., & Höfle, B. (2023). Classification of structural building damage grades from multi-temporal photogrammetric point clouds using a machine learning model trained on virtual laser scanning data. International Journal of Applied Earth Observation and Geoinformation, 122, 103406. https://doi.org/10.1016/j.jag.2023.103406

Zerafa, S., & Azzopardi, B. (2022). Evaluating urban solar resource using drones and real-data techniques. 13th Mediterranean Conference on Power Generation, Transmission, Distribution and Energy Conversion (MEDPOWER 2022), 540–544. https://doi.org/10.1049/icp.2023.0050

Zhang, J., Xu, S., Zhao, Y., Sun, J., Xu, S., & Zhang, X. (2023). Aerial orthoimage generation for UAV remote sensing: Review. Information Fusion, 89, 91–120. https://doi.org/10.1016/j.inffus.2022.08.007

Zhuang, L., Zhang, Z., & Wang, L. (2020). The automatic segmentation of residential solar panels based on satellite images: A cross learning driven U-Net method. Applied Soft Computing, 92, 106283. https://doi.org/10.1016/j.asoc.2020.106283

2024

DETECTION AND MAPPING OF SOLAR PANELS BASED ON DEEP LEARNING, INSTANCE SEGMENTATION AND VERY HIGH-RESOLUTION MULTIBAND UNMANNED AERIAL VEHICLE (UAV) PHOTOGRAMMETRIC SURVEY.

Nnadozie Uzoma Onyeukwu.

Masters Program in **Geospatial Technologies**

Supported by:
Education and Culture
ERASMUS MUNDUS