Building extraction and roof type classification of aerial images for maximal PV panel installation

Dr P Uma Maheswari Shruthi Muthukumar Gayathri Murugesan Jayapriya M

***Abstract*—Environmental issues are a major source of concern these days, as they threaten the existence of humankind. The transition to renewable energy sources such as wind and solar has begun as mankind faces the issue of meeting its energy demands in a more sustainable manner. Several studies are being conducted on optimizing the use of solar energy through the use of highly efficient solar panels. However, this is associated with a lot of concerns on its own. Traditional approaches, such as online assessment, however pose significant challenges as they are time-consuming and costly. It also demands a large amount of human labor, with competent individuals visiting the site to assess each building and determine the type of roof to determine the PV panel installation arrangement.**

In accordance, this project focuses on creating a comprehensive and detailed pipeline to address the problem of classifying and manually assessing rooftops for solar panel placement. For this purpose, understanding urban dynamics, and accurate mapping of buildings from satellite images is crucial. Thus, given a satellite image with wide coverage and spatial resolution of 7.5cm, we employ a deep-learning model called MultiRes-UNet, which is an improved version of the original UNet network for building segmentation. The experimental results show an average IoU of 95.25% and average dice coefficient value of 97.56% for building segmentation using MultiRes UNet. The rooftops of segmented buildings are then extracted by applying image processing techniques like finding contours, color filling and background subtraction.

The extracted rooftops are manually classified into three categories: flat, gable, and hip, and then sent into the second stage for roof type classification. Various deep learning models including a customized CNN model and three transfer learning approaches (ResNet50, EfficientNetB4, and VGG16) are used for classification and a comparative analysis is made between the models. Although the shallow CNN model has a slightly lower accuracy of 78%, it is still acceptable given the very limited data. With the AIRS dataset, we conclude that the transfer learning models have an average accuracy of 91.93%, indicating that the models have been fine-tuned to work with the sparse data we have. Because different DL models performed well on different portions, majority voting was employed as the assembling method. The results show that using majority voting improved classification accuracy by 5.67%.

This is finally followed by a simulation of modular layout of PV panels on rooftops based on customized inputs on the dimensions and tilt angle of the panel. The arrangement of PV panels also takes into consideration the obstructions (tree shadows, obstacles) to ensure maximal consumption of energy. This pipeline, thus, resolves the time-consuming, laborious task of manual assessment of rooftops.

***Index Terms*—Building Segmentation, AIRS, MultiRes UNet, Rooftype classification, PV panel simulation**

I. Introduction

In this chapter, we will be discussing the need for a sustainable environment and potential methods undertaken to resolve climate change issues. This is followed by understanding the importance of remote-sensing in getting to know better about urban dynamics, and terrestrial mapping and building types. An overall synopsis of the project is briefly mentioned out.

A. Problem Statement

Sustainable environment is required for the advancement of economic development, to improve energy security, access to energy, and mitigate climate change. Energy production sources such as coal, oil, and natural gas are responsible for one-third of global greenhouse gas emissions, and there is a growing need for everyone to switch to solar power for electricity generation. Estimating a roof's solar potential, on the other hand, is a time-consuming process that requires manual labor and site inspection. The current algorithms only work with LIDAR data and do not predict the number of solar PV panels based on the type of rooftops. Furthermore, mapping urban buildings for rooftop segmentation presents its own set of issues, since aerial satellite photos are typically of low resolution.

In this project, we use the AIRS dataset that provides a wide coverage of aerial imagery with 7.5 cm resolution and propose a mechanism to address the above problem: State-of-the-art MultiRes U-Net architecture is used for building detection as a first step to identify and segment buildings from aerial image. This is followed by classifying rooftops from the extracted buildings with different deep learning and transfer learning algorithms and detecting the boundaries of rooftop. Finally, a fitting algorithm is used to provide a simulation of PV panels fitted on top of roofs. This proposed solution uses a single drone/satellite image to recognize and classify rooftops for buildings dispersed throughout an area, thus automating the entire process and reducing human labor.

B. Sustainable Environment

Over the last few decades, climate change has become a global problem, affecting a variety of life forms. Sustainable development is thus the need of the hour for the survival of mankind. There is a rising awareness that harnessing renewable resources is the way to go in order to construct a sustainable environment. Potential tapping of solar power for generating electricity has gained enormous popularity

and attention in recent years, and people are increasingly gravitating toward the PV revolution.

However, traditional approaches, such as online assessment are time-consuming and expensive, and they require a significant amount of human effort, as concerned experts must visit the site to inspect the building in order to determine whether PV panels should be installed and the modular layout of PV panel installation. However, by automating the process of building roof extraction for PV panel placement, a lot of money and time can be saved. In this way, PV sales can be made more efficient by using an automatic PV system design.

C. Overall Objectives

To simulate the placement of PV panels on rooftop of buildings from aerial images based on roof type for maximum energy consumption, the following steps are to be performed:

- To segment and detect buildings in a given satellite image using MultiRes UNet model and perform background subtraction to extract the rooftops.
- To label the extracted building rooftops into different classes - flat, gable, and hip and create a dataset and train different DL models for roof type classification.
- To perform edge detection on the extracted rooftops to mark boundaries to find the necessary area for PV panel simulation.

D. Scope

The primary goal of the project is to classify roof tops and provide a simulation of modular layout of PV panels fitted on top of roofs. Even though, manual assessment is more reliable, it is often time consuming and labor intensive when it comes to covering numerous buildings in a particular area. That is why automating the entire process is necessary. We must note that so far no automated approaches have been used to detect buildings from a wide coverage satellite image and classify roof types. This is where our one-shot pipeline comes into play, which can segment and detect buildings from aerial/satellite images, extract each of the rooftops and classify them based on their type in order to provide a simulation of PV panels fitted on them. This can easily automate the entire process of assessing hundreds of roof tops.

II. Related Work

Over the years, image databases and computer vision have gotten a lot of attention. Many breakthroughs in image segmentation have been made as a result of this. Bio-medical and remote sensing are the two industries where segmentation is critical in understanding the images.

With respect to this, a multi-scale convolutional neural network that adopts an encoder-decoder U-Net architecture is utilized by X. Li, Y. Jiang, H. Peng and S. Yin (2019) for building segmentation. Here, a U-Net is constructed as the main network, and the bottom convolution layer of U-Net is replaced by a set of cascaded dilated convolution with different dilation rates and an auxiliary loss function added after the cascaded dilated convolution. The proposed method has achieved an IOU of 74.24% on the whole dataset that covers regions like Austin, Chicago, and Vienna. The proposed model has aided in network convergence, and a major advantage is that it does not involve manual features and does not involve preprocessing or post-processing steps. However, the segmentation of middle parts in buildings are misaligned and the bulges on the boundaries are lost. Yet another major issue is that the algorithm performs well only in one subset (countryside and forest) but not in another.

By scraping data from Google Maps, V. Golovko, S. Bezobrazov, A. Kroshchanka, (2017) developed a CNN-based solar PV panel recognition system. Pre-processing techniques such as image scaling and sharpening are used, followed by the training of a 6-layer CNN model. Instead of using high resolution color satellite orthoimagery, the authors employed low-quality satellite imagery (Google Maps satellite photos), which allows them to reduce the approach's requirements. Because some solar panels resemble roof tops, poor quality satellite pictures taken from Google Maps have led to erroneous classification and there is no validation on the dataset.

A mask R-CNN with three steps is proposed to extract buildings in the city of Christchurch from aerial images post-earthquake by Chen, Mengge and Jonathan Li (2019) to recognize small detached residences to understand the havoc caused by it. Feature extraction with ResNet is the first step. The RPN (Regional Proposal Network) is then utilized to locate RoI and filter out the irrelevant bounding boxes using object and background classification, as well as bounding box regression. The background and buildings are then identified by object classification. The RoIAlign method used instead of the RoI Pool gives better feature extraction. However due to a small training dataset, the model was unable to successfully demarcate building edges, resulting in low accuracy and precision when compared to other SOTA models.

A combination of image processing techniques, including Adaptive Edge Detection and contours are used by Kumar, Akash & Sreedevi, Indu. (2018) to segment out rooftop boundaries and obstacles present inside them along with polygon shape approximation. It provides a comparative analysis of the solar potential of buildings. Several types of the rooftop are considered to learn the intra-class variations. Because Google Maps India's satellite resolution is so low, the edges aren't fully identified, and there are outliers plotting solar panels outside of the building's rooftop area.

The problem of semantic segmentation of buildings from remote sensor imagery is addressed by a novel framework called the ICTNet proposed by B. Chatterjee and C. Poullis (2019). ICTNet: a novel network with the underlying architecture of a fully convolutional network, infused with feature recalibrated Dense blocks at each layer is combined with dense blocks, and Squeeze-and-Excitation (SE) blocks. Dense blocks connect every layer to every other layer in a feed-forward fashion. Along with good gradient propagation they also encourage feature reuse and reduce the number of parameters substantially as there is no need to relearn the redundant feature maps which allows the processing of large patch sizes.

Reconstruction is done by extruding the extracted boundaries of the buildings and comparative analysis is made between the two. With no 3D information on the buildings, the authors have used the building boundaries as a proxy for the reconstruction process and have got better overall IoU compared to other methods. The main limitation here is that there is no loss function for the reconstruction accuracy. Furthermore, due to the fact that ground truth photos used for training contain mistakes and are manually generated, there is a large variance in per-building IoU. In addition, the reconstruction accuracy is consistently lower than classification accuracy by an average of 4.% ± 1.65%.

Satellite images from Google Earth for the city of Heilbron are collected and manually labeled following which object proportion distribution is used in image-level by Peiran Li, Haoran Zhang, Zhiling Guo, Suxing Lyu (2021) and object occurrence possibility at pixel level is statistically analyzed. SOTA PV segmentation model (DeepSolar) is used to extract visual features. Local Binary Pattern (LBP) is used for texture feature extraction & color histograms for color feature extraction. The authors have addressed the issue of class imbalance of PV and non-PV panels on rooftops by hard and soft sampling. The major drawback is that lighting conditions caused distinct color clustering groups in PV/Non-PV color clustering, resulting in misclassification along with IOU being less than the acceptable range (0.5) for 1.2m resolution images.

Rooftop classification is the important step in identifying the type of PV panels that can be fitted. M. Buyukdemircioglu, R. Can, S. Kocaman (2021) have undertaken research to generate a roof type dataset from very high-resolution (10 cm) orthophotos of Cesme, Turkey, and to classify the roof types using a shallow CNN architecture. UltraCam Falcon large-format digital camera is used to capture orthophotos with 10cm spatial resolution and roofs are manually classified into 6 different labels. The prediction is investigated by comparing with three different pre-trained CNN models, i.e. VGG-16, EfficientNetB4, and ResNet-50.

Simple CNN models are hence easier to implement and require nominal hardware specifications. The shallow CNN model has achieved 80% accuracy. As the roof images were clipped automatically from orthophotos, there are few buildings with overlap. Half-hip roofs are not classified properly and the F1 score obtained for them is very low. The authors haven't experimented with alternate hyperparameter tweaking for the shallow CNN architecture, which is a serious flaw.

SVM based machine learning approach is used to classify building roofs in relation to their solar potential. The SVM classifier used by Nahid Mohajeri, Dan Assouline, Berenice Guiboud (2018) on an average produces 66% accuracy and is able to classify rooftop types into 6 major classes. The authors have calculated the ratio of useful roof area for each type of roof shape to that of the corresponding building footprint area and the results are close to a value of 1. This indicates that better the segmentation of building, maximum is the solar potential of each of the rooftop areas.

Although the above models address the issue of building segmentation to a great extent, there are certain major flaws. In situations where shadows, vegetation and parking lots and other obstacles enclose buildings, the above frameworks showed poor results in building detection. In addition, there was no clear distinction to demarcate adjacent buildings and boundaries of buildings were not properly detected. Furthermore, when a satellite or aerial image spanning a large area with several buildings was provided, only manual cropping was done to extract roofs for classification, and there were no automated mechanisms in the approaches outlined above.

Some of the studies mentioned calculate the discrete number of PV modules that should be placed on the available roof space, but none of them computed the actual modular layout that would fit on each roof segment. Additionally, shadows and other obstacles were often neglected while considering the placement of PV panels which resulted in bad prediction of energy consumption.

To address the above problems, we propose a pipeline for building segmentation and roof type classification with a simulation of PV panels fitted on it. In this work, we adapt MultiRes UNet, with additional convolution layers to extract spatial features at various scales. The skip connections in UNet are replaced with a chain of convolution operations to reduce the semantic gap between encoder and decoder features and hence building segmentation results can be improved drastically. This is then followed by a technique called background subtraction to extract the rooftops automatically which resolves the problem of manual cropping and selection of roofs. Roof type classification is performed with deep learning models, following which we use edge detection algorithms to simulate a modular layout of PV panels fitted on top of roofs which considers several attributes including obstructions to maximize the solar energy consumption.

III. System Design

A. System Architecture

The proposed model's overall system architecture is depicted in fig. 1.

The AIRS dataset contains satellite images of Christchurch in New Zealand with 7.5cm resolution. The first module mainly emphasizes building detection and rooftop extraction from satellite images. The first stage of building segmentation includes two pre-processing steps: clipping satellite images and the ground truth masks, followed by scaling and normalization. The normalized images are then trained using the MultiRes UNet architecture for segmenting the buildings. Subsequently, image processing techniques like finding contours, drawing bounding boxes and background subtraction are performed to extract the rooftops of buildings.

The extracted rooftops from the previous module are sent to the second stage of the pipeline. The rooftops are manually labeled into 3 different classes: namely flat, gable, and hip which are fed to four deep learning models for classification and a comparative analysis is made. An ensemble method (majority voting) is used to further improve the accuracy of classification. Following that, white hazing and auto canny

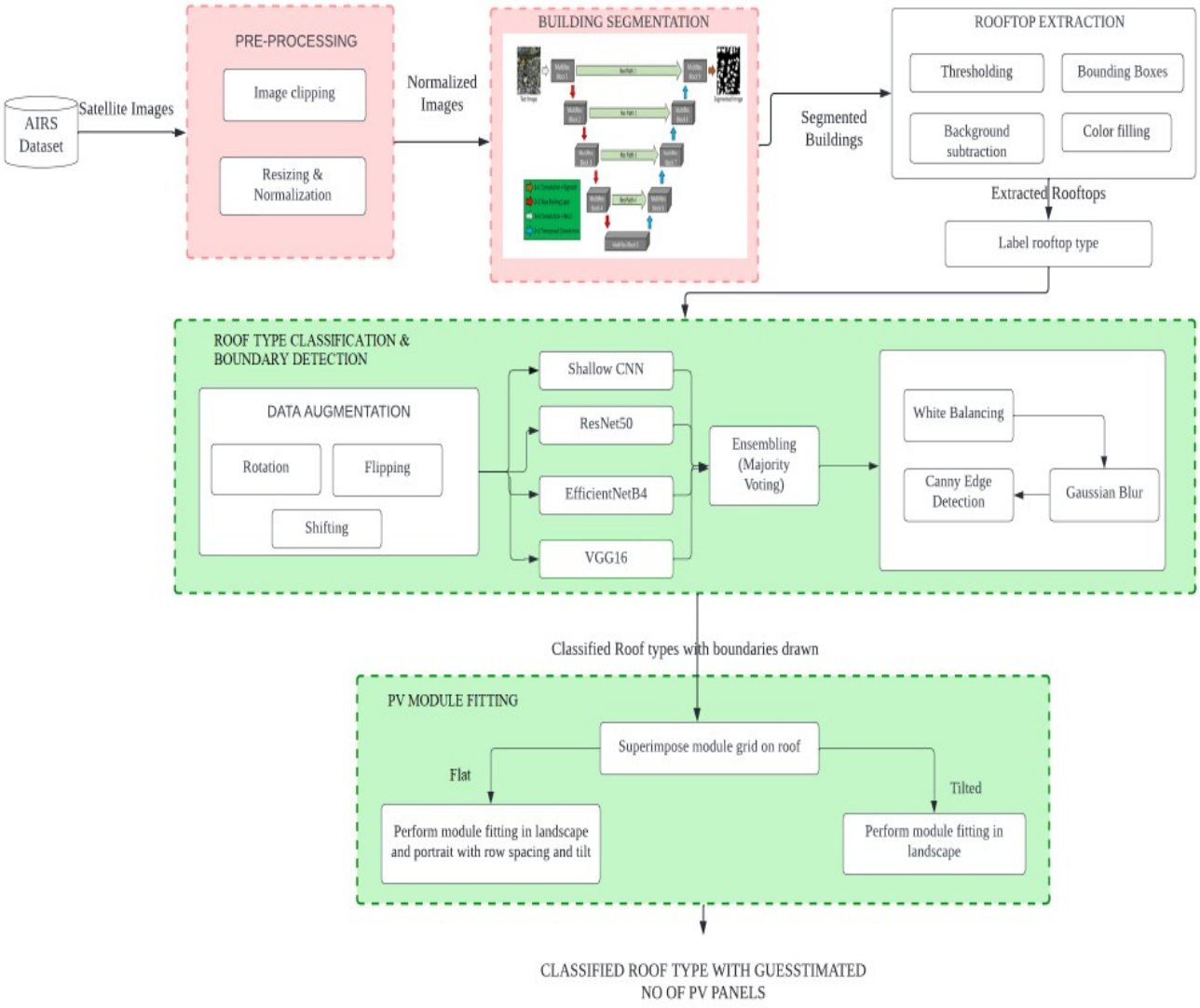


Fig. 1. Overall system architecture of building extraction and roof type classification for maximal PV panel installation

edge detection is performed to identify and mark boundaries on rooftops for fitting PV module grids.

The final module resorts to providing a simulation of modular layout of PV panels fitted on top of roofs. This is accomplished by applying a panel fitting algorithm which considers the dimensions (length, width, tilt angle) of the PV panel, type of roof and obstructions on roof top to find the placement of panels.

B. Detailed Module Design

This chapter gives a description on the different modules involved in this project and an overview on what each module intends to do along with module wise block-diagrams and intermediate deliverables of each module.

I. BUILDING DETECTION

1) ***Pre-Processing****:* Before building segmentation, a set of pre-processing steps are to be followed. Aerial satellite images from AIRS dataset are of very large dimensions (10000 * 10000). Owing to the hardware constraints for training the model, as a first step in pre-processing we clip the images into smaller images called 'patches' or 'tiles' to (1536 * 1536) dimensions. By applying this approach, we generate 1548 images for the training set and 144 images each in the testing and validation set. This is followed by resizing the images to 256 * 256 dimensions and further applying MinMax scaler as the normalization technique. Small patches of normalized images are sent to the next layer for model training.

2) ***Building Segmentation****:* Once preprocessing is done, the images are fed into the MultiRes UNet model. The architecture of the model is discussed in the Implementation section below. The model is trained for 100 epochs and Adam optimizer is used for stochastic gradient descent. After tweaking the hyperparameters, we found that a batch size of 8 with learning rate = 0.0001 works well for our model. The

loss function used here is binary cross entropy as we have two classes here - building and the background. ReLu is used as the activation function in the top layers and sigmoid is used as the activation function at the last layer. The segmented masks of buildings are compared with original ground truth masks and performance analysis is carried out.

3) ***Rooftop Extraction****:* The segmented masks of buildings from previous layers undergo thresholding here. Thresholding is done to detect the edges more accurately. After this, contours around the buildings are identified and a bounding box is drawn to the buildings. Following this, background subtraction takes place to extract rooftops from widely dispersed satellite images and they are saved and stored in a new database.

II. ROOFTYPE CLASSIFICATION AND BOUNDARY DETECTION

4) ***Data Augmentation****:* The extracted rooftops from previous steps are manually labeled into four classes - Flat, Gable, Complex, and Hip. Data augmentation is done here to increase the size of the dataset for the roof type classification to increase the accuracy as well as to prevent over-fitting. The techniques used for data augmentation include rotation, shifting, flipping.

5) ***Classification of roof type images with ensembling approach****:* Four deep learning architectures are used here for roof type classification. Shallow CNN is used as a baseline model. Three pre-trained models ResNet50, EfficientNetB4 and VGG16 are also used and a comparative analysis is made. Further, majority voting is used as an ensembling approach to predict the roof type of unseen images.

6) ***Boundary Detection****:* The extracted roof tops are now further enhanced by applying white balancing to remove haze. This is followed by applying Gaussian blur and using Auto Canny Edge Detection for finding the rooftop boundaries.

III. PV MODULE FITTING

The boundary detected rooftops are superimposed with a rectangular PV module shaped grid. Based on the type of roof, module fitting happens. PV module dimensions (length, width, tilt angle) along with the type of roof is taken into consideration and a simulation of PV modules fitted on top of roofs is provided as the final end output.

IV. Methodology

A. Dataset

The dataset used in this study is AIRS (Aerial Imagery for Roof Segmentation) dataset, which covers the full area of Christchurch in New Zealand and provides a wide coverage of aerial imagery with a spatial resolution of 7.5 cm and spatial dimensions of 10000 * 10000. The training set has 857 images and the validation and testing set each contains 90 images. The dataset has aerial satellite images along with the corresponding ground truth mask images. The roof type classification dataset has 1115 images divided into three subcategories: Flat, Gable, and Hip, with 1020 images in the training set, 50 images in the validation set, and 45 images in the testing set.

B. Experimental Setup

The whole process of training and testing the MultiRes UNet network for building detection and different transfer learning models for rooftop classification is executed under TensorFlow backend and Keras framework in Colab Pro with a memory of 16 GB, T4 GPU.

C. Building Detection

Satellite and aerial images are typically too large to be segmented directly and hence are clipped into smaller patches or tiles. The 10000 × 10000 dimensions satellite images are clipped to 1536 × 1536 pixels using a sliding window technique and consequently we obtain 36 patches for a single satellite image. This is done for training, validation, testing images and masks resulting in a dataset with 1548 images in the training set, 72 images in validation and 144 in the testing set. Images of size 1536 × 1536 are still huge to process and train on complex neural nets as we do not have the required hardware specifications. As a result, the images are resized to 256 × 256 dimensions with INTER_CUBIC interpolation method and normalization was performed.

This is then followed by adapting the MultiRes UNet architecture for building segmentation. The architecture of the MultiRes UNet network consists of 2 important blocks: the MultiRes Block (fig. 2) and the Res Path(fig. 3). Four MultiRes blocks are each used in the encoder and decoder stage. The number of filters in each of the MultiRes blocks is based on the formula: $W = \alpha \times U$ where α is the scalar coefficient whose value is set to 1.67 and U refers to the no of filters. The parameter W preserves an analogous connection between the suggested MultiRes-UNet network and the main UNet network. After every pooling or transposing of layers, the value of W became double, similar to the original UNet network. The values of U = [32, 64, 128, 256, 512] are set as follows. We also allocated filters of [*W/6*], [*W/3*], and [*W/2*] to the three succeeding convolutions respectively. The filters used in the Res Path are as follows: 64, 128, 256, and 512. Batch Normalization is performed in each of the blocks to avoid overfitting. ReLU is used as the activation function in all the convolution layers and the last layer employs the Sigmoid activation function as we have only binary classes here (0 indicating background, 1 indicating the building). The hyperparameters used for the model are as follows:

TABLE I
HYPERPARAMETER VALUES USED IN TRAINING OUR MULTIRES UNET MODEL

Hyperparameters	Values
Learning rate	0.0001
Epochs	100
Batch size	8
Optimizer	Adam
Loss	Binary cross entropy

After training the images with the MultiRes UNet model, some pixel values lie in the range between (0, 1) and these pixels do not clearly indicate whether they belong to foreground

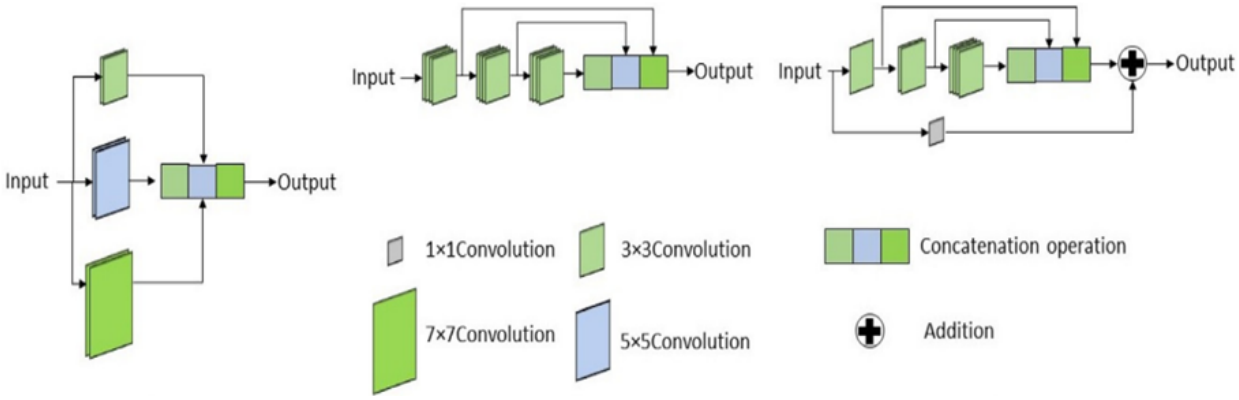


Fig. 2. Architecture of MultiRes block

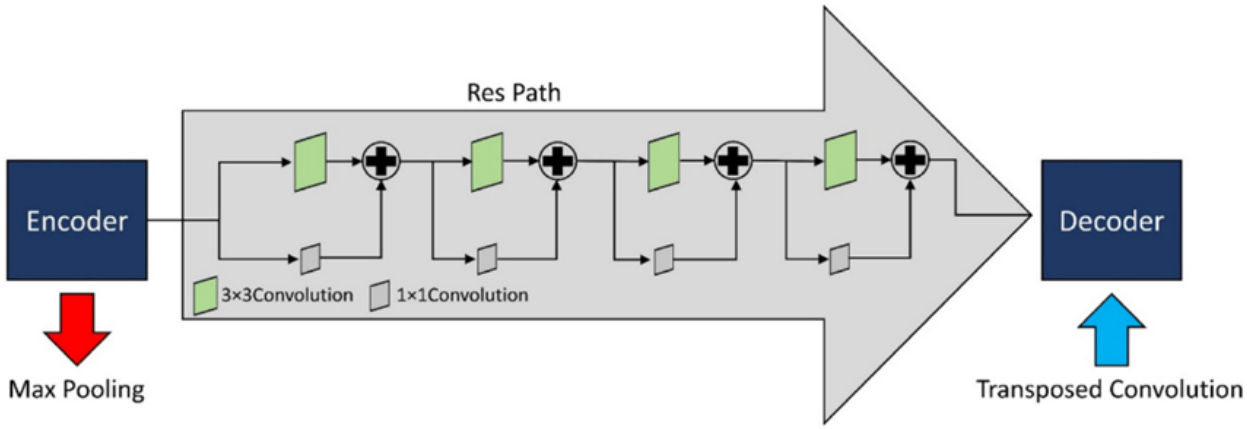


Fig. 3. Architecture of ResPath

or background. In order to resolve this issue and delineate the boundaries of buildings properly, we use simple thresholding. The threshold value is set to 0.5, and all values greater than 0.5 are assigned to the foreground, while values less than 0.5 are assigned to the background. Fig. 4 provides a comparison between the ground truth images with our prediction masks before and after applying threshold. After thresholding, our buildings are more effectively demarcated.

Following building segmentation, rooftop extraction from the mask image is performed using background subtraction. The contours of each of the buildings are identified along with the the number of buildings in a given satellite image. in the image. The detected contour is enclosed within a rectangle using bounding boxes which are used to highlight the regions of interest. The top-left (x, y) coordinates along with the width and height of each box is stored in a vector. Finally, background subtraction takes place where the original satellite image is subtracted from the segmented masks with bounding box. The extracted building rooftops are stored in a database which is later used in the second stage for classification.

D. Rooftype Classification and Boundary Detection

The second phase in the pipeline is classifying the type of roofs and drawing boundaries on the rooftop. After conducting rooftop extraction in the preceding step, a dataset including 1115 rooftop photos is populated into three different categories: Flat, Gable, and Hip. After reading numerous publications, we discovered that the authors used manual roof type labeling in every case. Hence, we resorted to the same technique of manual labeling of roof tops as there was no supervised dataset for roof type classification available. The task of manual labeling does not require any prior knowledge of the field and roof types in most cases can be easily identified by laymen. Thus, we referred to the sample images provided by M. Buyukdemircioglu et al. for labeling our dataset.

The dataset was cleaned up by removing roof types with poor resolution. A balance between the several roof type classifications was sought as much as feasible when compiling the dataset. Despite this, the number of rooftops of type gable is considerably higher than the other two classes due to the varying numbers of instances for each roof type in the research area. The distribution of different roof types are presented in table II.

Image number: 0
IOU Score: 0.959441065788269
Dice Coefficent: 0.9793001413345337
MCC: 0.9720540046691895

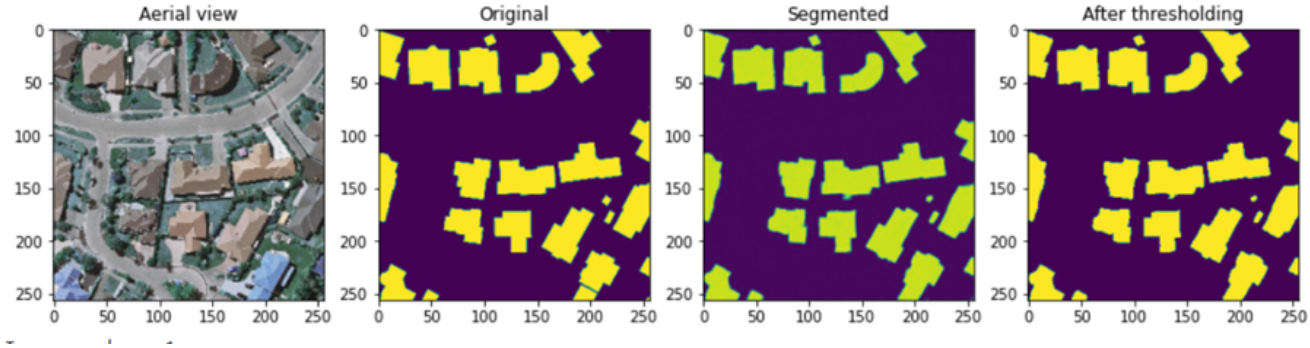


Image number: 1
IOU Score: 0.9527599215507507
Dice Coefficent: 0.975807785987854
MCC: 0.9687466621398926

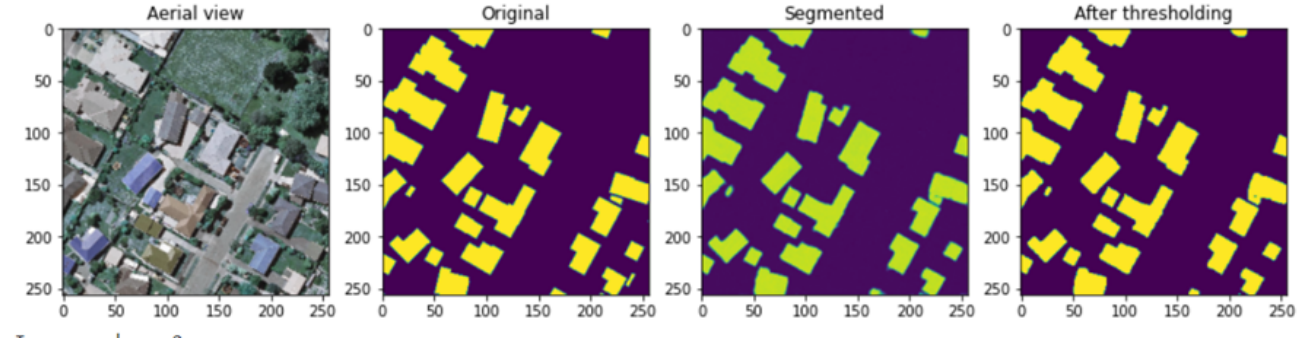


Image number: 2
IOU Score: 0.951558530330658
Dice Coefficent: 0.9751772880554199
MCC: 0.9668542742729187

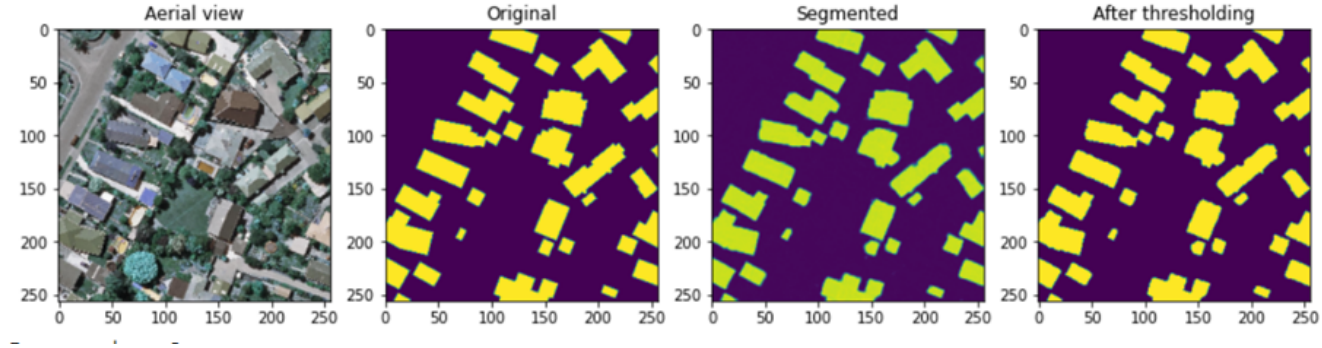


Fig. 4. Comparison of ground truth images with predictions from MultiRes before and after applying threshold

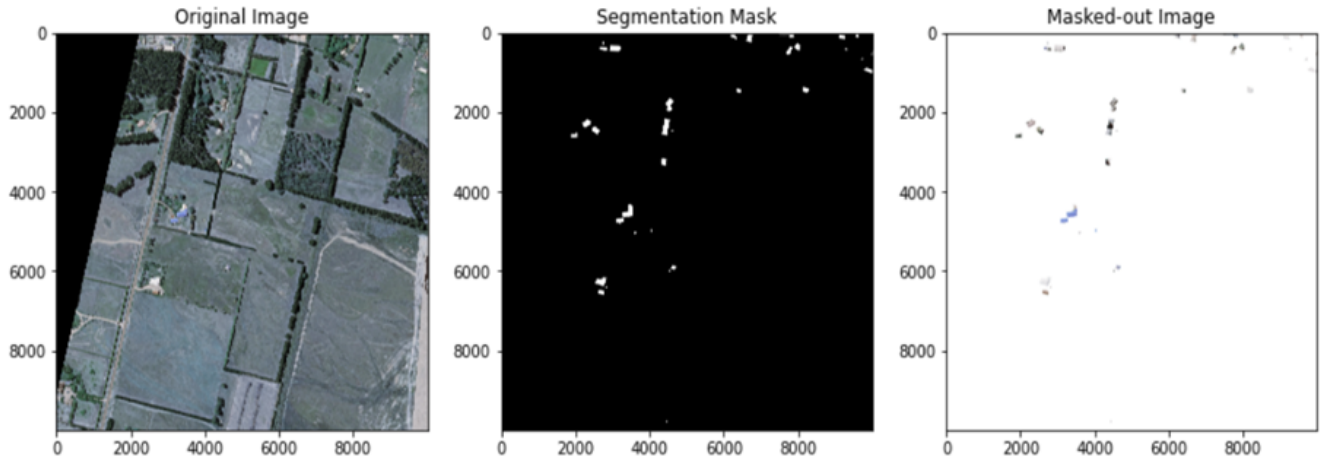


Fig. 5. Original satellite image (left), predicted masked image from MultiRes UNet model (center), extracted rooftops (right)

TABLE II
DISTRIBUTION OF ROOF TYPE DATASET

Roof Type	Training (91%)	Val (5%)	Testing (4%)	Total
Flat	303	14	18	335
Gable	410	23	18	451
Hip	307	13	9	329

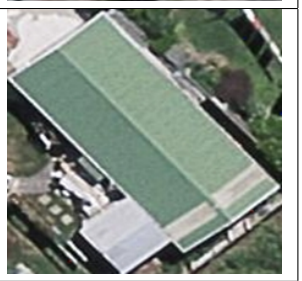
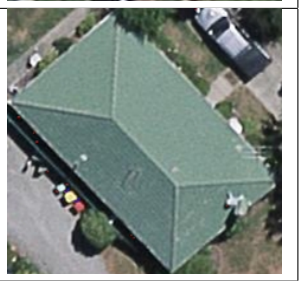
Table III gives a sample representation of the different roof tops belonging to each of the 3 classes. In order to improve the performance of model training, rooftops of different layouts and orientations were carefully picked, analyzed and labeled.

For training neural network models, the size of the dataset needs to be large. However, owing to time constraints we were able to manually label only 1115 images. As a result, data augmentation approaches were employed to increase the dataset size for better model training. Rotation, shifting and flipping are the three main techniques used here.

Following data augmentation, the model is trained with different DL models to classified different types of roofs. CNN model along with additional transfer learning models like VGG16, ResNet50, and EfficientNetB4 are used for training. Following this, majority voting is applied as the ensembling approach to combine learnings from different models for a better result.

Boundary detection is the next step after rooftop classification. White balancing is the initial step in boundary detection of roof tops. White balancing is used to remove any colored haze from an image so that it seems to be under white light. Due to the highly variable lighting conditions under which satellite imagery is gathered, this is an extremely impactful step on the resulting quality of edge detection. This is followed by Gaussian blurring which is used to reduce various types of noise and accentuates edge contrast within the image. Finally, a multi-stage canny edge detection is used. This algorithm looks for intensity gradients, reduces noise, and suppresses hysteresis.

TABLE III
SAMPLES OF IMAGES OF EACH CLASS FROM CHRISTCHURCH, NEW ZEALAND

Flat		
Gable		
Hip		

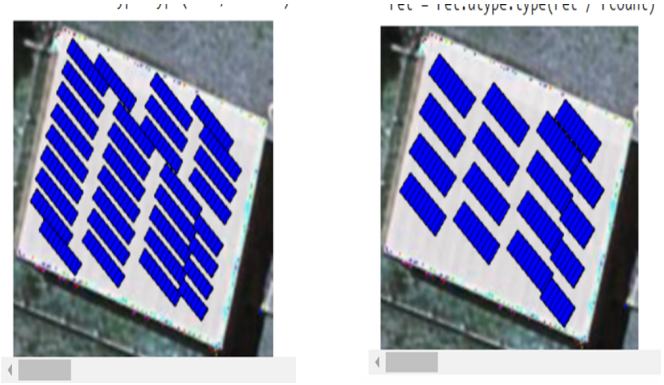


Fig. 6. PV panels with dimensions length - 20mm; tilt angle - 40°; width - 10 mm (left) and length - 20mm; tilt angle - 40°; width -20mm (right).

Fig. 7. Placement of PV panels when there are obstacles.

E. PV Module Fitting

The first stage in PV module installation is to identify the contours and detect obstacles on the rooftop using the edge sharpened image acquired by auto canny edge technique in the previous step.Since most of the obstacles on top of roof top occupy a smaller area, all contours whose area is less than 5, are marked as obstacles or patches having small areas where panels cannot be fitted and these points are not considered for panel placement.

This is followed by obtaining the dimensions of solar panels including length, width and tilt angle from the user. The tilt angle for gable and hip roofs is typically kept between 0° and 10°. For rooftops of flat type, the preferred tilt angle lies around 20°- 45°.

PV panels that have zero tilt angle can be mounted directly on rooftops and moved. However, for panels that have non-zero tilt angle, a projection of panels needs to be made. Let us assume the initial coordinates of PV panels to be (x, y). The rotated coordinates (x', y') are determined depending on the tilt angle using (1).

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x' \\ y' \end{bmatrix} \tag{1}$$

The rotated points are now checked to see if they lie within the boundary of rooftops. For each of the rotated solar PV panels, pointpolygontest is undertaken to ensure that all points of solar PV panels lie inside the rooftop. If the rotated panel coordinates lie outside, then the panels are not placed. Otherwise, the PV panels are placed inside the roof tops.

This is then followed by using Bresenham's algorithm to identify points in straight line segments so that PV panels can be placed one after the other along the line segment. This procedure can be used to remove smaller patches of area and other obstructions. The number of points between two end points are determined by the above step and a threshold value of 10 is set, over which indicates the presence of a large number of intermediary points, which in most circumstances could constitute impediments.

A simulation of PV panel arrangement on flat roof tops is shown in fig. 6. It can be observed from fig. 6 that the number of solar panels and their position/alignment change as the width of PV panels change.

The effectiveness of our algorithm in placing PV panels on rooftops when there are obstacles can be noticed from fig. 7. In this case, the tree occupies a large portion of the rooftop and becomes the obstacle. The PV panels are oriented in such a way to ensure there is maximum energy consumption and does not occupy the place where there are obstacles.

V. Experimental Results

*A. **Performance results on building segmentation***

The first module which involves building segmentation uses IoU/ Jaccard Coefficient and Dice coefficient as the primary metrics. Along with this, pixel accuracy and MCC was also used to evaluate the training and validation results.

The graphs in fig. 8 allows us to easily interpret the training of the MultiRes UNet model. The blue line indicates the training details while the red line shows the validation details. The MCC value is close to 98% at the end of 100 epochs which indicates that buildings have been identified correctly. IoU, which penalizes heavily than dice coefficient, has achieved a remarkable value of 96.27% at the end of 100 epochs. It can be seen that the Dice coefficient starts with 58% at the beginning of training and gradually attains a value of 98.5% with no major fluctuations. As can be seen from fig. 9, the training and validation curve coincides with each other and the difference in loss is very meager which shows the model does not overfit.

Furthermore, from table IV, we infer that the model has performed remarkably well as the training loss hovers around 0.173 and validation loss is close to 0.2. This is a reasonably good value which indicates that our model performs well in segmenting the buildings accurately. Average IoU after training 1548 images for 100 epochs comes to 93.53% and the validation IoU of 95.25% demonstrates that the model performs well even after the images are resized and normalized.

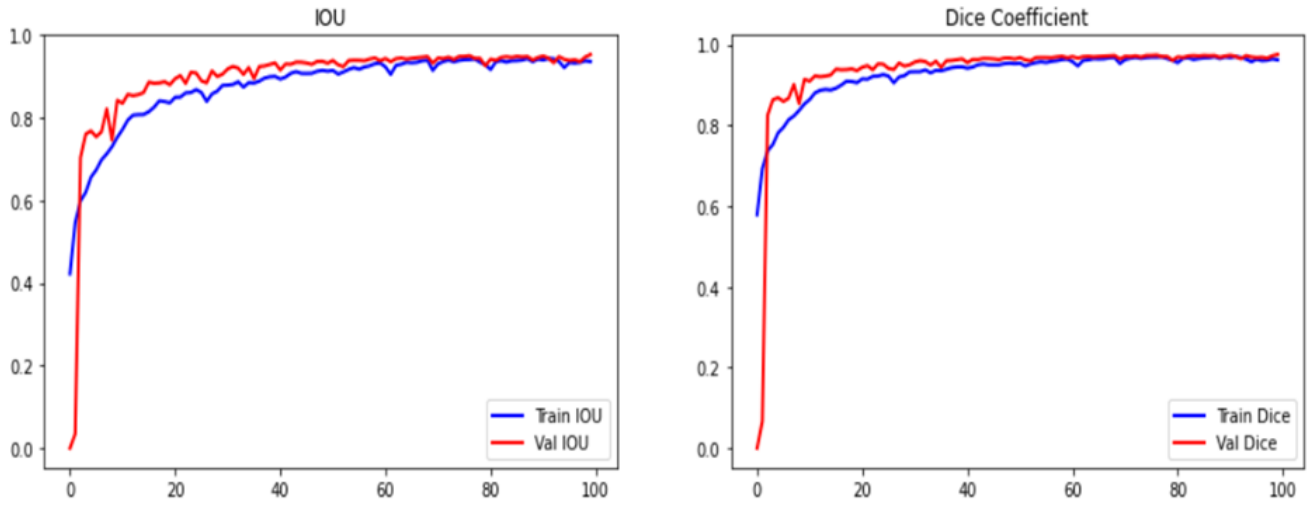


Fig. 8. IoU and Dice coefficient graphs on training and validation set

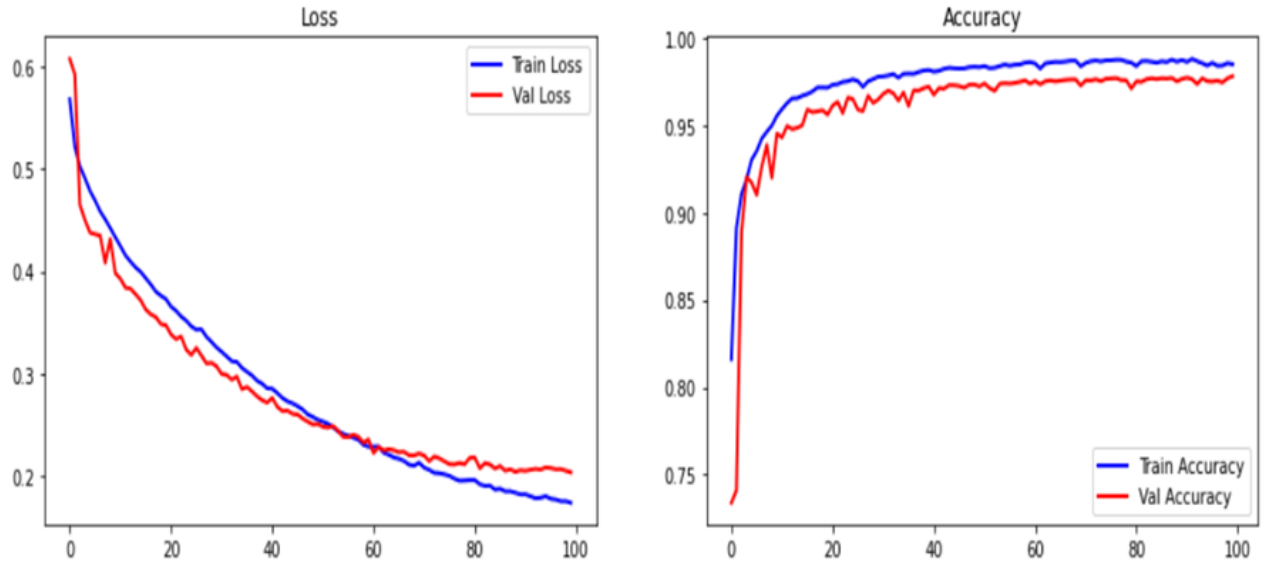


Fig. 9. Loss and accuracy graphs on training and validation set

TABLE IV
PERFORMANCE RESULTS OF MULTIRES MODEL

Performance Metrics	Training Set	Validation Set
IOU (%)	93.53	95.25
Dice Coefficient (%)	96.25	97.56
MCC (%)	95.87	96.74
Accuracy (%)	98.51	97.83
Loss	0.1734	0.2033

B. *Performance results on roof type classification*

CNN model along with additional transfer learning models like VGG16, ResNet50, and EfficientNetB4 are used for roof type classification. The performance results for each of the models on the manually labeled dataset is discussed below.

*1) **Performance of Shallow CNN Model**:* The CNN model achieved an accuracy of 78% on the training set and 74% on the validation set. From fig.10, it can be inferred that there is no generalization gap because of dropout and data augmentation. The model was likewise trained for 50 epochs at first, but because it didn't produce satisfactory results, the model was further trained for another 50 epochs. The confusion matrix in fig. 11 reveals that gable and hip classes are accurately classified to a larger extent than other classes. An equal number of flat classes, on the other hand, are misclassified as gables. The results also show that hip class has the highest recall value of 81% followed by gable (67%) and then flat (55%). The overall accuracy on the test samples

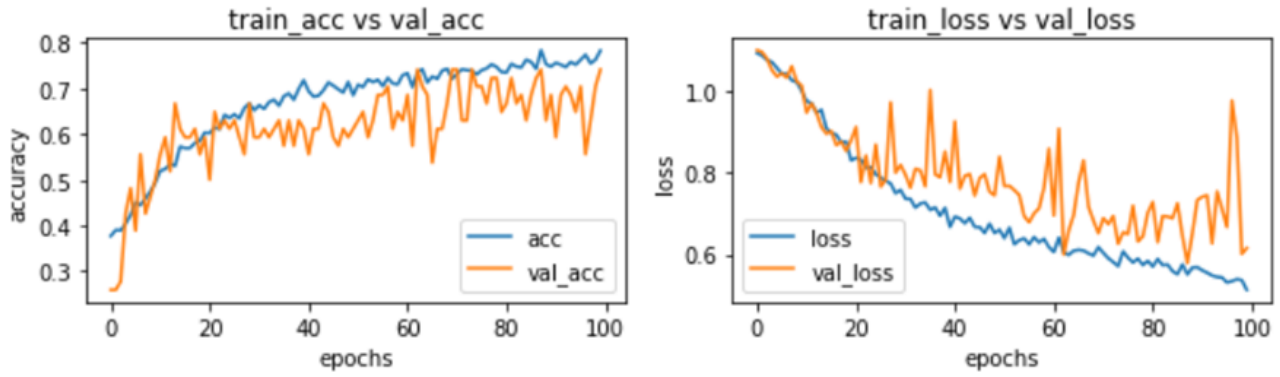
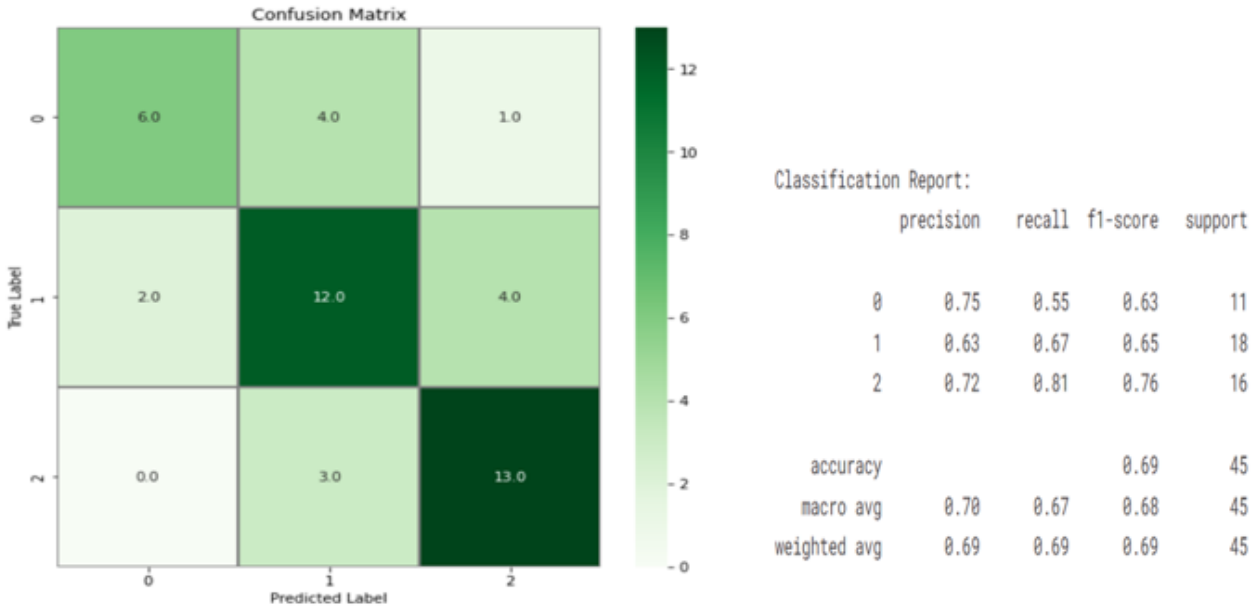


Fig. 10. Graphs showing training & validation accuracy, loss for CNN



Classification Report:

	precision	recall	f1-score	support
0	0.75	0.55	0.63	11
1	0.63	0.67	0.65	18
2	0.72	0.81	0.76	16
accuracy			0.69	45
macro avg	0.70	0.67	0.68	45
weighted avg	0.69	0.69	0.69	45

Fig. 11. Confusion matrix & classification report on test data for CNN

accounts to 69%.

*2) **Performance of fine-tuned ResNet-50 model**:* The ResNet-50 has yielded a very good accuracy of 91.59% on the training set and 92.5% on the validation set. From fig. 12, it can be witnessed that the loss is also very less at only 0.245 on the training set. All transfer learning models are only trained for 50 epochs as training for 100 epochs led to overfitting.

From the confusion matrix in fig. 13, it can be inferred that the hip class is 100% accurately classified and nearly all gable classes are correctly identified. The classification of flat classes is better compared to the previous CNN model and the misclassification of a few flat classes as gable could be due to the slight tilt angle of a few rooftop slopes, making it difficult for the model to classify it to the appropriate class. The overall accuracy on the test samples is 89% with F-1 score averaging to 91%.

*3) **Performance of fine-tuned EfficientNetB4 model**:* The graphs in fig. 14 exhibits that the accuracy of the model has steadily grown from 40% to 89% and the model loss has significantly reduced from 1.25 to 0.33 at the end of 50 epochs. When compared with the ResNet-50 model which had higher accuracy in correctly classifying hip class, EfficientNetB4 model has achieved higher F-1 score and accuracy in identifying flat roof tops (fig. 15). The average F-1 score, precision, and recall hover around 91%.

*4) **Performance of fine-tuned VGG16 model**:* VGG-16 model has achieved the highest accuracy compared to the other three models. The classification accuracy on the validation set is 94.45% while the training accuracy is around 97%. In fig. 16, it can be noticed there is a gradual decline in

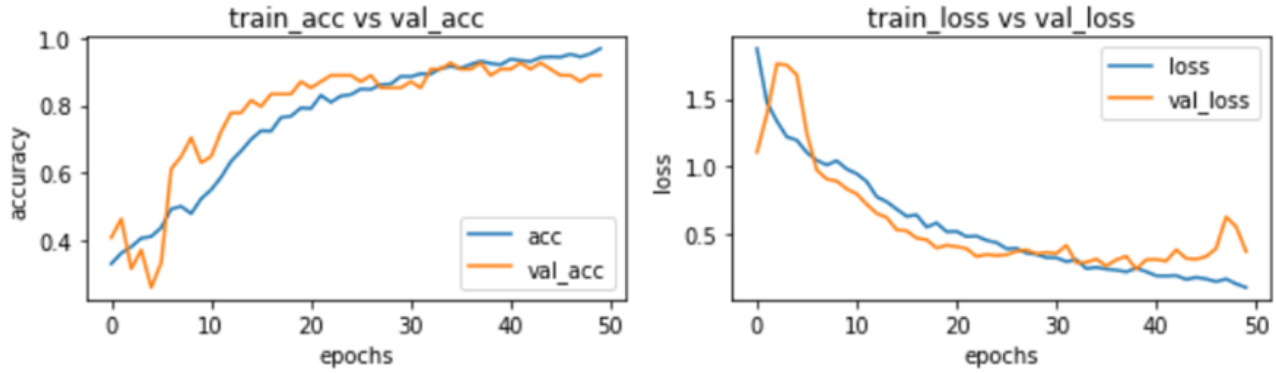


Fig. 12. Graphs showing training & validation accuracy, loss for ResNet50

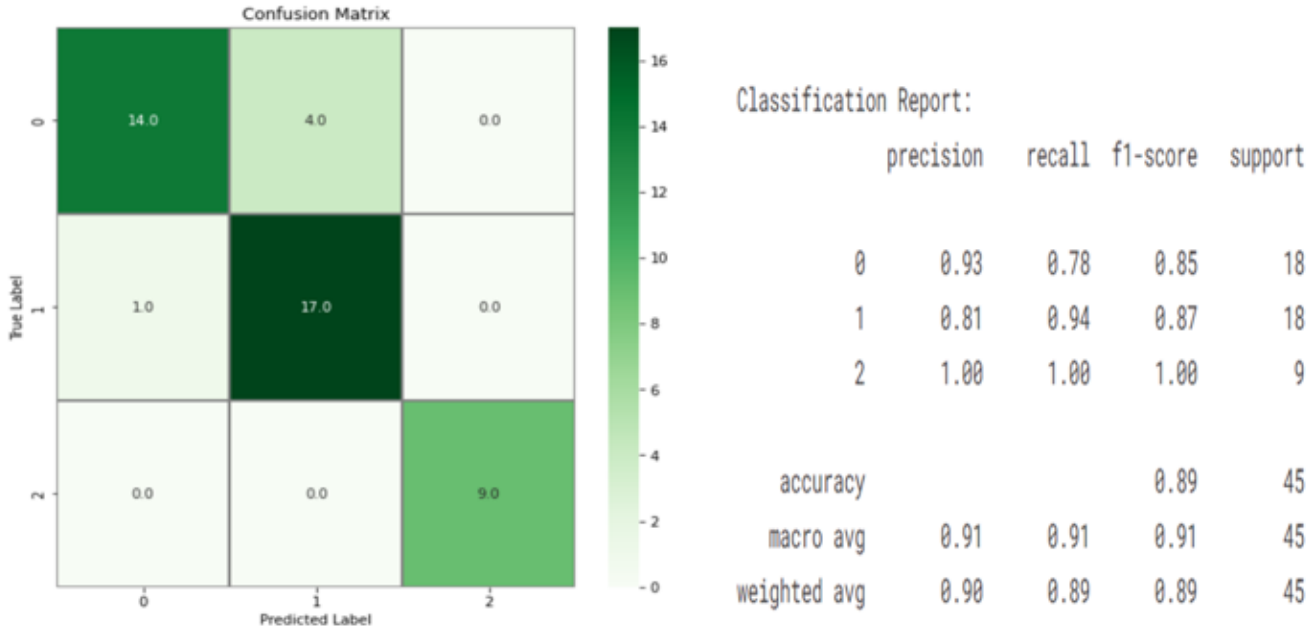


Fig. 13. Confusion matrix & classification report on test data for ResNet50

training loss over 50 epochs while there are some fluctuations with validation loss. From classification report in fig. 17, it is observed that the F-1 score for flat class is the highest with VGG16 compared to other models. The overall classification accuracy has totalled to 89%.

*5) **Majority Voting**:* It is inferred that transfer learning methods result in higher classification accuracy when compared to the shallow CNN model. However in these transfer learning models, it can be noticed that each model performs classification well on only some parts. Hence an ensembling approach is used as it can make better predictions and achieve better performance than any single contributing model. Ensembling also reduces the spread or dispersion of the predictions and model performance. The ensembling method applied here is majority voting. VGG16 was efficient in identifying hip roof types. Similarly, ResNet very well classified flat roofs from other rooftop types and EfficientNet was able to classify a large chunk of gable roof types correctly. Majority voting is thus incorporated to further boost the performance. The accuracy has increased by 5.67%

VI. Conclusion

A 3-step pipeline was proposed in this research to automate the process of PV panel placement on rooftops given a satellite image covering a wide range and different types of buildings.

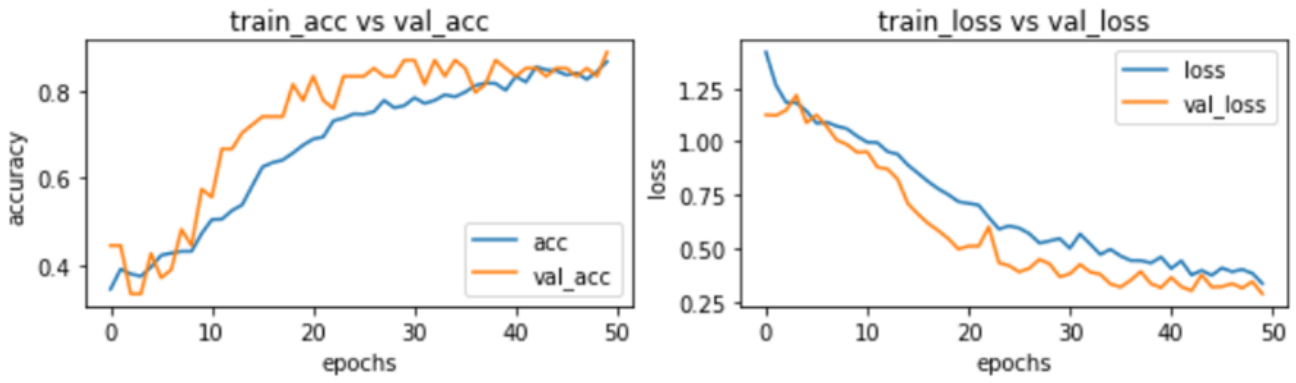


Fig. 14. Graphs showing training & validation accuracy, loss for EfficientNetB4

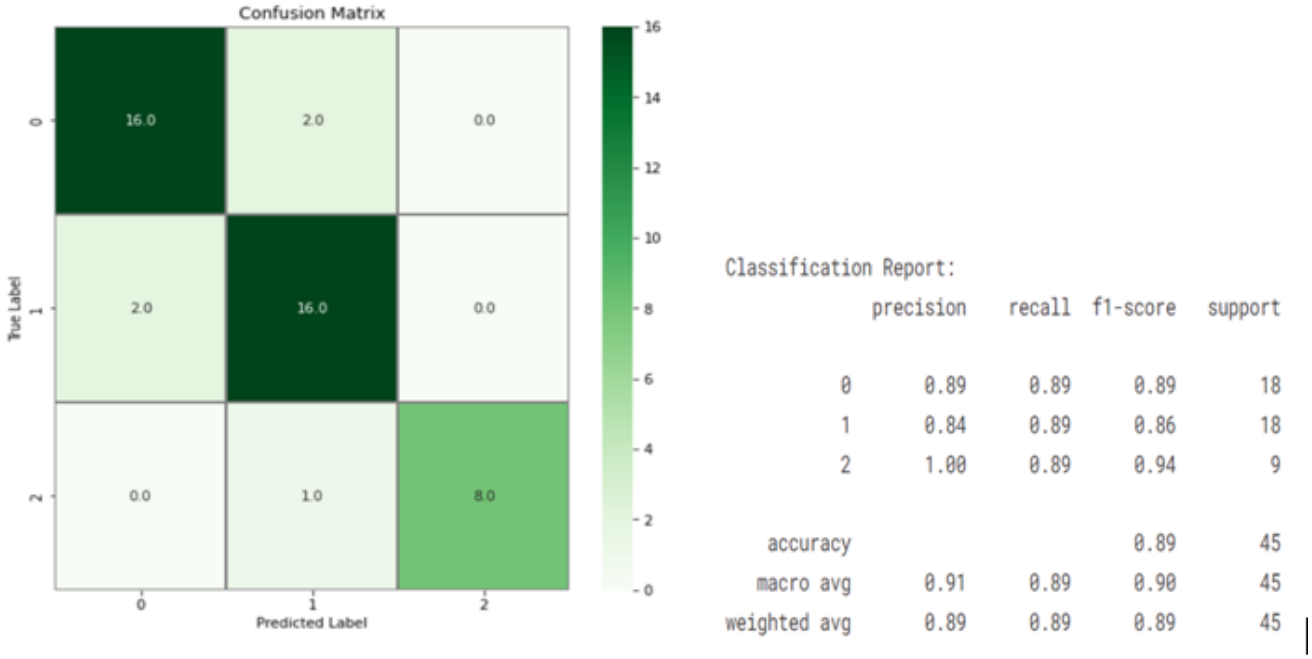


Fig. 15. Confusion matrix & classification report on test data for EfficientNetB4

The first phase involved building segmentation on AIRS dataset with the MultiRes UNet model that utilizes MultiRes block to adapt spatial features from various scales and Res path for transferring encoder features to decoder features using a set of convolution operations. Despite resizing the images to 256×256 dimensions after clipping the large satellite images, the findings show that the model works extremely well, with an average IoU of 95.25% and a Dice Coefficient of 97.56%.

Proper segmentation of buildings led to proper extraction of different roof types which were manually categorized into 3 classes: Flat, Gable and Hip. Different deep learning models were trained on the 1115 images and a comparative analysis was performed. With an F1 score of 0.69, the shallow CNN model achieved a classification accuracy of 79.25%. Fine-tuned transfer learning models, on the other hand, fared well in classifying roof types. ResNet50 and EfficientNetB4 both achieved an F1 score of 0.91 while VGG16 obtained a score of 0.89.

We discovered that different models performed well in classifying different classes of roof tops with ResNet performing well in identifying flat roofs, EfficientNetB4 in classifying gable roof tops and VGG16 with hip roofs and hence majority voting was used as an ensembling technique in combining the predictions of different learning models. Validation was performed on the supervised Potsdam dataset, with an accuracy of 80.34% after applying majority voting. The comparatively low accuracy could be due to the fact that the training was done on high-resolution satellite photos from the AIRS

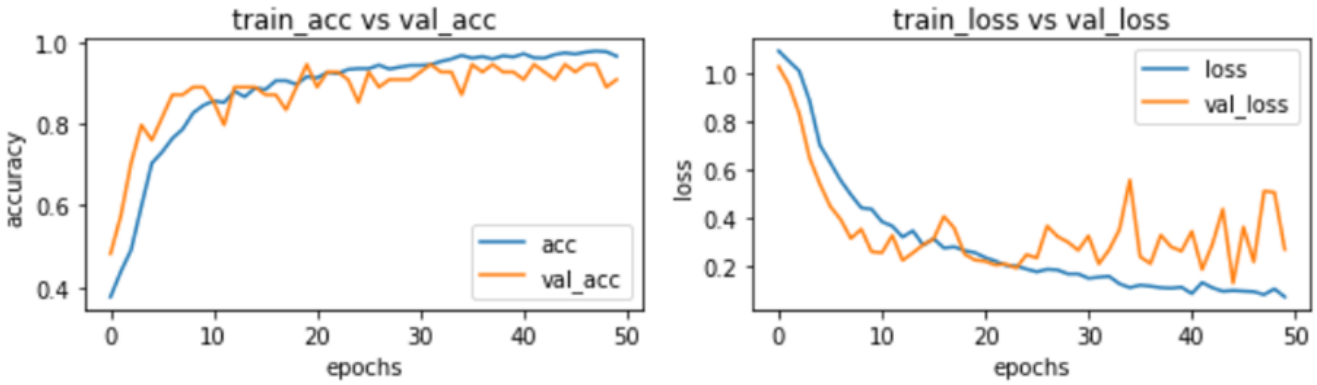
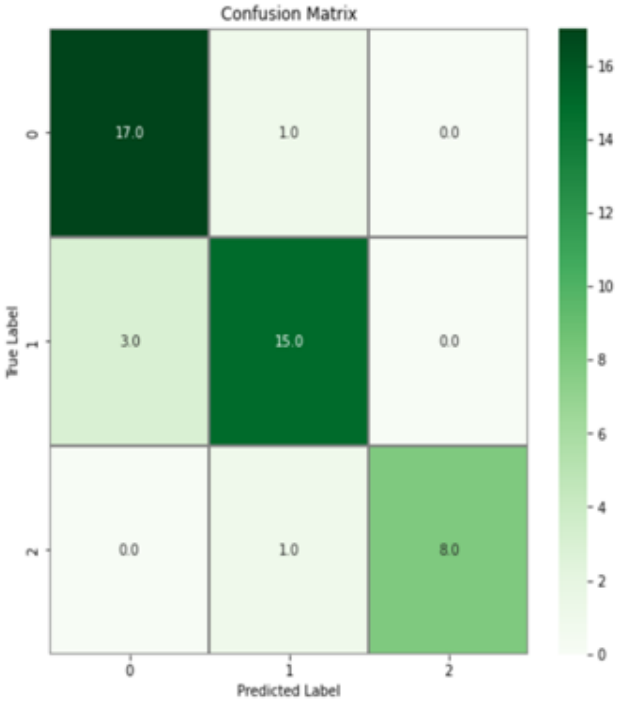


Fig. 16. Graphs showing training & validation accuracy, loss for VGG16



Classification Report:

	precision	recall	f1-score	support
0	0.85	0.94	0.89	18
1	0.88	0.83	0.86	18
2	1.00	0.89	0.94	9
accuracy			0.89	45
macro avg	0.91	0.89	0.90	45
weighted avg	0.89	0.89	0.89	45

Fig. 17. Confusion matrix & classification report on test data for VGG16

dataset, whereas the images on Potsdam were somewhat low-resolution. The classification results can further be improved by using more data.

Finally, boundaries were drawn on rooftops using auto canny edge detection mechanism. Based on the type of roof and user input dimensions (length, width, tilt angle), a simulation of PV panels fitted on top of roofs are produced as the final end output. We discovered that the algorithm was pretty good in identifying obstacles or trees (shadows) on rooftops and hence not placing PV panels on obstructions. This module provides an additional flexibility of trying out the simulation with different customized values.

To summarize, the proposed 3-step mechanism takes a single satellite image as input, performs segmentation to detect buildings from the image, extracts rooftops alone using background subtraction technique, classifies the roof type as flat, gable, or hip, and then provides a simulation of PV panel placement on roof tops.

VII. Future Work

The MultiRes model was trained with only 1548 images owing to memory and hardware constraints. Training with even more images can lead to an even better accuracy. The

classification of roof types is one of the areas of anticipated future development. We primarily seek to improve the outcomes of shallow CNN models by expanding the dataset size, either through labeling or by obtaining equivalent training data from other sources. The final module so far only considers the length, width, tilt angle and obstructions on roof tops. It is planned to further include additional parameters like solar irradiation angle, climatic conditions and other necessary factors and validate the results from the final module with a solar expert.

REFERENCES

[1] Abdollahi, A., Pradhan, B., Gite, S., and Alamri, A., (2020), "Building footprint extraction from high resolution aerial images using generative adversarial network (GAN) architecture,"IEEE Access, Article 209517-209527.

[2] Axelsson, M., Soderman, U., Berg, A. and Lithen, T., (2018), "Roof Type Classification Using Deep Convolutional Neural Networks on Low Resolution Photogrammetric Point Clouds from Aerial Imagery," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp.1293-1297.

[3] Badrinarayanan, V., Kendall, A., and Cipolla, R., (2017), " Segnet: A deep convolutional encoder–decoder architecture for image segmentation," IEEE Transactions on Pattern Analysis Machine Intelligence, 39, 2481–2495.

[4] B. Chatterjee and C. Poullis, (2019), "On Building Classification from Remote Sensor Imagery Using Deep Neural Networks and the Relation Between Classification and Reconstruction Accuracy Using Border Localization as Proxy," 16th Conference on Computer and Robot Vision (CRV), pp. 41-48.

[5] Buyukdemircioglu, Mehmet and Can, Recep and Kocaman, Sultan. (2021), "Deep learning based roof type classification using VHR aerial imagery," The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. XLIII-B3-2021. 55-60. 10.5194/isprs-archives-XLIII-B3-2021-55-2021.

[6] Chen, Mengge and Jonathan Li., (2019), "Deep convolutional neural network application on rooftop detection for aerial image," ArXiv abs/1910.13509.

[7] Edun, Ayobami and Harley, Joel and Deline, Chris and Perry, Kirsten, (2021), "Unsupervised azimuth estimation of solar arrays in low-resolution satellite imagery through semantic segmentation and Hough transform," Applied Energy. 298. 10.1016/j.apenergy.2021.117273.

[8] Jordan M Malof, Rui Hou, Leslie M Collins, Kyle Bradbury, and Richard Newell, (2015), "Automatic solar photovoltaic panel detection in satellite imagery," International Conference on Renewable Energy Research and Applications (ICRERA), IEEE, 1428–1431.

[9] Kumar, Akash and Sreedevi, Indu, (2018), "Solar Potential Analysis of Rooftops Using Satellite Imagery," ArXiv abs/1812.11606.

[10] Nahid Mohajeri, Dan Assouline, Berenice Guiboud, Andreas Bill, Agust Gudmundsson, and Jean-Louis Scartezzini, (2018), "A city-scale roof shape classification using machine learning for solar energy applications," Renewable Energy Volume 121. Pages 81-93. ISSN 0960-1481.

[11] Peiran Li, Haoran Zhang, Zhiling Guo, Suxing Lyu, Jinyu Chen, Wenjing Li, et al. (2021), "Understanding rooftop PV panel semantic segmentation of satellite and aerial images for better using machine learning," Advances in Applied Energy, Elsevier. Volume 4, 100057, ISSN 2666-7924.

[12] Qi, Chen and Wang, Lei and Wu, Yifan and Wu, Guangming and Guo, Zhiling and Waslander, Steven, (2018), "Aerial Imagery for Roof Segmentation: A Large-Scale Dataset towards Automatic Mapping of Buildings," ISPRS Journal of Photogrammetry and Remote Sensing, Elsevier. Volume 147, pp. 42-55.

[13] Q. Li, Y. Feng, Y. Leng and D. Chen, (2020), "SolarFinder: Automatic Detection of Solar Photovoltaic Arrays,", 19th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN), pp. 193-204.

[14] V. Golovko, S. Bezobrazov, A. Kroshchanka, A. Sachenko, M. Komar and A. Karachka, (2017), "Convolutional neural network based solar photovoltaic panel detection in satellite photos," 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), pp. 14-19.

[15] X. Li, Y. Jiang, H. Peng and S. Yin, (2019), "An aerial image segmentation approach based on enhanced multi-scale convolutional neural network," IEEE International Conference on Industrial Cyber Physical Systems (ICPS), pp. 47-52.