ARTICLE TEMPLATE

Exploring RAU-Net for Semantic Segmentation of Philippines Satellite Images in Identification of Building Density

Joseph Jessie S. Oñate[ab] and Marianne Ang-Tolentino[b]

[a]Camarines Sur Polytechnic Colleges, Nabua, Camarines Sur Philippines; [b]Ateneo de Naga University, Naga City, Philippines

ABSTRACT
This study explores RAU-Net convolutional network architecture on satellite images by classifying geospatial objects on satellite images such as roofs to help in determining the building density of a specific location. This study developed a satellite image dataset in the Philippines and trained the dataset in an RAU-Net Convolutional Neural Network using Tensorflow Keras API. This study proves that U-Networks with a few datasets and provides acceptable performance scores. Furthermore, a new way to calculate Building Density in terms of Building Coverage Ratio(BCR) using satellite images was also presented in this study. The model and the analysis were integrated into an application that determines the BCR of a specific location through a satellite map.

1. Introduction

In recent decades, satellite imaging and remote sensing have been widely employed to identify land coverage and land usage. Aside from this, analyzing satellite images could lead to many different fields of study, such as Building Density. The density of buildings in a community impacts how crowded or built-up it appearsPeiser (2015). There are two ways to estimate building density: Building Coverage Ratio (BCR) and Floor Area Ratio (FAR)Pan et al. (2008). BCR refers to the area of the building divided by the area of the land. The floor space of a structure, when viewed from above, is referred to as its building area, while FAR refers to the proportion of total floor area to land area. The total floor area of a building is the sum of all the floor space. However, in this study, Building Coverage Ratio will be used to estimate the building density.

Techniques and tools were developed to solve complex real-world problems such as remote sensing and analysis of satellite images, especially in Artificial Intelligence (AI). One of the current trends in AI that is rapidly evolving is computer vision. Computer vision is an artificial intelligence field that trains computers to interpret and comprehend the visual world. It also allows machines to recognize and classify

---

CONTACT JJS Oñate. Email: josephjessie97@cspc.edu.ph

objects accurately and respond to what they see, using digital images from cameras and videos. Some techniques evolved, such as Image Classification, Object Detection, Image Segmentation, etc. Also, advanced computer vision techniques evolved over the years, and these techniques could be implemented using deep learning models that have developed significantly in the last decades.

Deep learning is an AI feature that imitates the human brain's workings in processing data for object detection, speech recognition, language translation, and decision making. Many deep learning model architectures were developed to improve this concept further. Nevertheless, this analysis will concentrate only on a particular architectural model of deep learning, the U-Net. U-net was initially developed and used mainly for the segmentation of biomedical images. Its architecture can be commonly seen as a network of encoders followed by a network of decoders.

Nonetheless, numerous successful uses of U-Net have been seen in various challenges and researches, especially in satellite image segmentation. Satellite Image Segmentation image involves detecting and classifying geospatial objects such as buildings, roads, cars, trees, etc., within the image. However, in the Philippines, no dataset was developed, and only several pieces of research or studies were conducted on satellite image segmentation. Improvements in U-Net architecture was seen in studies. One of which is Residual Attention U-Net (RAU-Net).

This study aims to train a Residual Attention U-Net Convolutional Neural Network (RAU-Net) Model using satellite images in the Philippines by classifying geospatial objects such as roofs to help determine the building density in a specific location.

This study is structured as follows; Section 2 explains the literature and studies related to the topics of this study. Subsequently, Chapter 3 presents the theories and concepts to understand the existing knowledge related to the topic of the study. Chapter 4 expounds on the research process, deep learning model, and evaluation metrics used in the conduct of this study. Chapter 5 elucidates the experimental results of the study guided by the research process by explaining each phase. Lastly, Chapter 6 discussed the summary and contributions of the study as well as the recommendations for future researches.

2. Literature Review

2.1. *Building Density*

Densification inwardly constructs the city is then seen as a housing strategy while battling the propensity of cities to spread and reclaim valuable lands. Swedish National Board of Housing (2017)

Some books and studies presented why density is important, such as in the book Higher-Density Development: Myth and Fact Haughey (2005), which presented some facts about high-density development: (1) there is a rising diversity of populations. Today, many families favor homes of greater density, including suburban areas; (2) people from any income level prefer housing with higher density; and (3) higher density development fascinates good residents and tenants and copes with emerging communities.

The population density is growing in accordance with extreme internal and external relocation of individuals to major cities. As a result, a growing number of new multi-story constructions and high-rise structures and high-speed expansion of the engineering and transportation infrastructure Giyasov and Giyasova (2018). These infrastructures can be measured by building density, an important measure of population

distribution since all human beings dwell and perform operations in buildings Zhou et al. (2016).

Building density is an important factor for urban design, planning, management, and urban environmental studies Pan et al. (2008);Yu et al. (2010). Such is the study of Rode et al. (2014), it was revealed that built type and density are among the most significant factors when it comes to per capita carbon emissions. Comp compact and tall building styles were found to have the highest heat-energy efficiency at the neighborhood scale, while detached housing had the lowest. This was also supported by Resch et al. (2016), which agree that for low urban energy usage, a much denser and taller city layout than what is currently common in cities appears to be optimal.

In measuring building density, there are indicators in measuring building density such as Floor Area Ratio (FAR), which is the ratio of the gross floor area of all buildings to the total area of the interest area, and Building Coverage Ratio (BCR), which is the ratio of the total standing area of all buildings to the total area of the interest area Pan et al. (2008). Some studies used different approaches such as Geographical Information System (GIS) that was utilized in the study of Ye and Van Nes (2013) in visualizing and quantifying spatial properties such as density of buildings, etc. in urban maturation processes. Aside from this, Meinel, Hecht, and Herold (2009) used GIS to compute building density by filtering outbuilding stock utilizing digital image processing. The general physical structure and block boundaries are then classified after the buildings are surveyed, their relationships with neighboring buildings are also interpreted. After which, density can now then be calculated either for the block boundaries or a reference geometry of one's choice. Further settlement structure parameters can be determined by combining building form with standard reference density values for floor number and building area-related dwelling and occupant densities. Another advanced approach is in the study of Yu et al. (2010), in which building density indicators are numerically and automatically derived from high-resolution airborne LiDAR data. These studies are some of the many approaches employed in measuring building density which gave insights to a researcher in the conduct of this study.

2.2. *Deep Learning in Extracting Region of Interest (ROI)*

ROI extraction was used in many ways to recognize the object in a picture, including medical imaging, security monitoring, database, and remote sensing imagery Kiadtikornthaweeyot and Tatnall (2016). One of the popular techniques in ROI extraction is Image segmentation. It is the process where an image is partitioned into several segments. It is usually used for finding representations of objects and boundaries Tan (2016). This technique is also a central subject in image processing and computer vision, among many others, applications such as scene recognition, medical imaging research, robotic perception, video detection, virtual reality, and image compression, Minaee et al. (2020). Many techniques on image segmentation were used to solve different real-world problems such as k-means and subtractive clustering Dhanachandra, Manglem, and Chanu (2015); different thresholding methods such as spatial Mardia and Hainsworth (1988), histogram Tobias and Seara (2002), recursive Cheriet, Said, and Suen (1998); region growing Shih and Cheng (2005) and many others. However, deep learning networks have created a new generation of image segmentation models over the past few years with impressive performance improvements, achieving the highest accuracy rates, which results in what many consider to be a paradigm shift in the field.

3

Many deep architectural networks were utilized in image segmentation. One of many networks is the SegNet Badrinarayanan, Kendall, and Cipolla (2017). SegNet can solve various image segmentation problems, particularly in brain tumor segmentation Alqazzaz et al. (2019), colon gland segmentation Tang, Li, and Xu (2018), oil spill detection Guo, Wei, and An (2018), etc. Nonetheless, it was also utilized on satellite image segmentation Barthakur and Sarma (2019); Zhu et al. (2018). Another common network architecture used in image segmentation is Fully Convolutional Networks (FCNs) Long, Shelhamer, and Darrell (2015). FCN were also used for semantic segmentation to analyze VHR remotely sensed images Sun and Wang (2018), slum areas Wurm et al. (2019) and building extraction Li et al. (2018).

Another popular architectural network in image segmentation is U-Net. U-Net is a convolutional neural network that was originally designed to segment biomedical images Ronneberger, Fischer, and Brox (2015a). The advantage of U-net to other models is that from very few images, the network can be trained end-to-end and outperforms the best method developed before this.

With the high IoU results achieved by the U-Net, many pieces of research utilized U-Net in different applications such as detection of a brain tumor Dong et al. (2017), cell counting, detection, and morphometry Falk et al. (2018), Pulmonary nodules segmentation Tong et al. (2018), etc. Many kinds of research also explored U-Net by rethinking the u-net architecture or merging it to other techniques such as nested u-net Zhou et al. (2018), U-Net with VGG11 Encoder Iglovikov and Shvets (2018), and many others.

Throughout the literature, there is a piece of consistent evidence that U-Net can be used in other fields aside from bio-medicine and can be trained with a small number of datasets which could help in attaining the objectives of this study.

2.3. *U-Net and its Advancements*

Apart from using U-net in biomedical images, it has also been shown that U-net can be used in other application fields. Nowadays, U-Net is also being employed on satellite image segmentation. The study of Ivanovsky et al. Ivanovsky et al. (2019a) used U-net to detect buildings on aerial images. The result shows that U-net performed well in the segmentation with 96.31% accuracy against LinkNet, which gets the only accuracy of 95.85%. AlexNet was also outperformed by U-net in the study of Freudenberg et al. Freudenberg et al. (2019a) on the detection of large-scale palm trees, having 88.6% accuracy against 75%. Besides, the study Bai et al. Bai, Mas, and Koshimura (2018) explored u-net in operational satellite-based damage-mapping and achieved an accuracy of 70.9%. Many other kinds of research and studies exploring the u-net have not been mentioned in this study. Nonetheless, the aforementioned studies prove that U-Net has been one of the best architecture in deep learning image segmentation. These facts led the researcher to use the u-net as network architecture for this study.

Today, U-Net was improved by integrating more blocks and functions to further improve its performance, one of the recent advancement of U-Net is Residual Attention U-Net (RAUNet). RAUNet was implemented in numerous paper such as the study of Chen, Yao, and Zhang (2020) for automated multi-class segmentation of COVID-19 chest CT images; extraction of liver and tumor in CT scans Jin et al. (2020); and nuclei segmentation in histology images Zeng et al. (2019) which has achieved competitive results in image segmentation.

3. Theoretical Framework

3.1. *Satellite Images*

According to Wu et al. (2019), when compared to conventional images, satellite imagery contains more organized and uniform data. It also contains rich context information such as roofs, buildings, roads, etc. This was supported by Ohleyer and Paris-Saclay (2018) which used two semantic classes, the building and the not building, in his study on segmentation of satellite images which is also the focus of this study. There is much available satellite imagery dataset. However, the location of these satellite imagery is limited to a specific location globally, and there is no publicly available dataset for the Philippines. Hence, in this study, the dataset will be collected and will be manually labeled by humans. Hernandez also used this technique Hernandez (2019) in semantic segmentation of Mammographies. Furthermore, only a few studies were conducted on the analysis of Philippine Satellite Images, such as the study of Tingzon et al. (2019) which estimates the poverty in the Philippines using satellite images and crowd-sourced geospatial information. Mentioned above are the influential studies that support the concept and the gaps that will be addressed in this study.

3.2. *Data Augmentation*

Another useful technique to further improve the number of data is Data augmentation. According to Lux (2018), data augmentation is a technique often used when a training dataset is too small. The goal is to produce new data from the same distribution of probabilities as test samples to increase the efficiency of the neural network. In addition, a variety of functions are expanded by creating new samples in the training dataset by this method. Many methods of data augmentation are based on random improvements to current samples Lux (2018).

3.3. *U-Net Model*

Ronneberger, Fischer, and Brox (2015b) developed the U-Net architectural model shown in Figure 1 at the Computer Science Department of the University of Freiburg, Germany, to segmentation biomedical images. Despite that it was developed to segment biomedical images, further testing and researches proved that it could be used in other fields such as satellite image segmentation. U-Net consists of two paths, the encoder, and the decoder path. Encoder path, which captures the context of the feature maps that produce the image. The encoder path is nothing more than a stack of convolution layers and max pooling. The encoder path is composed of four (4) blocks. Each block consists of two 3 x 3 convolution layers plus a ReLU activation function with batch normalization and one 2 x 2 max-pooling layer. In the decoder path, using transposed convolutions, the decoder path used to allow precise localization. The decoder path consists of 4 blocks. Each block contains a deconvolution layer with stride 2, concatenation with the corresponding cropped feature map from the contracting path, and two 3 x 3 convolution layers plus ReLU activation function with batch normalization.

Many implementations have proven that U-net contributed significantly to the field of image segmentation stipulated in the literature review. The original paper has achieved an IoU average of 92% on PhC-U373 and 77.5% on DIC-HeLa datasets surpassing the IoU rating of other models. Moreover, in the study of Hernandez Hernandez
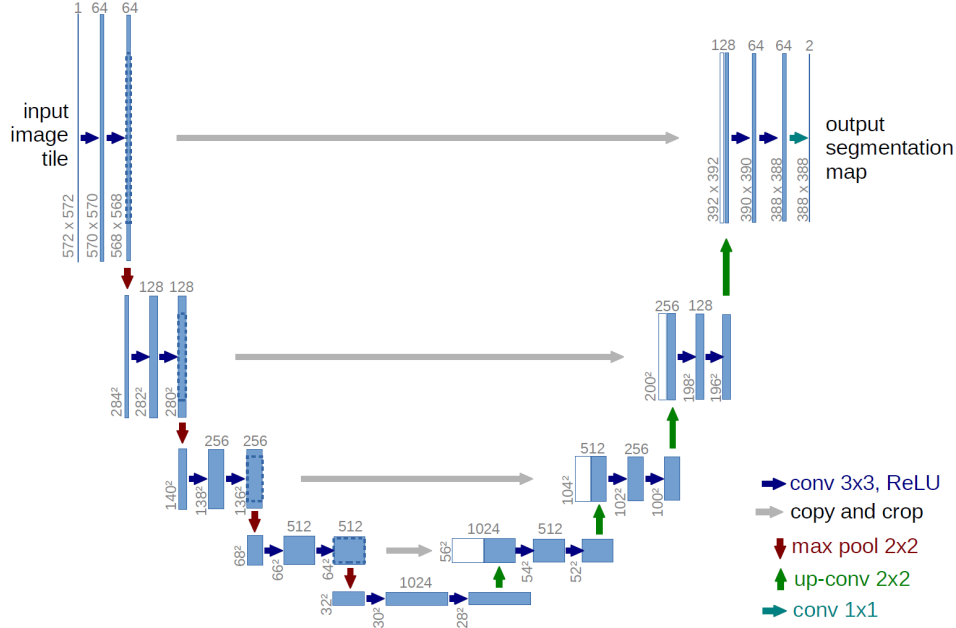
Figure 1. U-Net Architectural ModelRonneberger, Fischer, and Brox (2015b)

(2019), it was proven that U-Net could learn with just a few labeled images that are well suited for image segmentation in different fields, especially in satellite image segmentation presented in the study of Ulmas and Liiv (2020); Freudenberg et al. (2019b) with the accuracy of 74.9% and 93.9%, respectively. Given all the advantages of the U-Net model, the U-Net model will be used as the convolutional neural network architecture in this study and could be implemented in Tensorflow introduced by Abadi et al. (2016). TensorFlow is a dataflow software library and programming that can be distinguished between various activities. Tensorflow is a library that is also used in machine learning like neural networks. TensorFlow provides developers new optimizations and algorithms for model training and evaluation. TensorFlow can be used to develop and train the U-Net Model easily and effectively.

3.4. *Residual Attention U-Net (RAUNet)*

Today, many advances of U-Net architecture has been implemented by various scholars in different fields such as Residual Attention U-Net (RAUNet) Ni et al. (2019). RAUNet is a U-Net architecture integrated with Residual Layer and Attention Layer. Residual U-Net (ResUnet) was introduced by Zhang, Liu, and Wang (2018) for road extraction. ResUnet shown in Figure 2 makes the U-Net network training easier. Furthermore, the skip connections within a residual unit and between low and high levels of the network will facilitate information propagation without degradation, allowing a neural network to be designed with far fewer parameters while still achieving comparable or even better semantic segmentation performance.

On the other hand, Attention U-Net was also developed by Oktay et al. (2018) for Pancreas segmentation by applying attention gate (AG) model. In the proposed AG shown in Figure 3, the attention coefficients (α) produced are used to scale the input features (x^i). The activations and contextual information supplied by the gating signal (g), which is gathered on a coarser scale, are used to identify spatial areas. Trilinear
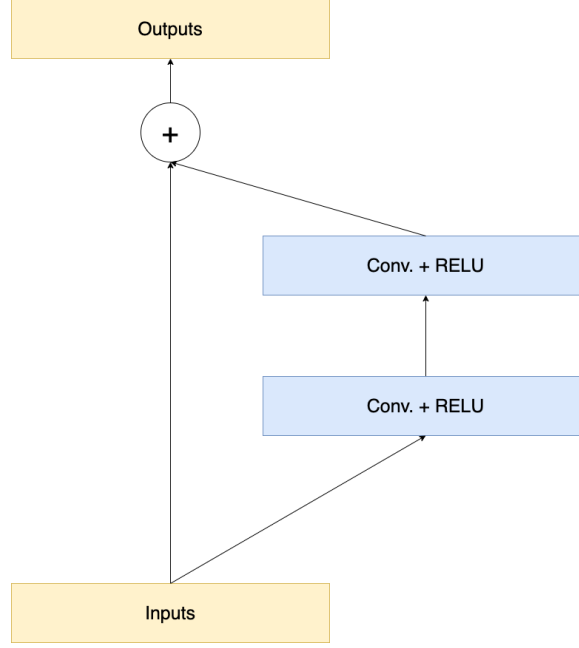
6

Figure 2. Residual U-Net (ResUnet)

interpolation is also used to grid resample attention coefficients.
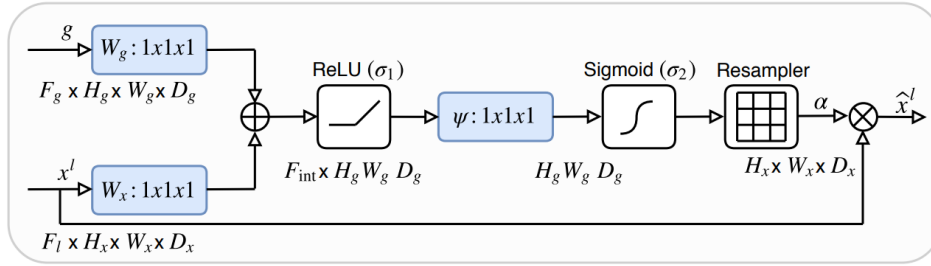

Figure 3. Attention Gate

3.5. *Evaluation Metrics*

For tasks such as segmentation, object detection, and tracking, intersection over union (IoU), also known as the Jaccard index, is one of the most common evaluation metrics developed Jaccard (1912). To apply Intersection over Union in evaluating the prediction of a segmented image, more formally, the Equation 1 will be followed. A will set as the ground truth image, and B will set as the predicted image. Simply, by getting the intersection value between A and B and divide it with the union value of A and B.

$$IoU = \frac{|A \cap B|}{|A \cup B|} \tag{1}$$

Another common evaluation metric is Dice Coefficient developed by Sørensen (1948). Dice Coefficient is similar to IoU however, Dice Coefficient doubles count the

intersection shown on the Equation 2.

$$DC = \frac{2|A \cap B|}{|A| + |B|} \tag{2}$$

Furthermore, model will also be evaluated using Macro F1 shown on Equation 3. Macro F1 is the harmonic mean between precision and recall, where the average is calculated per label and then averaged across all labels.

$$\text{Macro F1-score} = \frac{1}{N} \sum_{i=0}^{N} \text{F1-score}_i \tag{3}$$

F1 Score shown on Equation 4 is being computed as . This is the harmonic mean of Precision and Recall, and it provides a more accurate picture of instances that were erroneously categorized than the Accuracy Metric.

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

Precision shown in Equation 5 is the ratio of correctly predicted positive observations to the total predicted positive observations.

$$Precision = \frac{TruePositive}{(TruePositive + FalsePositive)} \tag{5}$$

Recall shown in Equation 6 is the ratio of correctly predicted positive observations to the all observations in actual class.

$$Recall = \frac{TruePositive}{(TruePositive + FalseNegative)} \tag{6}$$

3.6. *Building Detection*

Similar to this study is the paper of Ivanovsky et al. (2019b) which segments Inria satellite images in detecting buildings using U-Net however, the dataset was only in selected USA locations with a Sorensen-Dice coefficient (DSC) of 77%. Hence, this study wants to address the gap in which there is no current exploration or study on building detection of locations in the Philippines using U-Net.

On the analysis of building density, according to Pan et al. (2008), Building Coverage Ratio (BCR) is generally computed by the ratio of the sum of the areas of all building denoted with F divided by the area of land they occupy denoted by A shown on Equation 7.

$$BCR = \sum F/A \tag{7}$$

Furthermore, the study of de Bellefon et al. (2019) introduced a new dartboard technique for delineating urban areas based on comprehensive building location details, which is implemented using a map of all buildings in France. The new methodology introduced gives a new foundation on how to measure building density. However, the process is tedious and requires a lot of work. Hence, this methodology will be improved by adding additional methods to compute the building density using the segmentation of satellite images.

4. Methodology

This study will follow the process enumerated below to attain the objectives of this study. The process is compacted phases viz, dataset gathering, data processing, augmentation supervised learning, model evaluation, and building density metric.

(1) **Dataset Gathering.** In this phase, the collection of data will be conducted. Satellite images will be scraped from different random locations in the Philippines.

(2) **Data Pre-processing and Augmentation.** After which, the collected images will be labeled. Once the labeling is done, the dataset will be processed and arranged according to the input specification of the model.

(3) **Supervised Learning.** Subsequently, once the dataset was established, training of the model will be conducted through supervised learning. Model training aims to find a collection of weights and biases with low losses, on average, across all classes.

(4) **Model Evaluation.** After training and testing, the model's accuracy will be evaluated to estimate the generalization accuracy of a model on future data.

(5) **Building Density Metric.** Ensuring that the model works properly, achieved a high accuracy percentage based on the evaluation, the model will be deployed considering an intuitive and user-friendly approach in this phase.

4.1. *Data Gathering*
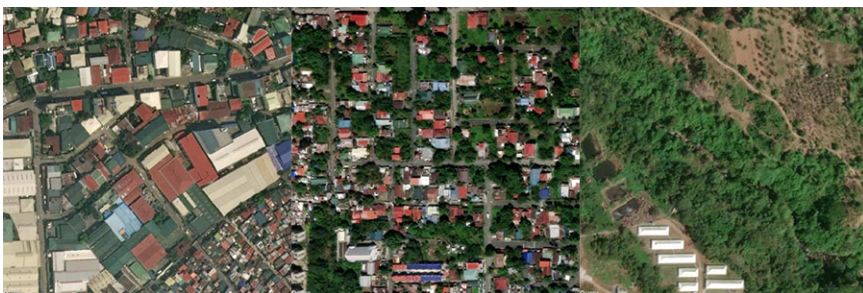
Figure 4. Sample of Satellite Image Dataset

Since there is no readily available satellite image dataset for the Philippines, this study collected 200 RGB-Satellite images in JPEG format using Google Earth Pro on random places in the Philippines, such as urban cities like Manila and Pasig, provinces such as Laguna and Cavite, etc. shown on Figure 4. Images were saved in a resolution

of 3840x2160(4K UHD) pixels, resized, and cropped to 2160x2160 pixels using Python code.


Figure 5. Labelling of Image Dataset

After the collection, these satellite image datasets were uploaded to Labelbox Labelbox (2020) to manually label the building roofs and decks are shown in Figure 5. After labeling, Labelbox automatically generates a mask for the labeled satellite image, and the data were exported to a JavaScript Object Notation (JSON) file. Since the image and generated mask cannot be exported in image form, the researcher has written a Python code that parses the exported JSON file and automatically arrange and download the image dataset and the labeled mask. The code will arrange the dataset as a ready input for the model.

4.2. *Dataset Pre-Processing and Augmentation*

The image and mask dataset was pre-processed to 256x256 pixel and converted into NumPy array with 256 x 256 x 3 shape for satellite image and 256 x 256 x 2 shape for mask image in Figure 6.

| | ((a)) Satellite Image | ((b)) Ground Truth Mask |
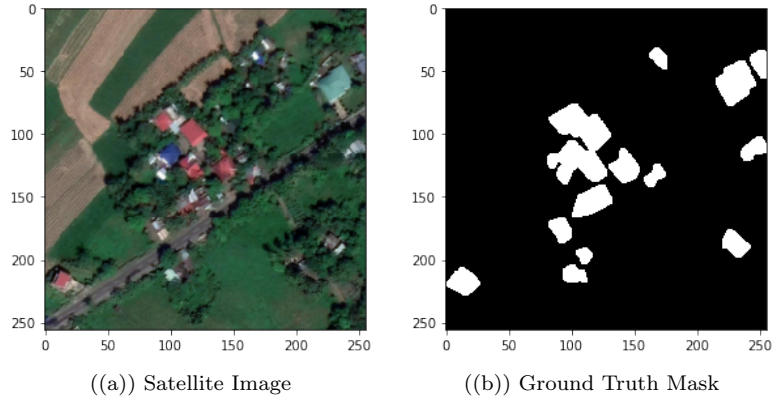
Figure 6. Pre-processed Image Dataset

Subsequently, the one hundred and forty image data was augmented by rotating, horizontally flipping, and vertically flipping to produce more input image data shown in Table 1. Data augmentation help reduce overfitting when training a model.

Table 1. Training and Testing Dataset Size

Dataset	Original	Augmented
training dataset	140	700
validation dataset	60	300
Total	200	1000

4.3. Supervised Learning

In extracting the ROI in an image, supervised learning will be used to map an input satellite image to an output based on trained input-output pairs (satellite image and ground truth mask).

On each block of architecture, a dropout of 10-30% was added to ensure that the model will not overfit. Overfitting occurs when a model forecasts an overly noisy trend of data. This is due to a model with too few parameters, which is overly complex. An overfitted model is wrong, as the trend does not represent the reality in the data.

Deep neural networks require the stochastic optimization of the gradient so that the cost function of their parameters is minimized. To estimate the parameters, we have adopted the adaptive time estimator (Adam). Adam typically uses the first and second phases of gradients to update and correct the average moving gradients. The model was trained with the same parameter in 100 epochs with batch size of 5. Since the researcher has no high-end device in training the model, Google Colab Pro was used to train and test the models which provide faster GPUs.

Also, to measure how good the prediction model predicts the expected outcome, this study used the binary cross-entropy as the loss function since the prediction is in binary form. Cross-entropy for a binary or two-class prediction problem is actually calculated as the average cross-entropy across all examples.

4.4. Model Evaluation

To evaluate the model, intersection over union (IoU), Dice Coefficient, Macro F1, Precision and Recall will be used to measure the model's overall performance. This metrics is substantial measure of how well a model of image segmentation performs in

11

Table 2. Building Coverage Pixel Ratio Scale.

Building Coverage Pixel Ratio(%)	Density Classification
81-100	High Density
61-80	Medium-High Density
41-60	Medium Density
21-40	Medium-low Density
0-20	Low Density

all the classes the model would like to identify.

4.5. *Building Density Metric*

In this study, BCR will be computed using the Equation 8 by getting the total number of pixel of segmented building pSI and divide it by the total number of pixel of the original image pOI shown on Equation 8 since the study is limited to unlabelled density on satellite image dataset.

$$BuildingCoveragePixelRatio(\%) = \frac{pSI}{pOI}x100 \qquad (8)$$

It will be assessed using the Five-Point Scale to determine the assessment criteria of the segmented image Landcom (2011). Categories consist of low, medium, and high density. Two categories were added, medium-high density and medium-low density. Since there are five classifications, the range is statistically computed for each classification shown in Table 2.

Subsequently, it will be applied to a system to make the model and the analysis more intuitive and user-friendly. Flask, a python microframework, and web technologies such as HTML, CSS, etc., will be used to utilize and integrate the model into the system easily.

5. Experimental Results

5.1. *Model Performance*

Table 3. Segmentation Performance of RAU-Net against Base U-Net

Method	Dice(%)	IoU (%)	Macro F1 (%)	Param
U-Net	83.14	77.70	87.76	31.40M
RAUNet	86.43	80.30	90.03	39.09M

After the training, the model was evaluated using the testing dataset. Table 3 shows that RAUNet outperforms the base U-Net including the U-Net with Attention and U-Net with Residual with performance scores of 86.43%, 80.30%, and 90.03% in Dice Coefficient, IoU and Macro F1, respectively. The training time reaches to approximately 1 hour with 39.09 million parameters. Figure 7 is the vizualization of RAU-Net

Segmentation Result. On the first column is the satellite image, ground truth mask on the second column, and segmentation result on the third column.

5.2. *Ablation Study*

Table 4. Segmentation Performance of Ablation Study

Method	Dice(%)	IoU (%)	Macro F1 (%)	Param
U-Net	83.14	77.70	87.76	31.40M
AU-Net	84.80	78.94	88.86	37.33M
RU-Net	86.22	80.14	89.97	33.15M
RAUNet	86.43	80.30	90.03	39.09M

This study also tries to determine the performance score of each added components in U-Net. We have conducted a separate training for the added Attention component and Residual component. The training parameters were identical to ensure establish a fair comparison. Table 4 shows that U-Net with Residual achieved performance scores of 86.22%, 80.14%, and 89.97% in Dice Coefficient, IoU, and Macro F1, respectively with a 33.15 million parameters. For the U-Net with Attention, it has achieved performance scores of 84.40%, 78.90%, and 88.86% in Dice Coefficient, IoU and Macro F1, respectively with 37.33 million parameters. It also shows that the two added component in base U-Net plays a significant role in improving the performance of the model.

5.3. *Model Testing*

The model was tested by inputting new data, shown in Figure 7 are the sample image and the segmented result. The first column is the input satellite image, the ground truth mask is shown on the second column, and the segmented image using the model is shown on the third column. It shows that the model is segmenting the satellite image correctly validated by good accuracy scores.

5.4. *Model Application*

Ensuring that the trained U-Net Model obtained a high segmentation accuracy, an application coined as BuilDense was developed to test the model on a real map. Figure 8 is the actual screenshot of the application, which shows the satellite map on the right part. The application also has the capacity to search for locations and automatically set the satellite map. The map is set to default minimum zoom of 16. Based on Mapbox, the geographical distance is 1.194 meters/pixel (3.92 feet/pixel). It can also be adjusted up to 17, which the geographical distance is 0.597 meters/pixel (1.96 feet/pixel). Once the analyze button is clicked, a segmentation result will appear on the right part and the building pixel density.

6. Discussion and Conclusions

In this study, the main goal is to train a RAU-Net model using Philippine Satellite Images by segmenting geospatial objects such as roofs to predict the building density

((a))

((b))

((c))

((d))

((e))
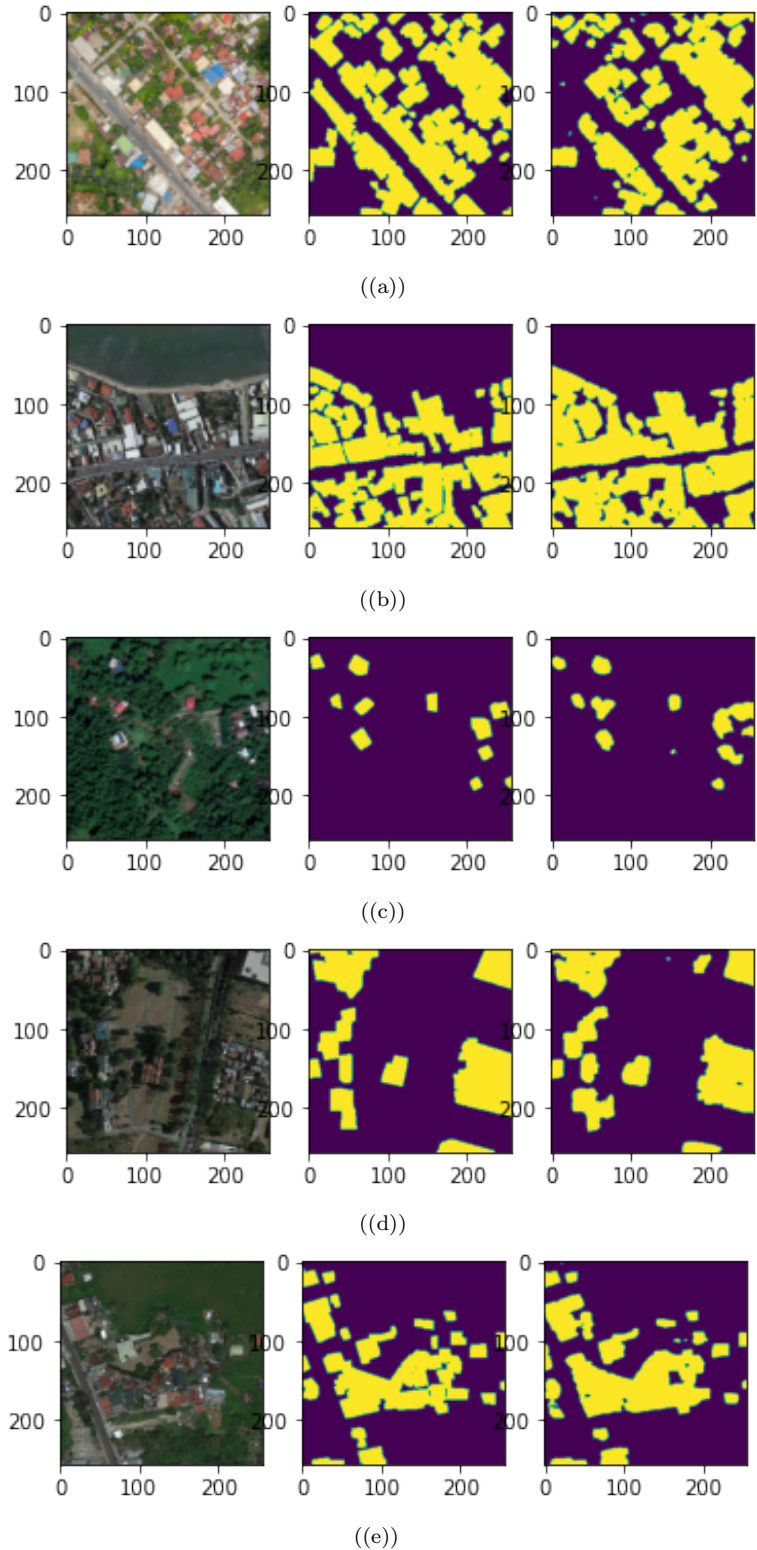
Figure 7. RAUNet Model Visualization of Segmentation Result

14
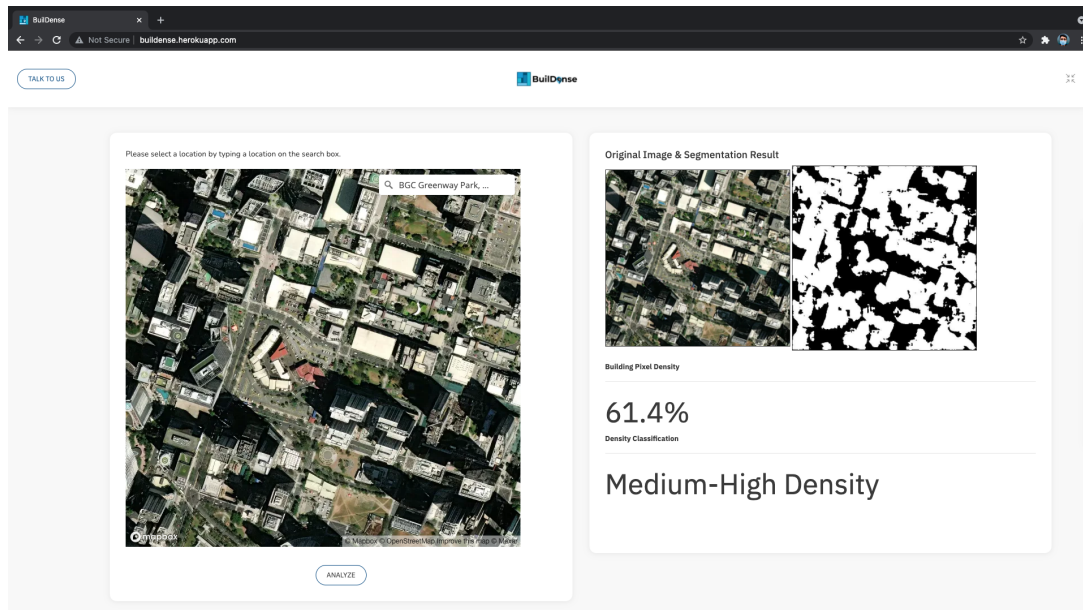
Figure 8. BuilDense Application

automatically.

Satellite Images were collected using Google Earth Pro in different locations in the Philippines, such as Laguna, Cavite, etc., and manually labeled by humans using Labelbox to generate ground truth mask images. A total of 200 satellite images were collected, which is relatively small compared to the other publicly available dataset. However, the satellite image dataset and the ground truth mask were augmented to produce more image datasets. After the augmentation, the image dataset increased to 400%.

A Residual Attention U-Net (RAU-Net) model was trained using the satellite image dataset and the ground truth mask as the input to the model. The trained model automatically detects building. The accuracy of the trained U-Net model was determined by getting the Macro F1, Intersection over Union and Dice Score as the chosen evaluation metrics. After a tedious training process, the trained U-Net model achieved performance scores of 86.43%, 80.30%, and 90.03% in Dice Coefficient, IoU and Macro F1, respectively.

Moreover, an application coined as BuilDense was also developed to test the trained U-Net Model into a live map using Flask, a python micro-framework combined with web technologies such as HTML, CSS. This study also presents a way to calculate the building density by calculating the building coverage pixel ratio of the segmented image.

After a thorough analysis of the study, our future work would be improving the number of the image dataset since the number of developed Philippines satellite images is relatively small. Also, The researcher focuses on Residual Attention U-Net Convolutional Neural Network, one of many other segmentation techniques hence we would be testing the dataset on other state-of-the-art image segmentation techniques. Another limitation of this study is that it was focused on the Building Coverage Ratio (BCR) hence, the researcher would also further extend the study by analyzing Floor Area Ratio (FAR) using satellite images or other data available.

15

Acknowledgement(s)

References

Abadi, Martín, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. 2016. "TensorFlow: A System for Large-Scale Machine Learning." In *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, Savannah, GA, Nov., 265–283. USENIX Association. https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi.

Alqazzaz, Salma, Xianfang Sun, Xin Yang, and Len Nokes. 2019. "Automated brain tumor segmentation on multi-modal MR image using SegNet." *Computational Visual Media* 5 (2): 209–219. https://doi.org/10.1007/s41095-019-0139-y.

Badrinarayanan, V., A. Kendall, and R. Cipolla. 2017. "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (12): 2481–2495.

Bai, Yanbing, Erick Mas, and Shunichi Koshimura. 2018. "Towards Operational Satellite-Based Damage-Mapping Using U-Net Convolutional Network: A Case Study of 2011 Tohoku Earthquake-Tsunami." *Remote Sensing* 10 (10): 1626. https://doi.org/10.3390/rs10101626.

Barthakur, M., and K. K. Sarma. 2019. "Semantic Segmentation using K-means Clustering and Deep Learning in Satellite Image." In *2019 2nd International Conference on Innovations in Electronics, Signal Processing and Communication (IESC)*, 192–196.

Chen, Xiaocong, Lina Yao, and Yu Zhang. 2020. "Residual Attention U-Net for Automated Multi-Class Segmentation of COVID-19 Chest CT Images." .

Cheriet, M., J. N. Said, and C. Y. Suen. 1998. "A recursive thresholding technique for image segmentation." *IEEE Transactions on Image Processing* 7 (6): 918–921.

de Bellefon, Marie-Pierre, Pierre-Philippe Combes, Gilles Duranton, Laurent Gobillon, and Clément Gorin. 2019. "Delineating urban areas using building density." *Journal of Urban Economics* 103226. https://www.sciencedirect.com/science/article/pii/S0094119019301032.

Dhanachandra, Nameirakpam, Khumanthem Manglem, and Yambem Jina Chanu. 2015. "Image Segmentation Using K -means Clustering Algorithm and Subtractive Clustering Algorithm." *Procedia Computer Science* 54: 764 – 771. Eleventh International Conference on Communication Networks, ICCN 2015, August 21-23, 2015, Bangalore, India Eleventh International Conference on Data Mining and Warehousing, ICDMW 2015, August 21-23, 2015, Bangalore, India Eleventh International Conference on Image and Signal Processing, ICISP 2015, August 21-23, 2015, Bangalore, India, http://www.sciencedirect.com/science/article/pii/S1877050915014143.

Dong, Hao, Guang Yang, Fangde Liu, Yuanhan Mo, and Yike Guo. 2017. "Automatic Brain Tumor Detection and Segmentation Using U-Net Based Fully Convolutional Networks." In *Communications in Computer and Information Science*, 506–517. Springer International Publishing. https://doi.org/10.1007/978-3-319-60964-5$_4$4.

Falk, Thorsten, Dominic Mai, Robert Bensch, Özgün Çiçek, Ahmed Abdulkadir, Yassine Marrakchi, Anton Böhm, et al. 2018. "U-Net: deep learning for cell counting, detection, and morphometry." *Nature Methods* 16 (1): 67–70. https://doi.org/10.1038/s41592-018-0261-2.

Freudenberg, Maximilian, Nils Nölke, Alejandro Agostini, Kira Urban, Florentin Wörgötter, and Christoph Kleinn. 2019a. "Large Scale Palm Tree Detection In High Resolution Satellite Images Using U-Net." *Remote Sensing* 11 (3): 312. https://doi.org/10.3390/rs11030312.

Freudenberg, Maximilian, Nils Nölke, Alejandro Agostini, Kira Urban, Florentin Wörgötter,

and Christoph Kleinn. 2019b. "Large Scale Palm Tree Detection in High Resolution Satellite Images Using U-Net." *Remote Sensing* 11 (3). https://www.mdpi.com/2072-4292/11/3/312.

Giyasov, Botir, and Irina Giyasova. 2018. "The Impact of High-Rise Buildings on the Living Environment." *E3S Web of Conferences* 33: 01045. https://doi.org/10.1051/e3sconf/20183301045.

Guo, Hao, Guo Wei, and Jubai An. 2018. "Dark Spot Detection in SAR Images of Oil Spill Using Segnet." *Applied Sciences* 8 (12): 2670. https://doi.org/10.3390/app8122670.

Haughey, R. 2005. *Higher-Density Development: Myth and Fact.*

Hernandez, Albert Bou. 2019. "Deep Learning models for semantic segmentation of mammography screenings." .

Iglovikov, Vladimir, and Alexey Shvets. 2018. "TernausNet: U-Net with VGG11 Encoder Pre-Trained on ImageNet for Image Segmentation." .

Ivanovsky, Leonid, Vladimir Khryashchev, Vladimir Pavlov, and Anna Ostrovskaya. 2019a. "Building Detection on Aerial Images Using U-NET Neural Networks." In *2019 24th Conference of Open Innovations Association (FRUCT)*, Apr. IEEE. https://doi.org/10.23919/fruct.2019.8711930.

Ivanovsky, Leonid, Vladimir Khryashchev, Vladimir Pavlov, and Anna Ostrovskaya. 2019b. "Building Detection on Aerial Images Using U-NET Neural Networks." In *2019 24th Conference of Open Innovations Association (FRUCT)*, 116–122.

Jaccard, Paul. 1912. "THE DISTRIBUTION OF THE FLORA IN THE ALPINE ZONE.1." *New Phytologist* 11 (2): 37–50. https://doi.org/10.1111/j.1469-8137.1912.tb05611.x.

Jin, Qiangguo, Zhaopeng Meng, Changming Sun, Hui Cui, and Ran Su. 2020. "RA-UNet: A Hybrid Deep Attention-Aware Network to Extract Liver and Tumor in CT Scans." *Frontiers in Bioengineering and Biotechnology* 8. https://doi.org/10.3389/fbioe.2020.605132.

Kiadtikornthaweeyot, Warinthorn, and Adrian R. L. Tatnall. 2016. "REGION OF INTEREST DETECTION BASED ON HISTOGRAM SEGMENTATION FOR SATELLITE IMAGE." *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLI-B7: 249–255. https://doi.org/10.5194/isprs-archives-xli-b7-249-2016.

Labelbox. 2020. "Labelbox: The leading training data platform for data labeling." https://labelbox.com/. (Accessed on 10/23/2020).

Landcom. 2011. "Density Guide Book." https://www.landcom.com.au, May. (Accessed on 09/22/2020).

Li, Weijia, Conghui He, Jiarui Fang, and Haohuan Fu. 2018. "Semantic Segmentation Based Building Extraction Method Using Multi-source GIS Map Datasets and Satellite Imagery." 07.

Long, Jonathan, Evan Shelhamer, and Trevor Darrell. 2015. "Fully Convolutional Networks for Semantic Segmentation." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June.

Lux, Filip. 2018. "Segmentation of Dense Cell Populations using Convolutional Neural Networks." https://is.muni.cz/th/v35vy/DP$_L$ux$_F$ilip.pdf.

Mardia, K. V., and T. J. Hainsworth. 1988. "A spatial thresholding method for image segmentation." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 10 (6): 919–927.

Meinel, Gotthard, Robert Hecht, and Hendrik Herold. 2009. "Analyzing building stock using topographic maps and GIS." *Building Research & Information* 37 (5-6): 468–482. https://doi.org/10.1080/09613210903159833.

Minaee, Shervin, Yuri Boykov, Fatih Porikli, Antonio Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos. 2020. "Image Segmentation Using Deep Learning: A Survey." .

Ni, Zhen-Liang, Gui-Bin Bian, Xiao-Hu Zhou, Zeng-Guang Hou, Xiao-Liang Xie, Chen Wang, Yan-Jie Zhou, Rui-Qi Li, and Zhen Li. 2019. "RAUNet: Residual Attention U-Net for Semantic Segmentation of Cataract Surgical Instruments." .

Ohleyer, Sébastien, and ENS Paris-Saclay. 2018. "Building segmentation on satellite images." 1–5. http://arxiv.org/abs/1703.06870.

Oktay, Ozan, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, et al. 2018. "Attention U-Net: Learning Where to Look for the Pancreas." .

Pan, Xian-Zhang, Qi-Guo Zhao, Jie Chen, Yin Liang, and Bo Sun. 2008. "Analyzing the Variation of Building Density Using High Spatial Resolution Satellite Images: the Example of Shanghai City." *Sensors* 8 (4): 2541–2550. https://www.mdpi.com/1424-8220/8/4/2541.

Peiser, Richard. 2015. "Real Estate Development." In *International Encyclopedia of the Social Behavioral Sciences (Second Edition)*, edited by James D. Wright, second edition ed., 12–19. Oxford: Elsevier. https://www.sciencedirect.com/science/article/pii/B9780080970868740347.

Resch, Eirik, Rolf André Bohne, Trond Kvamsdal, and Jardar Lohne. 2016. "Impact of Urban Density and Building Height on Energy Use in Cities." *Energy Procedia* 96: 800–814. Sustainable Built Environment Tallinn and Helsinki Conference SBE16, https://www.sciencedirect.com/science/article/pii/S1876610216307810.

Rode, Philipp, Christian Keim, Guido Robazza, Pablo Viejo, and James Schofield. 2014. "Cities and Energy: Urban Morphology and Residential Heat-Energy Demand." *Environment and Planning B: Planning and Design* 41 (1): 138–162. https://doi.org/10.1068/b39065.

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. 2015a. "U-Net: Convolutional Networks for Biomedical Image Segmentation." In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, edited by Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, Cham, 234–241. Springer International Publishing.

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. 2015b. *U-Net: Convolutional Networks for Biomedical Image Segmentation*, 234–241. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4_28.

Shih, Frank Y., and Shouxian Cheng. 2005. "Automatic seeded region growing for color image segmentation." *Image and Vision Computing* 23 (10): 877 – 886. http://www.sciencedirect.com/science/article/pii/S0262885605000673.

Sørensen, T. 1948. *A method of establishing groups of equal amplitude in plant sociology based on similarity*. Vol. 5. København: I kommission hos E. Munksgaard.

Sun, W., and R. Wang. 2018. "Fully Convolutional Networks for Semantic Segmentation of Very High Resolution Remotely Sensed Images Combined With DSM." *IEEE Geoscience and Remote Sensing Letters* 15 (3): 474–478.

Swedish National Board of Housing. 2017. *Urban Density Done Right*. https://www.boverket.se/globalassets/publikationer/dokument/2017/urban-density-done-right.pdf.

Tan, Ying. 2016. "Chapter 11 - Applications." In *Gpu-Based Parallel Implementation of Swarm Intelligence Algorithms*, edited by Ying Tan, 167 – 177. Morgan Kaufmann. http://www.sciencedirect.com/science/article/pii/B978012809362750011X.

Tang, J., J. Li, and X. Xu. 2018. "Segnet-based gland segmentation from colon cancer histology images." In *2018 33rd Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, 1078–1082.

Tingzon, I., A. Orden, K. T. Go, S. Sy, V. Sekara, I. Weber, M. Fatehkia, M. García-Herranz, and D. Kim. 2019. "MAPPING POVERTY IN THE PHILIPPINES USING MACHINE LEARNING, SATELLITE IMAGERY, AND CROWD-SOURCED GEOSPATIAL INFORMATION." *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLII-4/W19: 425–431. https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLII-4-W19/425/2019/.

Tobias, O. J., and R. Seara. 2002. "Image segmentation by histogram thresholding using fuzzy sets." *IEEE Transactions on Image Processing* 11 (12): 1457–1465.

Tong, Guofeng, Yong Li, Huairong Chen, Qingchun Zhang, and Huiying Jiang. 2018. "Improved U-NET network for pulmonary nodules segmentation." *Optik* 174: 460 – 469. http://www.sciencedirect.com/science/article/pii/S003040261831235X.

Ulmas, Priit, and Innar Liiv. 2020. "Segmentation of Satellite Imagery using U-Net Models for Land Cover Classification." .

Wu, M., C. Zhang, J. Liu, L. Zhou, and X. Li. 2019. "Towards Accurate High Resolution Satellite Image Semantic Segmentation." *IEEE Access* 7: 55609–55619.

Wurm, Michael, Thomas Stark, Xiao Xiang Zhu, Matthias Weigand, and Hannes Taubenböck.

2019. "Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks." *ISPRS Journal of Photogrammetry and Remote Sensing* 150: 59 – 69. http://www.sciencedirect.com/science/article/pii/S0924271619300383.

Ye, Yu, and Akkelies Van Nes. 2013. "Measuring urban maturation processes in Dutch and Chinese new towns: Combining street network configuration with building density and degree of land use diversification through GIS." *Journal of Space Syntax* 4 (1): 1 8– 37. http://www.journalofspacesyntax.org/.

Yu, Bailang, Hongxing Liu, Jianping Wu, Yingjie Hu, and Li Zhang. 2010. "Automated derivation of urban building density information using airborne LiDAR data and object-based method." *Landscape and Urban Planning* 98 (3): 210–219. Climate Change and Spatial Planning, https://www.sciencedirect.com/science/article/pii/S016920461000191X.

Zeng, Zitao, Weihao Xie, Yunzhe Zhang, and Yao Lu. 2019. "RIC-Unet: An Improved Neural Network Based on Unet for Nuclei Segmentation in Histology Images." *IEEE Access* 7: 21420–21428. https://doi.org/10.1109/access.2019.2896920.

Zhang, Zhengxin, Qingjie Liu, and Yunhong Wang. 2018. "Road Extraction by Deep Residual U-Net." *IEEE Geoscience and Remote Sensing Letters* 15 (5): 749–753. https://doi.org/10.1109/lgrs.2018.2802944.

Zhou, Yi, Chenxi Lin, Shixin Wang, Wenliang Liu, and Ye Tian. 2016. "Estimation of Building Density with the Integrated Use of GF-1 PMS and Radarsat-2 Data." *Remote Sensing* 8 (11): 969. https://doi.org/10.3390/rs8110969.

Zhou, Zongwei, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. 2018. "UNet: A Nested U-Net Architecture for Medical Image Segmentation." In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, 3–11. Springer International Publishing. https://doi.org/10.1007/978-3-030-00889-5_1.

Zhu, B., H. Gao, X. Wang, M. Xu, and X. Zhu. 2018. "Change Detection Based on the Combination of Improved SegNet Neural Network and Morphology." In *2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC)*, 55–59.