Høgskolen i Østfold

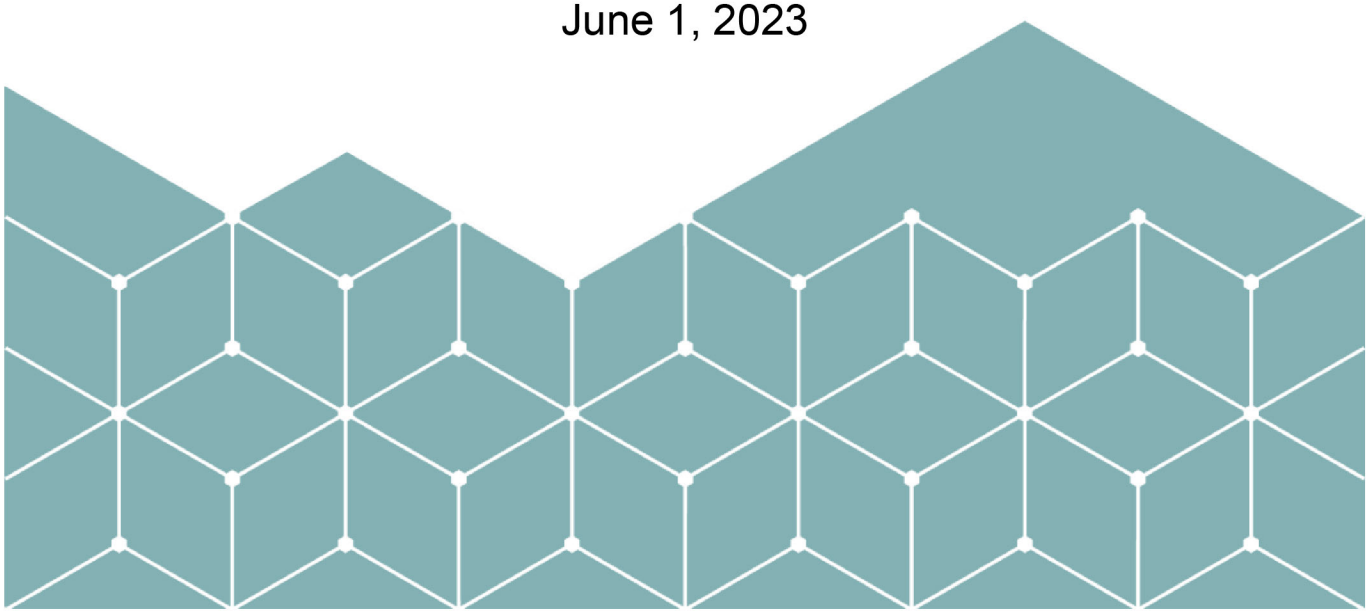
Master's Thesis

A Comparative Study on Machine Learning Approaches for Semantic Segmentation of Land Cover

Ahmad Bilal Aslam

School of Computer Sciences
Østfold University College
Halden

June 1, 2023



Master's Thesis

A Comparative Study on Machine Learning Approaches for Semantic Segmentation of Land Cover

A thesis presented for the degree of Master in Applied Computer Science

Ahmad Bilal Aslam

School of Computer Sciences
Østfold University College
Halden
June 14, 2023

Abstract

Power line inspection has a critical role in providing uninterrupted power supply. Vegetation surrounding power lines can cause power outages by impacting them. Utility owners spend a lot of resources on the surveillance of transmission lines. In this thesis, we performed land cover semantic segmentation using satellite imagery obtained from International Society for Photogrammetry and Remote Sensing (ISPRS) 2D Semantic Labeling Contest for Potsdam and Vaihingen. We analyzed the performance of different deep neural architectures on these dataset. This information can help utilities identify the area requiring their attention.

Keywords: Aerial Imagery, Power line Inspection, Satellite Images, Semantic Segmentation

Acknowledgments

I am deeply grateful to my supervisor, Lars Magnusson, for his invaluable guidance, and to my family, friends, and colleagues for their unwavering support and enriching discourse, all of which have been instrumental in the successful completion of this thesis.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Grid-lines are an important part of the infrastructure that provides power to our homes, companies, and communities. Regular inspections and timely maintenance are crucial for lowering resource input, maintaining cheap pricing, and restoring power quickly. However, the current methods are insufficient. Most power grid operators use on-the-ground staff and low-flying helicopters to monitor their electrical wires. Drones are also being used to inspect electrical lines. These devices can be outfitted with cameras or other sensors to inspect powerlines from above or below, providing detailed views as well as any faults discovered. They are very useful when investigating gridlines in troubled places, such as high voltage conditions.

Figure 1.1: Power Line Inspection using Helicopter *Image credit:* [37]

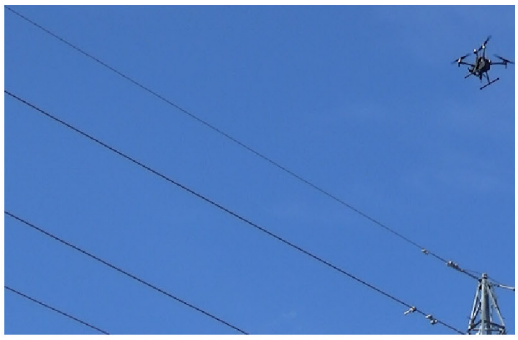


Figure 1.2: Power Line Inspection using Drone *Image credit:* [23]

Satellite imagery is becoming more widely available, and it can be a helpful resource for grid line assessment. Satellite photography, which uses computer vision and machine learning algorithms, provides a more faster and more precise way of analyzing grid lines than traditional approaches. These algorithms can be used to evaluate gridline pictures or videos for potential flaws or defects, such as wear or corrosion. This allows maintenance personnel to detect and rectify defects instantly, ensuring that gridlines are not jeopardized in a safe and effective manner. By tracking gridlines over time, satellite imaging provides proactive maintenance solutions that reduce unexpected outages or other issues, whereas satellite tracking provides another monitoring method that monitors gridlines continuously

to allow frequent repairs that prevent outages before they occur.

Figure 1.3: Power line view from a satellite image*Image credit:* [31]

1.1 Machine Learning and Satellite Imagery for Gridlines Inspection

In order to detect and assess faults and errors quickly along a gridline, machine learning algorithms are combined with satellite imagery. A more in-depth exploration of their roles in gridline assessment can be found here:

Image analysis is one of the primary applications of machine learning with satellite imagery. A collection of annotated photos showing various forms of flaws or issues on gridlines such as wear or damage can be used to train machine learning algorithms. Subsequently, algorithms can be utilized to analyze new gridline satellite photos and predict any flaws or abnormalities present within them. This allows maintenance employees to quickly discover any faults in gridlines that require repairs so that corrective actions may be taken and ensure their operation safely and efficiently. Predictive maintenance is another application of machine learning with satellite imagery. Training a machine learning model on images and data collected from sensors installed near gridlines allows one to predict when these might experience issues or require maintenance. This model can be applied to new satellite photos of gridlines to estimate their likelihood of faults or defects using attributes from its training dataset. This allows maintenance professionals to address potential problems before they become serious issues ensuring gridlines operate safely and efficiently.

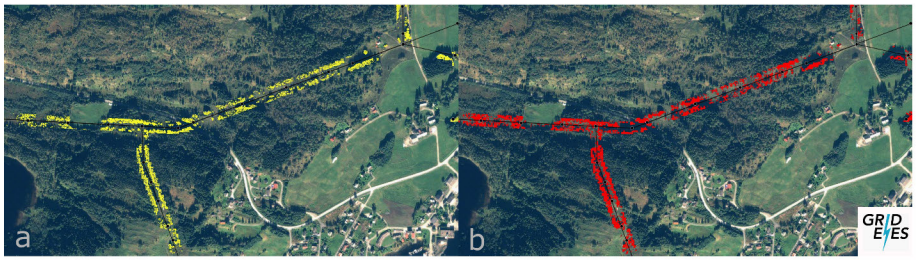


Figure 1.4: AI tool result of detection of a Power line from a satellite image *Image credit:* [31]

Machine learning algorithms can also detect abnormalities or strange patterns in satellite imagery that could indicate potential faults or defects, by training a machine learning model on an initial dataset of normal photos and then applying that model to detect anomalies in new gridline satellite images. This enables maintenance employees to promptly discover potential issues and take appropriate actions quickly in order to ensure gridlines continue operating safely and effectively. Overall, the combination of machine learning algorithms with satellite imagery provides a powerful tool for gridline inspection that efficiently detects any faults or defects quickly and precisely. This can help maintenance employees ensure that gridlines operate safely and effectively, maintaining continuous electricity distribution to customers.

1.2 Research Objectives

The goal of this research is to compare the performance of deep learning algorithms conducting Land Cover Semantic Segmentation. The performance of the two datasets, Potsdam and Vaihingen, will be compared. The dataset was acquired via the 2D Semantic Labeling Contest by the International Society for Photogrammetry and Remote Sensing. The background chapter provides a detailed description of the algorithms and datasets.

RQ 1 How do different deep neural architectures perform on land cover semantic segmentation?

RQ 2 How does adding infrared data affect the comparison of modern deep state of art of architecture?

This study is organized into five chapters: an Introduction chapter covers its motivation and goals; a Background chapter delves deeply into Semantic Segmentation and deep learning algorithms; the Related Work chapter offers a thorough review of its literature; experiments are described thoroughly in an Experimental Setup chapter, followed by Results chapter with illustrations depicting experimental results, Discussion section discussing them further, and ultimately concluding results being provided as reported in Conclusions chapter.

Chapter 2

Background

This chapter presents a comprehensive introduction to Gridlines Inspection, Satellite Imagery, Land Cover Semantic Segmentation, and Deep Learning.

2.1 Gridlines Inspection

The practice of inspecting power lines, also known as gridlines, that transmit electricity from power plants to homes and businesses is known as gridline inspection. This is done to detect and assess potential flaws. Gridlines are an essential component of modern society's infrastructure, and they must run efficiently and safely. Their visual inspection consists mostly of visually inspecting lines and checking for wear and deterioration. Sensors and other technologies are also utilized to keep track on the lines' status. Sensors, drones, robotics, and computer vision algorithms are a few examples.

Figure 2.1: Manual grid line inspection *Image credit:* [20]

Maintenance staff used to physically inspect electrical cables for signs of corrosion or wear. The lines were visually inspected for flaws such as cracks, bends, and other irregularities. To obtain a better view at the lines, maintenance workers may use binoculars

and telescopes. Maintenance workers were frequently expected to work in hazardous environments or at heights, which made their occupations dangerous and time-consuming.

Drones are increasingly being used in grid line inspection since they are a quick and safe technique of inspecting gridlines and spotting possible faults or deficiencies. Drones, or autonomous tiny aircraft outfitted with sensors or cameras, can be used to monitor gridlines from a safe distance. You can see the gridlines from above and notice any faults such as wear and damage. Drones can monitor big gridlines or gridlines in difficult-to-reach areas such as subterranean or high-voltage situations. Drones are an excellent tool to inspect grid lines in regions that would be difficult or perhaps dangerous for maintenance staff to access. Drones, for example, can be used to survey gridlines in challenging situations such as high mountains or dense forest. Gridlines can be maintained safely and effectively without placing maintenance employees in danger.

2.2 Satellite Imagery

Satellite imagery is a sort of remote sensing data gathered by Earth-orbiting satellites equipped with cameras or other sensors that snap photographs or collect data about the Earth's surface and atmosphere before relaying it back to Earth for analysis and usage in various ways.

Satellite imagery can help gridline inspection by remotely monitoring and analyzing their condition. Through techniques such as image analysis, predictive maintenance and anomaly detection, satellite imagery can detect potential issues or defects with gridlines. A machine learning model trained on labeled images or sensor data installed near gridlines could then analyze satellite images taken at different points along the line to predict defects or issues more rapidly and accurately - helping maintenance personnel quickly recognize potential issues and take corrective actions quickly ensuring safe operation of gridlines.

2.3 Land Cover Semantic Segmentation

Land cover semantic segmentation is the practice of classifying various types of land covers such as vegetation, water bodies, urban areas and bare land using satellite or aerial images. Images are divided into segments or classes to represent these different forms of coverage - each segment representing one type of cover. Land cover semantic segmentation is used for various applications, including:

- **Environmental monitoring:**
  It can be used to track changes in land cover over time. For example, urbanization and vegetation loss as a result of deforestation. This is helpful in assessing the environmental impact of human activities and identifying places that are at risk of degradation.

- **Land use planning:**
  Land cover segmentation, which gives information on the types and locations of various land cover categories, can help influence land use planning. It can assist

planners in making sound decisions about land use, such as agriculture, urbanization, and conservation.

- **Disaster management:**
  Land cover segmentation is an effective method for assessing the impact of natural disasters such as wildfires and floods on land cover. This can help emergency responders discover sites in danger and in need of assistance. It can also aid in rehabilitation and reconstruction efforts.

- **Natural resource management:**
  It can detect and map natural resources such as lakes and forests, as well as monitor their condition and utilization. This can be used to encourage resource management and sustainable land usage.

2.4 Deep Learning

Deep learning is a subset of machine learning derived from brain function and structure. [22] Artificial neural networks are used to recognize patterns within data and make decisions based on these findings. Deep learning algorithms are named such due to being composed of layers of artificial nodes or neurons that receive and process data before passing it along to another layer; their output serves as input for further computation allowing algorithms to recognize subtler patterns within it and make decisions accordingly.

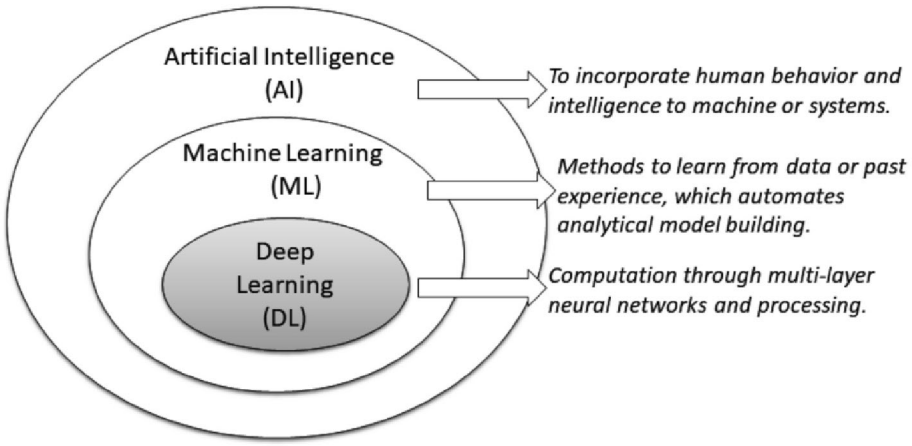


Figure 2.2: Deep Learning Illustration *Image credit:* [22]

Deep learning systems have the ability to outperform humans in categories like language translation, speech recognition and image recognition. Trained on large datasets, deep learning algorithms often outshone traditional machine-learning techniques - often outpacing them entirely! Deep learning algorithms are widely utilized for computer vision applications as well as audio recognition and natural language processing tasks in applications like healthcare banking and transportation industries.

- **Convolutional neural networks (CNNs):**
  Image categorization and object detection are common applications for these techniques. They excel in analyzing and interpreting visual data and are commonly employed in computer vision and image recognition applications.

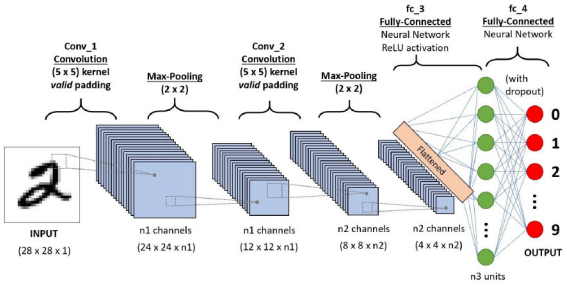


Figure 2.3: Architecture of a CNN. *Image credit:* [36]

- **Recurrent neural networks (RNNs):**
  These algorithms are designed to handle sequential data, such as time series or natural language. They are often employed in the fields of language translation and speech recognition.

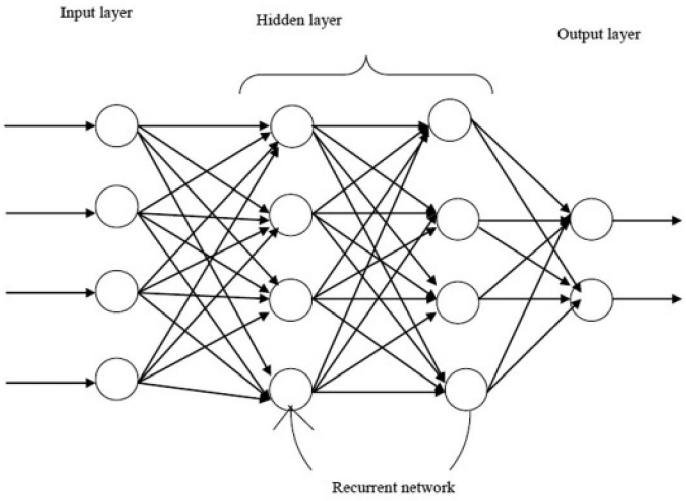


Figure 2.4: Architecture of a RNN classifier. *Image credit:* [28]

- **Autoencoders:**
  These techniques are used for a variety of purposes, including data compression and feature extraction. They consist of an encoder and a decoder, which work together to learn a compact representation of the input data.

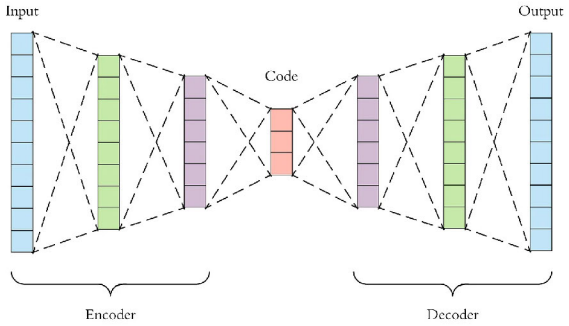


Figure 2.5: Architecture of an Autoencoder. *Image credit:* [4]

- **Generative adversarial networks (GANs):**
  These algorithms, among other things, are employed for image production and data augmentation. They are composed of two neural networks, a generator and a discriminator, which work together to generate new data that is similar to the training dataset.

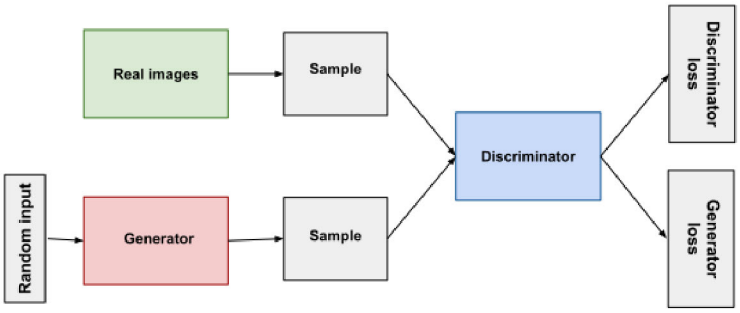


Figure 2.6: Overview of a GAN. *Image credit:* [35]

- **Self-organizing maps (SOMs):**
  These algorithms are used for a variety of purposes, including data visualization and clustering. They are composed of a grid of neurons arranged in a two-dimensional map that is used to organize and group data into similarity-based groups.

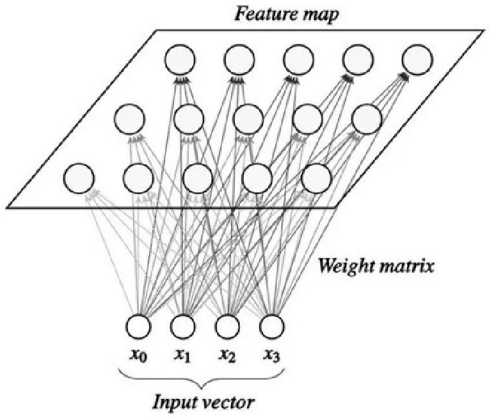


Figure 2.7: Overview of a SOM. *Image credit:* [12]

2.5 Dataset Description

The International Society for Photogrammetry and Remote Sensing (ISPRS) is an international organization that promotes photogrammetry, remote sensing, and related spatial information sciences. The 2D Semantic tagging Contest entails the development and testing of algorithms for semantic tagging on aerial photography. Semantic Labeling is the process of assigning a class to each pixel based on the items or attributes in a scene. A semantic labeling system, for example, might classify pixels as belonging to one of several classifications, such as "building", "road", "vegetation" or "water." This study's dataset is based on two different places, namely Potsdam and Vaihingen.

2.5.1 Potsdam

The Potsdam dataset is part of the International Society for Photogrammetry and Remote Sensing's 2D Semantic Labeling Contest. It includes high-resolution aerial photographs of Potsdam, Germany. These photos have a 5cm per pixel resolution and include Red, Green, Blue, and Near Infrared channels. This dataset is used for semantic segmentation tasks. These include categorizing every pixel in an image, such as "building", "tree", "car", "road", and so on. The collection contains aerial laser scan data, which is used to generate a Digital Surface Model, or DSM, which represents the height of structures at each site. This includes buildings and trees. The Potsdam dataset contains ground truth data that serves as semantic labels to each pixel. This is essential for the training and evaluation semantic segmentation models. The dataset covers an area of about 6 square kilometers. It is divided into 38 tiles measuring 6000x6000 pixels each.

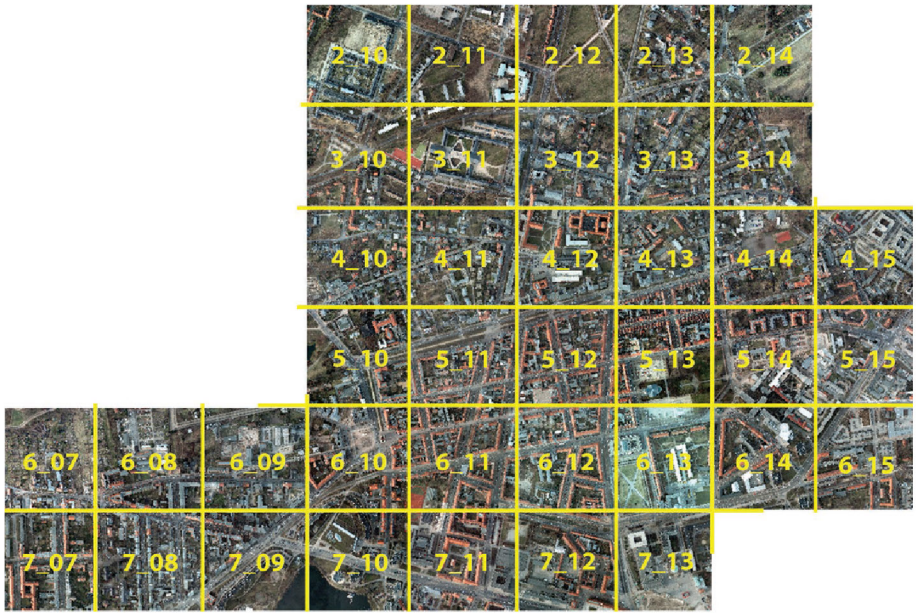


Figure 2.8: Overview of Potsdam dataset. *Image credit: ISPRS* [32]

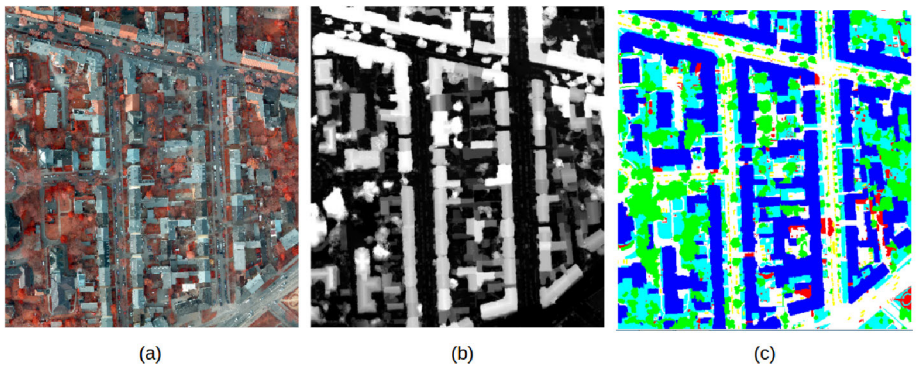


Figure 2.9: Sample patches of the semantic object classification contest with (a) true orthophoto, (b) DSM, and (c) ground truth. *Image credit: ISPRS* [32]

2.5.2 Vaihingen

The dataset was created in Vaihingen by the German Association of Photogrammetry and Remote Sensing. (DGPF) for testing digital aerial cameras. The test dataset is divided into three major areas: "Inner City", "High Riser", "Residential Area" each of which showcases different urban structures and settings. A larger "Roads", test site is also available to evaluate urban road extraction methods. The dataset contains digital aerial images, orienta-tion parameters, and Airborne Laserscanner Data from the Leica ALS50 System. Multiple images are available for each test area, with an average density of 4 points/m2 in each strip. Since January 2013, a TrueOrtho image, a DSM (digital surface model) from image matching and a DSM (digital surface model from image comparison), as well as the original point cloud, are also provided.[32]

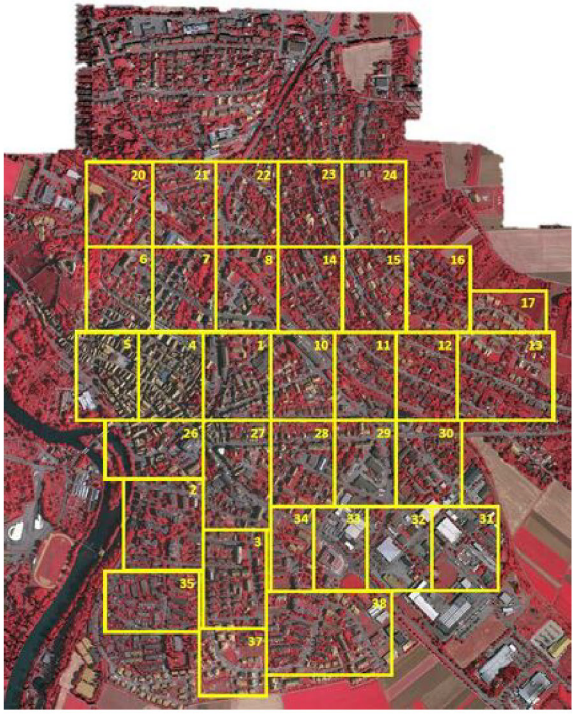


Figure 2.10: Overview of Vaihingen dataset. *Image credit: ISPRS* [32]

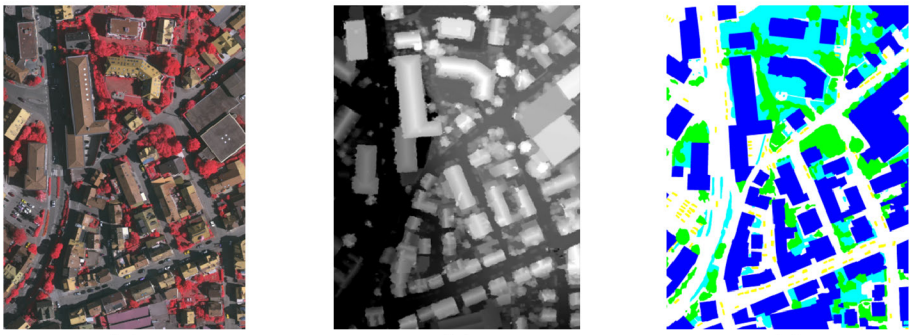


Figure 2.11: Sample patches of the semantic object classification contest with (a) true orthophoto, (b) DSM, and (c) ground truth. *Image credit: ISPRS* [32]

2.6 Machine Learning Algorithms

This study aims to evaluate how accurately machine learning algorithms can perform semantic segmentation. Semantic segmentation is used to classify land cover types. The study has used the following algorithms.

- ResNet50
- ResNet101
- FC-ResNet50
- VGG-19

- UNetFormer

2.6.1 ResNet-50

ResNet50 was created by Microsoft researchers to be a component of the ResNet family of models, which are renowned for their complexity and capacity to efficiently learn from enormous datasets. It was created by Microsoft researchers to be a component of the ResNet family of models, which are renowned for their complexity and capacity to efficiently learn from enormous datasets.It consists of two independent parts, an encoder and a decoder, which cooperate to carry out image recognition.[3] Here is a more thorough breakdown of the encoder and decoder components of ResNet50:

An Encoder employs convolutional layers applied successively to an input image in order to extract its low-level features. Batch normalization layers may also be included within these convolutional layers in order to improve model stabilization and performance, with feature maps then downsampled using pooling layers after passing through many convolutional layers - this reduces both feature map size and computational load on the model; each convolutional layer block making up ResNet50 encoder includes numerous convolutional layers; pooling layers downsample their feature maps created by these layers while following each convolutional block is followed by numerous convolutional layers that create feature maps which is followed by downsampling pooling layers that downsampled feature maps created from convolutional layers - size reduction is reduced while computational load on model reduced accordingly. As they progress through the convolutional layers, the feature maps become more abstract and capture intricate elements of input images. The encoder component is made up of a collection of feature map that captures the key aspects of an input image.

The ResNet50 Decoder first upsamples the feature maps that the encoder has created using a transposed layer of convolutional. The feature maps are enlarged and combined with low-level features the encoder learned. The final output is generated by passing upsampled features maps through convolutional layers that are trained to combine high-level and lower-level information. The decoder part of the model produces a set of feature map that is used to make the final prediction. This prediction can be either a label for a class or a series of bounding box boundaries to detect objects, depending on the task that the model has been trained to perform.

2.6.2 ResNet-101

ImageNet was used to train ResNet101 (a neural network). This is a huge image collection with labels for 1,000 different classes. It was designed by Microsoft researchers as a ResNet model, which is noted for its depth and capacity to learn from massive datasets.[3]

ResNet101, a deep CNN model with 101 layers, is one of the more complicated ResNet models. Because of its additional layers, it can extract more complicated features from input data. This can boost its performance in tasks like object detection and image classification. It consists of two distinct components that work together to conduct image

recognition: an encoder (which translates images) and a decoder (which interprets them).

It is the encoder's responsibility to extract features from the input image. By using convolutional layers and pooling, the image is made smaller and more complicated. The encoder extracts data from the input image that the decoder uses to classify it. The features are combined so that the neural network can predict the class of an input image. The architecture of the encoders and decoders will vary depending on how ResNet101 has been implemented. In general, however, the encoder includes a few convolutional layers and a pooling layer, while the decoder contains a few fully connected layers.

2.6.3 FC-ResNet50

FC-ResNet50 is similar to ResNet50 but without the FC layer. This means that its output is the feature map produced by the decoder portion of the network rather than final predictions. This output captures high-level features of an input image and can be used as input into other models or applications such as semantic segmentation which assigns class labels for every pixel in an image. FC-ResNet50 [2] comprises several layers such as convolutional layers, pooling layers and batch normalization layers; here is an in-depth breakdown of its architecture here:

An input image is processed through several convolutional layers that learn to extract low-level features from it, including optional normalization layers that help stabilize and enhance performance. After being passed through multiple convolutional layers, feature maps are downsampled using pooling layers in order to reduce both map size and computational load. By doing this, both the size of feature maps and the computational load on the model can be reduced significantly. The FC-ResNet50 encoder comprises multiple convolutional layer blocks containing multiple convolutional layers for maximum reduction in size and computation load. Pooling layers, which downsample feature maps created by convolutional layers, follow each block of convolutional layers. As they move through each of these blocks of convolutional layers, their feature maps become increasingly abstract and capture more complex aspects of an input image. An encoder component of the model's output comprises of feature maps that capture key aspects of an input image, followed by upsampling with transposed convolutional layers by FC-ResNet50 decoder. As a result, feature maps expand in size, merging low-level features that the encoder has learned with high-level features that the modeler has learnt. Finally, the output of the model is produced by passing upsampled feature maps through convolutional layers trained to combine high-level and low-level information into one coherent output. FC-ResNet50 decoder produces feature maps which highlight key aspects of an image, for use in semantic segmentation - for instance assigning class labels to pixels within it - or other tasks as input for models or programs. These feature maps may then be utilized as input into other models or used directly as part of these tasks.

ResNet50's overall architecture is similar to FC-ResNet50, but without FC layers at the top. This indicates that feature maps generated by ResNet50's decoder component are used as output rather than as predictions; when performing tasks such as semantic segmentation where class labels must be assigned for every pixel in an image, feature maps

generated by decoder can help capture high-level features from an input image.

2.6.4 VGG-19

The VGG-19 model is a convolutional neural network (CNN) architecture developed by researchers at the University of Oxford. It was trained on the ImageNet dataset, which consists of a large number of images categorized into 1000 different classes. The VGG model family, to which VGG-19 belongs, is well-known for its effectiveness in image classification tasks. The VGG-19 architecture is composed of various layers, including convolutional layers, pooling layers, and fully connected (FC) layers. [1]

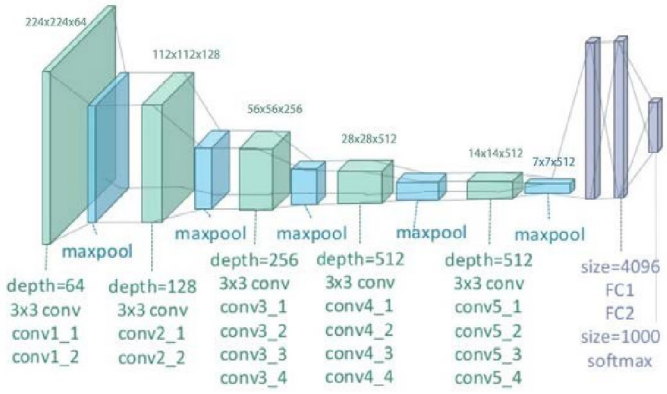


Figure 2.12: Architecture of VGG *Image credit:* [11]

The process involves applying convolutional layers successively on an input image so they can learn to recognize low-level features, followed by pooling layers to reduce feature map sizes created by convolutional layers. At each block of convolutional and pooling levels, additional convolutional layers and pooling levels will be added, increasing as feature maps progress through the network - this allows the model to learn complex aspects of an image while giving predictions based on features learned.

VGG-19 is a relatively simple architecture consisting of convolutional, pooling and FC layers. It is often used in image classification tasks and performs well across several benchmarks.

2.6.5 UNetFormer

U-Net is a convolutional neural system (CNN) that has been modified to perform efficient semantic segmentation for remote sensing data of urban scenes. This variant is also known as a "UNet-like Transformer." Semantic segmentation assigns class labels to each pixel of an image using the features that the model has acquired. Remote sensing urban scene images are used to describe images of cities captured by satellites or aircraft. They can be used in a number of ways, such as identifying buildings and classifying land uses.

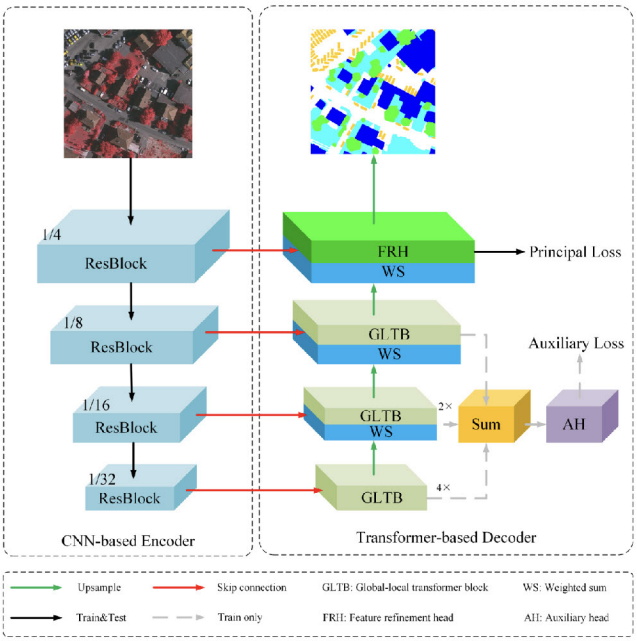


Figure 2.13: Architecture of UNetFormer *Image credit:* [30]

The UNetFormer uses a hybrid architecture consisting of a Transformer decoder and CNN encoder. We create a global local Transformer block (GLTB), which is used to build the encoder, and we use ResNet18 for the decoder. The GLTB suggests creates a global-local attention system with an attentional local branch and a convolutional global branch that captures both global and local contexts of visual perception. This is in contrast to the five self-attention blocks used by the regular Transformer.[30]

2.7 Evaluation Criteria And Comparisons of Algorithms

2.7.1 Evaluating the performance of models

Jaccard coefficient and validation loss have been used for evaluating the performances of models.

The Jaccard coefficient is also known as Intersection over Union. It is a statistical tool used to evaluate model performance for tasks like picture segmentation or object detection. It calculates the ratio between the intersection of two set to their union in order to determine the similarity of their sets. In the context of a validation set, the Jaccard coefficient measures the accuracy of the predictions of the model in comparison with the ground truth labels. The Jaccard coefficient can be computed in image segmentation by comparing the predicted masks with the real masks of the validation set.It quantifies the degree of overlap or agreement between the expected and true masks.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{2.1}$$

- A∩B represents the intersection of sets A and B (the elements common to both sets).

- |A∩B| represents the cardinality (number of elements) of the intersection.
- A∪B represents the union of sets A and B (all unique elements from both sets).
- |A∪B| represents the union of sets A and B (all unique elements from both sets).

Validation loss measures the performance of machine learning models during their validation phase. It represents the model's average loss or error on the validation dataset. The validation loss can be estimated by applying the model gained during training to a second dataset (known as the validation dataset) that was not utilized during training. The model's ability to generalize to previously unseen data is evaluated. When the validation loss is low, the model's validation dataset performs well.

$$\text{val\_loss} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] \tag{2.2}$$

- N is the total number of validation samples
- y_i represents the ground truth (actual) label for the i-th validation sample (either 0 or 1)
- $\hat{y}_i$ represents the predicted probability of the positive class (between 0 and 1) by the model for the i-th validation sample

2.7.2 Comparisons of Algorithms

When comparing algorithms, the t-test have been used to determine whether there is a statistically significant difference in their performance. Typically, this is accomplished by measuring a certain metric or performance indicator (such as accuracy, error rate, or execution time) for each algorithm on a specified dataset or set of tasks.

The unpaired t-test is a statistical test used to compare the means of two independent groups. The formula for calculating the t-value in an unpaired t-test is given by:

$$t = \frac{\text{mean}_1 - \text{mean}_2}{\sqrt{\frac{\text{variance}_1}{\text{sample size}_1} + \frac{\text{variance}_2}{\text{sample size}_2}}}$$

where mean_1 and mean_2 are the means of the two groups, variance_1 and variance_2 are the variances of the two groups, and sample size_1 and sample size_2 are the sample sizes of the two groups.

The calculated t-value can then be compared to the critical t-value from the t-distribution to determine if the difference between the means is statistically significant.

Chapter 3

Related Work

This chapter provides a fair review of the literature on image analysis of satellite data for power line inspection.

Regular inspections and timely maintenance of gridlines are critical for reducing resource input, keeping prices down, and swiftly restoring power. Traditional techniques of inspection, including foot patrol and helicopter-assisted surveys, have typically been used, which are cumbersome, expensive, and potentially hazardous. Recently, drone technologies have been introduced for inspections which are better than the conventional techniques but it introduces new issues related to operations, flying time, and mission planning. If we could conduct grid lines surveillance using satellite imagery this could be a substantial contribution. There are a few studies that have addressed the application of performing grid inspection using image analysis techniques on satellite data.

3.1 Literature Survey

A thorough literature survey was carried out to discover potentially relevant studies. The databases incorporated in the primary literature survey consisted of ACM Digital Library, IEEE Xplore and Google Scholar. Since, the concept of using image analysis and deep learning model on satelite data is relatively new, so those studies were particulary considered which were conducted recently in past seven years. Additionaly, for the sake of credibility studies having decent number of citation were given preference. Keywords were used to narrow down the search and find the most relevant articles.

3.1.1 Image Analysis, Satellite Data and Grid Lines Inspection

This section presents a number of studies that have investigated the application of power line inspection using image analysis techniques on satellite data.

Table 3.1: Summary of Database Search Results-1

Database	**Keywords**	**No. of Matches**
ACM Digital Library	Machine Learning, Satellite Images, Power line Inspection	1
IEEE Xplore	Machine Learning, Satellite Images, Power line Inspection	2
Google Scholar	Machine Learning, Satellite Images, Power line Inspection	6

A Review of Vegetation Encroachment Detection in Power Transmission Lines using Optical Sensing Satellite Imagery

This study by F. M. E. Haroun et al. focuses on monitoring vegetation encroachment along the power lines corridor using satellite images.[16] They state the importance of vegetation monitoring alongside power lines. The most commonly used strategies for detecting vegetation using satellite pictures are the object-based detection method, the Vegetation Index-based method, Stereo matching-based techniques, and others.

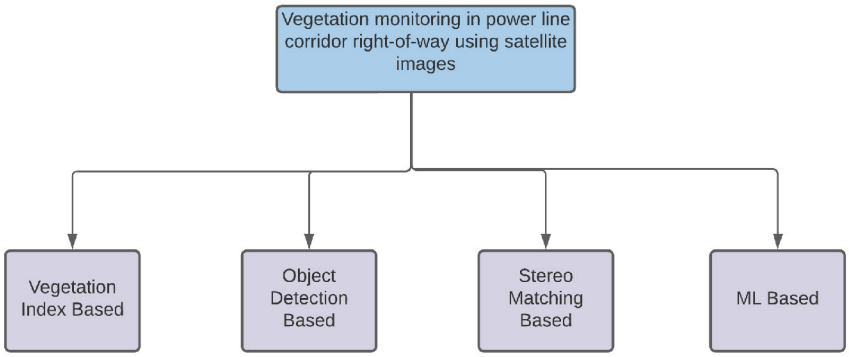


Figure 3.1: Techniques for Vegetation Encroachment Detection *Image credit:* [16]

The current detection approaches rely on manually setting threshold values, which makes it impossible to detect vegetation in varying resolutions and testing situations in a dynamic manner. Deep Learning, on the other hand, can be quite effective in detecting several objects in satellite photos with high classification accuracy. This opens the door to

a possible option for monitoring vegetation encroachment in power line corridors.

An intelligent identification and acquisition system for UAVs based on edge computing using in the transmission line inspection

This research proposes an intelligent acquisition system for UAVs used in transmission line inspection, designed to automate the image collection process and reduce manual PTZ camera control reliance.[15] The system comprises of front-end edge computing detection module using Single Shot Multibox Detector algorithm as well as Pan Tilt Zoom camera control module. Experimental results demonstrate substantial improvements in recognition accuracy, with an overall success rate of 73% for identifying transmission line equipment. By integrating cutting-edge technologies such as edge computing and the SSD algorithm, the system streamlines transmission line inspections, enhancing their efficiency and precision. The findings contribute to the field of UAV-based inspection systems, highlighting the potential of automated image acquisition techniques in revolutionizing transmission line inspections.

Automated Power Lines Vegetation Monitoring Using High-Resolution Satellite Imagery

This research by M. Gazzea et al. focused on vegetation survelliance using satellite images. [26] The research was carried out a power distribution system operator (DSO) in Norway's western region. It highlights that whenever vegetation gets in the way of electrical lines, it poses a threat to people's safety, the economy, and the envi-ronment. LiDAR scans performed by helicopters or drones are commonly used for vegetation monitoring. If a large transmission or distribution business performs Li-DAR-based line monitoring, it is usually done seldom, once every 5 to 10 years. The cost of satellite imaging has decreased dramatically in recent years as launching prices have decreased and the number of satellites and mini-satellites has increased. Using high-resolution satellite photos, this research presents a system for monitoring vegeta-tion near power lines. It's a hybrid of a supervised machine learning method with a deep unsupervised architecture. A geolocation map for vegetation-related risks near power lines is the result of the proposed approach.

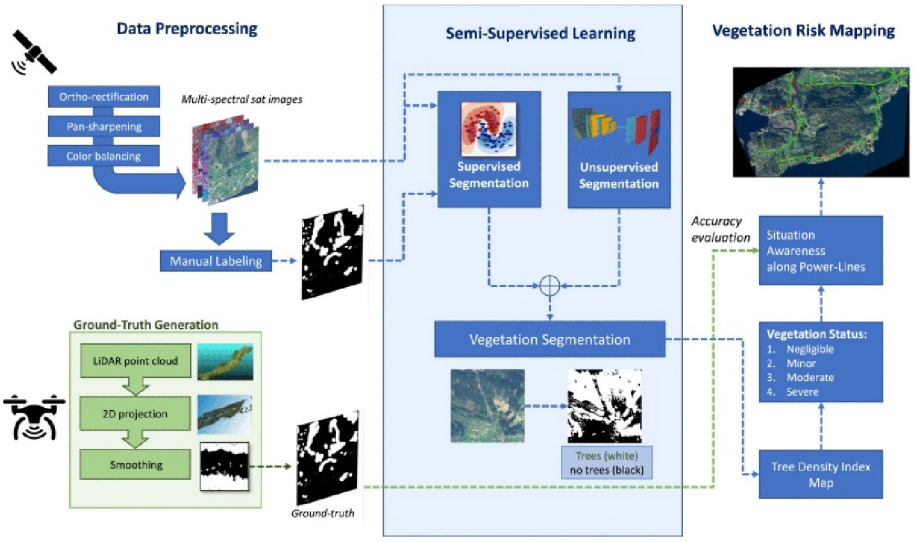


Figure 3.2: Overview of Proposed Algorithm for Vegetation Monitoring *Image credit:* [26]

The framework can be separated in different blocks: Data Pre-processing Block where labelling of training and testing of data is performed. Semi-supervised consisting of supervided and unsupervised segmentation. Supervised Image Segmentation Block extracting texture features, spectral features, and a Gaussian kernel. A fully convolutional neural network (FCN) extracts features, and a superpixel refinement method self-trains the model in the Unsupervised Segmentation Block. They evaluated the picture segmentation method for a power grid in western Norway using aerial LiDAR. According to preliminary findings, this method can correctly identify vegetation danger zones in this area with an accuracy of 84 percent. In 92 per-cent of situations, no-risk areas are appropriately identified. These preliminary find-ings show that this satellite-based architecture has a lot of promise. approach.

Aerial video inspection of greek power lines structures using machine learning techniques

Pioneering research investigates the use of unmanned aerial vehicles (UAVs) equipped with camera sensors, coupled with deep learning algorithms, to automate and enhance power line inspections.[29] Key attention is given to the fine structure of power lines, with a focus on an effective deep learning architecture capable of segmenting these thin structures and reducing background noise. The study found that Deep Neural Networks (DNNs) using dilated convolutions perform impressively, particularly the D-LinkNet architecture. Two public datasets, representative of urban and mountainous settings, were utilized for training, and tests were conducted on videos from real UAV flights in challenging environments. To overcome the limitation of small training datasets, data augmentation techniques were used. The study underscores the significance of dilated convolutions in maintaining image resolution, which is essential for power line structures. D-LinkNet outperformed other architectures across different datasets, showing higher precision and recall. This research showcases the potential of UAVs and deep learning in power line inspection, with promising implications for real-time fault detection and alerts in future research, thereby augmenting

the safety and efficiency of power line inspections.

Automatic autonomous vision-based power line inspection: A review of current status and the potential role of deep learning

V.N. Nguyen et al. examine existing strategies for automated electrical line inspection and the possible use of deep learning.[10] The benefits and cons of helicopter-assisted inspection, climbing robots inspection, foot patrol inspection, automated helicopter-assisted inspection, and unmanned aerial vehicle (UAV) inspection are examined.

The main tasks of an inspection are mapping and inspecting power line components, monitoring vegetation encroachment, detecting and measuring icing, and monitoring disasters. Synthetic aperture radar pictures, optical aerial images, optical satellite images, ultraviolet images, airborne laser scanner data, thermal images, land-based mobile mapping data, and unmanned aerial vehicle data are all conceivable inspection data sources. They address related studies briefly and critique them, with a particular emphasis on navigation and inspection using UAV and optical imaging. They discovered that no completely automated autonomous vision-based inspection system capable of detecting a wide range of faults has been developed successfully. Deep learning vision-based algorithms for navigation and inspection, as well as their difficulties and potential solutions, are reviewed.

The authors propose to combine GPS waypoints-based, pole-detection-based and power-line detection-based navigation with an autopilot in order to create a hybrid system. To detect faults with varying shapes, sizes, and conditions (e.g. cracked insulators and rotten poles), (i) use multistage objects detection, classification, segmentation pipelines. (ii), use contextual information to enhance fault detection performance (e.g. eliminating invalid faults); and (iii), combine multiple data sources to detect complex faults (e.g. cracked insulators and rotten Poles).

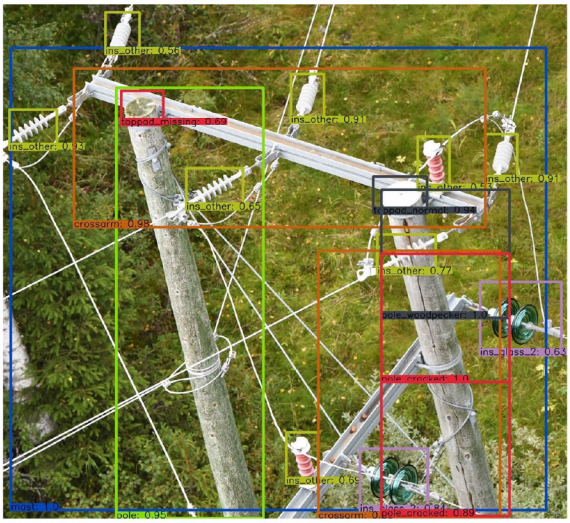


Figure 3.3: Detection of different components of a utility pole *Image credit:* [10]

The Future Application of Transmission Line Automatic Monitoring and Deep Learning Technology Based on Vision

The paper under review offers an intensive exploration of automatic vision monitoring technology as an efficient and safer alternative to traditional manual and helicopter-assisted methods for monitoring power transmission lines and distribution grids.[14] The authors dissect existing automatic vision monitoring systems, highlighting their strengths and weaknesses. In response to the identified shortcomings, a fresh scheme for automated transmission line vision monitoring is proposed, grounded on optical images and deep learning for data analysis. Deep learning algorithms such as SSD, fast R-CNN, YOLO, R-FCN, and ResNet are highlighted for their impressive target detection and classification capabilities. Furthermore, DPN and Mask R-CNN are pointed out for their semantic segmentation capabilities when coupled with traditional background removal techniques. Despite the potential of deep learning in this domain, the paper acknowledges challenges such as scarcity of training data, ineffective detection of small targets, and complexities in detecting transmission lines in complex backgrounds. To combat these, the authors suggest image processing techniques, localizing region of interest in images, and a comprehensive scheme of multiple line detection methods, respectively. Ultimately, this study paints a hopeful future for deep learning-based automatic vision monitoring of power transmission lines while providing solutions to potential roadblocks.

Vision-based autonomous navigation approach for unmanned aerial vehicle transmission-line inspection

The research presents an autonomous navigation approach for Unmanned Aerial Vehicles (UAVs) inspecting electricity lines.[8] A perspective navigation model is used in the strategy to improve 3D direction perception and safety during inspection. For transmission tower detection, the system employs a Faster Region-Based Convolutional Neural Network (Faster R-CNN) in conjunction with Kernelized Correlation Filters (KCF). Fully Convolutional Networks (FCNs) are also utilized for transmission line extraction, which is critical for estimating flight direction via the vanishing point (VP). The method also considers instances in which a VP is not present. The suggested method's usefulness is demonstrated by experimental findings in a practical context, marking the first time a transmission tower-based navigation approach has been presented and implemented. However, the authors note that keeping a safe distance from transmission lines and traversing more complex situations may demand the usage of GPS. Future work will focus on establishing a transmission-line tracking algorithm and an online transmission-line fault diagnosis system.

A survey of intelligent transmission line inspection based on unmanned aerial vehicle

The article provides a comprehensive survey of intelligent transmission line inspection using unmanned aerial vehicles (UAVs).[33] It outlines the development of this practice, its current processes, and potential challenges and future solutions. The process involves integrated navigation via differential GPS information from an RTK(Real-Time Kinematic) base station to select and calculate patrol points, followed by path planning, trajectory

tracking, and finally, fault detection and diagnosis using images captured by the UAV and intelligent algorithms. The survey identified several challenges: improving autonomous navigation technology and obstacle avoidance, enhancing the stability and accuracy of trajectory tracking, expanding fault detection beyond insulators, implementing advanced image processing technology for small equipment detection, handling strong electromagnetic interference around high voltage lines, and improving UAV flight endurance. Despite the challenges, the study concludes that intelligent power inspection is a long-term development process requiring multi-disciplinary cooperation. The increasing application of UAVs in this area is expected to remain a research hotspot for a significant period.

Review of data analysis in vision inspection of power lines with an in-depth discussion of deep learning technology

The article provides a thorough review of current literature and challenges in the field of power line inspection data analysis using unmanned aerial vehicles (UAVs).[17] The main data sources for this process include image data and non-image data (mainly airborne laser scanner (ALS) data). The paper focuses more on image data, particularly visible images due to their prevalence, practicality, flexibility, low-cost, and high-quality acquisition. The authors gave an in-depth overview of deep learning approaches and their application in this industry, including the use of current frameworks, extracting deep features, network cascading, resolving data scarcity, and domain knowledge-based enhancements. The paper also contains a proposed deep learning-based system for inspection data analysis, which covers data preparation, component detection, fault diagnosis, and model training and optimization. The authors finish the work by outlining future research directions, such as dealing with data quality issues, small item detection, embedded applications, and establishing evaluation baselines.

3.1.2 Dataset and Land Cover Semantic Segmentation

Table 3.2: Summary of Database Search Results-2

Database	Keywords	No. of Matches
ACM Digital Library	ISPRS, land cover semantic segmentation	1
IEEE Xplore	ISPRS, land cover semantic segmentation	4
Google Scholar	ISPRS, land cover semantic segmentation	7

Semantic Segmentation in Aerial Images Using Class-Aware Unsupervised Domain Adaptation

This research presents an unsupervised domain adaptation (UDA) framework for deep neural network-based semantic segmentation of aerial imagery.[21] By learning class-aware distribution discrepancies between the source and target domains, this innovative approach overcomes the issue of domain shift, a typical problem in aerial photographs due to major visual appearance changes. On the target domain, entropy minimization is also used to generate high-confidence predictions. In comparison to prior methods, the unique strategy, which incorporates class-aware distribution alignment and entropy minimization, is non-adversarial, simpler to train, and requires less parameters for training.

Experimental results demonstrate that this method outperforms the current state-of-the-art methods on the ISPRS segmentation challenge dataset. Importantly, the proposed method demonstrates significant improvements in the segmentation of hard-to-detect objects. The paper indicates that this unique strategy of class-wise distribution alignment paired with entropy minimization increases domain adaption performance for semantic segmentation in aerial photos significantly. The authors propose that this work be extended to semi-supervised learning scenarios using a small set of labeled data.

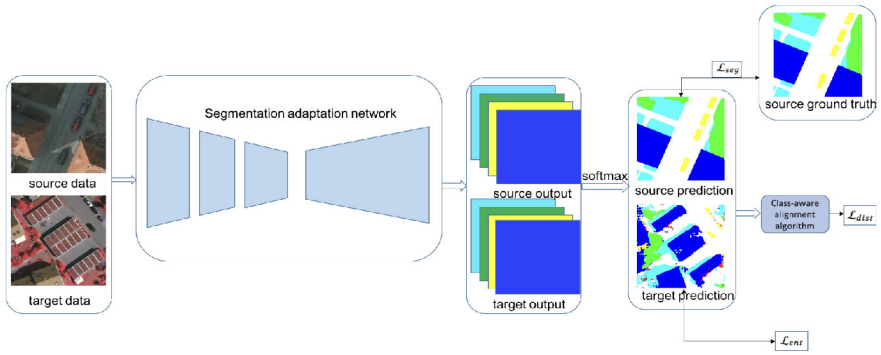


Figure 3.4: Proposed network architecture *Image credit:* [21]

A Comparison of Deep Learning Architectures for Semantic Mapping of Very High Resolution Images

The article conducts an in-depth comparison of advanced deep learning segmentation architectures used for semantic mapping in high-resolution aerial images.[9] The authors specifically test and analyze models such as PSPNet, GCN, DUC, FCN, U-Net, and SegNet using the ISPRS Potsdam dataset. These models were evaluated based on their ability to conduct semantic segmentation of the data, with image patches and a series of augmentations applied to the dataset to ensure accurate analysis. The AdaDelta optimizer and a 2D cross-entropy loss function with median frequency balancing were used for model training. The results showed that the DUC model achieved the best mean F1 score of 88.2% and mIoU of 79.3%, though performance was closely followed by the PSPResNet50 and SegNet-VGG19 models. The authors concluded that models using dilated convolution such as DUC are promising for handling multi-scale and large objects, while models using global average pooling methods like PSP could be beneficial for detecting smaller objects. The study was the first of its type, concentrating on distant sensing semantic mapping with a variety of leading deep learning architectures from the most recent Pascal VOC competition.

CTMFNet: CNN and Transformer Multiscale Fusion Network of Remote Sensing Urban Scene Imagery

The research investigates the possibilities of Convolutional Neural Networks (CNN) and Transformers in the semantic segmentation of remotely sensed urban scene photos.[34] The researchers suggest a CNN and Transformer Multiscale Fusion Network (CTMFNet) with a dual backbone attention fusion module (DAFM) for combining local and global context information, as well as a multilayer dense connection network (MDCN) to bridge the semantic gap between scales. Their experiments using the International Society of Photogrammetry and Remote Sensing (ISPRS) Vaihingen and ISPRS Potsdam datasets show that their method outperforms current methods. The suggested method, however, has only been validated for semantic segmentation of remotely sensed urban scene images, with future enhancement plans to accommodate more distant sensing vision tasks.

Aerial Image Semantic Segmentation Using Spatial and Channel Attention

In the study, A new land cover categorization approach is proposed that makes use of spatial and channel attention inside an Encoder-Decoder network structure.[13] This technique employs Deep Convolutional Neural Networks (DCNNs), ResNet_v2-101 as the feature extractor, Atrous Spatial Pyramid Attention (ASPA) as the context encoder, and a Spatial Attention module for decoding. The ISPRS Potsdam 2D-Semantic Segmentation Challenge Dataset was used to test the approach, which contains 38 hand-annotated aerial photos, 32 for training and 6 for testing. Subimages for training and testing were extracted using data augmentation techniques. The performance metrics revealed that the proposed Spatial and Channel Atrous Spatial Pyramid Attention Network (SC-ASPA-net) outpaced the established Deeplab_v3+ algorithm, exhibiting a 5% enhancement in mean Intersection over Union (mIoU).

Semantic Segmentation of Aerial Images With Shuffling Convolutional Neural Networks

In the research, a novel method of performing semantic segmentation on aerial imagery using shuffling Convolutional Neural Networks (CNNs) is proposed.[6] The paper introduces two versions of the Shuffling CNNs (SCNNs): Naive-SCNN and Deeper-SCNN, both proficient at detecting small objects. A Field-of-View (FoV) enhancement technique, suitable for various networks, is proposed to improve predictions. The study leverages ISPRS Vaihingen and Potsdam datasets to evaluate the proposed models, with performance gauged on F1-scores across various categories. Results indicate that SCNNs significantly surpass baseline models such as RDM and RDM-ASPP, chiefly by learning to upsample, thereby achieving smoother and more accurate semantic segmentation. Particularly, SCNNs outperform RDMs by more than 10% and 6% in detecting small objects for the Vaihingen and Potsdam sets, respectively. The models also displayed considerable efficiency, with SCNNs notably faster than FPL and NDFCN networks. The study introduces an ensemble method that improves the overall precision by averaging the score maps from different model checkpoints, leading to further enhancements in performance. However, while the Atrous Spatial Pyramid Pooling (ASPP) was tested within the SCNNs, its improvement was found to be limited.

Land-Use Mapping for High-Spatial Resolution Remote Sensing Image Via Deep Learning: A Review

The review article, discusses the use of deep learning (DL) techniques in land-use mapping (LUM) with high-spatial resolution remote sensing images (HSR-RSIs).[25] It thoroughly reviews different high spatial resolution datasets, various basic DL methods used in LUM, and the performance of different models on the ISPRS Vaihingen and Potsdam datasets. The methods reviewed include supervised, semi-supervised, and unsupervised learning techniques, and both pixel-based and object-based approaches. They tested four state-of-the-art architectures for semantic segmentation: SegNet, U-Net, FCN-32s, and FCN-8s, all using VGG-16 as the backbone. The results suggest that DL-based LUM methods have significantly improved land-use mapping for HSR-RSIs, with continued advancements from semantic segmentation models. The encoder-decoder structures such as SegNet and

U-Net were noted as particularly promising, and U-Net's ability to train well with small datasets was highlighted. However, the study identified challenges in handling interclass homogeneity, small object segmentation, accurate edge segmentation, and insufficient training labeled HSR-RSIs for semantic segmentation.

Supervised methods of image segmentation accuracy assessment in land cover mapping

This article examines the present state of picture segmentation accuracy assessment and future research needs in the context of land cover mapping.[7] From 2014 to 2015, the authors evaluated literature from three major remote sensing publications, concluding that while qualitative visual interpretation is often utilized, several quantitative methods exist. The geometric and non-geometric categories of supervised methods are thoroughly investigated. The authors disclose that the application of these approaches is still in its early stages, owing to a lack of a solid basis in image segmentation accuracy assessment, challenges in implementing the methods, and a lack of credible suggestions for method selection. Key considerations for selecting methods are proposed, emphasizing application goals, the relative importance of under- and over-segmentation errors, and the pros and cons of the methods. The authors call for more exhaustive testing and comparison of supervised methods in different contexts and a reevaluation of over- and under-segmentation error concepts. They conclude that further research is needed to elevate the standard of image segmentation accuracy assessment.

RAANet: A Residual ASPP with Attention Framework for Semantic Segmentation of High-Resolution Remote Sensing Images

The article introduces an improved deep learning model, RAANet (Residual ASPP with Attention Net), for semantic segmentation of high-resolution remote sensing images.[27] RAANet aims to improve classification accuracy of land-use types, capitalizing on the atrous-spatial pyramid pooling (ASPP) framework and incorporating an attention module and residual structure to capture important semantic information and reduce network complexity. The model was tested using the land-cover domain adaptive semantic segmentation (LoveDA) and ISPRS Vaihingen datasets and compared against PSPNet, U-Net, and class-wise FCN (C-FCN) models. The results showed RAANet, specifically its convolutional block attention model (CBAM) variant, outperformed these models in terms of mIoU, mRecall, and mPrecision, producing higher prediction accuracy. Notably, the mIoU score of the LoveDA dataset and the ISPRS Vaihingen dataset was 2.94% higher than DeeplabV3plus and 1.12% higher than C-FCN respectively.

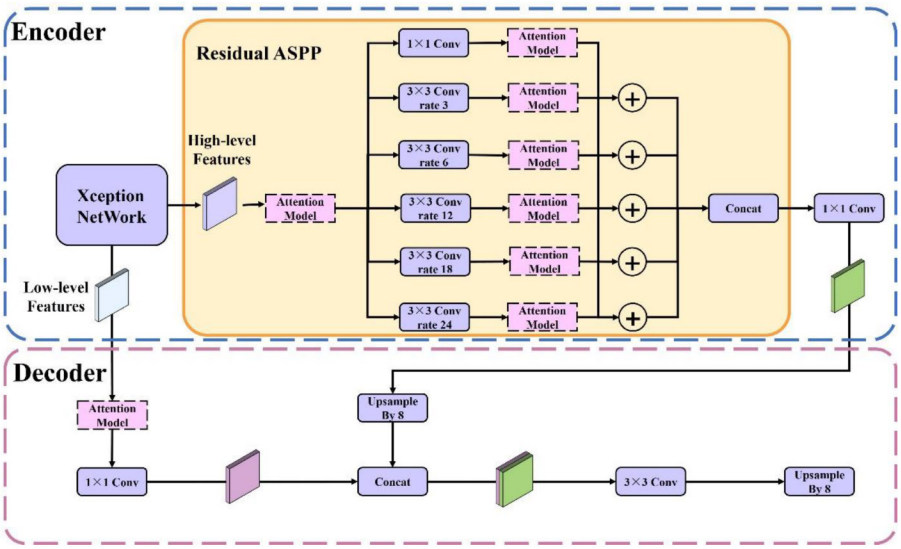


Figure 3.5: Structure of the RAANet *Image credit:* [27]

FRF-Net: Land Cover Classification From Large-Scale VHR Optical Remote Sensing Images

The article presents a novel deep learning approach, called the full receptive field (FRF-Net), for the classification of large-scale and very high-resolution (VHR) land cover from optical remote sensing images.[18] Leveraging the ResNet-101 backbone, the FRF-Net uses a self-attention mechanism to generate an ensemble feature that captures long-range semantics and a fusion attention mechanism that integrates low-level and high-level features to offer a refined semantic description for precise land cover mapping. The algorithm was tested on two datasets, namely the GID and ISPRS, using several state-of-the-art deep learning networks (Deeplab v3+, GCN, PSPNet, U-Net, and Seg-Net) for comparison. The experiments included various data augmentation strategies and were evaluated on the mean of classwise Intersection over Union (mIOU) and Pixel Accuracy (PA) metrics. The FRF-Net outperformed other models on both datasets, achieving an mIOU of 66.71% and 64.17%, and PA of 86.04% and 76.24% on the ISPRS and GID datasets, respectively. Crucially, the FRF-Net offered these superior results at a lower computational cost compared to competing models.

Transformer Meets Convolution: A Bilateral Awareness Network for Semantic Segmentation of Very Fine Resolution Urban Scene Images

In the research paper, a Bilateral Awareness Network (BANet) is proposed to effectively deal with the challenges in semantic segmentation from very fine resolution (VFR) urban scene images.[24] The BANet contains two key components, a dependency path and a texture path. The dependency path is designed based on ResT, a Transformer backbone with a memory-efficient multi-head self-attention mechanism to capture long-range relationships, while the texture path utilizes stacked convolution operations to capture fine-grained details. The authors further devise a feature aggregation module with a linear attention mechanism to fuse dependency features and texture features. To validate the effectiveness

of BANet, comprehensive experiments were performed on three large-scale urban scene image segmentation datasets, namely the ISPRS Vaihingen dataset, ISPRS Potsdam dataset, and UAVid dataset. Results showed that the BANet outperformed the baselines, achieving a 64.6% mean Intersection over Union (mIoU) on the UAVid dataset. The proposed method holds potential applications in urban planning, land cover classification, autonomous driving, and other related urban applications.

Learning Aerial Image Segmentation From Online Maps

In this study, the authors investigate the challenge of acquiring enough annotated training data for deep learning algorithms, such as convolutional neural networks (CNNs), used in high-resolution aerial image segmentation.[5] They suggest using large volumes of readily available data from legacy sources or crowd-sourced maps, despite their potential noise and inaccuracies, as an alternative to manually labeling extensive datasets. The researchers employed a leading CNN architecture designed for semantic segmentation of buildings and roads in aerial images, and examined its performance when trained on various datasets, including pixel-accurate ground truth from the same city and automatic training data obtained from distant locations via OpenStreetMap. The outcomes highlight that the volume of such extensive, public datasets can make up for their lower accuracy. Also, training data covering multiple cities improved the model's ability to generalize to new, unseen locations. These findings advocate for the use of large-scale, "weakly" labeled training data as a feasible approach to attain satisfactory performance and enhance the generalization capability of models in aerial image segmentation tasks.

Segmentation of Satellite Imagery using U-Net Models for Land Cover Classification

In the research paper the authors developed machine learning models utilizing a modified U-Net structure for creating land cover classification maps from satellite imagery.[19] The goal was to enhance the accuracy of existing land cover maps and aid in detecting land cover changes. The study relied on two datasets, namely the BigEarthNet satellite image archive and a self-curated set featuring a Sentinel2 image with a CORINE land cover map of Estonia. The convolutional models performed admirably, exhibiting a high overall F1 score of 0.749 on multiclass land cover classification with 43 possible image labels and indicating high IoU scores for specific land cover classes such as forests, inland waters, and arable land. The research also highlighted noise in the BigEarthNet dataset due to possible mislabeled images, emphasizing the need for class-based analysis and accuracy measurement. The paper achieved its objectives, setting a stage for future research directions like adjusting classes for better segmentation results and considering the hierarchical structure of the CORINE land cover classification to refine results.

3.2 Summary

The decision to use the Potsdam and Vaihingen satellite image datasets from the International Society for Photogrammetry and Remote Sensing (ISPRS) was primarily influenced by the findings of the article "Learning Aerial Image Segmentation From Online Maps" [5]. This paper highlighted the advantages of using large-scale publicly available labels to replace a

substantial part of the manual labeling effort while still maintaining satisfying performance. The insights from this article guided the selection of these particular datasets, which come with pre-labeled information about different land types.

In the choice of machine learning models - ResNet50, ResNet101, FC-ResNet50, VGG-19, and UNetFormer - multiple articles underscored the effectiveness of these models for image segmentation tasks. Specifically, research papers like "Deep Residual Learning for Image Recognition" [3] and "Fully Convolutional Networks for Semantic Segmentation" [2] illuminated the capabilities of ResNet and FC-ResNet architectures, respectively, in handling high-dimensional data and achieving precise image recognition results. The usage of VGG-19 was inspired by "Very Deep Convolutional Networks for Large-Scale Image Recognition" [1], which showcased the model's robustness in learning from large image datasets. Lastly, the application of the UNet model stemmed from the understanding gained from "Segmentation of Satellite Imagery using U-Net Models for Land Cover Classification" [19] about its effectiveness in handling semantic segmentation tasks, especially in the context of aerial images.

After applying these models, their performances were compared through a statistical lens, revealing the most accurate model for this specific task. This comparative analysis was shaped by numerous scientific studies that emphasize the importance of proper evaluation metrics in assessing the performance of machine learning models.

Chapter 4

Experimental Setup

The experimental setup used for the study is fully described in this chapter.

4.1 Dataset Preprocessing

Potsdam and Vaihingen datasets were specially prepared for ISPRS(Semantic Labeling) competition, consisting of aerial photos taken of German cities of Potsdam and Vaihingen with associated ground truth data that labeled each pixel according to class.[32] The aim is for algorithms that can accurately categorize these datasets using available ground truth information.

4.1.1 Potsdam

In Potsdam, there were 38 patches in the data set (all of the same size), each of which is a true orthophoto (TOP) that was taken from a larger TOP mosaic. Images were provided as TIFF files with various channel compositions:

1. IRRG: 3 channels (IR-R-G)

2. RGB: 3 channels (R-G-B)

3. RGBIR: 4 channels (R-G-B-IR)

The size of each patch/image was 6000 x 6000, which was not suitable for training. Some images were (5999, 5999) pixels, which had to be padded. 33 images were used for training and 5 for testing. Each image was further divided into 400 small images of size 300 x 300. The pixels of each small image were divided by 255. It is a typical preprocessing step in computer vision which helps in scalability of the data and enhances the efficiency of machine learning algorithms. The number of classes in the dataset were 6 i.e 'Impervious', 'Building', 'Low vegetation', 'Tree', 'Car', and 'Clutter'.

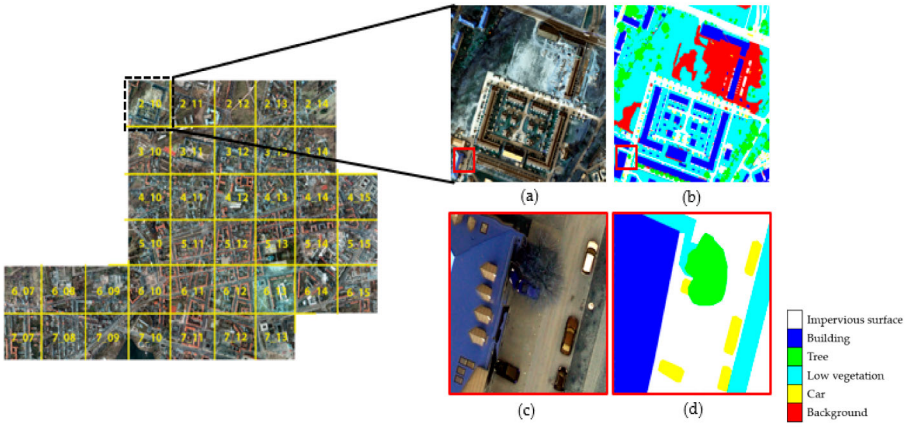


Figure 4.1: Example of Dataset *Image credit:* [32]

4.1.2 Vaihingen

In Vaihingen, there were 33 patches in the data set. Each of which is a true orthophoto (TOP) that was taken from a larger TOP mosaic. Images were provided as TIFF files with various channel compositions:

1. IRRG: 3 channels (IR-R-G)

2. RGB: 3 channels (R-G-B)

3. RGBIR: 4 channels (R-G-B-IR)

The resolution of images in the dataset were of 2250x2569 and 1800 x 1919 pixels. The different sized images were handled by resizing all the images to 2560 x 1920 pixels. 28 images were used for training, and 5 for testing. Each image was further divided into 150 small images of size 256 x 128. Again, pixels of image were divided by 255, which enhances performance of the model and the results can be more easily understood. The number of classes in the dataset were 9 i.e 'Powerline', 'Low vegetation', 'Impervious', 'Tree', 'Car' , 'Fence', 'Roof', 'Facade', and 'Shrub'.

4.2 Deep Learning Algorithms

Following algorithms were used on both datasets i.e Potsdam and Vaihingen with the same setting but both have different data pre-processing.

4.2.1 Resnet 101

Hyperparameter tuning was used to determine the optimal configuration of the ResNet 101 model for image classification. As a starting point for our tuning process, we selected the initial parameters of a learning rate set at 0.0001, batch size 20, and Adam Optimizer. The initial values were influenced based on previous experience. However, they were not the final values. Keras Tuner's hyperparameter tuning tool allowed us to test a wide range of parameters, such as the learning rate and batch size. We also tested the optimizer type

and number of epochs. Our goal was to maximize the validation Intersection Over Union (IoU) of our model, our chosen evaluation metric.

These were not the final values. We tested hyperparameters such as the learning rate and batch size. The optimizer type was also tested. Our goal was to maximize the validation Intersection Over Union (IoU) of our model, our chosen evaluation metric. The best results were achieved by the Adam optimizer when it was paired with the learning rate of 0.001. We found that after several experiments, 30 epochs produced the best performance.

Hyperparameter tuning led to significant enhancements of model performance. The final model configuration included a learning rate of 0.001, batch size of 8, Adam as the optimizer, and training for 30 epochs.

4.2.2 Resnet-50

GeoPandas, Rasterio and Shapely were employed to manage data processing and model training tasks efficiently in Python. For geospatial data processing tasks, Sklearn's train_test_split function played an integral part; its partition of dataset into training and validation sets using train_test_split was essential as well. h5py was utilized as an additional safeguard against potential operational mishaps with weight-saving models while TensorFlow provided crucial machine learning library support in constructing and training machine learning models using TensorFlow's well-established machine learning library facilitating construction and training of machine learning models using TensorFlow's well-established machine learning library which helped in building and training machine learning models with ease.

We created a parameter grid containing values for key hyperparameters, including batch size, number of epochs, learning rate, network architecture and learning rate schedule. We then utilized the Random Search Algorithm to exhaustively search through all possible combinations of these parameters while simultaneously evaluating model performance against an objective scoring metric such as accuracy or loss.

Cross-validating different parameter combinations allowed us to identify the optimal set of hyperparameters with cross-validation using the Grid-Search algorithm, then using cross-validation again we determined which set were performing best based on the highest achieved performance metric score. These optimal hyperparameters included 32 batch size, 30 epochs, learning rate of 0.001, constant learning rate schedules, and data augmentation; which allowed us to attain the best performance on the validation set.

4.2.3 VGG-19

Our TensorFlow-based image classification model training process involved extensive hyperparameter tuning to achieve an ideal configuration. To begin the process, initial parameters were established - such as learning rate of 0.0001 and batch size of 20 in combination with Adam optimizer settings - that were not chosen arbitrarily, but were determined based on past experience and commonly adopted practices in our field.

Although these settings were used as the starting point for tuning, which was assisted by Keras Tuner. Experimentation involved testing different hyperparameters such as learning rate, batch size, optimizer type and configurations of convolutional and dense layers - with the goal being to find an ideal combination that maximized validation Intersection over Union (IoU), our chosen evaluation metric. Out of the various optimizers tested - SGD and RMSprop among them - Adam Optimizer with a learning rate of 0.001 produced better results.

We significantly improved the performance of the model through an iterative tuning process. The final model was created using the Adam Optimizer, with a learning rate 0.001, batch size 8, and optimal configurations of convolutional and dense layer.

4.2.4 FC-Resnet-50

In our experiment, we optimized a Fully-Convolutional Network (FCN), which was based on ResNet architecture. Several trials were conducted for hyperparameter optimization to find the best values for key parameters. We explored a search space systematically using Keras Tuner to optimize the model's performance.

The optimal learning rate was identified as 0.001, which struck a balance between model stability and learning speed. Adam was the most efficient and effective optimizer for our model and data. DataGenerator's batch size was tweaked in order to optimize the tradeoff between computational efficiency, and generalization of the model.

Hyperparameters for the ReduceLROnPlateau callback were adjusted in order to adapt effectively to the model's learning progression. We decided on decreasing learning rate by 50% when validation loss didn't improve for three epochs; minimum learning rate set at 0.000001; optimal number of epochs determined at 30; this allowed the model to identify patterns without overfitting.

4.2.5 Unet-Former

Google Colab's computational capabilities were used in this experiment. The first step was to install Conda to manage library dependency. Pip then installed specific libraries like Rasterio and Pillow, while Conda enabled the installation of PyTorch-related packages. Google Drive was mounted in Colab to enable direct data access. The GeoSeg repository from GitHub, which contains network architectures, was cloned and DataLoader was installed for efficient data handling.

In order to meet the requirement, GeoSeg installed additional Python libraries. Due to the large size of the image, the patch-based approach was chosen as the best strategy. To achieve more efficient processing during both training and validation, smaller patches from large images were generated using the potsdam_patch_split.py script, splitting off sets for this purpose. To train the model, UNet-Former was used in conjunction with a configuration file specific to the Potsdam dataset. Once training began, metrics like F1 score, overall Accuracy (OA) and mean intersection over Union (IoU) were used to evaluate the performance of the model. These measurements provide insight into the segmentation

capability of the model.

Inference was then performed using the UNet-Former trained model on a large-scale image. The predictions were then saved and compared visually by loading the predicted masks and the ground truth masks in a separate database. This allowed a qualitative evaluation of its performance.

Chapter 5

Results

This chapter is devoted to discussing the findings and is organized around the research questions established in the introduction chapter.

5.1 Results on Potsdam dataset

This table displays the performance of several algorithms on the Potsdam dataset, which is another often used benchmark for measuring the performance of algorithms in semantic segmentation. The table depicts the mean intersection over union (IOU), F1 score, and accuracy, as well as the performance for each specific class.

Table 5.1: Performance Comparison on Potsdam dataset

	ResNet-101	ResNet-50	FC ResNet-50	VGG-19	UNetFormer
Impervious	0.844	0.844	0.817	0.840	0.846
Building	0.822	0.822	0.895	0.900	0.899
Low vegetation	0.744	0.773	0.798	0.833	0.798
Tree	0.698	0.698	0.814	0.724	0.717
Car	0.924	0.924	0.932	0.931	0.835
Clutter	0.703	0.704	0.769	0.723	0.669
Mean IOU	0.543	0.543	0.629	0.825	0.790
F1 score	0.765	0.765	0.821	0.809	0.880
Accuracy	0.775	0.770	0.846	0.820	0.888

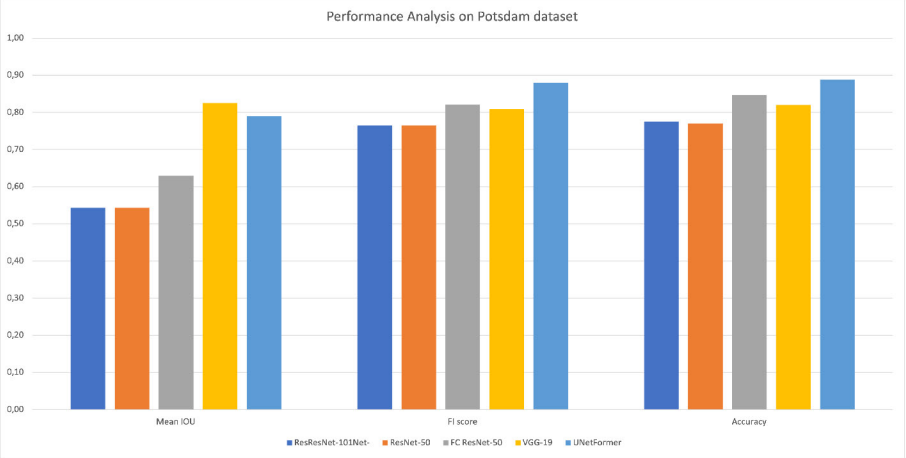


Figure 5.1: Performance Analysis on Potsdam

UNetFormer is the top performing algorithm in the table, with the highest F1 scores and accuracy of all methods. UNetFormer excels in the majority of classes. It has strong Impervious and Building ratings. ResNet-101, ResNet-50, and UNetFormer all do well in most classes but have lower total results. FC ResNet-50, VGG-19, and ResNet-101 all perform worse overall and in several classes.

5.2 Results on Vaihingen dataset

This table shows the results for the Vaihingen dataset. The table displays accuracy for each class, the F1 score, accuracy, and average intersection over union (IOU).

Table 5.2: Performances of Algorithms on the Vaihingen dataset

	ResNet-101	**ResNet-50**	**FC ResNet-50**	**VGG-19**	**UNetFormer**
Powerline	0.997	0.997	0.998	0.998	x
Low Vegetation	0.881	0.487	0.420	0.704	0.770
Impervious	0.942	0.750	0.660	0.917	0.891
Car	0.389	0.363	0.412	0.352	0.724
Fence	0.98	0.975	0.998	0.989	x
Roof	0.950	0.725	0.835	0.851	0.932
Facade	0.975	0.995	0.997	0.979	x
Shrub	0.995	0.997	0.998	0.989	x
Tree	0.885	0.709	0.630	0.858	0.792
Mean IOU	0.726	0.455	0.402	0.656	0.800
F1 score	0.928	0.799	0.743	0.893	0.892
Accuracy	0.927	0.787	0.730	0.893	0.920

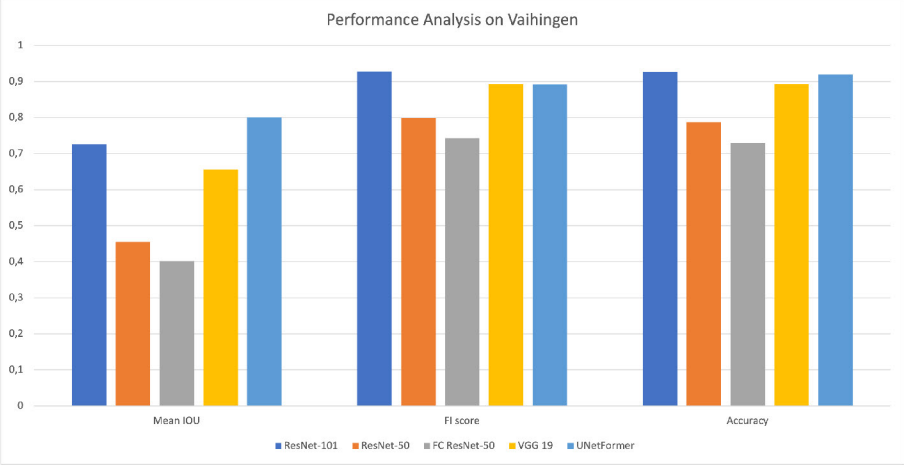


Figure 5.2: Performance Analysis on Vaihingen

Among the algorithms in the table and graph, UNetFormer has the highest accuracy, F1 score, and mean IOU. UNetFormer excels in the majority of classes. It has high Powerline and Low Vegetation ratings. ResNet-101, VGG-19, and VGG-19 also perform well in some classes but score lower than UNetFormer.

5.3 Comparison on Potsdam dataset

On the Potsdam dataset, Following table compares the performance of various models, including UNET Former ResNet-101 ResNet-50 VGG-19 and FC ResNet-50. The t-test findings are used to make the comparison. This is a statistical strategy for comparing two distinct samples or groups.

In each row, the t-value reflects how much the UNET Former mean performance differs from the other models. The larger the t value, the bigger the gap in mean performance between the two groups.

Table 5.3: Comparison of UNetFormer with other models on the Potsdam dataset

	ResNet-101	ResNet-50	VGG-19	FC ResNet-50
T-Value	24.47	24.47	16.62	15.41
P	< 0.0001	< 0.0001	< 0.0001	< 0.0001

According to the table above, the t value is large in all comparisons. It ranges from 15.41 and 24.47. This shows that UNET Former outperforms the other models on the Potsdam dataset. This table compares the performance of UNET Former, ResNet-50, and VGG-19 on the Vaihingen dataset. The p-value (0.0001) is relatively low, indicating that the differences observed are statistically significant. This comparison is based on the t-test results.

5.4 Comparison on Vaihingen

On the Vaihingen dataset, this table compares the performance of several models, including UNET Former ResNet-101 ResNet-50 VGG-19 and FC ResNet-50. This comparison is based on the t-test results. The t-test is a statistical procedure used to compare two independent samples or groups.

Table 5.4: Comparison of UNetFormer with other models on the Vaihingen dataset

	ResNet-101	ResNet-50	VGG-19	FC ResNet-50
T-Value	9.643	83.6	29.6	102.5
P	< 0.0001	< 0.0001	< 0.0001	< 0.0001

The table demonstrates that the t value is large in all comparisons. It ranges between 9.643 and 102.5. This suggests that UNET Former performs much better than the other models on the Vaihingen dataset. In all comparisons, the p value is quite low (0.0001). This means that the differences detected are statistically significant.

We can see, in particular, that FC ResNet-50 is the one with the highest t-value. This indicates the biggest difference in performance when compared to UNET Former. ResNet-50, VGG-19 and UNET Former also have a significant difference in performance. ResNet-101 shows the lowest t-value indicating the least difference in performance when compared to UNET Former. However, it is still significant with a 0.0001 value.

5.5 Performance of ResNet-101 Models on Potsdam Dataset: RGB And IRRG

The table below shows the performance of two ResNet-101 models using the Potsdam dataset. The first ResNet-101 is trained using RGB images while the second ResNet-101 is trained with RGB and infrared images.

Table 5.5: Performance comparison of algorithms on the Potsdam dataset

Performances of Algorithms on Potsdam dataset		
	ResNet-101 (RGB)	ResNet-101 (IRRG)
Impervious	0.825	0.858
Building	0.841	0.939
Low vegetation	0.744	0.807
Tree	0.740	0.723
Car	0.931	0.949
Clutter	0.695	0.820
Mean IOU	0.563	0.632
F1 score	0.775	0.854

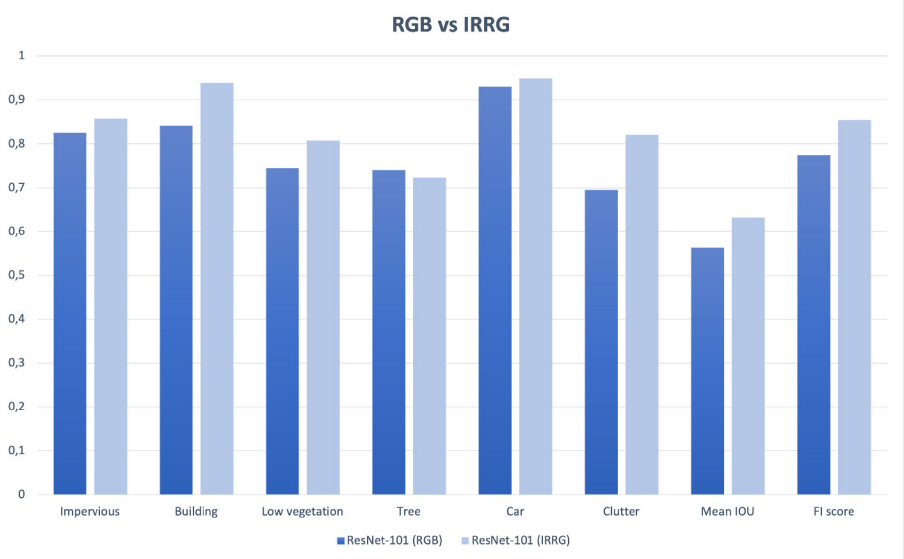


Figure 5.3: Analysis of Resnet-101 on RGB and IRRG Potsdam datset

In the table and graph, it is clear that the ResNet-101 model (IRRG), in terms of IOU and F1 scores, outperforms ResNet-101 model (RGB). The ResNet-101 model (IRRG), in comparison to the ResNet-101 model (RGB), achieves higher scores on the Impervious, Low Vegetation, Car and Clutter classes. ResNet-101's (RGB model) performs better in the Tree class than ResNet-101's (IRRG model). The ResNet-101 model's mean IOU score and F1 scores are higher than the ResNet-101 model (RGB), indicating the ResNet-101 model (IRRG), is better at segmenting the Potsdam dataset.

5.6 Comparison of Resnet-101 on RGB and IRRG

The following table compares two ResNet 101 models that were trained on the Potsdam dataset. One model uses RGB images, and the other RGB and infrared images (IRRG). The comparison is based upon the results of the t-test. This is a statistical method used to

compare two independent samples or groups.

Table 5.6: Comparison of ResNet-101 RGB on the Potsdam dataset

	ResNet-101 IRRG
T-Value	2.80114
P-Value	0.01897

The table shows that the t-value for ResNet-101 RGB is 2.80114. This indicates a statistically significant but small difference in performance. The p value is 0.01897 which is less than 0.05 and indicates that the observed performance difference is statistically significant.

Chapter 6

Discussion

The discussion of the findings from the Results chapter is included in this chapter. It is organized in accordance with the introduction chapter's research questions.

6.1 Performance of Modern Neural Networks in Land Cover Segmentation

The performance of UnetFormer (a deep learning-based technique) was compared to that of typical unsupervised picture segmentation techniques in this study. The research also looked into how UnetFormer's performance differs based on the type of image or dataset being studied. UnetFormer will be compared to regularly used deep-learning-based approaches such as ResNet-50, VGG-19, and FC ResNet-50.

Techniques based on deep learning are becoming more popular in sectors such as remote sensing, computer vision, and medical imaging. The study discovered that UnetFormer outperformed typical unsupervised techniques, ResNet-101 and ResNet-50, in terms of accuracy and F1 scores. The study also emphasizes the need to select the segmentation technique based on the type of image and dataset being studied. The study's findings may be useful to researchers and practitioners in selecting the optimum picture segmentation method for their application.

Even though UnetFormer has been identified as the most effective model, other architectures or methods that have not yet been explored may produce better results. The exploration of the integration of infrared was limited to one source. This leaves the question of the performance improvements that can be achieved by using other data types open. The results may also vary depending on the datasets used, which is another factor that requires further investigation. In looking back at the research process, it was found that the choice of deep learning architecture and the inclusion of infrared images were major influences on performance. These insights will be valuable in future efforts to optimize the performance of such tasks. Future improvements may incorporate advanced hyperparameter tuning methods such as grid searching, random searching, or Bayesian Optimization to further enhance the model's performance.

6.2 Effect of Inclusion of Infrared Data on Modern Deep Learning Architectures' Performance

This study looked at how adding infrared information to ResNet-101 models affected image segmentation accuracy and efficiency, as well as how using both RGB images and IRRG affected image segmentation accuracy and efficiency. The high-resolution Potsdam dataset was used to test the ResNet 101 models. Individuals who were only trained with RGB images outperformed those who were only trained with IRRG images, earning higher IOU, F1, and accuracy values. ResNet-101 had a median IOU of 0.632 and an F1 of 0.854 after being trained on IRRG pictures. It also had an accuracy of 0.868.

Incorporating infrared (IR) data into deep learning networks can increase their performance in remote sensing and image processing greatly. Thermal information captured by IR data, allows the identification of temperature fluctuations and patterns, which can be critical in a variety of applications such as environmental monitoring, agricultural, and infrastructure inspection. This data can aid in the detection of hotspots or anomalies, assuring the dependability and safety of important systems. The combination of IR and visible spectrum data can improve object detection and classification tasks, resulting in a more complete picture of the scene. IR data can also resolves restrictions associated with visible spectrum data alone, especially under difficult lighting circumstances. Additional convolutional layers and pooling levels will be added at each block of convolutional and pooling levels, rising as feature mappings pass through the network this helps the model to learn complicated parts of an image while providing predictions based on features learned.

Deep learning models can give more accurate and resilient solutions for remote sensing, surveillance, and infrastructure monitoring applications by harnessing the capabilities of both spectra. Using IR data opens up new avenues for enhanced analysis and insights, resulting in better decision-making processes and assuring the effectiveness and reliability of important systems and activities.

According to the findings, integrating IR data can be a useful strategy for increasing deep-learning model performance in semantic segmentation. utilizing IRRG images is more efficient in terms of accuracy, computing time, and efficiency than utilizing RGB images. These findings have far-reaching implications for a variety of applications requiring high accuracy and efficiency in semantic segmentation tasks. Our study adds to the literature on the use of alternative information sources, such as IR images, to improve ResNet-101's performance in semantic segmentation tasks.

Chapter 7

Conclusion

This study reviewed the performance of deep-neural architectures in land-cover semantic segmentation tasks, focusing on UnetFormer ResNet-101 ResNet-50 FC ResNet-50 and VGG-19. These models were assessed for their performance in land cover segmentation as well as the effects of integrating infrared pictures. According to the findings, UnetFormer was the best model in terms of many criteria. Infrared data was also integrated into ResNet-101, which improved performance. This underscores the significance of multidimensional data for such jobs. These findings are critical because they demonstrate the power of deep-learning models for semantic segmentation. Potential uses include remote sensing and medical imaging.

The study also addressed the research objectives by providing a comparison of the performance and effects of infrared data in land cover segmentation tasks. These results set the stage for future explorations that may investigate the integration of different types of data, or the use novel deep learning architectures in order to improve the performance of semantic segments. This research has a significant impact on the field of semantic segmentation, enhancing the literature and providing practical implications to industry practitioners. This study, which concludes this article, has demonstrated the effectiveness of UnetFormer for land cover segmentation tasks. It also highlighted the potential to use infrared images to improve model performance. This study could be viewed as a major step forward in the domain of semantic segmentation and opens the door for future research to optimize performance.

Bibliography

[1] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[2] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[4] A. Dertat, "Applied deep learning - part 3: Autoencoders," *Medium*, 2017, Oct 3, 2017. [Online]. Available: `https://towardsdatascience.com/applied-deep-learning-part-3-autoencoders-1c083af4d798`.

[5] P. Kaiser, J. D. Wegner, A. Lucchi, M. Jaggi, T. Hofmann, and K. Schindler, "Learning aerial image segmentation from online maps," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 11, pp. 6054–6068, Nov. 2017. DOI: `10.1109/TGRS.2017.2719738`.

[6] K. Chen, K. Fu, M. Yan, X. Gao, X. Sun, and X. Wei, "Semantic segmentation of aerial images with shuffling convolutional neural networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 2, pp. 173–177, Feb. 2018. DOI: `10.1109/LGRS.2017.2778181`.

[7] H. Costa, G. M. Foody, and D. S. Boyd, "Supervised methods of image segmentation accuracy assessment in land cover mapping," *Remote Sensing of Environment*, vol. 205, pp. 338–351, 2018. DOI: `10.1016/j.rse.2017.11.024`. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/S0034425717305734`.

[8] X. Hui, J. Bian, X. Zhao, and M. Tan, "Vision-based autonomous navigation approach for unmanned aerial vehicle transmission-line inspection," *International Journal of Advanced Robotic Systems*, vol. 15, p. 172 988 141 775 282, 2018. DOI: `10.1177/1729881417752821`.

[9] Q. Liu, A.-B. Salberg, and R. Jenssen, "A comparison of deep learning architectures for semantic mapping of very high resolution images," in *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, Valencia, Spain, 2018, pp. 6943–6946. DOI: `10.1109/IGARSS.2018.8518533`.

[10] V. N. Nguyen, R. Jenssen, and D. Roverso, "Automatic autonomous vision-based power line inspection: A review of current status and the potential role of deep learning," *International Journal of Electrical Power & Energy Systems*, vol. 99, pp. 107–120, 2018, ISSN: 0142-0615. DOI: 10.1016/j.ijepes.2017.12.016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0142061517324444.

[11] Y. Zheng, C. Yang, and A. Merkulov, *Breast cancer screening using convolutional neural network and follow-up digital mammography*. 2018, DOI: https://doi.org/10.1117/12.2304564.

[12] A. Ali. "Self organizing map (som) with practical implementation." (May 2019), [Online]. Available: https://medium.com/machine-learning-researcher/self-organizing-map-som-c296561e2117.

[13] Z. Lan, Q. Huang, F. Chen, and Y. Meng, "Aerial image semantic segmentation using spatial and channel attention," in *2019 IEEE 4th International Conference on Image, Vision and Computing (ICIVC)*, Xiamen, China, 2019, pp. 316–320. DOI: 10.1109/ICIVC47709.2019.8981028.

[14] D. Li and X. Wang, "The future application of transmission line automatic monitoring and deep learning technology based on vision," in *2019 IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*, Chengdu, China, 2019, pp. 131–137. DOI: 10.1109/ICCCBDA.2019.8725702.

[15] L. Yue, W. Wanguo, X. Ronghao, L. Zengwei, and T. Yuan, "An intelligent identification and acquisition system for uavs based on edge computing using in the transmission line inspection," in *Proceedings of the 2019 4th International Conference on Robotics, Control and Automation*, Jul. 2019, pp. 205–209.

[16] F. M. Haroun, S. N. Deros, and N. M. Din, "A review of vegetation encroachment detection in power transmission lines using optical sensing satellite imagery," *ArXiv*, 2020. [Online]. Available: https://doi.org/10.30534/ijatcse/2020/8691.42020.

[17] X. Liu, X. Miao, H. Jiang, and J. Chen, "Review of data analysis in vision inspection of power lines with an in-depth discussion of deep learning technology," *ArXiv*, 2020. [Online]. Available: https://doi.org/10.1016/j.arcontrol.2020.09.002.

[18] Q. Sang, Y. Zhuang, S. Dong, G. Wang, and H. Chen, "Frf-net: Land cover classification from large-scale vhr optical remote sensing images," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 6, pp. 1057–1061, Jun. 2020. DOI: 10.1109/LGRS.2019.2938555.

[19] P. Ulmas and I. Liiv, "Segmentation of satellite imagery using u-net models for land cover classification," *ArXiv*, 2020, arXiv:2003.02899.

[20] M. Chen, Y. Tian, S. Xing, *et al.*, "Environment perception technologies for power transmission line inspection robots," *Journal of Sensors*, vol. 2021, pp. 1–16, Mar. 2021. DOI: 10.1155/2021/5559231.

[21] Y. Chen, X. Ouyang, K. Zhu, and G. Agam, "Semantic segmentation in aerial images using class-aware unsupervised domain adaptation," in *Proceedings of the 4th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery*, ser. GEOAI '21, Beijing, China: Association for Computing Machinery, 2021, pp. 9–16, ISBN: 9781450391207. DOI: 10.1145/3486635.3491069. [Online]. Available: https://doi.org/10.1145/3486635.3491069.

[22] I. H. Sarker, "Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions," *SN Computer Science*, vol. 2, no. 6, p. 420, 2021, ISSN: 2661-8907. DOI: 10.1007/s42979-021-00815-1. [Online]. Available: https://doi.org/10.1007/s42979-021-00815-1.

[23] Tokyo Electric Power Company Holdings, Inc. and Blue Innovation Co., Ltd. and TEPCO Systems Corporation and TEPCO Power Grid, Inc. "Development and introduction of an "autonomous flight system for transmission line inspection drones" that enables autonomous drones to follow and photograph transmission lines." Retrieved from https://www.tepco.co.jp/en/hd/newsroom/press/archives/2021/20210511_01.html. (May 2021).

[24] L. Wang, R. Li, D. Wang, C. Duan, T. Wang, and X. Meng, "Transformer meets convolution: A bilateral awareness network for semantic segmentation of very fine resolution urban scene images," *Remote Sensing*, vol. 13, no. 16, p. 3065, 2021. DOI: 10.3390/rs13163065.

[25] N. Zang, Y. Cao, Y. Wang, B. Huang, L. Zhang, and P. T. Mathiopoulos, "Land-use mapping for high-spatial resolution remote sensing image via deep learning: A review," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 5372–5391, 2021. DOI: 10.1109/JSTARS.2021.3078631.

[26] M. Gazzea, M. Pacevicius, D. O. Dammann, A. Sapronova, T. M. Lunde, and R. Arghandeh, "Automated power lines vegetation monitoring using high-resolution satellite imagery," *IEEE Transactions on Power Delivery*, vol. 37, no. 1, pp. 308–316, Feb. 2022. DOI: 10.1109/TPWRD.2021.3059307.

[27] R. Liu, F. Tao, X. Liu, *et al.*, "Raanet: A residual aspp with attention framework for semantic segmentation of high-resolution remote sensing images," *Remote Sensing*, vol. 14, no. 13, p. 3109, 2022. DOI: 10.3390/rs14133109.

[28] D. K. Sharma, M. Chatterjee, G. Kaur, and S. Vavilala, "3 - deep learning applications for disease diagnosis," in *Deep Learning for Medical Applications with Unique Data*, D. Gupta, U. Kose, A. Khanna, and V. E. Balas, Eds., Academic Press, 2022, pp. 31–51, ISBN: 978-0-12-824145-5. DOI: https://doi.org/10.1016/B978-0-12-824145-5.00005-8. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128241455000058.

[29] A. Tsellou, K. Moirogiorgou, G. Plokamakis, G. Livanos, K. Kalaitzakis, and M. Zervakis, "Aerial video inspection of greek power lines structures using machine learning techniques," in *2022 IEEE International Conference on Imaging Systems and Techniques (IST)*, Kaohsiung, Taiwan, 2022, pp. 1–6. DOI: 10.1109/IST55454.2022.9827761.

[30] L. Wang, R. Li, C. Zhang, *et al.*, "Unetformer: A unet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 190, pp. 196–214, 2022.

[31] European Space Agency. "Feasibility study." (2023), [Online]. Available: https://business.esa.int/projects/grideyes (visited on 06/09/2023).

[32] International Society for Photogrammetry and Remote Sensing. "Isprs - urban semantic lab - detection and reconstruction." (2023), [Online]. Available: https://www.isprs.org/education/benchmarks/UrbanSemLab/detection-and-reconstruction.aspx#VaihigenDataDescr (visited on 06/09/2023).

[33] Y. Luo, X. Yu, D. Yang, *et al.*, "A survey of intelligent transmission line inspection based on unmanned aerial vehicle," *Artificial Intelligence Review*, vol. 56, pp. 173–201, 2023. DOI: 10.1007/s10462-022-10189-2.

[34] P. Song, J. Li, Z. An, H. Fan, and L. Fan, "Ctmfnet: Cnn and transformer multiscale fusion network of remote sensing urban scene imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, no. Art no. 5900314, pp. 1–14, 2023. DOI: 10.1109/TGRS.2022.3232143.

[35] G. Developers, "Overview of gan structure," [Online]. Available: https://developers.google.com/machine-learning/gan/gan_structure.

[36] M. L. .-. Paperspace. "Convolutional neural network (cnn)." (), [Online]. Available: https://machine-learning.paperspace.com/wiki/convolutional-neural-network-cnn.

[37] L. Dzierzak. "Epic job: Meet a high-voltage line inspector." (Year), [Online]. Available: https://gearjunkie.com/adventure/high-tension-wire-inspector-job-epic-occupation.