TECHNICAL UNIVERSITY OF CRETE

DEPT. OF ELECTRICAL & COMPUTER ENGINEERING

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

Thesis committee:
Prof. Michalis Zervakis

Prof. Euripides Petrakis

Prof. Dionissios Hristopulos

Author:
Aikaterini Tsellou

Master's Thesis

Chania, 2024

Abstract

Power line inspection is a crucial task for the uninterrupted operation of an electricity distribution network. Till date, it is mainly carried out using manned helicopters or foot patrol. However, autonomous, intelligent inspection using unmanned aerial vehicles (UAVs) equipped with camera sensors has come to the fore lately as it can offer an advantageous automated way to deliver the task of inspection. For the accurate detection of the power lines in the imagery acquired, different state-of-the-art semantic segmentation techniques have been used. In this work, attention is mainly paid to the structure of the power lines, in order to find a proper deep learning architecture that can segment them efficiently, preserving their thin shape and reducing background noise. It is found out that DNNs that employ dilated convolutions can reach this goal and achieve high performance. The architectures in this work were evaluated in both literature datasets and videos collected by HEDNO S.A. (Hellenic Electricity Distribution Network Operator S.A.) using UAVs. Results show that, out of the four deep learning-based segmentation architectures used in the experiments, the D-LinkNet architecture, first introduced for road segmentation purposes in high-resolution satellite imagery, outperformed the others in terms of F'l-Score in various background scenarios.

Περίληψη

Ο έλεγχος των ηλεκτρικών γραμμών αποτελεί ζωτικής σημασίας εργασία για την αδιάκοπη λειτουργία ενός δικτύου διανομής ηλεκτρικής ενέργειας. Μέχρι σήμερα, πραγματοποιείται κυρίως με επανδρωμένα ελικόπτερα ή πεζοπερίπατο. Ωστόσο, η αυτόνομη, έξυπνη έλεγχος χρησιμοποιώντας μη επανδρωμένα αεροσκάφη (UAVs) εξοπλισμένα με αισθητήρες κάμερας έχει προσεγγίσει πρόσφατα το προσκήνιο, καθώς μπορεί να προσφέρει ένα πλεονέκτημα ως προς το αυτοματοποιημένο τρόπο εκτέλεσης της εργασίας έλεγχου. Για την ακριβή ανίχνευση των ηλεκτρικών γραμμών στις λήψεις εικόνας, έχουν χρησιμοποιηθεί διάφορες τεχνικές σημασιολογικής εγκάλισης που βρίσκονται στην αιχμή της τεχνολογίας. Σε αυτήν την εργασία, επιδίδεται μεγάλη προσοχή στη δομή των ηλεκτρικών γραμμών, προκειμένου να βρεθεί μια κατάλληλη αρχιτεκτονική εκμάθησης βαθιάς μάθησης που να μπορεί να τις εγκαλίσει αποτελεσματικά, διατηρώντας το λεπτό τους σχήμα και μειώνοντας τον θόρυβο του φόντου. Έχει διαπιστωθεί ότι οι ΔΝΝ (Δίκτυα Νευρωνικών Δικτύων) που χρησιμοποιούν διευρυμένες συνελίξεις μπορούν να επιτύχουν αυτόν τον στόχο και να πετύχουν υψηλή απόδοση. Οι αρχιτεκτονικές που χρησιμοποιήθηκαν σε αυτήν την εργασία αξιολογήθηκαν τόσο σε σύνολα δεδομένων από τη βιβλιογραφία όσο και σε βίντεο που συλλέχθηκαν από την HEDNO S.A. (Διαχειριστής Ελληνικού Δικτύου Διανομής Ηλεκτρικής Ενέργειας Α.Ε.) με τη χρήση UAVs. Τα αποτελέσματα δείχνουν ότι, από τις τέσσερις αρχιτεκτονικές εκμάθησης βαθιάς μάθησης που χρησιμοποιήθηκαν στις πειραματικές δοκιμές, η αρχιτεκτονική D-LinkNet, που εισήχθη αρχικά για σκοπούς εγκαλίσεως δρόμων σε εικόνες υψηλής ανάλυσης από δορυφορικές λήψεις, ξεπέρασε τις άλλες όσον αφορά το F'l-Score σε διάφορα σενάρια φόντου.

Contents

Chapter 1

Introduction

1.1 Problem Statement

1.1.1 Power Line Inspections Today

Regular inspections and timely maintenance play a pivotal role in minimizing resource input, controlling costs, and ensuring a swift restoration of electricity within the complex landscape of power grid operations. However, the current methodologies employed are far from optimal, as many power grid operators still depend on a combination of ground personnel and low-flying helicopters for the thorough inspection of the power lines. This reliance on traditional methods introduces several significant drawbacks that disrupt operational efficiency and compromise overall effectiveness.

One critical challenge lies in the inefficiency of outage management due to the suboptimal use of helicopters for power line inspections. Helicopters are not well-suited for this complex task due to their size and weight, especially when dealing with bad weather or navigating through narrow spaces. In cases where helicopters lack integrated cameras, inspectors are forced to physically ascend masts to identify faults. Even when equipped with cameras, helicopters require significant manual effort to analyze the captured photographs, leading to delays in identifying and fixing potential issues.

Furthermore, the reliance on helicopters introduces a significant risk to workforce safety, particularly when operating in challenging terrains and adverse weather conditions. As reported by T&D World, the utility line work sector ranks among the top ten most dangerous jobs in America, with an alarming 30 to 50 workers out of every 100,000 facing fatal accidents annually. This highlights the immediate need for a safer, more efficient, and technologically advanced approach to power line inspections.

Another significant concern arises from the existence of disparate systems throughout the inspection process. Utilities commonly employ different systems from the planning phase to the detailed ground crew work orders, resulting in isolated data flow. This lack of synchronization in the system portfolio makes the analysis process more difficult and time-consuming than needed. A complete solution includes bringing all stages of the inspection process together into a well-coordinated system. This helps streamline operations, improve efficiency, and allows for a quicker response to potential issues in the power grid infrastructure.

1.1.2 Benefits of using drones for power line inspection

Using drones for power line inspection goes beyond just improving safety and efficiency; it also delves into the realm of technological advancement and data analytics. Drones equipped with cutting-edge sensors and cameras can capture not only visual data but also thermal and multispectral imagery. This expanded data set allows for a more thorough analysis of power line components, identifying potential issues such as overheating, corrosion, or other anomalies that may not be immediately visible.

Furthermore, the real-time capabilities of drone technology provide instant insights into the condition of power lines. Live video feeds enable operators and inspectors to make informed decisions on the spot, facilitating quicker response times to emerging issues. This real-time monitoring capability is a substantial improvement over the intermittent and delayed feedback inherent in traditional inspection methods.

Drones also contribute to predictive maintenance strategies by collecting historical data over time. This data can be analyzed to identify patterns and trends, allowing utility companies to proactively address potential problems before they escalate. Predictive maintenance not only reduces the risk of unexpected failures but also extends the lifespan of power line infrastructure, resulting in long-term cost savings.

Additionally, the integration of artificial intelligence (AI) and machine learning algorithms enhances the analytical capabilities of drone-collected data. These technologies can automatically detect and classify defects or abnormalities, further streamlining the inspection process and ensuring a higher level of accuracy in identifying potential risks.

The environmental impact of power line inspections is also mitigated through drone usage. Traditional methods often involve transportation of more personnel and heavier equipment to remote or difficult-to-reach locations, leading to increased carbon emissions.

Drones, on the other hand, significantly reduce the need for such travel, since lighter equipment is being transported and potentially a smaller team is needed. Therefore, the overall ecological footprint of power line inspection operations is minimized.

In summary, the adoption of drones in power line inspection not only improves safety and operational efficiency but also opens up avenues for advanced data analytics, real-time monitoring, predictive maintenance, and environmentally conscious practices. The combination of drone technology with sophisticated sensors and AI capabilities positions power line inspections at the forefront of innovation within the utility industry.

1.1.3   Sensor Technologies

Power line inspection drones utilize a variety of sensors and technologies to acquire comprehensive data for assessing the condition of infrastructure. Among the key data types acquired are RGB imagery, providing standard color images for visual inspection and identification of issues such as damaged equipment, vegetation encroachment, or structural anomalies.

Thermal imagery, captured by infrared cameras, detects abnormal temperature patterns along power lines and equipment. Elevated temperatures may signal potential electrical issues or equipment failures. A visual representation of this thermal analysis is illustrated in Fig. 1.1, providing a sample image that highlights temperature variations along power lines.

LiDAR sensors contribute by generating highly accurate 3D point clouds, facilitating the creation of detailed maps of power lines, terrain, and structures. This technology proves particularly valuable in analyzing spatial relationships between objects and assessing topography. An illustrative example of the power of LiDAR technology is showcased in Fig. 1.2.

Combining LiDAR and RGB imagery enables the creation of vegetation maps, aiding in the identification of areas where plants may pose a hazard to power lines. This information guides maintenance planning for vegetation clearance, ensuring the safety and reliability of the power infrastructure.

Digital Elevation Models (DEMs) derived from LiDAR data offer accurate representations of terrain and ground elevation. These models enhance landscape analysis and provide insights into potential challenges related to accessing certain areas.

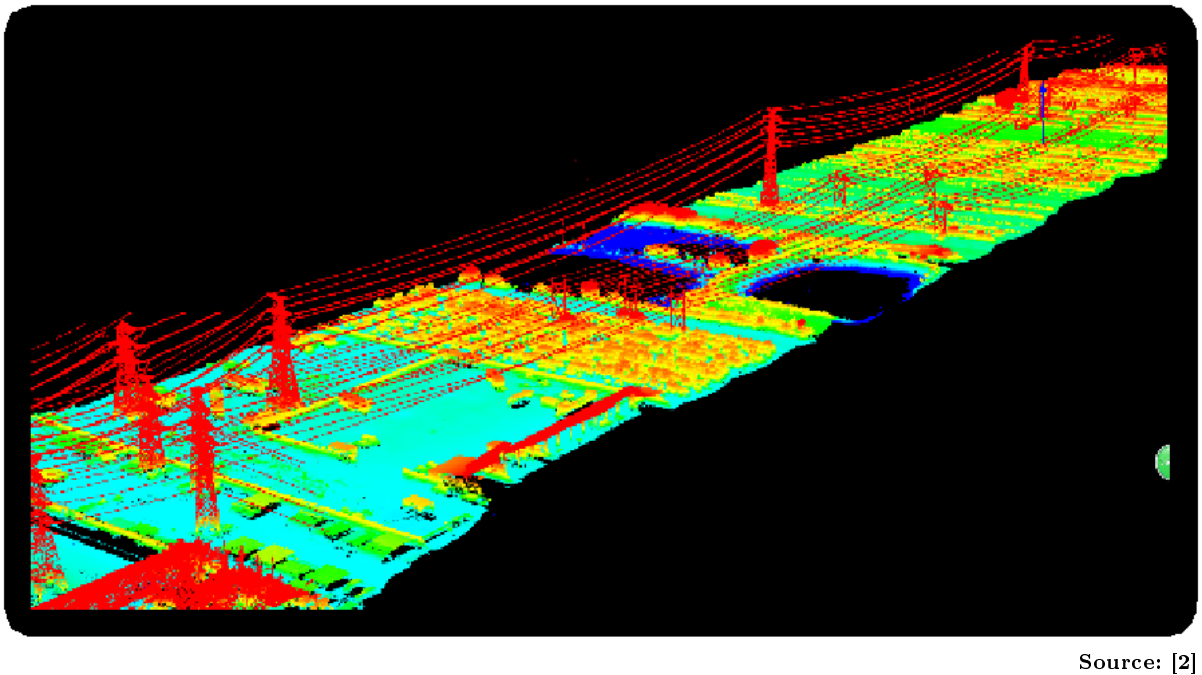Figure 1.1: Thermal Image of High Voltage Power Lines

Figure 1.2: Lidar Image of High Voltage Power Lines

Orthomosaic maps, created by stitching together high-resolution images, offer georeferenced, detailed maps of the surveyed area. These maps provide a broader perspective for overall analysis and documentation, contributing to a more holistic understanding of the power line infrastructure.

Drones equipped with specialized sensors for corrosion inspection contribute to identifying potential issues with the coating on power line structures. This aspect is crucial for maintaining the structural integrity of the infrastructure, ensuring its long-term reliability.

Combining data from these sources allows for a thorough evaluation of power line infrastructure. This, in turn, helps utility providers in recognizing potential issues, strategizing maintenance activities, and maintaining the reliability and safety of their systems.

1.2 Related Work

Power line inspection in a regular basis and timely maintenance are considered as necessary tasks for running an efficient electricity distribution network. A thorough power line inspection for fault detection is a proactive step that can enhance the reliability of electricity distribution network operators. Existing methods for this task include on-theground staff, low-flying helicopters, and/or crawling robots [3], [4], [5], [6]. These solutions are time-expensive, dependent on human visual observation skills and sometimes dangerous for the crew because they have to perform intensive inspection at distances close to the power lines. As an alternate solution, airborne light detection and ranging (LiDAR) technology can be used [7], [8], [9], but it usually comes with high cost. On the other hand, the use of unmanned aerial vehicles (UAVs) can provide robust solutions for an efficient power line inspection. More specifically, offline inspection can be performed using UAVs equipped with camera sensors collecting image or video data of the power lines network. The cost of this kind of solution can be quite low, the safety of the crew high and the visual information that is collected can be detailed because the UAVs can fly relatively close to the power lines [10], [11]. However, it is still challenging to design intelligent, autonomous power line inspection systems using UAVs capable to process online the imagery data and provide output about faults existence. Since the power line detection task is formulated as a binary classification problem, i.e. each pixel is labeled as line or non-line, the presence of background noise in UAV videos, caused by leaves, grass and poor lighting conditions, makes the detection a quite complex task. Many studies propose computer vision techniques for lines detection in the same or similar application field [12], [13], [14],

[15], [16]. In [17], a line segment detection (LSD) method based on an adaptive Gaussian pyramid technique was proposed to initially identify candidate regions for power lines. Following this, a combination of Gaussian mixture model (GMM) and weighted region adjacency graph (WRAG) was employed to construct an object-based Markov random field (OMRF). The final step involved utilizing the least-squares method for fitting power lines. Despite these efforts, traditional methods like these exhibit poor portability and robustness across different scenarios, and their real-time performance and accuracy often do not meet the requirements of specific tasks.

It is a fact that the accuracy of lines detection along with the complexity of the proposed methodologies are significantly related to the reliability of the proposed intelligent inspection systems. Under this consideration, the use of Convolutional Neural Networks (CNNs) seem to contribute to both accuracy and time efficiency. There have already been several attempts to detect power lines using deep learning models. According to the literature, it is common to train a deep learning model using power line images and their corresponding binary masks. As a next step, the pre-trained network is able to predict a binary mask for all new images [18]. Work in [19] utilizes the GoogleNet pre-trained model, and retrain only the last layers exploiting the power line images dataset. Another approach, combines the final layer's output along with the feature maps in each stage to produce highlevel predictions [20]. A feed-forward fully CNN-based architecture, LS-Net, is also presented, achieving a performance of 21.5 frames per second on a state-of-the-art GPU [21]. In their work [22], Yang et al. introduced a power line segmentation network using an encoder-decoder architecture, enabling end-to-end extraction of power lines from aerial images. Abdelfattah et al. proposed PLGAN for power-line segmentation in aerial images, employing adversarial training for accurate predictions. Despite its effectiveness, PLGAN is task-specific and not a versatile segmentation model [23].

For the specific application of the power lines inspection for the Greek Electricity Distribution Network, it is important to point out that it is performed in a yearly basis as a preventive maintenance procedure, while corrective maintenance is performed in cases of storms or fires. The pilot site of Chania area is characterized by cases of heavy load or places with high vegetation, both requiring more frequent inspection in order to check the condition of the network. So, regular inspections in an automated, intelligent way is important in order to ensure a secure and uninterrupted network operation.

In the present study, the dilated convolutional layers technique is exploited, as a more accurate way, among other solutions, to gather context without reducing feature map

size. To that end, two semantic segmentation networks that make use of dilated con-
volutional layers were implemented: D-LinkNet [24] and DeepLabv3 [25]. Also, a Fully
Convolutional Network (FCN) structure similar to the original one presented in [26], has
been constructed, with the difference that some of its convolutional layers were replaced
by a dilation block similar to D-LinkNet's. Finally, for comparison purposes, UNet [27],
which is among the most popular architectures for semantic segmentation, is also trained
and tested using the same datasets. Parts of this work have appeared in [28], [29].

1.3   Purpose, Objectives & Innovation

1.3.1   Purpose

The primary purpose of this research is to address the essential need for efficient
power line inspection within electricity distribution networks. Traditionally carried out
via manned helicopters or ground patrols, this study seeks to examine the capabilities of
unmanned aerial vehicles (UAVs) equipped with camera sensors, aiming to revolutionize
the landscape of autonomous and intelligent inspection. By leveraging state-of-the-art
deep learning architectures, specifically focusing on semantic segmentation techniques,
the goal is to enhance the accuracy of power line detection in acquired imagery. This
research is dedicated to actively contributing to the development of automated methods
that streamline the inspection process, ensuring the uninterrupted operation of electricity
distribution networks.

1.3.2   Objectives

The research includes several key objectives aimed at advancing the field of power line
inspection. Firstly, it seeks to evaluate the effectiveness of dilated convolutional layers,
focusing on their ability to provide contextual information without compromising fea-
ture map size. The implementation of two semantic segmentation networks, D-LinkNet
and DeepLabv3, along with a modified Fully Convolutional Network (FCN) and the
widely used UNet, forms another objective. Through thorough evaluation using litera-
ture datasets and real-world videos collected by HEDNO S.A., the study aims to compare
and assess the performance of these architectures in diverse scenarios. Furthermore, the
research aims to identify the architecture that best preserves the thin structure of power

lines while minimizing background noise, thus contributing to the optimization of inspection processes.

1.3.3 Innovation

This research significantly advances power line segmentation from UAV images through the application of deep learning methodologies. Well-established frameworks like UNet, DeepLab, and DLinkNet were skillfully tailored for this particular assignment, demonstrating their flexibility and precision in addressing the distinctive challenges presented. Significantly, the built-in dilated convolution layers in both DeepLab and DLinkNet architectures were purposefully utilized, their suitability for line segmentation tasks being acknowledged, with consideration given to the slender structure of power lines. To further explore the effectiveness of dilated convolutions, a deliberate modification was introduced into a common Fully Convolutional Network (FCN), incorporating these layers. This enhancement aimed to optimize network architectures for superior performance in power line segmentation, particularly in capturing the thin and elongated structure of power lines. The thorough evaluation of these adapted architectures significantly contributes to the exploration of efficient models tailored to this specialized task. Emphasis was placed on evaluating and comparing the efficiency of dilated convolution layers, providing valuable insights into their suitability for linear structures and enriching the broader field of image segmentation. Additionally, an innovative image splitting and concatenation approach was introduced, showcasing its potential to enhance efficiency and accuracy in power line segmentation tasks. Innovation is introduced by emphasizing the use of dilated convolutional layers in deep learning architectures for power line segmentation, showcasing their application in autonomous UAV-based inspections. The introduction of D-LinkNet, originally designed for road segmentation in high-resolution satellite imagery, for power line segmentation represents a novel adaptation of existing architectures, contributing to the field by demonstrating superior performance, especially in various background scenarios. Overall, the research underscores the application of cutting-edge techniques to enhance the efficiency and accuracy of power line inspection processes.

1.4 Thesis Outline

The thesis is organized as follows:

1. In Chapter 2, foundational concepts in image segmentation and deep learning are introduced. It covers various types of image segmentation tasks and traditional approaches. The basics of deep learning and its role in image segmentation are explained.

2. Chapter 3 outlines the research approach, beginning with an overview of the datasets used for evaluation. It proceeds to explore key components of the chosen architectures, such as dilated convolution, DeepLab, and D-LinkNet, along with their practical implementation.

3. Chapter 4 focuses on the experimental results. It provides a comprehensive overview, revealing the outcomes of using deep learning networks for segmentation. Each architecture is thoroughly examined in different settings, applying evaluation metrics to assess model performance.

4. Chapter 5 conclusively wraps up the thesis by presenting a concise summary of the pivotal findings, capturing the core contributions and insights derived from the research.

Chapter 2

Background

2.1 Image Segmentation

Image segmentation involves dividing a digital image into segments, reducing its complexity for subsequent processing. It entails assigning labels to pixels to identify objects, people, or other significant elements within the image.

Image segmentation is frequently employed in object detection, where the approach involves using a segmentation algorithm to identify objects of interest within the image. This enables subsequent operation of an object detector on a pre-defined bounding box established by the segmentation algorithm. This strategy enhances accuracy and reduces inference time by avoiding the need to process the entire image.

Furthermore, image segmentation is a fundamental component in various computer vision technologies and algorithms. It finds practical use in medical image analysis, computer vision for autonomous vehicles, face recognition and detection, video surveillance, as well as satellite image analysis.

2.1.1 Types of Image Segmentation tasks

Semantic segmentation

Semantic segmentation is the process of classifying individual pixels in an image into distinct semantic categories. Each pixel is exclusively assigned to a specific class without considering additional context.

However, this task becomes challenging when there are closely grouped instances of the same class in an image, leading to a less detailed prediction. For example, in an image
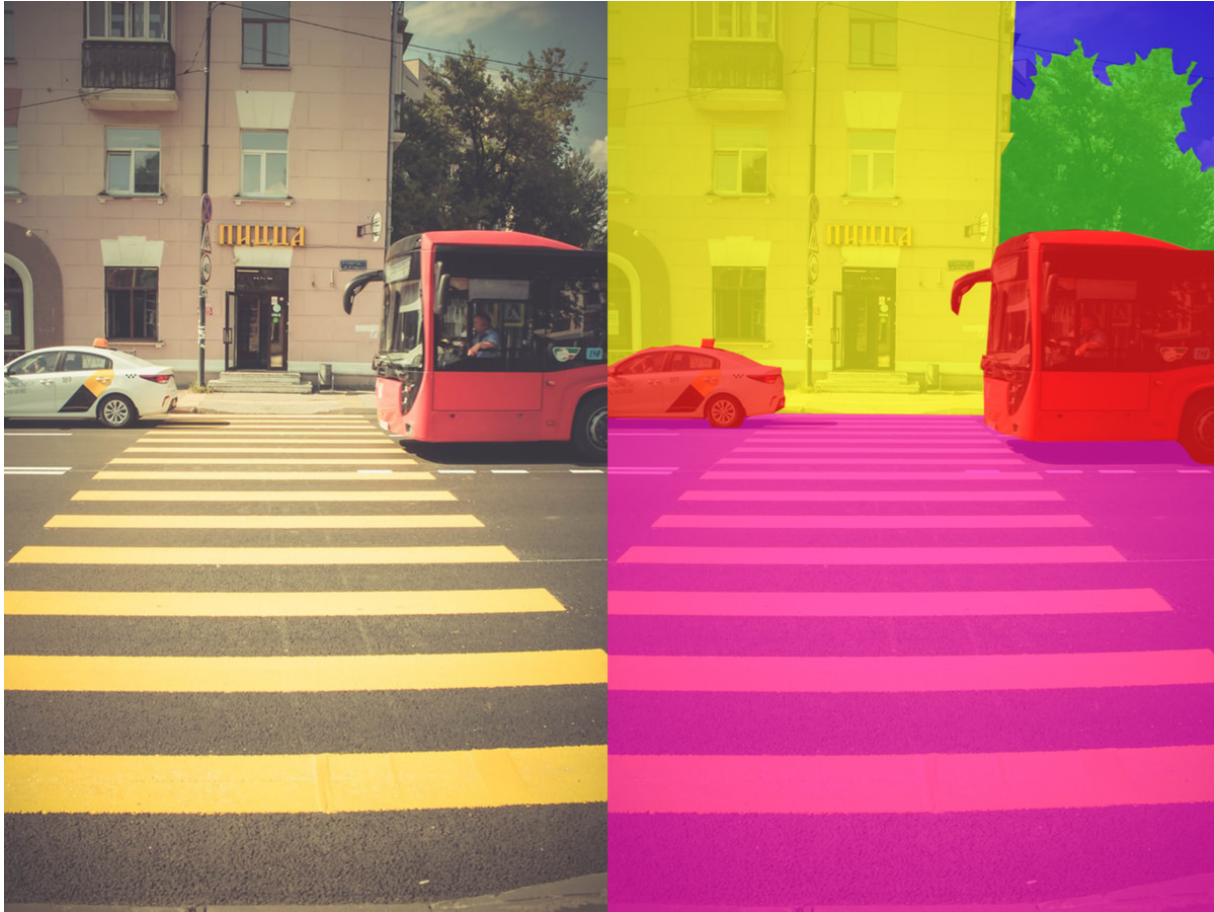
Figure 2.1: Semantic Segmentation

of a crowded street, a semantic segmentation model may predict the entire crowd region as belonging to the "pedestrian" class, providing limited in-depth information about the image.

In Fig. 2.1, we observe an exemplary instance of semantic segmentation, where the pixels have been categorized into distinct semantic classes. Notably, the model has successfully identified and labeled five specific classes: road, cars, the sky, buildings, and trees. This segmentation allows for a granular understanding of the scene, enabling the recognition and differentiation of specific objects and elements within the image.

Instance segmentation

Instance segmentation models categorize pixels based on instances rather than predefined classes.

Unlike classification models, instance segmentation algorithms do not identify the class of a segmented region. Instead, they excel at distinguishing between overlapping or similar
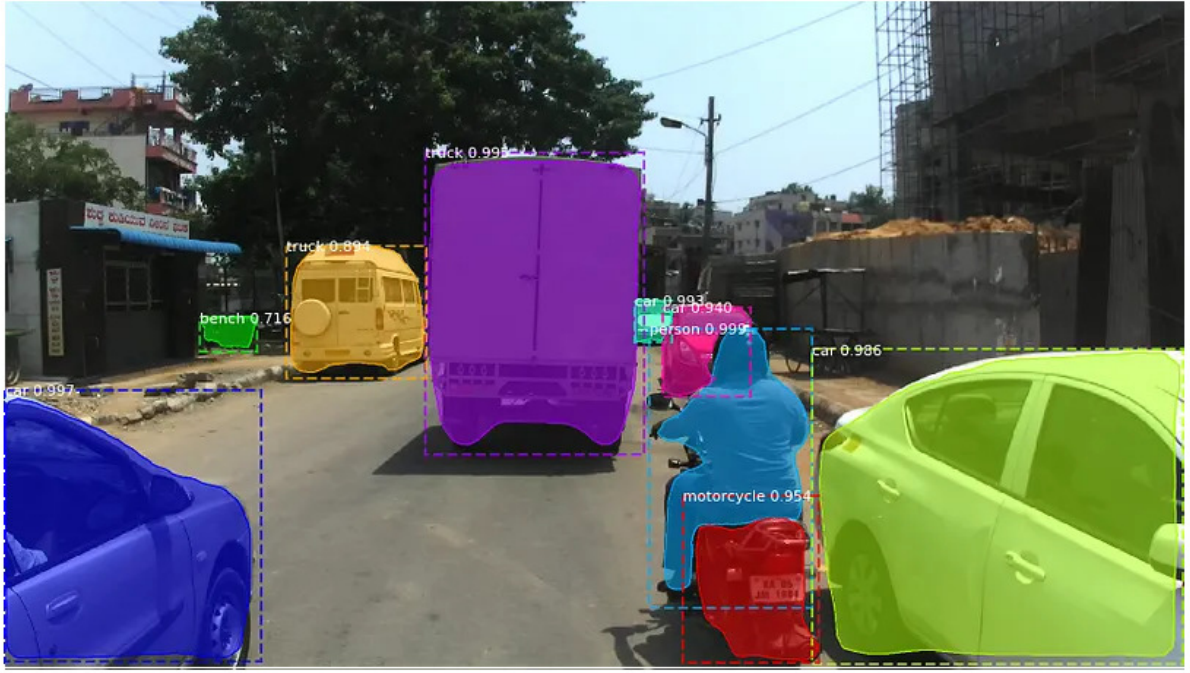
Figure 2.2: Instance Segmentation

object regions by focusing on their boundaries.

In the captivating road scene captured in Fig. 2.2, the instance segmentation model was able to segregate various objects, including cars, trucks, motorcycles, and even a bench.

Panoptic segmentation

Panoptic segmentation, the latest advancement in segmentation tasks, can be expressed as the combination of semantic segmentation and instance segmentation. In this approach, the goal is to not only categorize the pixels in an image semantically but also to delineate each instance of an object and predict its identity.

Panoptic segmentation algorithms have broad applications, particularly in tasks such as self-driving cars, where it is crucial to capture extensive information about the surrounding environment through a continuous stream of images. An example of panoptic segmentation is illustrated in Fig. 2.3, where the distinct regions for object instances and stuff classes are seamlessly segmented within the scene.

Figure 2.3: Panoptic Segmentation

2.1.2 Traditional Approaches to Image Segmentation

Edge-Based Segmentation

Edge-based segmentation is a widely used method in image processing that aims to detect and highlight the boundaries of different objects within an image. This method is essential for identifying important aspects of objects by using details obtained from their edges. By identifying edges, redundant information in images can be eliminated, leading to a reduction in size and facilitating more efficient analysis.

Edge-based segmentation algorithms identify edges based on contrast, texture, color, and saturation variations. They can accurately represent the borders of objects in an image using edge chains comprising the individual edges. Fig. 2.4 shows an example of edge-based segmentation with thresholding.

Threshold-Based Segmentation

Thresholding stands out as the most straightforward image segmentation approach, categorizing pixels depending on their intensity relative to a specified threshold value. This method is effective for separating objects with higher intensity from backgrounds or other objects. The threshold value, denoted as T, can remain constant in scenarios with low noise. Alternatively, dynamic thresholds may be employed in certain cases. In essence,

Figure 2.4: Edge-Based Segmentation

thresholding transforms a grayscale image into two segments by classifying pixels based on their relationship to the threshold, resulting in a binary image. An example of image segmentation by the thresholding method is shown in Fig. 2.5.

Region-Based Segmentation

Region-based segmentation algorithms operate on the principle that pixels within the vicinity of a particular region exhibit similar intensity values. In general, region-based segmentation is a pixel-wise operative algorithm in which each pixel is compared to its neighborhood pixels. If a predefined homogeneity condition is met, the pixel and its neighbors are categorized as part of a specific object. Region-based techniques can be divided into two categories: region growing and region splitting-merging algorithms.

The region growing segmentation starts with a set of seed points as the initial region hypothesis and then proceeds to grow the regions by adding adjacent pixels to the regions based on a predefined criteria. This process continues until no more adjacent pixels meet the criteria. An example of region growing segmentation is depicted in Fig. 2.6. The effectiveness of region growing algorithms is influenced by the initialization of seed points and the similarity criteria.

On the other hand, the region splitting-merging algorithm is an iterative approach to region-based segmentation, following a top-to-bottom strategy. The splitting-merging algorithm compares adjacent regions and merges them if they are homogeneous. It starts

Figure 2.5: Edge-Based Segmentation

with the whole image as a single region and then splits it into subsequent four disjoint quadrants. This process iteratively continues for each subsequent region until pixels within the regions no longer satisfy a certain similarity constraint, such as gray-level homogeneity. The splitting-merging algorithm generates adjacent regions with very similar properties.

Cluster-Based Segmentation

Clustering algorithms serve as unsupervised classification techniques designed to unveil latent patterns within images, enhancing human perception by discerning clusters, shades, and structures. These algorithms operate by partitioning images into clusters of pixels sharing analogous features, effectively segregating data elements and assembling akin elements into coherent clusters. This process, known as clustering segmentation, facilitates the extraction of meaningful information by isolating regions of similarity, thereby aiding in the comprehensive analysis and interpretation of complex visual data.

Watershed Segmentation

Watersheds are transformations applied to grayscale images. In watershed segmentation algorithms, images are conceptualized as topographic maps, where the brightness of each pixel dictates its elevation or height. This approach identifies ridges and basins in the image, delineating the regions between watershed lines. The algorithm effectively partitions the image into multiple segments based on pixel intensity, grouping together pixels that share similar gray values.
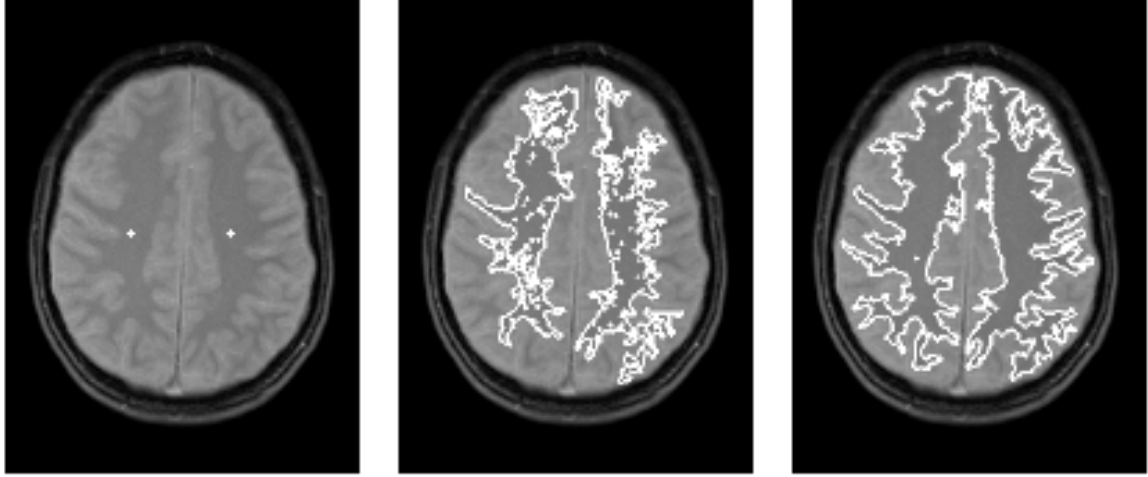
Figure 2.6: Region-Based Segmentation

The watershed technique finds significant applications, particularly in medical image processing. For instance, in the analysis of MRI scans, it proves valuable in distinguishing variations between lighter and darker regions, potentially aiding in the diagnostic process by highlighting important features within the medical imagery.

2.2   Deep Learning

2.2.1   Introduction to Deep Learning

In recent years, the field of artificial intelligence has undergone a revolutionary transformation, driven largely by the advent of deep learning. Deep learning is a branch of machine learning which is based on artificial neural networks. It has demonstrated remarkable success in solving complex problems across various domains, ranging from natural language processing to image recognition and beyond. As shown in Fig. 2.7, the diagram illustrates the relationships and interdependencies among key concepts in the fields of Artificial Intelligence (AI), Machine Learning (ML), Data Science, and Deep Learning.

Deep learning is based on the concept of neural networks, which are computational models inspired by the structure and functioning of the human brain. Nodes, or artificial neurons, are organized into layers, forming a network that learns and extracts features from data. Training involves adjusting the weights and biases of these connections to minimize the difference between predicted and actual outcomes.
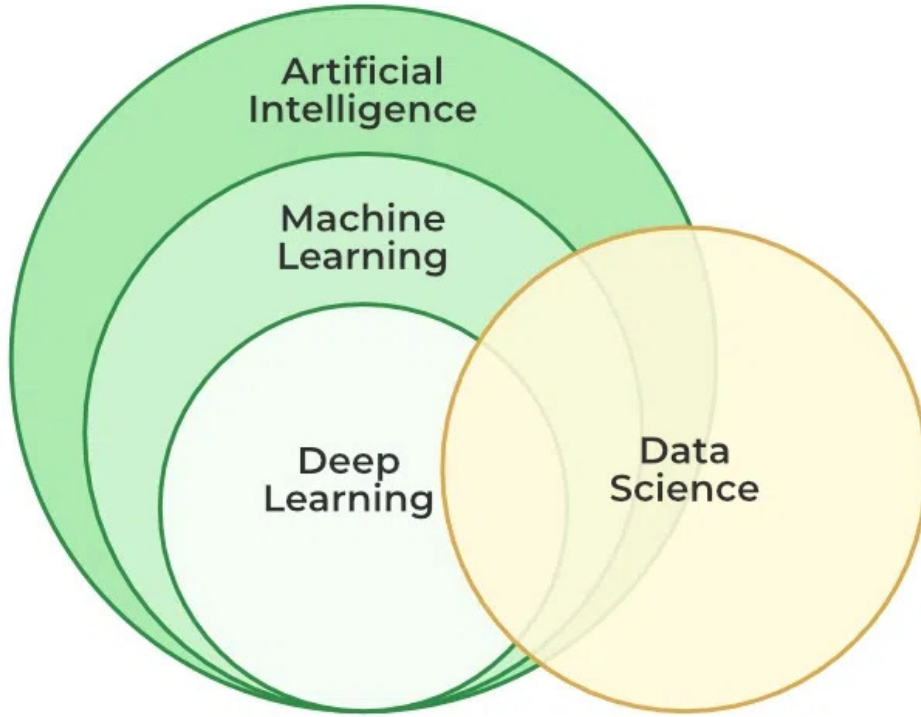
Figure 2.7: Interconnected Concepts: AI, ML, Data Science, and Deep Learning

Various neural network architectures exist, each tailored to specific tasks. Convolutional neural networks (CNNs) excel in image processing, recurrent neural networks (RNNs) are effective for sequence data, and transformer architectures have shown great promise in natural language processing tasks. Understanding the architecture that suits the problem at hand is crucial for achieving optimal results.

The training of deep neural networks involves feeding them with labeled data, adjusting weights through backpropagation, and iteratively optimizing the model. Deep learning frameworks, such as TensorFlow and PyTorch, have simplified the implementation and training of complex neural network architectures.

Deep learning's adaptability has resulted in its extensive use across various applications, such as speech recognition, natural language processing, autonomous vehicles, and healthcare.

2.2.2 Deep Neural Networks for Image Segmentation

Image processing involves the manipulation and analysis of visual data, and deep learning has proven to be a transformative force in this field. Traditional image processing

techniques often rely on handcrafted features, whereas deep learning enables the automatic extraction of relevant features directly from the data.

Convolutional Neural Networks (CNNs) have become the fundamental building blocks of deep learning in image processing. Their ability to automatically learn hierarchical representations of visual features makes them well-suited for tasks such as image classification, object detection, and image segmentation. Understanding the architecture of CNNs and their application in image processing is essential for harnessing their power effectively.

Image Classification

One of the primary applications of deep learning in image processing is image classification. Deep neural networks can learn to classify images into predefined categories, achieving human-level or even superhuman performance in tasks like recognizing objects, animals, and scenes. The chapter explores the principles behind image classification using deep learning and delves into real-world examples.

Object Detection

Object detection involves identifying and locating objects within an image. CNNs equipped with techniques like region proposal networks and anchor boxes have demonstrated exceptional accuracy in object detection tasks. This section discusses the challenges and advancements in deep learning-based object detection, including applications in fields like autonomous vehicles and surveillance.

Image Segmentation

Image segmentation divides an image into meaningful segments, assigning each pixel to a specific category. Deep learning techniques, especially convolutional neural networks with skip connections, have significantly improved the accuracy and efficiency of image segmentation. The chapter explores the principles of image segmentation and its applications in medical imaging, satellite imagery analysis, and more.

Chapter 3

Approach

This research has made significant contributions to the domain of power line segmentation from UAV images through the application of deep learning methodologies. Established networks such as UNet, DeepLab, and DLinkNet were systematically adapted for the specific task of power line segmentation, highlighting their adaptability and fine-tuning for a novel use case. Notably, a deliberate choice was made to leverage the inherent inclusion of dilated convolution layers in both DeepLab and DLinkNet architectures, as they are deemed ideal for line segmentation tasks. Furthermore, to explore the effectiveness of dilated convolutions, a modification was introduced into a common Fully Convolutional Network (FCN), incorporating these layers. This strategic enhancement aimed to optimize network architectures for enhanced performance in power line segmentation. The thorough evaluation of these adapted architectures contributes substantially to the exploration of efficient models tailored to this specialized task. The focus on the evaluation and comparison of the efficiency of dilated convolution layers in power line segmentation provides valuable insights into their suitability for linear structures, enriching the broader field of image segmentation. Additionally, an innovative image splitting and concatenation approach was introduced, demonstrating its potential to improve efficiency and accuracy in power line segmentation tasks. In summary, this work represents a novel and tailored solution to the real-world challenge of power line segmentation from UAV images, emphasizing the strategic choice of leveraging dilated convolution layers inherent in DeepLab and DLinkNet architectures, along with the deliberate inclusion in the modified FCN, for their efficiency in capturing linear structures.

3.1 Power Line Datasets

3.1.1 Literature Datasets

Two publicly available datasets were used jointly to train the suggested network architectures, namely the PLDU and PLDM. The former consists of urban scenes, while the latter one consists of mountain scenes [20]. Sample images of both datasets along with their annotations are shown in Fig. 3.1 & Fig. 3.2. The annotation of the training dataset was performed by the online tool LabelMe [37]. All annotated images were converted into binary masks.

(a)   (b)

(c)   (d)
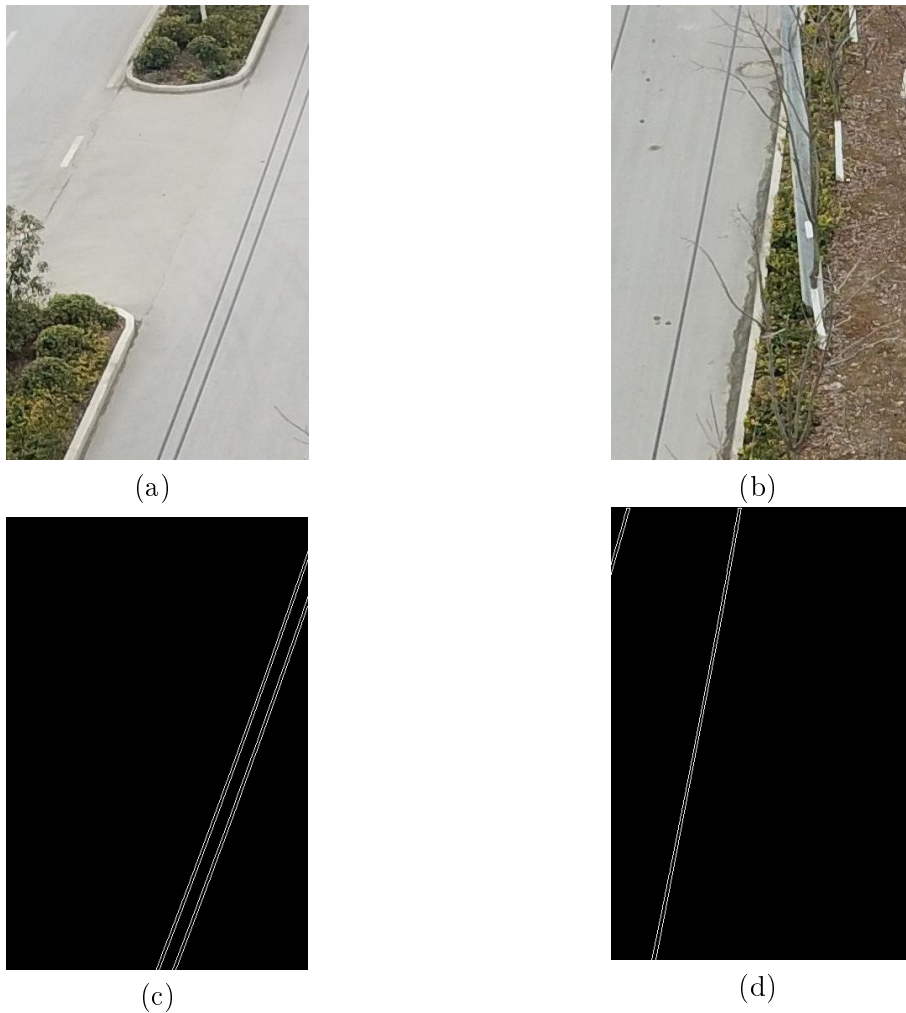
Figure 3.1: PLDU dataset samples, (a), (b) Power line images & (c), (d) corresponding binary masks

(a)

(b)

(c)

(d)
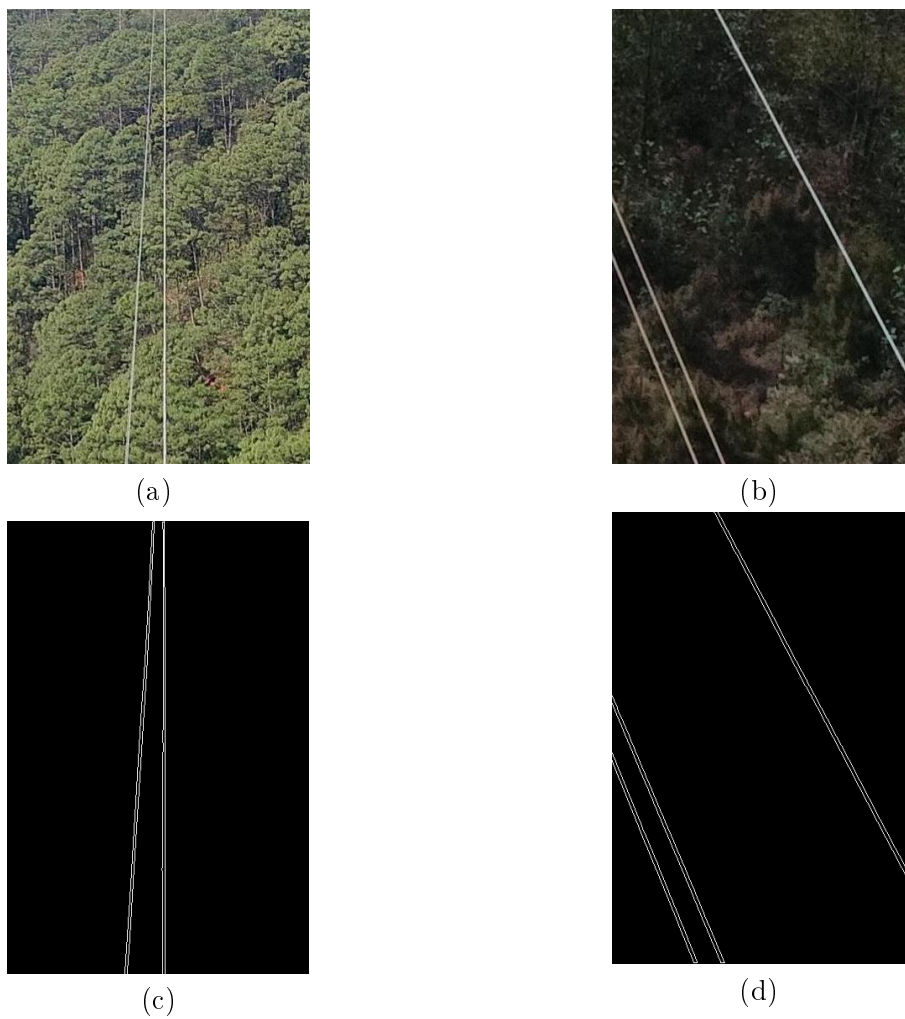
Figure 3.2: PLDM dataset samples, (a), (b) Power line images & (c), (d) corresponding binary masks

3.1.2   HEDNO Dataset

Apart from the test images of the PLDM and PLDU datasets, the trained networks were also tested on videos provided to us by the HEDNO S.A. (Hellenic Electricity Distribution Network Operator S.A.) department that administers the network in Chania area, Crete island, Greece. The videos were acquired using UAVs and the selected site for this study is characterized by thick vegetation. Sample frame of the videos acquired by HEDNO S.A. is shown in Fig. 3.3.

The frames within the dataset were annotated using the same annotation tool employed for annotating the literature datasets. Fig. 3.4 provides a screenshot showcasing the annotation environment.

3.1.3   Data Augmentation Techniques

Due to the limited size of the training dataset, we employed data augmentation techniques to artificially increase the diversity of the training samples. Data augmentation is performed on-the-fly during each training epoch, generating slightly modified versions of existing images to enrich the training set. The following augmentation techniques were applied:

- Random Rotation: Images are rotated by a random angle to introduce variations in orientation.

- Flipping: Random horizontal and vertical flips are applied to simulate different viewpoints.

- Zoom: Random zooming is employed to replicate instances where power lines appear at varying distances.

- Contrast Adjustment: Contrast levels are randomly adjusted to account for different lighting conditions.

- Brightness Adjustment: Random adjustments to image brightness help the model generalize across different illumination scenarios.

Dataset Information

Table 3.1 provides detailed information about the composition of our datasets, including the number of samples for both the training and testing sets. Our datasets consist of

(a)

(b)
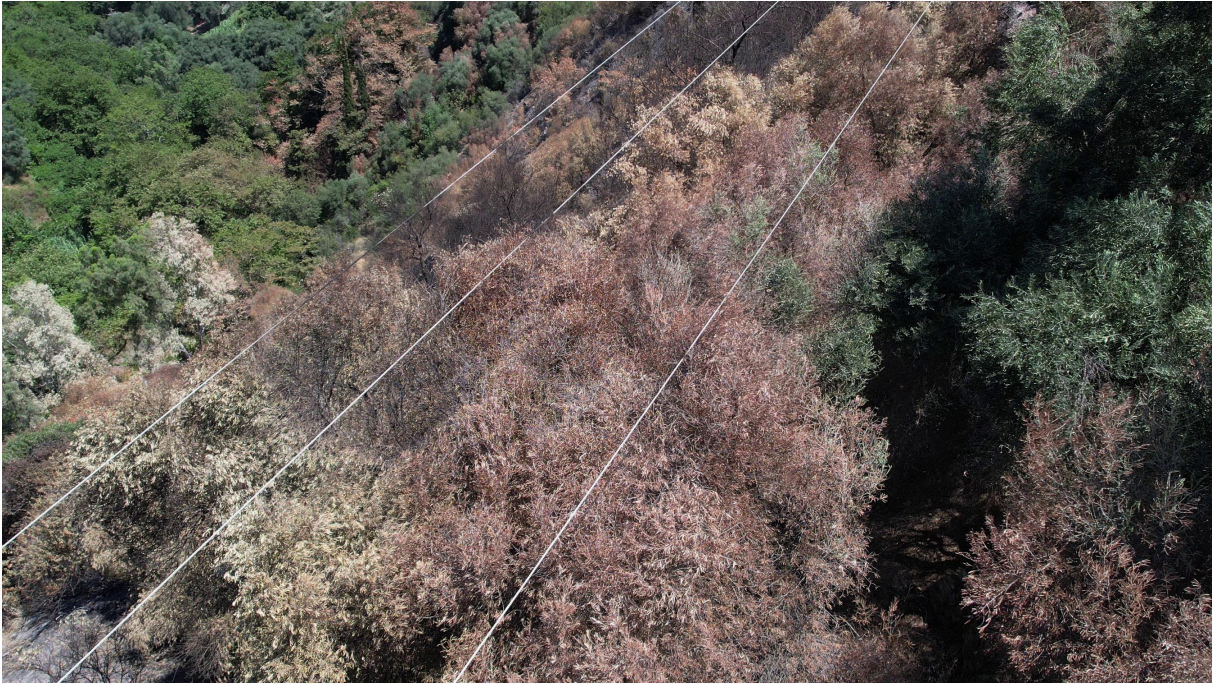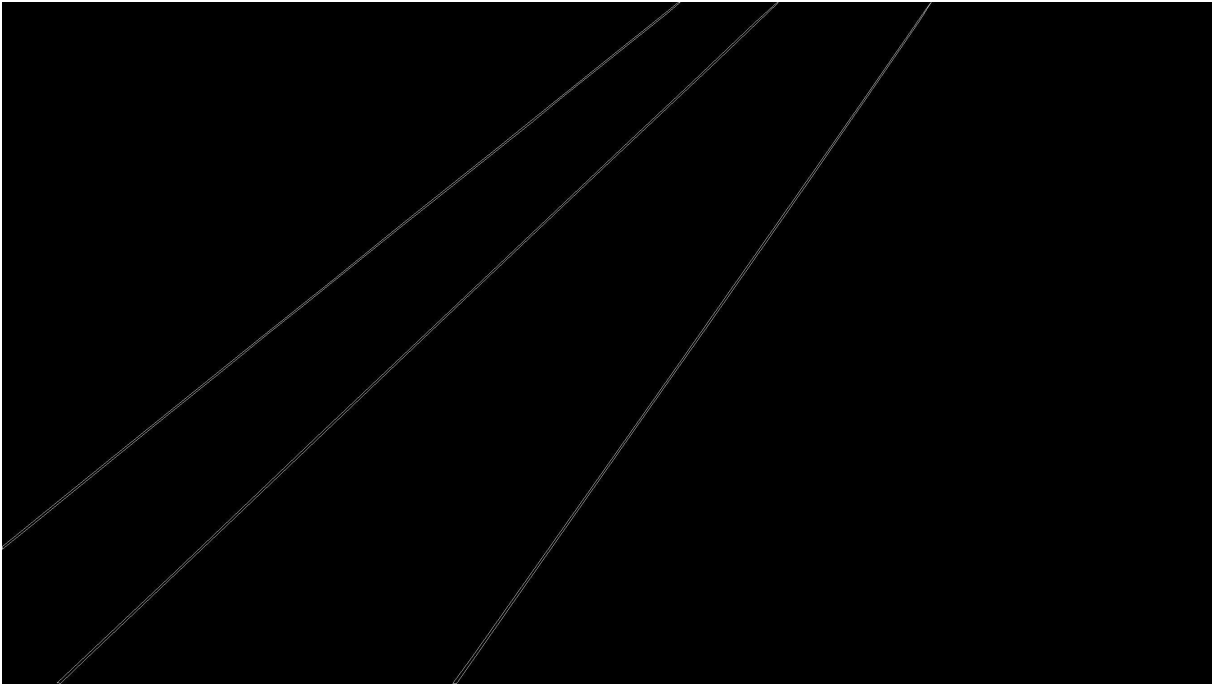
Figure 3.3: Dataset sample, (a) video frame & (b) corresponding binary mask

Figure 3.4: LabelMe Annotation Tool

three main components: PLDU, PLDM, and Video Frames. The PLDU dataset comprises 453 samples for training and 120 samples for testing, while the PLDM dataset includes 237 training samples and 50 testing samples. The Video Frames dataset consists of 81 samples for training and 15 samples for testing.

To create a comprehensive dataset for our experiments, we combined these individual datasets into a Final Dataset, which includes 771 samples for training and 185 samples for testing. The training set is further enriched through on-the-fly data augmentation in each training epoch, resulting in a dynamically augmented training set equivalent to the product of the epoch number and the original dataset size (Epoch Number * 771), while the testing set remains unchanged at 185 samples.

This diverse dataset, along with the augmentation strategies, ensures a robust training process and thorough evaluation of our models' performance across different scenarios.

Table 3.1: Dataset Information [28]

Dataset	Train Samples	Test Samples
PLDU	453	120
PLDM	237	50
Video Frames	81	15
Final Dataset	771	185
Augmented Dataset	Epoch Number * 771	185

3.2   Architecture

3.2.1   Dilated Convolution

Pooling layers are a common practice in convolutional architectures to reduce dimensionality, preserving relevant features and removing irrelevant details. However, in power line detection and segmentation, this approach is unsuitable due to the narrow width of power lines, risking the loss of crucial segments. Traditional pooling, being irreversible, may compromise spatial information, particularly for small objects. As an alternative to pooling layers, dilated convolutional layers [38] emerge as a powerful tool for preventing the loss of resolution. By expanding the convolutional kernel through strategically placed holes, dilated convolution covers a larger area of the input at each step, preserving fine details. Figure 3.5 illustrates the comparison between dilated convolution and normal convolution.

An additional parameter l (dilation factor) tells how much the input is expanded. In other words, based on the value of this parameter, $(l-1)$ pixels are skipped in the kernel. In essence, normal convolution is just a 1-dilated convolution.

Convolution equation:

$$(f * g)[n] = \sum_{k=-\infty}^{\infty} f[k] \cdot g[n-k] \tag{3.1}$$

Dilated convolution equation with dilation factor l:

$$(f *_{\text{dilated}} g)[n] = \sum_{k=-\infty}^{\infty} f[kl] \cdot g[n-kl] \tag{3.2}$$

The importance of dilated convolution in power line segmentation is evident. Power lines, characterized by their slim structures, present a unique challenge in computer vision, requiring precise spatial information. Unlike traditional pooling layers that may compromise such details, dilated convolution proves crucial in capturing expansive contextual information without sacrificing resolution. This is especially significant in scenarios where power lines demand meticulous attention to detail, and adaptability to varying scales of structures within an image is paramount. Dilated convolution, by preventing the loss of spatial precision, stands as an indispensable tool for enhancing the efficacy and accuracy of power line segmentation models, establishing itself as a pivotal component in advancing the state-of-the-art in this critical application.
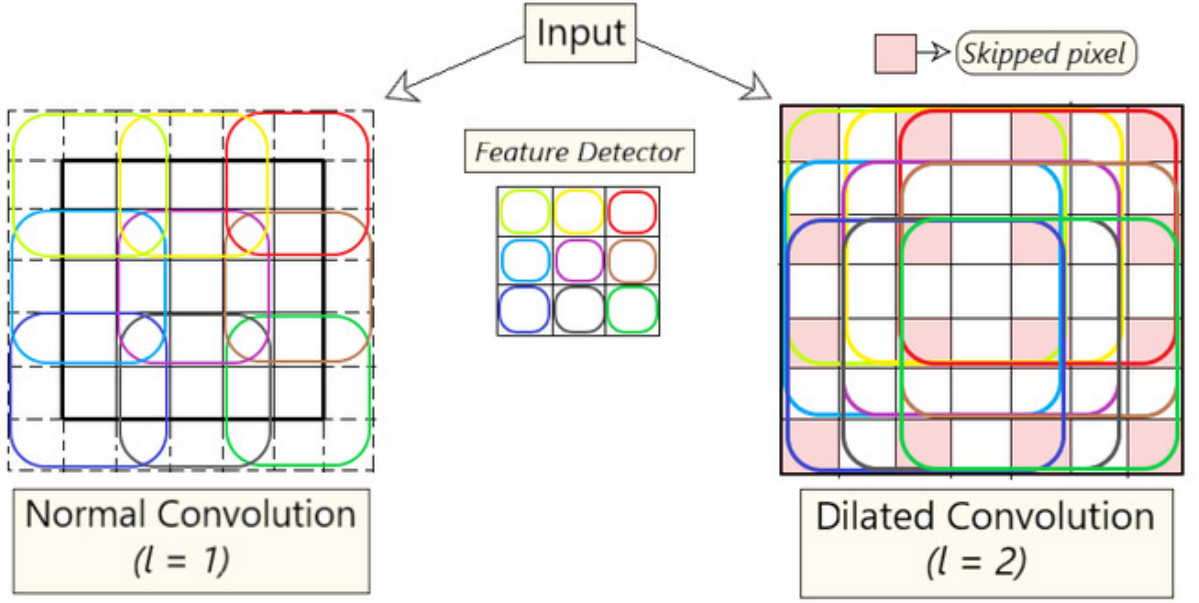
Figure 3.5: Normal Convolution Vs Dilated Convolution

3.2.2   DeepLab

The DeepLab model addresses the challenge of information loss using atrous/dilated convolutions and Atrous Spatial Pyramid Pooling (ASPP) modules. Its architecture has evolved over several versions; namely, DeepLabV1, DeepLabV2 and DeepLabV3. DeepLabV1 [40] and DeeplabV2 [41] use dilated convolutions and Fully Connected Conditional Random Field (CRF) to extend the receptive field of the network and get more contextual information. DeepLabV2 is superior to DeepLabV1 as it makes also use of ASPP modules. DeepLabV3 [25] is the most improved version of DeepLab series. It gets rid of CRF used in V1 and in V2 and it uses an improved ASPP module. DeepLabV3 was tested in this work for the task of power line segmentation and it is illustrated in Fig. 3.6.

3.2.3   D-LinkNet

D-LinkNet architecture, was first introduced in [43] for the purpose of road segmentation, whose structure is identical to power lines. In this work, D-LinkNet is used as a starting point to investigate the effectiveness of dilated convolutional layers in power line segmentation task. The network is based on LinkNet architecture [44], which is considered as an efficient method for semantic segmentation, and has dilated convolution layers in its center part.
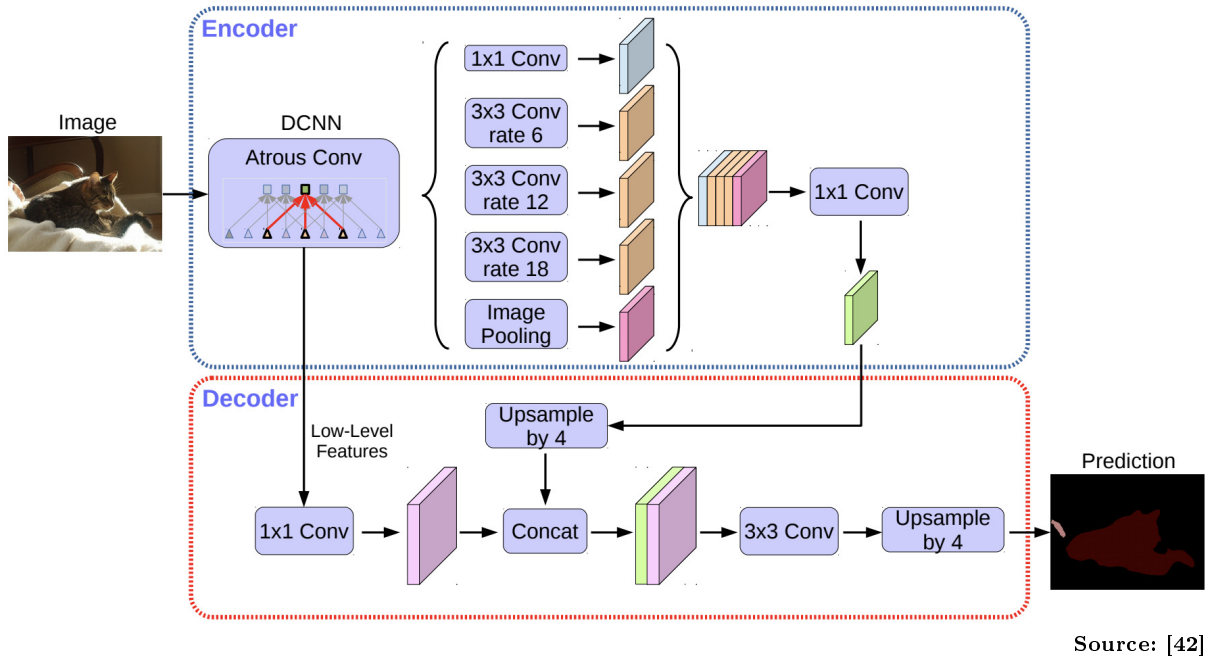
Figure 3.6: DeepLabV3 Architecture
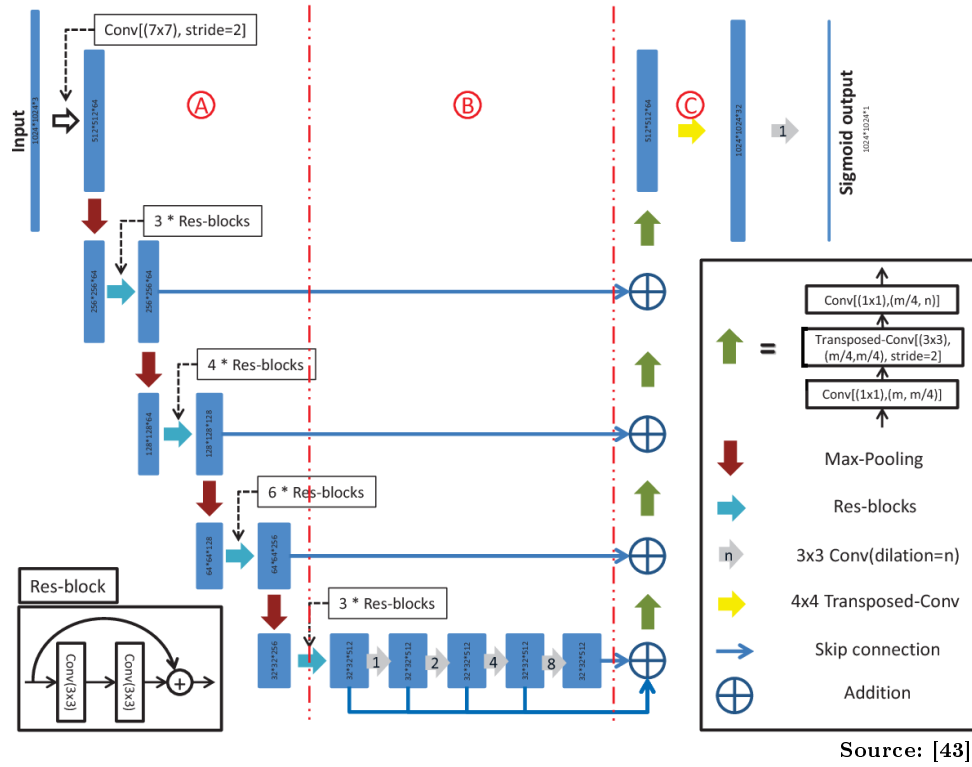
Figure 3.7: D-LinkNet architecture
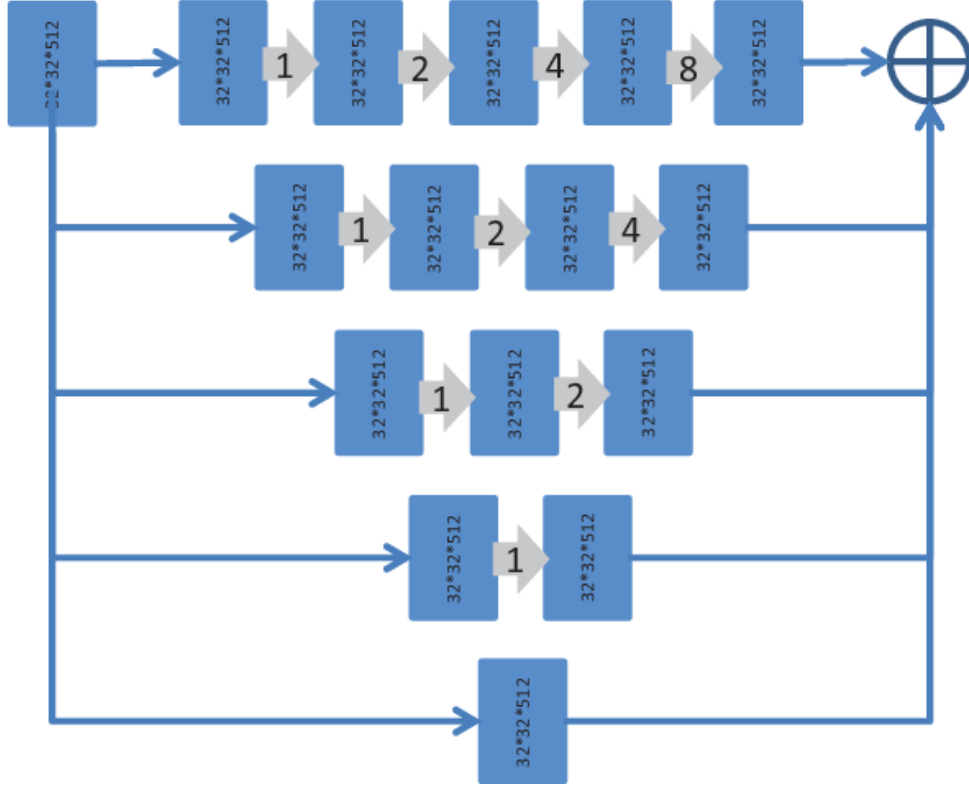
Source: [43]

Figure 3.8: D-LinkNet's center dilation block

D-LinkNet employs ResNet34 [45] pretrained on the ImageNet [46] dataset as its encoder, which consists of five downsampling layers. In the central part, D-LinkNet utilizes dilated convolution layers with dilation rates of 1, 2, 4, and 8. Consequently, the feature points on the last center layer observe a receptive field of 31×31 on the first center feature map, covering the primary portion of the first center feature map. D-LinkNet capitalizes on multi-resolution features, and the central part can be viewed as a parallel mode, as illustrated in Fig. 3.8.

The decoder of D-LinkNet remains consistent with the original LinkNet. It is computationally efficient and uses transposed convolution layers for upsampling, restoring the resolution of the feature map from 32×32 to 1024×1024.

3.2.4 Modified FCN with Dilated Convolution Module

In our experiments, we introduced modifications to the Fully Convolutional Network (FCN) model, specifically targeting the center part of the architecture. The primary innovation involves the incorporation of a dilation block, as illustrated in Figure 3.9. This modification aims to address the challenges associated with power line detection and seg-
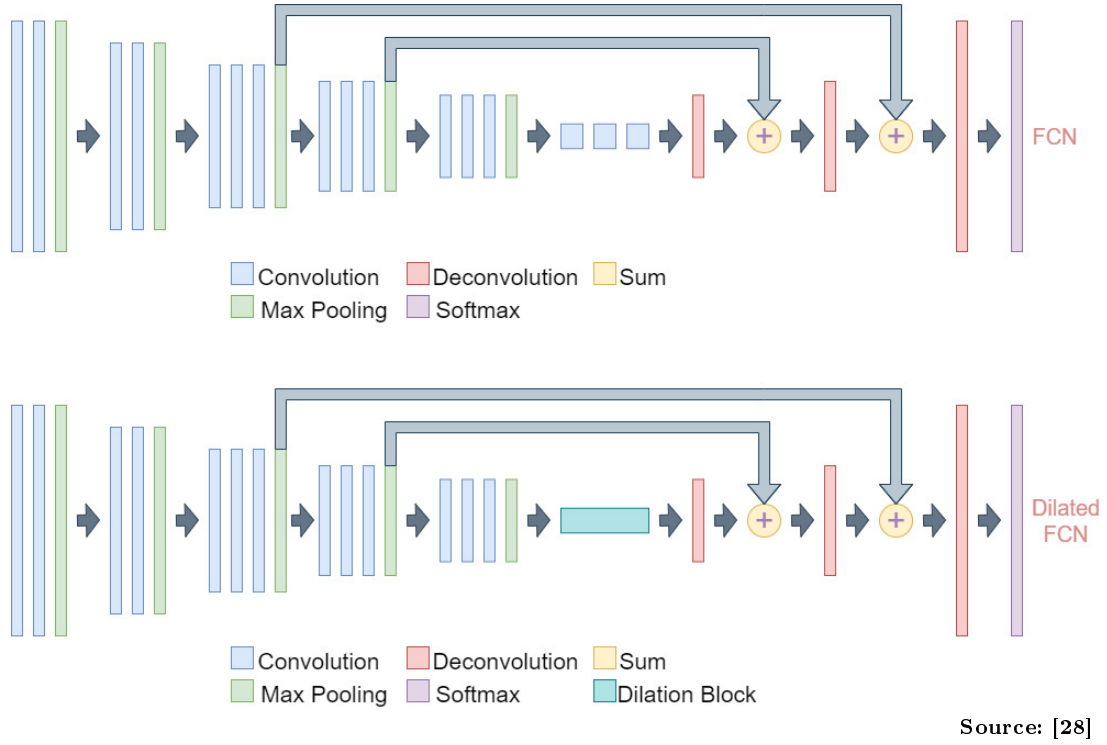
Figure 3.9: Original FCN Architecture vs. Proposed FCN Architecture

mentation by preserving spatial information and preventing the loss of resolution. This alteration proves crucial in scenarios where power lines are only a few pixels wide, ensuring that fine details are retained during the convolution process.

The dilation block in the modified FCN architecture closely aligns with the one in D-LinkNet's architecture, featuring a sequence of dilated convolutional layers with rates of 1, 2, 4, and 8. Each dilated convolutional layer captures a progressively wider contextual field, allowing the model to recognize intricate structures and long-range dependencies in power line images.

In the architecture of the modified Fully Convolutional Network (FCN) dilated convolutional layers play a pivotal role in capturing both local and global contextual information. The network structure consists of encoder blocks, the Dilation Block, decoder blocks, and a final upsampling layer.

The encoder blocks, implemented as *conv*3*x*3_*block*_*x*2, initiate the network by extracting hierarchical features from the input image. These blocks progressively increase the receptive field and channel complexity, capturing low-level and mid-level features. The Dilation block is a critical addition, featuring dilated convolutions with different rates (1, 2, 4, and 8). This allows the model to incorporate a broader range of contextual information, aiding in the recognition of power line structures across varying scales.

Skip connections are established between corresponding encoder and decoder blocks to facilitate the flow of high-resolution features. These connections alleviate the vanishing gradient problem and aid in the precise localization of power lines. The decoder consists of upsampling layers to restore the spatial resolution of the feature maps.

The final upsampling layer increases the resolution to match the original input image size, refining the segmentation output and producing a pixel-wise prediction for power line presence.

The incorporation of dilated convolutional layers in the dilation block serves to capture a larger contextual field without significantly increasing computational complexity. This strategy enables the model to recognize long-range dependencies and intricate structures in power line images, enhancing segmentation performance. Skip connections between encoder and decoder blocks contribute to the model's ability to capture both fine details and global context.

3.3   Segmentation process

3.3.1   Proposed grid approach

The UAV-captured video content is encompassed by a range of perspectives, from close-up shots to long-distance views of power lines. To effectively handle this diversity, a grid segmentation approach is employed. Each frame is broken down into a grid of smaller cells, for instance, a 4x5 grid, effectively dividing the image into multiple sections. Within each of these grid cells, predictions are made using a deep learning network. The network assesses its respective grid cell and predicts a binary mask individually, where each pixel in the mask is assigned a value of "1" to indicate the presence of a power line or "0" to signify background. These individual predictions are crucial for accuracy. The binary masks for all the grid cells of the frame are then combined back together to create a final output binary mask that accurately identifies power lines in the entire image. The versatility of this method lies in the ability to adjust the size of the grid cells based on the distance between the UAV camera and the power lines to be detected. Using a smaller grid cell size results in thicker lines in the output mask, while a larger grid cell size yields thinner and more precise lines. This adaptability allows for accurate power line detection in a wide range of scenarios and distances from the camera. An illustration of the proposed 4x5 grid grid is depicted in Fig. 3.10.
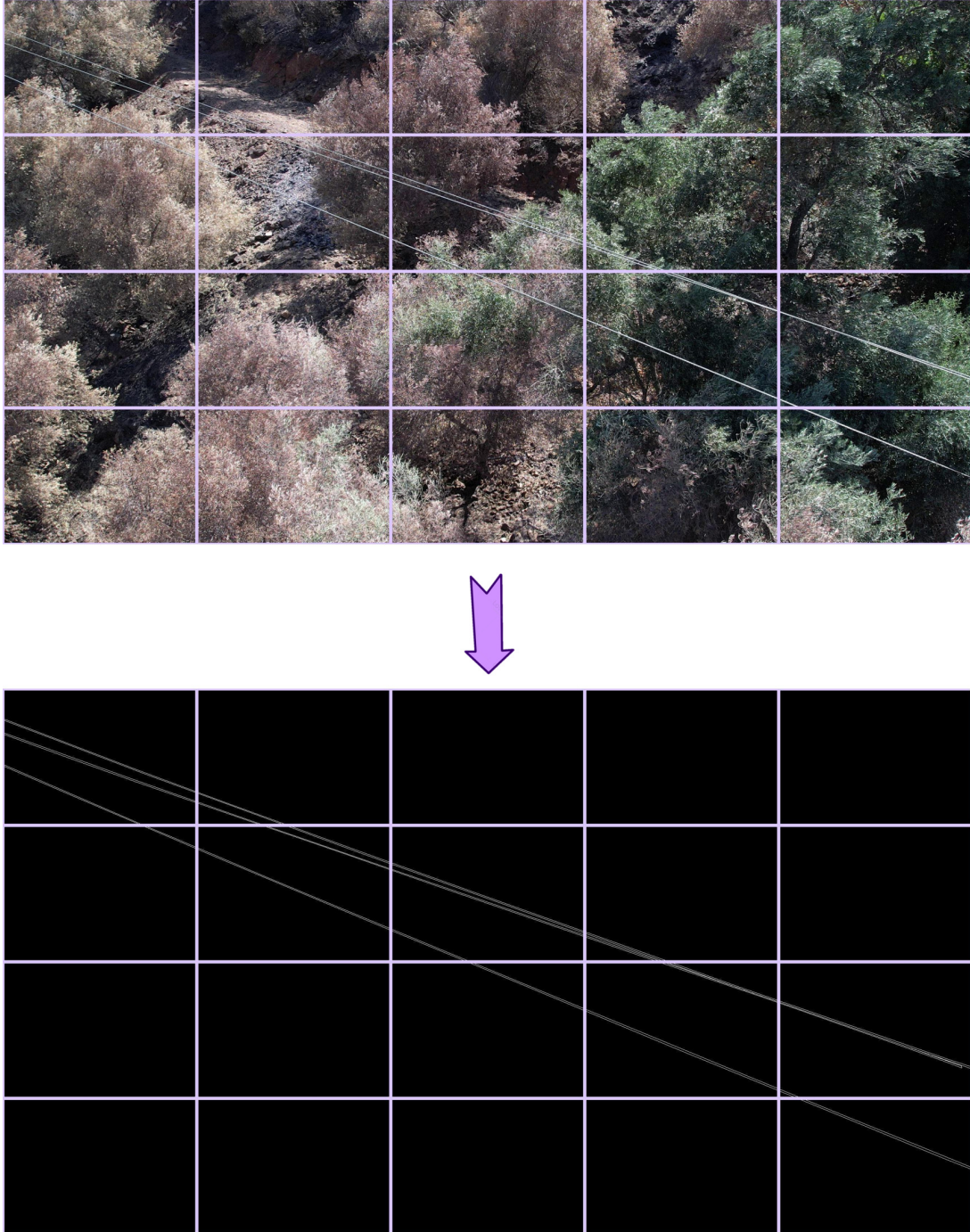
Figure 3.10: Illustration of the proposed 4x5 grid approach for the segmentation process.

3.3.2   Otsu's Threshold

In the final stage of the segmentation process, the deep learning network produces a feature map as the output, encapsulating learned patterns and relevant details. To convert this continuous-valued feature map into a binary image suitable for segmentation, Otsu's thresholding method was integrated. Otsu's method optimally determines a threshold by maximizing the variance between two classes in the image histogram. Mathematically, the method seeks the threshold t that maximizes the intra-class variance

$$\sigma_w^2(t) = w_0(t) \cdot w_1(t) \cdot [\mu_0(t) - \mu_1(t)]^2$$

where $w_0(t)$ and $w_1(t)$ are the probabilities of occurrence of the two classes split by the threshold t, and $\mu_0(t)$ and $\mu_1(t)$ are the means of pixel intensities in the respective classes. Applying Otsu's threshold to the feature map allows for an automatic and adaptive binarization, effectively distinguishing between foreground and background regions.

3.4   Implementation

The pre-mentioned architectures were tested along with the U-net which was also implemented for comparison reasons. U-Net was first developed and used for biomedical image segmentation. It is a fully convolutional neural network that is designed to learn from fewer training samples. In particular, U-Net is an improvement over the existing FCN. Its architecture consists of two basic components, encoder and decoder, that are connected via a bridge. The encoder network halves the spatial dimensions and doubles the number of filters at each encoder block. Likewise, the decoder network doubles the spatial dimensions and halves the number of feature channels. Fig. 3.11 provides an illustration of the U-Net model architecture.

The networks were implemented in Python using Pytorch [48], on an NVIDIA GeForce GTX 1650 Ti GPU. All four networks were trained and tested using the same training sets for comparison purposes. The learning rate for all networks is equivalent to 0.001. Also, Adam optimiser was used in all cases. Binary Cross-Entropy (BCE) loss function was used to gauge the error between the prediction output and the provided target value. U-Net needed 10 epochs to achieve results comparable with the other 3 networks, which were trained for 6 epochs. D-LinkNet uses ResNet18 as encoder, while DeepLabV3 uses MobileNet. FCN and U-Net were implemented without backbone.
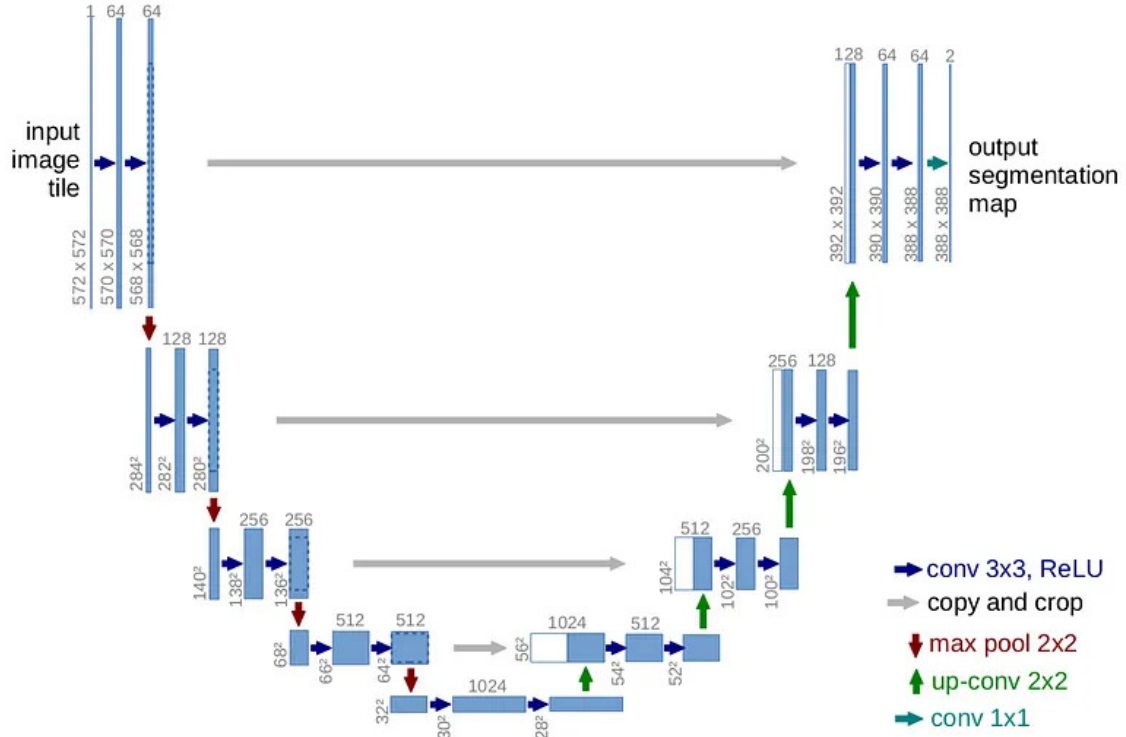
Source: [47]

Figure 3.11: U-Net Architecture

3.4.1 Binary Cross Entropy (BCE) loss function

The Binary Cross-Entropy Loss, also known as logistic loss or log loss, is a commonly used loss function for binary classification problems. It measures the difference between the predicted probability distribution and the actual binary output.

The Binary Cross-Entropy Loss for a single example is defined as:

$$\text{Binary Cross-Entropy Loss} = -\left(y \cdot \log(p) + (1 - y) \cdot \log(1 - p)\right)$$

where: y is the true class label (either 0 or 1), p is the predicted probability of the instance belonging to class 1.

For a dataset with N examples, the overall Binary Cross-Entropy Loss is the average over all examples:

$$\text{Overall Binary Cross-Entropy Loss} = -\frac{1}{N} \sum_{i=1}^{N} \left(y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i)\right)$$

This loss function penalizes models more when their predicted probabilities diverge

from the true labels, encouraging accurate probability estimates for binary classification tasks.

3.4.2 Adam optimizer

The Adam optimizer is a widely used optimization algorithm in machine learning, particularly for training deep neural networks. It combines ideas from RMSprop and Momentum to achieve efficient and adaptive learning rates.

The update rule for the parameter θ at time step t is given by:

$$
\begin{aligned}
m_t &= \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot \nabla_\theta J(\theta) \\
v_t &= \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot (\nabla_\theta J(\theta))^2 \\
\hat{m}_t &= \frac{m_t}{1 - \beta_1^t} \\
\hat{v}_t &= \frac{v_t}{1 - \beta_2^t} \\
\theta_{t+1} &= \theta_t - \frac{\alpha}{\sqrt{\hat{v}_t} + \epsilon} \cdot \hat{m}_t
\end{aligned}
$$

where:

α is the learning rate.
β_1 and β_2 are exponential decay rates for the first and second moments, respectively.
m_t and v_t are the first and second moments of the gradients.
$\hat{m}_t$ and $\hat{v}_t$ are bias-corrected estimates of the moments.
ϵ is a small constant to avoid division by zero.

The Adam optimizer adapts the learning rates for each parameter individually, facilitating the training of deep neural networks.

Chapter 4

Results

At first, the trained deep neural networks were evaluated on the PLDU and PLDM datasets' test images. Metrics in Table 4.1 and 4.2 indicate the performance of all networks, in both datasets, in terms of Precision, Recall and F1-Score which are given in Eqs. 4.1, 4.2 and 4.3 respectively.

Precision is the ratio of true positive predictions to the total number of positive predictions made by the model. In the context of power line segmentation, precision quantifies the accuracy of the model in identifying pixels as power lines, representing the proportion of correctly predicted power line pixels among all pixels predicted as power lines by the model.

$$Precission = 2 * \frac{TP}{TP + FP} \tag{4.1}$$

- TP (True Positives): Number of pixels correctly identified as power lines.

- FP (False Positives): Number of pixels wrongly identified as power lines.

High precision indicates that when the model predicts power lines, it is likely to be correct.

Recall, also known as sensitivity or true positive rate, is the ratio of true positive predictions to the total number of actual positive instances in the ground truth. In the context of power line segmentation, recall quantifies the model's effectiveness in detecting actual power lines, representing the proportion of successfully detected power lines among all the actual power lines present in the ground truth.

$$Recall = 2 * \frac{TP}{TP + FN} \tag{4.2}$$

39

- FN (False Negatives): Number of pixels that are part of actual power lines but were not detected by the model.

High recall indicates that the model is effective at capturing most of the actual power lines.

F1-Score is the harmonic mean of precision and recall. It provides a balanced measure that considers both false positives and false negatives. F1-Score is especially useful when there is an imbalance between the classes.

$$F1 - Score = 2 * \frac{Recall * Precission}{Recall + Precission} \tag{4.3}$$

F1-Score ranges between 0 and 1, where a higher value indicates a better balance between precision and recall.

Results in Tables 4.1 and 4.2 show that all three implemented networks using the dilated convolution technique, outperform the standard U-Net in terms of F1-Score, precision and accuracy. Moreover, D-LinkNet's results are close to identical. U-Net needed more epochs, specifically 10, to achieve results comparable to the other networks. However, it still achieved inferior performance. The visual representation of these results, depicted in Figs. 4.1 & 4.2, provides an insightful complement to the numerical findings.

Table 4.1: Results - PLDU Dataset [28]

Architecture	F1-Score	Precision	Recall
D-LinkNet	0.9648	0.9387	0.9965
DeepLabV3	0.8864	0.8264	0.9750
FCN-8s	0.8669	0.8441	0.9126
UNet	0.7712	0.7405	0.8340

Table 4.2: Results - PLDM Dataset [28]

Architecture	F1-Score	Precision	Recall
D-LinkNet	0.9862	0.9774	0.9965
DeepLabV3	0.9451	0.9329	0.9763
FCN-8s	0.9418	0.9537	0.9411
UNet	0.8682	0.8529	0.9156

In order to investigate further the dilated convolutions' superiority, networks were tested in video frames, acquired by HEDNO S.A. To make the system responsive to both

Figure 4.1: Visualizing PLDU: Segmentation Insights with 4 Deep Learning Networks. From Top to Bottom: Original Images, UNet Segmentation, Modified FCN Segmentation, DeepLab Segmentation, DLinkNet Segmentation.

Figure 4.2: Visualizing PLDM: Segmentation Insights with 4 Deep Learning Networks. From Top to Bottom: Original Images, UNet Segmentation, Modified FCN Segmentation, DeepLab Segmentation, DLinkNet Segmentation.

close and long-distance shots of the power lines, the frames used for testing, were split in 20 smaller tiles and segmented separately. Table 4.3 shows the results in terms of the evaluation metrics used above.

Table 4.3: Results - HEDNO Dataset [28]

Architecture	F1-Score	Precision	Recall
D-LinkNet	0.9978	0.9983	0.9974
DeepLabV3	0.9540	0.9254	0.9849
FCN-8s	0.9627	0.9493	0.9769
UNet	0.7560	0.6303	0.9734

To get a better idea for the performance of the implemented networks, sample results are visualized in Figs. 4.3, 4.4, 4.5 and 4.6. Within the output frame of D-LinkNet, the lines are not only detected efficiently but also exhibit a remarkable absence of significant gaps or instances of falsely detected lines. The outputs from FCN and DeepLab also showcase effective line detection, demonstrating a commendable performance with minimal noise and only occasional small gaps. However, in the context of U-Net, the detected lines are enveloped in a perceptible level of noise, and the gaps in the line segments that went undetected are notably more evident, contributing to a less precise delineation of the desired features.

It is obvious that in networks where the dilated convolution technique is utilized, results are far better than U-Net, where regular convolution is being used. The networks were trained using a dataset consisting of various background scenarios, making the network responsive and accurate in many test images with different content and background (urban and mountain, high and low vegetation). Furthermore, the existence of the images acquired in both near and far distances to the power lines enhances the performance of the network.

The Precision and Recall Curves, depicted in Figs. 4.7, 4.8 and 4.9, are presented for each dataset utilized in our study, namely the PLDU dataset, the PLDM dataset, and the HEDNO S.A. dataset. The tradeoff between the true positive rate and the positive predictive value for our model using different probability thresholds is important to be examined in order to evaluate it. The higher the area under each curve the higher recall and precision, where high precision relates to a low false positive rate, and high recall relates to a low false negative rate. For all three datasets, the area under the D-LinkNet model is higher than the area under the rest architectures, which is promising for the

Figure 4.3: Segmentation Result of D-LinkNet. Detected lines are highlighted in red, while gaps and noise are highlighted in yellow.

Figure 4.4: Segmentation Result of DeepLabV3. Detected lines are highlighted in red, while gaps and noise are highlighted in yellow.

Figure 4.5: Segmentation Result of Dilated FCN. Detected lines are highlighted in red, while gaps and noise are highlighted in yellow.

Figure 4.6: Segmentation Result of UNet. Detected lines are highlighted in red, while gaps and noise are highlighted in yellow.

future, i.e. for testing our model to more pilot cases, with complex background terrain or noisy video acquisition.

(a)
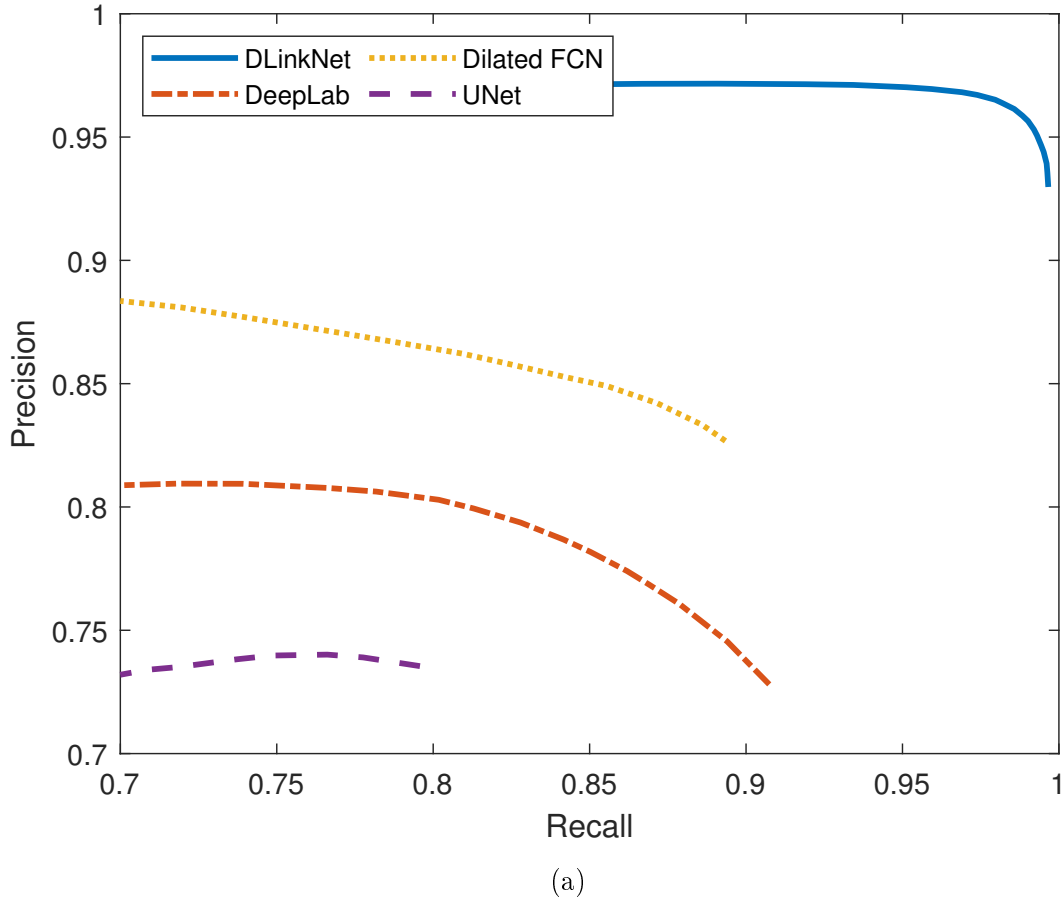
Figure 4.7: Precision & Recall Curve for PLDU Dataset
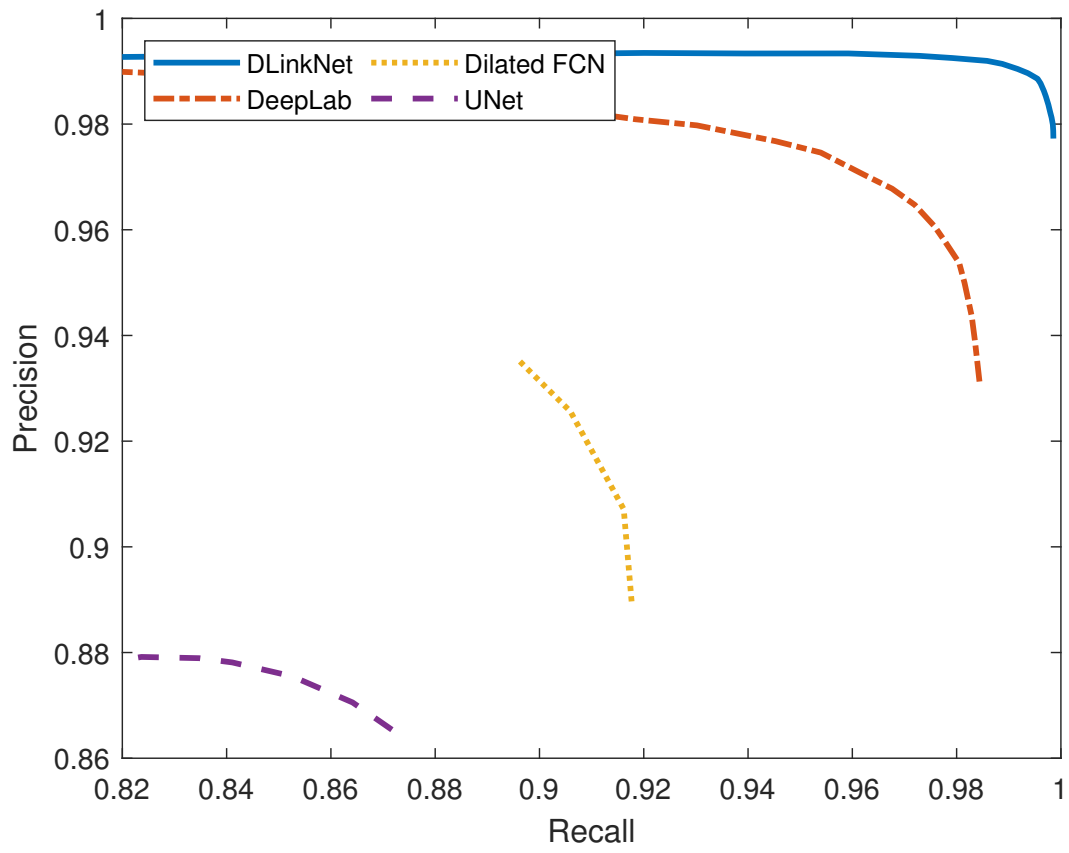
(b)

Figure 4.8: Precision & Recall Curve for PLDM Dataset
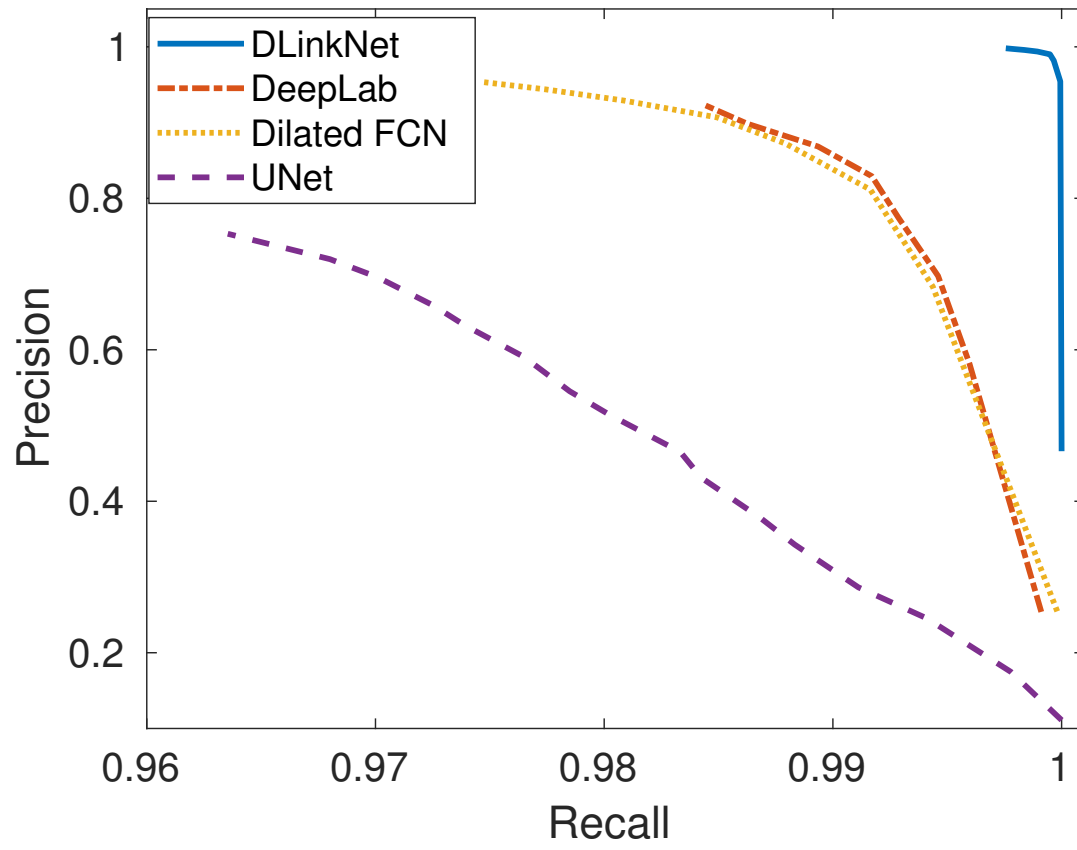
(c)

Figure 4.9: Precision & Recall Curve for Unseen Video Frames

Chapter 5

Conclusion

The dilated convolution technique seems to be a good solution when it comes to power lines or other similar structures. It is important to keep in mind the structure of the object when looking for the appropriate segmentation model. In the case of aerial power line detection, it is important to remember that their shape is only a few pixels wide. Thus, an appropriate approach is required in order to lose as less information as possible. At a first glance, some lost segments may not seem to be crucial in the performance of the system. However, segmentation is only a first step; especially, when it comes to cases such as fault detection in power lines, a lost segment may prove to cause false fault detection.

Our proposed model is designed towards robustness; it was compared to other popular models used in similar applications and the datasets we used contain images with complex background terrain acquired in both urban and mountain areas. Finally, it is suitable for real-time operation providing us the opportunity to further expand our system in order to produce alerts upon faults detection during UAV's flight.

Acknowledgements

This thesis becomes a reality with the kind support of many individuals. I would like to thank all of them.

First and foremost, I would like to express my sincere appreciation to Professor Michalis Zervakis for providing me with the initial opportunity to work on this thesis and for his valuable guidance along the way.

I extend my deepest thanks to Mrs. Konstantia Moirogiorgou for her continuous support and guidance at every step of the road.

I would also like to thank the rest of my thesis committee: Prof. Euripides Petrakis and Prof. Dionysios Christopoulos.

Special gratitude is extended to the Hellenic Electricity Distribution Network Operator S.A. for the generous provision of information and data crucial to the completion of this thesis.

And finally, last but by no means least, I would like to thank my family for supporting me all these years and my friends for always believing in me.

Appendix A

Additional segmentation results
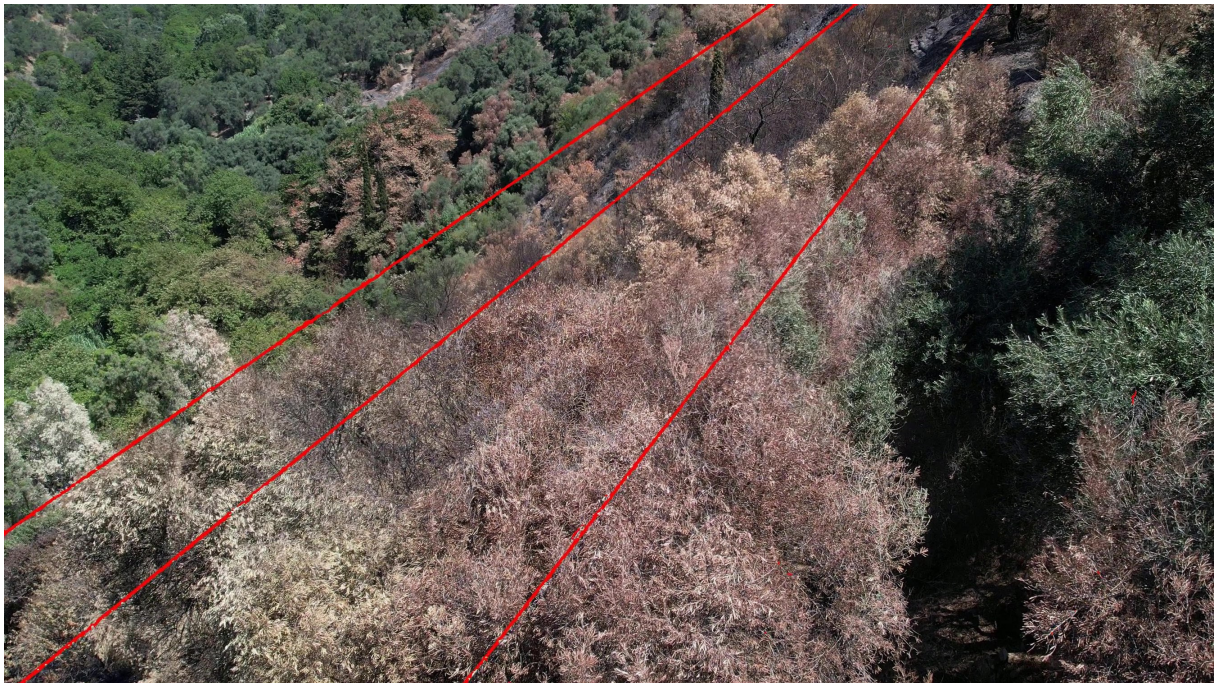
Figure A.1: Example of detected power lines using D-LinkNet (a)

Figure A.2: Example of detected power lines using D-LinkNet (b)

Figure A.3: Example of detected power lines using D-LinkNet (c)
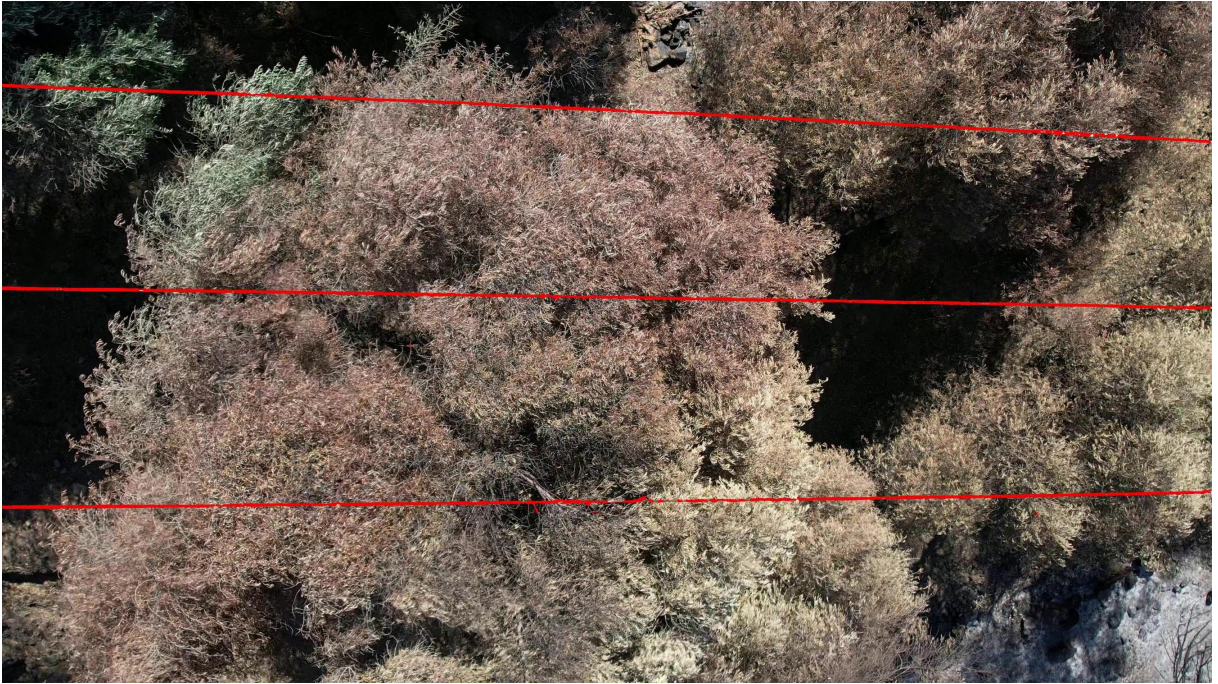
Source: [28]

Figure A.4: Example of detected power lines using D-LinkNet (d)

Figure A.5: Example of detected power lines using D-LinkNet (e)
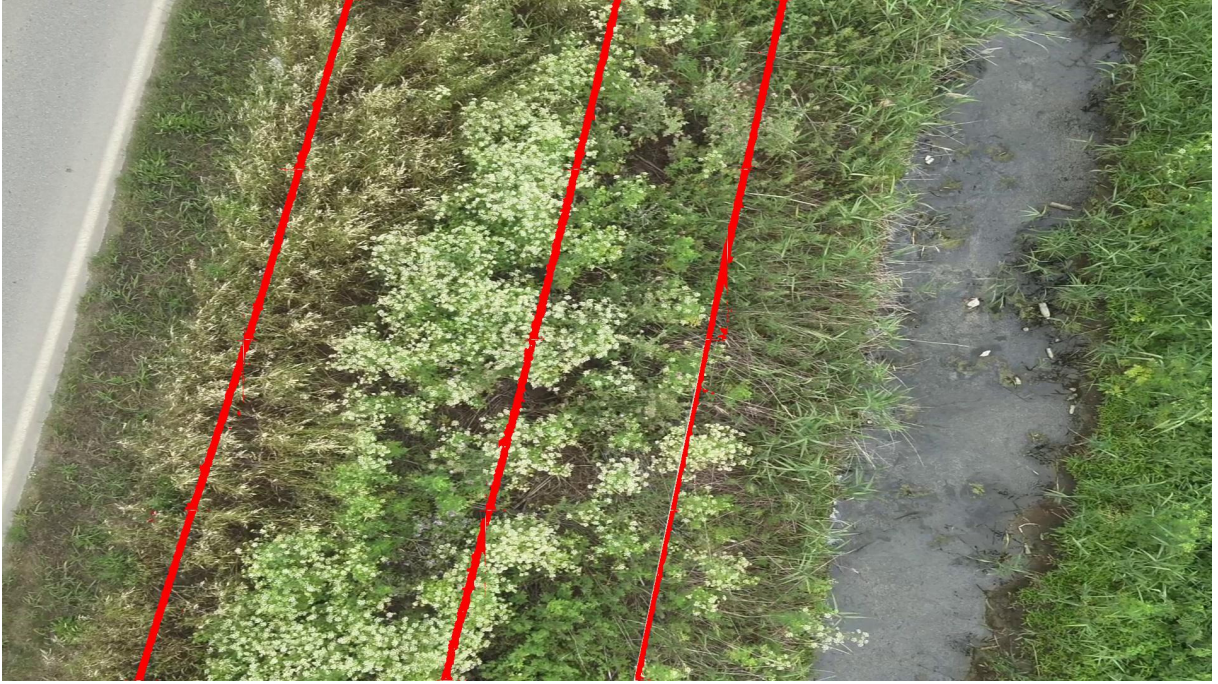
Figure A.6: Example of detected power lines using D-LinkNet (f)

Figure A.7: Example of detected power lines using D-LinkNet (g)

Figure A.8: Example of detected power lines using D-LinkNet (h)

Figure A.9: Example of detected power lines using D-LinkNet (i)
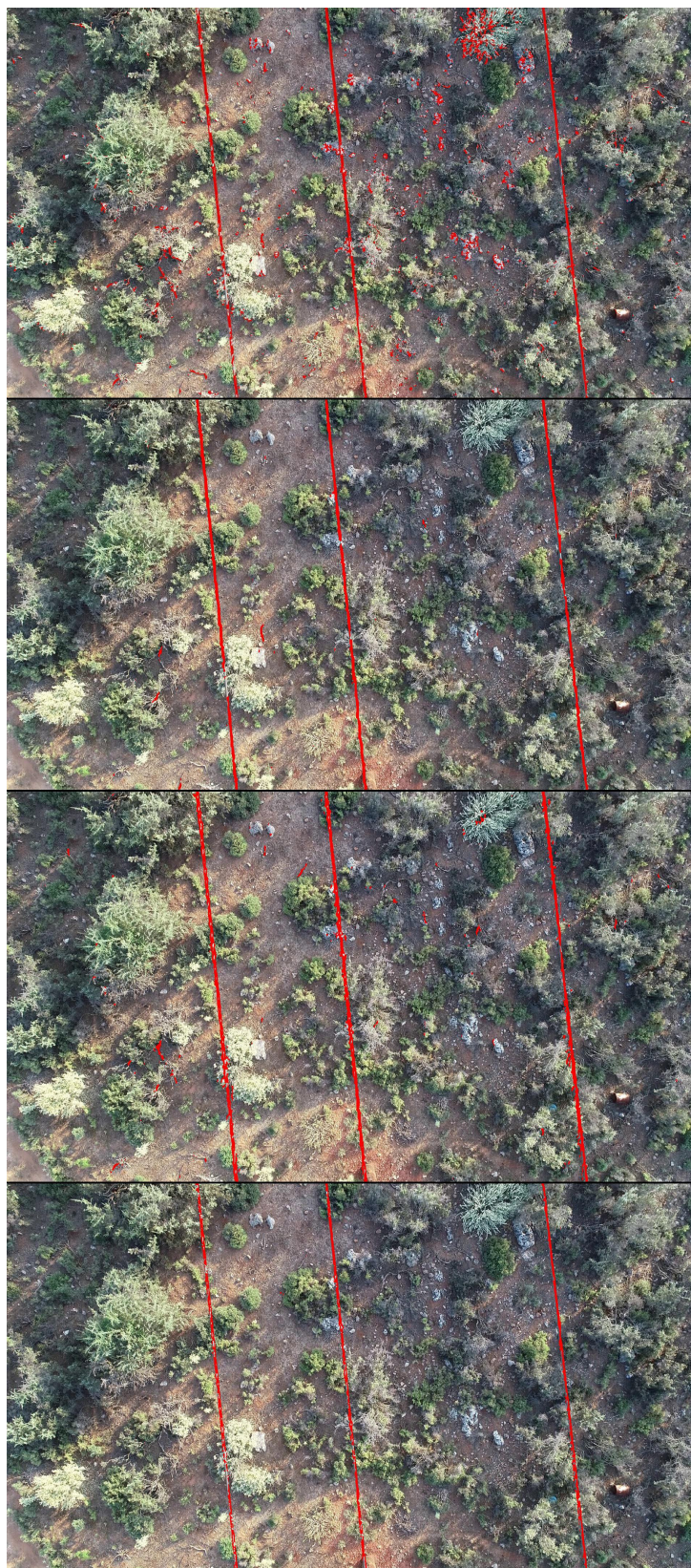
Figure A.10: Example (a) - HEDNO Dataset: Segmentation Insights with 4 Deep Learning Networks. From Top to Bottom: Original Images, UNet Segmentation, Modified FCN Segmentation, DeepLab Segmentation, DLinkNet Segmentation.

Figure A.11: Example (b) - HEDNO Dataset: Segmentation Insights with 4 Deep Learning Networks. From Top to Bottom: Original Images, UNet Segmentation, Modified FCN Segmentation, DeepLab Segmentation, DLinkNet Segmentation.
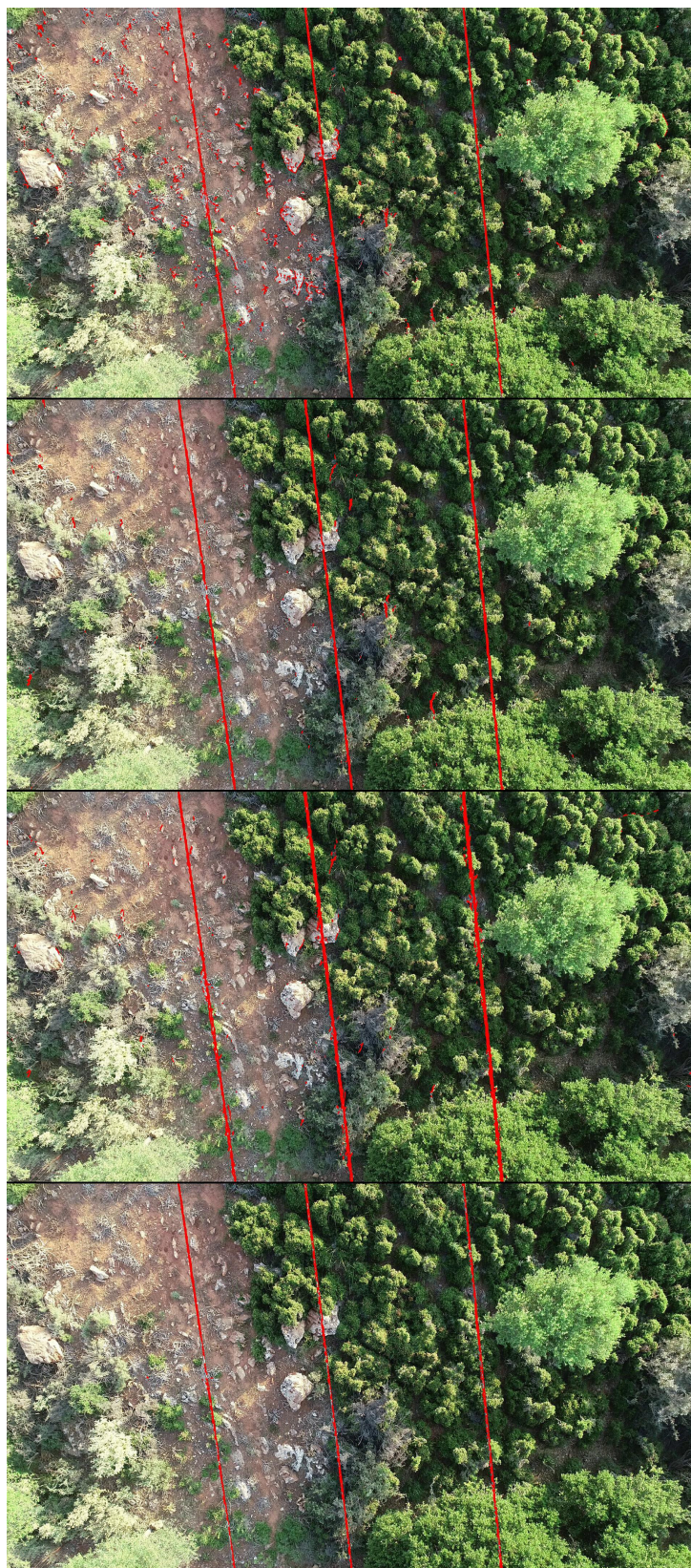
Figure A.12: Example (c) - HEDNO Dataset: Segmentation Insights with 4 Deep Learning Networks. From Top to Bottom: Original Images, UNet Segmentation, Modified FCN Segmentation, DeepLab Segmentation, DLinkNet Segmentation.

Figure A.13: Example (d) - HEDNO Dataset: Segmentation Insights with 4 Deep Learning Networks. From Top to Bottom: Original Images, UNet Segmentation, Modified FCN Segmentation, DeepLab Segmentation, DLinkNet Segmentation.

Figure A.14: Example (e) - HEDNO Dataset: Segmentation Insights with 4 Deep Learning Networks. From Top to Bottom: Original Images, UNet Segmentation, Modified FCN Segmentation, DeepLab Segmentation, DLinkNet Segmentation.

Figure A.15: Example (f) - HEDNO Dataset: Segmentation Insights with 4 Deep Learning Networks. From Top to Bottom: Original Images, UNet Segmentation, Modified FCN Segmentation, DeepLab Segmentation, DLinkNet Segmentation.

Figure A.16: Example (g) - HEDNO Dataset: Segmentation Insights with 4 Deep Learning Networks. From Top to Bottom: Original Images, UNet Segmentation, Modified FCN Segmentation, DeepLab Segmentation, DLinkNet Segmentation.

Figure A.17: Example (h) - HEDNO Dataset: Segmentation Insights with 4 Deep Learning Networks. From Top to Bottom: Original Images, UNet Segmentation, Modified FCN Segmentation, DeepLab Segmentation, DLinkNet Segmentation.

Bibliography

[1] *Drone for Power Line Inspection: the Benefits, Applications, and More.* URL: `https : / / www . jouav . com / industry / power - line - inspection ? fbclid = IwAR1QShwEAo9y4fTwUu1rDt6bwOxMiOZ64txm1cb3ablLX16g5LJ7VPl4a8A`.

[2] Owensong. *Lidar UAV Drone in Powerline Inspection Industry.* 2019. URL: `https : / / www . uavfordrone . com / lidar - uav - drone - in - powerline - inspection - industry/`.

[3] Guangjian Yan et al. "An airborne multi-angle power line inspection system". In: *2007 IEEE International Geoscience and Remote Sensing Symposium.* Ieee. 2007, pp. 2913–2915.

[4] Lizhong Zhang, Huilin Jiang, and Lixin Meng. "Research of the observation suspended bin for helicopter power line inspection". In: *2009 International Conference on Mechatronics and Automation.* IEEE. 2009, pp. 1689–1694.

[5] Ludan Wang et al. "Development of a practical power transmission line inspection robot based on a novel line walking mechanism". In: *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems.* IEEE. 2010, pp. 222–227.

[6] Giuseppe Silano et al. "A multi-layer software architecture for aerial cognitive multi-robot systems in power line inspection tasks". In: *2021 international conference on unmanned aircraft systems (ICUAS).* IEEE. 2021, pp. 1624–1629.

[7] Xionggang Li and Yanming Guo. "Application of LiDAR technology in power line inspection". In: *IOP Conference Series: Materials Science and Engineering.* Vol. 382. 5. IOP Publishing. 2018, p. 052025.

[8] S Pu et al. "Real-time powerline corridor inspection by edge computing of UAV Lidar data". In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 42 (2019), pp. 547–551.

[9]   Fábio Azevedo et al. "Lidar-based real-time detection and modeling of power lines for unmanned aerial vehicles". In: *Sensors* 19.8 (2019), p. 1812.

[10]   Zhenyu Zhou et al. "Energy-efficient industrial Internet of UAVs for power line inspection in smart grid". In: *IEEE Transactions on Industrial Informatics* 14.6 (2018), pp. 2705–2714.

[11]   Tong He, Yihui Zeng, and Zhuangli Hu. "Research of multi-rotor UAVs detailed autonomous inspection technology of transmission lines based on route planning". In: *IEEE Access* 7 (2019), pp. 114955–114965.

[12]   Zhitao Guan et al. "Toward delay-tolerant flexible data access control for smart grid with renewable energy resources". In: *IEEE Transactions on Industrial Informatics* 13.6 (2017), pp. 3216–3225.

[13]   MH Nasseri et al. "Power line detection and tracking using hough transform and particle filter". In: *2018 6th RSI International Conference on Robotics and Mechatronics (IcRoM)*. IEEE. 2018, pp. 130–134.

[14]   Mcebisi Solilo, Wesley Doorsamy, and Babu Sena Paul. "Uav power line detection and tracking using a color transformation". In: *2021 International Conference on Electrical, Computer and Energy Technologies (ICECET)*. IEEE. 2021, pp. 1–5.

[15]   Milan Banić et al. "Intelligent machine vision based railway infrastructure inspection and monitoring using UAV". In: *Facta Universitatis, Series: Mechanical Engineering* 17.3 (2019), pp. 357–364.

[16]   Mehrez Marzougui et al. "A lane tracking method based on progressive probabilistic Hough transform". In: *IEEE access* 8 (2020), pp. 84893–84905.

[17]   Wenbo Zhao, Qing Dong, and Zhengli Zuo. "A method combining line detection and semantic segmentation for power line extraction from unmanned aerial vehicle images". In: *Remote Sensing* 14.6 (2022), p. 1367.

[18]   Sumeet Saurav et al. "Power line segmentation in aerial images using convolutional neural networks". In: *Pattern Recognition and Machine Intelligence: 8th International Conference, PReMI 2019, Tezpur, India, December 17-20, 2019, Proceedings, Part I*. Springer. 2019, pp. 623–632.

[19] Jayavardhana Gubbi, Ashley Varghese, and P Balamuralidhar. "A new deep learning architecture for detection of long linear infrastructure". In: *2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA)*. IEEE. 2017, pp. 207–210.

[20] Heng Zhang et al. "Detecting power lines in UAV images with convolutional features and structured constraints". In: *Remote Sensing* 11.11 (2019), p. 1342.

[21] Van Nhan Nguyen, Robert Jenssen, and Davide Roverso. "LS-Net: Fast single-shot line-segment detector". In: *Machine Vision and Applications* 32 (2021), pp. 1–16.

[22] Lei Yang et al. "Vision-based power line segmentation with an attention fusion network". In: *IEEE Sensors Journal* 22.8 (2022), pp. 8196–8205.

[23] Rabab Abdelfattah, Xiaofeng Wang, and Song Wang. "Plgan: Generative adversarial networks for power-line segmentation in aerial images". In: *IEEE Transactions on Image Processing* (2023).

[24] Lichen Zhou, Chuang Zhang, and Ming Wu. "D-LinkNet: LinkNet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction". In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2018, pp. 182–186.

[25] Liang-Chieh Chen et al. "Rethinking atrous convolution for semantic image segmentation". In: *arXiv preprint arXiv:1706.05587* (2017).

[26] Jonathan Long, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 3431–3440.

[27] Alejandro Frangi. "Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015". In: *Lecture Notes in Computer Science*. 2015.

[28] Aikaterini Tsellou et al. "Aerial video inspection of Greek power lines structures using machine learning techniques". In: *2022 IEEE International Conference on Imaging Systems and Techniques (IST)*. IEEE. 2022, pp. 1–6.

[29] Aikaterini Tsellou et al. "A UAV Intelligent System for Greek Power Lines Monitoring". In: *Sensors* 23.20 (2023), p. 8441.

[30] B. Palac. *Segmentácia obrázku*. Wikimedia Commons. Wikimedia Foundation. 2020. URL: https://commons.wikimedia.org/wiki/File:Image_segmentation.png.

[31] Ravitejarj. *Indian driving dataset: Instance Segmentation with Mask R-CNN and Tensorflow*. Published on Medium. 2019. URL: https://medium.com/analytics-vidhya/indian-driving-dataset-instance-segmentation-with-mask-r-cnn-and-tensorflow-b03617156d44.

[32] Harshit Kumar. *Introduction to Panoptic Segmentation: A Tutorial*. https://kharshit.github.io/blog/2019/10/18/introduction-to-panoptic-segmentation-tutorial. Technical Fridays. 2019.

[33] Sahar Zafari et al. "Segmentation of overlapping convex objects". In: (2014).

[34] EA Zanaty and Said Ghoniemy. "Medical image segmentation techniques: an overview". In: *International Journal of informatics and medical data processing* 1.1 (2016), pp. 16–37.

[35] Teaching Assistants. *Week 6: Region Growing and Clustering Segmentation*. Online teaching materials by the teaching assistants in the Biomedical Engineering department at Cairo University. Correspondence: Asem Alaa, Eslam Adel, Ayman Anwar. 2019. URL: https://sbme-tutorials.github.io/2019/cv/notes/6_week6.html.

[36] Saumya Saxena. "Introduction to Deep Learning". In: *GeeksforGeeks* (2023). Contributed by Saumya Saxena. URL: https://www.geeksforgeeks.org/introduction-deep-learning/.

[37] Kentaro Wada. *Labelme: Image Polygonal Annotation with Python*. DOI: 10.5281/zenodo.5711226. URL: https://github.com/wkentaro/labelme.

[38] Fisher Yu and Vladlen Koltun. *Multi-Scale Context Aggregation by Dilated Convolutions*. arXiv:1511.07122. Nov. 2015.

[39] GeeksforGeeks. *Dilated Convolution*. 2022. URL: https://www.geeksforgeeks.org/dilated-convolution/#:~:text=Dilated\%20Convolution\%3A\%20It\%20is\%20a,larger\%20area\%20of\%20the\%20input..

[40] Liang-Chieh Chen et al. "Semantic image segmentation with deep convolutional nets and fully connected crfs". In: *arXiv preprint arXiv:1412.7062* (2014).

[41] Liang-Chieh Chen et al. "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs". In: *IEEE transactions on pattern analysis and machine intelligence* 40.4 (2017), pp. 834–848.

[42] Fezan. *Review DeepLabv3 Semantic Segmentation*. 2017. URL: https://medium.com/swlh/review-deeplabv3-semantic-segmentation-52c00ddbf28d.

[43] Lichen Zhou, Chuang Zhang, and Ming Wu. "D-LinkNet: LinkNet with Pretrained Encoder and Dilated Convolution for High Resolution Satellite Imagery Road Extraction". In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. June 2018, pp. 192–1924.

[44] Abhishek Chaurasia and Eugenio Culurciello. "LinkNet: Exploiting Encoder Representations for Efficient Semantic Segmentation". In: *IEEE Visual Communications and Image Processing (VCIP)*. Dec. 2017, pp. 1–4.

[45] Kaiming He et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[46] Jia Deng et al. "Imagenet: A large-scale hierarchical image database". In: *2009 IEEE conference on computer vision and pattern recognition*. Ieee. 2009, pp. 248–255.

[47] Jeremy Zhang. "UNet — Line by Line Explanation: Example UNet Implementation". In: *Towards Data Science* (2019). URL: `https : / / towardsdatascience.com/unet-line-by-line-explanation-9b191c76baf5`.

[48] Adam Paszke et al. "PyTorch: An Imperative Style, High-Performance Deep Learning Library". In: *Proceedings of the 33rd International Conference on Neural Information Processing Systems (NeurIPS)*. Dec. 2019.

List of Figures

List of Tables