Deep Learning Course - Final Project Report

Transferability of photovoltaic (pv) arrays from high resolution satellite imagery data for semantic segmentation

Submitted By: Inbal Karibian & Arad Peleg & Amir Boger & Dolev Hagbi

31 August 2024

Abstract

This study evaluates the performance of U-Net, LinkNet, FPN, and PSP-Net models, each utilizing ResNet-18 backbones, for segmenting solar panels from satellite imagery across varying domains. At first, we set a baseline by training and testing the models on Australian (AUS) data, achieving mIoU scores around 0.6 to 0.7, to understand the models' performance in a consistent environment. Next, we tested raw transferability by applying these models to U.S. (USA) data, which revealed a significant drop in mIoU scores by 10% to 20%, highlighting the domain gap caused by differences in data distribution across geographical environments. To address this, we applied AdaptSegNet for domain adaptation, which aimed to help the models learn domain-invariant features. This led to modest improvements, with mIoU increasing by about 1% in certain cases. Finally, we combined AdaptSegNet with fine-tuning, further refining the models on the target domain, though the overall gains were slight, demonstrating the ongoing challenges in overcoming domain shifts. For the detailed project code and more, visit our GitHub repository.

1 Introduction

The rapid growth of distributed solar photovoltaic (PV) arrays necessitates high-resolution data collection on their quantity and energy generation [5]. Semantic segmentation, which assigns a category label to each pixel of an image [4], is a powerful technique used in this context to accurately detect and delineate PV arrays. Recent advancements show that deep learning, particularly convolutional neural networks (CNN), significantly outperforms traditional methods in detecting PV arrays using high-resolution satellite imagery [6]. Additionally, domain adaptation (DA) techniques enhance the transferability of these models across different regions [7].

1.1 Domain adaptation

Domain adaptation (DA) is a technique designed to handle situations where training and test sets can originate from different feature spaces or distributions, leading to a phenomenon known as domain shift. It aims to learn a model from a source labeled data that can be generalized to a target domain by minimizing the difference between domain distributions. In other words, DA addresses the challenge of applying a model trained on one dataset (the source domain) to another dataset (the target domain) when the distributions of these datasets differ [2].

The main objective of DA is to minimize the discrepancy between the source domain (with labeled data) and the target domain (which may have unlabeled data). The general objective can be expressed as:

$$\min_{\mathbf{G_f},\mathbf{G_y}} \max_{\mathbf{G_d}} \left[\mathcal{L}_\mathbf{y}(\mathbf{G_y}(\mathbf{G_f}(\mathbf{x_s})), \mathbf{y_s}) - \lambda \mathcal{L}_\mathbf{d}(\mathbf{G_d}(\mathbf{G_f}(\mathbf{x})), \mathbf{d})\right] \tag{1}$$

where G_f is the feature extractor, G_y is the label predictor, and G_d is the domain discriminator. The term x_s represents the source domain data, and y_s denotes the labels associated with this data. The data from both the source and target domains is represented by x. The function $\mathcal{L}_y$ is the classification loss, such as cross-entropy, for the source domain, while $\mathcal{L}_d$ is the domain classification loss, like binary cross-entropy. The domain label d indicates whether the data is from the source (0) or target (1) domain, and λ is a trade-off parameter that balances the two objectives.

1.2 Adversarial Training

The approach we use, adversarial DA, leverages the principles of generative adversarial networks (GANs) to minimize the distribution discrepancy between source and target domains, creating domain-invariant features. This approach involves a generator producing features similar to the target domain, while discriminator attempts to differentiate between source and target features [2].

In adversarial training, the generator (feature extractor) and the discriminator (domain classifier) play a min-max game:

$$\min_{\mathbf{G_f}} \max_{\mathbf{G_d}} \mathcal{L}_{\mathbf{d}}(\mathbf{G_d}(\mathbf{G_f}(\mathbf{x})), \mathbf{d}) \tag{2}$$

- The **generator** G_f (feature extractor) tries to fool the discriminator by making the features from the source and target domains indistinguishable.
- The **discriminator** G_d tries to correctly classify the features as coming from either the source or target domain.

1.3 DA General Idea

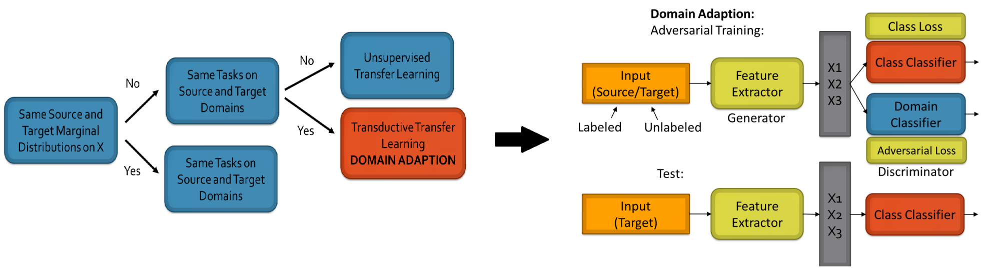

Figure 1: The left side of the image illustrates the decision process for determining the appropriate learning approach based on whether the source and target domains share the same marginal distributions and tasks, leading to either unsupervised transfer learning or transductive transfer learning (domain adaptation). The right side demonstrates the adversarial domain adaptation process, where labeled and unlabeled data from the source and target domains are processed through a feature extractor and then classified by both a class classifier and a domain classifier, with adversarial training used to reduce domain discrepancy.

1.4 Objectives

This project aims to develop a scalable, high-precision method for PV array detection, leveraging deep learning and domain adaptation to improve accuracy and applicability across diverse datasets.

The specific aims are to:

1. Investigate non-trivial changes to the existing model architectures and training methods to enhance transferability further. This might include experimenting with different backbone networks or combining multiple domain adaptation strategies.
2. To compare the modules in terms of performance, training time and explainability.
3. Develop or improve and test new objective functions that could better capture the domain-specific features and improve the generalization capability of the models.

2 Related Work

2.1 Renewable Energy

Over the past few years, renewable energy has increased rapidly worldwide. In fact, according to the International Energy Agency, it was predicted that total renewable-based power capacity will grow by 50 between 2019 to 2024, which is roughly equivalent to the total existing power capacity of the United States alone. Solar photovoltaic (PV) generation is expected to account for 60% of this rise, making it today's fastest-growing form of energy generation [1]

2.2 Semantic Segmentation

Semantic segmentation is a challenging task in computer vision, with numerous methods developed to address it across various fields such as autonomous vehicles, human-computer interaction, robotics, medical research, and agriculture. Many of these approaches leverage the deep learning paradigm [3]. Semantic segmentation assigns each pixel in an image to a predefined category, dividing the image into meaningful regions. Unlike image classification, which labels the entire image, or object detection, which identifies objects and their locations, semantic segmentation precisely delineates object boundaries. This makes it more challenging as it bridges the gap between low-level image features and high-level semantic understanding. Deep neural networks have shown great promise in this area by learning to map pixels to semantic labels through training on large datasets. Additionally, newer tasks like instance and panoptic segmentation extend these principles to handle individual object segmentation and the integration of both semantic and instance-level information [4].

2.3 PV Arrays Semantic Segmentation

Various remote sensing systems have been employed to gather information related to photovoltaic (PV) systems. A common method involves using satellites to estimate ground irradiation, which is then applied in models to assess the potential for solar PV energy generation [5]. Detecting solar panels from satellite images is challenging due to their diverse shapes, sizes, and colors, as well as the varying angles of rooftop installations. Additionally, the task becomes even more difficult when using a device with limited computational power. Segmenting satellite images offers a simple and cost-effective method for detecting solar arrays installed on rooftops and on the ground across a region. Identifying solar panels is a crucial first step in the image-based estimation of energy generation from distributed solar arrays connected to a conventional electric grid [9].

2.4 Domain Adaption For Segmentation

Remote Sensing (RS) semantic segmentation/classification tasks. require large training datasets and are generally known for lack of transferability due to the highly disparate RS image content across different geographical regions. DA corrects biases between the source (training) and target (testing) domains at the image, feature, or classifier levels. Image-level DA uses style transfer, feature-level DA aligns features, and classifier-level DA trains a shared classifier for both domains. In remote sensing, DA addresses biases such as radiometry, spectrum, or scale differences, which lead to varying domain gaps. For example, images taken from neighboring areas might have "light" domain gaps, while images from distant regions present "large" domain gaps. Additionally, data captured at different times can create significant domain gaps, as the conditions between the source and target domains may differ substantially [7].

In the study of [8], the authors address the challenge of semantic segmentation, which often struggles to generalize to unseen image domains due to its reliance on pixel-level ground truth. Given the labor-intensive nature of labeling, they propose an adversarial learning method for domain adaptation. By leveraging spatial similarities between source and target domains, adversarial learning is applied in the output space to improve adaptability across domains.

2.5 Adapt Structured Output Space (AdaptSegNet)

Adapt Structured Output Space single-level adversarial learning approach to enhance the adaptation of the segmentation network to new domains. This approach focuses on training a discriminator that operates on the segmentation outputs of the network. The goal of the discriminator is to differentiate between outputs generated from the source domain (where the model was originally trained) and those from the target domain (which may be new or unseen). By doing so, the segmentation network is encouraged to produce outputs that are similar and consistent across both domains. This method leverages the spatial and local similarities in segmentation outputs, ensuring that the network can generalize better to different image domains, even when they vary significantly from the source domain [8].

3 Materials And Methods

In this chapter, we discuss how we applied DA to test the ability of our four different models with back bone **ResNet-18** encoder: **U-Net, Linknet, Feature Pyramid Network (FPN) and PSP-Net** to generalize across different datasets. We conducted experiments to see how well models trained in one environment perform in another, using DA to reduce differences between the training and testing data. The focus is on understanding and improving the adaptability of models to ensure they work effectively across various conditions.

3.1 Data Acquisition

We collected data from two regions: Australia and the United States. The dataset from Australia consists of **3,567** images, each with a corresponding manually created mask that precisely labels the regions of interest (pv-arrays). The dataset from the United States includes about **864** images, also derived from satellite imagery, and each image has a corresponding mask that we got from the website: Earthexplorer That you can down load from here: USA-Data

3.2 Networks Used

Explanation on networks and general architecture:

- **U-Net:** was originally designed for medical image segmentation, uses an encoder-decoder architecture. The encoder, composed of convolutional and pooling layers, extracts features and down samples the input image. The decoder then restores the resolution and performs pixel-wise classification. Skip connections between corresponding encoder and decoder blocks improve localization and segmentation accuracy. Finally, convolution layer and sigmoid function are applied for pixel classification [9].

- **Linknet:** Instead of concatenating features like U-Net, LinkNet adds the upsampled feature maps to the original resolution information, making it a more efficient network. This approach allows LinkNet to use information from both the earlier and later layers of the network to produce the final output [10].

- **FPN:** creates two pyramids—a bottom-up and a top-down pathway—linked by lateral connections to generate feature-rich segmentation maps at each level. The bottom-up pathway consists of convolutional and pooling layers, with each stage producing a feature map for the pyramid. The top-down pathway upsamples these features, enhanced by lateral connections from the bottom-up stages. Finally, the modules with 1/4 of the input resolution are concatenated to produce the final result [10].

- **PSP-Net:** was developed for scene parsing and incorporates a pyramid pooling module (PPM) to aggregate context and detect small objects within a scene. The PPM applies different pooling rates in parallel to generate feature maps from various sub-regions, creating pooled representations from different spatial positions. These operations produce feature maps of varying sizes, which are then upsampled using bilinear upsampling to generate the final segmentation maps [9].

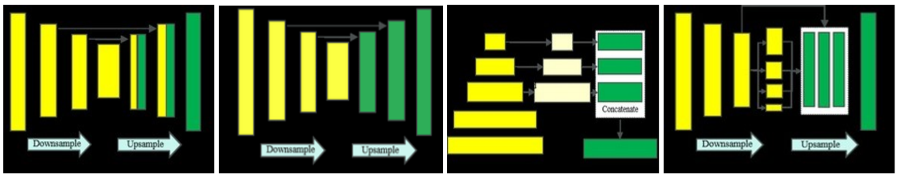

Figure 2: visual comparison of four segmentation architectures from left to right: UNet, Linknet, FPN, and PSPNet. Each diagram illustrates how these architectures handle downsampling, upsampling, and the combination of feature maps for segmentation tasks [10].

3.3 DA Architecture And Components

3.3.1 AdaptSegNet

AdaptSegNet is a DA architecture made for semantic segmentation, particularly when there's a significant domain shift between training and testing data. Unlike feature-based adaptation used in image classification, which can be less effective for complex segmentation tasks, AdaptSegNet aligns the segmentation maps (output space) between domains through adversarial learning. By focusing on the structured output space, which contains crucial scene layout and context information, AdaptSegNet effectively reduces the domain gap, ensuring that target domain segmentations closely resemble those from the source domain [8].

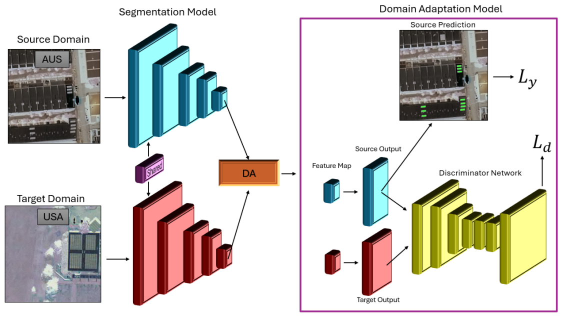

Figure 3: AdaptSegNet architecture, where a segmentation network processes inputs from both source and target domains. The outputs are refined by a discriminator network to minimize domain differences.

3.3.2 Components For Adversarial Training

- **Discriminator:** helps adversarial learning by distinguishing between segmentation outputs from the source and target domains. This process encourages the segmentation network to produce outputs for the target domain that are indistinguishable from those of the source domain, helping to bridge the domain gap and improve performance on the target data.

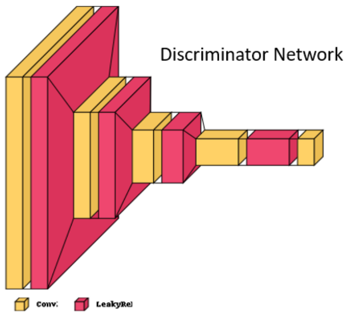

Figure 4: The Discriminator Network architecture in the figure consists of a series of convolutional layers (Yellow blocks) that reduce the input's spatial dimensions while applying Leaky ReLU activations (Red blocks) to introduce non-linearity.

- **Segmentation Model:** generates the segmentation outputs that need to be adapted across domains. By incorporating modifications that help maintain high-resolution feature maps and better context aggregation, this network ensures that the segmentation results are robust, even when applied to data from a different domain than the one it was trained on.

From here we can say that In the Discriminator aims to maximize the loss by correctly identifying whether segmentation outputs are from the source or target domain. Conversely, the Segmentation Network works to minimize this loss, adjusting its outputs to make them indistinguishable between the domains, thereby improving domain adaptation.

3.4 Model Evaluation Metrics

- **Precision:** measures how accurate your model's positive predictions are.
  PV arrays: It tells you the percentage of identified PV arrays that are actually correct.
  Formula:

$$\text{Precision} = \frac{\text{True Positives (correctly identified PV arrays)}}{\text{True Positives} + \text{False Positives (wrongly identified areas as PV arrays)}} \tag{3}$$

- **Recall:** measures the ability of your model to identify all relevant positive instances.
  PV arrays: It shows how many of the actual PV arrays present in the image were correctly detected by the model.
  Formula:

$$\text{Recall} = \frac{\text{True Positives (correctly identified PV arrays)}}{\text{True Positives} + \text{False Negatives (missed PV arrays)}} \tag{4}$$

- **F1-Score:** is the harmonic mean of Precision and Recall.
PV arrays: The F1-Score provides a single metric that balances how well the model identifies PV arrays correctly without missing many or falsely identifying non-PV areas.
Formula:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

- **mIoU:** measures the overlap between the predicted and actual areas.
PV arrays: mIoU assesses how closely the predicted PV array areas match the actual PV arrays on the ground.
Formula:

$$\text{IoU} = \frac{\text{Intersection of predicted and actual PV array areas}}{\text{Union of predicted and actual PV array areas}} \tag{6}$$

$$\text{mIoU} = \text{Average IoU across all classes} \tag{7}$$

4 Experiments And Results

4.1 Expirment 1 - Baseline Evaluation: AUS Data (Train) to AUS Data (Test)

First we conducted an initial experiment without using transferability techniques. It is important to set a baseline for how well the models perform on the Australian dataset. This helps you see how each model handles the data without any domain shift and allows you to understand the natural strengths and weaknesses of each model. It provides a clear reference point for later comparison:

Model/Metrics	mIoU	Precision	Recall	F1-Score
U-Net	0.7181	0.7613	0.5056	0.6083
Linknet	0.6847	0.5246	0.5586	0.5411
FPN	0.6295	0.7071	0.2917	0.4130
PSP-Net	0.6224	0.7169	0.2724	0.3948

Table 1: Model evaluation results metrics on the Baseline evaluation AUS data.

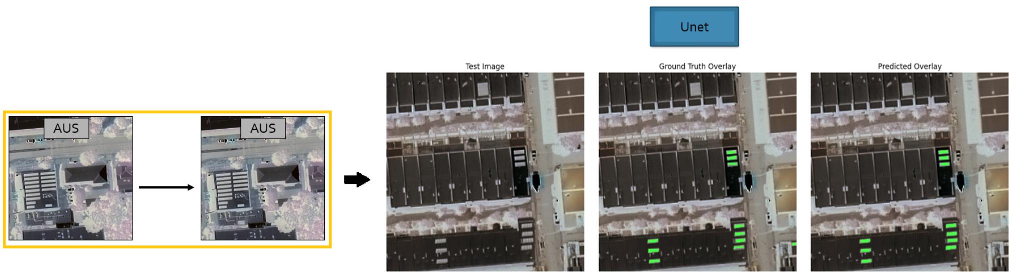

Figure 5: Example for U-Net prediction on AUS data in Baseline expirment.

4.2 Expirment 2 - Raw Transferability: AUS Data (Train) to USA Data (Test)

Testing raw transferability, is crucial for evaluating how well a model generalizes to different domains with varying characteristics. It helps identify domain gaps, benchmark performance, and assess the model's real-world applicability, guiding the development of more robust and adaptable solutions.The results showed a significant drop of 10% to 20% in mIoU due to domain shifts and differences in data distribution, highlighting the challenges of different geographical domains.

Model/Metrics	mIoU	Precision	Recall	F1-Score
U-Net	0.5154	0.5592	0.0365	0.0670
Linknet	0.5106	0.0471	0.0911	0.0621
FPN	0.5282	0.3460	0.0638	0.1141
PSP-Net	0.5091	0.3510	0.0232	0.0434

Table 2: Model evaluation results metrics on the Raw Transferability.

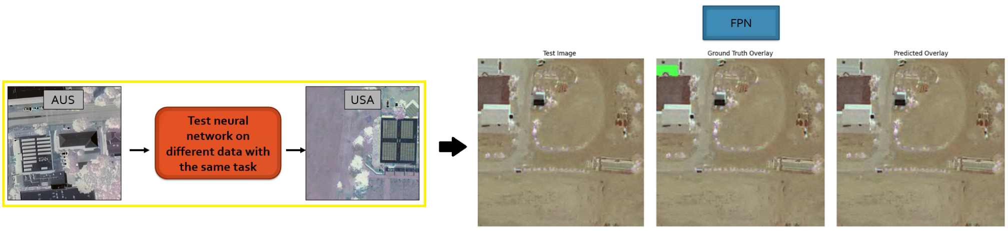

Figure 6: Example for FPN prediction on USA data in Raw Transferability expirment.

4.3 Expirment 3 - DA, AdaptSegNet: AUS Data (Train) to USA Data (Test)

This experiment assesses how well the model can adapt to the domain shift between these two datasets, which differ in geographical and visual characteristics. The goal is to determine if AdaptSegNet improves performance on the USA data by effectively transferring knowledge from the AUS dataset. By attempting to resolve the domain gap challenge and create invariant features through adversarial training, we hoped to see significant improvements. However, the results barely went up,this highlights how challenging it is to close the gaps between different domains, even when using DA approaches

Model/metrics	mIoU	Precision	Recall	F1-Score
Unet	0.5166	0.6383	0.0385	0.0727
Linknet	0.5300	0.6133	0.0660	0.1191
FPN	0.5097	0.5458	0.0233	0.0447
PSPNet	0.5181	0.7557	0.0400	0.0759

Table 3: Model evaluation results metrics on AdaptSegNet DA.

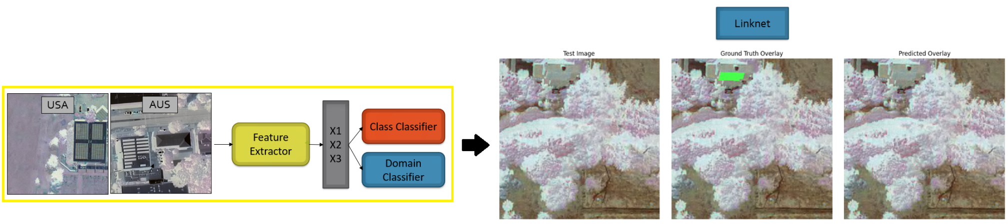

Figure 7: Example for Linknet prediction on USA data in AdaptSegNet expirment.

4.4 Expirment 4 - DA, AdaptSegNet And Fine Tuning (FT): AUS Data (Train) to USA Data (Test)

In this experiment, we explored how combining AdaptSegNet with fine-tuning can improve model performance when the data changes between domains. We first trained the model on source domain data, then used the saved weights for adversarial training with improved discriminator that includes sigmoid activation on the last layer and 2D convs layers with different input and out put channels on target domain data. Adversarial training helps the model learn features that are similar across both domains, making it difficult for the discriminator to tell them apart. The combination with fine-tuning allows the model to produce better, more domain-invariant features by specifically refining its focus on the target domain. In addition we added early stoping function to prevent over fit. However, even with this approach, the results barely improved, indicating that there is still much to be explored and understood in the DA area.

Model/metrics	mIoU	Precision	Recall	F1-Score
Unet	0.5260	0.4308	0.0610	0.1069
Linknet	0.5218	0.3944	0.0508	0.0900
FPN	0.5104	0.5132	0.0248	0.0474
PSPNet	0.5174	0.6607	0.0394	0.0743

Table 4: Model evaluation results metrics on AdaptSegNet DA and FT.

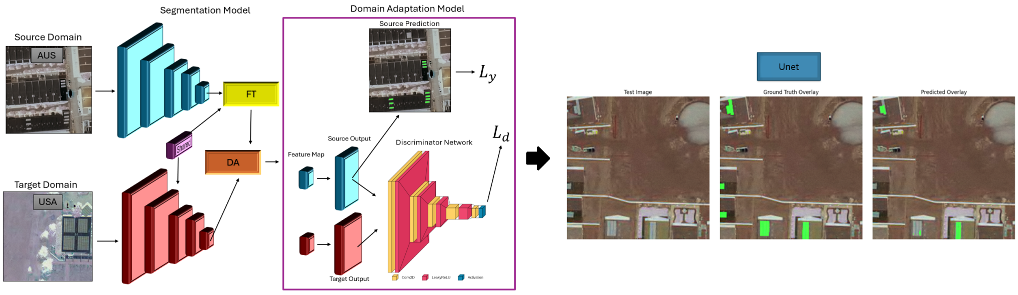

Figure 8: Example for U-Net prediction on USA data with AdaptSegNet, fine tuning and improved discriminator.

5 Discussion And Conclusions

In this project, we evaluated the effectiveness of deep learning models like U-Net, LinkNet, FPN, and PSPNet, all all based on ResNet-18, for segmenting solar panels from satellite imagery across different domains. Our study involved four different experiments to assess the models' performance in the context DA.

Experiment 1: We trained and tested the models on AUS data to establish a baseline. U-Net achieved a high mIoU score of 0.7181, thanks to its ability to preserve spatial details. LinkNet and FPN also performed well, leveraging their own unique approaches to feature preservation. PSPNet's global context awareness allowed it to stay competitive, although it didn't offer a major advantage in this homogenous dataset.

Experiment 2: We tested the models' transferability by applying them to USA data without retraining. This led to a significant drop in performance, with mIoU scores decreasing by 10% to 20%. The models struggled to generalize due to differences in lighting, panel types, and satellite characteristics between the two regions. U-Net and LinkNet, which rely on preserving specific spatial features, were particularly affected. FPN and PSPNet handled the domain shift slightly better, but still faced challenges.

Experiment 3: We applied AdaptSegNet for DA, which aims to align feature distributions across domains using adversarial training. While this slightly improved the mIoU scores (e.g., LinkNet improved from 51% to 53%), the overall impact was minimal. The models' architecture limitations, such as U-Net's focus on local features, restricted their ability to fully benefit from domain adaptation.

Experiment 4: Finally, we combined AdaptSegNet with fine-tuning, where models were refined on the target domain after initial training. These small improvements suggest that the models had already reached their potential for learning domain-invariant features. The ResNet-18 architecture might have been a limiting factor, and more complex architectures could offer better generalization.

These findings help us understand the importance of using DA to improve model generalization, especially in efforts to close domain gaps and create similar distributions for data from different geographical regions. The results also highlight the architectural challenges in adapting models designed for homogenous data to varied and complex real-world scenarios.

For future research, we recommend exploring additional DA techniques, such as multi-level adversarial training or incorporating self-supervised learning approaches. Another area of interest could be experimenting with different backbone architectures or ensemble methods to further enhance the models' robustness across diverse datasets. Finally, expanding the dataset to include more diverse regions and environments could provide a more comprehensive evaluation of the models' generalization capabilities.

References

[1] Julian de Hoog, Stefan Maetschke, Peter Ilfrich, and Ramachandra Rao Kolluri. Using satellite and aerial imagery for identification of solar pv: State of the art and research opportunities. In *Proceedings of the Eleventh ACM International Conference on Future Energy Systems*, pages 308–313, 2020.

[2] Abolfazl Farahani, Sahar Voghoei, Khaled Rasheed, and Hamid R Arabnia. A brief review of domain adaptation. *Advances in data science and information engineering: proceedings from ICDATA 2020 and IKE 2020*, pages 877–894, 2021.

[3] Yanming Guo, Yu Liu, Theodoros Georgiou, and Michael S Lew. A review of semantic segmentation using deep neural networks. *International journal of multimedia information retrieval*, 7:87–93, 2018.

[4] Shijie Hao, Yuan Zhou, and Yanrong Guo. A brief survey on semantic segmentation with deep learning. *Neurocomputing*, 406:302–321, 2020.

[5] Jordan M Malof, Kyle Bradbury, Leslie M Collins, and Richard G Newell. Automatic detection of solar photovoltaic arrays in high resolution aerial imagery. *Applied energy*, 183:229–240, 2016.

[6] Jordan M Malof, Leslie M Collins, Kyle Bradbury, and Richard G Newell. A deep convolutional neural network and a random forest classifier for solar photovoltaic array detection in aerial imagery. In *2016 IEEE International conference on renewable energy research and applications (ICRERA)*, pages 650–654. IEEE, 2016.

[7] Rongjun Qin, Guixiang Zhang, and Yang Tang. On the transferability of learning models for semantic segmentation for remote sensing data. *arXiv preprint arXiv:2310.10490*, 2023.

[8] Yi-Hsuan Tsai, Wei-Chih Hung, Samuel Schulter, Kihyuk Sohn, Ming-Hsuan Yang, and Manmohan Chandraker. Learning to adapt structured output space for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7472–7481, 2018.

[9] M Arif Wani and Tahir Mujtaba. Segmentation of satellite images of solar panels using fast deep learning model. *International Journal of Renewable Energy Research (IJRER)*, 11(1):31–45, 2021.

[10] Junshi Xia, Naoto Yokoya, Bruno Adriano, Lianchong Zhang, Guoqing Li, and Zhigang Wang. A benchmark high-resolution gaofen-3 sar dataset for building semantic segmentation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:5950–5963, 2021.

Individual Contributions

Inbal Karibian: Implemented and experimented with the LinkNet model, ran experiments to assess the model's performance on the dataset, wrote the related work section in the report, and collected literature references for the work. explored different DA approaches and decided which one we will use.

Arad Peleg: Implemented the U-Net model, wrote the majority of the codebase for the project, ensured consistency and integration across all models, provided the code and support for other team members to run their models, debugged and refined the overall project, was responsible for designing and conducting the experiments, analyzing the results, writing the abstract and conclusion, creating the figures, and collecting literature references for the work and created GitHub repository.

Amir Boger: Handled the implementation and setup of the PSP-Net model, conducted experiments to evaluate the model's effectiveness, was responsible for data acquisition and processing, wrote the Materials and Methods section in the report, and combined all the different parts of the report into one document.

Dolev Hagbi: Implemented and configured the FPN model, ran experiments to test the model's performance,was responsible for data acquisition and processing, proposed the idea of combining fine-tuning with domain adaptation, wrote the introduction of the report.