Enhancing Power Line Segmentation for UAV Inspection utilizing Synthetic Data

Georgios Kalitsios
Aristotle University of Thessaloniki, Department of Informatics
Thessaloniki, Greece
kalitsios@csd.auth.gr

Vasileios Mygdalis
Aristotle University of Thessaloniki, Department of Informatics
Thessaloniki, Greece
mygdalisv@csd.auth.gr

Ioannis Pitas
Aristotle University of Thessaloniki, Department of Informatics
Thessaloniki, Greece
pitas@csd.auth.gr

***Abstract*—The task of autonomous inspection of the power transmission infrastructure includes a robotic system that is able to track the power line. This work addresses the problem of the visual localization of the power lines from RGB cameras, using deep neural networks. For challenging-to-label tasks like this, simulators can efficiently generate large amounts of labeled data. In this work, a large-scale annotated synthetic power lines dataset has been generated[1]. To address domain shift between real and synthetic domain, input-level adaptation was performed. Additionally, a new power line segmentation loss developed to mitigate the effects of unbalanced pixel distributions among power lines and background. Experiments demonstrate that our approach achieves state-of-the-art performance on power line segmentation task.**

***Index Terms*—Power line segmentation, Synthetic dataset, Domain adaptation, UAVs inspection, Aerial images**

I. Introduction

Transmission line networks have almost expanded everywhere due to rising electrical consumption. To perform the actual inspection autonomously, it is important that the robotic system is able to identify the power lines. From the computer vision perspective, this task can be addressed either as a line detection problem [1], or as a pixel-level segmentation problem [2], which is the focus of this work. The task is particularly challenging because the power lines are difficult to detect even for the human eye, since they are small typically appear visually similar with other high-frequency feature elements, e.g., a plowed field, roof ridges and road lines and they must be distinguished among many different backgrounds and vegetation. Therefore, training a neural network that will be robust for this task, requires vast amounts of data for training purposes.

Power line segmentation is a crucial component of the Unmanned Aerial Vehicles (UAVs) intelligent power line inspection process for power-grid security and low-altitude safety. The effectiveness of segmentation methods is dependent on the availability of labeled training data. Collecting and annotating large datasets for any new segmentation task is costly, time-consuming, labor-intensive, and error-prone. In this work, we created a large-scale annotated synthetic dataset, that captures the diversity of the appearance across various flight altitudes and geographic locations. Nevertheless, the diversity of the different backgrounds can not be optimally modeled in a simulated environment. In order to address the domain shifts, we have employed Domain Adaptation (DA) to bridge the distribution gap between the real domain and the synthetic domain, hence boosting the generalization capabilities of learned models on real data.

The following are the major contributions of this work: *i) a large-scale synthetic power lines dataset.* To the best of our knowledge, this is the first synthetic dataset including power lines that will be made publicly available, and it constitutes a significant contribution for both the scientific community and industry. *ii) a new DA framework for power line segmentation*, particularly input-level adaptation performed utilizing a lightweight fourier-based image-to-image translation strategy. This strategy outperforms well-known domain adversarial adaptation strategies in terms of training speed, ease of implementation, and memory. *iii) a weighted power line segmentation loss function* combining focal loss and dice loss was developed to mitigate the impact of the unbalanced distribution of pixels between power lines and background on segmentation accuracy. *iv) The proposed framework achieve SOTA performance* in the task of power line segmentation when applied to the well-known TTPLA power lines dataset [3].

II. Related Work

A. Power line Segmentation

Traditional approaches identify power lines primarily by recognizing power line features [1] or objects associated [4] with power lines based on the assumption that power lines are either straight lines or polynomial curves that are parallel to each other. Such methods begin by distinguishing potential power line pixels from the background utilizing edge detector, then they employ hough transformation [5], followed by previous knowledge to refine the detected outcomes. Most of the previous approaches rely on complex parameter adjustments, making them less stable in practice.

Since Deep Neural Networks (DNNs) can automatically learn useful features and enable an end-to-end solution, researchers have started to build power line inspection systems leveraging DL. In [6] authors proposed an attentional convolutional network for pixel-level, which consists of an encoder-decoder information fusion module as well as an attention

[1]that will be made publicly available along with the unity scenes upon paper acceptance.

module. A fast single-shot line segment detector trained with artificially generated power line images proposed in [7]. In [2] an approach based on Generative Adversarial Networks (GANs) is described for segmenting power lines from aerial images.

B. Domain Adaptation

Adaptation methods can be applied at several levels, including the input-image level, internal-feature representations, and output-level. Recently, the majority of approaches have relied on adversarial learning [8] enabling pixel or feature-level adaptations or self-training [9] through the refinement of pseudolabels. In adversarial learning, the core issue is the necessity of learning multiple networks in addition to the target one, and the challenge of stabilizing adversarial training, but also the self-training approach that is self-referential, demands careful design to avoid error propagation.

III. INTRODUCED SYNTHETIC DATASET

There are two works with artificial datasets containing power lines [7], [10], but both of them are not very photo-realistic and are not publically available. The previous issues served as inspiration for this work. First and foremost, there is a demand for a publicly available large-scale Synthetic Power Lines (SPL) dataset consisting of thousands of RGB images. Moreover, this dataset can expand at any time, at any desired resolution, without requiring the sensor to be adjusted or the rebuilding of any previously used environments, and it is also considerably more photorealistic as shown in Fig. 1. The SPL

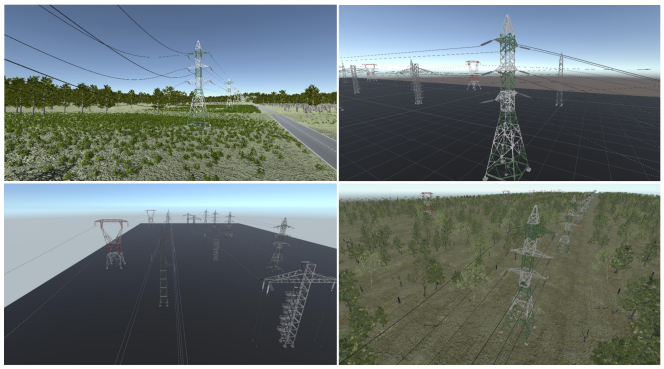

Figure 1: Images from the synthetic power lines dataset captured in various environments, angles, and camera distances.

dataset was generated utilizing unity, one of the most popular gaming engines in the world, and unity's perception package. The introduced dataset containing aerial images of power lines in every possible combination of different locations, lighting conditions, tilts, angles, FOVs. All power lines in SPL are recorded from several viewpoints, including front view, top view, and side view. As a result, there are almost no occlusions and UAVs can fly in any direction without concerned about detection accuracy. Power lines must be correctly detected against backgrounds, the SPL includes a large number of images of power lines with noisy backgrounds, making the process of extracting power lines challenging and close to a real-world scenario. The SPL dataset statistics are summarized in Table 1, and were divided into 70% for training, 10% for validation, and 20% for testing.

TABLE I: SPL Dataset Statistics.

Resolution	World 1	World 2	Total Images	Train set	Validation set	Test set
1,920×1,080	5,770	7,080	12,850	8,995	1,285	2,570

IV. PROPOSED FRAMEWORK

A. Fourier-based Adaptation:

Although the majority of inter-domain low-level statistical differences lack any semantic importance, they are likely to cause an unexpected performance drop on target samples, even though the source and target images share a higher level of semantic similarity in terms of scene structure and content. This is important as it seems that DNNs do not really transfer well between various low-level statistics. Inspired by [11] we employ a Fourier-based translation technique at the image input level by swapping a synthetic image's amplitude spectrum with that of a random real image. This method, does not employ discriminators that align pixel/feature-level distributions or image-to-image translation networks to generate training images. Fourier DA utilized as a separate step and does not require any training to achieve domain alignment, instead relying on a simple Fourier Transform and its inverse.

In DA, given a source dataset $\mathcal{D}_s = \{(x_i^s, y_i^s)\}_{i=1}^{N_s}$ and a target dataset $\mathcal{D}_t = \{(x_i^t, y_i^t)\}_{i=1}^{N_t}$, where $x^s, x^t \in \mathbb{R}^{H\times W\times 3}$ is a RGB image, and $y^s, y^t \in \mathbb{R}^{H\times W}$ is the segmentation map associated with x^s, x^t accordingly. Utilizing Fourier-based DA, we aim to bridge the domain discrepancy between the two datasets and enhance performance on the $\mathcal{D}_T$. The amplitude and phase components of the Fourier Transform are defined as:

$$\Pi(x)(k,l) = \sqrt{\mathcal{R}^2(x)(k,l) + \mathcal{I}^2(x)(k,l)} \tag{1}$$

$$\Phi(x)(k,l) = \arctan\Big[\frac{\mathcal{I}(x)(k,l)}{\mathcal{R}(x)(k,l)}\Big] \tag{2}$$

where the real and imaginary components of the Fourier Transform $\mathcal{F}\ (x)$ are represented by $\mathcal{R}\ (x)$ and $\mathcal{I}\ (x)$, respectively. Assuming that the image's center is $(0,0)$, we indicate with $\mathcal{M}\ (\beta)$ a mask whose value is zero everywhere but not in the region where $\beta \in [0,1]$. The altered spectral representation of x^s, indicated as $\mathcal{X}(x^s, x^t)$, in which the low frequency component of the amplitude of source image $\Pi\ (x^s)$ is swapped with that of the target image $\Pi\ (x^t)$, can be formalized as:

$$\mathcal{X}(x^s, x^t) = \mathcal{M}\ (\beta) \cdot \Pi\ (x^t) + (1 - \mathcal{M}\ (\beta)) \cdot \Pi\ (x^s) \tag{3}$$

Fourier-based adaptation formalized given a set of randomly sampled images x_s, x_t as:

$$x^{s\to t} = \mathcal{F}^{-1}\ (\big[\mathcal{X}(x^s, x^t), \Phi\ (x^s)\big]) \tag{4}$$

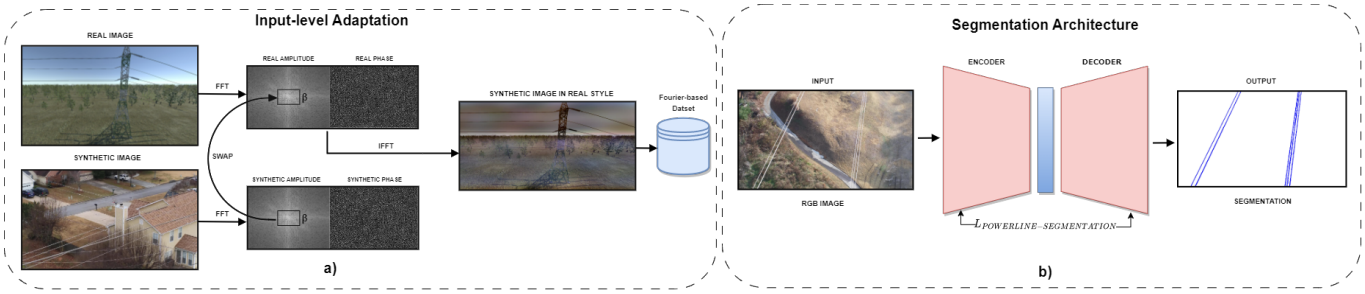

Figure 2: The proposed framework consists of two components: a) an input-level DA module that employs a Fourier-based image translation strategy, b) a high-performance segmentation architecture trained with a power line segmentation loss.

Where the altered spectral representation $\mathcal{X}(x^s, x^t)$ is projected back to the image $x^{s \to t}$ while keeping the phase component $\Phi\ (x^s)$ unaltered, whose content is similar to x^s, but whose appearance is similar to a sample from $\mathcal{D}_T$.

B. Segmentation Architecture:

DeeplabV3+ [12], the most recent DeepLab family version that combines a wide range of strategies such as skip connections, dilated convolution, global context, and robust backbone network, was selected in this this work. As an encoder, DeeplabV3+ employed the DeepLabV3 [13]. The segmentation architecture, as shown in Fig. 2, was trained using the proposed power line segmentation loss.

C. Power line segmentation loss:

Power line inspection is challenging due to poor visual appearance and complex backgrounds. Dice loss [14] considers both local and global loss information, has no trouble learning from classes with less spatial representation inside an image, and focuses mostly on mining the foreground during the training phase. Dice loss, formulated as follows:

$$\mathcal{L}_{dice} = 1 - \frac{2 \cdot \sum_{i=1}^{H \cdot W} p_i \cdot g_i}{\sum_{i=1}^{H \cdot W} p_i^2 + \sum_{i=1}^{H \cdot W} g_i^2 + \epsilon} \tag{5}$$

where p_i is the i-th pixel's estimated probability and g_i is the i-th pixel's ground truth. The ϵ term is employed in this case to guarantee the stability of the loss function by addressing the numerical problem of dividing by zero. However, it only solves the problem among foreground and background while ignoring another imbalance among easy and hard examples.

To address this restriction of dice loss, we employ focal loss [15], which emphasizes on examples where the model is inaccurate rather than examples where it can reliably estimate, allowing predictions on difficult examples to get better over time rather than the model being overconfident with easy ones. This is accomplished by a process known as down weighting. Focal loss, formulated as follows:

$$\mathcal{L}_{focal} = \begin{cases} -\alpha \cdot \sum_{i=1}^{H \cdot W} (1 - p_i)^{\gamma} \cdot \log(p_i), & g_i = 1 \\ -(1 - \alpha) \cdot \sum_{i=1}^{H \cdot W} p_i^{\gamma} \cdot \log(1 - p_i), & g_i = 0 \end{cases} \tag{6}$$

The weight vector $\alpha = [0.25, 075]$ for background and power lines class, respectively. The higher the value of focusing parameter γ, the greater the attention paid to misclassified hard examples (power lines) and the smaller the loss propagated from simple examples (background), based on our experiments, we set $\gamma = 3$.

In this work, in order to enhance the performance of power line segmentation on aerial images we introduced a weighted loss function for power line segmentation that combines focal loss and dice loss to benefit from the advantages of both. When the loss functions are combined, we get our complete learning objective as:

$$\mathcal{L}_{total} = \lambda_1 \cdot \mathcal{L}_{focal} + \lambda_2 \cdot \mathcal{L}_{dice} \tag{7}$$

where λ_1, λ_2 regulates the compromise between the dice loss and the focal loss. Based on the results of our experimental analysis, λ_1, λ_2 are both set equal to 0.5.

V. EXPERIMENTAL EVALUATION

A. Datasets and Implementation Details

There are not many publically available datasets for power lines. For experimental evaluation, the SPL dataset presented in this work and the well-known real-world TTPLA dataset [3] were employed. We utilized 905 images from the TTPLA dataset for training, 110 images for validation, and 217 images for testing. We resized the images from both datasets to 1024×1024.

The ResNet-50 backbone pre-trained on ImageNet was utilized for feature extraction. To optimize our network, we utilize SGD, using an initial learning rate of 0.01, momentum of 0.9, and weight decay of 4e-5. The "poly" learning rate policy is employed to control how the learning rate decays during training. The model was trained on a single NVIDIA RTX 2080 TI GPU with 11GB of VRAM for 100 epochs.

B. Baselines

Table II, second row, demonstrates that training with just the SPL dataset decreases TTPLA performance compared to training with just the TTPLA dataset, Table II, first row. This happens as a result of the domain shift from synthetic to real. However, when both the TTPLA and SPL datasets are combined, we can observe in the third row that there is

TABLE II: Datasets and Loss functions Experiments.

Training Dataset	SPL mIOU (%)		TTPLA mIOU (%)		Loss function	
	Val	Test	Val	Test	Standard CE	Power Line Segmentation
TTPLA	**1.60**	**2.03**	**47.92**	**44.51**	√	
SPL	**56.66**	**55.14**	**18.93**	**17.32**	√	
TTPLA & SPL	**50.05**	**49.70**	**53.11**	**52.83**	√	
TTPLA & SPL	**51.89**	**50.34**	**57.58**	**55.24**		√

a considerable improvement in TTPLA, Table II, first row, particularly $+5,19\%$ on the val set and $+8,32\%$ on the test set.

C. Loss function:

We compare the power line segmentation loss introduced in these experiments to the standard CE loss used in semantic segmentation. The power line segmentation loss improved performance is $+4.47\%$ in the validation set and $+2.41\%$ in the test set, as shown in Table III

D. Domain adaptation:

We tune β until artifacts in the transformed images are visible, which occurs when β exceeds 0.25. According to our findings, an intermediate value for $\beta = 0,01$ yields the greatest outcomes (58.47% in the TTPLA test set). We gain $+2.10\%$ on the validation set and $+3.23\%$ on the test set with input-level DA, as shown in Table IV.

TABLE III: Input image level Fourier-based DA Experiments.

Domain Adaptation	TTPLA (Real dataset) mIOU	
	Val	Test
Without Domain Adaptation	**57.58 %**	**55.24 %**
Fourier Adaptation ($\beta = 0,001$)	**57.73 %**	**56.20 %**
Fourier Adaptation ($\beta = 0,01$)	**58.39 %**	**58.47 %**
Fourier Adaptation ($\beta = 0,25$)	**59.68 %**	**56.09 %**

E. Comparison with SOTA method on TTPLA dataset:

To ensure fairness, the same ResNet-6 backbone and 512×512 resolution are chosen. Based on our previous results, we apply fourier-based adaptation with $\beta = 0,01$. On the TTPLA test set, our method outperforms the recently presented PLGAN architecture [2] by $+3.82\%$, as shown in Table V.

TABLE IV: Comparison on TTPLA dataset.

Method	Backbone Network	Image Resolution	TTPLA (Real dataset) Test set mIOU
PLGAN [2]	**ResNet-6**	**512x512**	**53.30 %**
Ours	**ResNet-6**	**512x512**	**57.12 %**

VI. CONCLUSION

In this work, a novel large-scale synthetic dataset for training and testing power line segmentation approaches generated, reducing the need to gather and label a large number of real-world images. Furthermore, a novel power line segmentation DA framework that can close the gap produced by the domain shift between the synthetic and real image presented, and it is a highly promising direction based on our experimental results. In addition, a new weighted loss function proposed that combines focal loss and dice loss to improve the performance of power line segmentation on aerial images. Finally, our framework achieves state-of-the-art performance for power line segmentation task on the well-known real-world TTPLA dataset.

ACKNOWLEDGMENT

This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 871479 (AERIAL-CORE).

REFERENCES

[1] Z. Li, Y. Liu, R. Walker, R. Hayward, and J. Zhang, "Towards automatic power line detection for a uav surveillance system using pulse coupled neural filter and an improved hough transform," *Machine Vision and Applications*, vol. 21, no. 5, pp. 677–686, 2010.

[2] R. Abdelfattah, X. Wang, and S. Wang, "Plgan: Generative adversarial networks for power-line segmentation in aerial images," 2022. [Online]. Available: https://arxiv.org/abs/2204.07243

[3] ——, "Ttpla: An aerial-image dataset for detection and segmentation of transmission towers and power lines," in *Proceedings of the Asian Conference on Computer Vision*, 2020.

[4] J. Zhang, H. Shan, X. Cao, P. Yan, and X. Li, "Pylon line spatial correlation assisted transmission line detection," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, no. 4, pp. 2890–2905, 2014.

[5] L. A. Fernandes and M. M. Oliveira, "Real-time line detection through an improved hough transform voting scheme," *Pattern recognition*, vol. 41, no. 1, pp. 299–314, 2008.

[6] Y. Li, Z. Xiao, X. Zhen, and X. Cao, "Attentional information fusion networks for cross-scene power line detection," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 10, pp. 1635–1639, 2019.

[7] V. N. Nguyen, R. Jenssen, and D. Roverso, "Ls-net: Fast single-shot line-segment detector," *Machine Vision and Applications*, vol. 32, no. 1, pp. 1–16, 2021.

[8] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *The journal of machine learning research*, vol. 17, no. 1, pp. 2096–2030, 2016.

[9] Y. Zou, Z. Yu, B. Kumar, and J. Wang, "Unsupervised domain adaptation for semantic segmentation via class-balanced self-training," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 289–305.

[10] R. Madaan, D. Maturana, and S. Scherer, "Wire detection using synthetic data and dilated convolutional networks for unmanned aerial vehicles," in *Proceedings of (IROS) IEEE/RSJ International Conference on Intelligent Robots and Systems*, September 2017, pp. 3487 – 3494.

[11] Y. Yang and S. Soatto, "Fda: Fourier domain adaptation for semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 4085–4095.

[12] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.

[13] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.

[14] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. Jorge Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings 3*. Springer, 2017, pp. 240–248.

[15] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.