UAV-based Unsupervised Domain Adaptation for Road Extraction

Gustavo Rota Collegio[1], Antonio Gaudencio Guimarães Filho[2], Aluir Porfírio Dal Poz[1], Ayman Habib[3]

[1] Department of Cartography, Faculty of Sciences and Technology, São Paulo State University (UNESP), São Paulo 19060-900, Brazil - (gustavo.collegio, aluir.dal-poz)@unesp.br
[2] DSG, Brazilian Army Geographic Service, Brasília - DF, Brazil - guimaraesfilho.antonio@eb.mil.br
[3] Lyles School of Civil Engineering, Purdue University, West Lafayette, IN 47907, USA – ahabib@purdue.edu

Keywords: Road Extraction, UAV imagery, Unsupervised Domain Adaptation, Domain Adversarial Neural Network.

Abstract

Despite advances in Deep Learning (DL) for road extraction, this task remains challenging. First, domain shifts in data distribution hinder the inference of pre-trained models to new areas, leading to a drop in classification accuracy. Second, DL-based models require a large amount of labeled training data to achieve robust performance. To address these challenges, this work proposes an Unsupervised Domain Adaptation (UDA) approach leveraging the Domain Adversarial Neural Network (DANN) strategy applied to Unmanned Aerial Vehicle (UAV) imagery. While most existing approaches rely on satellite imagery, they may not generalize well to UAV data, as very high-resolution images with fine-grained road details introduce additional domain adaptation challenges. Furthermore, since DANN operates at the feature level, the design of the feature extractor plays a key role to achieve the domain alignment. To investigate this, we evaluate our approach with three segmentation models: DeepLabv3+, ResU-Net, and Att-ResU-Net, the latter incorporating attention-enhanced skip connections. Experimental results demonstrate that UDA effectively deals with domain shift, improving road extraction performance by 1.79-6.07% on F1-Score and 2.12-7.85% on IoU when tested on the target domain without labeled training data. Among the evaluated architectures, Att-ResU-Net achieves the highest UDA performance. The qualitative analysis through further illustrates how architectural differences impact UDA for UAV-based road extraction.

1. Introduction

Road extraction is an active and popular research topic. Since the road network is constantly in change, keeping it updated is one of the main motivations of the current approaches, in addition to the possibility of dealing with a wide range of applications (e.g., urban management, logistics, infrastructure, etc). Field surveys and manual labeling are the conventional methods to keep the road network up-to-date; however, they are cost and labor-intensive. Thus, given the possibility of automatic execution, high-resolution images have become the main source for road network mapping (Chen et al., 2022).

Several approaches have been proposed since the 1970s. However, the emergence of deep learning (DL) in the last decade has opened up a new era for road extraction (Filho et al., 2023). Taking Deep Convolutional Neural Networks (DCNNs) as a paradigm, such models are dominant and can outperform handcrafted methods. On the other hand, despite their advantages, road extraction remains a challenging task. In this regard, the inherent variability and the complex context are the main challenges that DL-based algorithms have to deal with. For instance, urban areas are more subject to changes compared to rural areas, making them more dynamic. Besides, variations in road geometries, occlusion cases, and objects that resemble road structures highlight diverse and complex contextual conditions.

According to the state-of-the-art, current approaches tackle these complexities based on strategies such as the attention mechanism (Dai et al., 2023) in DCNNs or Transformer-based models (Meng et al., 2024). Yet, two limitations persist. Firstly, deploying a pre-trained model in new areas is often unfeasible. Since DL-based models rely on implicitly capturing a dataset's probability distribution, changing the road network scenario or using different datasets leads to a shift in data distributions. As a result, the classification accuracy in the inference phase is significantly decreased. Secondly, a mass of labeled training data is a prerequisite to training and generalizing those models. Given that such labels are commonly unavailable and pixel-wise road annotation is a time-consuming process, current methods are limited to evaluations on benchmark datasets.

This statistical divergence that hinders a training set's ability to generalize well on a testing set is called domain shift in the literature (Wang and Deng, 2018), and datasets are typically referred to as the source and the target domains, respectively. In situations where both source and target road labels are available, transfer learning, which pre-trains a model on the source domain and fine-tuning it on the target domain, can overcome the domain shift problem. However, when target labels are unavailable, Unsupervised Domain Adaptation (UDA) becomes a viable approach.

A UDA-based model has to be trained to perform well in the target domain without the need for extra labeled data in that domain (Vega et al., 2021). There have been several studies on UDA in the field of semantic segmentation (Ma et al., 2023). However, few efforts have been made to investigate its application to road extraction (Iqbal et al., 2023), particularly using UAV imagery. Compared to satellite data, UAV imagery introduces additional challenges, such as increased local variations, more affected by occlusions and shadows, and complex road structures.

In addition, applying UDA for road extraction also depends on the DL-based strategy used, as different models may extract and represent features differently under domain shift. Whereas DL-based models are data-driven, the existing architectures are also shaped by assumptions about the nature of the task. For

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-3-W4-2025-79-2026

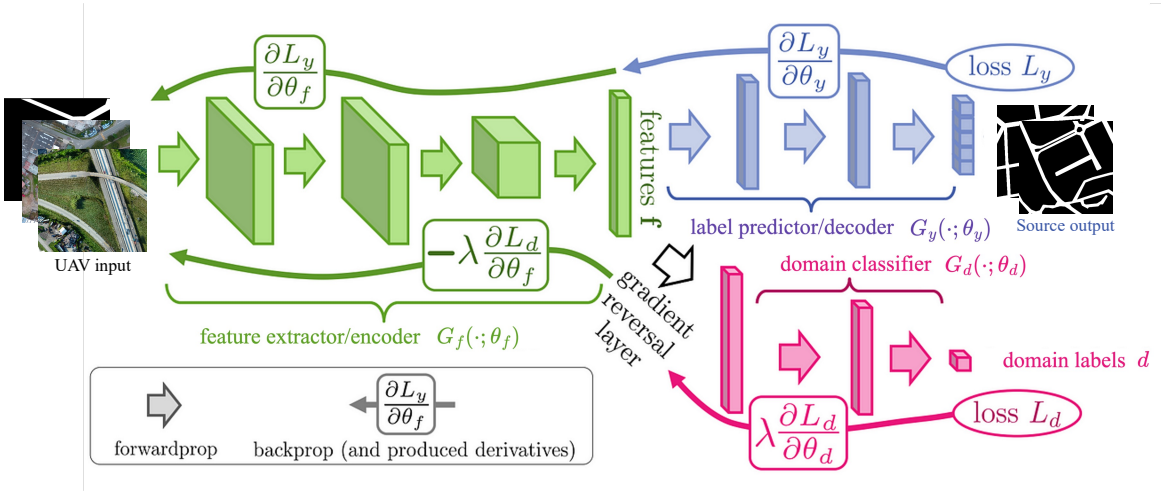

Figure 1. A segmentation model (encoder-decoder) adapted according to the DANN. Source: Adapted from Ganin et al. (2015).

instance, U-Net adopts the strategy of skip connections to preserve local spatial details (Ronneberger et al., 2015) that are important for road connectivity, while DeepLabv3+ leverages global multi-scale information (Chen et al., 2018). Attention mechanism, in turn, can further refine feature extraction, dynamically filter task-relevant features. However, how these architectural strategies influence model behavior when integrated with UDA remains unclear.

One widely adopted UDA framework is the Domain Adversarial Neural Network (DANN) (Ganin et al., 2016), which minimizes the divergence between two probability distributions via adversarial learning. Since DANN operates at the feature level (Figure 1), the design of the feature extractor plays a key role to achieve domain alignment, especially in domains like UAV imagery, where very high-spatial resolution makes the adaptation more challenging.

Motivated by the hypothesis that different segmentation architectures interact differently with DANN's feature extraction process, we evaluate three semantic models under a unified DANN-based UDA framework for UAV-based road extraction: DeepLabv3+ ResU-Net, and Attention ResU-Net (Att-ResU-Net). The main contributions of this work are threefold:

1. We propose a UAV-based UDA approach for road extraction using DANN, addressing domain shift without requiring target domain labels.

2. We analyze how different architectural strategies (i.e., skip connections, attention mechanisms, and multi-scale context) affect domain alignment and performance under UDA.

3. We provide both quantitative and qualitative evidence to assess the feature alignment behavior across architectures and its impact on road extraction.

2. Proposed Method

2.1 Fundamentals of DANN for UDA

DANN, proposed by (Ganin et al., 2016), introduces an adversarial learning strategy that combines discriminative and domain-invariance feature representation. Given two different probability distributions from the source and the target domains, the goal is to minimize their divergence through a deep neural network. To do so, DANN employs three modules (see Figure 1): a feature extractor $G_f(\cdot,\theta_f)$, a label predictor $G_y(\cdot,\theta_y)$, and a domain classifier $G_d(\cdot,\theta_d)$. In short, the first module extracts features from both domains to learn the data representation for a specific task (e.g., capturing road-like patterns for road extraction). The label predictor module acts as a supervised classifier, where only the source features are forwarded since their labels are known. In the domain classifier, however, both source and target features are forwarded, aiming to identify the domain to which an input sample belongs. The optimal function parameters for each module $(\theta_f^*, \theta_y^*, \theta_d^*)$ are given by (Ganin et al., 2016):

$$(\theta_f^*, \theta_y^*) = \text{argmin}_{\theta_f,\theta_y} E(\theta_f, \theta_y, \theta_d^*) \quad (1)$$

$$(\theta_d^*) = \text{argmax}_{\theta_d} E(\theta_f^*, \theta_y^*, \theta_d) \quad (2)$$

where the DANN loss function (E) to be optimized is expressed as:

$$E(\theta_f, \theta_y, \theta_d) = \mathcal{L}_y(\theta_f, \theta_y) - \lambda \mathcal{L}_d(\theta_f, \theta_d) \quad (3)$$

In (3), $\mathcal{L}_y$ is the loss for the label predictor, and $\mathcal{L}_d$ for the domain classifier. The feature mapping parameters θ_f play a dual role by minimizing $\mathcal{L}_y$ (i.e., ensuring discriminative features) while maximizing $\mathcal{L}_d$ (i.e., making domain-invariance features). Meanwhile, the parameters θ_d aims to minimize $\mathcal{L}_d$, and the parameters θ_y minimize $\mathcal{L}_y$.

The network is stimulated to learn domain-agnostic features across domains by multiplying the domain classifier loss by $-\lambda$. Such an operation is called the Gradient Reversal Layer (GRL), which acts as an identity mapping during the forward pass but reverses the gradients that come from the domain classifier. Finally, the saddle point of Eqs. (1) and (2) is obtained through stochastic-based optimization, as defined in the following equations:

$$\theta_f \leftarrow \theta_f - \mu \left(\frac{\partial \mathcal{L}_y}{\partial \theta_f} - \lambda \frac{\partial \mathcal{L}_d}{\partial \theta_f} \right) \quad (4)$$

$$\theta_y \leftarrow \theta_y - \mu \frac{\partial \mathcal{L}_y}{\partial \theta_y} \quad (5)$$

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-3-W4-2025-79-2026 | © Author(s) 2026. CC BY 4.0 License.

Table 1. Characteristics of Source and Target Domains.

Feature	Source Domain	Target Domain
Location	El Retiro, Colombia	I-65, USA
Context	Urban	Suburban
GSD	7 cm	3 cm
Bands	RGB (8 bits)	RGB (8 bits)

$$\theta_d \leftarrow \theta_d - \mu\lambda\frac{\partial \mathcal{L}_d}{\partial \theta_d} \tag{6}$$

where μ is the learning rate.

2.2 UAV datasets

The proposed method is based on UAV data, ensuring a consistent acquisition method across domains despite technical differences. As summarized in Table 1, the source domain consists of the publicly available El Retiro dataset (Ballesteros et al., 2022), an urban area with orthomosaic and road reference data. In contrast, the target domain corresponds to a survey conducted along Interstate 65 (I-65) in the United States (mile markers 141 and 178). The dataset corresponds to a suburban environment, where the target roads were manually labeled for evaluation.

2.3 Architectures

In our work, the proposed approach addresses UAV data based on different encoder-decoder architectures to incorporate contextual information strategies and explore the spatial relationship between objects in the scene, namely DeepLabv3+ (Chen et al., 2018), ResU-Net (He et al., 2015), and Att-ResU-Net, with the introduction of an attention block (Oktay et al., 2018) in ResU-Net's skip connections. Each architecture adopts a unique strategy, but all use the ResNet-101 backbone (He et al., 2015) to enhance the feature extraction and ensure fair comparisons. The encoders are adapted following the DANN's feature extractor, while the decoders work as the label predictor module. Additionally, a domain classifier is implemented, transforming the architectures into UDA approaches.

According to (Chen et al., 2018), DeepLabv3+'s encoder employs a module called Atrous Spatial Pyramid Pooling (ASPP) to capture contextual information. ASPP varies dilation rates (i.e., 1, 6, 12, 18) to adjust the filter's field of view of the parallel convolutions, improving multi-scale representation. Smaller receptive fields focus on local features, whereas large ones capture global context, complemented by the global average pooling operation. The decoder, in turn, refines the segmentation results along object boundaries.

In ResU-Net (He et al., 2015), as the encoder layers progress deeper into low-level features, important features are preserved; however, this process may result in a loss of spatial information related to the road network. To address this, the network leverages skip connections to recover encoded information at each level by linking the encoder and decoder layers. Since ResU-Net does not focus entirely on capturing contextual information, we introduce the attention block proposed by (Oktay et al., 2018) right before concatenating the information into the decoder. Specifically, in Att-ResU-Net, the attention block focuses on spatial regions, filtering out the feature maps to retain only the most relevant activations for road extraction.

The domain classifier, with the same setting across all three models, consists of the following layers: a 3x3 convolution with 256 filters, followed by a Leaky-ReLU activation, a flattening operation, a fully connected layer with 2 units, and a sigmoid activation to estimate the probability of features originating from the source or the target domain.

Moreover, we adopt an alternate update strategy during backpropagation. First, the encoder and decoder modules are updated while domain classifier gradients are calculated to compute the total loss function but are not updated. These gradients are recalculated and updated in the second step using a separate learning rate. This procedure aims to stabilize the adversarial training process.

3. Experiments

3.1 Protocol

Our experiments were conducted with an NVIDIA RTX 3090 24GB DDR6x graphics processing unit (GPU) using TensorFlow 2.15 version. With input patches of 1024x1024 pixels, 125 images were used to train (70%), 21 to validate (10%), and 30 to test (20%). Data augmentation was performed with the following operations: random flip (horizontal and vertical), 270° rotation, and image transposing.

Training employed the Focal Tversky loss (Abraham and Khan, 2018) (with α, β, and γ set to 0.3, 0.7, 3/4, respectively) which controls the classes' unbalancing of the roads for the decoder, and Binary Cross-Entropy (BCE) for the domain classifier (domain labels of 0 and 1). Stochastic Gradient Descent (SGD) optimized encoder-decoder modules with a batch size of 2. DeepLabv3+ decayed exponentially the learning rate from $1e{-}2$ up to $1e{-}3$ over 300 epochs, while ResU-Net and Att-ResU-Net from $1e{-}2$ to $1e{-}4$ over 200 epochs. The domain classifier of all models was optimized with the Adam optimizer (learning rate of $1e{-}4$). DeepLabv3+ achieved better performance with λ value set to 0.0001, while ResU-Net and Att-ResU-Net performed best with a value of 0.1.

The UDA-based models were evaluated against two baselines: the source-trained baseline tested on the target domain (lower bound) and the fully supervised target-trained model (upper bound). The UDA-based models and their respective baselines shared the same hyperparameter setting within each architecture. However, distinct architectures required specific hyperparameter tuning to achieve optimal performance. Accordingly, the results are discussed in two parts: first, we analyze the UDA and source-only models; then, we assess the models trained under full supervision on the target domain.

Quantitative analysis was carried out using F1-Score (F1) and Intersection over Union (IoU) metrics, as defined in (Filho et al., 2023). In addition, qualitative analysis was supported by the Score-CAM method (Wang et al., 2019), which provides a visual explanation through class activation mapping based on activation map weights. Since the encoder is the primary module of DANN for generating domain-agnostic features, only its layers are analyzed.

3.2 Performance of UDA Across Architectures

Under the same conditions, the quantitative evaluation results are summarized in Table 2. Our three UDA approaches outperform the baseline algorithms. By aligning feature distributions across domains, UDA mitigates domain shift and improves the

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-3-W4-2025-79-2026 | © Author(s) 2026. CC BY 4.0 License.

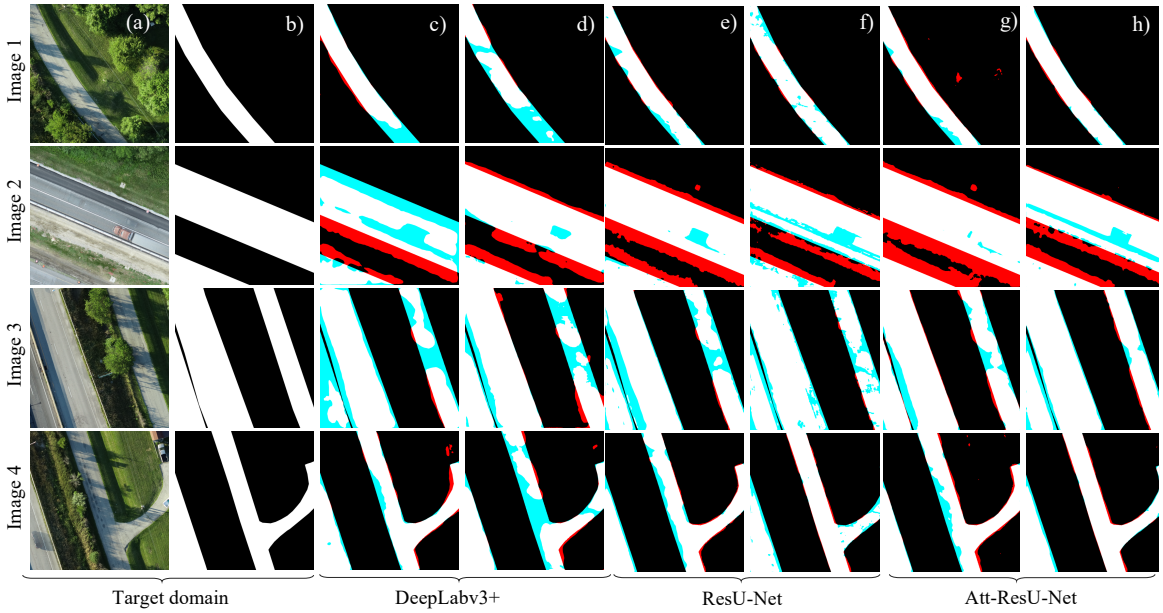

Figure 2. Predictions on the target domain (I-65, USA). Columns: (a) Target image; (b) Reference data; (c)-(d) Baseline and UDA for DeepLabv3+; (e)-(f) Baseline and UDA for ResU-Net; (g)-(h) Baseline and UDA for Att-ResU-Net. False positives (FP) are shown in red, and false negatives (FN) in cyan.

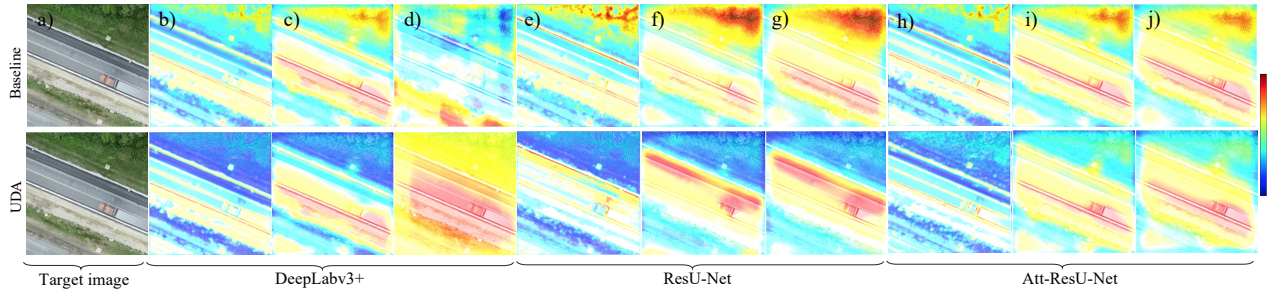

Figure 3. Encoder feature maps' visualization by Score-CAM. Weighted class activation maps on image overlap from the coarse to the fine level of the encoder. Columns: (a) Target image; (b)-(d) 2nd conv. block, 4th conv. block, and ASPP for DeepLabv3+; (e)-(g) 2nd, 4th, and 5th conv. block for ResU-Net; (h)-(j) 2nd, 4th, and 5th conv. block for Att-ResU-Net.

Table 2. UDA target domain's prediction metrics. Best UDA results are in bold and second ones are underlined.

Model		F1 (%)	IoU (%)
DeepLabv3+	Baseline	75.44	64.79
	UDA	79.12	69.2
ResU-Net	Baseline	77.39	67.13
	UDA	<u>81.06</u>	<u>72.28</u>
Att-ResU-Net	Baseline	79.72	70.52
	UDA	**81.51**	**72.64**

performance of road extraction by 1.79-6.07% on F1-Score and 2.12-7.85% on IoU when testing on the target domain without labeled training data. Specifically, DeepLabv3+ shows an improvement of 3.68% in F1-Score and 4.41% in IoU over its baseline. ResU-Net and Att-ResU-Net (the best performance) also benefit from domain alignment, with UDA improving their F1-Score by 3.67% and 1.79%, respectively, and IoU by 5.15% and 2.12%.

Qualitatively, Figure 2 depicts the predictions for the three architectures. Challenges such as different illumination conditions, shadows, and occlusion by cars over the road, typically observed in UAV imagery, indicate the occurrence of false positives and false negatives that affect road network extraction. While the UDA approaches improve road connectivity compared to the baselines, they remain sensitive to these factors. Notably, the UDA performance can vary depending on the architectural strategy. Taking Image 1 of Figure 2 (d), (f), and (h) as an example, DeepLabv3+ is more likely to struggle with these challenges when compared to ResU-Net and Att-ResU-Net. This suggests that the domain alignment primarily focuses on matching feature distributions rather than explicitly refining class boundaries to improve road delineation.

To further support this observation, Figure 3 illustrates the Score-CAM results at the encoder levels. From broader road-related features to more specific ones, the stronger activations observed in UDA-based models indicate that domain alignment enhances the network's ability to recognize domain-invariant road features. On the other hand, baselines show less focus on road features, often highlighting non-road regions. This could be due to the baseline network paying attention to features common across domains, such as vegetation.

At deeper encoder levels, different UDA strategies behave differently within UDA (cases (d), (g), and (j) of Figure 3), poten-

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-3-W4-2025-79-2026 | © Author(s) 2026. CC BY 4.0 License.

Table 3. Supervised training results on the target domain. Best values are in bold.

Model	F1 (%)	IoU (%)
DeepLabv3+	**89.05**	**82.53**
ResU-Net	85.31	78.41
Att-ResU-Net	86.74	80.14

tially affecting the final performance of road extraction. For instance, skip connections tend to activate more directly on road areas, unlike DeepLabv3+. Since the ASPP module in DeepLabv3+ provides a broader context, the high abstraction level may make distinguishing road features more challenging, introducing difficulties in refining localized road features under domain shift. In contrast, ResU-Net and Att-ResU-Net appear to refine road-specific features more effectively after UDA, with Att-ResU-Net's attention mechanism achieving the highest metrics.

3.3 Performance Under Full Supervision: An Upper-Bound Comparison

As expected, fully supervised training on the target domain outperforms UDA, since the model leverages full availability of labeled data. When compared to the first analysis (Table 2), Table 3 presents that DeepLabv3+ achieves the highest performance, with an F1-Score of 89.05% and an IoU of 82.53%. Att-ResU-Net follows closely, while ResU-Net shows the lowest scores among the three. These results suggest that road extraction through UAV imagery benefits more from multi-scale context (DeepLabv3+) in a fully supervised training, while only the skip connections strategy is less effective to handle road features.

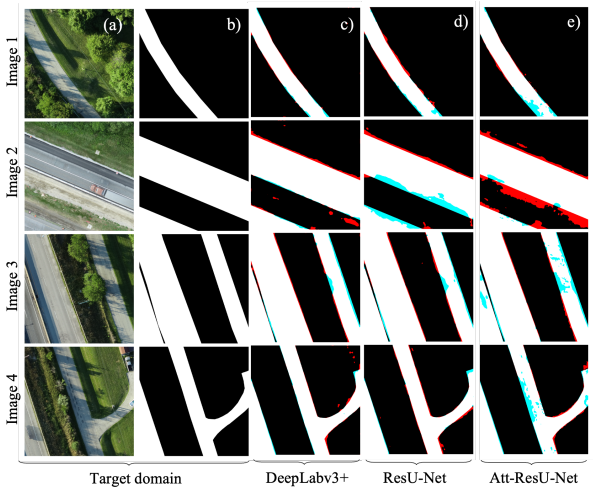

Figure 4. Fully supervised model's prediction on target domain.

Figure 4 illustrates the target predictions in this scenario. Compared to the UDA results (Figure 2), these predictions show fewer false negatives and false positives, resulting in improved road network connectivity. Moreover, while DeepLabv3+ achieves the highest performance under full supervision, it is outperformed by Att-ResU-Net in the UDA setting. This contrast suggests that architectural strengths vary depending on the training strategy and the availability of labeled data. Multi-scale information extracted by DeepLabv3+ benefits more from direct supervision, effectively learning global contextual information. However, Att-ResU-Net, attention-enhanced skip connections, indicates to be better suited to handle road extraction and keep focus on domain-invariant representations under domain shift.

4. Conclusion

This work proposed a UDA-based approach for road extraction using the DANN framework on UAV imagery. Leveraging very high-resolution images, different feature extraction strategies demonstrated the ability to handle such data, with Att-ResU-Net and ResU-Net, based on the skip connections strategy, achieving the best performance. Qualitative results showed that incorporating target-domain information, even unsupervised, improves activation on road areas. Although fully supervised approaches outperform UDA, our method does not require target-labeled data, making it valuable when such data are inaccessible.

While focused on urban and suburban UAV imagery, the approach generalizes across very high-resolution domains, indicating potential for extension to other platforms and broader geographic scales. Future work will explore more challenging environments, such as dense forests, where secondary roads increase domain shift, and investigate strategies to extend applicability beyond urban and suburban scenarios.

References

Abraham, N., Khan, N. M., 2018. A Novel Focal Tversky loss function with improved Attention U-Net for lesion segmentation. *Proceedings - International Symposium on Biomedical Imaging*, 2019-April, 683-687. https://arxiv.org/abs/1810.07842v1.

Ballesteros, J. R., Sanchez-Torres, G., Branch-Bedoya, J. W., 2022. A GIS Pipeline to Produce GeoAI Datasets from Drone Overhead Imagery. *ISPRS International Journal of Geo-Information 2022, Vol. 11, Page 508*, 11, 508. https://www.mdpi.com/2220-9964/11/10/508/htm https://www.mdpi.com/2220-9964/11/10/508.

Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11211 LNCS, 833-851. https://arxiv.org/abs/1802.02611v3.

Chen, Z., Deng, L., Luo, Y., Li, D., Junior, J. M., Gonçalves, W. N., Nurunnabi, A. A. M., Li, J., Wang, C., Li, D., 2022. Road extraction in remote sensing data: A survey. *International Journal of Applied Earth Observation and Geoinformation*, 112, 102833.

Dai, L., Zhang, G., Zhang, R., 2023. RADANet: Road Augmented Deformable Attention Network for Road Extraction From Complex High-Resolution Remote-Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, 61.

Filho, A., Shimabukuro, M., Poz, A. P. D., 2023. Deep Learning Multimodal Fusion for Road Network Extraction: Context and Contour improvement. *IEEE Geoscience and Remote Sensing Letters*.

Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., Lempitsky, V., 2016.

Domain-Adversarial Training of Neural Networks. *Advances in Computer Vision and Pattern Recognition*, 17, 189-209. https://arxiv.org/abs/1505.07818v4.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-Decem, 770-778. https://arxiv.org/abs/1512.03385v1.

Iqbal, J., Masood, A., Sultani, W., Ali, M., 2023. Leveraging topology for domain adaptive road segmentation in satellite and aerial imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 206, 106-117.

Ma, X., Zhang, X., Wang, Z., Pun, M. O., 2023. Unsupervised Domain Adaptation Augmented by Mutually Boosted Attention for Semantic Segmentation of VHR Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, 61.

Meng, Q., Zhou, D., Zhang, X., Yang, Z., Chen, Z., 2024. Road Extraction From Remote Sensing Images via Channel Attention and Multilayer Axial Transformer. *IEEE Geoscience and Remote Sensing Letters*, 21, 1-5.

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B., Rueckert, D., 2018. Attention U-Net: Learning Where to Look for the Pancreas. https://arxiv.org/abs/1804.03999v3.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv:1505.04597 [cs]*. http://arxiv.org/abs/1505.04597 http://files/21/Ronneberger et al. - 2015 - U-Net Convolutional Networks for Biomedical Image.pdf. Comment: conditionally accepted at MICCAI 2015.

Vega, P. J. S., da Costa, G. A. O. P., Feitosa, R. Q., Adarme, M. X. O., de Almeida, C. A., Heipke, C., Rottensteiner, F., 2021. An unsupervised domain adaptation approach for change detection and its application to deforestation mapping in tropical biomes. *ISPRS Journal of Photogrammetry and Remote Sensing*, 181, 113-128. https://linkinghub.elsevier.com/retrieve/pii/S092427162100232X.

Wang, H., Wang, Z., Du, M., Yang, F., Zhang, Z., Ding, S., Mardziel, P., Hu, X., 2019. Score-CAM: Score-Weighted Visual Explanations for Convolutional Neural Networks. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2020-June, 111-119. https://arxiv.org/abs/1910.01279v2.

Wang, M., Deng, W., 2018. Deep visual domain adaptation: A survey. *Neurocomputing*, 312, 135-153.