Convolutional Neural Networks for Road Detection: An Unsupervised Domain Adaptation Approach

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

Abstract

Due to the frequent road network changes, keeping them updated is fundamental for several purposes. Currently, models based on Deep Learning (DL), specifically, Convolutional Neural Networks (CNNs), such as encoder-decoder type, are state-of-the-art for this purpose. In this context, the high performance in CNNs has two aspects involved: the model needs a large labeled dataset, and the dataset belongs to the same probability distribution. In practical applications, however, this may not hold, since there is a domain shift effect, and it is not customary for the availability of labeled data. To approach these challenges, we propose to adapt the U-Net architecture (encoder-decoder) to the Unsupervised Domain Adaptation (UDA) that does not need labeling data to minimize the domain shift effect. Our results demonstrate that the proposed method contributes to road segmentation, whose model reaches 74.31% (IoU) and 85.04% (F1), against the same model without UDA that reaches 67.36% (IoU) and 80.02% (F1). This implies that the information that comes from the target domain, even unsupervised, contributes to adversarial learning, improving the generalization capacity of the model, enhancing aspects such as better discrimination surrounding classes, and in the geometric delineation of the road network.

1. INTRODUCTION

Due to the frequent road network changes, keeping them updated is fundamental for several purposes. The road network is one of the main modes of transport in the world, which provides many types of support, such as expansion or limitation among sites, urban planning, economy (e.g., logistics), autonomous navigation systems, smart cities' development, etc. (Wang et al., 2016). As the large availability of remotely sensed data, a variety of road network detection methods have been proposed since the 70s. Essentially, these methods are based on information about geometrical, radiometrical, topological, functional, scale, and contextual characteristics. Since the methods based on handcrafted features are a time-consuming and costly process to handle with all these characteristics, algorithms based on Deep Learning (DL) exploit them automatically in a unique process.

In this context, Convolutional Neural Networks (CNNs) have been making fundamental advances in Computer Vision over the last decade for the classification, detection, and segmentation of objects in images. In regards to the specific topic of road detection, CNNs, especially, the deeper ones, are the most utilized algorithms for this end (Filho et al., 2023). Algorithms like this are flexible to approach different data types, which increases the detection accuracy based on feature automatic extraction and in different levels of abstraction. Since the pioneering work in 2010 (Mnih and Hinton, 2010), different methods have emerged, such as patch CNNs, deep CNNs, and those based on adversarial training.

With the revolution in the semantic segmentation field by the Fully Convolutional Networks (FCNs) (Long et al., 2015), road detection based on patch CNNs (Li et al., 2016), which have suffered with computational resources with a larger patch, was gradually replaced for the networks with all convolutional layers (Zhong et al., 2016, Henry et al., 2018). Later, the encoder-decoder architectures gained attention, such as the U-Net (Ronneberger et al., 2015). This is due to the skipping connections between the low and high levels of the layers, in contrast to FCNs, which are not effective in preserving with accuracy the spatial details in the reconstruction of the masks.

Approaches based on different characteristics with encoder-decoder nets have been proposed. For example, in the case of multitasking, involving the surface, centerline, and the edges of the roads (Cheng et al., 2017, Lu et al., 2022); residual learning as the increases of the depth of layers (Zhang et al., 2018, Bandara et al., 2022); local and global attention blocks (Xu et al., 2018); multimodal fusion (Filho et al., 2023), adversarial networks (Abdollahi et al., 2021), etc. On the other hand, although DL-based methods are the mainstream approach (i.e., encoder-decoder architectures), the high performance of the existing works has two aspects involved: the model needs a large labeled dataset, and the dataset belongs to the same probability distribution.

The different benchmark sets available in the literature are an example of this assumption, whose approaches use labeled datasets such as the Massachusetts Dataset (Mnih, 2013), DeepGlobe (Demir et al., 2018), CasNet dataset (Cheng et al., 2017), for rural areas (Yang and Wang, 2020), and others. Therefore, once these datasets are split into train and test subsets, the high accuracy of the inference is strongly related because they come from the same probability distribution. In other words, even though they have the same distribution, DL-based models assume that the training data must have enough variability for the generalization at the inference (testing).

In practical applications in the road detection context, however,

the accuracy can decrease and even achieve overfitting when in the presence of a domain shift effect. For instance, a model is trained with a dataset, but it is applied to another one, e.g., Deep Globe and Massachusetts; this characterizes that there are two domains, a source, and a target domain, respectively. Then, the domain gap is due to the different probability distributions that generate the datasets, caused by different spatial resolutions, different locations, acquisition methods, data types, or other factors. Furthermore, once the algorithms work like in fully supervised training, another limitation is the pixel-level labeling in the target domain; in the real world, not always labeled data is available and to do this is a time-consuming process.

Challenges like those contextualize the usage of Domain Adaptation (DA) for road detection. Essentially, DA can minimize the difference in the distribution between the domains, using labeled data just in one domain, to complete the task in the target domain without labeling. As for the specific case of labeled data available just in the source domain, this appeal is called Unsupervised Domain Adaptation (UDA) (Wang and Deng, 2018), a particularity of transfer learning, and the hardest case of DA. UDA methods can be categorized into two categories: representation matching and appearance adaptation networks. While representation matching aims to build a latent space of invariant domain (Wittich and Rottensteiner, 2019), appearance adaptation networks concentrate on transforming images from the target domain similar to the source domain, like generative methods (Wittich and Rottensteiner, 2021). Besides, for both approaches, adversarial training is the most used (Goodfellow et al., 2014).

Despite the recent advances in Remote Sensing and Photogrammetry (RS&P) area, i.e., on the land use and land cover classification by UDA (Xu et al., 2022), for road network detection, which belongs to the semantic segmentation task, several challenges are still suggested, with few works proposed (Iqbal et al., 2023). Specifically, since the UDA's pioneer work (Ganin et al., 2015) in the Computer Vision field, which brought advantages for frameworks of DL and training steps, neither for the RS&P area nor for the road detection topic, this approach has been investigated for deeper CNNs (Elshamli et al., 2017, Soto et al., 2022). Therefore, in this paper, we propose a new approach, in order to contribute to road network detection: we adapt the U-Net architecture to UDA, especially by the strategy of a Domain Adversarial Neural Network (DANN) (Ganin et al., 2015). Thus, the contributions of our study can be summarized in the following: 1) building an encoder-decoder architecture proposed for UDA, whose method and datasets utilized are state-of-the-art; 2) discussing the challenges involved in adversarial training, concerning the optimization problem; and 3) a viable solution to minimize the domain shift problem, without the necessity of labeled data for the inference, which lead to the segmentation of the road network automatically.

2. METHODOLOGY

2.1 Study area and preprocessing stages

A benchmark dataset (Cheng et al., 2017) is used, i.e., the CasNet dataset. The set was originally proposed for road and centerline detection, through CasNet architecture. It contains aerial images collected from Google Earth, whose road segmentation reference maps were manually labeled. The road width is about 12-15 pixels, and the set has some occlusions of cars and trees as well as a highly residential area whose roofs

have similar radiometric characteristics to roads. The images have a Ground Sample Distance (GSD) of 1.2 m, radiometric resolution of 8 bits, and a total number of 224 samples, with dimensions of at least 600x600 pixels.

The CasNet dataset is chosen for some reasons: the delineation of the road network is uniform and homogeneous with little occlusions, the data have a lower GSD and require less storage (little amount), and it also supports comparatives (metrics) among diverse works. For the unsupervised adaptation strategy, the proposed approach considers as source domain RGB images converted into grayscale, and RGB images as the target domain, as shown in Figure 1.

Figure 1. Adapted CasNet dataset used for UDA.

The total amount of 224 samples of the benchmark set is already split into training (160), validation (20), and testing (44). However, as the images are at least in 600x600 px dimensions, train and test images are cropped in 512x512 px patches and, for training, the samples are padded with mirror mode. This increased the number of samples of the set to 252, 20, and 81, for training, validation (the same), and testing, respectively. Moreover, various operations of data augmentation based on geometric operations are applied; for instance, the images are randomly flipped from left to right and then, from up to down; operations like image rotation and image transposing are also made.

2.2 Proposed architecture

The proposed method adapts the U-Net for UDA as the DANN. The U-Net is a network for semantic segmentation. It consists of a contracting path to capture context and a symmetric expanding path that enables precise location (reconstruction of the masks). These paths are named as encoder and decoder, respectively, which yields the u-shaped architecture, and the net does not have fully connected layers. The DANN, a UDA approach, focuses on learning discriminativeness and domain-invariance features. Particularly, DANN can minimize the divergence between two probability distributions (source and target domain) parametrized by an encoder of a deep neural network, which is implemented with other two modules: a decoder (or classifier), and a domain discriminator. The structure of the architecture is presented in detail in Figure 2.

The encoder layers follow the standard U-Net (ImageNet's pretrained weights), and also for the decoder, but with batch normalization layers (Ioffe and Szegedy, 2015) as regularizers. In regards to the domain discriminator, in order to determine from which domain the sample belongs, domain labels are defined as
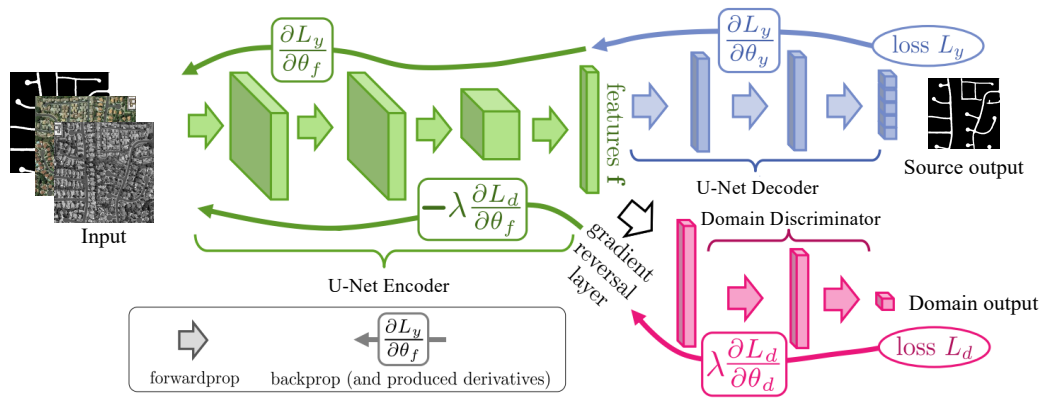
Figure 2. U-Net adapted according to the DANN. Source: Adapted from Ganin et al. (2015).

0 and 1 (for the source and the target domain, respectively), in addition to a Gradient Reversal Layer (GRL), which involves a domain regularizer hyperparameter.

The adapted net has two batches as input: source domain and target domain. They are convoluted separately by the encoder, but concatenated for the domain discriminator, while only the source pass goes to the decoder classification. In regards to the GRL layer, which acts as an identity transformation during the forward pass, and during the backward propagation like a negative scalar (λ), the proposed method implements it in the backward step, according to the gradients.

In the backward pass, gradients of the decoder and the domain discriminator are separately calculated with respect to the same module: the encoder (θ_f weights), as written in Equation 1. The gradients of the decoder loss with respect to its module are calculated and updated with the encoder module (1st step). Subsequently, the batch with source and target domain samples passes to the encoder and the domain discriminator and thus, the discriminator module is updated with another optimizer (2nd step).

$$E(\theta_f, \theta_y, \theta_d) = \mathcal{L}_y(\theta_f, \theta_y) - \lambda \mathcal{L}_d(\theta_f, \theta_d) \quad (1)$$

where $\mathcal{L}_y$ is the loss for the decoder (θ_y weights); and $\mathcal{L}_d$ is the loss for the domain discriminator (θ_d weights).

The whole network was implemented with TensorFlow framework (2.15 version) in the Python programming language (3.10 version). The network was trained on the Google Colab Pro environment, which utilizes the NVIDIA Tesla T4 16GB graphic card.

2.3 Evaluation of the proposed approach

The three modules of the architecture are used in training, but only the encoder and the decoder are used for the testing step. Based on this, the results can be evaluated. For both source and target domains, the data are split into training, validation, and testing. However, for a fair comparison, and in order to check the contribution of UDA for road detection, the results of the proposed method are compared from a baseline, which adopts the same segmentation model (U-Net) and settings but without UDA, and also, with a fully supervised training (on the target domain).

The evaluation consists of qualitative and quantitative analysis. The metrics Recall (Equation 2), Precision (Equation 3), F1-

Score (Equation 4), and Intersection over Union (IoU) (Equation 5) are used for this end, as the Equations are written below:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{F1} = \frac{TP}{TP + \frac{FP+FN}{2}} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

$$\text{IoU} = \frac{TP}{TP + FP + FN} \quad (5)$$

Where False Negatives (FN) are the number of pixels of "road class" but classified as "non-road class"; False Positives (FP) are the number of pixels of "non-road class" but classified as "road class"; True Negatives (TN) are the number of pixels correctly classified in "non-road class"; and True Positives (TP) are the number of pixels correctly classified in "road class".

Due to the classes' unbalancing of the roads, the Focal Tversky Loss (Abraham and Khan, 2018), written in Equation 6, is used for the decoder (prediction of the roads), and for the domain discriminator, two loss functions are analyzed: Binary Cross-Entropy (BCE), and the sigmoid with cross-entropy (S-CE).

$$L_{FTL}(\alpha, \beta) = 1 - \left(\frac{TP}{TP + \alpha * FP + \beta * FN} \right)^{\frac{1}{\gamma}} \quad (6)$$

where γ can range from [1,3].

3. RESULTS

For the selected domains from the CasNet dataset, it is considered that the adaptation setting includes different data types (i.e., grayscale and RGB images), different backgrounds, and types of road surfaces. For this, the batch size of training was fixed into 4 and data augmentation operations were made, such as images randomly flipped from left to right and up to down, image transpose, and rotation of 90° and 270°. An extensive inspection of hyperparameters was also made, which involved, mainly, finding an adequate learning rate, which optimizer to use, the total number of iterations (according to the batch size, epochs, and number of samples), and the λ value (domain regularizer hyperparameter).

	Grayscale → RGB					
Model name	Lr_e/Opt.	Lr_d/Opt.	$Loss_d$	It.	F1 (%)	IoU (%)
1a	SGD (lr=1e-4)*	SGD (lr=1e-4)	BCE	3780	77.27	63.48
1b	SGD (lr=1e-2 up to 1e-3)	SGD (lr=1e-3 up to 1e-4)	BCE	6300	81.73	69.65
1c	SGD (lr=1e-2)*	SGD (lr=1e-4)	S-CE	5040	**82.01**	**69.95**
1d	SGD (lr=1e-2 up to 1e-3)	SGD (lr=1e-4 up to 1e-5)	BCE	4410	81.06	68.72
2a	SGD (lr=1e-2 up to 1e-3)	Adam (lr=1e-4)	S-CE	4410	81.56	69.36
2b	SGD (lr=1e-2)*	Adam (lr=1e-4)	BCE	4410	80.99	68.67

Lr: learning rate; Opt.: optimizer; It.: iterations. *Exponential decay fixed in 1000 steps and decay rate of 0.96;

Table 1. Preliminary results of hyperparameters in the subtest set.

	Grayscale → RGB						
Model name	Lr_e*/Opt.	Lr_d/Opt.	$Loss_d$	λ**	Iterations	F1 (%)	IoU (%)
3a	SGD (lr=1e-2)	Adam (lr=1e-4)	BCE	[0.0001; 0.01]	5670	81.56	69.34
3b	SGD (lr=1e-2)	-	-	-	5670	81.15	68.63
3c	SGD (lr=1e-2)	SGD (lr=1e-4)	BCE	[0.0005; 0.01]	5670	79.84	67.13
3d	SGD (lr=1e-2)	SGD (lr=1e-4)	BCE	[0.0005; 0.1]	5670	82.68	71.02
3e	SGD (lr=1e-2)	SGD (lr=1e-4)	BCE	[0.0005; 0.0005]	6930	**85.04**	**74.31**
3f	SGD (lr=1e-2)	SGD (lr=1e-4)	BCE	[0.005; 0.005]	6930	84.92	74.14
3g	SGD (lr=1e-2)	SGD (lr=1e-4)	BCE	[0.001; 0.001]	6930	84.86	74.05

*Learning rate schedule: exponential decay from 1e-2 up to 1e-3; **The scalar concerns the encoder and the discriminator module, respectively.

Table 2. Preliminary results based on and domain regularization hyperparameter.
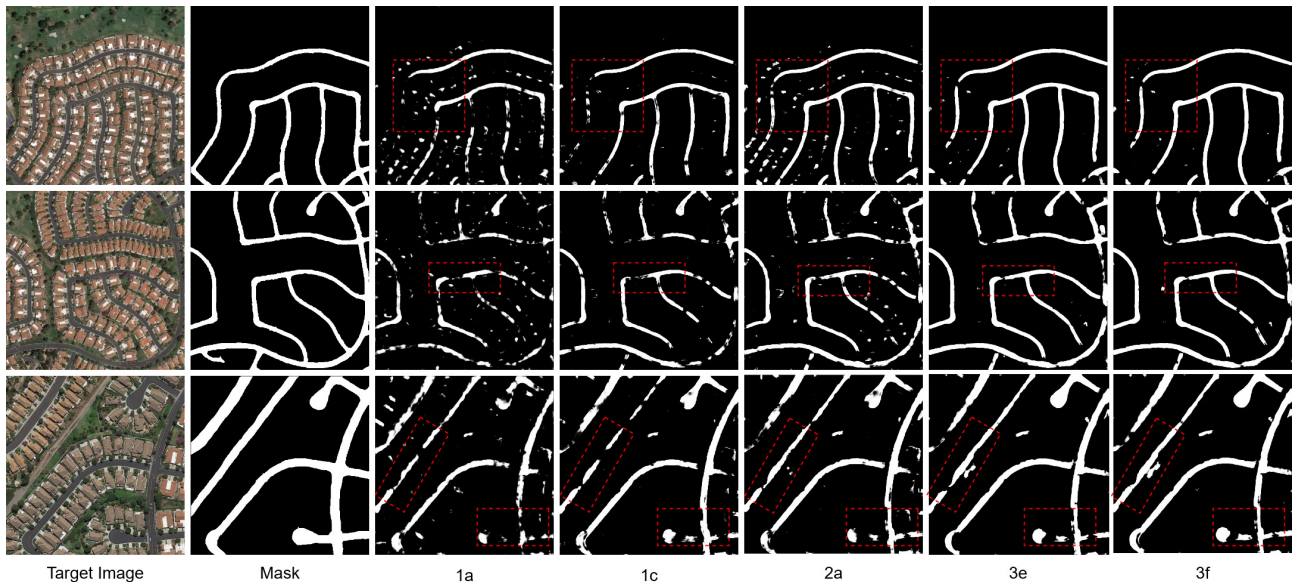

Figure 3. Results for the preliminary analysis. Columns 1a-2c: models from Table 1; Columns 3e-3f: models from Table 2.

Firstly, two optimizers are tested: Adam and Stochastic Gradient Descent (SGD) with exponential decay. For the domain discriminator, two loss functions are analyzed: the Binary Cross-Entropy, and the sigmoid with cross entropy (S-CE). Table 1 shows the results in terms of metrics for this preliminary analysis, where the subscript "e" and "d" refer to encoder-decoder, and discriminator modules, respectively.

Different settings for the model are presented in Table 1. The SGD optimizer with momentum (0.9) and exponential decay, which was chosen to update the encoder-decoder weights, emphasizes with a lower learning rate, the model reaches a plateau and does not converge (model 1a in Table 1). Then, a higher learning rate should be used, such as 10^{-2} with decay. Regarding the discriminator influence, Table 1 shows that an adaptive optimizer (Adam) or the SGD with or without scheduling for both loss functions, the prediction in the subset test in F1 and IoU values are around 82%, and 69%, respectively; but in visual analysis, more iterations can reduce the FPs in the prediction (model 1c and 2a). Figure 3 presents a comparison between them.

As presented in Figure 3, it is evident that more iterations are necessary for the experiment. For instance, the predictions have some shortcomings, like noises on the tiles, buildings are predicted as roads (similar in radiometric terms), and the roads are disconnected due to occlusion by the trees or cars. On the other hand, the failures might be related to the optimization problem; in adversarial training, it is essential to balance the weight between the label predictor and discriminator losses. In practice, it means that the label predictor should not be too good to make it difficult to predict the domains ("weaker" discriminator), and vice versa. In that regard, UDA architecture has a domain regularization hyperparameter, i.e., λ, which correctly needs to propagate the gradients to the network. Thus, different values were analyzed, for both the label predictor and the discriminator. Table 2 shows these analyses.

The results presented in the first preliminary analysis (Table 1) use a fixed λ = 1. However, according to Table 2, the regularizer with small values can better match distributions between source and target domains. Empirically, different values were tested, since there is a regularizer for the encoder-

Grayscale → RGB				
Model name	Precision (%)	Recall (%)	F1 (%)	IoU (%)
On Target (Cheng et al., 2017)*	83.26	97.01	89.56	81.86
On Target	88.89	86.68	86.98	77.52
Baseline	<u>83.92</u>	<u>78.18</u>	<u>80.02</u>	<u>67.36</u>
Proposed approach	**88.83**	**82.83**	**85.04**	**74.31**

*U-Net modified.

Table 3. Analysis of the proposed method and comparisons to baseline and training on target..
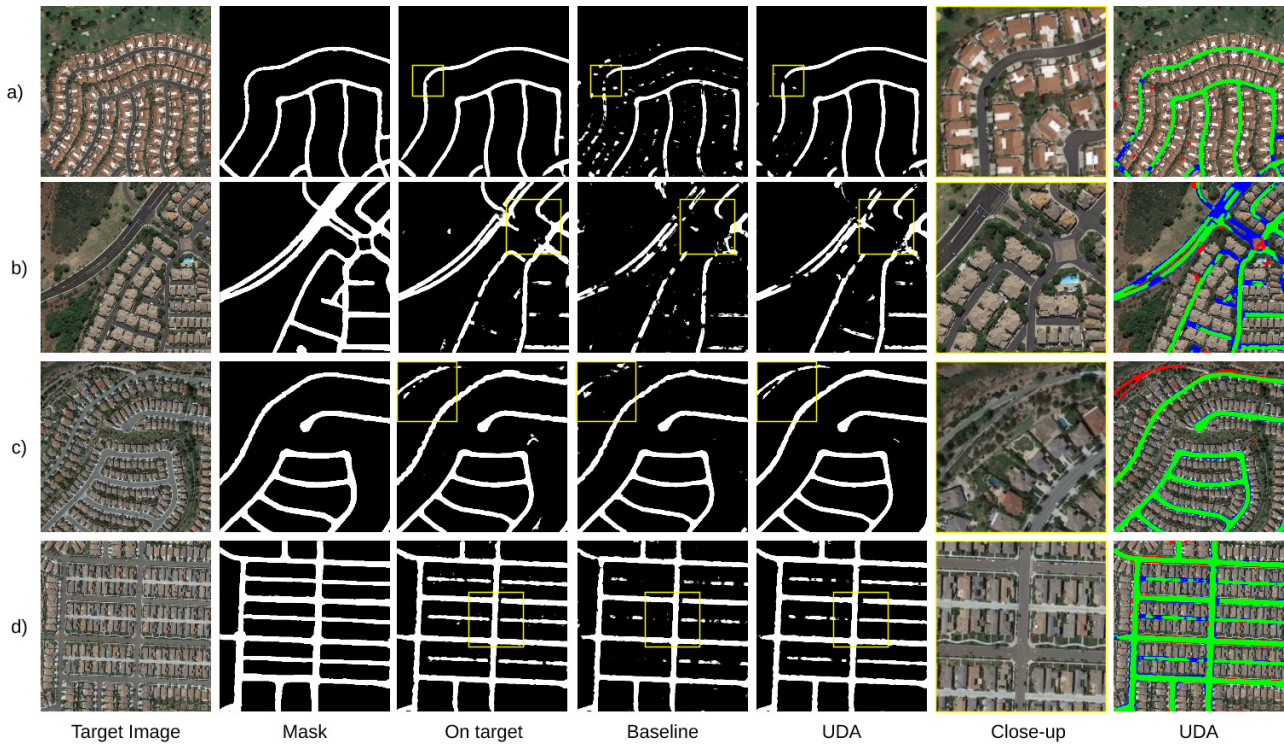

Figure 4. Visual comparisons of road detection results: on target, baseline, and UDA. The 6th column: close-ups images of the yellow boxes. The last column: FP - red color; FN - blue color; TP - green color.

decoder model and another for the discriminator module. Using different values for each one, in this case, may decrease the metric of the model. In contrast, as in model 3b, if the discriminator is not updated, F1 and IoU are affected. When using a small and the same value, the metrics increase, and more iterations are necessary, achieving approximately, 85% and 74% for F1 and IoU, respectively (e.g., model 3e). For qualitative analyses, Figure 3 (models 3e and 3f) shows the predictions for this second preliminary analysis.

The last two results presented in Figure 3 emphasize that a larger λ may be more sensitive to the gradients and even propagate them incorrectly. This means that a small and equal value is more suitable for the road detection problem (grayscale to RGB case). For example, when compared to the first analysis, now the model distinguishes the classes from the roads with more accuracy, without the noises presented previously, reducing the FPs on the segmentation, and better geometric delineation of the road network.

After these analyses, it is possible to compare the results to a baseline, in this case, the same model: U-Net. Also, the training and prediction in a fully supervised way is presented, i.e., with the RGB images, intending to check the contribution of UDA to this end. Table 3 shows the comparison.

As shown in Table 3, the proposed method performs favorably against the baseline, achieving 85.04% and 74.31% for F1-Score and IOU, respectively; the difference is around 6% and 10%, respectively. In terms of training (optimization), it emphasizes the importance of the hyperparameters calibration, and also, regularizers (e.g., λ), with a focus on adversarial training. Regarding the fully supervised training on the target domain, the UDA strategy still needs to be improved, whose performance achieves 89.56% and 81.86% (Cheng et al., 2017), and 86.98.56% and 77.52% (the same settings), for F1-Score and IoU, respectively. Figure 4 presents the comparing results in visual performance.

In visual performance (Figure 4), it is noted that the UDA strategy can significantly contribute to road segmentation. For instance, in the source domain with only information about grayscale (spectral information), the roads are easily confused with roofs of houses, while with the information from the target domain (RGB), the network can discriminate between these classes and improve their connectivity (topology). However, in the presence of trees or cars over the roads, where the network misses information, the probability of FN increases. Furthermore, the model is still sensitive to detecting roads and intersections, which means that does not have enough variability from grayscale and RGB domains. Different objects but radiometrically similar, like crosswalks, central trees, and different pave-

ments of roads are some examples of the main shortcomings of the proposed method.

Other tests are necessary for road segmentation, including more iterations for training. Despite the model still having some limitations in detecting the road network, it should be emphasized that the method improves the segmentation since training like transfer learning is not necessary, and some problems remain even in the fully supervised training (on target). In this sense, investigating other loss functions, another backbone, and other hyperparameter settings is some suggestion to overcome the current models' limitations and enhance the road segmentation of the proposed method.

4. CONCLUSION AND FUTURE WORK

This work approached an experiment in order to check the contribution of UDA in the road segmentation task, based on the strategy that available labeled road data can achieve a promising road network detection on unlabeled target images.

For this purpose, the achieved metrics have shown adequate results. The contribution of UDA is emphasized since the proposed method performs well against a baseline (the approach without UDA). The information that comes from the target domain, even unsupervised, contributes to adversarial learning, improving the generalization capacity of the model. Aspects such as better discrimination surrounding classes (contextual aspect) refine the geometric delineation of the road network and avoid the probability of false positives and false negatives. Despite some improvements that need to be made, such as in cases in which the roads are obstructed, or due to the different pavements, it is worth noticing that the proposed approach leads to an automatic road segmentation without labeled data in the target domain, once no further training or fine-tuning is required.

To state these perspectives, other approaches are proposed to check the contribution of UDA in a visual analysis, based on techniques for representation of data in the DL context, such as the t-SNE. Besides, since road detection comprehends multiscale information, another backbone, different hyperparameter settings, and even methods for multiscale and multilevel features extraction (attention) are some strategies to be investigated.

5. ACKNOWLEDGEMENTS

References

Abdollahi, A., Pradhan, B., Sharma, G., Maulud, K. N. A., Alamri, A., 2021. Improving Road Semantic Segmentation Using Generative Adversarial Network. *IEEE Access*, 9, 64381–64392.

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.

Bandara, W. G. C., Valanarasu, J. M. J., Patel, V. M., 2022. Spin road mapper: Extracting roads from aerial images via spatial and interaction space graph reasoning for autonomous driving. *2022 International Conference on Robotics and Automation (ICRA)*, IEEE, 343–350.

Cheng, G., Wang, Y., Xu, S., Wang, H., Xiang, S., Pan, C., 2017. Automatic Road Detection and Centerline Extraction via Cascaded End-to-End Convolutional Neural Network. *IEEE Transactions on Geoscience and Remote Sensing*, 55(6), 3322–3337.

Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Basu, S., Hughes, F., Tuia, D., Raskar, R., 2018. Deepglobe 2018: A challenge to parse the earth through satellite images. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 172–181.

Elshamli, A., Taylor, G. W., Berg, A., Areibi, S., 2017. Domain Adaptation Using Representation Learning for the Classification of Remote Sensing Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(9), 4198–4209.

Filho, A., Shimabukuro, M., Dal Poz, A. P., 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., 2015. Domain-Adversarial Training of Neural Networks. *Advances in Computer Vision and Pattern Recognition*, 17(9783319583464), 189–209. https://arxiv.org/abs/1505.07818v4.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative Adversarial Networks. *Communications of the ACM*, 63(11), 139–144. https://arxiv.org/abs/1406.2661v1.

Henry, C., Azimi, S. M., Merkle, N., 2018. Road Segmentation in SAR Satellite Images With Deep Fully Convolutional Neural Networks. *IEEE Geoscience and Remote Sensing Letters*, 15(12), 1867–1871. http://dx.doi.org/10.1109/LGRS.2018.2864342.

Ioffe, S., Szegedy, C., 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *32nd International Conference on Machine Learning, ICML 2015*, 1, 448–456. https://arxiv.org/abs/1502.03167v3.

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.

Li, P., Zang, Y., Wang, C., Li, J., Cheng, M., Luo, L., Yu, Y., 2016. Road network extraction via deep learning and line integral convolution. *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, IEEE, 1599–1602.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 07-12-June, 431–440.

Lu, X., Zhong, Y., Zheng, Z., Chen, D., Su, Y., Ma, A., Zhang, L., 2022. Cascaded Multi-Task Road Extraction Network for Road Surface, Centerline, and Edge Extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–14. https://ieeexplore.ieee.org/document/9756441/.

Mnih, V., 2013. Machine learning for aerial image labeling.

Mnih, V., Hinton, G. E., 2010. Learning to detect roads in high-resolution aerial images. *Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part VI 11*, Springer, 210–223.

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.

Soto, P. J., Costa, G. A., Feitosa, R. Q., Ortega, M. X., Bermudez, J. D., Turnes, J. N., 2022. Domain-Adversarial Neural Networks for Deforestation Detection in Tropical Forests. *IEEE Geoscience and Remote Sensing Letters*, 19.

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

Wang, W., Yang, N., Zhang, Y., Wang, F., Cao, T., Eklund, P., 2016. A review of road extraction from remote sensing images. *Journal of Traffic and Transportation Engineering (English Edition)*, 3(3), 271–282.

Wittich, D., Rottensteiner, F., 2019. ADVERSARIAL DOMAIN ADAPTATION FOR THE CLASSIFICATION OF AERIAL IMAGES AND HEIGHT DATA USING CONVOLUTIONAL NEURAL NETWORKS. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2-W7(2/W7), 197–204. https://doi.org/10.5194/isprs-annals-IV-2-W7-197-2019.

Wittich, D., Rottensteiner, F., 2021. Appearance based deep domain adaptation for the classification of aerial images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 180, 82–102.

Xu, M., Wu, M., Chen, K., Zhang, C., Guo, J., 2022. The Eyes of the Gods: A Survey of Unsupervised Domain Adaptation Methods Based on Remote Sensing Data. *Remote Sensing 2022, Vol. 14, Page 4380*, 14(17), 4380. https://www.mdpi.com/2072-4292/14/17/4380/htm https://www.mdpi.com/2072-4292/14/17/4380.

Xu, Y., Xie, Z., Feng, Y., Chen, Z., 2018. Road Extraction from High-Resolution Remote Sensing Imagery Using Deep Learning. *Remote Sensing 2018, Vol. 10, Page 1461*, 10(9), 1461. https://www.mdpi.com/2072-4292/10/9/1461/htm https://www.mdpi.com/2072-4292/10/9/1461.

Yang, C., Wang, Z., 2020. An ensemble Wasserstein generative adversarial network method for road extraction from high resolution remote sensing images in rural areas. *IEEE Access*, 8, 174317–174324.

Zhang, Z., Liu, Q., Wang, Y., 2018. Road Extraction by Deep Residual U-Net. *IEEE Geoscience and Remote Sensing Letters*, 15(5), 749–753.

Zhong, Z., Li, J., Cui, W., Jiang, H., 2016. Fully convolutional networks for building and road extraction: Preliminary results. *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, IEEE, 1591–1594.