Paving the Way for Automatic Mapping of Rural Roads in the Amazon Rainforest

Lucas Costa de Faria*, Matheus Brito*, Keiller Nogueira†, and Jefersson A. dos Santos*,‡
*Department of Computer Science, Universidade Federal de Minas Gerais, Brazil
†Computing Science and Mathematics, University of Stirling, United Kingdom
‡Department of Computer Science, University of Sheffield, United Kingdom

Email: {lucascostafaria, matheus, jefersson}@dcc.ufmg.br, keiller.nogueira@stir.ac.uk

***Abstract*—Amazon rainforest deforestation severely impacts the environment in many ways, including biodiversity reduction, climate change, and so on. A key indicator of deforestation is the sudden appearance of rural/unofficial roads, usually exploited to transport raw materials extracted from the forest. To early detect such roads and prevent deforestation, remote sensing images have been widely employed. Precisely, some researchers have focused on tackling this task by using low-resolution imagery, mainly due to their public availability and long time series. However, performing road extracting using low-resolution images poses several challenges, most of which are not addressed by existing works, including high inter-class similarity, complex structure, etc. Motivated by this, in this paper, we propose a novel approach to perform road extraction on low-resolution satellite images based on contextual and pixel-level decision fusion. We conducted a systematic evaluation of the proposed method using a new dataset proposed in this work. The experiments show that the proposed method outperforms state-of-the-art algorithms in terms of intersection over union and F1-score.**

I. Introduction

Road extraction (or mapping) aims at detecting and segmenting roads, usually by means of remote sensing images, mainly due to their wide coverage. As can be seen in Figure 1, such a task (and its developments and outcomes) can be directly linked to deforestation in the Amazon rainforest [1] which, in turn, is related to several environmental problems, such as biodiversity reduction and climate change. An important issue is that this type of road (presented in Figure 1) is usually opened illegally, without the knowledge or authorization of the responsible bodies.

Manual road extraction from images is a laborious and time-consuming process, particularly when considering the extension and sheer quantity of roads. Due to this, it is of paramount importance to develop automatic approaches capable of promptly identifying/extracting especially rural/unofficial roads in the Amazon forest ,and, consequently, predicting possible deforestation zones, thus assisting in the prevention. Towards this, several works have been proposed for automatic road mapping [2]–[6]. Most of these studies use high-resolution aerial images, mainly because of their fine level of detail. However, despite the advantages, such images are generally not publicly available and, since these sensors are relatively new technologies, they may not provide information over a long time series, crucial properties for understanding patterns and temporal dynamics of deforestation. Motivated by this, some researchers [4], [7] have focused on performing road mapping using low-resolution imagery, such as those from the Landsat and Sentinel constellations. In spite of the benefits, performing road extracting using low-resolution images poses several challenges, most of which are not addressed by existing works, including: (i) complex structure, given that these roads are very winding and close to each other, and (ii) high inter-class similarity, given that the roads can resemble other objects such as river banks.

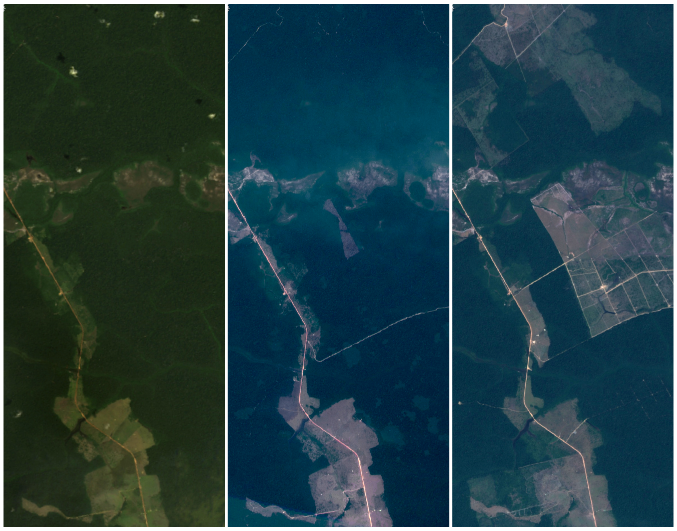

Fig. 1: Transamazônica highway near Jatuarana River in Amazon Rainforest. Note that there is no deforestation in the area in 2014. It is possible to observe the opening of rural roads in 2016. Many areas close to the roads have been deforested as observed in 2022.

In this paper, we propose a novel approach to perform road extraction using low-resolution satellite images that consists of a Context-to-Pixel fusion based on the results from a contextual module (classification backbone) and a pixel-wise module (segmentation backbone). Additionally, we also propose a new dataset for road extraction in the Amazon rainforest.

In practice, these are the contributions of this work:

- a novel method to identify roads in low-resolution remote sensing images (i.e., 10–20m pixel size) based on contextual and pixel-level decision fusion[1],
- a new dataset for road extraction in Amazon rainforest[2].

II. Related Work

Due to its importance, several approaches have been proposed for road mapping using remote sensing images [4], [8]–[11]. Most methods tackling road mapping use high-resolution remote sensing images [2], [3], [8]–[10]. Mosinska *et al.* [8] trained a U-net to perform road extraction using a new loss function that considers the street topology. They also employed an iterative procedure to recursively refine the predictions. Zhou *et al.* [9] proposed a new Convolutional Network, called D-LinkNet, capable of extracting roads in high-resolution images in a time-efficient manner. Such a network has additional dilated layers to help add more context [12]. In [10], the authors introduced a new Recurrent U-net to efficiently and effectively detect roads and centerlines. In [13], the authors proposed a novel encoder-decoder architecture. This network is then trained using a combination of the cross-entropy and dice losses in order to learn both local and global information. Wan *et al.* [2] designed a new dual-attention network that employs an attention mechanism on low and high-level features to aggregate more information and reduce discontinuous and incomplete results. Yang *et al.* [3] introduced a new network that combines an encoder-decoder convolutional architecture with Swin transformers [14] in order to aggregate more contextual information and reduce discontinuities.

As introduced, despite the advantages of high-resolution data, such images are generally not publicly available and may not provide information over a long time series. Therefore, due to its wide availability (mainly temporal), some studies have used low-resolution data to perform road extraction [4]–[6], [15]. In [15], the authors proposed a new U-net architecture that gradually incorporates all available Sentinel image bands into the network to generate better-quality results. Ayala *et al.* [6] combined super-resolution techniques with segmentation networks in order to be able to receive low-resolution remote sensing images but generate high-resolution road maps. Dixit *et al.* [5], the authors proposed a new architecture that combines Residual Networks, U-Net, and dilated convolutions to perform low-resolution road segmentation. Botelho *et al.* [4] trained a modified U-net architecture to detect rural roads in the Brazilian Amazon. They also employed a post-processing technique to refine and generate the final segmentation map.

In this work, we perform road extraction by combining contextual and pixel-level information. Different from previous approaches, the proposed method leverages contextual information to effectively identify regions likely to contain roads, thus reducing the number of false positives.

III. The AmazonWildRoads (AWR) Dataset

In this work, we proposed a novel dataset, called AmazonWildRoads, that focuses on (rural) roads in the Amazon rainforest. Precisely, this dataset is composed of 28 Sentinel 2 satellite images, obtained from the Google Earth Engine, covering various areas of the Amazon rainforest throughout several Brazilian states, as presented in Figure 2. In order to enhance the diversity of the dataset, distinct types of areas were selected and collected, including some with rivers, cities, vegetation, and so on. This process ensures that the dataset encompasses different environmental and urban elements, increasing its difficulty, but making it more appropriate to evaluate the task at hand.

In the Amazon region, there is a high incidence of clouds throughout the year. When acquiring the images, we selected all images from the year 2020, filtered the images that have cloud cover above 30% of the pixels of the total image area, and then applied the median to the set of all remaining images with Google Earth Engine tools [16].

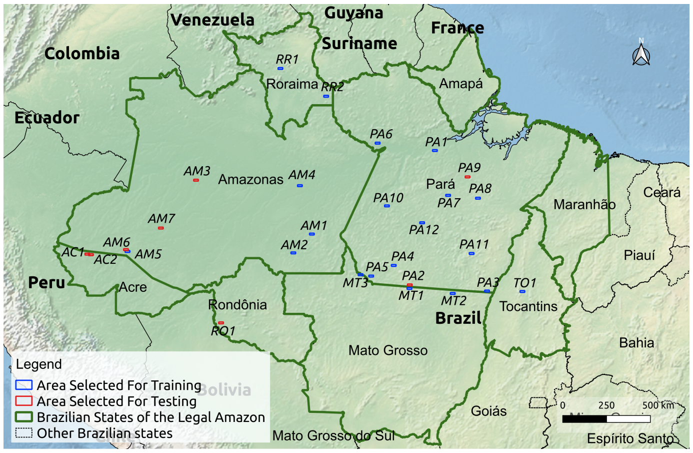

Fig. 2: Map of collected areas around the Amazon rainforest. In green, we have highlighted the states belonging to the Brazilian Legal Amazon. In red and blue, we can see all collected 28 areas.

The images in this dataset are composed of RGB bands, thus having a Ground Sampling Distance (GDS) of 10 meters per pixel. Furthermore, the images have an average resolution of $2,794 \times 1,059$ pixels, providing area coverage of approximately 290 km^2 each. Overall, the dataset covers an extensive area of approximately $8,120$ km^2 in total. Finally, the roads in these images were manually annotated by specialists.

IV. Proposed Approach

In this section, we describe the proposed approach to road extraction that combines contextual and pixel-level information, as presented in Figure 3. Section IV-A discusses the Contextual Road Indicator module, whereas the Pixel-wise Road Extraction module is presented in Section IV-B. Finally, Section IV-C introduces the Context-to-Pixel fusion strategy.

[1] https://github.com/lucascsfaria/ContextualPixelLevelRoadExtractionAmazon
[2] https://github.com/lucascsfaria/amazonwildroadsdataset

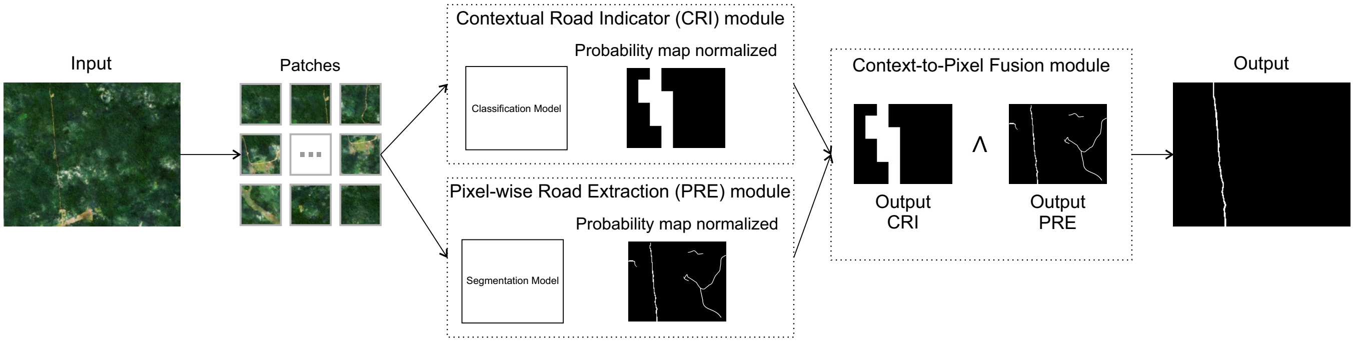
Input
Patches
Contextual Road Indicator (CRI) module
Probability map normalized
Classification Model
Pixel-wise Road Extraction (PRE) module
Probability map normalized
Segmentation Model
Context-to-Pixel Fusion module
Output CRI
∧
Output PRE
Output

Fig. 3: Overview of the proposed approach for extracting roads from low-resolution remote sensing images.

A. Contextual Road Indicator (CRI) module

The Amazon rainforest region has many areas of difficult access with no roads that may contain objects that, because of the context, could cause confusion in road extraction models, resulting in incorrect predictions (i.e., false positives). Due to this, it is vital that road extraction methods use strategies capable of discarding these regions in order to reduce false positives and improve general performance. Towards this, we propose a module, called Contextual Road Indicator (CRI), that exploits the contextual information to identify regions that are likely to contain roads. Such a module is based on neural networks, which are trained using overlapping patches/crops, generated from the original images. These patches are assigned one of two classes, road or non-road, depending on whether they have at least one pixel of road. By analyzing the context and extracting key features, the model learns to classify such patches into one of these two classes, enabling effective identification and differentiation of road and non-road areas.

During the inference, all patches of an image are classified by the trained model, which outputs a probability for each crop. These predictions (and corresponding probabilities) are then used to create a probability map for this entire image. To accomplish this, we: (i) consider that the likelihood predicted for an area is, in fact, associated with all its pixels, and (ii) add up the probabilities of overlapping regions. As a result, certain areas may have a probability greater than 1. To address this issue, the non-normalized probability map is further processed using the following function: $P(i,j) = \frac{X(i,j)}{max(X)}$ $\forall i, j \in X$, where X and $max(X)$ are the input non-normalized image and its maximum value, and P is the final normalized probability map. After this procedure, the context indicator image can finally be generated by binarizing the normalized probability map using a threshold of 0.5 for each pixel. This results in a map that highlights regions within the original image that are deemed highly likely to contain roads.

B. Pixel-wise Road Extraction (PRE) module

Besides the CRI, we also propose another module to generate a pixel-level segmentation of the input images. This module, called Pixel-wise Road Extraction (PRE), works similarly to the Contextual Road Indicator one, but instead of generating patch-level predictions, it generates pixel-level outcomes by using semantic segmentation networks.

Specifically, this module's networks are trained to segment input patches, i.e., to produce a dense prediction in which each pixel is associated with a probability of being a road. During the inference phase, the input patches are processed by the trained model, which generates dense predictions that are subsequently utilized to create a probability map for the entire image, following a procedure similar to that employed in the CRI module. Precisely, the only difference in this part between these modules is the normalization, which is carried out in the current component using the following function: $P(i,j) = \frac{X(i,j) - min(X)}{max(X) - min(X)}$ $\forall i, j \in X$, where P is the final normalized probability map, and X, $min(X)$, and $max(X)$ are the input non-normalized image, its minimum, and maximum values, respectively. After this, the final prediction map is generated by binarizing the normalized probability map using a threshold of 0.2 for each pixel that was defined experimentally.

C. Context-to-Pixel Fusion

As introduced, most of the area of the Amazon rainforest has no roads. However, because of similarity and other contextual characteristics, extraction methods tend to erroneously predict roads in such regions. Motivated by this, we propose a new module, called Context-to-Pixel Fusion, that merges the outcomes generated by the CRI and PRE components in order to better identify regions likely to have roads, thus reducing false positives and improving general performance.

Technically, the context indicator image generated by the CRI module and the final prediction map created by the PRE component are combined using an AND operator. By performing this combination, the method is capable of exploiting the contextual information extracted by the CRI module and the detailed pixel-level features captured by the PRE component to effectively identify regions likely to contain roads, thus reducing the number of false positives.

V. Experimental Setup

Experimental Protocol. Our approach was trained and evaluated using the AmazonWildRoads dataset, presented in Section III. We divided this dataset into training and testing sets, as depicted in Figure 2, by using 20 images for training and 8 for testing. It is important to note that we selected samples with very different properties for the test set, varying considerably in relation to road density (some do not even have a road), the presence of similar objects (such as rivers), etc. The main idea was to create a test set that would allow us to better understand the behavior of the methods in the most diverse scenarios found in the Amazon forest region. Results are presented using four distinct and complementary metrics, i.e., Intersection over Union (IoU), Precision, Recall, and F1 score.

Implementation. For the CRI module, we evaluated 4 classification networks: Resnet18 [17], VisionTransformer [18], ConvNext [19], and SwinTransformer [14]. All models were trained using the Binary Cross Entropy loss and the following hyper-parameters: patch size of 128 × 128 pixels (with an overlap of 64 pixels), 50 epochs, batch size equal to 64, SGD as optimizer, learning rate of 0.0001, and momentum of 0.9. For the PRE module, 3 different architectures were tested: U-net [20], U-net++ [21], and DeepLabv3+ [22]. For each one of these, 3 distinct encoders, pre-trained on the ImageNet dataset, were assessed: EficcientNet-b0 and EficcientNet-b7 [23], and ResNext101 [24]. All models were trained using a combination of the Jaccard and Focal losses [25], patch size of 128 × 128 pixels (with overlap of 64 pixels), 50 epochs, batch size equal to 64, Adam optimizer, and learning rate of 0.0003.

Baselines. We compared the proposed method with two state-of-the-art baselines: (i) D-LinkNet [9], a Convolutional Network that exploits dilated layers to help add more context to the learning process, and (ii) U-net road model [4], which combines a larger U-net architecture with post-processing techniques to refine and generate the final segmentation map.

VI. Results and Discussions

In Section VI-A, we evaluate the proposed method using different configurations and backbones. In Section VI-B we compare our results with state-of-the-art approaches.

A. Proposed approach evaluation

1) Architecture Analysis: Table I shows the results obtained for the proposed approach with different configurations for the pixel-wise and the contextual modules. We have analyzed and tested 36 configurations – **4** different backbones for the CRI module and **9** distinct architectures (3 networks times 3 backbones) for the PRE module. For comparison purposes, we also report the results of the PRE module (i.e., segmentation networks), but without any contextual information (i.e., without the CRI module). Note also that, for simplicity and clarity, only the most relevant results in terms of the F1-score metric were reported in this paper. Furthermore, since all the best results were produced using the EficcientNet-b7 [23] backbone for the PRE module, only these have been reported.

A first observation that we can elucidate from these results is that the Context-to-Pixel fusion we propose in this paper tends to bring effective gains, which can be observed when comparing the values of IoU, Precision, and F1-Score of the models exploiting contextual information with those without this component. However, this leads to an increase in the false negative rates, which is evidenced by the overall drop in Recall values. In summary, the use of context helps to remove incorrectly identified disconnected sections (false positives) from the main roads present in the images but, in contrast, small sections correctly detected are also lost.

TABLE I: Results for the proposed approach with different backbone combinations for the pixelwise and the contextual modules.

Pixelwise	Contextual	IoU	Precision	Recall	F1-Score
U-net++ [21]	-	0.539	0.624	0.799	0.701
	ConvNext	0.559	0.656	0.790	0.717
	ResNet	0.533	0.638	0.763	0.695
	SwinT	0.549	0.668	0.755	0.709
	ViT	**0.571**	**0.690**	0.768	**0.727**
U-net [20]	-	0.483	0.543	0.814	0.652
	ConvNext	0.522	0.596	0.808	0.686
	ResNet	0.489	0.569	0.776	0.657
	SwinT	0.517	0.613	0.769	0.682
	ViT	0.547	0.645	0.782	0.707
DeepLabv3+ [22]	-	0.483	0.533	**0.836**	0.651
	ConvNext	0.513	0.574	0.827	0.678
	ResNet	0.483	0.550	0.799	0.652
	SwinT	0.507	0.588	0.787	0.673
	ViT	0.528	0.608	0.802	0.692

Considering the performance of the evaluated architectures for the PRE module, U-net++ [21] presented leading outcomes among all combinations, with the best results being achieved with the Vision Transformers (ViT) backbone for contextual analysis, yielded 0.727 in terms of F1-Score. Note that Recall loss is also more pronounced. The best result in terms of Recall (0.836) was generated by the DeepLabV3+ architecture [22] without any contextual information. Such a model tends to identify a higher number of roads but, at the same time, generates a significant amount of noise or false positives (as can be seen by the Precision).

2) Qualitative Analysis: In Figure 4, we show visual results for three areas. Note that these results were generated using the best obtained model, i.e., a U-net++ [21] with EfficientNet-b7 [23] in the PRE module and ViT [18] in the CRI module.

The first image (PA2 – please, refer to Figure 2), has several roads and is quite degraded by deforestation. Few areas are discarded by our contextual fusion. At the same time, small road segments are improperly removed (false negatives). The second image (AM6) has some roads and a lot of degradation on the east half, but on the other side (west) just a few rivers. It is worth remembering that flooded areas and narrow rivers are easily confused with stretches of rural roads [1], [4]. In this case, the pixel-wise approach detected several segments of rivers and sandbanks as rural roads (false positive). Context played a key role in removing many of these segments in the final result. Finally, the third image (AC2) has a large east-west road and another, less obvious road, that follows part of a river running north-south. The rest of the image is of preserved forest. In this case, the proposed fusion is also quite effective, as it is able to remove several segments incorrectly detected as rural roads along the river.

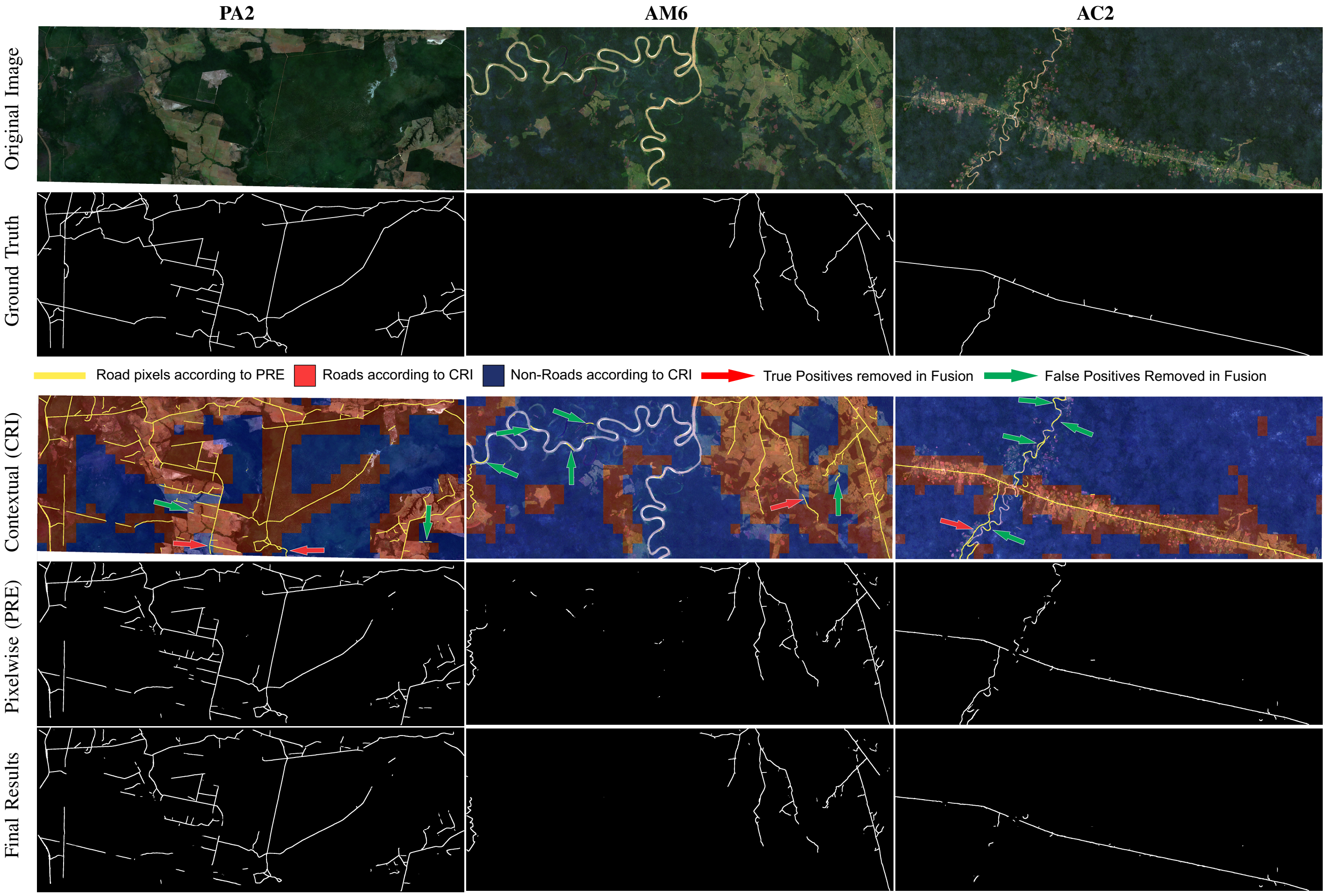
PA2
AM6
AC2
Original Image
Ground Truth
Road pixels according to PRE
Roads according to CRI
Non-Roads according to CRI
True Positives removed in Fusion
False Positives Removed in Fusion
Contextual (CRI)
Pixelwise (PRE)
Final Results

Fig. 4: Results of the proposed approach for distinct areas of the AmazonWildRoads dataset. *For the ground-truth, segmentation and final results*: white areas represent roads, while the black regions are non-road.

B. Comparison with baselines

In Table II we present results obtained by the proposed approach in comparison with state-of-the-art methods for road detection. For the D-LinkNet [9], we have tested ResNet-101 and ResNet-34 (original) and reported the best results. For the U-net road model [4], we used the same architecture presented in the original work. Considering the F1-score, IoU, and Precision values, our method achieved better results than the baselines. The proposed method achieved a slightly higher rate of false negatives compared to the U-net road model [4], as can be seen by the difference in Recall (0.823 versus 0.768).

In Figure 5, we show some results of obtained results of the proposed method in comparison with the baselines. Note how the proposed method is quite effective in removing sections of incorrectly identified roads compared to the other approaches.

TABLE II: Comparison with state-of-the-art baselines.

Method	IoU	Precision	Recall	F1-Score
D-LinkNet [9]	0.465	0.585	0.693	0.635
U-net road model [4]	0.454	0.503	**0.823**	0.625
Ours	**0.571**	**0.690**	0.768	**0.727**

VII. Conclusion

In this paper, we propose a novel approach for road extraction that combines contextual and pixel-level information. Experimental results, performed using the proposed Amazon Wild Roads (AWR) dataset, show that the method is robust to false positives and very effective in identifying road areas. This was reflected in the final results, where the proposed method generated the best outcomes in terms of IoU, Precision, and F1-score, outperforming other state-of-the-art baselines.

The contributions of this paper open new opportunities for mapping roads in the Amazon region. In future work, we intend to investigate large-scale aspects of unofficial road detection in the Amazon and carry out an analysis of the impact of these roads on deforestation.

Acknowledgment

We thank CAPES, FAPEMIG, CNPq (306955/2021-0), and the Serrapilheira Institute (grant R-2011-37776) for funding.

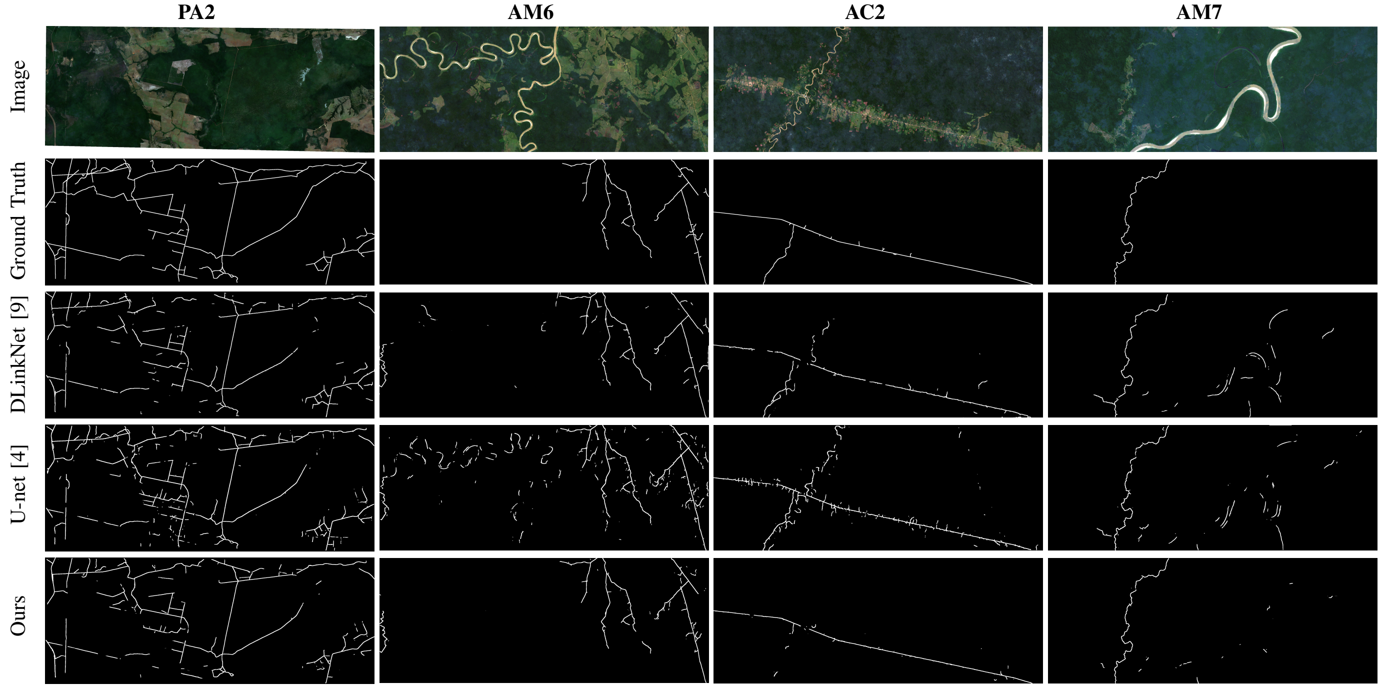

Fig. 5: Results obtained with the proposed approach and the state-of-the-art methods for distinct areas of the AmazonWildRoads dataset. White areas represent roads, while the black regions are non-road.

References

[1] C. P. Barber, M. A. Cochrane, C. M. Souza Jr, and W. F. Laurance, "Roads, deforestation, and the mitigating effect of protected areas in the amazon," *Biological conservation*, vol. 177, pp. 203–209, 2014.

[2] J. Wan, Z. Xie, Y. Xu, S. Chen, and Q. Qiu, "Da-roadnet: A dual-attention network for road extraction from high resolution satellite imagery," *JSTARS*, vol. 14, pp. 6302–6315, 2021.

[3] Z. Yang, D. Zhou, Y. Yang, J. Zhang, and Z. Chen, "Transroadnet: A novel road extraction method for remote sensing images via combining high-level semantic feature and context," *GRSL*, vol. 19, pp. 1–5, 2022.

[4] J. Botelho Jr, S. C. Costa, J. G. Ribeiro, and C. M. Souza Jr, "Mapping roads in the brazilian amazon with artificial intelligence and sentinel-2," *Remote Sensing*, vol. 14, no. 15, p. 3625, 2022.

[5] M. Dixit, K. Chaurasia, and V. K. Mishra, "Dilated-resunet: A novel deep learning architecture for building extraction from medium resolution multi-spectral satellite imagery," *Expert Systems with Applications*, vol. 184, p. 115530, 2021.

[6] C. Ayala, C. Aranda, and M. Galar, "Sub-pixel width road network extraction using sentinel-2 imagery," in *IGARSS*, 2021, pp. 2174–2177.

[7] W. Wang, N. Yang, Y. Zhang, F. Wang, T. Cao, and P. Eklund, "A review of road extraction from remote sensing images," *Journal of traffic and transportation engineering*, vol. 3, no. 3, pp. 271–282, 2016.

[8] A. Mosinska, P. Marquez-Neila, M. Koziński, and P. Fua, "Beyond the pixel-wise loss for topology-aware delineation," in *CVPR*, 2018, pp. 3136–3145.

[9] L. Zhou, C. Zhang, and M. Wu, "D-linknet: Linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction," in *CVPRW*, 2018, pp. 182–186.

[10] X. Yang, X. Li, Y. Ye, R. Y. Lau, X. Zhang, and X. Huang, "Road detection and centerline extraction via deep recurrent convolutional neural network u-net," *TGRS*, vol. 57, no. 9, pp. 7209–7220, 2019.

[11] A. Batra, S. Singh, G. Pang, S. Basu, C. Jawahar, and M. Paluri, "Improved road connectivity by joint learning of orientation and segmentation," in *CVPR*, 2019, pp. 10 385–10 393.

[12] K. Nogueira, M. Dalla Mura, J. Chanussot, W. R. Schwartz, and J. A. Dos Santos, "Dynamic multicontext segmentation of remote sensing images based on convolutional networks," *TGRS*, vol. 57, no. 10, pp. 7503–7520, 2019.

[13] A. Abdollahi, B. Pradhan, and A. Alamri, "Vnet: An end-to-end fully convolutional neural network for road extraction from high-resolution remote sensing data," *IEEE Access*, vol. 8, pp. 179 424–179 436, 2020.

[14] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *ICCV*, 2021.

[15] S. Oehmcke, C. Thrysøe, A. Borgstad, M. A. V. Salles, M. Brandt, and F. Gieseke, "Detecting hardly visible roads in low-resolution satellite time series data," in *IEEE ICBD*, 2019, pp. 2403–2412.

[16] L. Carrasco, A. W. O'Neil, R. D. Morton, and C. S. Rowland, "Evaluating combinations of temporally aggregated sentinel-1, sentinel-2 and landsat 8 for land cover mapping with google earth engine," *Remote Sensing*, vol. 11, no. 3, p. 288, 2019.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *CVPR*, 2016, pp. 770–778.

[18] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," in *ICLR*, 2020.

[19] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," in *CVPR*, 2022, pp. 11 976–11 986.

[20] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.

[21] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, 2018, pp. 3–11.

[22] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *ECCV*, 2018, pp. 801–818.

[23] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *ICML*, 2019, pp. 6105–6114.

[24] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *CVPR*, 2017, pp. 1492–1500.

[25] S. Jadon, "A survey of loss functions for semantic segmentation," in *IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology*, 2020, pp. 1–7.