

TOWARDS FINE-GRAINED ROAD MAPS EXTRACTION USING SENTINEL-2 IMAGERY

C. Ayala^{1,*}, C. Aranda¹, M. Galar²

¹Tracasa Instrumental, Calle Cabárceno, 6, 31621 Sarriguren, Navarra - (cayala, caranda)@tracasa.es

²Institute of Smart Cities (ISC), Public University of Navarre,
Campus de Arrosadía s/n, 31006, Pamplona, Spain - mikel.galar@unavarra.es

KEY WORDS: Sentinel-2, Remote Sensing, Road Network Extraction, Deep Learning, Convolutional Neural Networks.

ABSTRACT:

Nowadays, it is highly important to keep road maps up-to-date since a great deal of services rely on them. However, to date, these labours have demanded a great deal of human attention due to their complexity. In the last decade, promising attempts have been carried out to fully-automatize the extraction of road networks from remote sensing imagery. Nevertheless, the vast majority of methods rely on aerial imagery (< 1 m), whose costs are not yet affordable for maintaining up-to-date maps. This work proves that it is also possible to accurately detect roads using high resolution satellite imagery (10 m). Accordingly, we have relied on Sentinel-2 imagery considering its freely availability and the higher revisit times compared to aerial imagery. It must be taken into account that the lack of spatial resolution of this sensor drastically increases the difficulty of the road detection task, since the feasibility to detect a road depends on its width, which can reach sub-pixel size in Sentinel-2 imagery. For that purpose, a new deep learning architecture which combines semantic segmentation and super-resolution techniques is proposed. As a result, fine-grained road maps at 2.5 m are generated from Sentinel-2 imagery.

1. INTRODUCTION

In recent years, there has been a growing research interest in extracting valuable insights from remote sensing imagery. Occasionally assisted by semi-automatic tools, experts carefully observe aerial and satellite products to extract high-level features. Since, those labours demand a great deal of human attention, many attempts have been carried out to automatize them (Zhang et al., 2016).

Among the different objects that could be detected in remote sensing imagery, road networks have gained popularity in the last decade (Nachmany and Alemohammad, 2019). Ranging from urban planning and cadastral activities to route optimization, there are a plethora of services that rely on up-to-date road maps to properly work. However, nowadays it is practically impossible to keep road maps updated due to the great amount of resources demanded by those labours.

In recent years, deep learning has received a lot of attention in both scientific research and practical application. As a result, a great deal of automation attempts combining remote sensing imagery with deep learning models have been proposed (Zhu et al., 2017). Nevertheless, the feasibility to detect complex objects such as road networks, still largely depend on the spatial resolution of the imagery used. That is, the higher the spatial resolution is, the easier it will be to detect tricky objects.

The vast majority of works employ aerial imagery (< 1 m) to detect roads, whereas there are few who approach the problem using very high resolution imagery (< 10 m) (Hoesser and Kuenzer, 2020). However, increasing the spatial resolution result in higher costs, hindering its application on a daily basis.

In Europe, remote sensing data is becoming more accessible and affordable thanks to the Copernicus programme coordinated and managed by the European Commission in partnership

with the European Space Agency (ESA). Moreover, information produced in the framework of Copernicus is being made available free-of-charge to the public. Under the Copernicus programme ESA is currently developing seven Sentinel missions to monitor different Earth aspects. In this work we focus on Sentinel-2 (S2), a multi-spectral sensor that provides high-resolution optical images composed of thirteen bands, principally in the visible/near infrared (VNIR) and short-wave infrared spectral range (SWIR). Among the thirteen bands, only the Red, Green, Blue and Near Infrared ones are provided at the greatest resolution of 10 m. It must be noted that, the high revisit times provided by S2 allows one to monitor an specific zone up to 70 times per year in the equator.

Aiming at showing that using high resolution imagery (10 m) it is possible to accurately detect road networks, this paper proposes a novel deep learning architecture which fuses semantic segmentation and super-resolution techniques. As a result, not only associated costs could be reduced but also, the cartography update frequency could be increased due to the high revisit times of high resolution satellites.

The experimental study to validate the proposed approach consists of a set of 20 cities spread across the Spanish territory. The data-set has been handcrafted combining S2 imagery with OpenStreetMap (OSM) annotations. Taking into account that OSM is generated by volunteers, labeling errors may be present. However, OSM has been widely used to label remote sensing data-sets proving that it is possible to achieve a high performance using OSM noisy labels (Kaiser et al., 2017). Additionally, labelling errors in roads are fewer compared to other elements such as buildings. Our approach has been evaluated using the intersection over union (IoU) and F-score metrics, which are commonly used in semantic segmentation tasks.

The rest of this work is organized as follows. Firstly, previous works that uses deep learning techniques to detect roads are

* Corresponding author

briefly recalled in Section 2. Thereafter, our proposal for detecting road networks using high resolution imagery is presented in Section 3. Afterwards, the experiments are carried out and discussed in Section 4. Finally, Section 5 concludes this work and present some future research.

2. RELATED WORKS

Nowadays, deep learning-based methods have become the standard for image processing and computer vision tasks (Goodfellow et al., 2016). Accordingly, to deal with images, Convolutional Neural Networks (CNNs) (Lecun et al., 1998) are often considered. Remote sensing labours, such as generating road maps, have taken advantage of that breakthrough.

The road network extraction problem can be seen as a semantic segmentation task, since the aim is to assign a label (road or no-road) to every pixel in a image, thereby allowing the object to be extracted from the image.

The use of deep learning techniques to extract road networks from aerial images dates back to 2010 when Mnih et al. (Mnih and Hinton, 2010) proposed to utilize restricted Boltzmann machines (RBMs) to address this problem. Moreover, to refine the output mappings, they incorporated semantic structures such as the road connectivity through the inclusion of a post-processing network. Shortly thereafter, in 2012, Mnih and Hinton improved their methodology, making use of CNNs (Mnih and Hinton, 2012).

Promising road detection attempts started to emerge as research in CNNs progressed. Saito et al. (Saito et al., 2016) proposed a new technique to train CNNs efficiently for extracting roads and buildings simultaneously. Zhang et al. (Zhang et al., 2018) relayed on a modified version of the U-Net with residual connections to segment roads. Zhou et al. (Zhou et al., 2018) combined a LinkNet with a pretrained encoder and dilated convolutions to handle this task.

It must be noted that aforementioned works make use of aerial imagery (< 1 m). This is due to the complexity of detecting roads considering that tree occlusions, building shadows, and atmospheric and ground conditions can drastically increase its difficulty. However, the high cost and limited availability of aerial products hinders the possibility of applying proposed methodologies on a daily basis.

High resolution sensors (10 m) are not a priori the proper ones for detecting complex elements due to their limited spatial resolution (Hoeser and Kuenzer, 2020). For this reason few works have assessed their capabilities for the road detection task (Oehmcke et al., 2019, Radoux et al., 2016).

However, high resolution satellite imagery must be taken into account given not only their low cost but also their high revisit times. Since information produced in the framework of Copernicus is available free-of-charge, in this work we make use of S2 imagery. Moreover, S2 provides an average revisit interval of 4.5 days at the equator enabling a great deal of monitoring activities.

3. TOWARDS FINE-GRAINED ROAD MAPS EXTRACTION USING SENTINEL-2 IMAGERY

Considering the elevated costs of using aerial imagery (< 1 m) we have opted for using high resolution satellite imagery (10 m)

to detect road networks. A benefit of employing high resolution imagery is that the revisit times are higher, making it possible to keep road maps updated on a regular basis. However, the limited resolution of high resolution imagery drastically increases the complexity of this task.

Traditional fully convolutional neural networks (FCNs) (Long et al., 2015) output segmentation masks maintaining the resolution given at the input. That is, if we use a S2 image as input (10 m) the resulting road segmentation mask will have a spatial resolution of 10 m. However, it must be taken into account that in a pixel may coexist both (road and no-road) classes since roads can have sub-pixel width.

This work aims at showing that it is possible to generate semantic segmentation masks with greater spatial resolution than the input. Moreover, enhancing the output resolution allows one to detect roads whose width, depending on the sensor, could reach sub-pixel size. For that purpose, a novel deep learning architecture has been developed combining semantic segmentation and super-resolution techniques.

The most widely used architecture in semantic segmentation tasks is the U-Net fully convolutional network (Ronneberger et al., 2015). Despite, the U-net was originally invented and first used for biomedical image segmentation, it has been adapted for a wide variety of segmentation problems. The model proposed in this work is based on the U-Net as it can be seen in Figure 1.

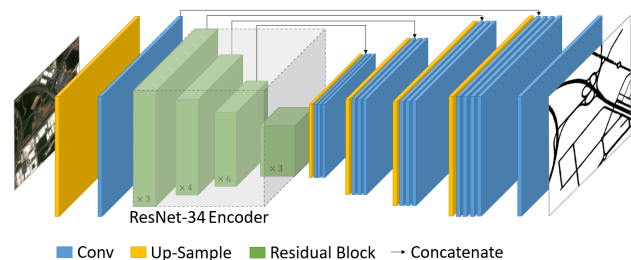


Figure 1. U-Net architecture with ResNet-34 as encoder and a bicubic interpolation layer prior to the feature extractor.

The model consist of a contracting path to capture context information, and a symmetric expanding path that enables precise localization. In classical fully convolutional architectures coarse segmentation masks are generated due to the loss of location information in the encoding path. The lack of fine detail is alleviating by the U-Net model, whose main contribution is to add skip connections. That is to combine high resolution features of the contracting path with the up-sampled output in the expanding path, avoiding losing pattern information.

In order to increase the resolution at the output of the network, deconvolutional layers are often used. However, when appending deconvolutional layers to the U-Net network, they cannot take advantage of the skip connections. For that reason, in this work, two main modifications have been made to the vanilla U-Net architecture:

- An up-scaling layer has been included before the encoder. As a result, the resolution at the input is quadrupled using the bicubic classical interpolation algorithm. This new outlook allows the network to provide an enhanced version of the input, making the most of the U-Net's skip connections and thus avoiding the loss of pattern information.

However, it must be noted that the increase in the input resolution has a negative effect on the computational cost since the number of parameters is also quadrupled. In the experiments we will compare this alternative with the traditional approach of adding deconvolutional layers at the output.

- The base encoder has been replaced with a ResNet-34 (He et al., 2016). Given the increase in the number of parameters due to the up-scaling layer, residual connections are included aiming at reducing the computational cost. Moreover, features extracted by residuals models are more precise than the ones extracted by the vanilla U-Net encoder, because residual connections exploit all the information available along the down-sampling process efficiently.

4. EXPERIMENTAL STUDY

4.1 Data-set

Since open road network extraction data-sets consist of hand-labeled aerial imagery (Demir et al., 2018, Cheng et al., 2017, Bastani et al., 2018), we have opted for generating our own data-set. Considering the free availability of their products as well as the high revisit times provided, S2 was chosen as high resolution sensor. Additionally, we have relied on OpenStreet-Map (OSM) to annotate the imagery on account of the quality of the road labels.

Overall, the pipeline for a generic area of interest is depicted in Figure 2. Firstly, S2 products are queried and downloaded from the Sentinels Scientific Data Hub (SciHub). Despite 13 bands are offered by S2, we will only make use of the Red, Green, Blue and Near Infrared bands, since they are the only ones provided at the greatest resolution of 10 m. Additionally, the Normalized Difference Vegetation Index (NDVI) is computed and concatenated to the other bands.

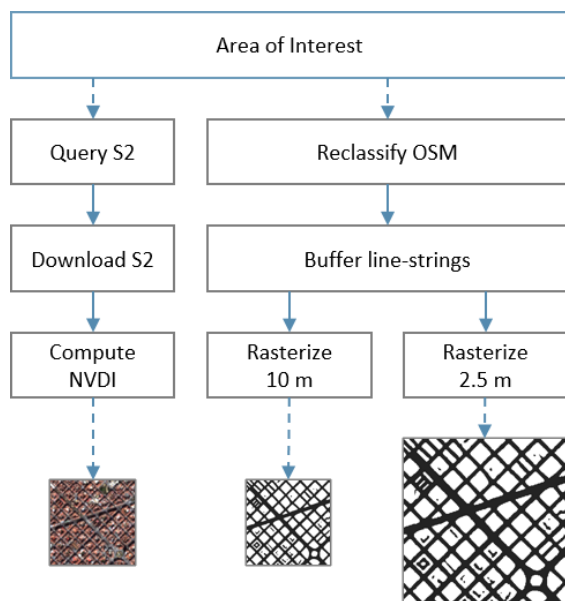


Figure 2. Data-set generation pipeline for a generic area of interest.

Regarding OSM, as a great deal of layers are provided it needs to be reclassified. That is, several road elements outlined in

Table 1 have been aggregated to build up the label road. Since OSM only provides coordinates of road center-lines, they were buffered to match S2 spatial resolution (10 m) before rasterizing (transforming to pixel coordinates). However, other approaches could have been considered such as determining an average road width for each category (Kaiser et al., 2017). Notice that when rasterizing vector data from OSM one can select the desired output resolution. In this case, we have both rasterized to 10 m and 2.5 m to compare the vanilla U-Net with 1x output with the modified versions with 4x output.

Code	Fclass	Description
5111	motorway	Motorway/freeway
5112	trunk	Important roads, typically divided
5113	primary	Primary roads, typically national
5114	secondary	Secondary roads, typically regional
5115	tertiary	Tertiary roads, typically local
5121	unclassified	Smaller local roads
5122	residential	Roads in residential areas
5123	living_street	Streets where pedestrians have priority
5124	pedestrian	Pedestrian only streets
5131	motorway.link	Roads connections (same of lower category)
5132	trunk.link	Roads connections (same of lower category)
5133	primary.link	Roads connections (same of lower category)
5134	secondary.link	Roads connections (same of lower category)

Table 1. OSM classes from the Road layer that have been aggregated to build up the label road.

For this study, we have selected 20 cities spread across the Spanish territory, which have been divided in two sets according to the machine learning standar guidelines (Ripley, 1996). That is, each whole city is assigned to either the training set or test set in order to prevent data leakage, as it is indicated in Table 2. Figure 3 shows how the data-set is geographically distributed.

City	Dimensions	Set
A coruña	704 × 576	Training
Alicante	1216 × 1472	Training
Barakaldo	1088 × 896	Training
Barcelona N.	1152 × 1728	Test
Barcelona S.	896 × 1088	Test
Castellón	1024 × 1024	Training
Córdoba	1088 × 1792	Training
Girona	1536 × 1216	Training
León	1216 × 768	Training
Logroño	768 × 960	Training
Madrid N.	1920 × 2688	Training
Rivas-vacía	1088 × 1088	Training
Salamanca	832 × 960	Training
San Sebastián	512 × 768	Test
Santander	1152 × 1216	Training
Sevilla	2176 × 2368	Training
Palma	1024 × 1344	Test
Pamplona	1600 × 1536	Test
Vitoria	576 × 896	Training
Zaragoza	2304 × 2752	Training

Table 2. Summary of the data-set. Overall, training set is composed by 15 zones (75%) whereas test set consists of 5 zones (25%)

It must be noted that open databases such as OSM usually contains labeling errors. Despite there is an evident lack of precision in rural areas annotations, OSM has been widely used to automatically annotate remote sensing data-sets. Moreover, deep learning-based models have been proved capable of learning from noisy data (Mnih and Hinton, 2012). Although in the future we plan to find ways for improving the generated data, in this work we rely on OSM annotations without performing any



Figure 3. Geographical distribution of the data-set (green training set / red test set).

pre-processing step.

4.2 Experimental framework

Keras (Chollet et al., 2015) has been chosen as deep learning framework to implement the architectures proposed in this work. All the models have been trained for 100K iterations taking batches of 24 samples of 128×128 S2 pixels. Moreover, samples have been randomly taken considering only samples containing at least 5% of road pixels. Adam (Kingma and Ba, 2015) was chosen as optimizer using a learning rate of $1e-3$.

Regarding the loss function, we use a combined loss function (Equation (1)) of Binary Cross-entropy (Ma Yi-de et al., 2004) and Dice Loss (Sudre et al., 2017). Briefly, the Dice loss controls the trade-off between false-negatives and false-positives whereas, the binary cross-entropy is used for curve smoothing.

$$\begin{aligned} L(y, \hat{y}) &= 0.5L_{BCE}(y, \hat{y}) + 0.5L_{DICE}(y, \hat{y}) \\ L_{BCE}(y, \hat{y}) &= -(y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})) \\ L_{DICE}(y, \hat{y}) &= \frac{2y\hat{y} + 1}{y + \hat{y} + 1} \end{aligned} \quad (1)$$

The different approaches have been evaluated using both the intersection over union (IoU) and F-score metrics. Closely related to the Dice coefficient, the IoU quantifies the percent overlap between the target mask and the predicted output. Both IoU and F-score (F1) are computed as indicated in Equation (2).

$$IoU(y, \hat{y}) = \frac{y \cap \hat{y}}{y \cup \hat{y}} \quad F1(y, \hat{y}) = \frac{2y\hat{y}}{y + \hat{y}} \quad (2)$$

Like other works (Mnih and Hinton, 2010, Zhang et al., 2018) we have performed precision relaxation aiming at reducing the impact of the low spatial resolution on the metrics. That is, we have discarded doubtful pixels located on the edges of the roads.

The experiments have been run on a computing node with an Intel Xeon E5-2609 v4 @ 1.70 GHz processor with 64 GB of RAM memory and 4x NVIDIA RTX2080Ti GPUs with 11 GB of RAM.

4.3 Results and discussion

As we have already mentioned in Section 3, the most widely used technique to increase the resolution at the output of a fully convolutional network is to append deconvolutional layers. Accordingly, the architecture proposed in this work has been compared not only to the vanilla U-Net but also to an evolved version with two extra deconvolutional layers.

In Table 3, we show the result in terms of IoU and F-score for all the models in the road detection task. Despite metrics are computed for each city from the test set individually, the overall performance is also included. Moreover, the best results achieved are presented in **boldface**.

In summary, the vanilla U-Net achieves an average IoU of 0.36 and F-score of 0.53. However, when the resolution at the output is enhanced, both metrics increase. Accordingly, the IoU gets almost doubled (0.61 and 0.68 for the Deconv and Bicubic models respectively) and the F-score also increases (0.75 and 0.81, respectively). This is mainly due to the coexistence of multiple classes in a single pixel derived from the limited spatial resolution. Moreover, as it can be observed, the results are consistent across all the cities in the test set.

Table 3 also reveals that up-scaling the input prior to the feature extractor results in a better performance compared to performing the up-scaling at the output through the use of deconvolutional layers (0.68 vs. 0.61 in terms of IoU and 0.81 vs. 0.75 in terms of F-score). Therefore, increasing the resolution at the input makes it possible for the U-Net network to learn not only how to semantically segmentate the images but also how to super-resolve them internally. Moreover, since our approach keeps the vanilla U-Net architecture intact, shared connections between the encoder and the decoder can refine the resulting segmentation masks. As a result, pattern information is conserved diminishing the lack of fine detail in the predictions.

Figure 4 visually compares the performance of the proposed architecture in some samples taken from the test set. Closely looking at this figure one draws the same conclusions as those looking at Table 3, with some additional information. Standard semantic segmentation models such as the U-Net struggle to detect sub-pixel sized elements. Therefore, when the output resolution is increased, models have more room for accurately defining small objects. However, when dealing with complex scenarios such as rural areas, only adding the deconvolutional layers produces a behaviour similar to the usage of Conditional Random Fields, removing noise and better defining the roads. On the contrary, the bicubic interpolation at the input allows one not only to remove noise, but also to detect roads with sub-pixel width.

Finally, to assess the generalization capability of the proposed architecture to other areas different from the training and testing ones, road networks have been extracted for the main cities in the Iberian peninsula. Accordingly, the complete map is available at our web page.¹

5. CONCLUSIONS AND FUTURE WORK

In this paper a new deep learning architecture to detect road networks in S2 imagery has been presented. Moreover, our model is capable of detecting roads regardless of their width. Results, demonstrates that increasing the resolution at the input of the feature extractor using a classical interpolation algorithm such as bicubic can boost both the IoU and F-score metrics.

Nevertheless, there are still several research lines on this topic that should be addressed in the future. The data-set could be

¹ <https://tracasa.es/sin-categoria-en/tracasa-succeeds-in-identifying-roads-and-highways-using-images-from-sentinel-satellites-super-resolved-with-artificial-intelligence/>

City	U-Net x1		U-Net + Deconv x4		U-Net + Bicubic x4 (ours)	
	IoU	F-score	IoU	F-score	IoU	F-score
Barcelona N.	0.2957	0.4564	0.6250	0.7692	0.7066	0.8280
Barcelona S.	0.3409	0.5085	0.5231	0.6869	0.6812	0.8103
San Sebastián	0.4044	0.5759	0.6199	0.7654	0.6966	0.8212
Palma	0.3956	0.5670	0.6430	0.7827	0.6677	0.8007
Pamplona	0.3921	0.5633	0.6400	0.7804	0.6950	0.8200
Overall	0.3657	0.5345	0.6102	0.7569	0.6894	0.8161

Table 3. Results obtained in test set with metric relaxation.

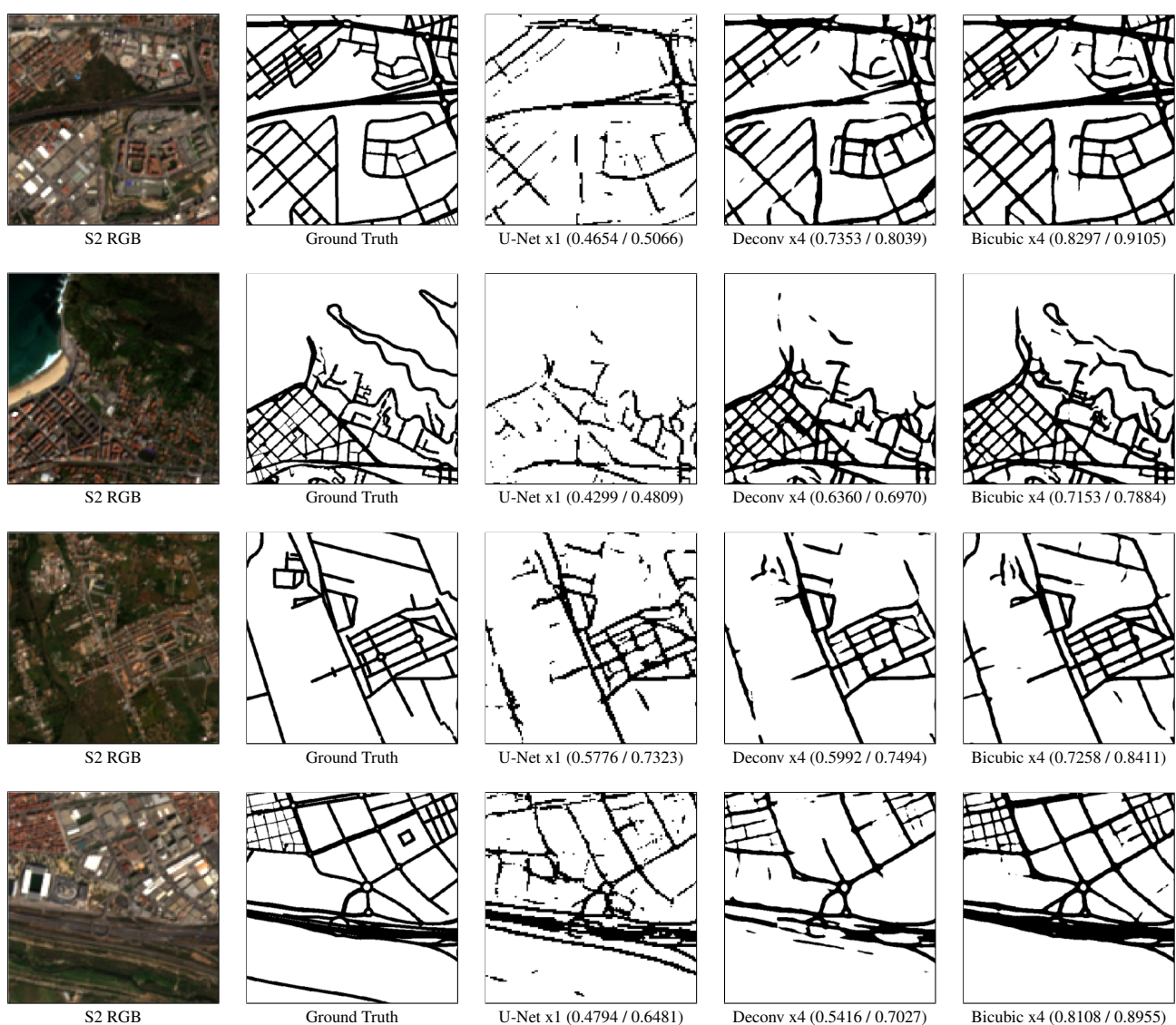


Figure 4. Visual comparison between the vanilla U-Net (col. 3), a modified version with two deconvolutional layers (col. 4), and the proposed architecture which applies a bicubic interpolation prior to the feature extractor (col. 5). The relaxed IoU and F-score metrics have been also included (IoU / F-Score).

improved including more images for training and testing. Additionally, the location of these images should cover different parts of the world to make the network more robust. Moreover, different images could be considered for a single zone taking advantage of the higher revisit times of S2.

With respect to the deep learning architecture, we would like to compare the proposed model with other state-of-the-art approaches such as HRNet. Likewise, it would be interesting to try out other feature extractors different from ResNet-34.

6. ACKNOWLEDGEMENTS

Christian Ayala was partially supported by the Government of Navarra under the industrial PhD program 2020 reference 0011-1408-2020-000008. Mikel Galar was partially supported by Tracasa Instrumental S.L. under projects OTRI 2018-901-073, OTRI 2019-901-091 and OTRI 2020-901-050.

REFERENCES

- Bastani, F., He, S., Abbar, S., Alizadeh, M., Balakrishnan, H., Chawla, S., Madden, S., DeWitt, D., 2018. Roadtracer: Automatic extraction of road networks from aerial images. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 4720–4728.
- 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*, PP, 1–16.
- Chollet, F. et al., 2015. Keras. <https://keras.io>.
- 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. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 172–17209.
- Goodfellow, I., Bengio, Y., Courville, A., 2016. *Deep Learning*. MIT Press. <http://www.deeplearningbook.org>.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.
- Hoeser, T., Kuenzer, C., 2020. Object detection and image segmentation with deep learning on Earth observation data: A review-part I: Evolution and recent trends. *Remote Sensing*.
- Kaiser, P., Wegner, J. D., Lucchi, A., Jaggi, M., Hofmann, T., Schindler, K., 2017. Learning Aerial Image Segmentation from Online Maps. *CoRR*.
- Kingma, D. P., Ba, J. L., 2015. Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*.
- Lecun, Y., Bottou, L., Bengio, Y., Ha, P., 1998. LeNet. *Proceedings of the IEEE*.
- 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*.
- Ma Yi-de, Liu Qing, Qian Zhi-bai, 2004. Automated image segmentation using improved pcnn model based on cross-entropy. *Proceedings of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing, 2004.*, 743–746.
- Mnih, V., Hinton, G., 2010. Learning to detect roads in high-resolution aerial images. *Computer Vision – ECCV 2010*, 6316, 210–223.
- Mnih, V., Hinton, G., 2012. Learning to label aerial images from noisy data. *Proceedings of the 29th International Conference on Machine Learning, ICML'12*, Omnipress, Madison, WI, USA.
- Nachmany, Y., Alemohammad, H., 2019. Detecting roads from satellite imagery in the developing world. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.
- Oehmcke, S., Thrysoe, C., Borgstad, A., Salles, M. A. V., Brandt, M., Gieseke, F., 2019. Detecting hardly visible roads in low-resolution satellite time series data. *2019 IEEE International Conference on Big Data (Big Data)*, 2403–2412.
- Radoux, J., Chomé, G., Jacques, D. C., Waldner, F., Bellemans, N., Matton, N., Lamarche, C., D'Andrimont, R., Defourny, P., 2016. Sentinel-2's Potential for Sub-Pixel Landscape Feature Detection. *Remote Sensing*, 8(6).
- Ripley, B. D., 1996. *Pattern Recognition and Neural Networks*. Cambridge University Press.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*.
- Saito, S., Yamashita, Y., Aoki, Y., 2016. Multiple Object Extraction from Aerial Imagery with Convolutional Neural Networks. *Journal of Imaging Science and Technology*, 60, 10402–1/10402.
- Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., Jorge Cardoso, M., 2017. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*.
- Zhang, L., Zhang, L., Du, B., 2016. Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art. *IEEE Geoscience and Remote Sensing Magazine*, 4(2), 22–40.
- Zhang, Z., Liu, Q., Wang, Y., 2018. Road Extraction by Deep Residual U-Net. *IEEE Geoscience and Remote Sensing Letters*, 15(5), 749–753.
- Zhou, L., Zhang, C., Wu, M., 2018. D-linknet: Linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 192–1924.
- Zhu, X. X., Tuia, D., Mou, L., Xia, G., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geoscience and Remote Sensing Magazine*.