

Exposure modelling of transmission towers using street-level imagery and a deep learning object detection model

Luigi Cesarini¹, Mario Martina
Scuola Universitaria Superiore IUSS, Pavia, Italy

Rui Figueiredo, Xavier Romão
CONSTRUCT-LESE, Faculty of Engineering, University of Porto, Porto, Portugal

ABSTRACT

Exposure modelling is a vital component of disaster risk assessments, providing geospatial information of assets at risk and their characteristics. Detailed information about exposure bring benefits to the spatial representation of a rapidly changing environment and allows decision makers to establish better policies aimed at reducing disaster risk. This work proposes and demonstrates a methodology aimed at linking together volunteered geographic information from OpenStreetMap (OSM), street-level imagery from Google Street View (GSV) and deep learning object detection models into the automated creation of exposure datasets of power grid transmission towers, an asset particularly vulnerable to strong wind among other perils. The methodology is implemented through a start-to-end pipeline that starting from the locations of transmission towers derived from the power grid layer of OSM's world infrastructure, can assign relevant features of the tower based on the identification and classification returned from an object detection model over street-level imagery of the tower, obtained from GSV. The initial outcomes yielded promising results towards the establishment of the exposure dataset. For the identification task, the YOLOv5 model returned a mean average precision (mAP) of 83.57% at intersection over union (IoU) of 50%. For the classification problem, although predictive performance varies significantly among tower types, we show that high values of mAP can be achieved when there is a sufficiently high number of good quality images with which to train the model.

Keywords: Risk assessment, exposure, OpenStreetMap, street-level imagery, object detection

INTRODUCTION

The built environment is constantly under the threat of natural hazards, and climate change will only exacerbate such perils (IPCC, 2021). The assessment of natural hazard risk requires exposure models representing the characteristics of the assets at risk, which are crucial to subsequently estimate damage and impacts of a given hazard to such assets (Figueiredo & Martina, 2016). Therefore, the exposure modelling of the built environment provides a basis for decision makers to establish policies and strategies to reduce disaster risk (Pearson & Pelling, 2015), offering geospatial information with a fine level of spatial detail regarding an often hard to track and rapidly changing built environment (Wieland et al., 2012).

Therefore, studies addressing the exposure modelling of the built environment are expanding, supported by the emergence of new modelling possibilities associated to technological progress. In fact, several works are introducing data collected from volunteered geographic information (VGI), user-generated content, and remote sensing data (Klonner et al., 2016). These methods generate large amounts of data that typically require a time-consuming extraction of the necessary information. This labour-intensive task is well suited for the capability of machine learning (ML) models, particularly deep learning models, to handle massive amounts of data (LeCun et al., 2015). For example, Geiß et al. (2017) use a ML classifier ensemble method to combine OpenStreetMap and remote sensing data and extract exposure information regarding the number of buildings

¹ Luigi Cesarini, luigi.cesarini@iusspavia.it

and the distribution of the population in the territory of Valparaíso, Chile. Several studies involve the use of deep learning models to extract information about the characteristics of buildings in urban areas (i.e., type of structures, numbers of floor, number of windows, type of roof etc.) that are then used to perform seismic risk assessments (Aravena Pelizari et al., 2021; Xu et al., 2019). Iannelli & Dell’Acqua (2017) use a deep learning model to estimate the number of floors for each building in an urban setting based on street-level imagery. Similarly, deep learning models also found application in the field of electrical infrastructure with a special focus on detection of transmission towers in images (Hu et al., 2018; Wang et al., 2019). However, these studies are either based on images retrieved via drone or are aimed at inspection and management of the tower rather than being used for risk assessment.

In this context, this work proposes a methodology that connects VGI obtained from OpenStreetMap (OSM), street-level imagery from Google Street View (GSV) and deep learning object detection models to create an exposure dataset of electrical transmission towers, an asset particularly vulnerable to strong winds among other perils (i.e., ice loads and earthquakes. (López et al., 2009; Rezaei et al., 2015; Tapia-Hernández & De-León-Escobedo, 2021)). The main objective of the study is to establish and demonstrate a complete pipeline that first obtains the locations of transmission towers from the power grid layer of OSM’s world infrastructure, and subsequently assigns relevant features of each tower based on the classification returned from an object detection model over street-level imagery of the tower, obtained from GSV.

The paper is structured as follows: *Methodology* introduces the key components of the pipeline, highlighting their features and criticality, describes the case study selected to pilot the methodology, the training procedures and the data used. The *Results* section highlights the main findings of the work, while *Conclusions* summarizes the most important take home message of the study, limitations, and future development that would improve the application of the methodology presented.

METHODOLOGY

The methodology proposed in this work connects VGI obtained from OpenStreetMap (OSM), street-level imagery from Google Street View (GSV) and deep learning object detection models into the creation of an exposure dataset of electrical transmission towers. First, OpenStreetMap VGI data are used to obtain the location of electricity transmission towers at the large scale. Street-level imagery from Google Street View is then used to collect photographs for each tower. Lastly, a deep learning detection model is used to classify each tower based on a set of classes that are previously defined based on a custom taxonomy. The workflow is represented in Figure 1.

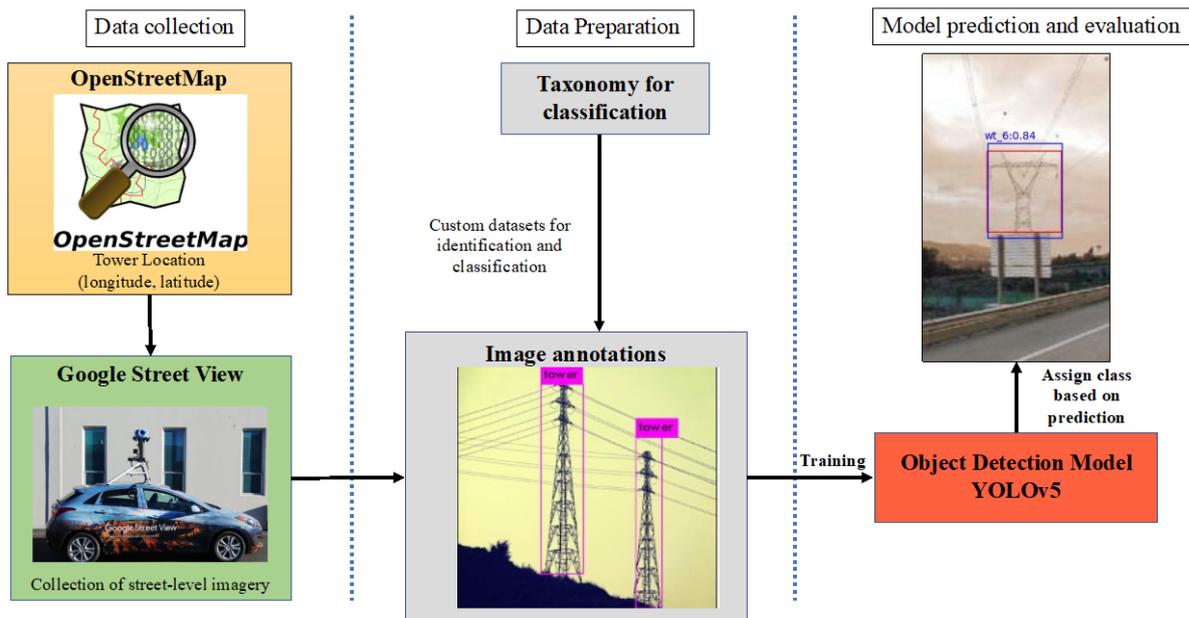


Figure 1. Workflow of the proposed methodology.

Data preparation

In the object detection field, it is common practice to train and test new advancements in architectures, loss functions and tweaks in the algorithms on well-established datasets (Everingham et al., 2010; Lin et al., 2015). This way, it is possible for different researchers to measure the improvements brought by their novelties and compare the performances of the model over the same baseline. On the other hand, when the objective of the study is not to bring improvements to the model but using it in real-life application (i.e., using the model to identify and classify towers), adopting an already existing datasets could be detrimental towards achieving the best overall possible performances. The dataset might not contain images for the class of interest, or the accuracy of the image in describing the classes might not be good enough to reflect the outcome expected, thus, the training of the model requires the usage of custom datasets (Chiu et al., 2020) .

In this work, two different tasks were carried out: identification and classification of transmission electrical towers. The first task was used to test the ability of a model to recognize whether a tower is present in an image, while the second task assigned a category to each tower based on a taxonomy derived from a compilation of the most used types of towers (Table 1).

Table 1. Taxonomy of the tower's classification

Name	Family	Label	Icon
Single level	Self-supporting	ss_1	
Double level	Self-supporting	ss_2	
Triple level	Self-supporting	ss_3	
Modified delta structure	Self-supporting	ss_4	
Delta	Waist-type	wt_5	
Portal	Waist-type	wt_6	
Tubular single level	Monopole	mono_7	
Tubular double level	Monopole	mono_8	
Tubular triple level	Monopole	mono_9	

Tubular modified delta structures	Monopole	mono_10	
---	----------	---------	---

The two tasks call for different levels of detail in the images used to train the models, therefore, two datasets were created and manually annotated. There is no consensus on the number of images required to properly train a model (Shahinfar et al., 2020). Nguyen et al. (2019) used 28674 images to develop their model used to monitor and inspect power line components, while Hui et al. (2018) used 600 images for training UAV to autonomously identify and approach towers to perform inspection. It has been proved that accurately annotating the ground truth when creating the datasets for training brings benefits to the performances of the model beyond the sheer numerosity of each class in the dataset (Zhang et al., 2016). However, when creating custom datasets for multiclass object detection, it is vital to have a balance in the numerosity of images per class (Olivier & Raynaud, 2021). The characteristics of the two datasets are reported in Table 2:

Table 2. Training datasets features

Task	# Images	Image size	# Classes
Identification	300	512x512	1
Classification	750	512x512	10

During the training, both datasets were partitioned into training, validation and testing in accordance with a 70/20/10 split. This step is crucial to avoid overfitting creating a model able to generalize outside of the set of data used for training (Reitermanova, 2010).

Model prediction and evaluation

Object detection models

Object detection is defined as the task of predicting the location of an object in an image along with the class associated with the object. In recent years, two main types of machine learning models, commonly referred to as object detectors, are used to perform such tasks: two-stage detectors and single-stage detectors (Lazoglou & Anagnostopoulou, 2017). In this work, the model You Only Look Once (YOLOv5) (Jocher et al., 2021) was chosen to represent the single stage detectors family.

YOLOv5 is the last iteration of the YOLO family models (Redmon et al., 2016) and, differently from two-stage detectors, treats the classification task as a regression problem. The model uses the image as an input and learns the class of the object and the coordinates of the bounding box as if they were the parameters of a regression (Soviany & Ionescu, 2018). For additional details on single-stage detectors and the YOLOv5 model, the reader is referred to Jocher et al., (2021).

Evaluation metrics

The output of an object detection model is a bounding box framing the object in the input image along with the class and a confidence associated to the assigned class. The goodness of such prediction can be evaluated using several metrics. Regardless of the metric adopted, each single prediction produced by the model falls into one of the following scenarios:

- True positive (TP): when the correct detection of a ground truth box occurs.
- False positive (FP): when the model detects an object that is not present or the model mislabel an existing object (e.g., the model assigns the label dog to an image of a cat).
- False negative (FN): when a ground truth bounding box goes undetected by the model.
- True negative (TN): when a bounding box is correctly not identified.

Noticeably, in the object detection field, true negatives are not taken into consideration when evaluating model performances, the reason being that there is an almost infinite number of bounding boxes that do not require detection in an image. Once the possible outcomes of a prediction are defined, it is necessary to establish what a correct detection is. The prevailing approach is using the intersection over union (IOU). The IOU is a measurement of the overlapping area between the ground truth bounding box and the detection bounding box, based on the Jaccardi Index (Ivchenko & Honov, 1998) and represented by the following equation (Eq.1):

$$J(B_p, B_{gt}) = IOU = \frac{area(B_p \cap B_{gt})}{area(B_p \cup B_{gt})} \quad (Eq. 1)$$

An IOU=1 indicates a perfect match between detection and ground truth, while IOU=0 means no overlap between the predicted bounding box and the box of the actual object. By virtue of using the value of IOU as a threshold to compute the metrics employed to evaluate the models, it is possible to have more strict or relaxed metrics.

Since identifying areas of the images without any object is irrelevant in the object detection field, most of the metrics used to evaluate these models do not take TN into consideration. As such, precision (P) and recall (R) are vastly adopted. Precision gives a measure of the percentage of times that the model is correct when making a prediction. Recall, on the other hand, measures the ability of the model to detect all relevant objects. The formulation of the two metrics is reported in Eq.2.

$$P_r = \frac{\sum_{n=1}^S TP_n}{\sum_{n=1}^S TP_n + \sum_{n=1}^{N-S} FP_n} = \frac{\sum_{n=1}^S TP_n}{all\ detections} \quad (Eq.2)$$

$$R_c = \frac{\sum_{n=1}^S TP_n}{\sum_{n=1}^S TP_n + \sum_{n=1}^{N-S} FN_n} = \frac{\sum_{n=1}^S TP_n}{all\ ground\ truths}$$

Precision and recall are used in conjunction in the precision-recall curve, a model-wide evaluation metric that highlights the tradeoff between precision and recall for different thresholds of the latter (Saito & Rehmsmeier, 2015).

In recent years, the average precision (AP) established itself as the golden standard in evaluating object detectors. The AP is obtained computing the area under the precision-recall curve after removing its typical zig-zag behavior generated by the single-value nature of precision and recall. To remove this effect, the precision is plotted as a function of a set number of recall values.

For models that are carrying out detection of several classes, the performances of the object detectors can be evaluated using the mean average precision, that is the arithmetic mean of the average precision for each class being detected. Being the goodness of the detection function of the IOU, it is possible to retrieve the AP across different values of IOU, evaluating the model in stricter or softer conditions. The metrics adopted in this study are: mAP@0.5 and AP[.5:.05:.95]. The mAP@0.5 returns the mean average precision for IOU set at 0.5 (or an overlap of 50%). This is the most basic declination of AP and can be used to compare the performances of the models with a broader number of object detectors models. The AP[.5:.05:.95] computes the average precision of each class at 10 different thresholds of IOU from 0.5 to 0.95. Adopting this metric is beneficial when one is interested in analyzing the ability of the model to perform accurate prediction in terms of precise identification of the ground truth.

Study area and data

The study area for the initial application of the methodology is the Porto district (Portugal), which has an area of around 1360 km² (Figure 4 in the results). The area was found to be representative given its diverse land use, containing both densely populated settlements and rural areas, and the different types of towers that can be found. Within this area, the power grid layer of OpenStreetMap contains 5789 towers distributed on almost 680 kilometres of electrical cables. Retrieving the corresponding street-level imagery for each tower required the adoption of different strategies to overcome problems related to the visibility. Using the building and road layer of OSM made it possible to remove those GSV shots that were obstructed by the face of an edifice by simultaneously finding the closest point on the road network. Furthermore, the SRTM 30m resolution digital elevation model (OpenTopography, 2013) was used to remove camera shots that were instead hindered by the orography.

RESULTS

The results are presented by firstly showcasing the performances of the object detection model on the two tasks previously introduced: identification and classification on the test partition of the datasets, thus, on images that the models never encountered during the training. Secondly, the results of the prediction over the study area are presented analyzing the spatial distribution of the accuracy of the prediction.

Identification

The identification was envisioned as the initial step into the investigation of the potentiality of the proposed methodology. Additionally, being able to recognize a tower in the image was also seen as a preparatory step towards a better outcome in the classification phase. Figure 2 shows the precision-recall curve for the identification task along with the mean average precision value of 83.57% at IoU of 50%. The high value of mAP shows the ability of the model in detecting, and correctly placing the bounding box around the tower in the image. Also, the high values of precision for increasing values of the recall demonstrates the ability of the model to recognize many relevant objects whilst identifying the majority of the towers.

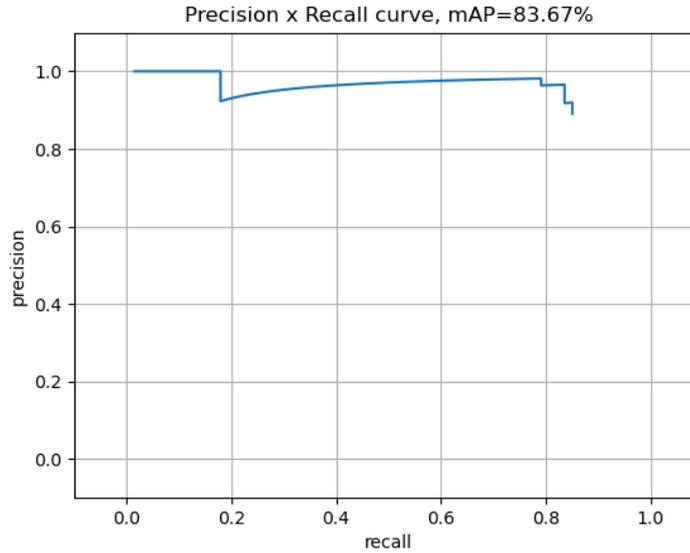


Figure 2. Mean average precision for YOLOv5 for the identification task.

Classification

Regarding the performance of the model for the classification task, Figure 3 shows the mAP for the 10 classes identified in Table 2. Although the drop in performances with respect to the identification task is evident, the breakdown of mAP by class displayed in Table 3, shed some light on the possible causes of such decrease. Indeed, high mAP is observed for most of the classes with especially good performances for ss_3 and mono_8, which are among the most common type of towers and therefore might be easy to recognize.

Table 3. Breakdown by class of the mean average precision for the classification task. The number of ground truth refers to the numerosity of each class in the images of the test partition.

Name	# Ground truth	Label	mAP@0.50
Single level	10	ss_1	3.33
Double level	15	ss_2	45.00
Triple level	48	ss_3	72.03

Modified delta structure	16	ss_4	59.09
Delta	25	wt_5	43.14
Portal	10	wt_6	63.33
Tubular single level	15	mono_7	13.33
Tubular double level	14	mono_8	92.86
Tubular triple level	9	mono_9	59.76
Tubular modified delta structures	5	mono_10	54.29
mAP over all classes	167		50.62

The large difference in overall mAP between classification and identification, can be attributed, in large portion, to the poor performance of two specific classes: the self-supporting single-level category and the monopole tubular single-level category (i.e., ss_1 and mono_7). The bad performances over ss_1 and mono_7 might be due to a low number of samples in the training set and/or to the quality of the image used. Without considering these two classes, the overall mAP would increase to an encouraging 61.2%. It must be noted that although the balance between classes is important, in the real world not all the types of towers are employed at the same rate, and consequently, also finding images for each class to annotate is not a trivial task.

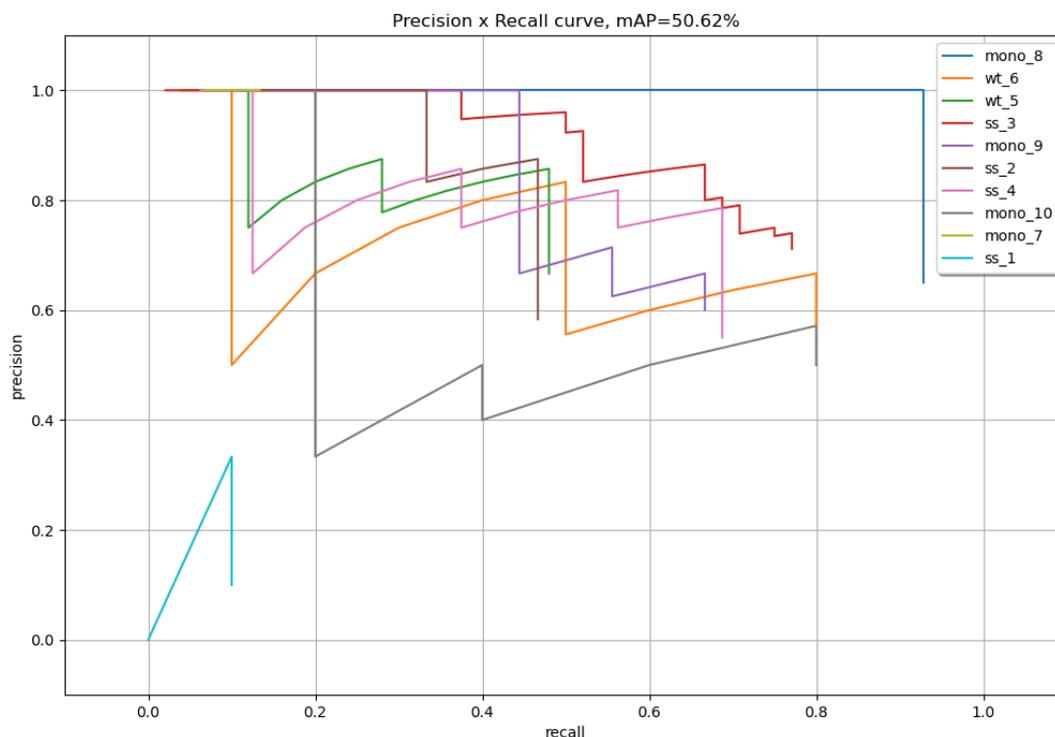


Figure 3. Mean average precision for YOLOv5 for the classification task.

Table 4 summarizes the metrics adopted to evaluate the performances of the object detection models for the two tasks. It is notable, and somehow expected, a decrease in performances going from the identification task to the classification one, even though, the sharp decrease in performances might be attributed to two specific classes rather than the overall performances of the model. Albeit the dip in mAP for the classification task is

undeniable, the values remain in line with other studies that used object detection models for customized tasks (Abdelfattah et al., 2020; Mittal et al., 2020).

Table 4. Summary of evaluation metric

Metric	Identification	Classification
mAP@0.50	83.57	50.62
AP[.5:.05:.95]	51.56	31.00

The results presented up to now were obtained from the test partitioning of the two datasets. The application of the trained models to the whole study area are instead reported in Figure 4. Similarly to the testing set, the images of the whole study area are never seen by the models, although in this case the bounding boxes of the ground truth are not available for the entire area. For the sake of brevity and clarity, the map shows the distribution of the confidence of the prediction over the entire study area only for the identification task. The points in red are the towers for which the identification did not occur. The reasons for this might be multiple: i) the model might not recognize any tower in the image, ii) the image does not contain the tower because no available image of the tower is available, or iii) the image of the tower might be obfuscated by an obstacle that was not taken into consideration in our methodology (i.e., a forest, a passing car). Nevertheless, it is possible to appreciate a large portion of dark blue points, which indicate the detection of towers with high levels of confidence (i.e., above 0.8). Notably, the distribution of the dark blue area is not homogenous. Both the northeast and southeast areas of the map show several towers that were not able to be identified. The missing predictions might be a consequence of the distance of the towers from the road and/or the morphology of the terrain. Another area where it is possible to notice a lower confidence in the predictions is close to the city of Porto. This might be expected due to the disturbance that the built environment inevitably brings to the images that are taken with Google Street View.

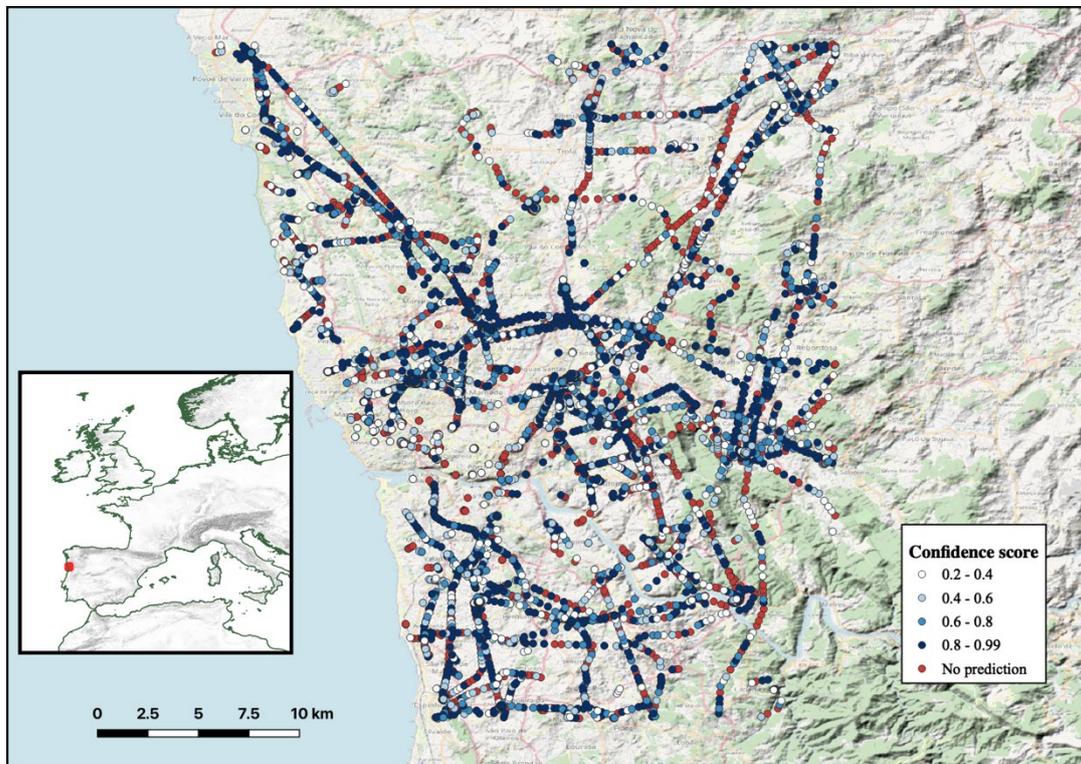


Figure 4. Spatial distribution of the prediction confidence over the study area for the identification task.

CONCLUSIONS

This study introduced a methodology to perform the automated classification of electricity transmission towers based on VGI data, street level imagery and a deep learning object detection model.

We showed that the deep learning model is able, in most of the cases, to identify the towers in the images with a high level of accuracy. This accuracy decreases when the model goes beyond the sole identification also assigning a class to the object, but the performances are still in line with similar studies. The outcomes of this work have clear ramifications on risk assessment for natural disaster, having as primary objective the improvement of the exposure component of this discipline that leans on methods to gather information that do not fully explore the information hidden in user generated content. Nonetheless, the methodology advocated in this paper still has limitations that need to be addressed but also bright possible developments.

The main limitations of this methodology can be divided into two categories: the ones related with the technological advancements and the ones inherently connected to the nature of the methodology and its subject. In fact, power grids are distributed all over the territory and not necessarily close to the roads. For this reason, it is particularly challenging to find clear shots of towers that are very far away from roads, and the chances of getting obstacles in between the shots increase when increasing the distance. In urban settlements, the problem shifts into the need to avoid buildings, which is not always possible in densely populated areas. Belonging to the first category are the restrictions imposed by the providers of the images, in this case, Google Street view. The quality of the images is not ideal and not in line with the instruments used to take the images, which have very high resolutions. Additionally, metadata regarding the camera used to take the picture or the original size in pixel of the image are not provided, making it impossible to derive critical information like the height of the towers. Also, it is worth mentioning that the size of the datasets, and the number of samples in each class have an influence on the training and subsequently on the performances of the models. Collecting and annotating images is a time-consuming task that is accentuated by the singular nature of the problem tackled. While the collection could be difficult to automate given the need of images containing specific types of towers, the annotation can be automated through the usage of unsupervised algorithms that might speed up the process of building larger training datasets. Finally, besides the limitations discussed so far, there is also room for improvements in the architecture of the model. The integration of additional layers might help increase the overall performances of the model at the expense of training and detection time, which for the task envisioned is not vital as it would be for real-time detection in videos. Furthermore, the implementation of attention mechanisms or innovative padding techniques aimed at obtaining the most important information in the input images might be beneficial to the performances of the model, as suggested by Zhu et al., 2021. Notwithstanding, the present work offers a glimpse into the potential of the proposed methodology for natural hazard risk assessment, laying the foundation for further development towards the construction of detailed exposure datasets.

ACKNOWLEDGEMENT

The 3rd and 4th authors would like to acknowledge the financial support of UID/ECI/04708/2020 - CONSTRUCT - Instituto de I&D em Estruturas e Construções, funded by national funds through the FCT/MCTES (PIDDAC).

REFERENCES

- Abdelfattah, R., Wang, X., & Wang, S. (2020). TTPLA: An Aerial-Image Dataset for Detection and Segmentation of Transmission Towers and Power Lines. *ArXiv:2010.10032 [Cs]*. <http://arxiv.org/abs/2010.10032>
- Aravena Pelizari, P., Geiß, C., Aguirre, P., Santa María, H., Merino Peña, Y., & Taubenböck, H. (2021). Automated building characterization for seismic risk assessment using street-level imagery and deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 180, 370–386. <https://doi.org/10.1016/j.isprsjprs.2021.07.004>
- Chiu, M. T., Xu, X., Wei, Y., Huang, Z., Schwing, A. G., Brunner, R., Khachatryan, H., Karapetyan, H., Dozier, I., Rose, G., Wilson, D., Tudor, A., Hovakimyan, N., Huang, T. S., & Shi, H. (2020). Agriculture-Vision: A

- Large Aerial Image Database for Agricultural Pattern Analysis. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2825–2835. <https://doi.org/10.1109/CVPR42600.2020.00290>
- Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2010). The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision*, *88*(2), 303–338. <https://doi.org/10.1007/s11263-009-0275-4>
- Geiß, C., Schauß, A., Riedlinger, T., Dech, S., Zelaya, C., Guzmán, N., Hube, M. A., Arsanjani, J. J., & Taubenböck, H. (2017). Joint use of remote sensing data and volunteered geographic information for exposure estimation: Evidence from Valparaíso, Chile. *Natural Hazards*, *86*(S1), 81–105. <https://doi.org/10.1007/s11069-016-2663-8>
- Hu, Z., He, T., Zeng, Y., Luo, X., Wang, J., Huang, S., Liang, J., Sun, Q., Xu, H., & Lin, B. (2018). Fast image recognition of transmission tower based on big data. *Protection and Control of Modern Power Systems*, *3*(1), 15. <https://doi.org/10.1186/s41601-018-0088-y>
- Hui, X., Bian, J., Zhao, X., & Tan, M. (2018). Vision-based autonomous navigation approach for unmanned aerial vehicle transmission-line inspection. *International Journal of Advanced Robotic Systems*, *15*(1), 172988141775282. <https://doi.org/10.1177/1729881417752821>
- Iannelli, G., & Dell'Acqua, F. (2017). Extensive Exposure Mapping in Urban Areas through Deep Analysis of Street-Level Pictures for Floor Count Determination. *Urban Science*, *1*(2), 16. <https://doi.org/10.3390/urbansci1020016>
- Ivchenko, G. I., & Honov, S. A. (1998). On the jaccard similarity test. *Journal of Mathematical Sciences*, *88*(6), 789–794. <https://doi.org/10.1007/BF02365362>
- IPCC, 2021: Summary for Policymakers. In: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press. In Press.
- Jocher, G., Stoken, A., Ayush Chaurasia, Borovec, J., NanoCode012, TaoXie, Yonghye Kwon, Kalen Michael, Changyu, L., Jiacong Fang, Abhiram V, Laughing, Tkianai, YxNONG, Skalski, P., Hogan, A., Jebastin Nadar, Imyhxy, Mammana, L., ... Wanghaoyang0106. (2021). *ultralytics/yolov5: V6.0 - YOLOv5n "Nano" models, Roboflow integration, TensorFlow export, OpenCV DNN support (v6.0)* [Computer software]. Zenodo. <https://doi.org/10.5281/ZENODO.5563715>
- Klonner, C., Marx, S., Usón, T., Porto de Albuquerque, J., & Höfle, B. (2016). Volunteered Geographic Information in Natural Hazard Analysis: A Systematic Literature Review of Current Approaches with a Focus on Preparedness and Mitigation. *ISPRS International Journal of Geo-Information*, *5*(7), 103. <https://doi.org/10.3390/ijgi5070103>
- Lazoglou, G., & Anagnostopoulou, C. (2017). An Overview of Statistical Methods for Studying the Extreme Rainfalls in Mediterranean. *Proceedings*, *1*(5), 681. <https://doi.org/10.3390/ecas2017-04132>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., & Dollár, P. (2015). Microsoft COCO: Common Objects in Context. *ArXiv:1405.0312 [Cs]*. <http://arxiv.org/abs/1405.0312>
- López, A. L., Rocha, L. E. P., Escobedo, D. de L., & Sánchez, J. (2009). *Reliability and Vulnerability Analysis of Electrical Substations and Transmission Towers for Definition of Wind and Seismic Damage Maps for Mexico*. 13.
- Mittal, P., Singh, R., & Sharma, A. (2020). Deep learning-based object detection in low-altitude UAV datasets: A survey. *Image and Vision Computing*, *104*, 104046. <https://doi.org/10.1016/j.imavis.2020.104046>
- Nguyen, V. N., Jenssen, R., & Roverso, D. (2019). Intelligent Monitoring and Inspection of Power Line Components Powered by UAVs and Deep Learning. *IEEE Power and Energy Technology Systems Journal*, *6*(1), 11–21. <https://doi.org/10.1109/JPETS.2018.2881429>

- Olivier, A., & Raynaud, C. (2021). *Balanced sampling for an object detection problem—Application to fetal anatomies detection*. 13.
- OpenTopography. (2013). *Shuttle Radar Topography Mission (SRTM) Global*. <https://doi.org/10.5069/G9445JDF>
- Pearson, L., & Pelling, M. (2015). The UN Sendai Framework for Disaster Risk Reduction 2015–2030: Negotiation Process and Prospects for Science and Practice. *Journal of Extreme Events*, 02(01), 1571001. <https://doi.org/10.1142/S2345737615710013>
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. *ArXiv:1506.02640 [Cs]*. <http://arxiv.org/abs/1506.02640>
- Reitermanova, Z. (2010). Data splitting. *WDS'10 Proceedings of Contributed Papers*, 10, 31–36.
- Rezaei, S. N., Chouinard, L., Legeron, F., & Langlois, S. (2015). *Vulnerability Analysis of Transmission Towers subjected to Unbalanced Ice Loads*. 8.
- Saito, T., & Rehmsmeier, M. (2015). The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLOS ONE*, 10(3), e0118432. <https://doi.org/10.1371/journal.pone.0118432>
- Shahinfar, S., Meek, P., & Falzon, G. (2020). “How many images do I need?” Understanding how sample size per class affects deep learning model performance metrics for balanced designs in autonomous wildlife monitoring. *Ecological Informatics*, 57, 101085. <https://doi.org/10.1016/j.ecoinf.2020.101085>
- Soviany, P., & Ionescu, R. T. (2018). Optimizing the Trade-off between Single-Stage and Two-Stage Object Detectors using Image Difficulty Prediction. *ArXiv:1803.08707 [Cs]*. <http://arxiv.org/abs/1803.08707>
- Tapia-Hernández, E., & De-León-Escobedo, D. (2021). Vulnerability of transmission towers under intense wind loads. *Structure and Infrastructure Engineering*, 1–16. <https://doi.org/10.1080/15732479.2021.1894183>
- Wang, H., Yang, G., Li, E., Tian, Y., Zhao, M., & Liang, Z. (2019). High-Voltage Power Transmission Tower Detection Based on Faster R-CNN and YOLO-V3. *2019 Chinese Control Conference (CCC)*, 8750–8755. <https://doi.org/10.23919/ChiCC.2019.8866322>
- Wieland, M., Pittore, M., Parolai, S., Zschau, J., Moldobekov, B., & Begaliev, U. (2012). Estimating building inventory for rapid seismic vulnerability assessment: Towards an integrated approach based on multi-source imaging. *Soil Dynamics and Earthquake Engineering*, 36, 70–83. <https://doi.org/10.1016/j.soildyn.2012.01.003>
- Xu, J. Z., Lu, W., Li, Z., Khaitan, P., & Zaytseva, V. (2019). *Building Damage Detection in Satellite Imagery Using Convolutional Neural Networks*. <https://doi.org/10.48550/ARXIV.1910.06444>
- Zhang, S., Benenson, R., Omran, M., Hosang, J., & Schiele, B. (2016). How Far are We from Solving Pedestrian Detection? *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1259–1267. <https://doi.org/10.1109/CVPR.2016.141>
- Zhu, L., Geng, X., Li, Z., & Liu, C. (2021). Improving YOLOv5 with Attention Mechanism for Detecting Boulders from Planetary Images. 19.