

# Detecting tree and wires entanglements with deep learning

Artur André Oliveira · Marcos S.  
Buckeridge · Roberto Hirata Jr.

Received: date / Accepted: date

**Abstract** Power and communication line corridors are usually mixed with urban trees, and this mixing can be the source of multiple issues like fires and communication failures. Nevertheless, urban trees are a valuable resource to the city as they dissipate heat island effects, reduce air pollution and increase general health perception. This work proposes a deep learning approach to detect trees entangled to power and communication lines using street-level imagery and perform quick quantitative and qualitative analyses based on the Grad-CAM++ method. Testing the method was performed using 1001 images from urban trees from the cities of São Paulo and Porto Alegre (both in Brazil). We found an overall accuracy of 74.6% (73.6% for São Paulo and 75.6% for Porto Alegre), suggesting that the methodology could be suitable in the future for city management to avoid risks of accidents due to contact between trees and electrical wiring. This text describes the method, a new data set of urban images, the experimental setup design and tests, and some possible future improvements.

**Keywords** Urban Trees · Power Lines · Google Street View · Deep Learning

---

Artur André Oliveira  
University of São Paulo, Institute of Mathematics and Statistics,  
Rua do Matão, 1010, São Paulo-SP, 05508-090, Brazil  
E-mail: arturao@ime.usp.br

Marcos S. Buckeridge  
University of São Paulo, Institute of Biosciences  
Rua do Matão, trav. 14, nº 321,  
São Paulo-SP, 05508-090, Brazil  
E-mail: msbuck@usp.br

Roberto Hirata Jr.  
University of São Paulo, Institute of Mathematics and Statistics,  
Rua do Matão, 1010, São Paulo-SP, 05508-090, Brazil  
E-mail: hirata@ime.usp.br

## 1 Introduction

Trees are essential to the life of our planet and also to the life of our cities. Urbanization is increasing rapidly on the Earth (Bazaz et al., 2018), which coincides with temperature increases due to the climate change effects (de Coninck et al., 2018). Urban trees are among the critical elements of adaptation of human populations to climate change due to the tree’s ability to decrease the local temperature, interfere with flooding, and help with pollution (Buckeridge, 2015; Locosselli et al., 2019, 2020). Cities vary considerably regarding the number of trees and their distribution (Lüttge and Buckeridge, 2020), probably because of differences in planning concerning their green areas - the quality of planning influences the quality of the environmental services provided by trees (Livesley et al., 2016; Escobedo et al., 2011).

The benefits of maintaining urban trees range from reducing air pollution to preserving wildlife species, to name but a few. However, the occurrence of trees near or on the sidewalks competes with electric power distribution. Moreover, city structures can also interfere with environmental services due to the competition for space between trees and the electrical wiring systems. In such cases, drastic pruning changes urban trees canopy, decreases the shading capacity and spreads diseases. This situation requires action by the public administration and has been generating reports that try to guide urban management (e.g., Most and Weissman 2012; Purcell 2015).

Trees entangled with power lines can conduct energy to the ground, causing electrical shock accidents and energy leakages, which can heat the wood, causing wildfires (Ma et al., 2020; Clarke and White, 2008), blackouts (Ahmad et al., 2013) and partial discharges (Pakonen, 2008). Thus, power line monitoring is an essential matter when it comes to maintenance (Cheng and Song, 2008; Jwa et al., 2009; Lopes et al., 2015). Multiple criteria and indicators are essential for the proper management and monitoring of urban forests (Kenney et al., 2011; Ordóñez and Duinker, 2013). Monitoring the interaction between power lines and trees may be used in conjunction with other criteria and indicators (c.f. for more criteria and indicators for urban forest management (Kenney et al., 2011)) in order to manage urban forests properly. Removing trees because of entanglements is not an acceptable solution, and the improper management of the urban vegetation is likely to be criticized by citizens (Fakes, 2000). From a manager’s perspective, human-centric monitoring is costly and time-consuming, and it may also be dangerous when dealing with power poles and lines (Jwa et al., 2009). Therefore, automatically finding entanglements or future entanglements is a desirable solution to prune tree branches and correctly avoid the problems mentioned above.

Regarding the interaction between vegetation and power lines, (Ma et al., 2020) analyze the ignition properties of vegetation in contact with power lines by applying and measuring (in a controlled environment) an electrical current on branches of different species of trees over time. Over time, the measured electrical current patterns reveal stages in which branches heat up, expel moisture, and finally flashover. (Ma et al., 2020) propose a machine learning

approach to identify when there is a high risk of ignition based on the observation that the time-series of current on the power line can indicate a bushfire's initial stages.

(Kobayashi et al., 2009) exploit the availability of satellite images to detect tall vegetation as trees close to power lines. The benefits of using satellite images include broad geographical extents and the easy inspection of otherwise difficult to access areas. By using multi-spectral images, they can spot healthy vegetation, and by using stereoscopic techniques, they can compute the height of the vegetation close to power lines and transmission towers, allowing one to distinguish between harmless understory vegetation in the ground as bushes and grass from tall trees that can come into contact with the power lines. Based on airborne LiDAR data (Jwa et al., 2009), they propose a method to detect and reconstruct (in 3D) power lines and transmission towers without previous data about tower position or number of lines. Due to the proximity to the ground in airborne imagery data and the structured laser from the LiDAR sensor, one can expect a better resolution in those images when comparing with satellite data, thus making it easier to spot issues arising from interference from tall vegetation and power lines.

Some authors propose methods to map vegetation and other urban objects along roads using ground Mobile Laser Scanner data (Yao and Fan, 2013) or even airborne laser scanners and LiDAR data (Clode and Rottensteiner, 2005; Jwa et al., 2009), providing both a high coverage area and accuracy. However, this kind of data availability is relatively smaller than simple RGB (i.e., colored images from regular cameras) street-level images provided by platforms such as Google Street View (Anguelov et al., 2010), Mappillary Vistas (Neuhold et al., 2017), and OpenStreetCam ( Grab and Contributors, 2019), for instance. Street-level images allow the visualization of tree entanglements with power lines from multiple points of view, leading to better results than aerial and satellite images in urban areas, especially in places where trees can be taller than pylons and cover wires above the above-ground images.

(Berland and Lange, 2017) study the applicability of the GSV as a tool for auditing trees through a virtual tour. Field audits are labor-intensive and costly, and by comparing results from the previous audits with those performed through images from GSV, they concluded that the agreement between tree counts is high for both audits. Identifying the genus and species for larger trees (i.e., with a larger diameter at breast height) also has a high agreement level between the audits, and finally that the time spent in the virtual audit was less than half of that of the field audit.

Tree detection and classification using images is a well-studied topic in computer vision, and multiple approaches exist nowadays. (Cheng and Song, 2008) and (Lopes et al., 2015) propose methods for detecting power poles in images. The former presents a graph-cut algorithm combined with a region proposal method to detect and extract image patches containing the power poles. The latter uses GSV images and a strategy using color, texture, and shape features as input to a multi-layer perceptron neural network. Motivated by the abundance of public geolocalized images, (Wegner et al., 2016) propose

a Convolutional Neural Network (CNN) (Goodfellow et al., 2016) architecture to detect and classify urban objects using aerial and street-level imagery from GSV. One of their conclusions is that using both image modalities will complement each other, and removing any one of the modalities decreases the mean average precision of the final classifier proposed.

Using GSV street-level images and a CNN architecture (Cai et al., 2018) propose a method to classify pixels and estimate the amount of greenery perceived by pedestrians (called Green View Index). They perform a comparison of three methods: a semantic segmentation unsupervised algorithm, a semantic segmentation CNN and a regression CNN, and show that the regression CNN outperforms the other methods. Because the results do not have a natural interpretation, they explore the gradient-weighted Class Activation Map (Grad-CAM++) (Chattopadhyay et al., 2017) visualization method to assess the main input image features (e.g., trees canopies and trunks) responsible for the Green View Index estimation qualitatively.

To the best of our knowledge, none of the works previously performed in the literature deals with the specific problem of detecting the entanglement between overhead power lines and tall vegetation in street-level imagery. This work proposes collecting geographical position data of trees and electricity poles, then images from these positions are collected. These images are labeled using an in-house application to alleviate the annotation process burden and use part of the images to adapt a pre-trained Convolutional Neural Network named MobileNetV2 (Sandler et al., 2018) to detect tree entanglements with power lines near poles in street-level RGB images. Using the tuned network, we achieve over 74% of test accuracy on a test set of 1001 images from two Brazilian cities. The main contributions of this paper are:

1. An approach to detect tree and wire entanglements in street-level imagery based on deep learning;
2. An annotated dataset of tree entanglements with power lines;
3. An open-source software for annotating street-level images.

## 2 Methodology

This section presents our solution to locate tree entanglements with power lines. We first present a dataset of more than six thousand images collected from São Paulo and Porto Alegre, two cities in the Southeast and South of Brazil, respectively, using Google Street View. We then present the strategy to annotate the images, and we finish the section presenting the strategy to tune a Convolutional Neural Network using transfer learning.

### 2.1 Image dataset

We start collecting geographical position data of 2981 trees from an open government public database called GeoSampa (Prefeitura de São Paulo, 2021),

with data for the São Paulo city, and 3111 positions of power poles at Porto Alegre from the energy company CEEE’s website (CEEE, 2016). Using these geographical positions, we applied INACITY to collect street-level images from these positions and create the training and testing datasets. INACITY (Oliveira and Hirata Jr., 2021) is a platform (available at <http://inacity.org/>) to collect and process image data for user-selected geographical regions or positions. The default image provider used in INACITY is Google Street View (GSV). The GSV platform uses an abstraction called panorama to encapsulate a set of stitched images obtained simultaneously with multiple cameras at the same spot but pointing in different directions. The stitched composition of the images creates a spherical ”panorama.”

To collect an image using INACITY, we first sample the nearest GSV panorama from the location  $L$  of some target object (i.e., trees from GeoSampa or power poles from CEEE’s dataset). After finding the nearest panorama from  $L$ , we extract an image from the panorama such that the target object is at its (horizontal) center. Figure 1 shows four examples of images from São Paulo city and four examples of images from Porto Alegre city. Notice that sometimes the target objects may not be visible due to occlusion, or they may not be there when the picture was taken.

## 2.2 Dataset labelling

The labeling problem can be seen as multiclass, and the strategy to label each image is to use a vector of three binary values, one to represent whether the image has a tree or not, one to represent whether the image has power lines or not, and one to represent whether the image has an entanglement between the tree and the power lines. For example, if an image contains a tree, overhead power lines, and no entanglements, the binary vector associated with this image will be  $(1, 1, 0)$ . We developed an open source application named ’Street Level Imagery Labeler (SLIL)’ (available at <https://github.com/arturandre/SLIL>) to mitigate the manual labeling of the images. Figure 2 shows a screenshot of the SLIL application while annotating a GSV image with trees, overhead power lines, and an entanglement between them.

Labeling can be challenging for some of the images, primarily due to the camera’s perspective. Figure 3 shows an example for which it is hard to decide whether the top of the tree is touching or not the power lines. This occurs because the wires and the trees seem to be entangled at the top of the image. However, at the bottom of the image, it is possible to see that they are very far apart from each other, leading to the opposite conclusion. Figure 4 shows another example where the camera is pointing in a direction parallel to the power lines, and a tree below the power lines is very far from the camera. In such images, the power lines are no longer visible when they intersect distant trees. The default decision for all cases of doubt is that there is no entanglement. The rationale for this decision is because captured images are from an unconstrained urban environment, and they usually present trees and

wires far away in the distance. Assigning a positive label to such challenging instances could increase the number of positive misclassifications due to the disproportionate number of such instances.

Figure 5 shows the distribution of labels for São Paulo city (left) and Porto Alegre city (right) in the test dataset. The labels 1 and 0 indicate the presence or absence of the objects: 'Tree,' 'Electric wire' and 'Entanglement'. To compute the histograms (Fig. 5), we consider how many images are labeled with each of the classes present in the test dataset (i.e., Tree, Electric wire, and Entanglement) from each city separately. If an image is labeled with a given class, it counts as one positive case of that class in the histogram. The clear unbalance between the number of labels with 'Tree' in São Paulo city and 'Electric wire' in Porto Alegre city is due to how the images were collected.

### 2.3 Transfer learning

In the last few years, the advances in neural network architectures, mainly for Computer Vision, made image classification an achievable task (Krizhevsky et al., 2012). The pipeline to build an image classifier was to extract image features heuristically and train a classifier using these features. The usual approach now is to train a Convolutional Neural Network on a large dataset of images, usually not related to the problem to be solved, and use the network's features to solve another classification problem. This approach is known as *Transfer Learning* (Pan and Yang, 2009), and this section presents the network architecture and the transfer learning approach used in this work.

MobileNetV2 (Sandler et al., 2018) is a Convolution Neural Network model used to perform object classification, detection, or segmentation. The architecture is based on MobileNet (Howard et al., 2017), a simplified model adapted to mobile devices with several hardware constraints as smartphones. Figure 6 shows a block diagram of the MobileNetV2 architecture used in this work. The above part of the diagram shows the main pipeline of the network. The network starts with a triplet composed of a convolutional filter layer, followed by a batch normalization layer and a Rectified Linear Unit (ReLU6) activation layer. After the triplet, the architecture has a set of seventeen Inverted Residual blocks (the bottom part of Fig. 6) and another triplet similar to the first one. A global average pooling (Lin et al., 2013) and a final dense layer (classification head) with 1000 neurons finish the model. An implementation of MobileNetV2 is available in the Github repository (tensorflow, 2018). The repository also has a set of weights obtained through training on the ImageNet dataset (Deng et al., 2009).

To classify images with or without the entanglement of power lines and trees, we opted for a binary classifier. Therefore, we replace the final Global Average Pooling (GAP) and the one thousand neurons dense layer with another GAP layer and a different classification head with three fully connected layers with 1080, 540, and one neuron, respectively, as a transfer learning strategy. Figure 7 shows the proposed classification head, and Fig. 8 shows

an overview of the modification performed (concerning the MobileNetV2). All the last three fully connected layers are equipped with the sigmoid function as their activations. To finish the transfer learning approach, we fine-tuned the weights of the last three layers (i.e., all previous layers' weights are the same as in the repository) using the dataset proposed in this paper. With this approach, it is possible to take advantage of a much bigger dataset (ImageNet) without training the network from scratch (Pan and Yang, 2009).

A fine-tuning process was performed by training the network for 1000 epochs, and we selected the epoch with the best validation accuracy to be the final network rather than the last epoch. This strategy is a way to avoid over-fitting and having a better accuracy in images other than those from the training dataset. This strategy is similar to early stopping, in which a set of images, called validation dataset, is kept separated from the training dataset and used to estimate the accuracy of the network in new images. The reason to choose this strategy rather than early stopping is due to the stochastic behavior introduced by the mini-batches approach used during training due to memory restrictions. We observed that small mini-batches cause the loss function to become noisier, as demonstrated both theoretically and empirically by (Qian and Klabjan, 2020).

## 2.4 Network Results Interpretation

Interpreting the results of a network is an essential topic of research, and there are several methods proposed in the literature (Zhang and Zhu, 2018) to help with that. We use a generalization of the Gradient-weighted Class Activation Map (Grad-CAM++) (Chattopadhyay et al., 2017), an improved modification of Grad-CAM (Selvaraju et al., 2017), to interpret and qualitatively assess the results. The Grad-CAM++ method's idea is to build an image with a weighted sum of the positive partial derivatives of the output of the network (just before the last sigmoid activation) with relation to the last convolutional layer feature maps (activation map) to show which pixels are relevant to classify the image according to a specific target label defined by the user or by the assessment. The images generated by this method, or saliency maps, are color mapped from blue to red, indicating regions that are less to more relevant to the classification output. Figures 9a and 9b show an example of a tree entanglement with power lines and its Grad-CAM++ output. The network correctly classified the image as having an entanglement and the Grad-CAM++ method correctly points the position of the pixels in the entanglement (reddish points). Figures 9c and 9d show another example with no tree entanglement and its corresponding Grad-CAM++ output, respectively. In this case, the network did not classify the image as having an entanglement, but it correctly identified the relevant regions (in red) for the classification, namely the image regions with wires and tree canopies. To assess the results qualitatively, one can visually inspect the Grad-CAM++ method's output map for a test image and check if the classification is correct due to the expected visual clues or not.

### 3 Results

This section presents the results of the modified MobileNetV2 architecture.

#### 3.0.1 Training and validation

To train the network, 5091 images (with corresponding labels indicating entanglements) were used, 2484 images from the São Paulo city dataset, and 2607 from Porto Alegre (POA) dataset explained in Section 2. The training process uses continuously 80% of the training set effectively for training and 20% of the training set to validate and control the whole process. The network took around 340 epochs to converge, and the final validation accuracy was 81.94% (on the 20% of the training set) while the training accuracy was 96.64% (on the 80% of the training set).

Table 1 presents a confusion matrix with the number of correctly and incorrectly classified images. When the class assigned by the classifier (rows) agrees with the ground truth (columns), we say the classification is correct (top-left position of the table, or true positive; or bottom right corner of the table, or true negative). When the class assigned by the classifier disagrees with the ground truth, we say the classification is incorrect (top-right position of the table, or false positive; or bottom left corner of the table, or false negative). A row (column) with value 1 represents images classified (manually annotated) as positive cases. The value 0 in a row (column) represents images classified (manually annotated) as negative cases.

#### 3.0.2 Test datasets

From the 6092 images collected (see Sec. 2.1), 1001 images compose the test dataset. There are 504 images from Porto Alegre city and 497 images from São Paulo city. The test dataset contains no images used in the training process, and it is intended to assess the capability of generalization of the network.

Table 2 shows three confusion matrices for the overall test set, the São Paulo test set, and the Porto Alegre test set, respectively, from left to right. We obtained an overall test accuracy of 74.63%. For São Paulo the test accuracy was 73.64% and for Porto Alegre city, it was 75.60%. The interpretation of the labels is the same as in Tab. 1, and the results are similar to the ones observed earlier.

### 3.1 Qualitative analyses of the errors

This section presents a qualitative analysis based on misclassified images using Grad-CAM++ (Sec. 2.4). Through these analyses, we can empirically find directions to improve the classification results, potentially obtaining a better generalization, that is, better accuracy in the test dataset.

Figures 10a and 10b show a saliency map of a positive case. Figure 10a was classified as a true positive; the entanglement was correctly identified. We can visually in Fig. 10b that the true positive classification was essentially due to the sizeable reddish region containing wires crossing through the canopy of a tree. This is an empirical evidence that the network can correctly detect entanglements because it is tuned to detect the relevant image features that characterize an entanglement.

Visual inspection of several misclassified images and their corresponding saliency maps shows that some errors are caused by:

- the viewpoint of the camera (see Figs. 11a and 11b);
- wires that seems to meet with trees far away from the camera (see Figs. 12a and 12b); and
- bad lighting conditions (e.g. poor contrast) as exemplified in Figs. 13a, 13b, 14a and 14b)

## 4 Discussion

The detection of entanglements between power lines and trees is a challenging problem, and the presented solution is just a step towards solving it. A possible improvement we plan to do is to break the problem into smaller tasks in a divide and conquer strategy. For example, we first can detect pylons and power lines (Cheng and Song, 2008; Lopes et al., 2015). Then, we can quantify the number of trees and the amount of greenery in ground-level images (Wegner et al., 2016; Cai et al., 2018). Finally, the results of these two steps can be used together with the input images in a classifier for the entanglements, possibly improving our results. A multi-step approach like that can be interesting from an explainability perspective, as well. One could explore how each component of the combined influences the final classification results.

Nevertheless, such analyses are more cumbersome than using a single algorithm to perform all these tasks. Moreover, the optimization of each component in the multi-step approach is more complex. The first stage components (e.g., the wires detectors and vegetation quantification) can be optimized individually since their tasks are independent, but the final classifier depends on both the results of the first stage components and the input images. Thus the optimization of the last classifier would need to be carefully designed and possibly would imply a joint optimization with the previous components.

In our dataset, the data collected was based on geographical information containing trees and pylons, such that nearly all the images contain at least one of these elements. An algorithm to classify images with or without trees (power lines) would face an extreme imbalance in this dataset. In our method, we avoid multi-step optimization and imbalance issues by optimizing a single deep learning convolutional network directly over the entanglement task of the input image. Notice that while the dataset is unbalanced regarding trees, pylons, and power lines, it is balanced with entanglements.

Although our method uses Google Street View images, it can also be generalized to any street-level imagery dataset. That is, by using a set of cameras and a GPS logger on a vehicle, one can create a geo-localized street-level imagery dataset. The INACITY documentation (Oliveira and Hirata Jr., 2018) shows how to plug other image databases into the system. Besides, the Python code used to train the network and the datasets are also available at <https://github.com/arturandre/tree-wires>.

Dataset labeling poses one of the biggest challenges to this problem because sometimes it is hard to decide whether an image has an entanglement (entanglement) between a tree and the electric wires. The main difficulty is due to the camera’s point of view, but the projection from 3D to 2D can pose some other challenges. The SLIL software is in continuous improvement, and we plan to implement a feature to change the perspective and mitigate the decision disambiguation.

Finally, similar works that deal with the classification problem of tree and wire entanglements in above-ground images (i.e., satellite and airborne images (Kobayashi et al., 2009; Jwa et al., 2009)) complement the one presented here. Their work captures a distinct point of view that could help in ambiguous cases in which the perspective from ground-level images makes it hard to tell whether there is an entanglement. Conversely, ground-level images provide information about entanglements below trees’ canopies, occluded in above-ground images, and usually have a better resolution due to the proximity to the trees and power lines in the ground.

## 5 Conclusion

This paper presented a new method for detection of trees possibly interfering with wires, or overhead power-lines. Our method reached 74,6% accuracy in detecting entanglements between trees and wiring and can be a valuable tool. This new method can be coupled with the platform INACITY, and together they can be a valuable tool for better management of the vegetation, such as prevention of accidents, pruning, maintaining the health of the trees to reduce the risk of diseases, and insects in the trees.

The results obtained show that the MobileNetV2 modified architecture can be generalized to new images in cities which similar visual characteristics as those composing our dataset. Nevertheless, our approach is general enough, and the fine-tuning can be used to train a classifier in different datasets, even with images with different vegetation visual characteristics.

Our qualitative analyses based on Grad-CAM++ reveals that the most relevant parts of the images for the network are indeed regions predominated by vegetation and power lines. It also worth noting that even in images misclassified due to an ambiguity in the projection from the 3D world to 2D images, or due to poor image conditions, the relevant parts still being trees and power lines, according to the grad-CAM++ heatmap.

Concerning the network, the dataset was not constrained to consider only the top view of the images. Consequently, features near the image floor were analyzed similarly to the features at the top of the images. In the future, we plan to experiment with this idea to check if this may be a possible source of errors.

**Funding** This research is part of the INCT of the Future Internet for Smart Cities funded by CNPq proc. 465446/2014-0, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001, FAPESP proc. 14/50937-1, and FAPESP proc. 15/24485-9.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001 and in part by the grant #2018/10767-0, São Paulo Research Foundation (FAPESP). This version of the article has been accepted for publication, after peer review and is subject to Springer Nature’s AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: <https://doi.org/10.1007/s00468-022-02305-0>

**Availability of data and material** Data available at: <https://github.com/arturandre/tree-wires>

**Code availability** Code available at: <https://github.com/arturandre/tree-wires>

**Authors’ contributions** All the authors contributed equally to the idea and to the writing of the paper. The first author wrote all the software and made the experiments. The first and second authors defined the experimental protocol, dataset labeling and review of the results.

### **Conflicts of interest/Competing interests**

The authors declare that they have no conflict of interest.

---

**Tables**

		GT (Train)				GT (Validation)	
		1	0			1	0
Ours	1	1127	137	Ours	1	238	122
	0	0	2808		0	62	597

Table 1: The confusion matrix showing the distribution of the predictions for the training (with 4072 images) and validation datasets (with 1019 images). In the top left corner are the number of true positives; in the top right false positives; in the bottom left the false negatives and at the right bottom the true negatives.

		GT (Test all)				GT (Test SP)				GT (Test POA)	
		1	0			1	0			1	0
Ours	1	373	86	Ours	1	143	20	Ours	1	230	66
	0	166	376		0	109	225		0	57	151

Table 2: The confusion matrices showing the distribution of the predictions for the test dataset (1001) composed with 497 images from São Paulo (SP) and 504 from Porto Alegre (POA).



Fig. 1: Sample images from (top row) São Paulo and from (bottom row) Porto Alegre.

## Figures

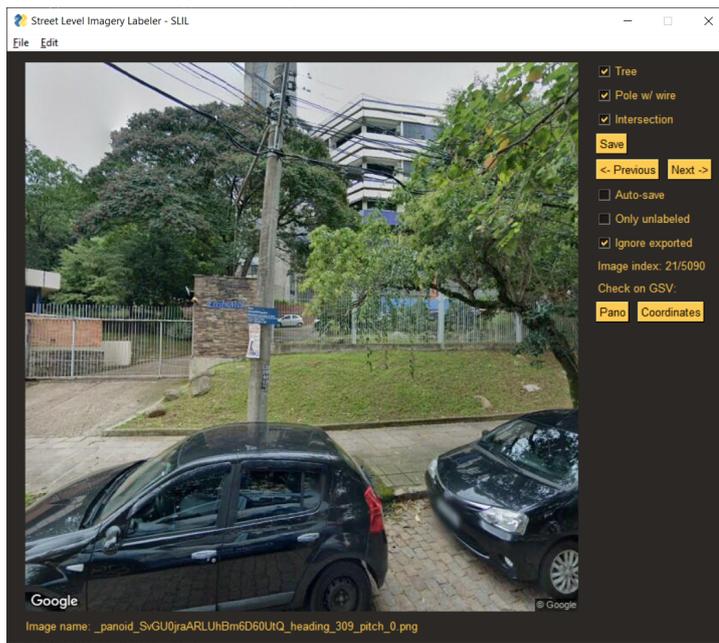


Fig. 2: Screenshot of the 'Street Level Imagery Labeler (SLIL)' application, used to annotate the GSV images in the dataset.

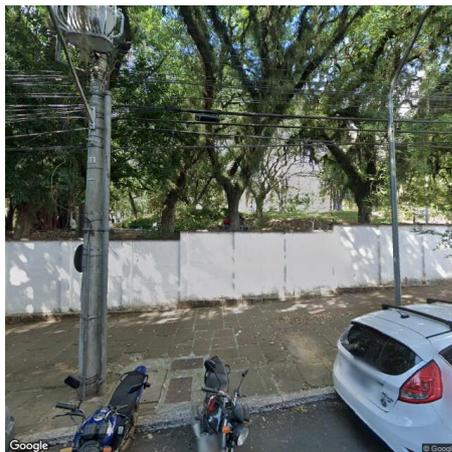


Fig. 3: Power lines in front of tree tops



Fig. 4: Power lines possibly intersecting with trees in the distance.

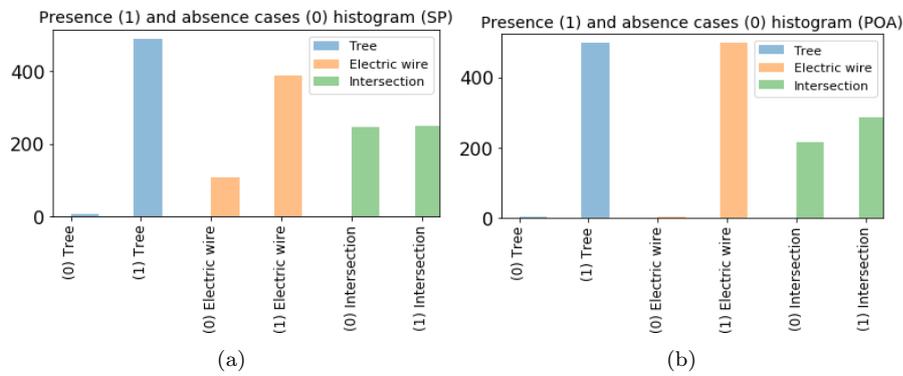


Fig. 5: Histogram of with the distribution of different classes for each test dataset in the images (a) from São Paulo and (b) from Porto Alegre.

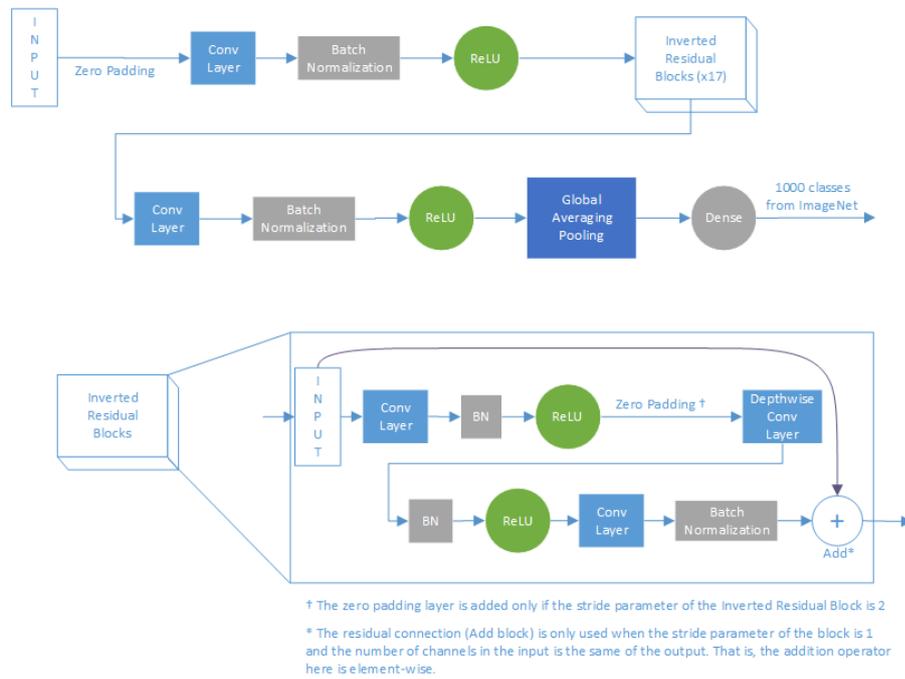


Fig. 6: MobileNetV2 architecture. The network has as output a vector representing of the probability distribution over the 1000 classes from the ImageNet used to pre-train the network.

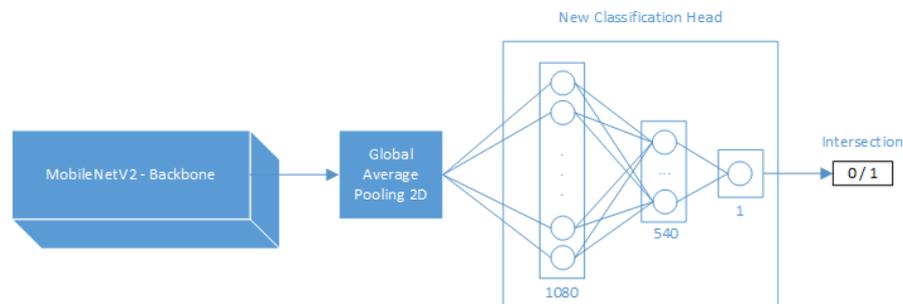


Fig. 7: New proposed classification head used on top of the MobileNetV2 backbone. We use two fully connected layers whose neurons have the sigmoid function as activation and a final layer with a single neuron and sigmoid activation function.

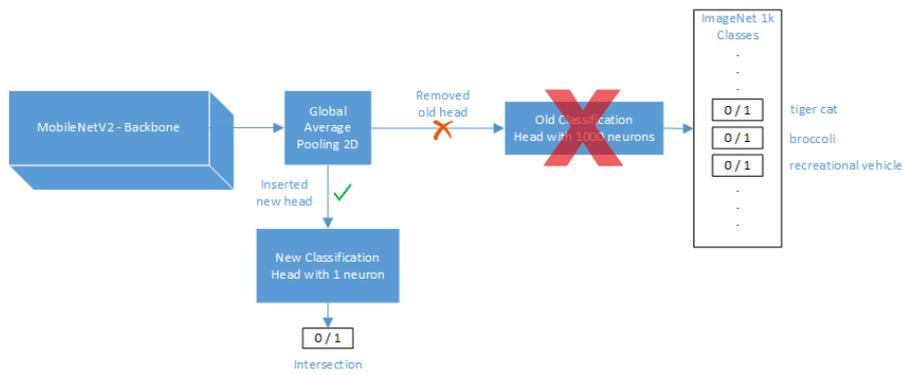


Fig. 8: MobileNetV2 architecture adaption. The network is adapted to output a single value instead of a probability distribution over the 1000 classes from the ImageNet used to pre-train the network.

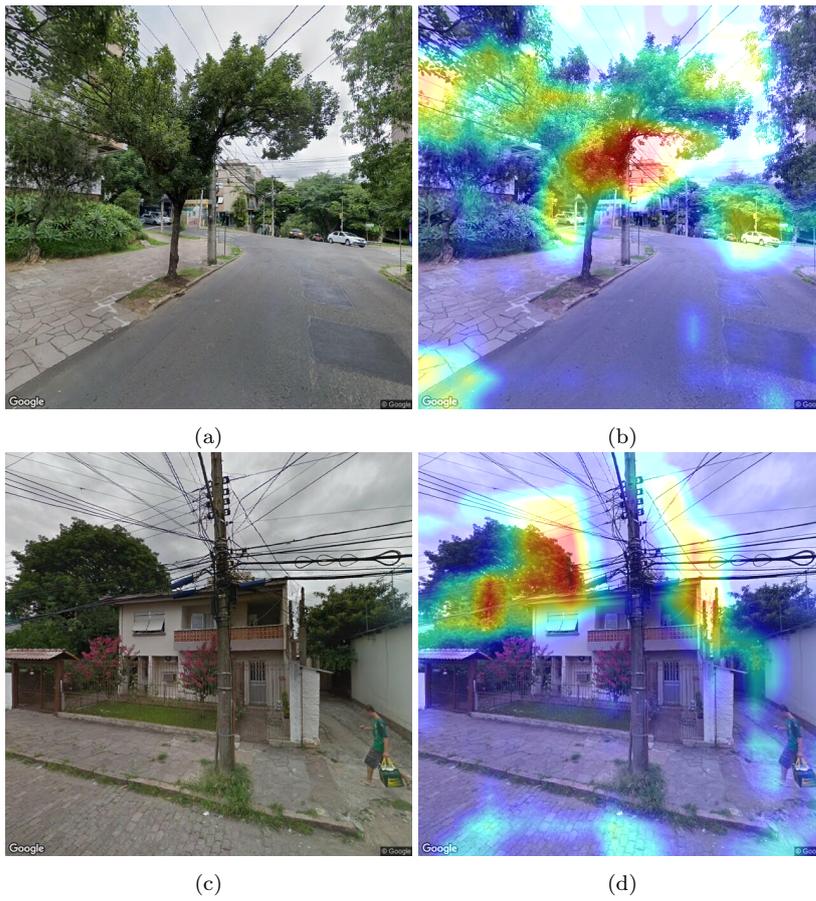


Fig. 9: (a): An entanglement between a tree and some overhead wires. (b): The Grad-CAM++ image with the dark reddish regions indicating the places relevant for the classification (positive) in this case. (c): An image without entanglements. (d): The corresponding Grad-CAM++ showing that the network correctly identified the regions in the image with trees and wires, even though they do not intersect with each other.

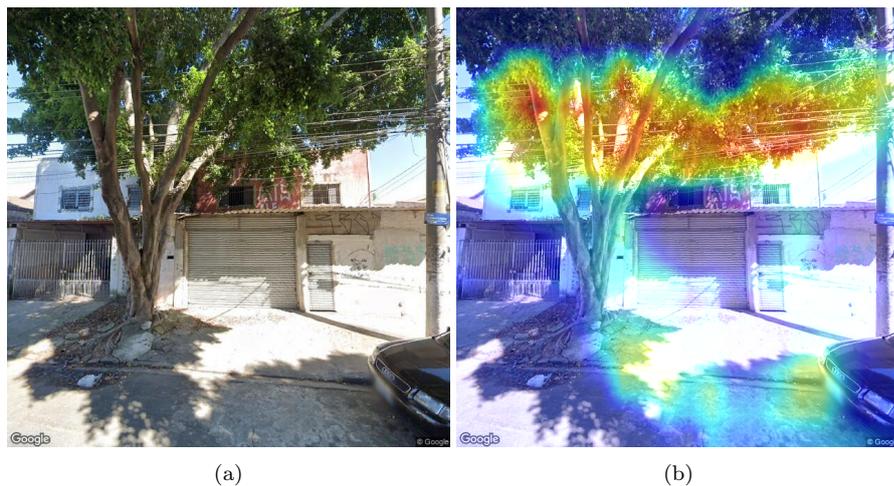


Fig. 10: (a): Image classified as a true positive. (b): The corresponding Grad-CAM++ image, showing that the whole extension of the wires passing through the canopy of the tree were identified and relevant for the classification.



Fig. 11: (a): An image misclassified as having an entanglement between tree and electric wires. Due to the position of the camera just below the foliage it seems that indeed there is an entanglement, but directly using Google Street View to navigate to another position reveals that the entanglement does not occur. (b): The salience map generated with the Grad-CAM++ method reveals that the main relevant regions (in red) observed by the network were the branches and the wires in the image.

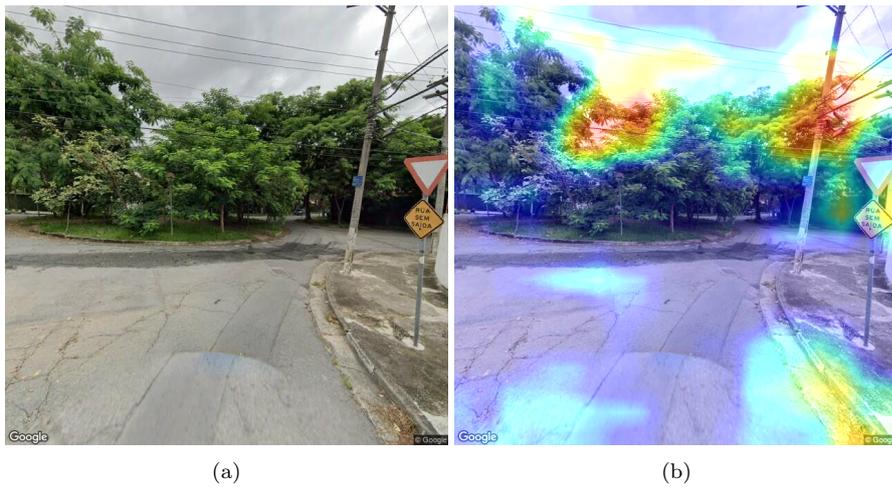


Fig. 12: (a): input image misclassified as a false positive. (b): Saliency map showing relevant regions with foliage of trees and overhead wires (in red). The wires do cross the branches of the trees on the image plane, but one can see that they are indeed apart from each other. The depth information is not available for the network during training so this kind of image is 'hard' to classify correctly.

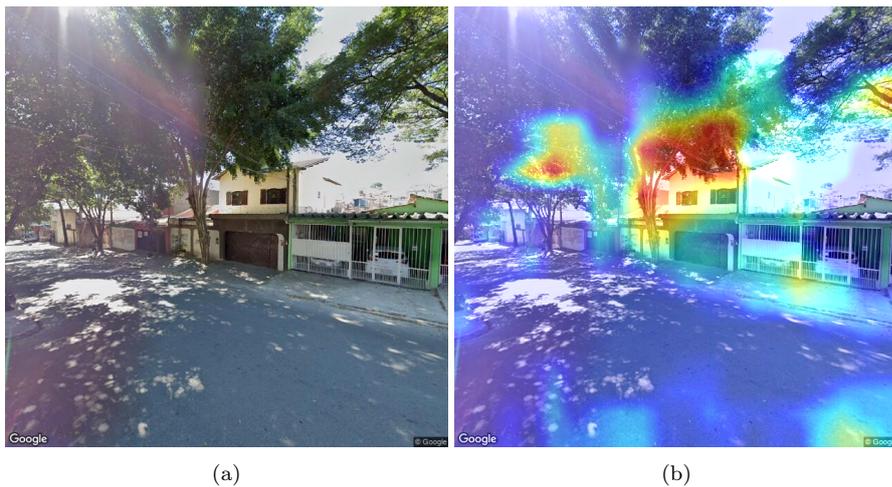


Fig. 13: (a): Input image misclassified by missing the entanglement, with a glare effect from the Sun. (b): The saliency map shows that the regions with the entanglement were indeed detected, but the classification was negative, possibly because the contrast between the wires and the trees just behind them is very poor due to the glare from the sunlight directly in the camera.

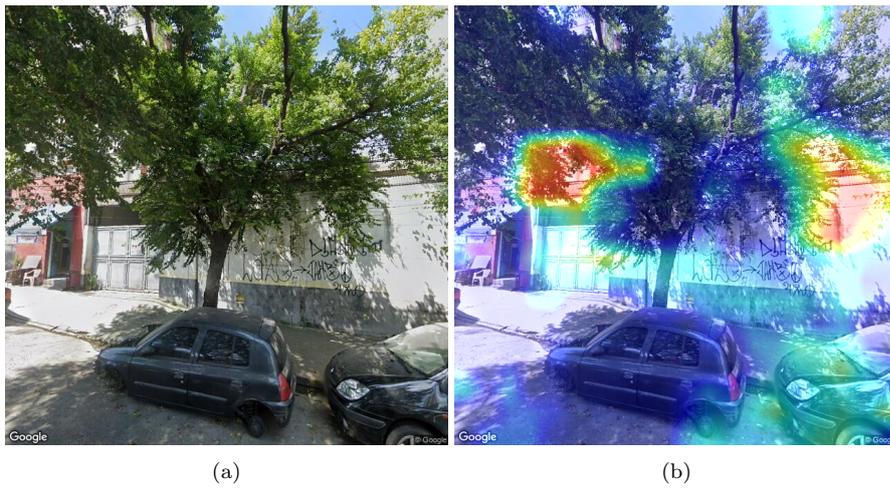


Fig. 14: (a): Input image misclassified by missing the entanglement of wires passing just under the shadow provided by the canopy. (b): The saliency map shows that the entanglement were detected, but similarly to the effect of the Sun glare, one can notice the low contrast caused by the shadow of the canopy over the wires.

## Legend to Figures

Fig. 1: Sample images from (top row) São Paulo and from (bottom row) Porto Alegre.

Fig. 2 Screenshot of the 'Street Level Imagery Labeler (SLIL)' application, used to annotate the GSV images in the dataset.

Fig. 3: Power lines in front of tree tops.

Fig. 4: Power lines possibly intersecting with trees in the distance.

Fig. 5 Histogram of with the distribution of different classes for each test dataset in the images (a) from São Paulo and (b) from Porto Alegre.

Fig. 6: MobileNetV2 architecture. The network has as output a vector representing of the probability distribution over the 1000 classes from the ImageNet used to pre-train the network.

Fig. 7: New proposed classification head used on top of the MobileNetV2 backbone. We use two fully connected layers whose neurons have the sigmoid function as activation and a final layer with a single neuron and sigmoid activation function.

Fig. 8: MobileNetV2 architecture adaption. The network is adapted to output a single value instead of a probability distribution over the 1000 classes from the ImageNet used to pre-train the network.

Fig. 9: (a): An entanglement between a tree and some overhead wires. (b): The Grad-CAM++ image with the dark reddish regions indicating the places relevant for the classification (positive) in this case.

(c): An image without entanglements. (d): The corresponding Grad-CAM++ showing that the network correctly identified the regions in the image with trees and wires, even though they do not intersect with each other.

Fig. 10: (a): Image classified as a true positive. (b): The corresponding Grad-CAM++ image, showing that the whole extension of the wires passing through the canopy of the tree were identified and relevant for the classification.

Fig. 11: (a): An image misclassified as having an entanglement between tree and electric wires. Due to the position of the camera just below the foliage it seems that indeed there is an entanglement, but directly using Google Street View to navigate to another position reveals that the entanglement does not occur. (b): The salience map generated with the Grad-CAM++ method reveals that the main relevant regions (in red) observed by the network were the

branches and the wires in the image.

Fig. 12: (a) input image misclassified as a false positive. (b): Saliency map showing relevant regions with foliage of trees and overhead wires (in red). The wires do cross the branches of the trees on the image plane, but one can see that they are indeed apart from each other. The depth information is not available for the network during training so this kind of image is 'hard' to classify correctly.

Fig. 13: (a): Input image misclassified by missing the entanglement, with a glare effect from the Sun. (b): The saliency map shows that the regions with the entanglement were indeed detected, but the classification was negative, possibly because the contrast between the wires and the trees just behind them is very poor due to the glare from the sunlight directly in the camera.

Fig. 14: (a): Input image misclassified by missing the entanglement of wires passing just under the shadow provided by the canopy. (b): The saliency map shows that the entanglement were detected, but similarly to the effect of the Sun glare, one can notice the low contrast caused by the shadow of the canopy over the wires.

## References

- Ahmad, J., Malik, A. S., Xia, L., and Ashikin, N. (2013). Vegetation encroachment monitoring for transmission lines right-of-ways: A survey. *Electric Power Systems Research*, 95:339–352.
- Angelov, D., Dulong, C., Filip, D., Frueh, C., Lafon, S., Lyon, R., Ogale, A., Vincent, L., and Weaver, J. (2010). Google Street View: Capturing the world at street level. *Computer*, 43(6):32–38.
- Bazaz, A., Bertoldi, P., Buckeridge, M., Cartwright, A., de Coninck, H., Engelbrecht, F., Jacob, D., Hourcade, J.-C., Klaus, I., de Kleijne, K., et al. (2018). Summary for urban policymakers: What the IPCC special report on global warming of 1.5°C means for cities.
- Berland, A. and Lange, D. A. (2017). Google Street View shows promise for virtual street tree surveys. *Urban Forestry & Urban Greening*, 21:11–15.
- Buckeridge, M. (2015). Árvores urbanas em São Paulo: planejamento, economia e água. *estudos avançados*, 29(84):85–101.
- Cai, B. Y., Li, X., Seiferling, I., and Ratti, C. (2018). Treepedia 2.0: applying deep learning for large-scale quantification of urban tree cover. In *2018 IEEE International Congress on Big Data (BigData Congress)*, pages 49–56. IEEE.
- CEEE, G. (2016). CEEE - Companhia Estadual de Energia Elétrica - Rio Grande do Sul. <https://web.archive.org/web/20191203114156/http://ceee.com.br/pportal/ceee/Component/Controller.aspx?CC=97626>. Access Jan 22nd, 2021.

- Chattopadhyay, A., Sarkar, A., Howlader, P., and Balasubramanian, V. N. (2017). Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. *CoRR*, abs/1710.11063.
- Cheng, W. and Song, Z. (2008). Power pole detection based on graph cut. In *2008 Congress on Image and Signal Processing*, volume 3, pages 720–724. IEEE.
- Clarke, D. J. and White, J. G. (2008). Towards ecological management of Australian powerline corridor vegetation. *Landscape and urban planning*, 86(3-4):257–266.
- Clode, S. and Rottensteiner, F. (2005). Classification of trees and powerlines from medium resolution airborne laserscanner data in urban environments. In *Proceedings of the APRS Workshop on Digital Image Computing (WDIC), Brisbane, Australia*, volume 21.
- de Coninck, H., Revi, A., Babiker, M., Bertoldi, P., Buckeridge, M., Cartwright, A., Dong, W., Ford, J., Fuss, S., Hourcade, J.-C., et al. (2018). Strengthening and implementing the global response.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*.
- Escobedo, F. J., Kroeger, T., and Wagner, J. E. (2011). Urban forests and pollution mitigation: Analyzing ecosystem services and disservices. *Environmental Pollution*, 159(8):2078–2087. Selected papers from the conference Urban Environmental Pollution: Overcoming Obstacles to Sustainability and Quality of Life (UEP2010), 20-23 June 2010, Boston, USA.
- Fakes, J. (2000). Practical issues in line clearance and street trees. *TREENET*, page 39.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press. <http://www.deeplearningbook.org>.
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications.
- Jwa, Y., Sohn, G., and Kim, H. (2009). Automatic 3D powerline reconstruction using airborne lidar data. *Int. Arch. Photogramm. Remote Sens*, 38(Part 3):W8.
- Kenney, W. A., van Wassenaeer, P. J., Satel, A. L., et al. (2011). Criteria and indicators for strategic urban forest planning and management. *Arboriculture & Urban Forestry*, 37(3):108–117.
- Kobayashi, Y., Karady, G. G., Heydt, G. T., and Olsen, R. G. (2009). The utilization of satellite images to identify trees endangering transmission lines. *IEEE Transactions on Power Delivery*, 24(3):1703–1709.
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105.
- Lin, M., Chen, Q., and Yan, S. (2013). Network in network.
- Livesley, S. J., McPherson, E. G., and Calfapietra, C. (2016). The urban forest and ecosystem services: Impacts on urban water, heat, and pollution cycles at the tree, street, and city scale. *Journal of Environmental Quality*,

- 45(1):119–124.
- Locosselli, G. M., de Camargo, E. P., Moreira, T. C. L., Todesco, E., de Fátima Andrade, M., de André, C. D. S., de André, P. A., Singer, J. M., Ferreira, L. S., Saldiva, P. H. N., et al. (2019). The role of air pollution and climate on the growth of urban trees. *Science of The Total Environment*, 666:652–661.
- Locosselli, G. M., Moreira, T. C. L., Chacón-Madrid, K., Arruda, M. A. Z., de Camargo, E. P., Kamigauti, L. Y., da Ferreira Trindade, R. I., de Fátima Andrade, M., de André, C. D. S., de André, P. A., et al. (2020). Spatial-temporal variability of metal pollution across an industrial district, evidencing the environmental inequality in são paulo. *Environmental Pollution*, page 114583.
- Lopes, A. K., Soares, F. A., Lopes, L. C., Laureano, G. T., Oliveira, L. L., Costa, R. M., and Soares, A. S. (2015). Segmentação de postes da rede elétrica a partir de imagen do Google Street View. *Anais do XII Simpósio Brasileiro de Automação Inteligente. Natal-RN: Sociedade Brasileira de Automação*.
- Lüttge, U. and Buckeridge, M. (2020). Trees: structure and function and the challenges of urbanization. *Trees*.
- Ma, J., Cheng, J. C., Jiang, F., Gan, V. J., Wang, M., and Zhai, C. (2020). Real-time detection of wildfire risk caused by powerline vegetation faults using advanced machine learning techniques. *Advanced Engineering Informatics*, 44:101070.
- Most, W. B. and Weissman, S. (2012). Trees and power lines: Minimizing conflicts between electric power. Access Jan 22nd, 2021.
- Neuhold, G., Ollmann, T., Rota Bulò, S., and Kontschieder, P. (2017). The mapillary vistas dataset for semantic understanding of street scenes. In *International Conference on Computer Vision (ICCV)*.
- Oliveira, A. A. A. d. M. and Hirata Jr., R. (2018). Inacity’s github repository. <https://github.com/arturandre/INACITY>. Access Jan 22nd, 2021.
- Oliveira, A. A. A. d. M. and Hirata Jr., R. (2021). Inacity - investigate and analyze a city. *SoftwareX*, 15:100777.
- Ordóñez, C. and Duinker, P. N. (2013). An analysis of urban forest management plans in Canada: Implications for urban forest management. *Landscape and Urban Planning*, 116:36–47.
- Pakonen, P. (2008). Characteristics of partial discharges caused by trees in contact with covered conductor lines. *IEEE Transactions on Dielectrics and Electrical Insulation*, 15(6):1626–1633.
- Pan, S. J. and Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359.
- Prefeitura de São Paulo (2021). Mapa digital da cidade de São Paulo. [http://geosampa.prefeitura.sp.gov.br/PaginasPublicas/\\_SBC.aspx](http://geosampa.prefeitura.sp.gov.br/PaginasPublicas/_SBC.aspx). Access Jan 22nd, 2021.
- Purcell, L. (2015). Trees and electric lines. <https://extension.purdue.edu/extmedia/FNR/FNR-512-W.pdf>. Access Jan 22nd, 2021.

- Qian, X. and Klabjan, D. (2020). The impact of the mini-batch size on the variance of gradients in stochastic gradient descent. *arXiv preprint arXiv:2004.13146*.
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520.
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626.
- tensorflow (2018). tensorflow/mobilenet\_v2.py at github. [https://github.com/tensorflow/tensorflow/blob/master/tensorflow/python/keras/applications/mobilenet\\_v2.py](https://github.com/tensorflow/tensorflow/blob/master/tensorflow/python/keras/applications/mobilenet_v2.py). Access Jan 22nd, 2021.
- ©Grab and Contributors, O. (2019). Openstreetcam. <https://openstreetcam.org/>. Access Jan 22nd, 2021.
- Wegner, J. D., Branson, S., Hall, D., Schindler, K., and Perona, P. (2016). Cataloging public objects using aerial and street-level images-urban trees. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6014–6023.
- Yao, W. and Fan, H. (2013). Automated detection of 3D individual trees along urban road corridors by mobile laser scanning systems. In *Proceedings of the International Symposium on Mobile Mapping Technology, Tainan, Taiwan*, pages 1–3.
- Zhang, Q. and Zhu, S. (2018). Visual interpretability for deep learning: a survey. *Frontiers of Information Technology & Electronic Engineering*, 19(1):27–39.