STN PLAD: A Dataset for Multi-Size Power Line Assets Detection in High-Resolution UAV Images

André Luiz Buarque Vieira-e-Silva*, Heitor de Castro Felix*, Thiago de Menezes Chaves*,
Francisco Paulo Magalhães Simões*†, Veronica Teichrieb*, Michel Mozinho dos Santos‡,
Hemir da Cunha Santiago‡, Virginia Adélia Cordeiro Sgotti‡, and Henrique Baptista Duffles Teixeira Lott Neto§

*Voxar Labs, Centro de Informática, Universidade Federal de Pernambuco, Recife, Brazil
{albvs,hcf2,tmc2,vt}@cin.ufpe.br
†Departamento de Computação, Universidade Federal Rural de Pernambuco, Recife, Brazil
francisco.simoes@ufrpe.br
‡In Forma Software, Recife, Brazil
{mmozinho,hsantiago}@informasoftware.com.br, vsgotti@informa.com.br
§Sistema de Transmissão Nordeste - STN, Recife, Brazil
hlott@stnordeste.com.br

Abstract—Many power line companies are using UAVs to perform their inspection processes instead of putting their workers at risk by making them climb high voltage power line towers, for instance. A crucial task for the inspection is to detect and classify assets in the power transmission lines. However, public data related to power line assets are scarce, preventing a faster evolution of this area. This work proposes the STN Power Line Assets Dataset, containing high-resolution and real-world images of multiple high-voltage power line components. It has 2,409 annotated objects divided into five classes: transmission tower, insulator, spacer, tower plate, and Stockbridge damper, which vary in size (resolution), orientation, illumination, angulation, and background. This work also presents an evaluation with popular deep object detection methods and MS-PAD, a new pipeline for detecting power line assets in hi-res UAV images. The latter outperforms the other methods achieving 89.2% mAP, showing considerable room for improvement. The STN PLAD dataset is publicly available at https://github.com/andreluizbvs/PLAD.

Fig. 1. A few image clippings from the proposed dataset.

I. INTRODUCTION

Nowadays, practically all human activities depend on the constant availability of electricity. Power transmission lines, which have an essential role in this task, are constantly exposed to the depreciating action of the environment. They have components that may break, rust, loosen or even go missing. The malfunction of such equipment affects the electricity grid, causing inefficiency in the power transmission and, sometimes, blackouts. According to [1], most of the power grids today are interconnected. Thus, these blackouts can initiate others, affecting even larger regions, like a cascade effect [2]. That can trigger catastrophic consequences such as shutting down hospitals, production at water supplies companies, and telecommunication services [3], which leads to significant economic losses for the energy company and, ultimately, severe social impacts [4], [5]. According to Bruch et al. [4], a power cut of only 30 minutes in the USA results in an average loss of over 15 thousand US dollars for midsize and large industrial clients and a loss of more than 90 thousand US dollars for an eight-hour interruption [1].

Given this scenario, constant maintenance of the equipment is necessary, replacing defective ones before they can cause any loss. Classically, a team has to check each station personally. That implies having someone climbing the transmission tower and checking the condition of its components, which is dangerous and time-consuming [6]. In other words, there is a cost to move people, and it takes time to get up at a station, apart from safety issues associated with the service. For example, Rahmani et al. [7] showed that there were 119 injured workers due to accidents between 2006 and 2012 in an Iranian electricity distribution company, with seven deaths.

Recent developments in computer vision can be applied in the field of Maintenance & Inspection, improving security and productivity. As an example of application in the security area, automatically detecting power line assets via UAVs eliminates much of the safety risks of manual inspection, as inspectors would not need to climb towers as often as before. Instead, much of the inspection could occur through a direct analysis of the detected components, which a human inspector could perform in a secure environment or by a fault classification method [8]. In addition to the security gains, there are also

financial and time gains, as the frequency of moving teams and providing the necessary equipment would decrease.

The dataset plays a major role in the training of deep learning networks, where its quality directly influences the accuracy. That is why we should pay more attention to data, as stated by [9]. Furthermore, public datasets play a fundamental role in a rapidly advancing area. They allow researchers to propose ideas and perform experiments even in scenarios where they cannot get the data by themselves. Also, those datasets usually serve as a benchmark for a specific task, providing a fair comparison among new techniques.

It is evident how quickly certain areas of computer vision evolved after the introduction of datasets such as CIFAR10 [10], ImageNet [11] and MNIST [12]. In that sense, public datasets on power line assets for object detection are extremely scarce, and the existing ones are quite limited, typically only supporting one asset type or two, at most [13]–[15]. That happens because most of the works in the area are privately funded by companies that want to maintain a competitive advantage by not making their datasets available.

As the main contribution, this paper introduces a new real-world, high-resolution, and multi-category dataset for multi-size power line assets recognition, the Power Line Assets Dataset, or STN PLAD. It serves as publicly available development data and benchmark for the computer vision community working on automatic power line inspection. In addition, experiments with state-of-the-art techniques show the dataset's strengths and limitations. These experiments used two popular general-purpose object detectors, namely SSD and Faster R-CNN. Based on its analysis, a variation of the training pipeline, which we call MS-PAD, is proposed to improve the overall object detection performance in the STN PLAD dataset.

This paper is organized as follows. The prior works are presented in section II. The properties of the new STN Power Line Assets Dataset are described in section III. section IV shows the methods used to evaluate the proposed dataset. Next, comparative and performance results of techniques applied in STN PLAD are presented in section V, followed by a discussion about what was seen in the tests in section VI and, lastly, the final remarks in section VII.

II. RELATED WORKS

This section is divided into two parts. First, the public datasets closely related to power lines are presented, along with their characteristics and limitations. Then, the existing methods that attempt to detect multi-size power line assets in high-resolution images are shown.

A. Public datasets related to power lines

A common problem found in the literature when using Deep Learning to detect power line objects is finding data. There are not enough publicly available datasets to feed detectors based on deep learning methods, or they do not cover enough power line components [1], [16]. Many similar works use private datasets, generally provided by the companies or government agencies financing the projects, which tend not to publish them

[1], [17]–[20]. Nevertheless, in the literature search presented by Liu et al. [16] and in the work of Abdelfattah et al. [15], a few publicly available datasets were found but with many limitations. Table I summarizes the main public image datasets related to power lines assets for the object detection task. The last line shows the proposed dataset for comparison purposes.

As evidenced in Table I, there is a minimal amount of public datasets related to power line asset detection. They target distinct tasks that can be detection, classification, or segmentation. For instance, the one in [21] is specifically related to conductor wires in low-resolution images for binary classification. Zhang et al. [22] propose two datasets of binarized masks of conductor wires of power lines in urban and mountain scenarios, respectively. Abdelfattah et al. [15] propose a dataset containing pixel-wise annotation (a.k.a. instance segmentation) of both transmission towers and power lines. However, the main competing datasets are CPLID [13], and the one from Tomaszewski et al. [14] as they are the only ones that use bounding box annotations.

CPLID [13] is a dataset related only to a specific type of insulator with a specific shape and size. Although it also has annotations of defects in some of those insulators, it lacks a diversity of data since there is only one type of power line asset. The defective samples are also limited because all of them are from data augmentation, i.e., a single faulty insulator was cropped from an image and then pasted into a limited set of backgrounds, like seen in Figure 2(a). On the other hand, STN PLAD provides asset variability in diverse scenarios.

The dataset in Tomaszewski et al. [14] has even more limitations. They mainly target data quantity (they reported 2630 images) rather than data variability, as they video recorded a ceramic long rod insulator hanging on an apparatus built by them and then extracted some of the frames using a stationary camera. These images only contain one of nine different backgrounds that do not correspond to real-world power line scenarios. An image from one of these nine variations is shown in Figure 2(b). From the perspective of deep learning techniques, this dataset has a very limited variability of scenes. In summary, although this dataset has a reasonable amount of data, the images taken in the same scenario have a high degree of similarity. Thus, the dataset ends up not being independent and identically distributed (IID) [23], [24], an essential dataset property. In practical terms, it is not feasible to use it in most techniques based on deep learning since it poses little to no challenge to these techniques. In STN PLAD, the images are collected by a drone in the field, providing several real-world scenarios with multiple objects (size, appearance, position, orientation, self-occlusion, background).

B. Detection of power line assets in high-resolution images

A few works address the issue of detecting objects related to power lines in high-resolution images. The first one, Zhang et al. [19], is a study on object detection in high-resolution images captured through Unmanned Aerial Vehicles (UAVs), using deep learning techniques. In their work, the authors propose the MOHR dataset. This private dataset has over ten

Dataset	#Assets	Instances/image (average)	Image size	Instances	Images	Background variation
CPLID [13]	1	1.9	1152×864	1569	848	Limited
Tomaszewski et al. [14]	1	1	5616×3744	2630[a]	2630[b]	Very limited
STN PLAD	**5**	**18.1**	**5472×3078 or 5472×3648**	**2409**	**133**	**Diverse**

[a] All the instances correspond to the same object.
[b] Images from this dataset are extremely similar and captured from just nine different points of view.

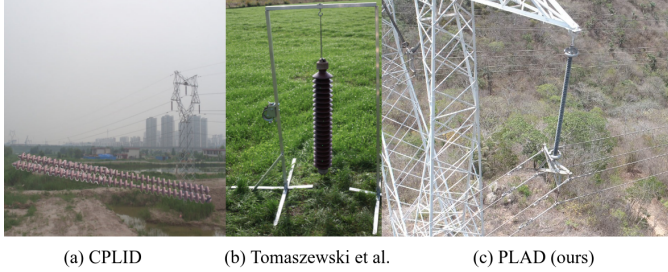(a) CPLID   (b) Tomaszewski et al.   (c) PLAD (ours)

Fig. 2. Sample images from public datasets as compared to our dataset STN PLAD in which (a) CPLID [13] with an asset inserted in a background, (b) [14] with an asset in a non-real-world scenario, and (c) our dataset (STN PLAD) with multi-objects in real-world scenes (zoomed in image).

thousand high-resolution UAV images with five classes: car, truck, building, collapse, and flood damage. The UAV altitudes are high, ranging from 200 to 400 meters, making the objects look quite small. The authors apply six general-purpose object detectors, SSD [25] and Faster R-CNN [26] included, to the MOHR dataset. The results suggest that detecting small object instances in high-resolution UAV images remains challenging since they perform poorly. The best mean Average Precision (mAP) achieved was 43.94%, yielded by RFCN-DF [27]. Those results reinforce the relevance of this task.

The works of Kong et al. [28] and Zhu et al. [29] share some of their authors and contents, indicating that one is an incremental improvement over the other. The former proposes a technique to detect small objects in high-resolution images. The technique, based on Faster R-CNN, is tested on a private dataset of 3700 high-resolution images. However, the proposed approach here is limited and prone to issues since only small objects inside the context of a large one are attainable. For instance, dampers are usually small independent objects far from larger ones. Another issue is the low Average Precision (AP) of some classes, such as the tower plate (73.2%), which the authors justify by saying they are too small.

Finally, Zhu et al. [29] attempts to improve the efficiency of their previous work by merging two stages in order to share early convolutional layers. They also use a private dataset with high-resolution images with six classes: electric tower, vibration damper, spacer, insulator, bird's nest, and tower plate. Despite the efficiency of this method, the same issues and limitations that existed before regarding object detection are maintained. The only difference is that the objects are not so small as their previous work, which is one of the main factors that positively impact the mAP.

These last two works, [28] and [29], are not reproducible since the datasets are private, and they are not open-source.

III. STN PLAD: DATASET DESCRIPTION

The images were captured using a DJI Phantom 4 Pro [1], and Figure 1 shows some image clippings. A set of policies for data collection was proposed to ensure data variability and consistency. First, the drone was handled by certified drone pilots, who were instructed to capture the images, always maintaining a similar distance to the transmission tower in a wide shot due to the high-resolution nature of the camera. In addition, the drone's viewing angles were varied to ensure better learning by models based on neural networks and diverse daytime, weather, angulation, and illumination conditions. Finally, several transmission towers were captured to obtain background and component variation. This data capture protocol provides images with a wide range and number of power line assets in each one of them, with a mean of 18.1 instances per captured image as can be seen in Table I .

The equipped camera is a DJI FC6310 and it can take pictures with a resolution of 5472×3078 (3:2) or 5472×3648 (16:9). Both aspect ratios were used during data collection. For annotating the 2409 objects in all 133 captured images, the LabelImg tool was used [30]. Two annotators were responsible for carefully surrounding each object with a bounding box. Each person took, on average, 10 minutes to annotate one image. Each image is assigned to only one annotator to perform all its annotations. To maintain the annotation consistency between different annotators, they labeled each assigned image with their highest possible scrutiny and were in touch during the entire annotation process.

The total amount of images in STN PLAD may appear small but, considering the employed data collection protocol, the camera's resolution, and, more importantly, the total amount of object instances, it can be seen that it has a reasonable amount of data. Images from STN PLAD have considerably more information than regular images from common datasets, such as ImageNet [11] and MSCOCO [31]. On average, the STN PLAD has more than 18 objects per image with an average area of at least 2.89×10^3 pixels. This 18 objects/image density is way above the related datasets, as seen in Table I. Finally, the STN Power Line Assets Dataset is publicly available in https://github.com/andreluizbvs/PLAD [2].

[1] https://www.dji.com/phantom-4-pro
[2] In case the article is accepted, the dataset will be posted on a web page with a structured presentation.

TABLE II
STN PLAD statistics.

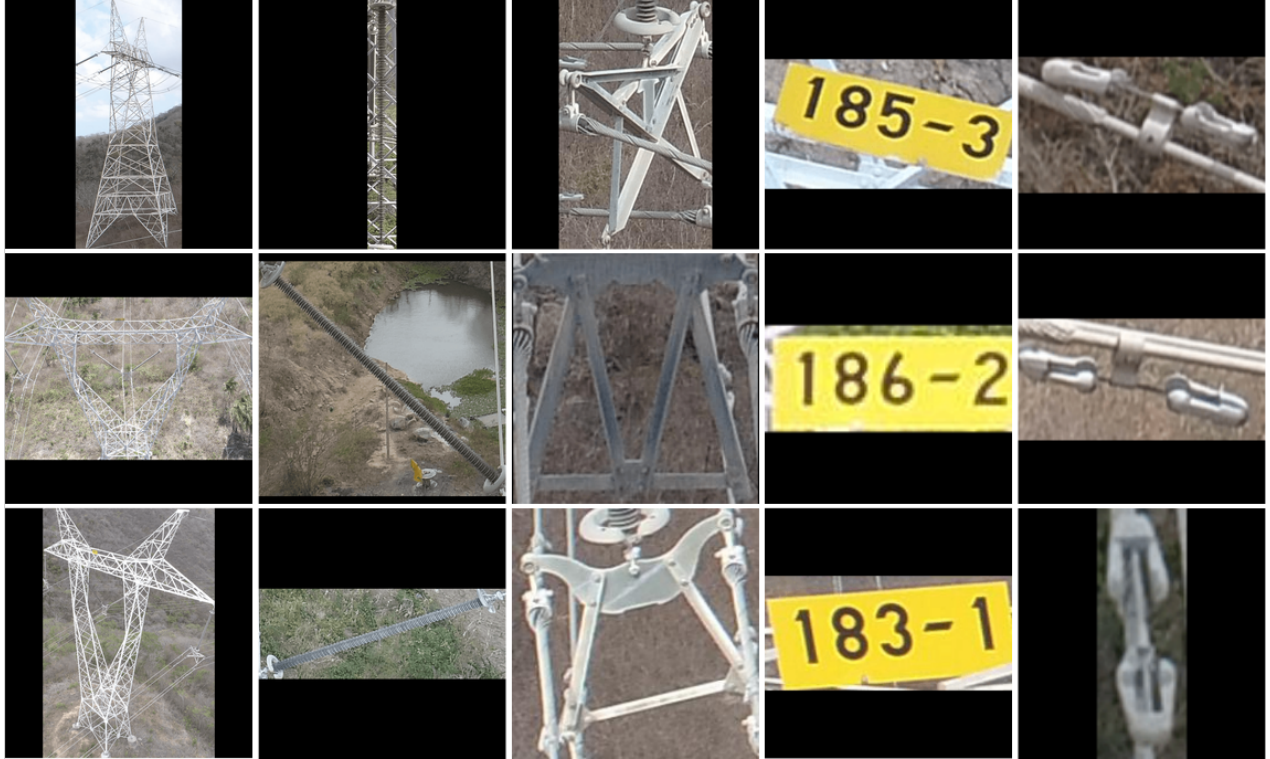| Class name	Label	Instances	Instances per image	Average Area (px)	Standard Deviation (px)
Transmission tower	*tower*	253	1.9	2.61×10^6	3.12×10^6
Insulator	*insulator*	312	2.3	8.84×10^4	8.55×10^4
Spacer	*spacer*	253	1.9	2.82×10^4	2.41×10^4
Tower plate	*plate*	86	0.6	9.42×10^3	1.11×10^4
Stockbridge damper	*damper*	1505	11.3	2.89×10^3	5.78×10^3

Fig. 3. Examples of all five classes of power line assets in STN PLAD. Each column shows instances from one class. From left to right: Transmission tower, Insulator, Spacer, Tower plate, and Stockbridge damper.

IV. METHODS

This section describes two techniques that are often used to validate object detection datasets [13], SSD and Faster R-CNN. Their performance on STN PLAD is presented in the next section and discussed later. The observed limitations in dealing with the proposed dataset inspired the creation of a pipeline called MS-PAD, which is also detailed here.

A. Single Shot MultiBox Detector (SSD)

In the context of power line inspection, SSD is one of the suggested techniques of two recent reviews [1], [16] to target the problem of detecting assets on power transmission towers. Both reviews have the same context as this work, focusing on inspecting power line assets from UAV images. Moreover, they analyze Deep Learning techniques applied to solve problems in the area. Some of the mentioned problems are assets detection, assets segmentation, assets fault identification.

The parameters of the SSD used are the same as in the original work by Liu et al. [25], such as the backbone, VGG16,

and all the parameters and dimensions for the convolutional layers. In the original work, two different input layers were proposed, 300×300 and 512×512. In our experiment, the latter was used since it achieved better accuracy than the former, according to the original results. Also, the images used for this experiment have a higher resolution. This high-resolution implies that the larger the size of the input layer, the less the resizing effect will affect the input image quality. Finally, weights pre-trained with the COCO Dataset [31] were used.

B. Faster R-CNN

The objective of including the Faster R-CNN [26] in the tests was to use a recent technique of object detection to obtain results close to the current state of the art. Faster R-CNN-based networks are also suggested to detect and inspect power transmission towers according to the same reviews mentioned in subsection IV-A. They are also well consolidated, have performed well in object detection competitions, and are used by similar works [1], [16].

The network used for this experiment was the Feature Pyramid Network (FPN) Faster R-CNN [32]. FPN aims to improve the detection of small objects, as it uses multi-scale feature maps and higher resolution layers to build new semantically rich layers. Thus, information from the initial layers is used. These layers are traditionally less condensed, but even so, they already have a high semantic level. ResNet-101 was used as a backbone, which had the best detection result in its publication [32]. Also, the input resolution was chosen in order to decrease the image resizing impact. The input image is resized to 2736×1824, representing a downscaling factor of 4 when compared to the original size, which is much less than the downscaling factor of approximately 19 used in the SSD experiment. All other parameters were kept as the original.

C. MS-PAD

After observing the results related to the SSD and Faster R-CNN methods, it was noticed that a simple pipeline modification could enhance the overall performance of power line assets detection in STN PLAD. This approach takes advantage of the images' high resolution, where information is lost after resizing. In the Multi-Size Power line Asset Detection (MS-PAD) workflow, represented in Figure 4, two independent networks are trained separately. The SSD was chosen because it performed better, as it will be shown in the next section.

The first of these two networks uses the original pipeline that resizes its input, but this time trained without the Stockbridge damper class. The damper asset was excluded because it has a much smaller size, being harder to identify after image resizing. This strategy can be applied to other assets not contemplated in this work that are small. An important note is that, although the tower plate is also small, it still had enough features after resizing that made it highly recognizable.

The second network is responsible for detecting small objects, in this case, the Stockbridge damper class. It goes through a different process, where the initial image is split following a grid. The image is divided into 16 smaller ones with a fixed resolution of 1368×769 or 1368×912, depending on the original image, which can be 5472×3078 (3:2) or 5472×3648 (16:9). This 4×4 division is constant since the drone pilots followed a data collection protocol, in which the drone had to stay at similar distances from the transmission towers, as described in section III.

In summary, only the Stockbridge damper class is submitted to the second step of MS-PAD, which divides the high-resolution image in a 4×4 grid. This choice is based on the average area of the classes of Table II and the AP in the original image-resizing approach. The Stockbridge damper has the lowest average area and was not well detected using the previously mentioned methods, indicating the need to receive extra attention compared to other classes.

V. EXPERIMENTAL RESULTS

This section presents the two conducted experiments and their results. The first one is responsible for comparing the performance of the two mentioned object detectors in the
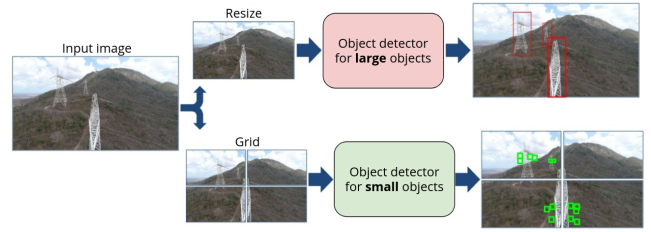

Fig. 4. MS-PAD pipeline. The input image is resized for the first network. Meanwhile, the image is divided into a grid, resulting in the second network's inputs that generate different bounding boxes.

proposed dataset. The second one demonstrates another way to deal with the input data, using MS-PAD, which was detailed in subsection IV-C. In all experiments, the standard metric of Average Precision (AP) is used to evaluate object detection performance on STN PLAD. In order to validate and obtain a greater degree of confidence in the results of the proposed MS-PAD, the Monte Carlo cross-validation method [33] was chosen and implemented in its experiments presented in subsection V-B. This method creates k random splits of train and test sets of the whole dataset. Then, the model is trained and tested for each k split, and the final result is the average. In the end, this section shows which object detector and which pipeline present the best results in the described scenario according to the considered metric.

For the experiments, STN PLAD was split in a standard 80/20 proportion for the training and test sets, respectively. Also, to consider that an object was correctly detected, the Intersection over Union (IoU) between the ground truth and the predicted bounding box had to be equal to or larger than 0.5. Also, data augmentation is already an embedded stage in both implementations of the techniques. Finally, the experiments were performed on a desktop running the Ubuntu 18.04 Operating System, powered by an Nvidia RTX 2080 Ti GPU (11 GB of VRAM) and an Intel Core i7 - 4790K CPU @ 4.00 GHz with 32 GB of available RAM.

A. SSD and Faster R-CNN results

For this test, both detectors were trained once and for the same period, about two days. The mAP results of using the MS-PAD approach for SSD and Faster R-CNN were 90.2% and 88.6%, respectively, showing that SSD has a slight advantage over Faster R-CNN. These methods were also applied to the two main dataset competitors of the proposed STN PLAD. In CPLID [13], SSD and Faster R-CNN achieved 98.17% and 98.31% mAP, respectively. Regarding Tomaszewski et al. [14], both detectors reached 100% mAP.

Figure 5 and Figure 6 show the visual results regarding the detected objects by the SSD and the Faster R-CNN methods, respectively, using the original approach, which only resizes the input image. In the images, the bounding boxes' colors are connected with the dataset classes: blue is for the Insulators; yellow is for the Spacers; green is for the Stockbridge dampers; red is for the Tower plate; white is for the Transmission tower. It is possible to see in both figures how
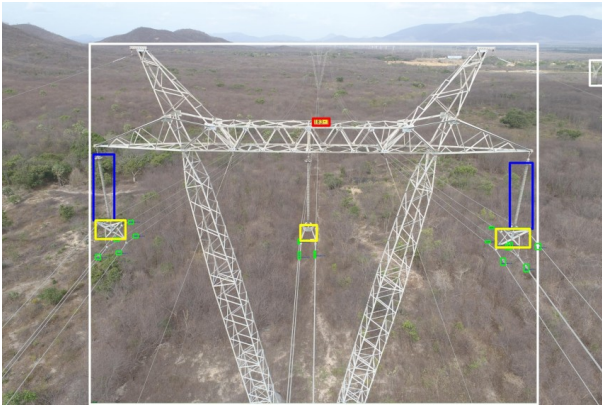
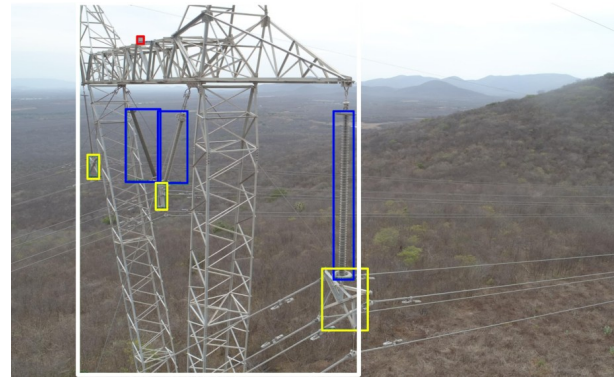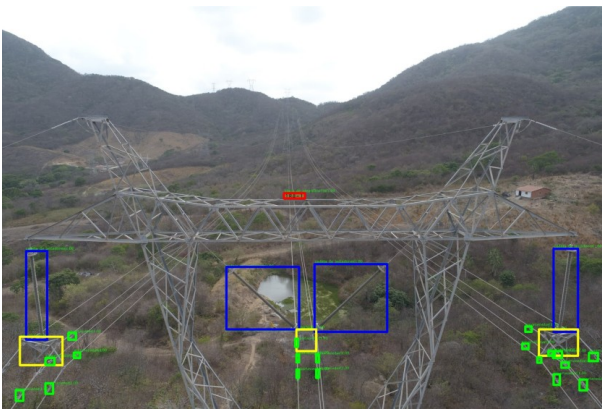Fig. 5. Qualitative detection results of the SSD technique using the original image-resizing approach.

Fig. 6. Qualitative detection results of the Faster R-CNN technique using the original image-resizing approach.

Fig. 7. Qualitative detection results of the MS-PAD pipeline for the big object classes.

Fig. 8. Qualitative detection results of the MS-PAD pipeline for the small object classes.

small the Stockbridge dampers are related to other objects. These images also illustrates failure cases, like the middle insulator in Figure 5 and the transmission tower in Figure 6.

B. MS-PAD results

For this experiment, $k = 5$, so five splits with randomly selected samples were used, in which each split is used twice: one time for the original image-resizing approach and another time for MS-PAD. The total amount of iterations for each training session is fixed at 20,000. The obtained results regarding AP for each split are shown in Table III. In addition, Table IV shows the average results from Table III reached by each approach side-by-side, in a direct comparison.

Figure 7 and Figure 8 present the qualitative performance of MS-PAD for big and small objects, respectively. The color codes previously mentioned are maintained for these images.

VI. DISCUSSION

This section details and discusses all results presented in section V, also giving insights into the usage of the MS-PAD approach in the proposed STN Power Line Asset Dataset.

A. STN PLAD strengths & limitations

The proposed STN PLAD is the first public power line assets dataset with multiple objects in real-world scenarios. It contains five classes of entirely different objects with multiple instances each, in varied real backgrounds. The data collection protocol allows for a balance in data quantity and variability since the captured images vary in illumination, backgrounds, and weather conditions. Also, the drone position is not fixed in order to obtain objects data from different perspectives. Another STN PLAD challenging characteristic is that there are many objects per image (18.1, on average) compared to the related public datasets, which commonly have a small instance per image rate (1 and 1.9, on average). Thanks to this process, STN PLAD poses a reasonable challenge to recent deep learning techniques, as observed in the section V, in which the best of the tested approaches achieved an 89.2% mAP, leaving considerable room for improvement.

Although the proposed STN PLAD provides new grounds in the power line area and stimulates the development of power line asset detection methods, it still has limitations. The main one is related to its total amount of images. That prevents some data-hungry object detectors from performing successfully since they would require a more extensive dataset. Another disadvantage is that the images were only collected from one private transmission line. Even though different transmission

TABLE III
COMPARISON OF DETECTION RESULTS OF THE ORIGINAL IMAGE-RESIZING APPROACH (ORIGINAL) AND THE MS-PAD PIPELINE (OURS) OF EACH
MONTE CARLO CROSS-VALIDATION SPLIT (k) RELATIVE TO THE AVERAGE PRECISION (AP). THE BEST MAP RESULTS ARE IN BOLD.

	$k = 1$		$k = 2$		$k = 3$		$k = 4$		$k = 5$	
	Original	Ours	Original	Ours	Original	Ours	Original	Ours	Original	Ours
Transmission tower	0.885	0.905	0.883	0.901	0.875	0.874	0.920	0.883	0.945	0.938
Insulator	0.825	0.938	0.924	0.866	0.931	0.893	0.839	0.884	0.874	0.889
Spacer	0.917	0.810	0.789	0.850	0.914	0.910	0.863	0.805	0.853	0.905
Tower plate	0.932	0.994	0.830	1.00	0.941	0.990	0.984	0.997	0.995	0.876
Stockbridge damper	0.189	0.829	0.201	0.882	0.189	0.870	0.264	0.824	0.227	0.787
mAP	0.750	**0.895**	0.725	**0.900**	0.770	**0.907**	0.774	**0.879**	0.779	**0.879**

TABLE IV
DETECTION AVERAGE RESULTS FROM TABLE III OF BOTH APPROACHES,
SIDE-BY-SIDE. THE BEST RESULTS FOR EACH CLASS AND THE MAP IS IN
BOLD.

	Original	Ours
Transmission tower	**0.902**	0.900
Insulator	0.879	**0.894**
Spacer	**0.867**	0.856
Tower plate	0.936	**0.971**
Stockbridge damper	0.214	**0.838**
mAP	0.755	**0.892**

lines tend to be similar, it would be better to have images of several transmission lines in other places to reduce the bias of background, environment, and electrical assets appearance. Finally, the images belong to a power line in Brazil, which may not apply to other countries.

B. SSD and Faster R-CNN comparison in STN PLAD

This discussion is related to the experiment in subsection V-A. According to the proposed methodology, when performing this experiment, it was expected that the results related to the Faster R-CNN would surpass the results from the SSD network considering the applied metrics. However, it can be observed in the results reported in subsection V-A that it did not occur. This result was obtained due to the limitations of the used data. Deep learning techniques benefit from the use of large amounts of data. According to Ng [34] [35] when the number of data limits a deep learning technique, shallower techniques can obtain comparable or even better results than deeper techniques. The used Faster R-CNN is much deeper and has more trainable parameters than the used SSD network. Therefore, for a limited amount of data, the learning of the used Faster R-CNN is limited.

The other results in this comparison were related to the SSD and Faster R-CNN performance in the competing datasets. In [14], both methods achieved 100% mAP, as expected due to the reasons presented in subsection II-A. In [13], the original SSD and Faster R-CNN obtained performances above 98% mAP. The high mAP values obtained by both object detectors in both competing datasets showed how well-resolved their challenges already are.

C. MS-PAD in STN PLAD

It is possible to see in Table III the mAP values reached by MS-PAD are higher than all mAP values of the Original approach in at least ten percentage points ($k = 5$) and at most 17.5 percentage points ($k = 2$). Table IV shows a direct comparison, in which the values for each approach are an average of the five splits showed in Table III. The best values for each class AP and mAP are in bold. MS-PAD yields the best AP result in three out of the five total classes, and there is a gap of 13.7 percentage points between mAPs. That gap is primarily due to the Stockbridge damper AP improvement, which grew 62.4 percentage points using MS-PAD.

It is noteworthy that the performance of large objects changes when comparing the Original and the MS-PAD. That may happen because during the MS-PAD resize branch training, one less class is considered (Stockbridge damper). That directly influences how the network learns since there is a different amount of objects and classes, directly impacting the final performance of large assets. Also, it is important to note that there is no guarantee that the performance impact will be positive or negative when training with one less class.

VII. CONCLUSIONS

This work proposes a new public real-world high-resolution power line asset dataset with multiple assets categories, called STN Power Line Assets Dataset (PLAD). Its images were captured by an Unmanned Aerial Vehicle (UAV) following a data collection protocol to ensure data variability in order to benefit deep learning models. STN PLAD contains 2409 annotated objects across 133 images divided into five classes with different shapes and sizes. It has the biggest amount of power line asset types among all public power line assets datasets, with the highest density of objects per image between them as well. The latter is possible due to its images having far above average resolutions, 5472×3078 and 5472×3648, more precisely. After evaluating STN PLAD in recent general-purpose object detectors, a different pipeline called MS-PAD is proposed. This pipeline contains a simple modification that allows for an mAP improvement from 75.5% to 89.2%. STN PLAD is publicly available to mitigate the lack of data in the power line inspection area and provide a new challenge to the computer vision community in order to stimulate the proposition of new asset detection methods for power lines.

REFERENCES

[1] V. N. Nguyen, R. Jenssen, and D. Roverso, "Automatic autonomous vision-based power line inspection: A review of current status and the potential role of deep learning," *International Journal of Electrical Power & Energy Systems*, vol. 99, pp. 107 – 120, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0142061517324444

[2] Y. Pradeep, S. A. Khaparde, and R. K. Joshi, "High level event ontology for multiarea power system," *IEEE Transactions on Smart Grid*, vol. 3, no. 1, pp. 193–202, 2012.

[3] A. Castillo, "Risk analysis and management in power outage and restoration: A literature survey," *Electric Power Systems Research*, vol. 107, pp. 9 – 15, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378779613002435

[4] M. Bruch, V. Münch, M. Aichinger, M. Kuhn, M. Weymann, and G. Schmid, "Power blackout risks," in *Cro Forum*, 2011, p. 28.

[5] L. Li, H. Wu, Y. Song, and Y. Liu, "A state-failure-network method to identify critical components in power systems," *Electric Power Systems Research*, vol. 181, p. 106192, 2020.

[6] Y. Hu and K. Liu, *Inspection and Monitoring Technologies of Transmission Lines with Remote Sensing*. Academic Press, 2017.

[7] A. Rahmani, M. Khadem, E. Madresehc, H.-A. Aghaei, M. Raei, and M. Karchani, "Descriptive study of occupational accidents and their causes among electricity distribution company workers at an eight-year period in iran," *Safety and health at work*, vol. 4, no. 3, pp. 160–165, 2013.

[8] J. Li, H. Wu, C. Hu, and C. Yu, "A fault diagnosis system based on case decision technology for uav inspection of power lines," in *IOP Conference Series: Earth and Environmental Science*, vol. 632, no. 4. IOP Publishing, 2021, p. 042077.

[9] N. Sambasivan, S. Kapania, H. Highfill, D. Akrong, P. Paritosh, and L. M. Aroyo, ""everyone wants to do the model work, not the data work": Data cascades in high-stakes ai," in *proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–15.

[10] A. Krizhevsky, G. Hinton *et al.*, "Learning multiple layers of features from tiny images," *Master's thesis, Department of Computer Science, University of Toronto*, 2009.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.

[12] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[13] X. Tao, D. Zhang, Z. Wang, X. Liu, H. Zhang, and D. Xu, "Detection of power line insulator defects using aerial images analyzed with convolutional neural networks," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2018.

[14] M. Tomaszewski, B. Ruszczak, and P. Michalski, "The collection of images of an insulator taken outdoors in varying lighting conditions with additional laser spots," *Data in Brief*, vol. 18, pp. 765 – 768, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2352340918302701

[15] R. Abdelfattah, X. Wang, and S. Wang, "Ttpla: An aerial-image dataset for detection and segmentation of transmission towers and power lines," in *Proceedings of the Asian Conference on Computer Vision*, 2020.

[16] X. Liu, X. Miao, H. Jiang, and J. Chen, "Review of data analysis in vision inspection of power lines with an in-depth discussion of deep learning technology," 2020.

[17] X. Lei and Z. Sui, "Intelligent fault detection of high voltage line based on the faster r-cnn," *Measurement*, vol. 138, pp. 379–385, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0263224119300831

[18] Y. Yang, L. Wang, Y. Wang, and X. Mei, "Insulator self-shattering detection: a deep convolutional neural network approach," *Multimedia Tools and Applications*, vol. 78, no. 8, pp. 10097–10112, 2019.

[19] H. Zhang, M. Sun, Y. Ji, S. Xu, and W. Cao, "Learning-based object detection in high resolution uav images: An empirical study," in *2019 IEEE 17th International Conference on Industrial Informatics (INDIN)*, vol. 1, 2019, pp. 886–889.

[20] Z. A. Siddiqui, U. Park, S.-W. Lee, N.-J. Jung, M. Choi, C. Lim, and J.-H. Seo, "Robust powerline equipment inspection system based on a convolutional neural network," *Sensors*, vol. 18, no. 11, p. 3837, 2018.

[21] Y. Ö. Emre, G. Ö. Nezih *et al.*, "Powerline image dataset (infrared-ir and visible light-vl)," *Mendeley Data*, vol. 7, 2017. [Online]. Available: https://data.mendeley.com/datasets/n6wrv4ry6v/7

[22] H. Zhang, W. Yang, H. Yu, H. Zhang, and G.-S. Xia, "Detecting power lines in uav images with convolutional features and structured constraints," *Remote Sensing*, vol. 11, no. 11, p. 1342, 2019.

[23] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proceedings of COMPSTAT'2010*. Springer, 2010, pp. 177–186.

[24] A. Rakhlin, O. Shamir, and K. Sridharan, "Making gradient descent optimal for strongly convex stochastic optimization," in *Proceedings of the 29th International Coference on International Conference on Machine Learning*, 2012, pp. 1571–1578.

[25] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham: Springer International Publishing, 2016, pp. 21–37.

[26] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, Eds., vol. 28. Curran Associates, Inc., 2015, pp. 91–99. [Online]. Available: https://proceedings.neurips.cc/paper/2015/file/14bfa6bb14875e45bba028a21ed38046-Paper.pdf

[27] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei, "Deformable convolutional networks," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[28] L. Kong, X. Zhu, and G. Wang, "Context semantics for small target detection in large-field images with two cascaded faster r-CNNs," *Journal of Physics: Conference Series*, vol. 1069, p. 012138, aug 2018. [Online]. Available: https://doi.org/10.1088%2F1742-6596%2F1069%2F1%2F012138

[29] X. Zhu, L. Kong, G. Wang, Z. Hu, and S. Li, "Multi-size object detection assisting fault diagnosis of power systems based on improved cascaded faster R-CNNs," in *Tenth International Conference on Digital Image Processing (ICDIP 2018)*, X. Jiang and J.-N. Hwang, Eds., vol. 10806, International Society for Optics and Photonics. SPIE, 2018, pp. 342 – 351. [Online]. Available: https://doi.org/10.1117/12.2503064

[30] Tzutalin, "Labelimg," Git code, 2015, https://github.com/tzutalin/labelImg.

[31] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*. Springer, 2014, pp. 740–755.

[32] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2117–2125.

[33] W. Dubitzky, M. Granzow, and D. P. Berrar, *Fundamentals of data mining in genomics and proteomics*. Springer Science & Business Media, 2007.

[34] A. Ng, *Machine learning yearning*. Stanford Press, 2017, http://www.mlyearning.org/(96).

[35] A. Tang, R. Tam, A. Cadrin-Chênevert, W. Guest, J. Chong, J. Barfett, L. Chepelev, R. Cairns, J. R. Mitchell, M. D. Cicero *et al.*, "Canadian association of radiologists white paper on artificial intelligence in radiology," *Canadian Association of Radiologists Journal*, vol. 69, no. 2, pp. 120–135, 2018.