PDP-CNN: A Deep Learning Model for Post-hurricane Reconnaissance of Electricity Infrastructure on Resource-constrained Embedded Systems at the Edge

Ashkan B. Jeddi, *Student Member, IEEE,* Abdollah Shafieezadeh, *Member, IEEE*, Roshanak Nateghi, *Member, IEEE*

***Abstract*—Unmanned aerial vehicles (UAVs) equipped with cameras have provided new capabilities for the reconnaissance of disaster-stricken areas. Deep learning-based computer vision algorithms enable the analysis of the captured images and the detection of damage to the built environment. If such analyses are conducted onboard in UAVs, they will provide real-time actionable information that is critical for the accelerated restoration of systems. However, conventional deep learning algorithms are computationally demanding. Moreover, the deployment of deep learning models with a large number of parameters in scenarios that require low-latency inference is prohibitive. To address this fundamental gap, we develop an *efficient* deep learning-based computer vision model of power distribution poles (PDP) damage detection that is capable of onboard deployment in UAVs. Specifically, we propose a lightweight convolutional neural network (CNN) architecture called PDP-CNN that embodies multi-scale feature operations and anchor-less object detection. This model is applied to a dedicated image database from a post-hurricane reconnaissance in power distribution poles. Results of extensive experiments show that PDP-CNN is capable of achieving high throughput, competitive accuracy, and efficient memory utilization on power-constrained embedded systems.**

***Index Terms*—Computer vision, convolutional neural networks (CNNs), unmanned aerial vehicles (UAVs), damage detection, hurricanes, distribution systems, utility poles, reconnaissance data.**

I. Introduction

EXTREME weather events pose significant risks to the reliability and resilience of power systems, resulting in tens of billions of dollars in economic damage annually and adverse public health impacts [1]. According to the North American Electric Reliability Corporation (NERC), a third of all sustained outages in medium and high voltage systems in the period of 2013-2018 were caused by weather-related events [1]. Moreover, a recent analysis of power outage data collected by the US Department of Energy (DOE) reveals a 67% increase in major power outages from weather-related events since 2000 [2].

This work was supported by the National Science Foundation under Grant CMMI-2000156.

A. B. Jeddi and A. Shafieezadeh are with the Department of Civil, Environmental, and Geodetic Engineering, Ohio State University, OH, USA (e-mail: bagherijeddi.1@osu.edu; shafieezadeh.1@osu.edu). R. Nateghi is with the School of Industrial Engineering, Division of Environmental Engineering Purdue University, West Lafayette, IN, USA (e-mail: rnateghi@purdue.edu)

Overhead power distribution systems are highly vulnerable to high-wind hazards [3]. In fact, high wind-induced outages in distribution systems account for over 92% of electric service interruptions in the US [4]. The extensive failures in power distribution systems in the aftermath of extreme events often lead to prolonged power outages. For example, full restoration of the grid post-Hurricane Maria in 2017 took nearly 10 months [5]. Similarly, nearly 800,000 customers experienced a 10-day long blackout during Superstorm Sandy in 2012 [6]. The high vulnerability of our overhead distribution systems and the significant socio-economic costs of prolonged outages underpin the critical importance of improved grid resilience, where consumers' access to electricity is restored rapidly post-disasters. The critical importance of achieving climate resilience is not lost on the current administration as evidenced by the DOE's new *Building a Better Grid Initiative*, investing $2.3 billion to strengthen the grid against extreme weather [7].

Strategies such as grid hardening, dispatch of restoration crew, network reconfiguration, and deployment of distributed generation units [8]–[10] are often leveraged to enhance the resilience of power systems during disruptions. Besides these control-based strategies, prompt assessment of physical damages to components in the aftermath of extreme weather and climate events such as hurricanes can facilitate rapid power restoration. Specifically, analysis of the collected damage data enables fault detection. Although these methods can help reduce overall equipment downtime, they cannot accurately pinpoint the location of the fault, and the type and degree of sustained damages. Moreover, data acquisition systems are prone to failure during extreme events, leaving the possibility that faults may not be reliably detected.

Visual inspection of the power equipment also provides valuable information about the characteristics of physical damages. Several methods of visual inspection have been employed by utility companies during extreme events. Manual reconnaissance (i.e., foot patrol) is the most common method, where a group of experienced inspectors visits impacted regions and documents incurred damages. Although foot patrol has been commonly used for post-disaster reconnaissance, the impacted regions in most cases are not readily accessible due to varying degrees of failures in transportation systems as well as safety concerns to inspectors posed by potentially live conductors. Unmanned aerial vehicles (UAVs) are

a favorable alternative in these conditions. Deployment of UAVs has recently gained attention for damage assessment and monitoring of infrastructure systems [11], [12] due to their high maneuverability, the ability for remote control, and lower cost relative to conventional reconnaissance strategies.

Motivated by this paradigm shift, several recent studies have focused on leveraging UAV-captured imagery and deep learning-based computer vision for rapid autonomous damage assessment in power distribution and transmission systems. For example, the authors in [13] developed damage classification and estimation models based on four different convolutional neural network (CNN) models. The first CNN unit was designed to classify power distribution poles (PDP) into three classes of healthy, fallen, and burning poles. Other separate CNN units were designated to estimate the extent of falling and fire damages in poles. The CNN models were loosely based on the structure of a ResNet18 model and the four CNN models had a total of 37.6 million learnable parameters. Adopting a similar pre-trained ResNet18 model with 72 layers and 11 million parameters, [14] estimated the leaning angle of distribution poles using a dataset acquired from open-source Google Street View images. In addition, the authors proposed using AlexNet with 61 million parameters for material classification in distribution poles. Inference using these models was carried out on a GeForce 1050 Max-Q NVIDIA graphical processing unit (GPU). With a focus on the detection of faulty insulators on overhead transmission towers, the authors in [15] developed an improved variant of YOLOv3 and a densely connected feature pyramid network with a total of 51 million learnable parameters. Similarly, [16] proposed a multiscale residual model for insulator surface breakage detection. The results of the performance evaluation showed a competitive accuracy compared to the state-of-the-art architectures with a frames per second of 3.16 and 3.14GB of FLOPS. Semantic segmentation is another approach that has been applied for detecting and assessing distribution poles. For example, the authors in [17] used a modified version of the SegNet model for pixel-wise semantic segmentation to detect distribution poles and an image processing approach to estimate the incline angle. In their study, the inference was carried out in an end-to-end manner on a GeForce 1080ti NVIDIA GPU.

Training the most commonly used convolutional neural network architectures -such as the ones employed in the existing literature- requires GPUs with high computing power and memory resources. In lieu of these resources on reconnaissance sites, training and evaluation of computer vision models are often carried out after transferring the imagery dataset off-site to cloud datacenters. However, data transfer between the cloud and an edge device requires a reliable wireless internet connection with high bandwidth. This prohibits the use of cloud processing units for reconnaissance in regions struck by natural hazards with limited connectivity. In these situations, relatively powerful embedded systems are a favorable alternative for real-time analysis of data real-time. However, it is often impractical to deploy deep neural network architectures on resource-constrained embedded systems. In particular, the size of conventional models which is greatly larger than the on-chip local storage prohibits such applications. More importantly, large networks cannot be employed on low-power UAV-mounted embedded platforms due to their high power consumption. Reducing the number of learnable parameters by pre-training computer vision models is often selected as a remedy to this issue. The major shortcoming of this approach is poor detection performance in object-specific tasks as a result of transferring weights learned on generic object detection datasets.

A. Major Contributions

To address the limitations elaborated in the previous section, we develop an efficient CNN approach, PDP-CNN, specifically designed for damage detection and classification of distribution poles from post-hurricane reconnaissance imagery data at the edge. In particular, as the architecture of object detection models has a significant influence on both their accuracy and computational efficiency, our objective is to design an architecture that enables real-time perception capabilities at the edge. For this purpose, we propose a new feature extraction backbone, as well as a feature pyramid network, and a detection backend. In summary, our major contributions are as follows:

- We develop a novel deep learning-based computer vision model suitable for deployment on resource-constrained embedded systems such as unmanned aerial vehicles.
- We introduce a lightweight yet effective feature extraction backbone that leverages separable convolutions, short and long dense-like connections, and feature pyramid networks.
- We develop anchor-free detection heads that remove the need for postprocessing steps such as non-maximum suppression.
- Results of extensive experiments show computational efficiency and competitive performance of the developed model compared to state-of-the-art computer vision models.

B. Organization

This paper is organized as follows. Section II provides a background of object detection methods based on deep learning. Details of the proposed PDP-CNN model are elaborated in Section III. Data acquisition and image pre-processing methods are explained in Section IV. Results of training, validation, and testing of the developed PDP-CNN model are presented in Section V. The conclusions of this research are presented in Section VI.

II. Background

A. Object Detection

Deep learning-based object detection methods are classified into two major categories, namely, region proposal-based methods and classification-based methods. While region proposal-based models require a discrete step designed for region proposal, classification-based methods consist of a single CNN for the acquisition of representative features and localizing the objects of interest. A brief review of these categories is presented next.

1) Region Proposal-based Methods: The family of region-based convolutional neural networks including R-CNN, Fast R-CNN, and Faster R-CNN [18] belongs to the class of region proposal-based methods. These methods essentially decompose the process of object detection into two discrete stages and design separate algorithms for each stage. The first stage in this process consists of extracting regions of interest, i.e., regions of proposals. This stage is followed by a separate network that is responsible for the classification of each region proposal as well as predicting the bounding box. Typically, the first stage of region proposal-based methods is the computational bottleneck of these object detection algorithms. Faster R-CNN alleviates this issue by developing a region proposal network (RPN) which is a fully convolutional network. RPN shares full-image convolutional features with the detection network used in the second stage of detection and therefore enables nearly cost-free region proposals. Despite major advancements offered by methods such as Faster R-CNN, generating a large number of region proposals and classifying each region during inference is a computationally demanding process that introduces significant delays [19]. For this reason, classification-based detection methods have gained attention. In this paper, we exclusively focus on classification-based detector networks due to their desirable trade-off between accuracy and computational cost which makes them more suitable for deployment on low-power and resource-constrained embedded systems.

2) Single-shot detectors: Single-shot detectors such as different variations of YOLO [20] and SSD [21] combine the distinct components of region proposal-based methods into a unified single neural network. Through this process, the features of the input image are directly used for both the classification and regression of the bounding boxes. In particular, YOLO uses a spatial grid over the input image and estimates bounding box coordinates along with the object label and the presence probability of the object. The architecture of YOLO models is inspired by GoogleNet and has a total of 24 convolutional layers with two fully connected layers at the end. Despite the computational advantage of YOLO, their major drawback is that they are limited to detecting only a single object in each grid. On the other hand, SSD network uses convolutional layers of varying sizes and computes multiscale feature maps which in turn produces the bounding box and class predictions. To further enhance the computational efficiency of YOLO models, lightweight variants, i.e., tiny-YOLO, are proposed in the literature where smaller feature extraction layers are employed at the cost of a marginal drop in accuracy [20], [22], [23].

B. Benchmark CNN Architectures

CNN networks commonly consist of a stack of convolutional layers with the purpose of performance enhancement. VGG [24] expanded the depth of CNN models and introduced a variant of the model with 13 convolutional layers and 3 fully connected layers as VGG-16 and a deeper variant with 3 additional convolutional layers as VGG-19. To overcome the limitations of the conventionally used CNN architectures in the selection of kernel sizes, inception networks were introduced [25]. Inception networks leveraged kernels of multiple sizes to develop an inception module which results in a wide rather than a deep network. Besides the limitation on the selection of kernel sizes, extremely deep neural networks often suffer from vanishing and exploding gradients. To address this issue, He et al. [26] introduced the idea of *skip connections*. Through this modification, the activations of one layer are directly fed into another deeper layer to explicitly let these layers fit a residual mapping. This resulted in the introduction of the residual blocks that prevent performance reduction in deep networks. Leveraging the idea of skip connections, ResNet-50 and ResNet-101, with 50 and 101 convolutional layers, were introduced. DenseNets [27] extended the idea of skip connections and proposed concatenation of the output of each layer with the output from all previous layers. MobileNetV1 [28] and V2 [29] introduced *separable convolutions* which replace normal convolution operations by depthwise convolution followed by point-wise convolution.

III. PDP-CNN Architecture

The PDP-CNN architecture is composed of three main components: a novel feature extraction backbone, a feature pyramid network, and an anchor-free detection backend. The feature extraction backbone features a set of properties to achieve state-of-the-art performance and accurate representations while minimizing the computational cost for application in resource-constrained embedded systems. These features include separable convolutions, long skip connections between layers with the aim of creating a multi-scale set of features as well as short connections to enhance the diversity of learned feature maps. The FPN module merges the feature maps extracted by the backbone in a top-down manner from coarsest to finest and generates multi-scale feature maps. Subsequently, the pyramid features are fed into the designed detection backend which consists of two separate subnets for object detection and bounding box regression. To further enhance the computational efficiency, we design the detection backend such that it does not require anchors for bounding box predictions. In addition to the novel PDP-CNN architecture, we propose a direct set prediction loss function that includes two major components: a bipartite matching loss and a bounding box loss to improve localization. Fig. 1 illustrates the overall architecture of the proposed PDP-CNN model.

A. Feature Extraction Backbone

The performance of deep learning-based object detection algorithms hinges on the quality of representations learned by the feature extraction layers. Besides the importance of these layers in achieving high detection accuracy, the architecture of feature extraction layers also impacts inference performance in terms of throughput. Evidently, feature extraction networks with fewer parameters exhibit lower memory utilization and consequently lower power consumption. Therefore, given the significant influence of feature extraction layers on the overall performance of detection models as well as the resources consumed by these layers, it is imperative to design a powerful

yet lightweight network with a limited number of parameters. With these considerations, we design 10 convolutional layers as the feature extractor in the PDP-CNN backbone. Inspired by the performance of *separable convolutions* [29], we explicitly factor regular convolution into distinct steps of a depthwise convolution followed by a pointwise convolution. Depthwise convolutions perform a spatial convolution independently over each channel of input. In addition, pointwise convolutions perform a 1×1 convolution and project the outputs of the depthwise convolution onto a new channel space. This results in a substantial reduction of the parameter count by a factor of the depth of the output feature map and the size of the kernel, without a noticeable reduction in accuracy.

Progressive downsampling is essential for the reduction of memory usage and computational cost in the development of efficient object detection algorithms. However, in order to maintain the localized information on foreground objects required for detection networks throughout the feature extraction layers, excessive downsampling should be avoided. In this study, we empirically achieve a desirable trade-off between the preservation of localized feature points and computational efficiency. In this context, the initial convolutional layer in the PDP-CNN backbone as shown in Fig. 1 only downsamples the input image by a factor of 2 using strided convolutions. Furthermore, in the first block of layers, the first convolutional layer and the other three have 64 and 32 kernels, respectively, and with a kernel size of 5×5. Other convolutional layers employ kernels with a size of 3×3 with an increasing number of filters of up to 256. It is critical to limit the number of kernels in order to achieve higher performance in the first few layers where the resolution of the feature map is extremely high.

Inspired by DenseNets [27], in order to facilitate a fluent flow of gradients and feature propagation between the layers of PDP-CNN and to avoid problems such as vanishing or exploding gradients, we design short dense-like connections after concatenation layers and long dense-like connections placed between pooling and every other convolutional layer. This design further builds a more diverse set of visual features. This along with the fusion of additional information with less semantic meaning compensates for pooling operations which reduce the spatial resolution. Therefore, the PDP-CNN backbone is capable of achieving richer feature maps. Despite the high accuracy of DenseNet architectures, a major drawback of these networks is the progressive increase of their computational demand as the number of dense connections increases [30]. In order to alleviate this issue, we design PDP-CNN with only a selection of short and long connections instead of using the feature maps of all preceding layers as inputs. Based on the selected arrangement of the layers in the backbone architecture, we design a connection at every other block of convolutional layers. Moreover, a channel-wise concatenation merges the feature maps that are obtained from the preceding pooling layer.

The final output of the feature extraction backbone is a coarse feature map that essentially represents a downsample of the input image. As a result, a desirable object detection accuracy is out of reach by only using this coarse feature map. To resolve this issue, we propose combining feature maps from multiple layers through FPN. FPN provides a top-down pathway to construct higher-resolution layers from a semantically richer layer. To augment extracted feature maps and create multi-scale feature pyramids, we extract feature maps $M3$, $M4$, and $M5$ and note the downsampling rate of 2^l where l denotes the index of the feature map. Accordingly, feature maps $P3$, $P4$, and $P5$ are created using upsampling. In addition, $P6$ and $P7$ are built using a convolution with a stride of two. Ultimately, these feature maps, i.e., $P3$ through $P7$ are passed to the detection backend.

B. Detection Backend

Inspired by the computational efficiency of RetinaNet [31] and anchor-less object detection models, we propose a novel anchor-free detection backend that does not assume predefined anchors during the training process. In this process, the points on extracted feature maps throughout the FPN are labeled with positive or negative classes considering the sampling rate of 2^l. For anchor-based object detection algorithms, if the anchor falls in a ground truth box, and the intersection-over-union (IoU) metric is larger than a fixed threshold, we label the point as positive and otherwise negative. Similarly, we assume that the point (x, y) on the input image falls in a ground truth box with center coordinates of (x_g, y_g), w as the width, and h as the height. We define (a_1, a_2, a_3, a_4) as the distance of point (x, y) from the four sides of the ground truth box, i.e., left, top, right, and bottom, respectively. Then, we define a projected box with center point (x, y) and distances (b_1, b_2, b_3, b_4). Based on this definition, b_1 and b_3 are the average of a_1 and a_2, and similarly b_2 and b_4 are defined with respect to a_2 and a_4. Next, the intersection and union between these two boxes with distance sets of a and b are calculated and IoU is determined. Finally, the points on feature maps are labeled based on this value. Through this process, we remove the need for predetermined anchor boxes and their corresponding computational complexity. It is worth noting that there still exists a threshold for the labeling process similar to the anchor-based methods.

Separate fully convolutional subnetworks are ubiquitously employed in the literature as detection heads for classification and regression tasks (e.g., [32]). Adopting this scheme, we design a detection head that consists of two subnetworks with four convolutional layers and 256 filters of the size of 3×3. Following these layers, the classifier subnetwork has a convolutional layer with a kernel size of 3×3 and K filters, where K represents the number of target classes. On the other hand, the regressor subnetwork adopts 4 filters with the same size, where the four outputs represent the bounding box (a_1, a_2, a_3, a_4). It is worth noting that the two subnetworks do not share weights despite their architectural similarity. However, the weights corresponding to each subnetwork are shared between features $P3$ through $P7$ that are extracted from the FPN model.

C. Loss Function

In order to learn the task of distribution pole detection in an end-to-end fashion, we develop a set prediction loss

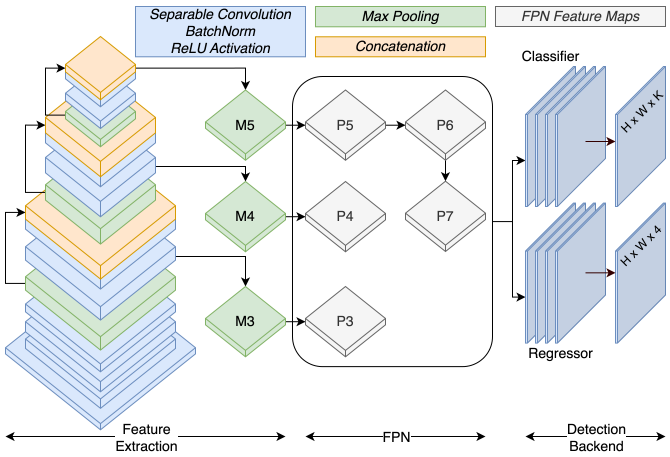

Fig. 1. An overview of the PDP-CNN consisting of a CNN-based feature extraction backbone, an FPN, and a detection backend for on-edge reconnaissance of power distribution poles. H and W denote the dimensions of the input image.

function. Inspired by the set-based global loss function in detection transformer (DETR) [33] and similar to methods such as Faster R-CNN [18] and RetinaNet [31], the proposed loss function consists of two components: bipartite matching loss and bounding box loss.

1) Bipartite Matching Loss: After passing through the detection backend, PDP-CNN makes N predictions where N represents the number of classes. Let y be the set of ground truths in the image, i.e., $y_i = (c_i, b_i)$, and $\hat{y} = \{\hat{y}_i\}_{i=1}^{N}$ be the set of N predictions of the network where $\hat{y_i} = (\hat{c}_i, \hat{b}_i)$ is the tuple consisting of the predicted class, c_i, and a bounding box $b_i = (\bar{x}_i, \bar{y}_i, w_i, h_i)$ where $(\bar{x}_i, \bar{y}_i)$ represents the midpoint of the bounding box, and w_i and h_i are the width and height of the bounding box, respectively. PDP-CNN finds a bipartite matching between these two sets of y and $\hat{y}$ using a matching function across a permutation of N elements, $\sigma \in \mathcal{S}_N$, with the lowest cost as follows:

$$\hat{\sigma} = \text{argmin}_{\sigma \in \mathcal{S}_N} \sum_{i}^{N} \mathcal{L}_{\text{match}}(y_i, \hat{y}_{\sigma(i)}) \tag{1}$$

where $\mathcal{L}_{\text{match}}$ is a pair-wise matching cost between ground truth y_i and prediction with index $\sigma(i)$. The matching cost accounts for both the class prediction and the similarity of predicted and ground truth boxes. For the prediction with index $\sigma(i)$, the probability of class c_i is defined as $\hat{p}_{\sigma(i)}(c_i)$ and the predicted box as $\hat{b}_{\sigma(i)}$. The matching cost, therefore, is then defined as

$$\mathcal{L}_{\text{match}}(y_i, \hat{y}_{\sigma(i)}) = -\mathbb{I}_{\{c_i \neq \emptyset\}} \hat{p}_{\sigma(i)}(c_i) + \mathbb{I}_{\{c_i \neq \emptyset\}} \mathcal{L}_{\text{box}}(b_i, \hat{b}_{\sigma(i)}) \tag{2}$$

where $\mathbb{I}_{\{\cdot\}}$ is an indicator function. The second step is to compute the loss function, which consists of the Hungarian loss for all pairs matched in the previous step. The loss is defined similarly to the losses of common object detectors as a linear combination of a negative log-likelihood for class prediction and a bounding box loss

$$\mathcal{L}_{\text{Hungarian}}(y_i, \hat{y}_i) = \sum_{i=1}^{N} [-\log \hat{p}_{\hat{\sigma}(i)}(c_i) + \mathbb{I}_{\{c_i \neq \emptyset\}} \mathcal{L}_{\text{box}}(b_i, \hat{b}_{\hat{\sigma}(i)})] \tag{3}$$

where $\hat{\sigma}$ is the optimal assignment computed in the first step via Eq. (1).

2) Bounding Box Loss: Another term in the loss function is the bounding box loss which estimates a score for the predicted bounding boxes. In this paper, a linear combination of the l_1 loss and the generalized IoU loss $\mathcal{L}_{\text{IoU}}$ that is scale-invariant is used as

$$\mathcal{L}_{\text{box}} = \lambda_{\text{IoU}} \mathcal{L}_{\text{IoU}}(b_i, \hat{b}_{\sigma(i)}) + \lambda_{l_1} ||b_i - \hat{b}_{\sigma(i)}|| \tag{4}$$

where λ_{IoU} and λ_{l_1} are hyperparameters. The bounding box loss function helps to predict the box directly without any anchor reference or scaling issue. These two losses are normalized by the number of objects inside the batch.

IV. Data Collection and Pre-processing

In this study, we use a post-hurricane reconnaissance imagery dataset collected by the StEER group of inspectors through both single-camera shots of utility poles and street-level imagery. After processing the street-level imagery and extracting all recorded distribution poles, the repository of images consists of 3182 poles that are acquired in the aftermath of three hurricane events (Irma 2017, Maria 2017, and Michael 2018) as well as an archive of past events [34]. A small set of damaged distribution poles (53 cases) is accompanied by a description reporting the state of the pole as one of the classes of intact, leaning, and fallen. In this labeling scheme, if structural damage is observable in the body of the pole, it is labeled as fallen and if only the pole is leaning and no structural damage is observed, the pole is labeled as leaning. In this study, these three classes of intact, learning, and fallen are hereinafter referred to as `T`, `D1`, and `D2`, respectively. Missing damage categories are manually annotated by the authors complying with the visual characteristics of the already labeled images by the inspectors. Samples of leaning and fallen poles are presented in Fig. 2.

We apply a scale augmentation to the input images in this study to a minimum and maximum dimension of 600 and 1024 pixels, respectively while keeping the original aspect ratio. The input images were horizontally flipped randomly with a probability of 0.5 as part of the data augmentation. Furthermore, in order to account for the variation in lighting conditions of the images, we apply a color jittering that randomly changes the brightness, contrast, saturation, and hue of an image. Moreover, we account for the potential low-quality of the captured images by the embedded systems by randomly blurring the images through a Gaussian blur.

V. Results and Discussion

A. Training, Validation, and Testing of PDP-CNN

We divide the labeled dataset of utility poles into parts containing 60%, 20%, and 20% of the size of the dataset and use them for training, validation, and testing, respectively. The purpose of defining a validation set is solely to tune the hyper-parameters of the PDP-CNN model via cross-validation. No repetitive sample was used in splitting the dataset for validation, and the evaluation metrics obtained through the

Fig. 2. Sample images of damaged utility poles from the gathered dataset. [35]

cross-validation folds were averaged to produce a single-point estimate. In addition, the training data have imbalanced representation from different classes, with the intact class having the largest number of data points, which can potentially induce a prediction bias towards the over-represented class of intact utility poles. Specifically, the populations of the classes of intact, leaning, and fallen in the dataset curated in this study are about 81%, 16%, and 3% of the size of the entire dataset (i.e., 3182 poles), respectively. To mitigate the problem of inducing bias, class-specific weights based on these populations are assigned to the loss function to impose an additional penalty for misclassifying an under-represented class [18]. The PDP-CNN model is trained with AdamW. The initial learning rate is set to 10^{-5} for the CNN model. Furthermore, a weight decay of 10^{-4} is applied to gradually reduce the learning weight along training epochs. The weights are initialized via Xavier initialization. The PDP-CNN model was implemented in PyTorch and was run on one GeForce 3090 NVIDIA GPU for 30 epochs with a batch size of 64 for about 40 minutes.

B. Metrics and Evaluation

1) Mean Average Precision (mAP): Precision and recall are two of the most common performance evaluation metrics that are inversely related. Therefore, impartial comparison between object detection methods is conventionally carried out without solely relying on either of these metrics. Alternatively, the area under the precision-recall curve is defined as average precision (AP). In the case of multi-class object detection tasks such as the problem in this study, mean average precision (mAP) is reported which is defined as the mean of AP values in the detection of individual classes.

2) Intersection over union (IoU): Intersection over union (IoU) is a well-established and common evaluation metric in anchor-based object detection methods. IoU investigates the similarity between arbitrary shapes. In the case of object detection, IoU is defined as the ratio of the area of overlap to the area of the union between the ground truth and estimated bounding boxes.

3) False positive per image (FPPI): FPPI is defined as the ratio of predicted false positives and the total number of images.

4) Miss rate (MR): MR is defined as the ratio between the number of missed objects and the total number of samples belonging to the same class.

5) Frames per Second (FPS): FPS is defined as the number of images that can be processed by the object detection model in a second during the inference phase (i.e., the rate of processing images).

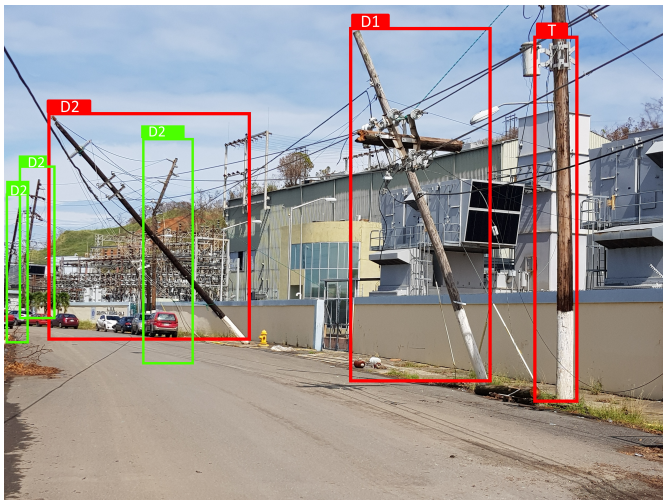

Fig. 3. Detected classes of damaged and undamaged utility poles in a single image. `T`, `D1`, and `D2` refer to classes of intact, leaning, and fallen, respectively. Identified classes have the highest probability among the three possible classes. (Different colors for overlapping bounding boxes are selected solely for better visualization. This image is captured during the reconnaissance after hurricane Michael in 2018. Raw image courtesy of [36])

C. Ablation Study

In this section, we perform an ablation study to determine the contribution of the building blocks of PDP-CNN in its performance in terms of mAP. Four variants of the model are considered in this analysis and the results are shown in Table I. The first variant of the model removes the loss function definition by performing a regression to estimate the bounding box coordinates. This base model further removes the defined short skip connections in the feature extraction backbone as well as the FPN and its long connections. Therefore, this model does not leverage diverse and semantically rich feature maps otherwise offered by PDP-CNN and can only achieve an mAP of 73.5%. The second variant uses the defined bipartite matching loss and bounding box loss and therefore investigates the contribution of the loss function formulation in the performance of PDP-CNN. As can be seen in Table I the introduced loss function significantly improves the performance of PDP-CNN in object localization and classification by 7.6%. To evaluate the significance of diverse feature maps, we add another variant that employs short skip connections as well as the defined loss function. This addition has resulted in a 1.8% increase in mAP. Finally, the last variant in this ablation study employs all features of the full PDP-CNN model and shows an mAP of 86.1%. This observation highlights the significance of

TABLE I
COMPARISON OF THE IMPACT OF MAIN COMPONENTS OF THE MODEL VIA ABLATION STUDY.

Loss Function		✓	✓	✓
Short Connections			✓	✓
Feature Pyramid Network				✓
mAP (%)	73.5	80.7	83.3	86.1

using multi-scale feature maps which is mainly attributed to the presence of objects with different scales in a single image similar to Fig. 3.

D. Comparison of Networks

Table II presents the overall performance of PDP-CNN in the detection and classification of the damages in power distribution poles. We compare this performance with a variety of state-of-the-art CNN benchmark architectures where all models are initialized with MSCOCO pre-trained weights and the same PDP-CNN training framework as described in Section V-A. This comparison also includes EfficientNet [37] in addition to the architectures described in Section II as one of the recent architectural advancements in object detection. EfficientNet introduces the idea of utilizing an effective compound coefficient to scale up CNNs. This model has shown state-of-the-art accuracy on benchmark datasets while decreasing the number of parameters by 20 times. In overall, the selected benchmark architectures are the most commonly used and well-established models in the field of object detection.

Performance of the proposed method is also presented for the individual classes of intact, leaning, and fallen in Table II in terms of the mean value and standard deviation of average precision (AP) over rounds of cross-validation. Based on the information presented in this table, no consistent pattern exists in the performance hierarchy of the baselines evaluated with respect to the AP value. To resolve the problem of inconclusive performance comparisons, it is common to average the AP values over all object classes and refer to mAP as a metric for the comparison of the efficiency of different detection algorithms. As can be seen in Table II, the proposed PDP-CNN backbone achieves an mAP of 86.1% which means that PDP-CNN is the second-best detection model, following YOLOv5 only by 1.2%. It is worth noting that the threshold of labeling in the anchor-free detection backend is a hyperparameter of the PDP-CNN model and it is set based on the performance of the model on the validation dataset. Based on our study, a labeling threshold of 0.4 results in the best mAP on the validation dataset. Moreover, in addition to the evaluation metrics defined earlier, we use accuracy contribution per parameter (APP) as a metric to quantify the contribution of parameters in the model in achieving high accuracy. This metric is defined as the ratio between an accuracy-related metric -mAP in this study- and the number of parameters in million. Based on APP, despite better mAP of YOLOv5, it suffers from poor memory utilization and therefore it achieves a significantly less APP compared to PDP-CNN. In fact, YOLOv5 is the largest CNN architecture among the benchmark architectures used in this analysis as it has the largest number of parameters. On the other hand, PDP-CNN exhibits an impressive APP value of 2.77 while all other models have an APP value of less than 1. This is further translated to the high FPS of PDP-CNN which is the highest among benchmark models. In terms of false positives and miss rate, PDP-CNN shows performance similar to the other benchmark architectures. The YOLOv5 model also shows the best performance with respect to FPPI which is mainly attributed to its larger number of parameters and the ability to achieve semantically richer feature maps. Contrary to large networks such as YOLOv5 and VGG16, PDP-CNN achieves high mAP through short skip connections, feature pyramid network, and anchor-free detection backend that result in more diverse and multi-scale feature maps. This demonstrates the effectiveness of the designed architecture.

Fig. 3 shows the detected and classified distribution poles in one of the test sample images. This sample is a great representation of the ability of PDP-CNN in the detection, localization, and classification of poles. As can be seen in this image, many poles are located in the background of the closer poles and show significant overlap. Even in this complex sample, PDP-CNN has identified and classified poles with multiple classes and at a far distance from the point where the image was captured. One of the failure cases of the proposed PDP-CNN model is predicting the pole class to be intact when the damage to the pole has caused detachment of the upper portion of the pole. However, this is one of the rare cases of damage in the employed dataset and in general does not control the performance of the model. Another limitation is the cases where the damaged poles are obstructed by environmental objects in still images. Both of these limitations can be resolved by increasing the size of the dataset and providing a stream of images to the model.

VI. CONCLUSION

The reliability and resilience of electric power distribution systems are exceedingly threatened by extreme weather events. In light of this increased vulnerability, energy systems' resilience has garnered more attention in recent times, especially since the reliance of communities on electric power is growing exponentially. A critical phase in the restoration of electric power to affected regions is the collection and processing of information about the state of the grid infrastructure in the aftermath of extreme events. This step is necessary to provide actionable information for effective crew dispatch management. Unmanned aerial vehicles (UAVs) can cruise autonomously and acquire critical visual information about the infrastructure. An autonomous damage detection model can subsequently facilitate rapid evaluation of the damaged assets. This paper developed a deep learning-based computer vision approach for rapid damage assessment of utility poles in the wake of hurricane events. In the broader domain of computer vision, this problem is regarded as an object detection problem. Although many object detection algorithms have been proposed in the literature, the inference phase of these methods when deployed on resource-constrained embedded platforms such as on-edge devices mounted on UAVs is time- and

TABLE II
COMPARISON OF NETWORKS IN TERMS OF ACCURACY METRICS, PERFORMANCE, AND PARAMETER EFFICIENCY.

Model	**Parameters** (M)	**mAP** (%)	**AP** (%) intact, T	**AP** (%) leaning, D1	**AP** (%) fallen, D2	**IoU** (%)	**FPPI**	**MR**	**APP** (mAP/MP)	**Memory** (MB)	**FPS** (1/second)
PDP-CNN	0.31	86.1	80.2 ± 0.1	88.5 ± 3.0	89.5 ± 2.2	64.2	0.18	0.14	2.77	0.815	28.1
VGG16 [24]	16.0	83.4	70.6 ± 3.5	90.0 ± 1.8	89.6 ± 1.55	60.4	0.22	0.11	0.05	63.6	4.5
ResNet-25 [26]	5.40	79.8	72.4 ± 0.8	74.7 ± 2.8	92.3 ± 0.48	60.1	0.24	0.17	0.15	21.3	14.1
ResNet-50 [26]	28.3	85.1	71.9 ± 4.5	88.0 ± 3.1	95.4 ± 0.74	61.7	0.19	0.13	0.03	113.26	6.1
MobileNetV1 [28]	5.60	83.6	92.4 ± 3.6	82.6 ± 2.1	75.8 ± 0.88	60.3	0.28	0.15	0.15	22.3	15.1
MobileNetV2 [29]	2.30	84.7	70.9 ± 4.7	90.6 ± 0.82	92.5 ± 0.95	62.5	0.27	0.13	0.36	21.3	16.3
YOLOv5 [20]	62.0	87.3	92.9 ± 1.0	74.9 ± 1.8	93.9 ± 2.3	65.0	0.16	0.12	0.01	247.7	3.8
EfficientNet-B0 [37]	5.50	85.2	85.8 ± 2.5	70.9 ± 2.6	98.8 ± 2.9	58.0	0.37	0.14	0.15	34.7	12.5
YOLOX [23]	54.2	83.5	88.9 ± 2.6	78.0 ± 0.33	83.6 ± 4.76	61.8	0.19	0.15	0.01	213.7	3.5
YOLOX-Tiny [23]	5.06	81.5	80.4 ± 1.53	75.1 ± 2.37	89.0 ± 4.0	56.0	0.39	0.14	0.10	25.4	14.2

resource-consuming and therefore is impractical.To overcome these performance-limiting disadvantages, in this study we introduced a lightweight deep learning-based computer vision model called PDP-CNN. PDP-CNN consists of a specially designed feature extractor with short dense connections, a feature pyramid network that leverages multi-scale feature maps, and an anchor-free detection backend. We conducted training, validation, and testing of the PDP-CNN model using post-hurricane reconnaissance imagery data acquired by structural extreme events reconnaissance (StEER) from over three historic hurricanes. Three damage categories of intact, leaning, and fallen were defined for the distribution poles. The results from performance evaluation on the test dataset showed that the proposed PDP-CNN architecture provides accuracy on par with state-of-the-art object detection methods while substantially outperforming these methods in terms of throughput. The outcomes of this study can facilitate rapid autonomous damage assessment of distribution poles and enhance the resilience of power distribution systems. With the recent advancements in transfer learning, the developed methodology can be extended to achieve competitive accuracy and efficiency in the task of detecting damages in other key assets of power distribution and transmission systems and for other hazards.

REFERENCES

[1] Executive Office of the President, "Economic Benefits of Increasing Electric Grid Resilience to Weather Outages," *IEEE USA Books & eBooks*, 2013.

[2] Climate Central, "Power OFF: Extreme Weather and Power Outages | Climate Central," Sept. 2020.

[3] T. Gwaltney, "Florida Power & Light Company Grid Hardening and Hurricane Response," *United States Department of Energy*, p. 22, 2018.

[4] United States Government Accountability Office (GAO), "Electricity and grid resilience: Climate change Is expected to have far-reaching effects and DOE and FERC should take actions," *GAO-21-423T*, no. GAO-21-423T, 2021.

[5] A. Kwasinski, F. Andrade, M. J. Castro-Sitiriche, and E. O'Neill-Carrillo, "Hurricane maria effects on puerto rico electric power infrastructure," *IEEE Power and Energy Technology Systems Journal*, vol. 6, no. 1, pp. 85–94, 2019. Publisher: IEEE.

[6] P. Hoffman and W. Bryan, "Comparing the impacts of northeast hurricanes on energy infrastructure," *Office of Electricity Delivery and Energy Reliability, US Dept. of Energy, Washington, DC*, 2013.

[7] Department of Energy (DOE), "Biden Administration Launches $2.3 Billion Program to Strengthen and Modernize America's Power Grid," Apr. 2022.

[8] N. L. Dehghani and A. Shafieezadeh, "Multi-stage Resilience Management of Smart Power Distribution Systems: A Stochastic Robust Optimization Model," *IEEE Transactions on Smart Grid*, 2022. Publisher: IEEE.

[9] N. L. Dehghani, A. B. Jeddi, and A. Shafieezadeh, "Intelligent hurricane resilience enhancement of power distribution systems via deep reinforcement learning," *Applied Energy*, vol. 285, p. 116355, 2021. Publisher: Elsevier.

[10] A. B. Jeddi and A. Shafieezadeh, "A Physics-Informed Graph Attention-based Approach for Power Flow Analysis," in *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 1634–1640, IEEE, 2021.

[11] H. Zhang, L. Wu, Y. Chen, R. Chen, S. Kong, Y. Wang, J. Hu, and J. Wu, "Attention-Guided Multitask Convolutional Neural Network for Power Line Parts Detection," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–13, 2022. Publisher: IEEE.

[12] R. Jenssen and D. Roverso, "Intelligent monitoring and inspection of power line components powered by UAVs and deep learning," *IEEE Power and energy technology systems journal*, vol. 6, no. 1, pp. 11–21, 2019. Publisher: IEEE.

[13] M. M. Hosseini, A. Umunnakwe, M. Parvania, and T. Tasdizen, "Intelligent damage classification and estimation in power distribution poles using unmanned aerial vehicles and convolutional neural networks," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 3325–3333, 2020. Publisher: IEEE.

[14] J. Kim, M. Kamari, S. Lee, and Y. Ham, "Large-scale visual data–driven probabilistic risk assessment of utility poles regarding the vulnerability of power distribution infrastructure systems," *Journal of Construction Engineering and Management*, vol. 147, no. 10, p. 04021121, 2021. Publisher: American Society of Civil Engineers.

[15] X. Zhang, Y. Zhang, J. Liu, C. Zhang, X. Xue, H. Zhang, and W. Zhang, "InsuDet: A fault detection method for insulators of overhead transmission lines using convolutional neural networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–12, 2021. Publisher: IEEE.

[16] L. She, Y. Fan, J. Wang, L. Cai, J. Xue, and M. Xu, "Insulator surface breakage recognition based on multiscale residual neural network," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–9, 2021. Publisher: IEEE.

[17] M. M. Alam, Z. Zhu, B. Eren Tokgoz, J. Zhang, and S. Hwang, "Automatic assessment and prediction of the resilience of utility poles using unmanned aerial vehicles and computer vision techniques," *International Journal of Disaster Risk Science*, vol. 11, no. 1, pp. 119–132, 2020. Publisher: Springer.

[18] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.

[19] Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE transactions on neural networks and learning systems*, vol. 30, no. 11, pp. 3212–3232, 2019. Publisher: IEEE.

[20] G. Jocher, A. Chaurasia, A. Stoken, J. Borovec, NanoCode012, Y. Kwon, TaoXie, J. Fang, imyhxy, K. Michael, Lorna, A. V, D. Montes, J. Nadar, Laughing, tkianai, yxNONG, P. Skalski, Z. Wang, A. Hogan, C. Fati, L. Mammana, AlexWang1900, D. Patel, D. Yiwei, F. You, J. Hajek, L. Diaconu, and M. T. Minh, "ultralytics/yolov5: v6.1 - TensorRT, TensorFlow Edge TPU and OpenVINO Export and Inference," Feb. 2022.

[21] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *European conference on computer vision*, pp. 21–37, Springer, 2016.

[22] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[23] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO Series in 2021," Aug. 2021. arXiv:2107.08430 [cs].

[24] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[25] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.

[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[27] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.

[28] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[29] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4510–4520, 2018.

[30] G. Huang, D. Chen, T. Li, F. Wu, L. Van Der Maaten, and K. Q. Weinberger, "Multi-scale dense networks for resource efficient image classification," *arXiv preprint arXiv:1703.09844*, 2017.

[31] Y. Li, A. Dua, and F. Ren, "Light-Weight RetinaNet for Object Detection on Edge Devices," in *2020 IEEE 6th World Forum on Internet of Things (WF-IoT)*, pp. 1–6, IEEE, 2020.

[32] Y. He, K. Song, Q. Meng, and Y. Yan, "An end-to-end steel surface defect detection approach via fusing multiple hierarchical features," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 4, pp. 1493–1504, 2019. Publisher: IEEE.

[33] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *European conference on computer vision*, pp. 213–229, Springer, 2020.

[34] "Home | StEER Website."

[35] A. L. Salman, "StEER - Hurricane Dorian: Field Assessment Structural Team (FAST-1) Early Access Reconnaissance Report (EARR)," Oct. 2019. Publisher: DesignSafe-CI Type: dataset.

[36] D. Roueche, J. Cleary, K. Gurley, J. Marshall, J.-P. Pinelli, D. Prevatt, D. Smith, A. Alipour, K. Angeles, B. Davis, C. Gonzalez, A. Lenjani, H. Mulchandani, M. Musetich, A. Salman, T. Kijewski-Correa, I. Robertson, and K. Mosalam, "StEER - HURRICANE MICHAEL: FIELD ASSESSMENT TEAM 1 (FAT-1) EARLY ACCESS RECONNAISSANCE REPORT (EARR)," Oct. 2018. Publisher: DesignSafe-CI Type: dataset.

[37] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*, pp. 6105–6114, PMLR, 2019.

Ashkan B. Jeddi received a B.S. degree in civil engineering from the Iran University of Science and Technology, Tehran, Iran, in 2015., and an M.S. in structural engineering from Sharif University of Technology, Tehran, Iran, in 2017. He is a Graduate Research Associate in the Department of Civil, Environmental and Geodetic Engineering, The Ohio State University, Columbus, OH, USA. His primary research interests include the application of machine learning techniques for the risk and resilience assessment of structural and infrastructure systems.

Abdollah Shafieezadeh received the B.S. and M.S. degrees in civil engineering from the University of Tehran, Tehran, Iran, in 2002 and 2006, respectively. He received a M.S. degree in structural engineering from the Utah State University, Logan, UT, USA, in 2008 and the Ph.D. degree in structural engineering with a minor in mathematics from the Georgia Institute of Technology, Atlanta, GA, USA, in 2011. He is the Lichtenstein Associate Professor in the Department of Civil, Environmental and Geodetic Engineering at The Ohio State University, Columbus, OH, USA. His primary research interests are in uncertainty quantification using machine learning, and resilience quantification and optimal management of infrastructure systems.

Roshanak Nateghi received the bachelor's degree in mechanical engineering from the Imperial College London, in 2006, and the M.S.E. and Ph.D. degrees in environmental engineering from Johns Hopkins University, in 2009 and 2012, respectively. Prior to joining Purdue in 2015, she was an NSF Science, Engineering and Education for Sustainability Fellow, jointly appointed between Johns Hopkins University and Resources for the Future. She is currently an Associate Professor in industrial engineering with Purdue University. Her research interests include developing interdisciplinary methods to model the sustainability, risk and resilience of critical infrastructure under natural hazards, and climate change.