Deep Learning for Component Fault Detection in Electricity Transmission Lines.

Iyke Maduako (imaduako@unicef.org)
University of New Brunswick Fredericton https://orcid.org/0000-0002-7260-3666
Chukwuemeka Fortune Igwe
NOVA University of Lisbon
James Edebo Abah
University of Nigeria - Enugu Campus
Obianuju Esther Onwuasoanya
University of Nigeria - Enugu Campus
Grace Amarachi Chukwu
University of Nigeria - Enugu Campus
Franklin Ezeji
Federal University of Technology Owerri
Francis Okeke
University of Nigeria Enugu Campus

Research

Keywords: Power-line Fault Detection, Deep Learning, UAV Imagery, Single Shot Multibox Detector.

Posted Date: November 17th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1028973/v1

Abstract

Fault identification is one of the most significant bottlenecks faced by electricity transmission and distribution utilities in developing countries to deliver efficient services to the customers and ensure proper asset audit and management for network optimization and load forecasting. This is due to data scarcity, asset inaccessibility and insecurity, ground-surveys complexity, untimeliness, and general human cost. In view of this, we exploited the use of oblique UAV imagery with a high spatial resolution and a fine-tuned and deep Convolutional Neural Networks (CNNs) to monitor four major Electric power transmission network (EPTN) components. This study explored the capability of the Single Shot Multibox Detector (SSD), a one-stage object detection model on the electric transmission power line imagery to localize, detect and classify faults. The fault considered in this study include the broken insulator plate, missing insulator plate, missing knob, and rusty clamp. Our adapted neural network is a CNN based on a multiscale layer feature pyramid network (FPN) using aerial image patches and ground truth to localise and detect faults via a one-phase procedure. The SSD Rest50 architecture variation performed the best with a mean Average Precision (mAP) of 89.61%. All the developed SSD based models achieve a high precision rate and low recall rate in detecting the faulty components, thus achieving acceptable balance levels of F1-score and representation. Finally, comparable to other works in literature within this same domain, deep-learning will boost timeliness of EPTN inspection and their component fault mapping in the long - run if these deep learning architectures are widely understood, adequate training samples exist to represent multiple fault characteristics; and the effects of augmenting available datasets, balancing intra-class heterogeneity, and small-scale datasets are clearly understood.

1. Introduction

Regular inspection of electric power lines has become an essential concern because virtually all human communities, processes, and mechanisms rely on electricity [1]. Generally, in many developing countries, the available electricity is unreliably characterised by households and businesses experiencing long and frequent power outages resulting from electricity demand exceeding available electricity supply caused by load shedding and/or technical failures [2], [3]. For example, Electric Utilities in Nigeria claimed that: some sections of the grid are outdated with inadequate redundancies; regular vandalization of the lines associated with low level of surveillance and security on all electrical infrastructure, and the serious lack of required modern technologies for communication and monitoring is causing more and more power outages [4], [5]. To tackle these challenges, there are different approaches, that have been developed for fault detection on the power transmission lines. Among these methods is the use of Machine Learning techniques on Very High Resolution (VHR) satellite imagery. This method has proven to be, more efficient and outperform manual inspection and traditional data analysis approach for detecting faults in power transmission line at large.

Remote sensing techniques have been very efficient in power line corrosion and mechanical loss detection. Inspection of electricity power transmission network (EPTN) especially in remote areas using remote sensing techniques requires very high-resolution imageries such as those gotten from aerial

survey, UAV images and Lidar point clouds data. Unmanned Area Vehicle (UAV) surveillance has become the state-of-the-art in power line inspection for defect and damage [6]. Many studies have also demonstrated the efficacy of high-resolution remote sensing techniques in power line inspection and monitoring. For example, Xue et al. [7], utilized SAR imagery to measure electricity towers' damage caused by landslide. Based on pixel resolutions, morphology algorithm, and location of the damage, the authors detected, and geotagged power lines damaged by landslides directly.

The use of high resolution TerraSAR-X imagery to track power line damages in natural disaster situations was discussed by Yan et al. [8]. SAR imagery was preferred to optical imagery as the author points out that SAR geometry makes it suitable to detect vertical, human-made objects, such as power towers. Sentinel-1 SAR satellite and other very high resolution (VHR) SAR images (TerraSAR, RADARSAR) that have spatial resolution higher than five metres have recently generated a lot of interest for surveillance, tracking, and monitoring of power transmission systems [9]. However, VHR SAR for EPTN seems impacted by distortions in the imagery, particularly pseudo-random variation of the different components' imprints; hence, making it semantically challenging to interpret [10]. They are also limited by coarse resolution to detect small defects in electricity transmission components, geometric deformations, strong noise-like effect creating false representations, and multi-path scattering [6], [7]. In light of this, various studies have adopted the use of multispectral images acquired from optical satellite remote sensing, as it allows for straightforward interpretation. However, these data are still restricted by lack of adequate spatial resolution for power line inspection.

Optical remote sensing has focused on fault diagnosis for the different EPTN components themselves because the ground sample data (GSD) is usually less than the individual components' size, especially for those caused by the adjoining environment. As a result, most power line inspection study using optical remote sensing is fixated on vegetation encroachment and minimum height and clearance distance [11], [12]. Additionally, a stereo pair of optical satellite images have been utilized to extract the canopy height model to monitor damages to EPTN caused by vegetation [13]. This has allowed the identification of individual overgrown trees affecting power lines with high accuracy. The studies on vegetation encroachment faults affecting transmission lines from satellite and aerial images have habitually utilized the classification of trees, vegetation indices, and segmentation approach. Vegetation invasion on transmission lines was studied by Ahmad et al. [12]. The paper explored the use of multispectral satellite stereo imagery to simulate transmission lines using a 3D digital elevation model (DEM) to detect dangerous vegetation branches that could affect power lines and cause blackouts. Apart from vegetation encroachment, a variety of papers addressed automatic inspection of insulators' condition. These techniques aimed to take images of the insulators periodically and use automated classification methods to identify damaged insulators. Reddy et al. [14], for example, used fixed cameras on poles. Jiang et al. [15], using a photogrammetric method, addressed flashover faults - pollution-related flashes affecting insulators. In the experiment, a sensing camera placed on a tripod was used. However, most remote optical sensing techniques are primarily restricted by the atmosphere. Consequently, using the Lidar method through airborne laser scanning or mobile laser scanning has also been used to

improve the shortcomings of multispectral optical images to detect and identify tall trees that may collapse through the conductor and exceed the required vegetation clearance.

Despite extensive studies on powerline inspection and fault detection, the advantages of using remote sensing in Sub-Saharan Africa remain unseen due to the data unavailability and peculiarity of the power line in this region. Many utility companies and investors rely on poorly collected data from ground-based surveys, multispectral visible colour images, and most recently video surveillance of transmission line fault inspection and monitoring [16]. UAV monitoring offers high-spatial multispectral images that deal with the limitation of other remote sensing methods because of the ability to capture accurate images of transmissions components at closer proximity [17]. UAVs are able to detect small-scale defects such as broken fittings and missing knobs and can be incorporated with other modes of remote sensing. In comparison to manual methods, with limited resources and man-hours, inspecting and monitoring long transmission line corridors for potential faults and damages becomes almost impossible.

For cost and time-effectiveness, electricity infrastructure inspection and fault diagnosis especially in transmission lines, the combination of UAV data and deep-learning techniques is imperative [6]. The advent of deep learning, which uses not only spectral information but also spatial, topological, spectral, and geometric properties of objects in images, is at the forefront of these developments. Deep learning has demonstrated potential promising advances in power line extraction and other study fields. Currently, improved algorithms and multilayer networks such as Convolution Neural Networks (CNNs), Recurrent Neural Networks (RNNs) and reinforcement learning have demonstrated more outstanding performance than standard approaches, particularly in power line identification, transmission components detection, as well as in, vegetation encroachment prevention [18]. Conversely, the traditional approach for pattern recognition depends on the continuous engineering of parameters that are well built by humans. Hence, making the manual extraction process inefficient, unfavourable, inadequate for generalization necessities and time consuming. With deep learning algorithms, visual perception to extract feature hierarchies and generalization ability is enhanced on several levels [19]. These algorithms have demonstrated that conventional learning methods are sluggish and unreliable; they require substantial post-processing attempts to differentiate between transmission infrastructure [20]. Succinctly, power transmission network mapping and fault inspection requires a more advanced adequate hybrid classifier that is way beyond task-based approaches, promoting the improved performance of visual recognition tasks and successfully adapts learning from multimodal data sensors for object detection.

Taking all these considerations into account, the use of deep learning technologies, together with the advantages of Unmanned Aerial Vehicles (UAVs), can play a fundamental role in relieving the current limitations of traditional power transmission monitoring, which are mainly based on manual operations such as electricity pole climbing, foot patrols, vehicle inspections, and field verification reports [21], [22]. Nevertheless, the operational applicability of Deep learning techniques on UAV surveillance has not yet been addressed in the context of power infrastructure inspection in developing countries, which is precisely the gap that motivates this work. This paper explores the use of deep learning techniques for

power line fault detection and inspection from UAV imagery through the lens of a case study in Nigeria. The major contributions of this paper are:

1. The development of a single-phase deep learning model for power line faulty component detection and classification pipeline for a series of faults (multiclass) that typically exists in power transmission components. Based on our understanding, very little research has been carried out in this area with the context of providing a contribution to the research communities and electricity infrastructure monitoring and management.
2. Exploring the feasibility of low-cost drone equipment to monitor electricity transmission infrastructure for faulty-component detection.
3. Empirical and comparative analysis of hyperparameters in CNN backbone architectures consisting of more than one electricity power line component-fault types to evaluate the effectiveness of our proposed approach.

2. Related Work

Most widely used models in literature for fault detection in power line components primarily involve clustering, mathematical-based techniques such as Hough transform, Gabon filters, knowledge-based techniques, and low-level filters. In detecting broken transmission line spacers, a Canny edge detector combined with Hough transform was used by Song et al. [23]. First, a scan window was formed in the path of the conductor and during the convolution process, if there are a candidate's spacers, they are recognized in all sliding windows. Finally, the shape configuration parameter is structured to decide whether the sensed spacer was broken based on the measurement of linked parts. Zhai et al. [24] exploited Saliency Aggregating Faster Pixel-wise Image (FPISA) for insulator extraction. Based on the colour channel in Lab colour space, the observed insulator's flashover region was extracted. The system was tested using 100 flashover fault insulating images and obtained a detection rate of 92.7%. Han et al. [26] and Zhai et al. [25] detected the faults associated with the missing cap of insulators based on saliency and adaptive morphology (S-AM). However, these approaches rely on traditional pattern recognition techniques, which are cumbersome and time consuming.

The most challenging type of faults to detect in power transmission line are faults with tiny aspect ratio on the EPTN components, for instance, power line fitting such as missing pin, nut, bolts, and a small degree of fault severity on some large components. To detect such kinds of faults, aerial images are captured close to the exact components containing the faults or the components (or faults) cropped from the original image manually [22], automatically, or via segmentation [27]. Fu et al. [28] implemented a dynamic model for the missing pin type of faults. The fitting is a combination of multiple sections, which include pin and nut. The haar-like attribute and Adaboost classifier was used to detect each part of the fitting. The methodology involved first extracting the segmented region and circles with LSD and Hough transform, respectively, to identify the missing pin. The missing pin fault was finally obtained and then observed based on the distance limit between the centre of the circle and the pin section. Other methods for the automated identification and monitoring of the electric power transmission network (EPTN) faults

have been implemented in recent years using supervised classification. The most used machine learning-based algorithms used as the feature classifier, primarily include Adaboost [28], and SVM [29], which have been applied successfully to detect foreign bodies on the conductor. These techniques have contributed to improve the accuracy of electrical transmission fault inspection. However, since they often include additional feature engineering, they become less attractive especially in imagery dataset.

One of the earliest works on fault detection using deep learning was detecting surface discoloration due to flashover on insulator using CNN classifier with pre-trained AlexNet published by Zhao et al. [30]. The experiments achieved a score of 98.71% mean Average Precision (mAP) on 1000 samples. The proposed architecture outperformed the conventional handcrafted approach but was limited to insulator condition inspection image classification, which demanded significant feature engineering. Additionally, Faster R-CNN was presented by Liu et al. [31] to identify insulators with missing caps. The system was tested for three different voltage transmission line levels with 1,000 training samples and 500 research samples prepared for each level. About 120 photographs (80 for training) were used to test the diagnosis of missing cap fault. The study also highlights the possibility of overfitting as data is small and employs data augmentation to physically extend the dataset. To handle similar faults across multi-scale level images, Jiang et al. [32] developed a novel approach using SSD as the meta-architecture for multi-level perception (low, mid, and high perception) based on ensemble learning to extract the missing insulator fault from the image resolution of 1920 × 1080-pixel. The middle and high-level perception images are made via the Region of Interests (ROIs) Union Extraction (RUE) image pre-processing. The proposed approach's absolute precision and recall rates were 93.69% and 91.23% respectively on the test image dataset with various perception levels containing missing cap insulator problems. However, these papers considered the contextual characteristics of one type of fault inspection that affect the insulator component across the transmission corridor neglecting other defects that coexist. In most cases, the features derived by such methods may not adequately reflect the insulators, and these approaches may need the imagery modified.

One particular issue in power line fault detection using deep learning CNN is data insufficiency. This is because the DL model is required to generalize the solution at the end of the training. To achieve this, a robust and large amount of dataset is usually required. In the previous papers to circumvent potential data shortages and expedite the creation of a reliable predictive model, attempts were made to synthesize the images (e.g., in [33]) and data augmentation (e.g., in [34]; [31]). Other researchers have examined the use of transfer learning and few-shot learning to identify fault types. For instance, for lack of sufficient training images, Bai et al. [35] utilized a transfer learning process using the ImageNet data kit, which included a 1.2 million samples dataset. This model was then trained, i.e., fine-tuned by the limited data set acquired containing the surface fault of insulators based on the Spatial Pyramid Pooling networks (SPP-Net) with transfer learning approaches. This allowed the weight optimization to begin at top layers (where there is a different feature complexity from the original training data utilized) in the 3D CNN of the SPP-Net adopted rather than for the whole model. The result showed the better performance of SPP-Net architecture with transfer learning over the RGB (red, green, and blue) imagery in a short computation

time. Although this model proved sufficient, the result was limited to a classification problem involving just the insulator fault.

In recent years, there have been few efforts to develop a deep learning approach, to identify several power lines faults simultaneously. Typically, a two-step object detection technique is commonly utilized: first, to identify the component and second, to detect the fault in those components. In this light, Tao et al. [33] developed two separate backbone models, Defect Detector Network (DDN) DDN and Insulator localizer Network (ILN) based on Visual Geometry Group (VGG) model and Residual Network (ResNet) model respectively, on the domain knowledge of the EPTN component's structure. To find a missing cap fault, a cascading architecture combining a custom developed ILN and a DDN models were utilized. The ILN identifies all the insulators in the aerial image and then cuts the detected areas and feeds them into the DDN. A total of 900 regular images were collected from UAV for this experiment and 60 defective images. Data insufficiency was tackled by segmenting the image using the U-net algorithm to divide the output of the ILN into insulator and background. The segmented insulator was then combined with distinct images of different backgrounds to mimic real-life background situations concerning insulator position. The result of this was then merged as input for the DDN model. Finally, about 1956 pictures for ILN (1186 for training) and 1056 images with missing caps (782 for training) were prepared. The DDN detection precision and recall are 0.91and 0.96. The resulting accuracy outperformed the direct use of existing frameworks. However, most related studies do not consider a single-phase approach and do not detect more than one fault simultaneously, but rather focus on video surveillance and single class fault detection on the transmission lines. Exploring the performance of different object detection deep learning models, the SSD meta-architecture utilized by Jiang et al. [32] performed well considering the multiscale camera imagery perception and model characteristics, SSD, therefore, will be utilized in this case study.

3. Study Area And Dataset

The sub-section 3.1 provides background into the study area chosen for this research. The second section (sub-section 3.2) describes the electricity transmission line dataset generated for this study. Finally, the four types of electricity transmission line faults considered, are described in the last sub-section, 3.3.

3.1 Study Area

The study area is made up of four different transmission line corridors of Nigeria. Different transmission line corridors were explored for feasibility regarding this study with the help of power-line engineers and photographs from reconnaissance surveys. Six transmission line corridors were investigated in total, and four corridors were selected after reconnaissance by the ground truth team. This decision was made based on the sites' usability for field experiments and based on the spatial resolution of the acquired imagery. These transmission corridors virtually have a connection with all the 36 states in Nigeria and the Federal Capital Territory. Nigeria lies between latitudes 4° and 14°N, and longitudes 2° and 15°E. The Nigerian power transmission network called the Transmission Company of Nigeria (TCN), is responsible

for the transmission of power in two phases, the 330kV - 132 kV and the 132kV-33kV through the transmission lines (otherwise referred to as conductors) [59]. In general, all transmission corridor in Nigeria shares similar structure, their infrastructure is radial and thus causes inherent problems without redundancies [36].

3.2 Datasets

The DJI Phantom (DJI FC330) fitted with high-resolution cameras was flown across the four transmission corridors namely, Shiroro-Kaduna, Lagos, Abuja, and Enugu overhead transmission lines to capture pylons, conductors, other components of power line/pylon accessories (e.g., insulators, fittings, cross arms) as well as the surrounding features (e.g., vegetation) from varying angles. The imagery is in three spectral bands (visible RGB) with high spatial resolution. The aerial survey was conducted from October 12, 2020, to October 22, 2020. A total of 140 large images tiles contained the study area and can be characterised as high-resolution oblique RGB images of dimension 4000 x 3000 pixels (72dpi). The mean pixel sensor resolution is 0.00124m. Generally, within the images, the most prominent objects are located and systematically distributed transmission conductor and pylons with dirt roads, small patches of natural forest, and grasslands.

3.3 Taxonomy of faults

The main purpose of this process is to classify the faults found in the transmission components. Each transmission line component like pylons, conductors, and pylon accessories or fittings (e.g., insulators, dampers, and fixtures), has different and unique faults.

Transmission line pylons are used to extend of the conductors over long distances, supporting lightning safety cables and other transmission elements. They ensure proper electrical transmission process of the other components by preserving the original design positioning and provide sufficient grounding against adjoining objects. Insulators are critical elements in a transmission line as they protect conductors by allowing lines to retain their expected electrical insulation strength [37]. As seen in Figure 2, the insulator has a repetitive, stacked cap structure. The colour, size, and string numbers of the insulators vary based on the transmission capacity and manufacturing design (e.g., single string and double strings). The pylon accessories, also called fittings, are the connectors of major components or elements seen in the electricity transmission lines. They mainly serve as support, inhibitors, connectors to the other transmission components. These include conductor clamps, dampers, splicing fitting, protective fittings, and guy wire fittings.

Consequently, most of these individual components have many different types of faults. For this research, the defects were divided taxonomically into four categories: missing insulator, broken insulator, rusty clamp, and broken dampers according to the contents of the captured aerial photographs. The detailed fault taxonomy discussed in this study is as follow:

i. Missing Insulator: These are glass insulators with missing insulator cap (plate); see Figure 1.

ii. Broken Insulator: This applies to those insulators that are made of porcelain or composite polymer plate or cap materials. In this case, the plate is incompletely destroyed by pressure exerted by external forces such as weather, especially thunder-strike and thaw.
iii. Rusty Clamp: The conductor clamp (strain or suspension clamp) helps to hold all components, especially the insulator, to the tower architecture based on its design. A faulty clamp can lead to the insulator's total malfunction, hence leading to transmission collapse.
iv. Broken Fitting: Broken fittings such as shown below in figure 4 where, the vibration damper is broken which could cause conductor fatigue and strand breakage.

4. Methodology

This section outlines the approaches and considerations for developing a predictive model for transmission line fault detection from high resolution imagery. This section also provides a description and architecture of actual Object detection and single shot detection models developed and designed for this study.

4.1. Convolutional Neural Networks

Convolutional Neural Networks (CNNs), which are specialized neural networks developed to exploit the two-dimensional nature of images, have in recent years advanced deep learning tasks (high-level vision) such as image classification, object detection, and image segmentation, as well as low-level vision tasks such as edge detection [38]. The deep learning tasks (deep convNet) was first developed for image classification problem based on the performance of convolution layers to recognize edges, patterns, context, and shapes resulting in a convolution feature map having spatial dimensions smaller and deeper than the original [39]. The progenitor of image classification architecture otherwise known as feature extractor in object detection problem is AlexNet with an 8-layer CNN, i.e., 5 convolutional layers + 3 fully connected layers developed by Krizhevsky et al. [40] in Imagenet challenge of 2012. Many improvements have been made to the architecture of Krizhevsky et al. [40] over the years. These include using a narrower receptive window and increasing the network depth.

Similarly, from the 2014 ImageNet contest, VGGNet metamorphosed with the aim of improving the work developed by Krizhevsky et al. [40]. This CNN architecture took first place in the localisation task and second place in the classification task. VGGNet's breakthrough is the mixture of kernel filters (3 x 3 filters) and deep neural networks (16-19 layers). The authors believed that 3 × 3 convolution layers have the same efficient receptive area as the 7 × 7 convolution layer, however, VGGNet's architecture is wider, with larger non-linearities, and fewer parameters to update [41]. This solidifies the concept that the best way to maximize the performance of CNNs is by increasing the depth and width of the CNNs.

Complexity of image classification problems increasingly calls for larger CNNs. However, deep CNNs with several layers can be difficult to train because of the problem of vanishing and exploding gradients. To handle this problem, the residual network learning called ResNet gained traction. Residual networks were

built with shortcuts to whole networks inspired by VGG networks by the subject of skipping [42]. To dissociate with the concept of increasing depth when creating CNN architecture, ResNet proposed a shallower network using shortcut connections, i.e., directly connecting the early layer's input to a later layer. The significant ability to train very deep CNNs in 50, 101, and 152 layers with great successful connections are attributed to the regular cut-off's connection (skipping) among the Deep CNN blocks [42].

The general tendency for network speeds has been to go deeper and more complex. This results in extended preparation and higher computing costs [43]. The aim of making low-latency models for mobile and embedded devices led Howard & Wang [44] to develop a lightweight deep neural network model referred to as Mobile networks (MobileNets). MobileNets and its derivatives were implemented to substitute a much deeper network constrained by the speed in achieving satisfactory output and real-time applications. This design's idea is that the regular neural network convolution layer is broken down into two filters, depth-wise convolution, and pointwise convolution [44]. The usual convolutional filter is more computationally complicated than depth-wise and pointwise convolutions. To achieve this model implementation, each channel is convolved with its kernel, called a depthwise convolution. Next, the pointwise (1 × 1) convolution is processed to abstract and integrate the individual intermediate output from the depth-wise convolution into a single feature layer.

Inspired by the success of CNNs in image classification and the need to adapt CNNs for more complex task other than classification problem, the object detection approach was conceived, which comprises of the classification of objects and finding object of interest positions in the image via regression. In line with this thought, the Faster R-CNN was developed utilizing a region-based CNN. Faster R-CNN performs object detection uses two major modules: a Regional Proposal Network (RPN) proposing regions, and a Region-CNN (R-CNN) detector classifying regions and refining boundary boxes. The model involves first the use of a base network, i.e., CNN architecture pretrained for classification to generate the necessary activation feature map [45]. Then, the extracted feature maps are passed through the RPN to generate object proposal. Each object proposal from the RPN, are then applied in the network by overlapping them over the existing convolutional feature map. This extracts various fixed feature maps of field of interest for each proposal. The final Region-based CNNs (R-CNN) incorporates the preceding output with class details based on regions proposal. Using the object proposals extracted via RPN and the extracted features for any one of the proposals (via ROI pooling), a final class and object localisation is achieved [45]. R-CNN is a model which attempts to simulate the final phases of CNN classification where a totally flattened layer is applied to generate a score for each conceivable object form [34]. R-CNN has two separate objectives: classify the proposal and modify the bounding box for the proposal according to the predicted class. Although faster R-CNN is extremely reliable but, it is very slow.

In the same vein, Region-based Fully Convolutional Network (R-FCN) was developed by Dai et al. [46] to tackle the shortcomings of the initially designed Faster R-CNN frameworks. Instead of using an inefficient sub network for each region hundreds of times, R-FCN adopts an entirely fully convolutional architecture over the whole image. In a way that allows totally network convolutions to carry out one calculation in detail and accurately, the R-FCN provides new location sensitive scoring maps. Also, the issue between

translation invariance and translation difference in recognising objects are addressed more effectively. Therefore, R-FCN integrates feature maps and applies convolution to construct positions sensitive score maps, which enable convolutional networks to successfully perform both classification and detection in a single evaluation. The position-sensitive ROI pooling is used to produce a vote array of the output size for any ROI to achieve a 2D score maps of each class. For regression of the boundary box, another convolution filter is used to construct a 3D output map on the final feature maps. Then, the ROI-pooling is used to measure a 2D array with each element that includes a boundary. The sum of these elements is the final bounding box estimate [46]. RFCN presents new position-invariant spatial scores which enable convolutional networks to successfully perform both classification and detection in a single evaluation. R-FCN incorporating these enhancements allows the framework to run faster about 2 to 20 times the speed and have better accuracy; therefore, the frameworks are quick and precise but have complicated pipelines.

To aid in real-time object detection maintaining a balance between time, speed and accuracy, many single-phase deep learning-based approaches, which detect multiple objects in a single scan, have been proposed. The two most popular single-shot models are the 'You Only Look Once' (YOLO) and Single-shot detector (SSD) frameworks. YOLO is a network that classifies bounding boxes in real-time [47]. To fulfil this, YOLO combines area proposal and region classification to form a single network and does this as the frame is simply regressing on box localization and related probabilities. YOLO uses a grid that separates the input image. The grids evaluate the bounding box position, assign confidence ratings, and conditional class probabilities. YOLO is incredibly fast because it is a single-threaded; however, YOLO lacks the precision seen in the two-phase frameworks such as R-FCN and Faster R-CNN preciously discussed. The SSD is a better approach as it is focused on a feed-forward-based convolution network generating a fixed-size bounding box set and scores of object instances present in these boxes, and a final detection process based on a Non-maximum Suppression (NMS) criterion [48]. The early network layers are constructed on a standard image-classification architecture known as the base network (i.e., the classification layer without the flattened fully connected layer).

SSD supersedes its counterpart, YOLO, by introducing several modifications: (i) multi-feature maps from subsequent networking stage are predicted to allow multiscale detection; (ii) object classes and offsets at bounding box locations are predicted using regular sized small convolutional filter; and (iii) after deriving final feature map, different predictors (classifiers) are used to identify objects at varying aspect ratios in the form of feature pyramids [49]. SSDs comprises two main parts: a feature map extractor and the convolution filter for object detection. SSD attaches additional convolutional layers (feature layers), i.e., multiscale features and default boxes, which causes a steady decrease in size up to the end of the primary network [48]. Hence, the predictions of detected objects are produced at multiple levels. Unlike YOLO, which uses a fully connected layer to make predictions, the SSD adds a series of small convolutional filters to each added feature layer (or an existing one in the base network optionally) and uses them in boundary box positions to predict classes and offsets of objects [50]. These changes improve both the speed and the accuracy of SSD over YOLO.

Undoubtedly, convolutional network tasks typically have a significant role in image classification and object identification. One of the ways in which CNN achieves this high performance is via the gradient-based learning process, more specifically loss computation and the loss function [38], [51], [52]. This is believed to be the object's real value, versus expected value. For instance, if the expected value ends up being 0.75, and the actual value is 1, the loss would be 0.25. As iterations continue, the model will better approximate the object's true value. In this respect, the optimisation process is employed so that the prediction capacity can be maximized. Mathematically, this implies that for neural networks, the loss is normally the sum of negative log probability and residual sum of squares for classification and regression part, respectively [53], [54]. After that, the key goal is to mitigate the loss with respect to model parameters by modifying the weight vector values using neural networks. For all object detection models, the loss function is a combination of the localization (bounding box regression) and the confidence loss (object classification).

4.2 Data Pre-processing and labelling

Pre-processing the dataset entails a series of steps aimed at cleaning and standardizing the raw data prior to modelling. Pre-processing is critical for increasing the sensitivity of the model and validating any model that uses aerial imagery for transmission line fault detection. The entire dataset is made up of 294 images. Due to the small-scale problem identified in some research [55], the dataset (132kV) was split into about 817 tiles centred on at least one components' fault of interest. For the other dataset representing the other 33kV transmission line, the non-destructive resize, i.e., resize and pad approach, is applied to preserve the image aspect ratio to preserve the geometric and spatial information. Moreover, the split and resized RGB images were normalized to the same size of 600 x 600 pixels following Huang et al. [49], combined to form a total of 1027 'Felect' dataset sample imagery. The data is divided into train, test, and validation sets. It was assured that 17% of the original dataset was allocated for the test dataset, and 83% of the dataset was reserved for training and validation. About 80% of training was used as the training samples, while the remaining 20% was dedicated to validation samples - Table 1 displays the data slicing information. The drone captured the 'Felect' dataset with numerous characteristics, including diverse perspectives, sizes, occlusion, background clutter, and intra-class variance.

Table 1: Data partition

Dataset	#Components' faults	%Components' faults	#Images	%Images	Missing knob	Broken ins	Missing ins.	Rusty clamp
Train	1198	62,7	646	62,9	490	223	259	225
Validation	372	19,5	207	20,0	151	70	82	68
Test	340	17,8	154	16,8	142	75	49	73
Total	1027	100	1027	100	783	368	390	366

Thus, a “stratified” data division is used, making the proportion of the faulty components for the dataset similar to the number of images, as well as the average number of components and the intraclass variation shared equally for samples with different types of difficulties to be learnt and appropriately located and classified.

Data annotation was carried out to identify and label the training dataset for model training. The bounding box approach and pixel-wise object segmentation are two approaches that can be used to annotate the main object on the image manually [54]. To annotate the faults, the ground truth annotation of actual components’ fault types was generated as a rectangular bounding box was used. A tool called ‘LabelImg’ was used to label the different component faults as shown in section 0. The details of the image, bounding box, and object class, along with shared characteristics, were stored as a VOC2007/extensible mark-up language (.xml) file. After annotating all the frames, the whole split dataset containing image patches tensor and their output label were converted into a TF record-oriented binary as depicted by Figure 5 to help dataset initialization and ease network architecture using the TFRecordWriter function.

4.3 Network Training

As stated in section 4.2, input patch images are first translated to tensors (TF records) with a [600 x 600 x 3] form prior to feeding it into the backbone architecture and are distributed by the action of the convolution layer to an intermediate layer consisting of a convolutional activation map. The head of the network architecture (backbone network) typically follows the patch-based CNN architecture. Therefore, image patches that contain either a single class of faults or a combination of different components’ faults centred in the pixel of interest, also termed as valid patches, were extracted. For, backbone neural network ResNet50, MobileNet, or ResNet101 are utilized for the first part of the SSD network as the head to develop three models.

This head is made of CNN that detects smaller characteristics (patterns and corners), and later layers detect higher characteristics successively. The image was resized first into 640px x 640px x 3 (RGB) and then translated into a *38 x 38 x 512* characteristic mapping through the backbone network passed to the Conv7 denoted as SSD 1 (auxiliary layer) in Figure 6. In all experimentation cases, the input patch tensor was abstracted into multi-level representations to classify the different faults after going through the backbone architecture (Figure 7: without a fully connected layer). As a deep neural network, the backbone algorithm derives semantic significance from the image while maintaining its spatial structure.

The series of auxiliary convolutional layers (SSD layers) introduced after the SSD model’s backbone allows the extraction of features at different scales as the input feature map decreases at each successive layer. This ensures the certainty of boundary variance and class prediction of targets at various scales. For each decreasing successive auxiliary layer (multi-scale feature maps), SSD networks grids the image and assign each grid with the task of detecting objects (Figure 8). After this, 3 × 3 convolution filters are applied to each cell to make predictions. If no object appears, the context class is

not considered, and the location is ignored. Each cell in the grid will decide the location and shape of the object inside it.

Immediately after gridding the auxiliary layer, i.e., feature map at multi-level, default boxes are generated at each grid cell for each convolution layer level using a defined scale value (Figure 9). This scale increases progressively towards the least spatial resolution feature map level (SSD 5). Next, bounding boxes are generated via a process called default box generation (prior). Default boundary boxes are selected explicitly, which are pre-computed, fixed-size boxes that closely fit the ground truth boxes. With the different experiment scale value, $s_{k,}$ and the aspect ratio, $ar \in \{1.0, 2.0, 0.5\}$, the default boxes sizes are built. To detect larger objects, SSD uses lower resolution layers such as the SSD 4 and SSD 5 layers in Figure 8. Each grid prediction composition includes a boundary box defined by c_x, c_y, w, h, and four scores for each class, i.e., components faults, in the prediction, with the highest-class score associated with the positioned default bounding box. The class score corresponds to object classification labelled in this research as “broken insulator,” “missing insulator,” “missing knob,” and “rusty clamp.” Having these several forecasts at once and awarding class scores to each is referred to as the Multibox. There are four predictions for every cell, regardless of the feature map’s spatial resolution, and an extra one prediction to represent objectness.

To improve the SSD to detect small-scale faults type, the Feature Pyramid Network (FPN) training structure is used in conjunction with the most immediate output feature map activated from the base network architecture. This method also imbues low-level CNN layers with more assertive semantic representation, such as layers near its head to detect small scale object labels. In particular, the default boxes are chosen so that their Intersection over Union (IoU) is greater than 0.6.

The Sigmoid function is then performed on the output feature map generated by the last CNN to obtain a class prediction score. Thereafter, the total loss is achieved by combining the two losses obtained for backpropagation. The two new losses measured by the network for each bounding box include:

a) The localisation loss is achieved using the weighted smooth-L_1 loss, calculated by comparing the generated default boxes (prior) against GT labels.
b) The confidence loss is achieved using a similar method applied in image classification, in this case, the weighted sigmoid focal.
The default boxes that did not get scored against any ground truth boxes are viewed as negatively matchbox and are applied to only the confidence loss, while the positive box is applied to the overall loss. This loss value is backpropagated to update the network parameters using different optimizer during experimentation.

4.4 Experimental Design

The current projects’ fundamental problems were related to the number of computing resources required and dataset’s limited size. In this study, the experiments - backbone architecture and meta-architecture

were built on the top of the deep learning framework of TensorFlow Object Detection API (TF 1) Model Zoo. Two separate outlets were utilized for execution, they include:

i) A physical computer with AMD Ryzen 5 3550H with Radeon Vega Mobile *Rfx* processor CPU with 7.81 GB for data processing, preparation, and model testing.
ii) Google Colab environment on the Google cloud server with 2 Intel(R) Xeon(R) @ 2.20GHz processor CPU with 13GB RAM (200 GB free space disk) and 1 GPU (Tesla K80) with 12.6 GB RAM for parallel processing for experimentation.
To ensure optimal experimentation with the data available, the validation dataset was utilized for evaluating the trained network. Due to computation cost and speed, the k fold cross-validation was not implemented. Hence, a hold-out validation with shuffling was used to generate an average detection result for all the models.

The training and test sets were used for the network training and testing, while the validation set was used to tune the hyperparameters. In the NMS process, 100 detections and an IoU threshold of 0.6 were maintained for each class. The momentum and the batch size were set as 0.9 and 8. The regularization value was set to 0.0004 as shown in Table 2. The warm-up learning rate of 0.0001333 was used to assist in the weight optimization after 5,000 training steps and at the end of the training period decay to zero. Batch normalization (BN) is used after the convolution layer and before nonlinearity layers to avoid overfitting and to save time during hyperparameter tuning [57]. During training, the data augmentation technique was used to increase samples' diversity because of insufficient training data. Six methods were employed for this data augmentation in the training phase: jitter boxes, horizontal flip, vertical flip, crop, pixel value, and rotation. To ensure guaranteed detection, the IOU confidence level is set at 0.6. Five measurements, including recall, precision, f_1 score, average accuracy, and mAP, are applied to evaluate the components' faults model performance.

Table 2: Training hyperparameters settings for CNN models

Hyperparameters	Values
Momentum	.9
L_2 regularization	0.0004
Batch size	8
IoU -Threshold	0.6
Min and max scale	3 - 7

5 Result And Discussion

This chapter presents and discusses the results of the experiment described in section 4. The three CNN architectures were trained with the same parameters and the same training, validation and testing

datasets and improved using hyperparameters tuning (see section 4.2). After running hyperparameter refinement simulations, the most optimum value was recorded and incorporated into each model to achieve the localisation and classification of the different EPTN component faults. Using these proposed SSD models with different backbones called SSD MobNet, SSD Rest101, and SSD Rest50, a four-class ETPN fault object detection was performed on our testing dataset containing 142 missing knobs, 75 broken insulators, 73 rusty clamps, and 45 missing insulator plate faults. The models were tested using three separate metrics, including F1-score and mAP. As previously mentioned in section 4.4, a holdout validation scheme was employed to produce an average detection result for all the utilized models in the study area. As observed, the CNN-based networks tested performs considerable well (regardless of the experimental setting considered), indicating the CNN superior capability to accurately detect faults on transmission assets in Nigeria using drone imagery. On the one hand, CNNs' remarkable ability to extract incredibly feature vectors from a neighbouring region enables the generation of more precise detections for a given pixel. On the other hand, the spatial resolution of drone imagery (in comparison to other conventional space borne sensors, for example, Landsat and sentinel) may make these convolutional features even more informative for identifying and diagnosing faults in the context of this work.

In this work, we have utilized three different optimizer (RMSprop, momentum, Adam) and the best average results were always achieved with the momentum optimizer. In general, it was observed that the momentum optimizer gave the best mAP across the different models using the default hyperparameter settings. SSD Rest50, SSD Rest101, and SSD MobNet achieved a mAP of 82.85%, 80.42%, 79.61%, respectively, using the momentum optimizer. The SSD Rest50 gained the highest accuracy when compared to the other two models. Also, the experiment showed that for the momentum optimizer, the validation, and total loss converge optimally. Furthermore, it has been expressively proven that the model's convergence is affected by the optimizer utilized. We observed that all the optimizers attain acceptable rates of accuracy, but one of the most glaring difference is the value of training loss and validation loss as well as the model convergence, i.e., the degree of loss range from zero. It can be inferred that the optimizer momentum with cosine learning rate is the one that provides the best results and the quickest to converge.

Using Momentum as the ideal learning algorithm, numerous learning rate settings were checked to improve the model performance. After several preliminary evaluations, it was confirmed that the best initial Learning rate (Lr) was 0.09. The first model, SSD MobNet, reached a mAP of 73.94%, 71.56%, 79.61%, and 82.52%, with the learning rate was 0.001, 0.01, 0.05, and 0.09, respectively, better performance of the model with increasing learning rate value. Similarly, the remaining two models: SSD Rest50 and SSD Rest101, demonstrated the greatest average mAP of 86.29% and 83.14%, with a learning rate of 0.09, which is 3.44% and 2.72% higher than those obtained when set to 0.05. The learning rate plays a significant role in the network's performance and how easily it can generalize [58]. Specifically, decreasing the learning rate beneath this value (0.09), which gives the fastest convergence outcomes, will improve the mAP to generalize, particularly for large, dynamic cases. The learning rate used for all models was 0.09 as they all performed better with this value.

The test results of the proposed single-phase components' faults identification and classification pipeline are shown in Table 3. It illustrates the precision, recall, f1 score, and accuracy of the three models, respectively. As can be seen from Table 5.1, the SSD ResNet holds the maximum overall mAP score of about 89.61% for the components' faults detected and properly classified. Low precision rates suggest that a significant number of false positive samples of the different EPTN component faults are generated when using the models for fault classification [22], [32], which is not the case here as the model generated fewer false positive samples of EPTN faults; hence the reason for the general precision rate being above 90.9%. By delving deeper into these results, we can develop a greater understanding of the contextual factors between the models and the various exploratory scenarios considered.

Table 3: Assessment of SSD Rest101, SSD Rest50, and SSD MobNet on the test dataset.

Models	EPTN Component Faults' Classes	Precision (%)	Recall (%)	F1-Score (%)	Avg F1 (%)	mAP
SSD Rest101	broken insulator	98.33	78.67	87.41	81.93	88.70
	missing insulator	100.00	67.34	80.49		
	missing knob	96.51	58.45	72.80		
	rusty clamp	98.27	78.08	87.02		
SSD MobNet SSD Rest50	broken insulator	100.00	72.00	83.72	82.54	89.61
	missing insulator	100.00	57.14	72.73		
	missing knob	97.22	73.94	84.00		
	rusty clamp	96.82	83.56	89.71		
	broken insulator	100.00	68.00	80.95	76.88	82.98
	missing insulator	95.83	46.93	63.01		
	missing knob	95.83	65.24	77.64		
	rusty clamp	93.55	79.45	85.93		

With regards to the research studies under consideration, there are more component faults not identified than misclassified, causing a lower recall rate, especially for the missing knob fault type as shown in Figure 10. From Figure 10, the recall rate of the SSD Rest50 is 57.14%, 73.94%, and 83.56% for missing knob and rusty clamp fault classes, respectively, which varies about 15.50%, and 5.48% to that detected and classified by SSD Rest101. Alternatively, the recall rate for the SSD Rest 101 is the greatest in identifying the broken and missing insulator faults. The SSD Rest50 achieved a better recall rate for broken insulator cap, missing insulator cap, missing knob, and rusty clamp component fault classes compared to SSD MobNet by 4.00%, 10.21%, 8.70%, and 4.11%. The SSD MobNet performs the least in detecting and classifying the missing insulator fault class compared to that of the SSD Rest101 and SSD Rest50 model. Generally, all models had a satisfactory recall in detecting and classifying each fault class, especially or identifying missing knob and rusty clamp faults. This reveals that the experimental single-

stage components' fault detection and classification pipeline can solve this identified problem by substantially increasing the model's performance in identifying and classifying the EPTN faults.

The SSD Rest101 is the second-best model with an overall mAP of 88.70%. Of the object detection methods tested, the one that delivered the least prediction (82.98%) was SSD MobNet. While the ResNet 101 derived model termed the SSD Rest101 has been noted to be the best in principle [33], [42]; however, in this case, the SSD model based on ResNet 50 contrasts conventional assumptions by revealing an improved result. The complexity of the network architecture can indeed justify the explanation behind the persistent lower results by ResNet 101 model, which is made of much deeper layers in contrast to the size of the training dataset; making the model characteristics over subsample and learns features; thus, affecting the performance in detecting different components' faults optimally. Furthermore, to intuitively reflect the proposed model's detection performance, the loss value graph was evaluated to understand, rationalize, and justify the proposed models' generalization ability. In general, we can assess the proposed model's performance using the loss graphs and examine the group of classification, localisation, and regularization loss (Liu et al., 2015; Shanmugamani et al., 2018). Figure 11 gives snapshots of the loss value sensitivity over the training and validation phase through the network trajectory.

A good performance is established based on the total and validation loss decrease until it becomes stable and the difference between both loss values reaches a minimum [58]. If the prediction errors are unbiased, the validation error should be near zero, and the validation loss decreases with a decrease in training loss. This can be seen distinctively by the loss graph of SSD Rest50, SSD Rest101, and SSD MobNet model. The Rest50 model represents a Deep network, the SSD Rest 101 serves as a super Deep Network, while the SSD MobNet is a shallow network.

The various weight optimizations associated with the training and validation of the dataset based on the model architecture show that the loss value remained relatively stable. In the experiments, the base and top CNN layers used the Rectified Linear Units (ReLUs) as activation functions over shuffled mini-batch gradient descent (batch size of 8) with the Adam optimization algorithm. The final output uses a sigmoid function for each decision node. Using the sigmoid activation, the final achieved pair losses, i.e., [validation loss, training loss] for the SSD MobNet, SSD Rest50, and SSD Rest101, were approximately equal to [0.281, 0.309], [0.378, 0.385] and [0.356, 0.342] respectively. In contrast to the SSD MobNet, SSD Rest50 and SSD Rest101 have higher orders of magnitude as they have more parameters due to having more layers and more filters per layer. This allowed the model to learn more complex features than the shallow network can provide. In the SSD Rest101, it is observable that the dataset was not sufficient to train the deeper network. The ResNet 50 backbone architecture, which represents the Deep Network, performs much better in minimizing the loss values than either previous network, achieving train and validation losses of 0.378 and 0.385, respectively, after 15 epochs. To better understand the algorithms proposed, some of the networks' training and development images output were examined. Finally, there is a strong link between training loss and validation loss. They both decrease and then become stable at a

constant value. This suggests that the model is correctly trained and has a high probability of working well on any dataset within this use case.

Overall, the proposed network consistently performs well in all tested scenarios, indicating that it is suitable for detecting faults on powerline in Nigeria using UAV imagery. The primary advantage of the proposed architecture over alternative methods is the SSD characteristic, which is based on its ability to effectively utilize a single-phase method for fault diagnosis of electricity transmission tasks and on its ability to effectively balance contextual constraints. Figure 12 provides an example of all the output images produced by all the models implemented. The sky-blue box denotes the missing insulator; the green box denotes the broken insulator, the turquoise box denotes the missing knob faults, while the white box bounds the rusty clamp defects. Each box is marked by the components' faults and its confidence score. The first column to the third column depicts the implemented method's performance, SSD MobNet, SSD Rest101, and SSD Rest50, respectively.

In the first row, the SSD MobNet (leftmost) gives an accurate detection of the missing insulator plate with a false positive identification of broken insulator, SSD Rest101 (middle) gives no result even with the presence of missing insulator plate and the SSD Rest50 (rightmost) achieves the best result with no false prediction. In the second row of Figure 6.4, the SSD Rest50 method detects the broken insulator fault, while the other implemented model leads to a wrong judgment with a false rate. In the third row, the model's performance behaves similarly to what is observed in the first row as the model is affected by the convoluted background interferences. The fourth row shows that all the implemented models had depicted the missing knob near perfectly with just one false positive of the missing knob faults for SSD MobNet (leftmost) and one false negative (rightmost).

6 Conclusion And Future Work

This study has expounded the incredible potential in combining UAV surveillance imagery and deep learning for automatic power transmission line inspection and fault detection, especially in developing countries. We proposed that for accurate power line fault detection the most suitable deep learning model is a single-shot object detection (CNN) model that has been adapted to the captured image constraints. The experimentation design of this study validates our proposed utilization of deep learning model on UAV imagery for power line fault detection. A comparative analysis of different state-of-the-art manual and deep learning-based power line fault detection techniques was carried out.

The findings of this study allow for the drawing of several significant conclusions about the general use of deep learning and UAV imagery for this application. First, transfer learning provided a better strategy to achieve a robust performance for all fault classes, being able to predict correctly more than half of their instances. Also, the adaptive optimizer, momentum with mini-batch SGD, allowed for the faster convergence of the proposed model and to automatically predict the optimum learning rate. Second, it was observed that higher learning rate achieved better mAP values across all the models implemented.

When examined more closely, each of the three approaches has a unique effect on each class, with SSDRest50 achieve the best performance.

With these considerations in mind, the presented modelling approach addresses the challenges of using UAV imagery in conjunction with data from developing countries to automate the monitoring of electrical power transmission faults in the future, thereby contributing to more reliable and formative transmission companies and power industry practices. In the future, the single-stage component identification and classification pipeline should be expanded to account for faults in different components' shapes and severity levels. Also, to measure the magnitude of the detected defects' scale, applying instance segmentation and using this knowledge to measure the scale and magnitude of the faults might suffice. In the future, given the limited data available for the fault inspection process, there are two methods to solve this problem. These include foreground and background superposition using segmentation networks and image processing techniques, and Generative Adversarial Networks (GANs) to create synthetic images. Additionally, extending this work to cover for real-time autonomous vision detection in the field incorporated with GPS-INS navigation.

Abbreviations

CNNs: Convolutional Neural Networks

EPTN: Electric Power Transmission Network

SSD: Single Shot Multibox Detector

FPN: Feature Pyramid Network

UAV: Unmanned Area Vehicle

VHR: Very High Resolution

SAR: Synthetic Aperture Radar

GSD: Ground Sample Data

DEM: Digital Elevation Model

RNN: Recurrent Neural Networks

ILN: Insulator Localizer Network

VGG: Visual Geometry Group

DDN: Defect Detector Network

RPN: Regional Proposal Network

RFCN: Region-based Fully Convolutional Network

RCNN: Region-based CNNs

RGB: Red, Green, Blue

IoU: Intersection over Union

mAP: mean Average Precision

ReLU: Rectified Linear Units

SGD: Stochastic Gradient Descent

GAN: Generative Adversarial Networks

BN: Batch Normalization

NMS: Non-maximum Suppression

Declarations

Ethics approval and consent to participate

NA

Consent for publication

NA

Authors' contributions

IM and FIO developed the conceptual framework. CFI, JEA, OEO, GAC and FE carried out the development and implementation under the supervision of IM and FIO. All the authors contributed to the writing of the paper. All authors read and approved the final manuscript.

Conflicting interests

There are no conflicting interests.

Acknowledgements

We want to acknowledge Nigerian Electricity Commission (NEC) for giving us access to the dataset.

Funding

NA

Availability of data and materials

The data is not available to public because NEC allowed us to use the data on the condition that the data shall not be distributed or shared.

References

[1] Alkire, S., Conconi, A., Robles, G. and Vaz, A. (2015). 'Destitution: Who and where are the poorest of the poor?', OPHI Briefing 35, Oxford Poverty and Human Development Initiative (OPHI), University of Oxford.

[2] S. Kufeoglu, *Economic Impacts of Electric Power Outages and Evaluation of Customer Interruption Costs*. Doctoral Dissertation in Permission of Aalto University, School of Electrical Engineering. 2015; 1-64.

[3] M. P. Blimpo and M. Cosgrove-Davies, *Electricity Access in Sub-Saharan Africa: Uptake, Reliability, and Complementary Factors for Economic Impact*. Africa Development Forum Washington, D.C. World Bank Group (2019).

[4] T. Ayodele, A. Ogunjuyigbe, and O. Oladele, "Improving the Transient Stability of Nigerian 330Kv Transmission Network Using Static Var Compensation Part I: the Base Study," *Niger. J. Technol.*, vol. 35, no. 1, p. 155, 2015, doi: 10.4314/njt.v35i1.23.

[5] P. Bertheau, C. Cader, and P. Blechinger, "Electrification modelling for Nigeria," *Energy Procedia*, vol. 93, no. March, pp. 108–112, 2016, doi: 10.1016/j.egypro.2016.07.157.

[6] L. Matikainen *et al.*, "Remote sensing methods for power line corridor surveys," *ISPRS J. Photogramm. Remote Sens.*, vol. 119, pp. 10–31, 2016, doi: 10.1016/j.isprsjprs.2016.04.011.

[7] Z. Xue, S. Luo, Y. Chen, and L. Tong, "The application of the landslides detection method based on sar images to transmission line corridor area," *2016 13th Int. Comput. Conf. Wavelet Act. Media Technol. Inf. Process. ICCWAMTIP 2017*, pp. 163–166, 2017, doi: 10.1109/ICCWAMTIP.2016.8079829.

[8] L. Yan, W. Wu, and T. Li, "Power transmission tower monitoring technology based on TerraSAR-X products," *Int. Symp. Lidar Radar Mapp. 2011 Technol. Appl.*, vol. 8286, p. 82861E, 2011, doi: 10.1117/12.912336.

[9] L. F. Luque-Vega, B. Castillo-Toledo, A. Loukianov, and L. E. Gonzalez-Jimenez, "Power line inspection via an unmanned aerial system based on the quadrotor helicopter," *Proc. Mediterr. Electrotech. Conf. - MELECON*, no. April, pp. 393–397, 2014, doi: 10.1109/MELCON.2014.6820566.

[10] M. Wang, W. Tong, and S. Liu, "Fault detection for power line based on convolution neural network," *ACM Int. Conf. Proceeding Ser.*, vol. Part F1285, pp. 95–101, 2017, doi: 10.1145/3094243.3094254.

[11] H. Guan, Y. Yu, J. Li, Z. Ji, and Q. Zhang, “Extraction of power-transmission lines from vehicle-borne lidar data,” *Int. J. Remote Sens.*, vol. 37, no. 1, pp. 229–247, 2016, doi: 10.1080/01431161.2015.1125549.

[12] J. Ahmad, A. S. Malik, M. F. Abdullah, N. Kamel, and L. Xia, “A novel method for vegetation encroachment monitoring of transmission lines using a single 2D camera,” *Pattern Anal. Appl.*, vol. 18, no. 2, pp. 419–440, 2015, doi: 10.1007/s10044-014-0391-9.

[13] X. Yu *et al.*, “Comparison of laser and stereo optical, SAR and InSAR point clouds from air- and space-borne sources in the retrieval of forest inventory attributes,” *Remote Sens.*, vol. 7, no. 12, pp. 15933–15954, 2015, doi: 10.3390/rs71215809.

[14] M. Jaya Bharata Reddy, B. Karthik Chandra, and D. K. Mohanta, “Condition monitoring of 11 kV distribution system insulators incorporating complex imagery using combined DOST-SVM approach,” *IEEE Trans. Dielectr. Electr. Insul.*, vol. 20, no. 2, pp. 664–674, 2013, doi: 10.1109/TDEI.2013.6508770.

[15] J. Jiang, L. Zhao, J. Wang, Y. Liu, M. Tang, and Z. Ji, “The electrified insulator paramater measurement for flashover based on photogrammetric method,” *MIPPR 2011 Multispectral Image Acquis. Process. Anal.*, vol. 8002, no. 86, p. 80021I, 2011, doi: 10.1117/12.902054.

[16] Y. Hu and K. Liu, Inspection and Monitoring Technologies of Transmission Lines with Remote Sensing, New York, NY, USA:Academic, pp. 257-264, 2017..

[17] A. Zormpas, K. Moirogiorgou, K. Kalaitzakis, G. A. Plokamakis, and P. Partsinevelos, “Power Transmission Lines Inspection using Properly Equipped Unmanned Aerial Vehicle (UAV),” *2018 IEEE Int. Conf. Imaging Syst. Tech.*, pp. 1–5, 2018.

[18] 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,” pp. 1–29, 2020, doi: 10.1016/j.arcontrol.2020.09.002.

[19] X. Yu, X. Wu, C. Luo, and P. Ren, “Deep learning in remote sensing scene classification: a data augmentation enhanced convolutional neural network framework,” *GIScience Remote Sens.*, vol. 54, no. 5, pp. 741–758, 2017, doi: 10.1080/15481603.2017.1323377.

[20] Y. Baştanlar and M. Özuysal, “Introduction to machine learning,” *Methods Mol. Biol.*, vol. 1107, pp. 105–128, 2014, doi: 10.1007/978-1-62703-748-8_7.

[21] Z. A. Siddiqui and U. Park, “A Drone Based Transmission Line Components Inspection System with Deep Learning Technique,” *Energies*, vol. 13, no. 13, pp. 1–24, 2020, doi: 10.3390/en13133348.

[22] J. Han *et al.*, “Search like an eagle: A cascaded model for insulator missing faults detection in aerial images,” *Energies*, vol. 13, no. 3, pp. 1–20, 2020, doi: 10.3390/en13030713.

[23] Y. Song *et al.*, “A vision-based method for the broken spacer detection,” *2015 IEEE Int. Conf. Cyber Technol. Autom. Control Intell. Syst. IEEE-CYBER 2015*, pp. 715–719, 2015, doi: 10.1109/CYBER.2015.7288029.

[24] Y. Zhai, H. Cheng, R. Chen, Q. Yang, and X. Li, “Multi-saliency aggregation-based approach for insulator flashover fault detection using aerial images,” *Energies*, vol. 11, no. 2, pp. 1–12, 2018, doi: 10.3390/en11020340.

[25] Y. Zhai, D. Wang, M. Zhang, J. Wang, and F. Guo, “Fault detection of insulator based on saliency and adaptive morphology,” *Multimed. Tools Appl.*, vol. 76, no. 9, pp. 12051–12064, 2017, doi: 10.1007/s11042-016-3981-2.

[26] J. Han *et al.*, “A method of insulator faults detection in aerial images for high-voltage transmission lines inspection,” *Appl. Sci.*, vol. 9, no. 10, pp. 1–22, 2019, doi: 10.3390/app9102009.

[27] Y. Liu *et al.*, “2016 4th International Conference on Applied Robotics for the Power Industry, CARPI 2016,” *2016 4th Int. Conf. Appl. Robot. Power Ind. CARPI 2016*, pp. 1–5, 2016.

[28] J. Fu, G. Shao, L. Wu, L. Liu, and Z. Ji, “Defect detection of line facility using hierarchical model with learning algorithm,” *High Volt. Eng. Dep. CEPRI*, vol. 43, no. 1, pp. 266–275, 2017, doi: 10.13336/j.1003-6520.hve.20161227035.

[29] T. Mao *et al.*, “Defect recognition method based on HOG and SVM for drone inspection images of power transmission line,” *2019 Int. Conf. High Perform. Big Data Intell. Syst. HPBD IS 2019*, no. 61701404, pp. 254–257, 2019, doi: 10.1109/HPBDIS.2019.8735466.

[30] Z. Zhao, G. Xu, Y. Qi, N. Liu, and T. Zhang, “Multi-patch deep features for power line insulator status classification from aerial images,” *Proc. Int. Jt. Conf. Neural Networks*, vol. 2016-Octob, pp. 3187–3194, 2016, doi: 10.1109/IJCNN.2016.7727606.

[31] X. Liu, H. Jiang, J. Chen, J. Chen, S. Zhuang, and X. Miao, “Insulator Detection in Aerial Images Based on Faster Regions with Convolutional Neural Network,” *IEEE Int. Conf. Control Autom. ICCA*, vol. 2018-June, pp. 1082–1086, 2018, doi: 10.1109/ICCA.2018.8444172.

[32] H. Jiang, X. Qiu, J. Chen, X. Liu, X. Miao, and S. Zhuang, “Insulator Fault Detection in Aerial Images Based on Ensemble Learning with Multi-Level Perception,” *IEEE Access*, vol. 7, pp. 61797–61810, 2019, doi: 10.1109/ACCESS.2019.2915985.

[33] 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 Trans. Syst. Man, Cybern. Syst.*, vol. 50, no. 4, pp. 1486–1498, 2020, doi: 10.1109/TSMC.2018.2871750.

[34] L. Ma, C. Xu, G. Zuo, B. Bo, and F. Tao, “Detection Method of Insulator Based on Faster R-CNN,” *2017 IEEE 7th Annu. Int. Conf. CYBER Technol. Autom. Control. Intell. Syst. CYBER 2017*, pp. 1410–1414, 2018,

doi: 10.1109/CYBER.2017.8446155.

[35] R. Bai, H. Cao, Y. Yu, F. Wang, W. Dang, and Z. Chu, “Insulator Fault Recognition Based on Spatial Pyramid Pooling Networks with Transfer Learning (Match 2018),” *ICARM 2018 - 2018 3rd Int. Conf. Adv. Robot. Mechatronics*, pp. 824–828, 2019, doi: 10.1109/ICARM.2018.8610720.

[36] T. Olawale and S. A. Ad, “REPORT ON STUDENT INDUSTRIAL WORK EXPERIENCE SCHEME (SIWES) AT Transmission Company of Nigeria (TCN) for the Student Industrial Work Experience Scheme (SIWES),” no. August 2018, 2018, doi: 10.13140/RG.2.2.35067.05929.

[37] Y. Hu and K. Liu and C. Mengqi 2019 Free vibration analysis of transmission lines based on the dynamic stiffness methodR. Soc. open sci.6181354181354.

[38] C. Francois, *Deep Learning with Python*, vol. 53, no. 9. 2019.

[39] D. of C. S. U. Stanford, “CS231n: Convolutional Neural Networks for Visual Recognition,” *STANFORD VISION AND LEARNING LAB*. https://cs231n.github.io/convolutional-networks/ (accessed Nov. 21, 2020).

[40] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Handbook of approximation algorithms and metaheuristics,” *ImageNet Classif. with Deep Convolutional Neural Networks*, pp. 1–1432, 2012, doi: 10.1201/9781420010749.

[41] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–14, 2015.

[42] J. W. Yun, “Deep Residual Learning for Image Recognition arXiv:1512.03385v1,” *Enzyme Microb. Technol.*, vol. 19, no. 2, pp. 107–117, 2015, [Online]. Available: https://arxiv.org/pdf/1512.03385.pdf.

[43] A. Rosebrock, “ImageNet: VGGNet, ResNet, Inception, and Xception with Keras,” *pyimagesearch*, 2017. https://www.pyimagesearch.com/2017/03/20/imagenet-vggnet-resnet-inception-xception-keras/ (accessed Nov. 25, 2020).

[44] 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:1704.04861, 2017.

[45] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN2015,” *Biol. Conserv.*, vol. 158, pp. 196–204, 2015.

[46] J. Dai, Y. Li, K. He, and J. Sun, “R-FCN: Object Detection via Region-based Fully Convolutional Networks. arXiv:1605.06409, Top of Form

2016.

[47] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 779–788,

2016, doi: 10.1109/CVPR.2016.91.

[48] W. Liu *et al.*, “SSD Net,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9905 LNCS, pp. 21–37, 2015.

[49] J. Huang *et al.*, “Speed/accuracy trade-offs for modern convolutional object detectors,” *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 3296–3305, 2017, doi: 10.1109/CVPR.2017.351.

[50] W. Shi, S. Bao, and D. Tan, “FFESSD: An accurate and efficient single-shot detector for target detection,” *Appl. Sci.*, vol. 9, no. 20, 2019, doi: 10.3390/app9204276.

[51] K. O’Shea and R. Nash, “An Introduction to Convolutional Neural Networks,” pp. 1–11, 2015, [Online]. Available: http://arxiv.org/abs/1511.08458.

[52] S. Saha, “A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way,” *Medium*, 2018. https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53 (accessed Nov. 15, 2020).

[53] N. K. Manaswi, “Deep Learning with Applications Using Python,” *Deep Learn. with Appl. Using Python*, pp. 91–96, 2018, doi: 10.1007/978-1-4842-3516-4.

[54] R. Shanmugamani, A. G. A. Rahman, S. M. Moore, and N. Koganti, *Deep Learning for Computer Vision: Expert techniques to train advanced neural networks using Tensorflow and Keras.*, 1st ed. Birmingham, UK: Packt Publishing Ltd, 2018.

[55] Y. Yang, L. Wang, Y. Wang, and X. Mei, “Insulator self-shattering detection: a deep convolutional neural network approach,” *Multimed. Tools Appl.*, vol. 78, no. 8, pp. 10097–10112, 2019, doi: 10.1007/s11042-018-6610-4.

[56] Arabi S, Haghighat A, Sharma A (2020) A deep-learning-based computer vision solution for construction vehicle detection. Comput Aided Civil Infrastruct Eng 35:753–767.

[57] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” *32nd Int. Conf. Mach. Learn. ICML 2015*, vol. 1, pp. 448–456, 2015.

[58] G. Perin and S. Picek, “On the Influence of Optimizers in Deep Learning-based Side-channel Analysis,” *Cryptol. ePrint Arch.*, no. Report 2020/977, pp. 1–22, 2020, [Online]. Available: https://eprint.iacr.org/2020/977.

[59] O. M. Komolafe and K. M. Udofia, “Review of electrical energy losses in Nigeria,” *Niger. J. Technol.*, vol. 39, no. 1, pp. 246–254, 2020, doi: 10.4314/njt.v39i1.28.

Figures

Figure 1

Missing glass insulator faults

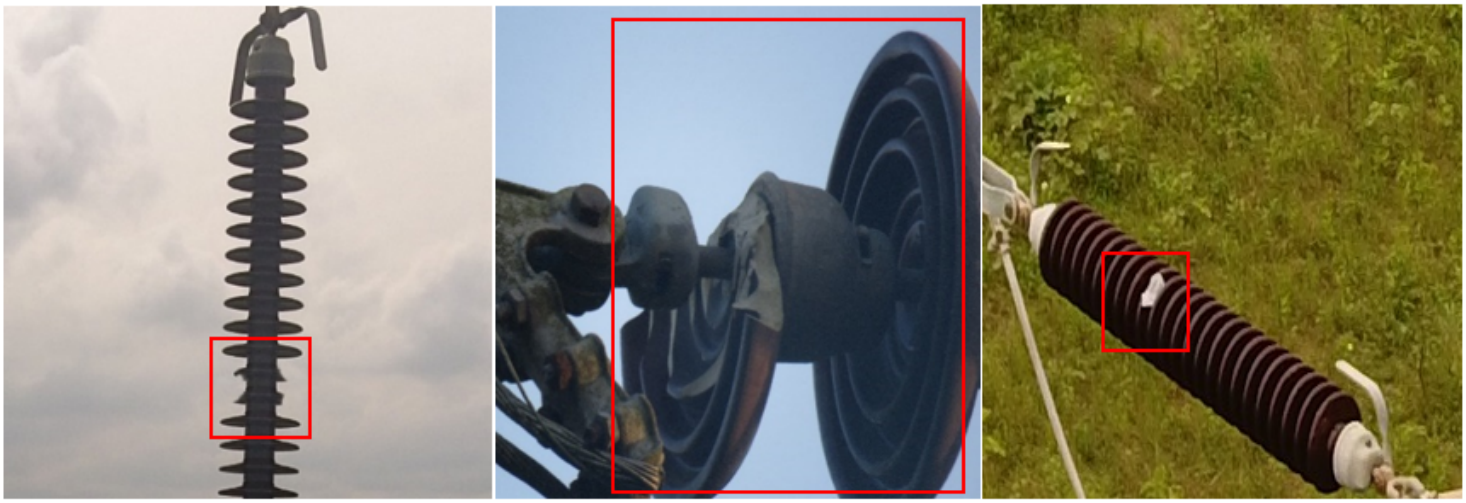

Figure 2

Broken insulator faults prominent with the porcelain or composite plate type insulator

Figure 3

Rusty strain (a) and suspension (b) clamp

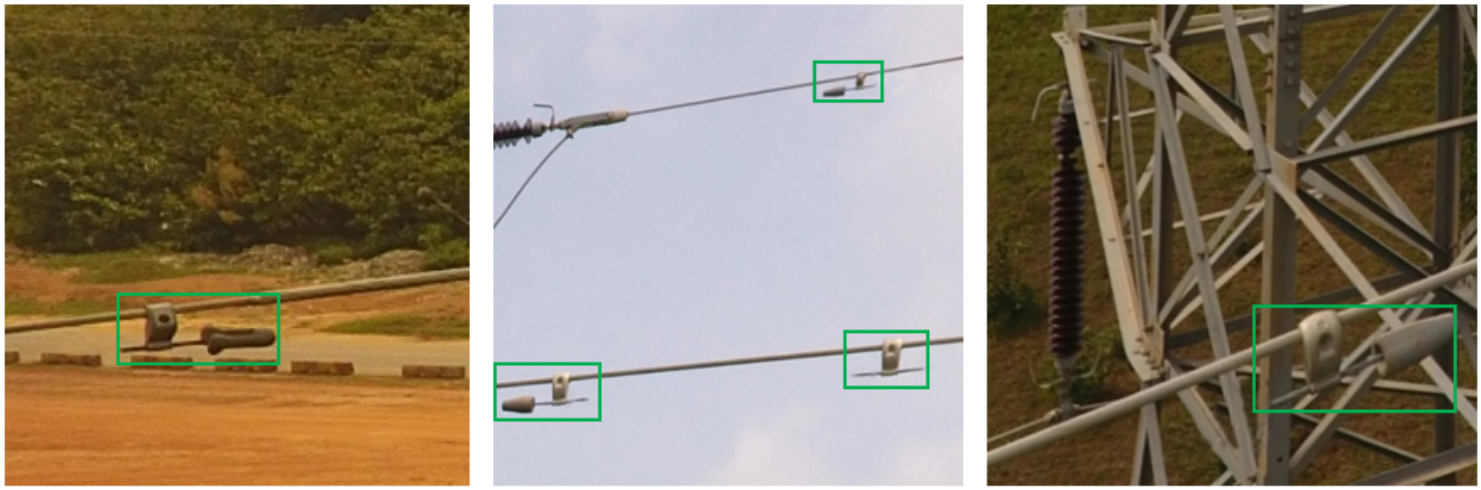

Figure 4

Broken fitting (vibration dampers)

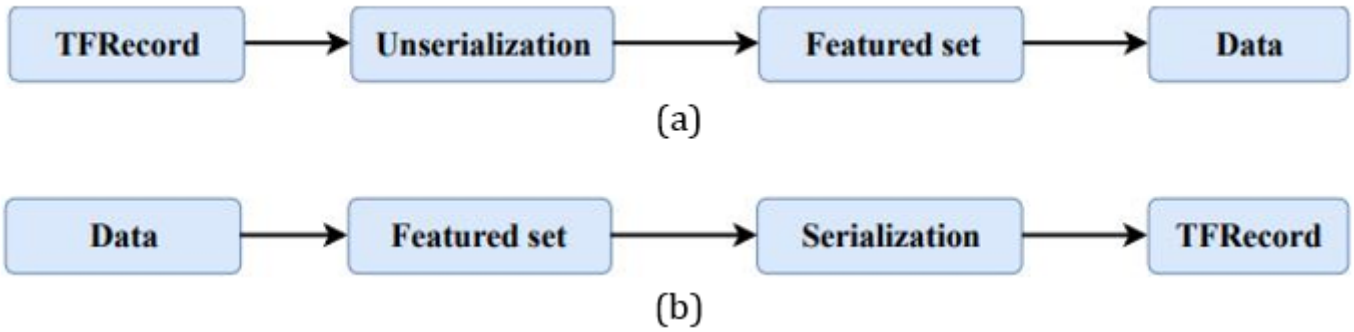

Figure 5

TFrecord (a) reading and (b) writing principle.

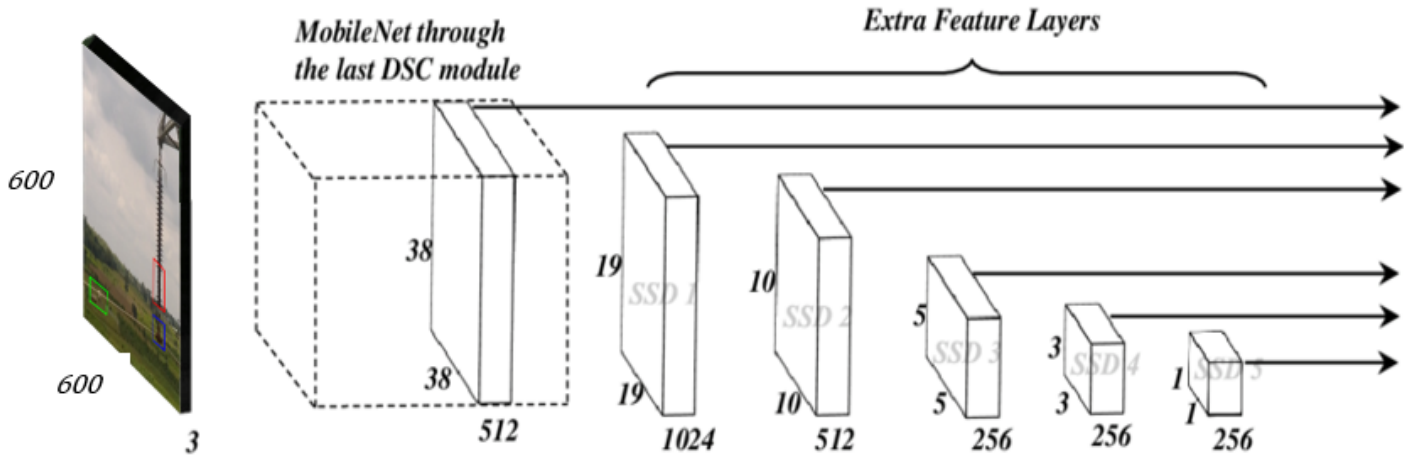

Figure 6

Model architecture [56].

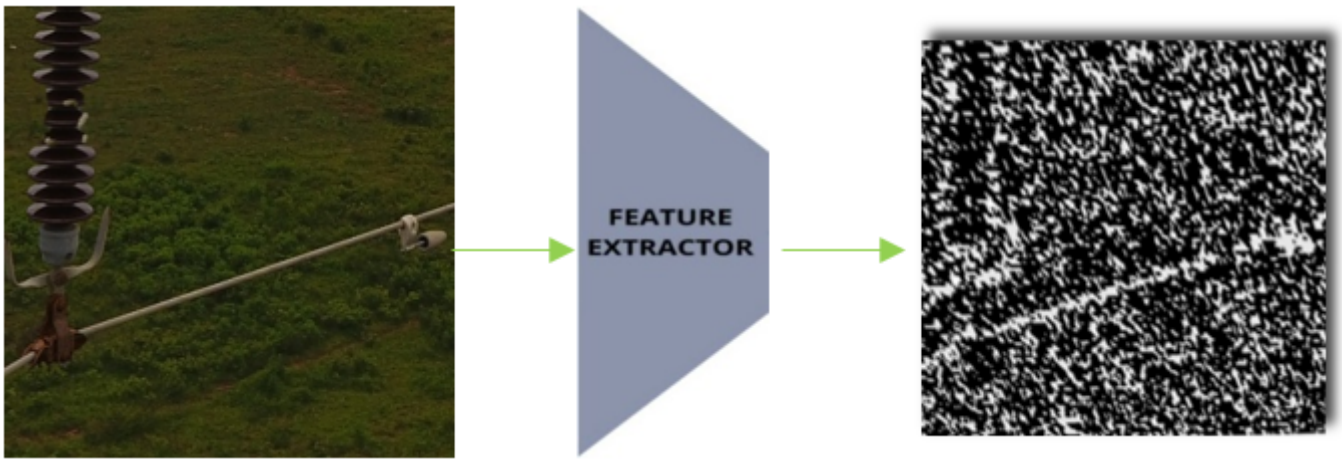

Figure 7

Input image patch and corresponding feature map generated by the feature extractor (backbone architecture).

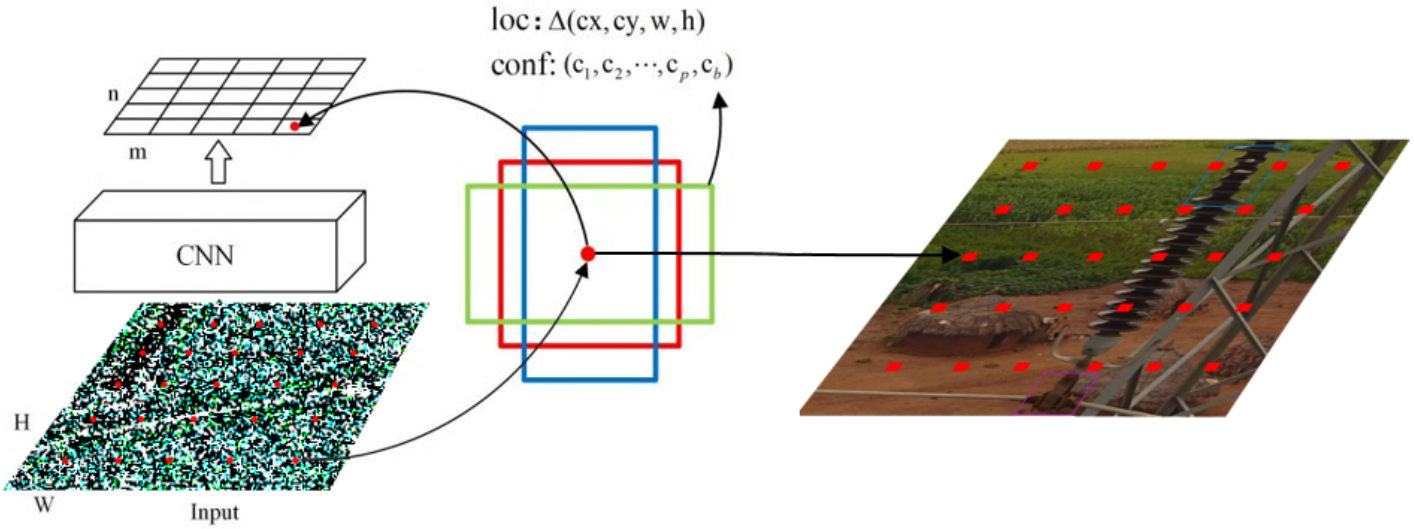

Figure 8

The default boxes generation for one cell over the backbone network feature map.

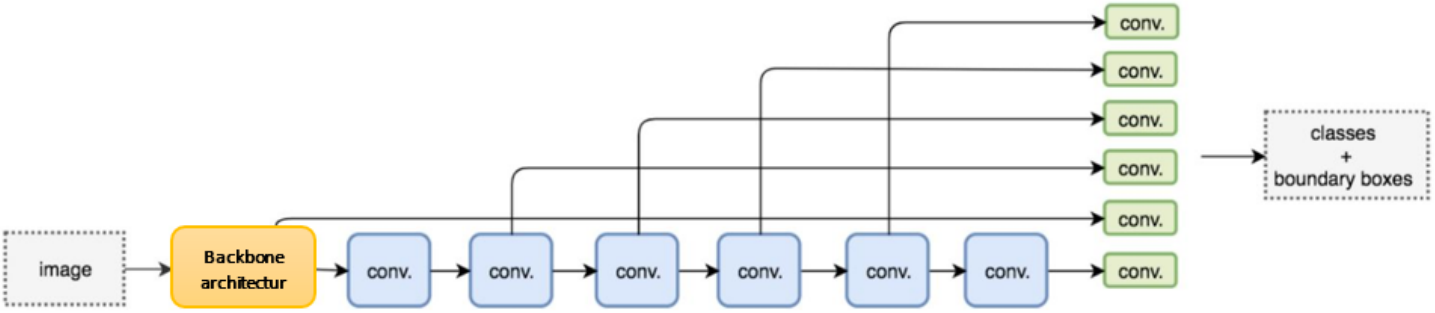

Figure 9

Multiscale downscaling layer (auxiliary layer) concept.

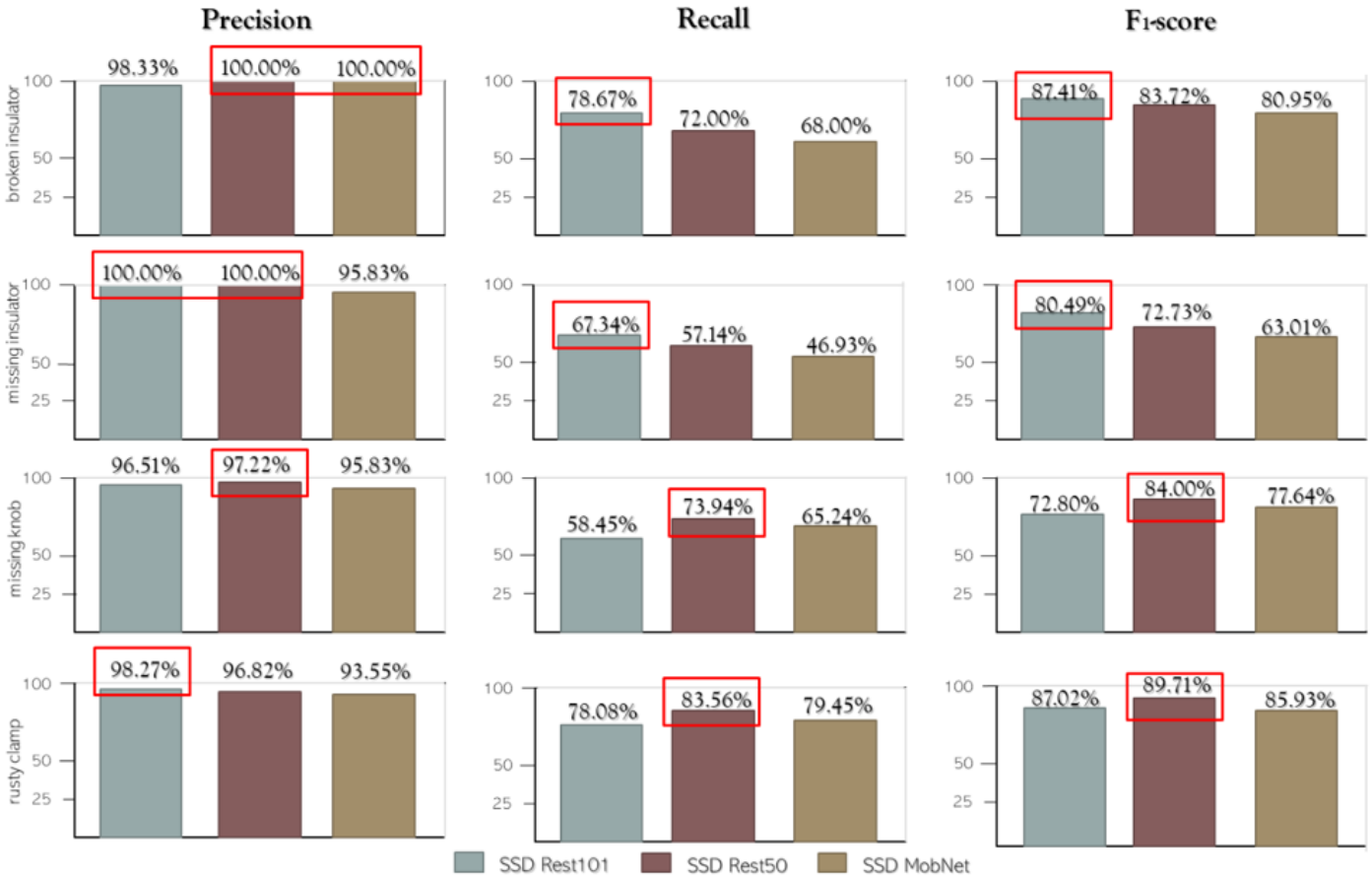

Figure 10

A comparative view of the performance metrics.

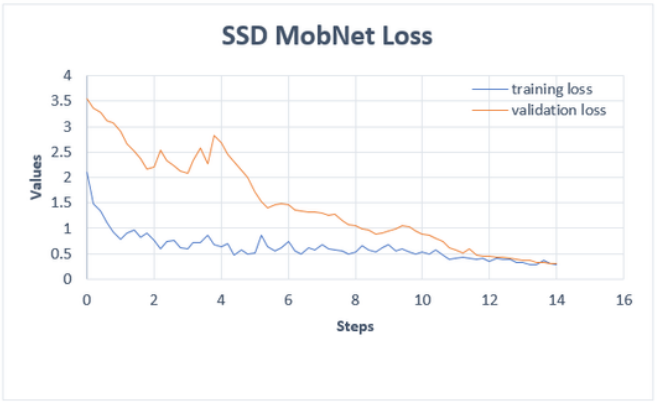

Figure 11

Epoch vs. Loss Graphs over time.

Figure 12

Experimental results of the four components faults. The first column to the third column depicts the proposed method's performance in each row, SSD MobNet, SSD Rest50, and SSD Rest101.