Received: 5 October 2024 | Accepted: 24 January 2025

DOI: 10.1111/mice.13434

INDUSTRIAL APPLICATION

COMPUTER-AIDED CIVIL AND INFRASTRUCTURE ENGINEERING WILEY

Automated corrosion surface quantification in steel transmission towers using UAV photogrammetry and deep convolutional neural networks

Pierclaudio Savino[1] | **Fabio Graglia**[2] | **Gabriele Scozza**[2] | **Vincenzo Di Pietra**[3]

[1]Department of Structural, Geotechnical and Building Engineering, Politecnico di Torino, Torino, Italy

[2]Rai Way S.p.A., Roma, Italy

[3]Department of Environment, Land and Infrastructure Engineering, Politecnico di Torino, Torino, Italy

Correspondence
Vincenzo Di Pietra, Land and Infrastructure Engineering, Politecnico di Torino, C.so Duca degli Abruzzi 24, 10129 Torino, Italy.
Email: vincenzo.dipietra@polito.it

Abstract

Corrosion in steel transmission towers poses a challenge to structural integrity and safety, requiring efficient detection methods. Traditional visual inspections are unsustainable due to the complexity and volume of structures. Their manual, qualitative, and subjective nature often leads to inconsistencies in maintenance planning. This study proposes a deep learning-based approach for semantic segmentation of corroded areas on steel towers. Using the DeepLabv3+ model, the network was trained and validated on 999 field photographs. MobileNetV2, serving as the feature extractor, was chosen for its optimal balance between accuracy and computational efficiency, achieving a validation accuracy of 90.8% and a loss of 0.23. The trained network was applied to real-world inspections using orthomosaics derived from photogrammetric reconstructions of the South-East tower at the Torino Eremo broadcasting center. These photogrammetric products not only enabled precise segmentation of corroded areas but also provided the foundation for corrosion quantification with metrical accuracy, a critical advantage for maintenance planning. Unlike traditional image segmentation methods, which lack a spatial reference and precise scaling, the photogrammetric approach ensures that the corrosion extent and distribution are quantified in exact physical dimensions, enhancing the reliability of the analysis. The results show that deep learning-based inspections can automate detection, providing reliable data and reducing reliance on manual inspections, enhancing efficiency, safety, and accuracy.

1 | INTRODUCTION

Structures and infrastructures, including bridges, tunnels, dams, energy plants, and telecommunications networks, are essential for human well-being and sustainable development. Modern societies depend on the reliable and safe operation of these systems, as malfunctions or service interruptions can cause significant losses and prevent progress. Throughout their lifecycle, these structures face environmental risks, human-induced threats, and performance deterioration, which can compromise their reliability and integrity. Therefore, it is crucial to implement effective control systems and consistent maintenance strategies to ensure efficient monitoring and management of structural integrity.

Visual inspection is currently the primary method used globally to assess the physical and functional conditions of most assets. It serves as the foundation for planning monitoring and maintenance activities, providing a preliminary evaluation of safety and structural conditions. These activities are still conducted mainly manually, with qualified inspectors examining on-site elements according to an inspection protocol. However, this practice is becoming increasingly unsustainable and costly on a large scale, given the number and complexity of the structures to be inspected. For example, in the case of telecommunications towers, inspectors often have to reach considerable heights, exposing themselves to significant safety risks. Additionally, the results of visual inspections tend to be qualitative and subjective, leading to potential discrepancies in inspection reports. The substantial volume of data generated necessitates an efficient information management system (Pezeshki et al., 2023). A maintenance management system should integrate a wide range of information, including drawings, inspection reports, manuals, and checklists. Often, this information is in text format and stored on paper documents, further complicating the process. These challenges can lead to a lack of crucial data for effective asset management and inadequate decision-making during the operation and maintenance phases. The series of challenges outlined above has driven academia and industry to develop a wide range of computerized tools to support structural inspections. Innovative technologies such as artificial intelligence (AI), laser scanning, unmade aerial vehicle (UAV), building information modeling, virtual reality, augmented reality, and mixed reality (MR) are advancing the digitalization of information management, processing, and visualization. Their application enables the resolution of various technical and complex tasks, including those related to the organization and management of constructions at all stages of the lifecycle.

One emerging technology that is gaining significant interest is the use of UAVs in construction and transportation engineering. Applications include traffic monitoring and surveillance (Belcore, Di Pietra, et al., 2022; Irizarry et al., 2014), road condition inventory and inspection (Barfuss et al., 2012), topographic surveying and mapping (Brooks et al., 2015), construction progress and status monitoring (Lin et al., 2015), earthwork volume estimation (Hugenholtz et al., 2015), and monitoring unstable slopes (Belcore, Piras & Pezzoli et al., 2022; Niethammer et al., 2010). This technology can effectively replace visual inspections of structures, offering significant advantages in terms of speed, safety, cost, and efficiency. It also facilitates instant information sharing with multiple stakeholders and the ease of maneuvering through automated flights. Specialized inspectors can safely conduct inspections by monitoring UAV video transmissions or accessing collected images and video later, eliminating the need to reach difficult locations as required in conventional procedures. UAV inspections provide a significant economic advantage by utilizing compact, portable, and low-cost devices, compared to the equipment needed for inspectors to physically access hard-to-reach investigation sites. The sensors mounted on drones, such as Red-Green-Blue (RGB), multispectral, thermal, or LiDAR, enable the collection of large amounts of inspection data, paving the way for increased digitalization of monitoring and inspection processes. Moreover, metrical accurate spatial information (e.g., dimensions, areas, volumes) can be derived from remote sensing data and enhanced with the spectral response of the objects under inspection. In recent years, there has been significant development in new inspection methods using UAVs, as highlighted in the literature. For example, Mandirola et al. (2022) proposed a practical approach for detecting and assessing bridges using aerial photogrammetry. Pinto et al. (2020) demonstrated a visual inspection and detection of two bridges using UAVs and the structure from motion (SfM) technique to build 3D models. Marchewka et al. (2020) presented a framework for monitoring steel bridges by measuring rivet displacement and corrosion using UAVs and image processing techniques. Sankarasrinivasan et al. (2015) utilized UAVs and image processing procedures to identify cracks and assess degraded areas in civil infrastructures. Reagan et al. (2016) proposed a method for reconstructing the surface deformation state of bridges using images captured by UAVs and 3D digital image correlation. Phung et al. (2017) developed a system for the automatic detection of cracks in bridges. UAV-captured images are processed with a peak detection algorithm for clustering and a thresholding technique for crack detection. Truong-Hong et al. (2018) proposed a method to extract a point cloud of a bridge deck from images captured by low-cost UAVs and to identify pavement cracks by comparing point cloud deviations with the flat surface of the deck. Khaloo et al. (2018) presented a case study of visual inspection of the Brighton Dam in Maryland, integrating UAV digital image acquisitions and photogrammetry for 3D point cloud reconstruction. Das and Woolsey (2019) proposed an algorithm for UAV inspection path planning for truss structures, such as steel bridges, based on a simplified model of the structure and the addition of navigation points around joints. On the same topic, Jeon et al. (2024) introduced the use of 3D LiDAR for autonomous flight to inspect transmission tower. The dense point cloud voxelization is used to plan the UAV's flight path and maintain alignment with transmission lines.

S. Chen et al. (2019) compared UAV-SfM with terrestrial laser scanning for image acquisition and 3D reconstruction of a bridge, finding significant advantages of UAV-SfM in terms of accuracy, cost, and survey time. Gillins et al. (2018) conducted a cost–benefit analysis of UAV structural inspections and estimated an average saving of about $10,000 per bridge inspection and a cost–benefit ratio of 9 when implementing a UAV bridge inspection program. Furthermore, the evolution of UAVs, combined with significant advancements in deep learning techniques in recent years, has increased the efficiency and accuracy of the inspection process. In the context of inspecting power transmission lines, a flight planning strategy has been developed to inspect both tower and transmission lines while maintaining a safe distance due to electromagnetic interference (Cui et al., 2017). To enhance operational efficiency and ensure greater safety in UAV control, Diniz et al. (2022) developed a deep learning-based strategy for detecting and tracking power transmission lines, along with a system to facilitate assisted control during UAV landing. Barbosa (2020) introduced a deep learning-based method for the automated inspection of electrical infrastructures using UAVs, enabling the identification and estimation of the position and size of various components such as insulators, poles, and transmission towers, ensuring comprehensive mapping of the structures. For residential building inspections, Shin et al. (2023) presented an integrated UAV-AI process that includes preliminary assessment, data acquisition, defect identification, and 3D model reconstruction. Chen et al. (2022) proposed a Building Information Modeling (BIM) assisted SfM pipeline to extract the structure of interest from aerial photographs and filter out irrelevant non-concrete areas to be forwarded to a U-Net with a VGG16 backbone for crack segmentation. For defect inspection in sewer pipelines, Ma et al. (2024) developed a complete system to create two digital replicas, including an attention mechanism for defect detection, a depth estimation network for generating depth maps, and 2D-to-3D mapping algorithm to transform segmentation results into 3D spaces. In the context of bridge inspections, AI applications have expanded significantly in recent years. These advancements include the classification of concrete surface defects (Aliyari et al., 2021; Jang et al., 2023; Savino & Tondolo, 2021), object detection models (Jiang et al., 2024), and pixel-level semantic segmentation (Jiang et al., 2023; Li et al., 2019; Savino & Tondolo, 2023; Yang et al., 2018; Zhang et al., 2023). Regarding damage detection based on the 3D reconstruction of bridges inspected by UAVs, Li et al. (2024) emphasize the importance of multi-sensor data integration for achieving high-resolution reconstructions, which significantly enhance the capability to detect surface damages. To address challenges in bridge crack segmentation, L. Sun et al. (2024) proposed CCSNet, an integration–competition network. This model incorporates grayscale-oriented adjustment to mitigate high-frequency light issues, an integration–competition mechanism to separate complex backgrounds and crack grayscale features, and an attention mechanism to enhance the extraction of shallow features in tiny cracks. Additionally, W. Sun et al. (2024) introduced a two-step rapid inspection method for underwater concrete bridge structures that combines sonar imaging, camera data, and deep learning techniques. This approach utilizes sonar data to localize potential areas of damage, which are subsequently inspected in detail using camera imaging and a deep learning model for defect classification and segmentation. Recent studies have proposed integrating UAVs for data collection, employing deep learning methods for corrosion segmentation on steel bridges, and applying MR for digital visualization (Montes et al., 2023). Hattori et al. (2024) developed a method for measuring the position and area of corroded sections on the underside of steel box girder bridges using semantic segmentation. The detected data are then integrated into a 3D BIM model by assigning area and coordinates in text format to a representative icon. For tunnel inspections, recent research has introduced innovative AI models. Q. Zhou et al. (2022) developed the YOLOv4-ED model, combining EfficientNet and depthwise separable convolution to detect water leakage, cracks, and exposed rebar. Z. Zhou et al. (2023) further refined crack identification for tunnel lining by integrating Swin Transformer and convolutional neural network (CNN) in a hybrid semantic segmentation model. Shim et al. (2023) developed a comprehensive automated tunnel inspection system, featuring a robot that autonomously navigates tunnels, defects concrete surface damage using a deep learning-based sensor fused with Generative Adversarial Network (GAN), and operates with a specifically designed manipulator. In the field of asphalt pavement inspections, deep learning models like U-Net and its variants have been widely utilized for crack segmentation. To address the challenges of pixel-level detection for thin cracks on road surfaces, Siriborvornratanakul (2023) introduced a new variant of a CNN named ThinCrack U-Net, demonstrating a significant improvement in performance, compared to existing U-Net variants. Yao et al. (2024) introduced a model incorporating a pyramid region attention module within the U-Net framework, leveraging Residual Network (ResNet)-34 for fast and high-precision segmentation. To reduce computational costs for deployment on robotic platforms, Zhu et al. (2024) proposed a novel network structure that uses a hybrid attention block to remove redundant feature channels and depth-wise separable convolutions. Similarly, Huang et al. (2024) introduced a lightweight feature attention fusion network, combining FasterNet as the backbone with a receptive field block to mitigate crack information loss and a feature fusion module to combine decoder outputs with encoder low-level features. For identifying underwater cracks in dams, Zhu et al. (2024) proposed a machine vision-based intelligent segmentation method integrating a swarm optimization algorithm and deep learning techniques. Their approach also incorporates a semantic compensation module in the decoder to fuse channel and spatial attention, improving multi-scale detail representation. In order to detect and assess fire damage in reinforced concrete structures, Wang et al. (2024) introduced an enhanced YOLOv5s-D network, incorporating a ShuffleNet module, an adaptive attention module, and a feature enhancement module. These innovations led to reduced network parameters, improved inference speed, and enhanced detection capabilities, particularly in complex backgrounds.

The advancement of UAV and AI technologies has opened new frontiers in the inspection and management of infrastructure, enabling the collection of detailed, real-time data more efficiently as well as remote sensing data from satellites (Belcore et al., 2020; Entezami et al., 2024). However, despite the progress highlighted in the literature, there are still challenges and limitations to be addressed. Most studies have primarily focused on the visual inspection of bridges and concrete structures, often neglecting damage detection for telecommunications towers. Additionally, while there have been numerous contributions regarding bridge inspection using UAVs and AI algorithms, few studies address the management of fragmented UAV images to achieve a comprehensive view of structural health.

In this context, this study addresses key gaps in the application of pretrained segmentation models for detecting and quantifying corrosion in transmission towers. To the knowledge of the authors, the performance of pretrained semantic segmentation models for detecting and quantifying corrosion in transmission towers has not been investigated. Unlike previous studies that primarily focused on other types of structures, this work specifically addresses the unique challenges associated with transmission towers, such as their complex geometries, environmental variations, and distinct corrosion patterns. Moreover, no prior research has investigated the generalization capabilities of such models on orthomosaic, which differ significantly from the used training datasets. Based on these gaps, this research proposed a DeepLabv3+ segmentation model to perform the semantic segmentation of images containing corrosion areas in steel transmission towers. The first objective was to train a robust neural network that is not affected by variations in image quality and is effective across a different set of UAV images and radiometric products. To achieve this, a dataset of 999 images with varying resolutions, collected from on-field tower inspections, was used to develop the neural network. The second objective was to identify the most suitable pretrained neural network for corrosion detection tasks using transfer learning technique. Finally, the practical utility of the proposed approach for on-site transmission tower inspections was also demonstrated by introducing a metrically accurate and user-friendly quantification procedure, simplifying the corrosion measurement process to prioritize ease of use and computational efficiency for practical applications. Semantic segmentation was applied to orthomosaics derived from the photogrammetric reconstruction process following UAV inspections. The results validate the trained model's ability to generalize well to unseen datasets, including orthomosaics, despite their distinct characteristics, compared to the training data. By bridging these gaps, the research contributes a novel framework for reliable, scalable, and practical corrosion detection in transmission towers, offering a significant advancement over existing techniques.

2 | METHODS

2.1 | Pixel-wise classification through neural networks

Neural networks, inspired by the human brain, are computational models composed of interconnected layers of nodes or "neurons." These networks are trained to recognize patterns and make predictions based on input data. Among the various architectures, CNNs are particularly well-suited for tasks involving spatial hierarchies, such as image classification, object detection, and image segmentation. CNNs operate by applying convolutional filters to the input images to create feature maps that capture edges, textures, and other patterns, progressively extracting higher-level features through multiple layers. While CNNs form the backbone of many computer vision tasks, they can also be extended for pixel-wise classification in semantic segmentation tasks. Semantic segmentation involves classifying each pixel of an image into a specific category to generate a segmentation map. While CNNs are often used as encoders for feature extraction, additional architectural elements like decoders are added to reconstruct spatial information in semantic segmentation tasks. In an encoder–decoder architecture, the encoder reduces the spatial dimensions while capturing high-level features, and the decoder upsamples the feature maps to restore the original resolution. Skip connections are often used to combine low-level spatial information from the encoder with high-level semantic information in the decoder, refining segmentation boundaries.

One of the most effective models for semantic segmentation tasks is DeepLabv3+ (L. -C. Chen et al., 2018), which includes the advantages of depthwise separable convolution, atrous spatial pyramid pooling (ASPP), and encoder–decoder structures in the DeepLab series algorithms (L. -C. Chen et al., 2016, 2017).

2.1.1 | Atrous separable convolution

The atrous separable convolution, a key innovation in DeepLabv3+, combines the advantages of atrous convolution with those of depthwise separable convolution. Atrous convolution introduces the concept of dilation rate to standard convolution operations, allowing the convolutional filter to have gaps between its weights. This dilation rate controls the spacing between the filter weights, enabling it to cover a larger receptive field without increasing the number of parameters or the amount of computation. Mathematically, for an input feature map x and a filter w with a dilation rate r, the atrous convolution output $y[i]$ is defined as

$$y\left[i\right] = \sum_{k} x\left[i + r \cdot k\right] \cdot w\left[k\right] \tag{1}$$

This operation effectively allows the network to capture multi-scale contextual information without expensive computational cost, which is particularly useful in tasks where understanding the context at various scales is important, such as image segmentation.

Depthwise separable convolution, on the other hand, decomposes the standard convolution operation into two simpler operations: depthwise convolution and pointwise convolution. The depthwise convolution applies a single filter per input channel (depth), thus significantly reducing the computational cost. Following this, the pointwise convolution (1 × 1 convolution) combines the outputs of the depthwise convolution. The combination of these two operations results in a more computationally efficient convolution, reducing both the number of parameters and the computational load, while maintaining the representational power of the network. Atrous separable convolution integrates these two techniques, applying atrous convolution in the depthwise part.

2.1.2 | ASPP

A critical component of DeepLabv3+ is the ASPP module, which captures multi-scale contextual information using parallel atrous convolutions with different dilation rates and global context pooling. This allows the network to effectively expand the receptive field without losing spatial resolution. The ASPP includes one convolution with kernel size 1 × 1, three convolutions with kernel size 3 × 3 and dilation rates 6, 12, and 18, alongside a global pooling layer that captures context from the entire input. The resulting features from the ASPP module are concatenated and passed through 1 × 1 convolution to form a feature map that combines diverse contextual information (Figure 1a). By integrating these features, ASPP enhances the network's ability to perceive objects and their surroundings at multiple scales, thus improving segmentation accuracy and performance.

2.1.3 | Model architecture

DeepLabv3+ employs an encoder–decoder structure designed to capture fine details and preserve spatial information (Figure 1b). In the encoder, the input image is processed by a deep convolution backbone network, extracting high-level features.

The ASPP module then captures features at multiple scales by applying parallel atrous convolutions with different dilation rates. This approach enables the network to handle objects of varying sizes effectively. The decoder module integrates low-level features from earlier layers of the backbone with the ASPP output, providing fine spatial details necessary for accurate boundary delineation. A series of convolutional layers in the decoder further refine the segmentation mask. The refined feature map is then upsampled to the original input image size, ensuring that the output segmentation mask matches the input image dimensions. The final 1 × 1 convolution layer produces the output segmentation mask, and the softmax activation function is applied pixel-wise to generate class probabilities. For a pixel at position i in the output feature map y, softmax is computed as

$$\sigma(y)_i = \frac{e^{y_i}}{\sum_{j=1}^{K} e^{y_{i,j}}} \tag{2}$$

where K is the number of classes. The softmax function is followed by cross-entropy loss, calculated between the predicted probability distribution $\hat{p}_{i,c}$ (obtained from softmax) and the ground truth label $p_{i,c}$:

$$Loss = -\sum_{c=1}^{K} p_{i,c} \log\left(\hat{p}_{i,c}\right) \tag{3}$$

During training, the objective is to minimize the average cross-entropy loss across all pixels in the training set. This process trains the DeepLabv3+ model to output probability distributions (via softmax) that closely match the ground truth labels for each pixel. Based on the DeepLabv3+ semantic segmentation network, this paper proposes a pixel-wise defect segmentation algorithm.

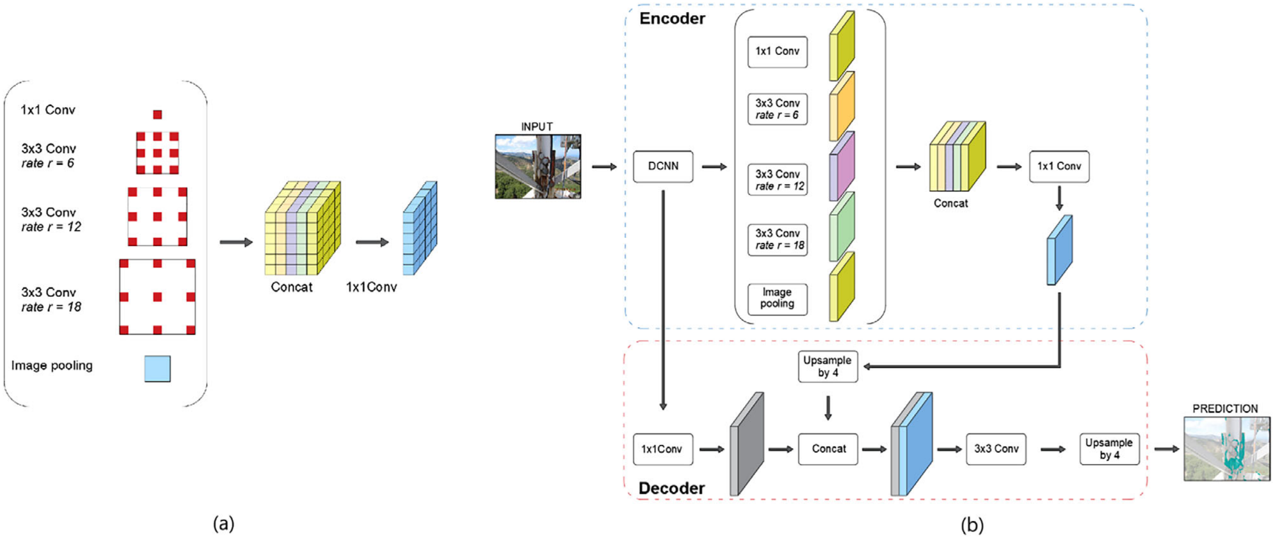


FIGURE 1 (a) Atrous spatial pyramid pooling. (b) Architecture of DeepLabv3+ with backbone network.

2.1.4 | Backbone networks

To enhance the performance of the proposed pixel-wise defect segmentation algorithm, various pretrained deep CNNs (DCNNs) were evaluated as backbone models within the DeepLabv3+ framework using transfer learning techniques (Torrey & Shavlik, 2010). This involves leveraging pretrained DCNN models that have been previously trained on large datasets and fine-tuning them for specific tasks such as defect segmentation. The idea is to transfer the knowledge captured by these pretrained models, which includes hierarchical feature representations learned through numerous layers of convolutional operations. These learned features are generic and broadly applicable across different computer vision tasks due to their ability to recognize edges, textures, and higher-level visual patterns. This approach typically leads to faster training convergence, requires less labeled data for the target task, and generally results in better performance, compared to training models from scratch. In this study, several state-of-the-art DCNN-based backbone architectures were evaluated, namely, ResNet-50 (He et al., 2016), VGG-16 (Simonyan & Zisserman, 2015), VGG-19 (Simonyan & Zisserman, 2015), MobileNetV2 (Sandler et al., 2018), Xception (Chollet, 2017), and InceptionResNetV2 (Szegedy et al., 2017). Each of these backbones brings unique strengths to the task.

ResNet-50, for instance, is a member of the ResNet family designed to address the vanishing gradient problem in very deep networks. It consists of 50 layers, including convolutional layers, batch normalization layers, activation functions, and shortcut connections (residual connections). The network starts with a 7 × 7 convolutional layer with 64 filters, a stride of 2, and padding to maintain spatial dimensions. This is followed by a max-pooling layer with a 3 × 3 kernel, a stride of 2, and padding. The core of ResNet-50 is built from a series of residual blocks, each containing convolutional layers and shortcut connections. These blocks are grouped into four stages, each with varying numbers of blocks: Stage 1 contains three residual blocks, each with 64, 64, and 256 filters; Stage 2 contains four residual blocks, each with 128, 128, and 512 filters; Stage 3 contains six residual blocks, each with 256, 256, and 1024 filters; and Stage 4 contains three residual blocks, each with 512, 512, and 2048 filters. Each residual block consists of three convolutional layers. The first layer is a 1 × 1 convolution that reduces the dimensionality, the second layer is a 3 × 3 convolution, and the third layer is another 1 × 1 convolution that restores the dimensionality. Batch normalization and Rectified Linear Unit (ReLU) activation functions follow each convolution. Shortcut connections directly connect the input of a block to its output, bypassing the intermediate convolutional layers. This helps mitigate the vanishing gradient problem, allowing for the training of much deeper networks. After the residual blocks, the network includes a global average pooling layer that reduces each feature map to a single value by averaging, thus reducing the overall size. The final layer is a fully connected layer with 1000 units, followed by a softmax activation function to produce the output probabilities for the 1000 classes in the ImageNet dataset.

VGG-16 and VGG-19 are deep CNNs known for their uniform architecture. VGG-16 consists of 16 weight layers, including 13 convolutional layers and three fully connected layers. It has five convolution blocks: The first two blocks contain two convolution layers each, while the next three blocks contain three convolution layers each. Each convolution operation uses a 3 × 3 kernel that automatically extracts features from the images. A ReLU activation function is applied after each convolution layer. Each convolution block is followed by a max-pooling layer, resulting in a total of five pooling layers. Each pooling operation uses a 2 × 2 kernel with a stride of 2 and no padding. At the end of the network, there are three fully connected layers; the first two layers have 4096 channels each, and the last layer has 1000 channels, corresponding to the number of classes in the ImageNet dataset. VGG-19 is a variant of the VGG architecture with 19 weight layers, including 16 convolutional layers and three fully connected layers. It also has five convolution blocks: The first two blocks are identical to VGG-16, each containing two convolution layers, while the last three blocks each contain an extra convolution layer, compared to VGG-16, making a total of four convolution layers per block. The activation functions, max-pooling layers, and fully connected layers in VGG-19 are identical to those in VGG-16.

MobileNetV2 is a CNN architecture designed for mobile devices and embedded system where computational resources and power consumption are limited. MobileNetV2 introduces two key innovations: depthwise separable convolutions and inverted residuals with linear bottlenecks. The network begins with a standard convolutional layer followed by a series of bottleneck layers. Each bottleneck layer consists of a 1 × 1 pointwise convolution that expands the input channels, a depthwise convolution that performs spatial filtering independently on each input channel, and another 1 × 1 pointwise convolution that projects the expanded channels back to a smaller number of output channels. This sequence is known as an inverted residual block because it starts with expanding the number of channels before reducing them, contrary to the traditional residual block. After the depthwise convolution, instead of using a non-linear activation function, a linear activation is applied. This helps in preserving information and avoiding the problem of non-linear transformations that can cause loss of useful information, especially in low-dimensional space. Another significant innovation is the shortcut connection used in the inverted residual blocks. This connection skips the intermediate layers and adds the input directly to the output of the bottleneck layer, which helps in preserving gradient flow during training. MobileNetV2 is designed to be efficient with respect to both the number of parameters and the computational cost. The architecture is modular, consisting of a series of these bottleneck layers with different number of filters, expansion factor, and stride to adapt to various tasks. The final part of the network includes a global average pooling layer, followed by a fully connected layer with 1000 units and a softmax activation function for classification.

Xception, short for “Extreme Inception,” is a DCNN architecture that builds upon the Inception model by incorporating depthwise separable convolutions. This design is inspired by the hypothesis that cross-channel correlations and spatial correlations can be entirely decoupled, a principle that guides the construction of the network. Xception begins with an entry flow consisting of standard convolutional layers followed by depthwise separable convolutions. The entry flow starts with two convolutional layers with a kernel size of 3 × 3 and an increasing number of filters. These layers are followed by a series of residual blocks that use depthwise separable convolutions. Each residual block in the entry flow contains a 1 × 1 pointwise convolution followed by a depthwise convolution. This is repeated with a gradually increasing number of filters and downsampling using strided convolutions. The middle flow of the network consists of multiple identical residual blocks that use depthwise separable convolutions. Each block includes a 1 × 1 pointwise convolution followed by a depthwise convolution and then another 1 × 1 pointwise convolution. This series of operations allows the network to capture intricate features while maintaining computational efficiency. In the exit flow, the network transitions from feature extraction to classification. It includes several residual blocks with depthwise separable convolutions, similar to the middle flow but with an increased number of filters. This is followed by a global average pooling layer that reduces each feature map to a single value by averaging, thus reducing the overall size of the output feature map. The final layer is a fully connected layer with 1000 units, followed by a softmax activation function to produce the output probabilities for the 1000 classes in the ImageNet dataset.

InceptionResNetV2 is a deep CNN that integrates the feature extraction capabilities of Inception modules with the training advantages of residual networks. The network begins with an initial convolutional layer followed by several Inception-ResNet modules. These modules are designed to capture multi-scale features using parallel convolutional layers with different kernel sizes, pooling operations and residual connections that add the input of the module to its output within the same module. The outputs from these parallel operations are concatenated, allowing the network to capture diverse features from the input data. This design helps in maintaining the gradient flow during backpropagation, thus enabling the training of very deep networks without suffering from the vanishing gradient problem. The architecture of InceptionResNetV2 can be divided into three main parts: the stem, the middle flow, and the reduction blocks. The stem consists of initial convolutional and pooling layers that process the input image into a lower resolution and higher dimensional feature map. Following the stem, the network includes several Inception-ResNet-A modules, followed by the first reduction block (Reduction-A), which reduces spatial dimensions while increasing the depth of the feature maps. The middle flow of the network consists of

TABLE 1 Pretrained networks properties example.

Networks	Depth	Size (MB)	Parameters (millions)
ResNet-50	50	96	25.6
VGG-16	16	515	138
VGG-19	19	535	144
MobileNetV2	53	13	3.5
Xception	71	85	22.9
InceptionResNetV2	164	209	55.9

multiple Inception-ResNet-B modules that continue to extract features with residual connections, allowing the network to learn complex patterns. Subsequently, another reduction block (Reduction-B) further reduces spatial dimensions. The final part of the network includes several Inception-ResNet-C modules that further refine the extracted features. This is followed by a global average pooling layer that reduces each feature map to a single value by averaging. The final layer is a fully connected layer with 1000 units, followed by a softmax activation function to produce the output probabilities for the 1000 classes in the ImageNet dataset. Table 1 provides additional details on the network architectures used in this study. Depth refers to the number of successive convolutional or fully connected layers from the input to the output layer.

These backbones were integrated into the DeepLabv3+ framework, and their performance was evaluated based on their ability to accurately segment defects at the pixel level. The comparative analysis focused on identifying the backbone that provides the optimal balance between segmentation accuracy and computational efficiency, thereby improving the overall performance of the defect segmentation algorithm.

2.2 | Corrosion metric quantification

Accurately measuring small corroded surface areas on steel transmission towers is crucial for assessing structural integrity and planning maintenance strategies. While image processing techniques and AI have been widely used for corrosion detection, they often do not deal with providing precise localization and metric quantification of corroded surfaces within complex structures. This limitation stems from the reliance on single-image processing and per-pixel analysis, which typically results in corrosion being quantified as a percentage of the total image area, with the area merely expressed in pixel units. Even when UAVs are employed for acquiring multiple images, the redundancy of visual data is primarily utilized to enhance the robustness of classification models rather than to achieve accurate metric measurements (Fei et al., 2021). Consequently, there is a need for methodologies that not only detect corrosion but also provide reliable metric assessments of its extent and distribution across intricate structural geometries.

UAV photogrammetry provides a more comprehensive solution to the challenges of corrosion quantification and localization in complex structures (Wu et al., 2023). By generating accurate 3D models and georeferenced data, UAV photogrammetry overcomes the limitations of traditional image processing techniques. The application of SfM from multiple aerial images enables the resolution of projective geometrical constraints between the 2D images captured by a moving camera and the 3D object in the real world. This approach, supported by positioning techniques, like GNSS and topographic land surveying, allows for precise metric measurements and detailed spatial analysis, ensuring that the corroded areas are not only detected but also accurately quantified and localized locally within the intricate geometries of steel transmission towers and globally among the earth (Liu et al., 2020).

Several methods are available to compute the surface areas of real objects from photogrammetric products, each with its strengths depending on the specific application and level of detail required. One widely used approach is the triangulated irregular network (TIN) method, which divides the surface into non-overlapping triangles (facets). This method is particularly effective for accurately representing complex terrains, including steep slopes and uneven surfaces, as it calculates the total surface area by summing the areas of all the individual triangles. The TIN method is especially advantageous in scenarios where the surface geometry is highly irregular or where detailed topographic information is critical.

Another commonly used technique is the grid-based method, which relies on the digital surface model (DSM) generated through photogrammetric procedures. In this method, each pixel or cell of the DSM represents an orthorectified portion of the object's surface. The slope of each cell is computed and used to adjust the planimetric area covered by the same cell.

In traditional land surveying, where aerial photogrammetry originated, adjusting the planimetric area based on slope is often deemed unnecessary due to the high resolution of DSMs and orthomosaics, typically with pixel sizes on the order of a few centimeters. In such cases, ensuring the validation of the orthophoto against ground-truth measurements, the difference between the adjusted and planimetric areas is within the uncertainties inherent to the photogrammetric process, making the adjustment negligible.

The same principle is true for close-range applications of UAV photogrammetry, such as communication tower inspections. In these scenarios, the proximity of the UAV

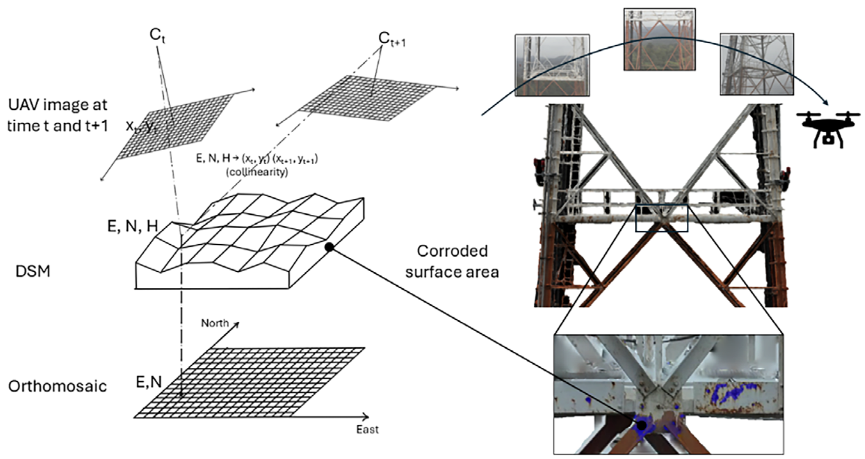


FIGURE 2 Workflow for corroded surface areas quantification on a steel transmission tower using UAV-based photogrammetry. The process begins with unmade aerial vehicle (UAV) imagery captured at different time intervals, followed by the creation of a digital surface model (DSM) and an orthomosaic. The corroded surface areas are then automatically detected through the analysis of these photogrammetric products. The final step involves quantifying the identified corroded regions, enabling precise localization, and assessment.

to the structure, combined with the high resolution of modern digital cameras, often results in DSMs with cell resolutions of just a few millimeters. As a result, the discrepancy between the adjusted and planimetric surface areas is minimal, falling within the inherent uncertainties of the photogrammetric procedure (Figure 2).

However, for vertical surfaces like transmission towers, planar orthomosaics inherently introduce approximation errors due to local surface curvature and irregularities of the corroded surface. For relatively flat or single-curved surfaces, these errors are minimal, but for highly irregular or double-curved geometries, the error can be significant. The error can be quantified by considering three main factors:

Local surface curvature K: The curvature-induced error ($\Delta A_{curvature}$) arises because the planar orthomosaic flattens the 3D geometry. This can be expressed as

$$\Delta A_{curvature} \approx \int_S \frac{K}{2} dA \tag{4}$$

where S is the corroded surface area, and K is obtained from the partial derivatives of the surface height $z(x, y)$ as

$$K = \sqrt{\left(\frac{\partial^2 z}{\partial x^2}\right)^2 + \left(\frac{\partial^2 z}{\partial y^2}\right)^2} \tag{5}$$

Resolution r: The resolution of the orthomosaic affects how accurately it captures the details of the corroded surface. Smaller resolution reduces error. The resolution error is expressed as

$$\Delta r \approx \frac{p \cdot r}{A} \tag{6}$$

where p is the corroder area perimeter, and A is the real 3D surface.

Noise and artifacts: imperfections in orthomosaic generation process can also contribute to deviations but are negligible with respect to the resolution and the curvature induced error.

Given these errors factors, the orthomosaic approximation error is computed as

$$\varepsilon = \frac{\Delta A_{curvature} + \Delta r}{A_{3D}} \tag{7}$$

This error grows with the curvature increasing and decreases with finer resolution. Considering a planar projection of the DSM into an orthorectified map, the 3D area of the corroded surface can be expressed as

$$A_{3D} = A_{ortho} + \varepsilon \tag{8}$$

Despite these sources of error, the high resolution of modern UAV orthomosaics (e.g., with cell sizes of a few millimeters) and the relatively flat geometry of the tower beams ensures that the discrepancies are generally within acceptable limits for practical applications. Consequently, the ε can be considered equal to zero and the corroded surface area is computed by summing the area of each pixel identified in the tower orthomosaic. The result is a planar surface area A_{ortho} calculated as

$$A_{ortho} = \int_S dxdy = \sum_{i=1}^{M} \sum_{j=1}^{N} \Delta x \Delta y \tag{9}$$

where M and N are the number of grid cells (pixels) in the orthomosaic in the x and y directions, respectively. Δx and Δy represent the dimensions of each grid cell in the x and y directions, expressed in real-world units. This formula sums the area of all the pixels representing the corroded region.

However, it is important to consider that in cases where sub-millimeter level precision is required, or where the surface geometry is highly complex, even these small discrepancies could become significant.

3 | TRAINING PROCESS

In this study, a DeepLabv3+ model was utilized for pixel-wise classification of images depicting steel transmission towers, with a specific focus on identifying and segmenting corroded areas. The networks were trained using manually labeled images, where each pixel was annotated as either “corrosion” or “background.” Through iterative optimization, the network learned to accurately predict the class of each pixel. By leveraging transfer learning, pretrained models on large image datasets were employed to enhance the network’s performance. This strategy allowed the network to benefit from previously learned features, thereby reducing the required amount of training data and computational resources. The neural network’s output is a segmentation map that provides a detailed and quantitative assessment of corrosion on the transmission towers. This pixel-wise classification enables precise localization and measurement of defects, thereby facilitating more effective monitoring and maintenance strategies.

To define the most effective architecture for the automatic identification of corroded areas on transmission towers, the study was conducted in two phases. The first phase aimed to identify the optimal neural network architecture through comparative analysis using a limited dataset of images. The second phase focused on hyperparameter tuning to optimize the performance of the network selected in the first phase, utilizing the entire dataset. The experiments were conducted with MATLAB (R2023a) platform functions on a Windows 10 PC equipped with an Intel Core i7-10750 processor, RAM of 32 GB, and an NVIDIA GeForce GTX 1650Ti graphical processing unit (GPU).

3.1 | Corrosion dataset

The performance of a neural network is closely related to the quality of its training dataset. To develop a robust dataset for training neural networks effective in segmenting images acquired by drones, it is essential to use real images captured during on-site inspections of transmission towers. Using realistic on-site images allows the dataset to capture the variability and operational challenges of actual inspection conditions, enhancing the ability of the model to generalize effectively. The collected images for this study encompass a range of conditions, including various angles, resolutions, lighting environments, paint conditions, and noise factors, all of which reflect the complexities of real-word inspections.

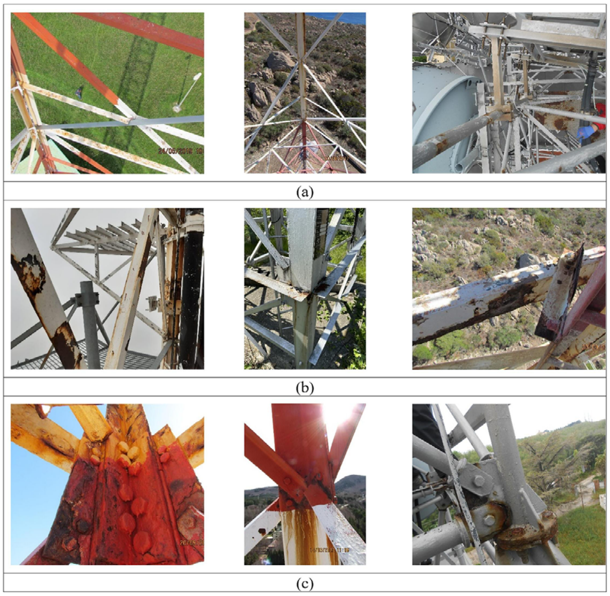


FIGURE 3 Examples of images collected from on-site inspection, used to build the datastore, representing: (a) the complexity and varying resolutions of analyzed area, (b) different types of backgrounds and steel elements, and (c) varying lighting conditions.

Figure 3 illustrates examples of these scenarios, showcasing the diversity typically encountered in visual inspections of transmission towers. This dataset is tailored to the requirements of transmission tower inspections, and it differs from other applications, such as steel bridge inspections or rebar corrosion detection in several key aspects. The image backgrounds vary significantly as transmission towers are typically located in open or rural environments, whereas bridges are often in urban or industrial settings. The structures themselves also differ: transmission towers tend to exhibit more complex geometry, compared to bridges or rebar frameworks. Additionally, the paint on transmission towers is distinctive, often employing specialized coatings for aviation visibility and extreme weather conditions, which differs from the paints commonly used for bridges. Furthermore, transmission towers often experience oxide flow caused by water percolation that creates irregular corrosion patterns. In contrast, corrosion on bridges or rebar is generally more localized, often forming in specific areas. These structural, environmental, and material differences underscore the relevance of this dataset for the specific task of detecting and quantifying corrosion on transmission towers. The dataset images had varying resolutions, ranging from a minimum of 640 × 480 pixels to a maximum of 5312 × 2988 pixels, presenting a realistic challenge for the neural networks. All images were carefully selected to ensure that the dataset reflects the diversity of real-world scenarios encountered during transmission tower inspections. The final dataset consists of 999 images from different sensors, referred to different communication tower located all around Italian territory and acquired from different angle and distance. The data were provided by the personnel in charge of inspect the infrastructure and were acquired in different period of the year along several years. The characteristic of the dataset ensures the required heterogeneity to train a model able to be generalized in every environmental condition, providing a robust foundation for training and validating the deep learning models. The data were partitioned such that 80% was used for training and 20% for validation. This partition allowed the performance of the model to be evaluated on unseen data, thus reducing the risk of overfitting. To prevent data leakage and ensure a robust evaluation, spatial and temporal independence was maintained, ensuring that same images were not shared between the training and validation sets.

Using a fixed input image size is crucial to ensure that the model learns features in a consistent and standardized manner, regardless of the original image dimensions. In computer vision model training, resizing images to smaller dimensions is a common practice to reduce the computational cost associated with processing large images. In this research, to optimize the use of the available database while preserving the rich information contained in high-resolution images, an alternative approach was adopted. Instead of resizing images to a lower resolution, a cropping process was implemented. This involved extracting several lower-resolution images from each original high-resolution image. This strategy is particularly effective in preserving the amount of information, as it maintains the total pixel count of the original resolution.

Considering that neural networks such as ResNet-50, VGG-16, VGG-19, and MobileNetV2 require input images of size 224 × 224, while Xception and InceptionResNetV2 operate with input images of size 299 × 299, two distinct databases were created. Both databases were constructed by cropping images from a total of 500 high-resolution images in order to reduce the computational cost and training time for the first phase of comparative analysis aimed at detecting the most effective architecture. The first database includes 60,567 images tailored for networks with an input size of 224 × 224, while the second database comprises 32,801 images suited for networks requiring an input size of 299 × 299. For the second phase of the study, aimed at hyperparameter tuning of the most effective architecture, cropping was performed on all 999 images, resulting in a third dataset of 105,754 images with 224 × 224 resolution.

Ideally, all classes should have the same number of observations. However, a common problem in datasets is the presence of a higher number of background pixels, compared to those of the object being classified. This occurs because the background typically occupies a larger portion of the image. If this imbalance is not adequately addressed, it can bias the learning process in favor of the dominant classes, thereby neglecting the minority class. To achieve a more equitable balance between classes, only images containing pixels related to corrosion were selected from the previously defined databases, while those containing only the "background" class were discarded. This process resulted in the creation of three databases with final sizes of 14,397, 9811, and 33,584 images. Table 2 provides the details of the resulting databases.

TABLE 2 Dataset properties.

Dataset	Images resolution	Images number	Pixel "corrosion"	Pixel "background"
1	224 × 224	14,397	179,324,264	540,153,576
2	299 × 299	9811	175,629,089	698,620,084
3	224 × 224	33,584	408,705,736	1,268,900,144

As shown in Table 2, despite the application of the filtering operation, an imbalance remains in favor of the "background" class, with a ratio of 3:1 in the first and last databases, and a ratio of 4:1 in the second database. To address this, the median frequency balancing method was used to improve the training, where the weights of the classes were calculated as

$$classWeights = \frac{median\,(frequency)}{frequency} \tag{10}$$

where *frequency* is the number of pixels for a class divided by the total number of pixels. This method assigns higher weights to minority class samples and lower weights to majority class samples when calculating the loss function. In this way, the model focuses more on the minority class, thereby improving predictions for this category.

3.2 | Experiment and result analysis

Once the dataset was constructed and the model architecture defined, the model hyperparameters must be configured to start the training process. As these are external to the networks, their values cannot be estimated directly from the data but can be set using heuristics. Thus, for the first phase of the preliminary comparative analysis, a fixed number of five epochs, a mini-batch size of eight images, a learning rate of 0.1, a momentum of 0.9, and an L2 regularization of 0.0001 were considered for the training process of each network. Gradient descent was used as the optimization method, iteratively optimizing weights and biases through the partial derivatives of the loss function.

A comparative study was first carried out by evaluating how well the trained neural network models performed on the validation dataset. This was done by taking into account the validation loss and calculating the percentage of correctly classified pixels according to:

TABLE 3 Neural networks performance.

Network	Validation loss	Validation accuracy	Intersection over union	Training time (min)
ResNet-50	0.37	80.4%	64.8%	144
VGG-16	0.31	83.8%	70.6%	190
VGG-19	0.38	89.1%	75.7%	236
MobileNetV2	0.31	87.9%	75.9%	92
Xception	0.26	89.2%	78.4%	221
InceptionResNetV2	0.31	90.8%	78.6%	504

$$GA = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

where GA is the global accuracy, TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives. In addition to validation accuracy and validation loss, which are metrics related to prediction accuracy and reliability, the intersection over union (IoU) metric was also used to evaluate the performance of the models without the influence of classes distribution. IoU, defined as the area of overlap between predicted segmentation and the ground truth divided by the area of their union, provides a more comprehensive understanding of how well the model predicts the spatial arrangement of pixels. Training time was also considered as a benchmark for evaluating computational efficiency. Reduced training time is particularly valuable in production environments with iterative development workflows, where models may need to be retrained frequently. In addition, the size of the model can impact both inference time and memory requirements, making it an important consideration for deployment in environments with limited resources or memory. Therefore, Table 3 provides a summary of the metrics considered for the comparative analysis of the models, including validation loss, validation accuracy, IoU, and training time.

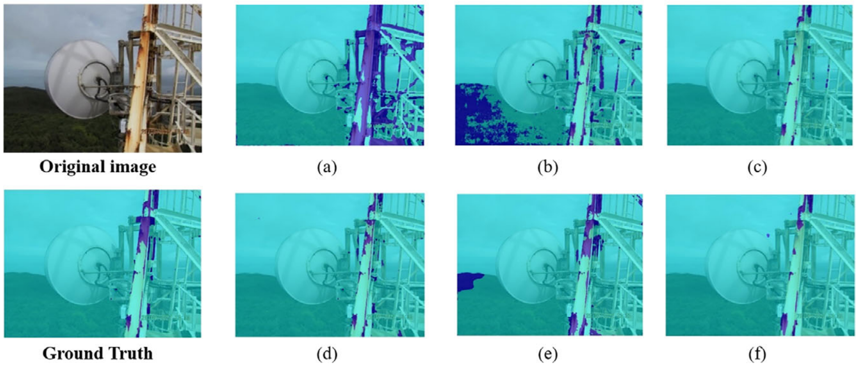
As reported in Table 3, InceptionResNetV2 and Xception emerged as the top performers, achieving validation accuracies of 91.8% and 91.0%, with validation losses of 0.31 and 0.26, respectively. They also demonstrated high IoU scores of 78.6% and 78.4%, indicating high performance in accurately delineating object boundaries. However, the MobileNetV2 network was 2.8% less accurate than the most accurate InceptionResNetV2 network but had the shortest training time among the evaluated networks, completing in just 92 min. In addition, the smaller model size of MobileNetV2 contributed to reduced memory requirements, making it more suitable for use in resource-constrained environments. Therefore, MobileNetV2 was chosen as the reference architecture for this study, balancing performance and efficiency and memory usage. As an example, Figure 4 shows the comparison of the segmentations performed by the neural networks listed in Table 3. This analysis is performed on a test image that was not previously used to train the networks.

The example shown in Figure 4 confirmed the specific performance of each network, demonstrating lower accuracy in the cases of ResNet-50 and VGG-16 (Figure 4a,b), while superior performance was highlighted for VGG-19, MobileNetV2, Xception, and InceptionResNetV2 (Figure 4c–f). Once the architecture for segmenting the corroded area was defined, empirical hyperparameter tuning was performed to optimize the performance of the MobileNetV2 network on Dataset 1 as detailed in Table 2. An extensive search was performed to define the best setting of these parameters. In order to determine the effect of each of these parameters, they were varied one at a time while keeping the other constant. The final configuration of tuned hyperparameters, which ensured the convergence behavior and avoided overfitting, resulted in a learning rate of 0.01 with a drop of 0.1 after a period of four epochs, a momentum of 0.9, a regularization of 0.0001, seven epochs and a mini-batch size of 32. The final validation loss and validation accuracies obtained after a training period of about 3 h were 0.22 and 90.6%, respectively. In the final phase of this research, the network defined in the previous phases was trained using the entire image dataset, according to Dataset 3 as defined in Table 2. The training process used the same set of hyperparameters defined in the previous experiment. Overall, the process required a training time of about 9 h and resulted in a validation accuracy of 90.8% (Figure 5) and a validation loss of 0.23 (Figure 6).

To evaluate the performance of the trained and validated network, Figure 7 shows some examples of segmented images that were never seen during the training process.

The first row shows the original images, the second row shows the ground truth labels, and the third row shows the prediction of the network. This visual comparison highlights the strengths and potential weaknesses of the network's performance, providing valuable insight into its practical applicability for tower monitoring and maintenance. By comparing the ground truth labels with

FIGURE 4 Example of semantic segmentation performed with (a) ResNet-50, (b) VGG-16, (c) VGG-19, (d) MobileNetV2, (e) Xception, and (f) InceptionResNetV2 network.



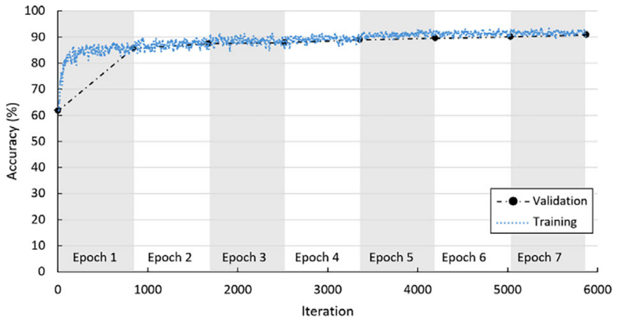


FIGURE 5 Accuracy of the network during training progress of the final experiment.

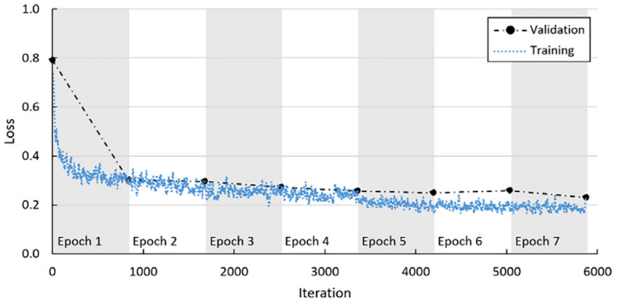


FIGURE 6 Validation loss of the network during training progress of the final experiment.

the network predictions, the effectiveness of the network in accurately segmenting corroded areas can be assessed. There is a remarkable agreement between the identified corroded areas, even in cases with complex thin areas, such as those surrounding bolts in the first example and the railings in the fourth. In some cases, the neural network outperforms manual labeling by detecting areas that were missed during manual labeling as can be seen on the bolt cap in the second image and around the first bolt in the last example. Conversely, there are cases where the neural network failed to detect certain corroded regions labeled in the ground truth, particularly noticeable in the sixth image, where corrosion along the edges of diagonal braces was overlooked. It is possible that the network interpreted this consistent deterioration along edges as part of the structure rather than as corrosion Furthermore, there are a few cases where the network misclassified background regions as corrosion, primarily in the first, second, and final images. This misclassification could be attributable to the cropping operation used during the construction of the dataset, which refines the network's ability to detect detailed features, but may disrupt the spatial relationship between corroded areas and the background. The missed regions along the diagonal braces may have significant practical implications, as such areas often represent critical points of structural vulnerability. Overlooking these regions could lead to underestimation of the overall corrosion extent, potentially affecting maintenance planning and prioritization. Additionally, misclassifications in the background, though less critical, might introduce noise into the quantification process, necessitating manual verification in certain cases. While detection failures are challenging to eliminate entirely due to dataset imbalance and annotation uncertainties, background misclassifications will benefit from the proposed method, where the integration of the 3D information helps to resolve spatial ambiguities.

In Section 4.3, the benefits of the UAV photogrammetric reconstruction in addressing this aspect will be described. In summary, the performance of the network can be considered acceptable, especially considering the complexity of the task, which involves irregular areas with no clearly defined boundaries for detection, often further complicated by oxide percolation.

4 | EXPERIMENT OF CORROSION DETECTION IN TRANSMISSION TOWER

In the practical application phase of this research, the trained neural network has been used for the semantic segmentation of corrosion on a transmission tower. The structure to be analyzed is the South-East tower of the Torino Eremo broadcasting center, located at Strada Comunale di Pecetto, 311/15, 10131–Torino, which belongs to Ray Way S.p.A. and support the Italian public radio television broadcasting. The Torino Eremo is a significant historical site as it represents the starting point of public

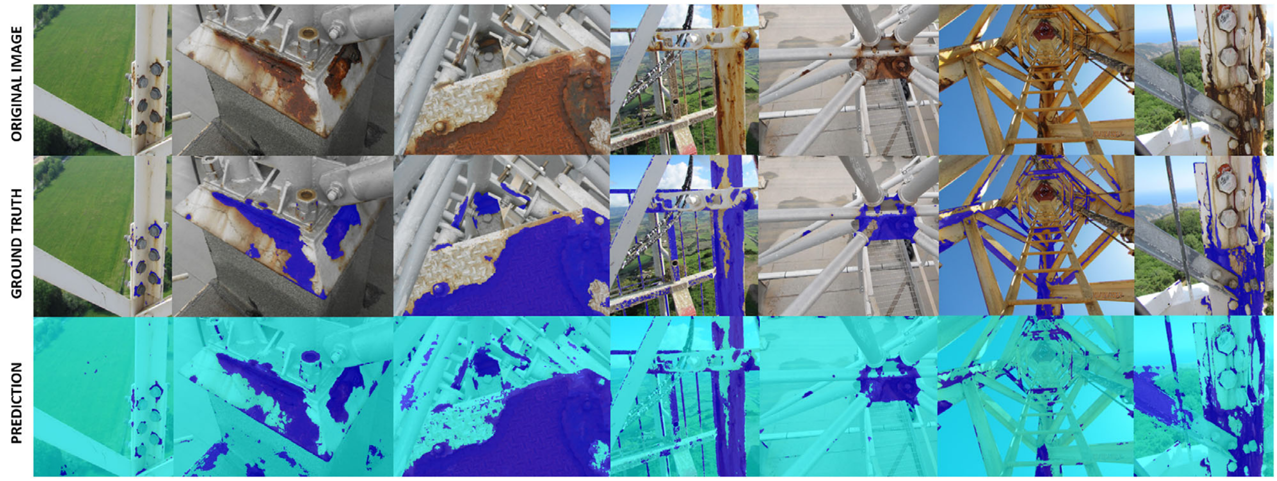


FIGURE 7 Examples of predictions generated by the proposed network.

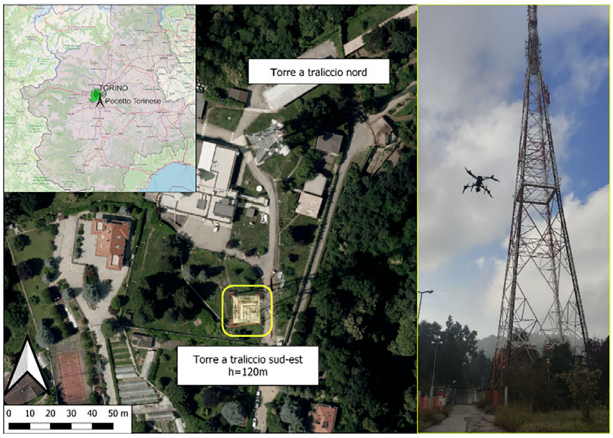


FIGURE 8 Study site and UAV inspection of the south-east steel transmission tower located at the Rai Way Transmitter Center of Torino Eremo, Italy. The map on the left provides a geographical context of the study area within the Piedmont region. The central aerial view highlights the northern and southern towers, with the southern tower (height = 120 m) marked for detailed inspection. On the right, an image captures the UAV in action during the structural inspection of the southern tower.

broadcasting in the country. Approximately 120 m high and currently out of service, the tower is protected by a white and red coating, making it an excellent case study for this analysis (Figure 8). In this real-world scenario, the network was used to accurately identify and segment areas affected by corrosion from high-resolution photogrammetric products, originated by the processing of tower pictures taken by drone under different lighting conditions and angles.

4.1 | Flight planning

The preliminary steps of the drone survey involved several important tasks. First, it was essential to confirm that the structure to be surveyed was not located within a no-fly zone. This was followed by a preliminary site inspection to identify potential risks and obstacles that could affect the survey. Subsequently, a detailed topographical survey was conducted to gather the necessary data for georeferencing and reconstruction error assessment. Extensive documentation and data collection were then performed to ensure all the necessary information was gathered. A number of criteria were used to select the UAV, including appropriate size for the environment, stable flight capability, obstacle proximity sensors, Real Time Kinematic (RTK) positioning accuracy, compatibility with different camera and equipment types, high autonomy, high-resolution cameras, and sufficient storage capacity. For this survey, the Matrice 300 RTK drone coupled with the ZENMUSE P1 camera was selected. The ZENMUSE P1 camera features a 45 MP CMOS optical sensor with a 35 mm lens, providing high-resolution imaging capabilities suitable for detailed aerial inspections and photogrammetric analysis.

However, the central aspect of the planning was the definition of the flight path, which is directly related to the fundamental principles of photogrammetric reconstruction. Planning the flight trajectory with an unmanned aerial system is a crucial step to ensure that the derived products meet the required project specifications due to its direct connection to the theoretical principles of photogrammetric reconstruction. Research in the field has led to the development of specific geometric acquisition schemes that are both metrically valid and energy-efficient (minimizing energy consumption during flight). These schemes, widely recognized and implemented in specialized algorithms, are traditionally applied to nadiral acquisition patterns aimed at surveying flat or large land areas.

The most important parameters and formulas for photogrammetric mission planning are well-documented in the literature (Kraus, 2007), with specific flight planning parameters summarized by Eisenbeiss (2009).

However, when surveying elevated structures, such as telecommunication towers, modern acquisition schemes are required. These advanced schemes are not yet fully implemented in most flight planning software and are particularly tailored to capturing vertical or irregularly shaped objects. The two primary acquisition schemes suitable for these applications are the spiral acquisition scheme and the ascending/descending trajectory scheme. In the first, the drone follows a helical path around the structure, capturing images from varying angles and distances, ensuring uniform coverage. In the second, the drone moves vertically along the structure in a systematic pattern, with the camera oriented forward (perpendicular to the tower's surface) to capture detailed images of the vertical geometry.

For the telecommunication tower in this study, the ascending/descending trajectory scheme was chosen due to its suitability for tall, linear structures. The flight path was manually planned by leveraging the expertise of both the Unmanned Aerial System (UAS) pilot and the photogrammetry specialist to adapt the scheme to the unique characteristics of the structure. This collaborative approach ensured that the flight parameters (e.g., horizontal distance, overlap, and altitude) were optimized to maintain a consistent ground sample distance (GSD) of approximately 1 cm/pixel, critical for detailed inspection and analysis. Moreover, the camera orientation was fixed forward-facing, directly targeting the tower's surface, to maximize the resolution of the captured features. The trajectory provided sufficient coverage of the entire structure, including protruding and inclined components, by adjusting the drone's flight path dynamically during the survey.

The survey must be planned to meet the project's specified accuracies σ_r and tolerances T_r. The key planning parameter to consider is the GSD, which must satisfy the following condition:

$$GSD \leq \sigma_r \tag{12}$$

The calculation of the GSD is performed using the following simplified formula:

$$GSD = \frac{H \cdot d_{pix}}{F_R} \tag{13}$$

where H is the flight distance from the object façade, d_{pix} is the pixel size of the specific camera, and F_R is the focal length of the camera optic. Considering the operational scenario and flight safety conditions, a minimum distance from the tower of 10 m is assumed ($H = 10\,\text{m}$). The sensor to be used for the photogrammetric survey is selected, and the parameters d_{pix} and F_R are determined. Considering the Zenmuse P1 optical sensor, the focal length is $F_R = 35$ mm and the pixel size is $d_{pix} = 4.4$ µm. The theoretical (nominal) GSD is calculated using Equation (13). Thus, the nominal GSD for this survey is 0.126 cm/pixel.

4.2 | Photogrammetric reconstruction

The photogrammetric survey aimed to acquire high-resolution RGB images for the radiometric and geometric reconstruction of the 3D object. Close-range aerial photogrammetry was employed, using SfM algorithms to reconstruct a digital model of the observed object from the 2D image content. This objective was achieved by accurately determining the camera positions during data acquisition (external orientation parameters), thus establishing the spatial relationship between the images and the object under analysis. All photogrammetric analyses were carried out using Agisoft Metashape v2.0.3 software, following a sequential procedure that included image matching, relative camera orientation, absolute block orientation via ground control point or image geotag, dense matching, and finally products generation. For each of these steps, photogrammetric outputs are produced like point clouds (sparse or dense), triangulated 3D mesh, textured 3D models (Figure 9a).

In addition, DSMs and orthophotos were also generated to provide an accurate representation of the surveyed area geometrically corrected. The topographic survey, performed with the integration of GNSS and Total Station, allows to validate the orthomosaic against ground-truth measurements with a root mean square error of 0.87 mm. This orthmosaic is critical for semantic segmentation of the corroded area, ensuring accurate representation for detailed analysis and inspection purposes. The orthomosaic was produced with a pixel resolution of 3 mm, about two times the theoretical GSD. With the same pixel resolution has been produced the DSM (Figure 9b). In this study, four high-resolution orthomosaics were generated, one for each façade of the transmission tower. The analysis was specifically concentrated on the outer surfaces of the structural beams, as these areas are most exposed to environmental factors and therefore more prone to corrosion. This focused approach ensured precise quantification of the corroded areas while minimizing interference from less relevant structural components. From a computational perspective, the photogrammetric processing was conducted on a workstation equipped with an Intel Core i7-7800X CPU (3.50 GHz, 12 cores), 96 GB of RAM, and two GPUs: an Intel UHD Graphics 630 (24 compute units, 19,600 MB global memory, OpenCL 2,1) and an NVIDIA Quadro P1000. The processing of 3182 images, each with a resolution of 8192 × 5460 pixels, required 7 h and 30 min

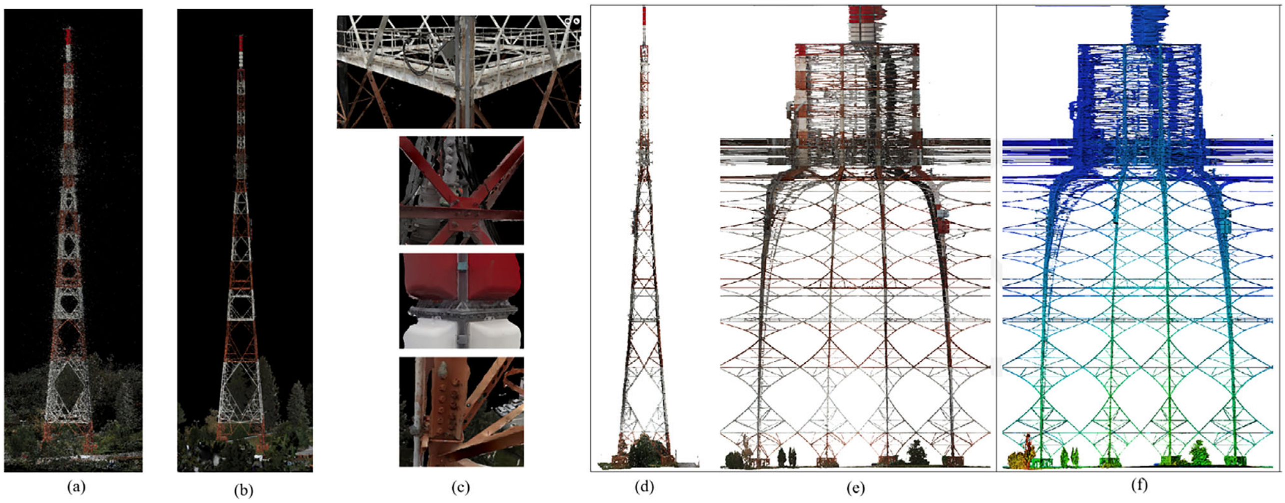


FIGURE 9 On the left: Photogrammetric products derived from a UAV-based survey. (a) A sparse point cloud. (b) A triangulated mesh. (c) Detailed close-up views of critical components, including the tower's joints. On the right: Comparative visualization of orthomosaic and DSM of the tower, (d) an orthomosaic showing the southern façade of the structure in high resolution; (e) a high-resolution orthomosaic developed in a cylindric projection. (f) A color-coded DSM illustrating the spatial distribution and distance of structural elements again in a cylindric projection.

for alignment, 10 h and 50 min for generating the 3D triangulated model, and an additional 9 h and 15 min to produce an 8K texture for the entire communication tower.

4.3 | Corrosion area calculation

Currently, the need and urgency for maintenance are often determined in an approximate and qualitative manner through rapid surveys of infrastructure. Each defect on the structure is assessed for its extent and intensity using constant coefficients without quantitative analysis. The complexity, level of detail, and cumbersome nature of traditional surveys are inversely related to the number of infrastructures to which they are applied and the certainty of the results. In the proposed approach, orthophotos generated from photogrammetric reconstruction were used, and the trained neural network was directly applied to segment the corroded areas on the transmission tower. Each orthophoto was processed to accurately identify and label the corroded areas. The duality between orthorectification and 3D reconstruction is demonstrated in Figure 10 where the segmented texture have been applied to the 3D model. Once the damage was detected, it could be possible to extract morphological information to determine durability, exposure conditions, and define economic and safety implications.

The segmented orthomosaics were then analyzed to calculate the total area affected by corrosion. This calculation involved summing the pixel areas classified as corroded and converting this sum into real measurements based on the scale and resolution of the orthophotos (Figure 10). As stated in Section 2.2, Equation (9) has been used to compute the corroded surface area. To this purpose, the software QGIS 3.40.1 has been used. The orthomosaic provides all the required metrical and radiometrical content to perform a quantification of the pixel corroded. In particular, the pixels labeled as corroded have been statistically analyzed with the *r.univar* tool from Geographic Resources Analysis Support System integrated in QGISS. The tool calculates univariate statistics from the non-null cells of a raster map (i.e., the corroded pixels) and in particular the sum of pixels of the cells inside a specific section. Being the orthomosaic georeferenced, it is possible to extract the coordinates of the cells labeled as corroded, and automatic vectorization procedure can provide the centroid of each corroded area with attached coordinates.

Additionally, since the communication tower was modeled through photogrammetry, the resulting DSM has been used to orthorectify the orthomosaic, ensuring an accurate projection of the structure. This approach effectively removes the background, as the orthomosaic focuses solely on the tower's geometry and features, effectively eliminating the risk of background misclassification. The results provided a quantitative assessment of the extent of corrosion, which is crucial for planning maintenance and repair strategies. Table 4 lists the amount of corroded area identified by the neural network for the locations shown in Figure 10.

While orthomosaics have been proven robust for surface measurements and areal quantification on quasi-nadiral object, they intrinsically deform complex 3D geometries during the 2D re-projections on a cartographic surface. The presence of DSM products, resulting from the SfM

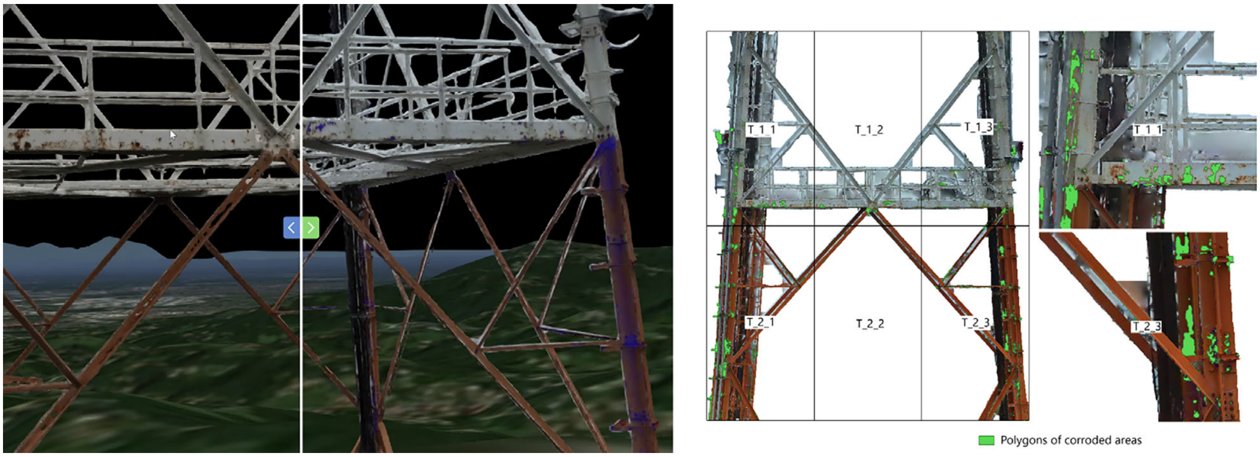


FIGURE 10 On the left: Texturized 3D model of the transmission tower at Torino Eremo. The sliding window visualization allows to compare the original model with the segmented one for corroded area identification. On the right: Detailed inspection of a steel transmission tower highlighting corroded areas. The image is divided into sections (T_1_1 to T_2_3), with green polygons indicating regions of corrosion detected through automated analysis. The enlarged views on the right side provide a closer inspection of specific corroded sections.

TABLE 4 Quantification of the corroded area.

Tassel	Tassel size	Structure pixel count	Corrosion pixel count	Corrosion percentage	Corrosion area
T_1_1	1003 × 1817 pixels	1,027,850 pixels	5094 pixels	0.50 %	0.046 m^2
T_2_1	1003 × 1817 pixels	435,743 pixels	1601 pixels	0.37 %	0.014 m^2
T_3_1	1003 × 1817 pixels	887,330 pixels	427 pixels	0.05 %	0.004 m^2
T_2_1	1003 × 1817 pixels	852,535 pixels	3080 pixels	0.36 %	0.028 m^2
T_2_2	1003 × 1817 pixels	84,119 pixels	129 pixels	0.15 %	0.001 m^2
T_2_3	1003 × 1817 pixels	648,039 pixels	3703 pixels	0.57 %	0.033 m^2

approach, mitigate partially these effects taking into account relief displacement and standardizing the scale. Additionally, with DSM-derived products, such as slope and aspect maps, it is possible to implement curvature correction algorithm in case of limited irregular geometries. Reticular steel communication towers, like the one analyzed in this study, predominantly consist of regular geometric structures with flat beams and minimal curvatures. These characteristics are advantageous for our analysis, allowing for accurate estimation of most corroded areas while ensuring comprehensive and consistent coverage of the pre-allocated orthophoto's projected space. An accurate design of the projection space is also required to avoid occlusion problems i.e. situations where objects in foreground block the view of objects in background.

However, the proposed methodology may face limitations when applied to part of the structures with highly irregular geometries or double-curved surfaces, such as connections, joints, or complex brackets often present in intricate steel frameworks. In such cases, the orthorectification algorithm as well as the interpolation methods used to map the pixel colors from the image coordinate system to the orthophoto coordinate system, introduces inaccuracies. Possible solutions include employing advanced algorithms for curvature-aware surface correction or estimating corroded areas directly on TIN, which could significantly enhance the accuracy of the quantification process.

The proposed deep learning-based inspection approach not only automates the process but also provides valuable data to reconstruct damage evolution without operator error. The neural network's ability to perform accurate segmentation on high-resolution orthophotos demonstrates its potential for practical applications in structural health monitoring and maintenance of transmission towers.

5 | CONCLUSION

Automatic corrosion segmentation and quantification is a leading research topic driven by advances in computer technology, particularly in AI. Despite significant progress, it remains an unresolved issue in the context of steel transmission towers. The challenges it poses for structural integrity and safety are substantial and ongoing. This study addresses the pressing need for efficient and accurate corrosion detection in steel transmission towers by using advanced deep learning techniques. The proposed approach uses a DeepLabv3+ model to achieve semantic segmentation of corroded areas on steel transmission towers. The network was trained and validated using a robust dataset of 999 field photographs of different resolutions, ensuring the model's adaptability to different imaging conditions. Among the various pretrained networks evaluated, MobileNetV2 emerged as the optimal choice due to its superior balance of accuracy and computational efficiency. Fine-tuning of the hyperparameters resulted in an acceptable validation accuracy of 90.8% and a validation loss of 0.23.

One of the major contributions of this research lies in the integration of UAV-based photogrammetric products, such as DSM and orthomosaics, with deep learning algorithms for corrosion detection and quantification. The network was used to process orthopmosaics generated from photogrammetric reconstructions of the south-east tower of the Torino Eremo broadcasting center. Thanks to the detailed and metrically accurate spatial representation of the structure, the methods ensure accurate surface area estimation. The accuracy of the orthomosaic was validated against ground-truth measurements using a topographic survey that integrated GNSS and Total Station, achieving a root mean square error of 0.87 mm. This high level of accuracy ensures that the spatial data derived from the orthomosaic is both reliable and accurate. The 3D model of the structure enables the neural network to focus only on the structural elements while eliminating background misclassification problems. Moreover, being georeferenced, the classified orthomosaic enables the extraction of corroded cells coordinates in a given reference system. This capability is highly advantageous for the development of geospatial databases and the creation of digital twins, providing precise spatial information essential for advanced analysis and monitoring. This practical application demonstrated the network's ability to accurately segment corroded areas in high-resolution images, providing a detailed and quantitative assessment of corrosion. Such precise measurements are essential for effective maintenance and repair strategies to improve the longevity and safety of critical infrastructure.

The automated nature of this deep learning-based inspection approach offers several advantages over traditional methods. It reduces the dependency on manual inspections, which are often time-consuming, labor-intensive, and subject to human error. By providing reliable and consistent data, the proposed method facilitates better decision-making and resource allocation in maintenance operations. Furthermore, the ability to monitor damage evolution over time without operator error is a significant step forward in the field of structural health monitoring. UAVs are capable of quickly capturing high-resolution images of hard-to-reach areas, offering comprehensive coverage of large and complex structures such as transmission towers. These images can then be analyzed by deep learning algorithms, enabling rapid and automated assessments across the entire asset. In contrast to traditional inspections, which require extensive preparation, such as scheduling personnel, gathering equipment, and ensuring site-specific safety measures, UAVs can be deployed with minimal setup, primarily involving flight path planning and sensor calibration. The imagery collected by the UAVs is processed either in real-time or post-flight using cloud-based or onboard systems, providing near-instantaneous feedback and offering immediate insights into the asset's condition. This allows for faster decision-making and more accurate assessments of structural health. With this new framework, the role of personnel evolves from traditional manual inspections to a more technical, supervisory, and analytical capacity. They now focus on overseeing the UAV and deep learning systems to ensure optimal performance, verifying data accuracy, generating detailed reports, and supporting informed decision-making.

However, some limitations of the presented procedure have emerged, opening up new areas of research that need to be explored. While the automation of estimating corrosion areas represents a crucial achievement, future research should focus on estimating corrosion in terms of volume reduction. This advancement could enable the estimation of cross-section reduction and provide insights into the impact on the bending capacity of structural elements. Another critical challenge lies in segmenting the edges of corroded areas. Unlike defects in other structures, such as cracks in reinforced concrete, the boundaries of corrosion are often diffuse and poorly defined. This issue is further exacerbated in some cases where the surface is coated with red paint. To address these challenges, future work will explore the integration of hyperspectral imaging alongside traditional RGB data. Hyperspectral imaging captures information across a wider range of wavelengths, providing a richer dataset that could improve neural networks' ability to differentiate corroded areas more effectively, even under challenging conditions. By leveraging this enhanced data, it is possible to achieve greater accuracy and reliability in corrosion detection, paving the way for further innovations in structural health monitoring. Deployment of UAV-based systems faces challenges, including regulatory restrictions on flight operations, weather dependency, and limitations in flight autonomy or payload capacity. However, the use of photogrammetric models mitigates several of these challenges by reducing prolonged flight durations and decreasing risks associated to manual piloting. Additionally, accurate and potentially automated flight planning can optimize data acquisition, ensuring that sufficient information is collected for post-processing analysis while reducing operational time and maintaining compliance with stringent regulatory requirements.

In conclusion, the integration of UAV-based photogrammetry and deep learning techniques represents a significant advancement in the inspection and maintenance of steel transmission towers. The successful application of the proposed neural network to real-world scenarios highlights its potential for wider adoption in structural health monitoring and maintenance programs. The evolution of personnel roles from inspectors to technical supervisors further underscores the transformative impact of this framework. This study lays the groundwork for a more efficient, accurate, and automated approach to maintaining the vital infrastructure that supports modern society.

ACKNOWLEDGMENTS

This work was supported in part by Rai Way S.p.A., Roma, Italy, which funded the research and provided data on the Eremo transmission tower. We would like to thank the technicians of Politecnico di Torino Geomatics Research Group for their support in collecting Eremo tower data.

CONFLICT OF INTEREST STATEMENT

The authors acknowledge that this research was partially funded by Rai Way S.p.A., Roma, Italy, and Fabio Graglia and Gabriele Scozza are employees of the company. The funding organization provided financial support for this research; however, it had no influence on the study design, data collection, and analysis, decision to publish, or preparation of the manuscript. All authors have fully disclosed these relationships and affirm that the research was conducted with transparency and integrity to minimize any potential bias.

REFERENCES

Aliyari, M., Droguett, E. L., & Ayele, Y. Z. (2021). UAV-based bridge inspection via transfer learning. *Sustainability*, *13*(20), 11359. https://doi.org/10.3390/su132011359

Barbosa, J. A. P. (2020). *Deep learning approach for UAV visual electrical assets inspection* (Publication No. 1131197) [Master's thesis, Instituto Superior de Engenharia do Porto]. ProQuest Dissertations & Theses Global. https://www.proquest.com/docview/2637685033/abstract/CD3225C07F0448E8PQ/1

Barfuss, S., Jensen, A., & Clemens, S. (2012). *Evaluation and development of unmanned aircraft (UAV) for UDOT needs*. Report No. UT-12.08. Utah State University and Utah Water Research Laboratory. https://rosap.ntl.bts.gov/view/dot/24691

Belcore, E., Di Pietra, V., Grasso, N., Piras, M., Tondolo, F., Savino, P., Polania, D. R., & Osello, A. (2022). Towards a FOSS automatic classification of defects for bridges structural health monitoring. In E. Borgogno-Mondino & P. Zamperlin (Eds.), *Geomatics and geospatial technologies* (pp. 298–312). Springer International Publishing. https://doi.org/10.1007/978-3-030-94426-1_22

Belcore, E., Piras, M., & Pezzoli, A. (2022). Land cover classification from very high-resolution UAS data for flood risk mapping. *Sensors*, *22*(15), 15. https://doi.org/10.3390/s22155622

Belcore, E., Piras, M., & Wozniak, E. (2020). Specific alpine environment land cover classification methodology: Google Earth Engine processing for Sentinel-2 data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLIII-B3-2020*, 663–670. https://doi.org/10.5194/isprs-archives-XLIII-B3-2020-663-2020

Brooks, C., Dobson, R. J., Banach, D. M., Dean, D., Oommen, T., Wolf, R. E., Havens, T. C., Ahlborn, T. M., & Hart, B. (2015). *Evaluating the use of unmanned aerial vehicles for transportation purposes: [Parts A-D]*. Tech Report No. RC-1616. Michigan Tech Research Institute & Michigan Technological University. https://rosap.ntl.bts.gov/view/dot/28859

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2016). *Semantic image segmentation with deep convolutional nets and fully connected CRFs*. arXiv. https://doi.org/10.48550/arXiv.1412.7062

Chen, L.-C., Papandreou, G., Schroff, F., & Adam, H. (2017). *Rethinking atrous convolution for semantic image segmentation*. arXiv. https://doi.org/10.48550/arXiv.1706.05587

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). *Encoder-decoder with atrous separable convolution for semantic image segmentation*. arXiv. https://doi.org/10.48550/arXiv.1802.02611

Chen, S., Laefer, D. F., Mangina, E., Zolanvari, S. M. I., & Byrne, J. (2019). UAV bridge inspection through evaluated 3D reconstructions. *Journal of Bridge Engineering*, *24*(4), 05019001. https://doi.org/10.1061/(ASCE)BE.1943-5592.0001343

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI (pp. 1800–1807). https://doi.org/10.1109/CVPR.2017.195

Cui, J., Zhang, Y., Ma, S., Yi, Y., Xin, J., & Liu, D. (2017). Path planning algorithms for power transmission line inspection using unmanned aerial vehicles. *2017 29th Chinese Control and Decision Conference (CCDC)*, Chongqing, China (pp. 2304–2309). https://doi.org/10.1109/CCDC.2017.7978899

Das, A., & Woolsey, C. A. (2019). Workspace modeling and path planning for truss structure inspection by unmanned aircraft. *Journal of Aerospace Information Systems*, *16*(1), 37–51. https://doi.org/10.2514/1.I010634

Diniz, L. F., Pinto, M. F., Melo, A. G., & Honório, L. M. (2022). Visual-based assistive method for UAV power line inspection and landing. *Journal of Intelligent & Robotic Systems*, *106*(2), 41. https://doi.org/10.1007/s10846-022-01725-x

Eisenbeiss, H. (2009). *UAV photogrammetry* [Doctoral dissertation, ETH Zurich]. https://doi.org/10.3929/ethz-a-005939264

Entezami, A., Behkamal, B., & De Michele, C. (2024). Pioneering remote sensing in structural health monitoring. In A. Entezami, B. Behkamal, & C. De Michele (Eds.), *Long-term structural health monitoring by remote sensing and advanced machine learning: A practical strategy via structural displacements from synthetic aperture radar images* (pp. 1–27). Springer. https://doi.org/10.1007/978-3-031-53995-4_1

Fei, Z., Yang, E., Yang, B., & Yu, L. (2021). Image enhancement and corrosion detection for UAV visual inspection of pressure vessels. In M. Fei, L. Chen, S. Ma, & X. Li (Eds.), *Intelligent life system modelling, image processing and analysis* (Vol. 1467, pp. 145–154). Springer. https://doi.org/10.1007/978-981-16-7207-1_15

Gillins, D. T., Parrish, C., Gillins, M. N., & Simpson, C. (2018). *Eyes in the sky: Bridge inspections with unmanned aerial vehicles*. Report No. FHWA-OR-RD-18-11. Oregon Department of Transportation. https://trid.trb.org/View/1502840

Hattori, K., Oki, K., Sugita, A., Sugiyama, T., & Chun, P. (2024). Deep learning-based corrosion inspection of long-span bridges with BIM integration. *Heliyon*, *10*(15), e35308. https://doi.org/10.1016/j.heliyon.2024.e35308

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV (pp. 770–778). https://doi.org/10.1109/CVPR.2016.90

Huang, Y., Liu, Y., Liu, F., & Liu, W. (2024). A lightweight feature attention fusion network for pavement crack segmentation. *Computer-Aided Civil and Infrastructure Engineering*, *39*(18), 2811–2825. https://doi.org/10.1111/mice.13225

Hugenholtz, C. H., Walker, J., Brown, O., & Myshak, S. (2015). Earthwork volumetrics with an unmanned aerial vehicle and softcopy photogrammetry. *Journal of Surveying Engineering*, *141*(1), 06014003. https://doi.org/10.1061/(ASCE)SU.1943-5428.0000138

Irizarry, J., Johnson, E. N., & Georgia Institute of Technology. (2014). *Feasibility study to determine the economic and operational benefits of utilizing unmanned aerial vehicles (UAVs)*. Report No. FHWA-GA-1H-12-38. Georgia Department of Transportation. https://rosap.ntl.bts.gov/view/dot/27333

Jang, K., Song, T., Kim, D., Kim, J., Koo, B., Nam, M., Kwak, K., Lee, J., & Chung, M. (2023). Analytical method for bridge damage using deep learning-based image analysis technology. *Applied Sciences*, *13*(21), 11800. https://doi.org/10.3390/app132111800

Jeon, M., Moon, J., Jeong, S., & Oh, K.-Y. (2024). Autonomous flight strategy of an unmanned aerial vehicle with multimodal information for autonomous inspection of overhead transmission facilities. *Computer-Aided Civil and Infrastructure Engineering*, *39*(14), 2159–2186. https://doi.org/10.1111/mice.13188

Jiang, J., Feng, X., Ye, Q., Hu, Z., Gu, Z., & Huang, H. (2023). Semantic segmentation of remote sensing images combined with attention mechanism and feature enhancement U-Net. *International Journal of Remote Sensing*, *44*(19), 6219–6232. https://doi.org/10.1080/01431161.2023.2264502

Jiang, T., Frøseth, G. T., Rønnquist, A., Kong, X., & Deng, L. (2024). A visual inspection and diagnosis system for bridge rivets based on a convolutional neural network. *Computer-Aided Civil and Infrastructure Engineering*, *39*, 3786–3804. https://doi.org/10.1111/mice.13274

Khaloo, A., Lattanzi, D., Jachimowicz, A., & Devaney, C. (2018). Utilizing UAV and 3D computer vision for visual inspection of a large gravity dam. *Frontiers in Built Environment*, *4*, 31. https://doi.org/10.3389/fbuil.2018.00031

Kraus, K. (2007). *Photogrammetry: Geometry from images and laser scans*. Walter de Gruyter.

Li, H., Chen, Y., Liu, J., Che, C., Meng, Z., & Zhu, H. (2024). High-resolution model reconstruction and bridge damage detection based on data fusion of unmanned aerial vehicles light detection and ranging data imagery. *Computer-Aided Civil and Infrastructure Engineering*, *39*(8), 1197–1217. https://doi.org/10.1111/mice.13133

Li, X., Ma, H., & Luo, X. (2019). Weaklier supervised semantic segmentation with only one image level annotation per category. *IEEE Transactions on Image Processing*, *29*, 128–141. https://doi.org/10.1109/TIP.2019.2930874

Lin, J. J., Han, K. K., & Golparvar-Fard, M. (2015). A Framework for model-driven acquisition and analytics of visual data using UAVs for automated construction progress monitoring. *2015 ASCE International Workshop on Computing in Civil Engineering, IWCCE 2015*, Austin, TX (pp. 156–164). https://doi.org/10.1061/9780784479247.020

Liu, Y.-F., Nie, X., Fan, J.-S., & Liu, X.-G. (2020). Image-based crack assessment of bridge piers using unmanned aerial vehicles and three-dimensional scene reconstruction. *Computer-Aided Civil and Infrastructure Engineering*, *35*(5), 511–529. https://doi.org/10.1111/mice.12501

Ma, D., Wang, N., Fang, H., Chen, W., Li, B., & Zhai, K. (2024). Attention-optimized 3D segmentation and reconstruction system for sewer pipelines employing multi-view images. Computer-Aided Civil and Infrastructure Engineering. Advance online publication. https://doi.org/10.1111/mice.13241

Mandirola, M., Casarotti, C., Peloso, S., Lanese, I., Brunesi, E., & Senaldi, I. (2022). Use of UAS for damage inspection and assessment of bridge infrastructures. *International Journal of Disaster Risk Reduction*, *72*, 102824. https://doi.org/10.1016/j.ijdrr.2022.102824

Marchewka, A., Ziółkowski, P., & Aguilar-Vidal, V. (2020). Framework for structural health monitoring of steel bridges by computer vision. *Sensors*, *20*(3), 3. https://doi.org/10.3390/s20030700

Montes, K., Liu, J., Dang, J., & Chun, P. (2023). Structure from segmented motion for bridge 3D damage detection using UAV, AI, and MR. *Intelligence, Informatics and Infrastructure*, *4*(2), 27–34.

Niethammer, U., Rothmund, S., James, M. R., Travelletti, J., & Joswig, M. (2010). *UAV-based remote sensing of landslides*. Commission V Symposium, Newcastle upon Tyne, UK.

Pezeshki, H., Adeli, H., Pavlou, D., & Siriwardane, S. C. (2023). State of the art in structural health monitoring of offshore and marine structures. *Proceedings of the Institution of Civil Engineers—Maritime Engineering*, *176*(2), 89–108. https://doi.org/10.1680/jmaen.2022.027

Phung, M. D., Hoang, V. T., Dinh, T. H., & Ha, Q. (2017). *Automatic crack detection in built infrastructure using unmanned aerial vehicles*. Proceedings of the 34rd ISARC, Taipei, Taiwan (pp. 823–829). https://doi.org/10.22260/ISARC2017/0115

Pinto, L., Bianchini, F., Nova, V., & Passoni, D. (2020). Low-cost UAS photogrammetry for road infrastructures' inspection. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIII-B2-2020*, 1145–1150. https://doi.org/10.5194/isprs-archives-XLIII-B2-2020-1145-2020

Reagan, D., Sabato, A., Niezrecki, C., Yu, T., & Wilson, R. (2016). An autonomous unmanned aerial vehicle sensing system for structural health monitoring of bridges. Proceedings of SPIE 9804, Nondestructive Characterization and Monitoring of Advanced Materials, Aerospace, and Civil Infrastructure 2016, Las Vegas, NV (pp. 244–252). https://doi.org/10.1117/12.2218370

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. *2018 IEEE/CVF Conference on Computer Vision and Pattern*

Recognition, Salt Lake City, UT (pp. 4510–4520). https://doi.org/10.1109/CVPR.2018.00474

Sankarasrinivasan, S., Balasubramanian, E., Karthik, K., Chandrasekar, U., & Gupta, R. (2015). Health monitoring of civil structures with integrated UAV and image processing system. *Procedia Computer Science*, *54*, 508–515. https://doi.org/10.1016/j.procs.2015.06.058

Savino, P., & Tondolo, F. (2021). Automated classification of civil structure defects based on convolutional neural network. *Frontiers of Structural and Civil Engineering*, *15*(2), 305–317. https://doi.org/10.1007/s11709-021-0725-9

Savino, P., & Tondolo, F. (2023). Civil infrastructure defect assessment using pixel-wise segmentation based on deep learning. *Journal of Civil Structural Health Monitoring*, *13*, 35–48. https://doi.org/10.1007/s13349-022-00618-9

Shim, S., Lee, S.-W., Cho, G.-C., Kim, J., & Kang, S.-M. (2023). Remote robotic system for 3D measurement of concrete damage in tunnel with ground vehicle and manipulator. *Computer-Aided Civil and Infrastructure Engineering*, *38*(15), 2180–2201. https://doi.org/10.1111/mice.12982

Shin, H., Kim, J., Kim, K., & Lee, S. (2023). Empirical case study on applying artificial intelligence and unmanned aerial vehicles for the efficient visual inspection of residential buildings. *Buildings*, *13*(11), 2754. https://doi.org/10.3390/buildings13112754

Simonyan, K., & Zisserman, A. (2015). *Very deep convolutional networks for large-scale image recognition*. arXiv. https://doi.org/10.48550/arXiv.1409.1556

Siriborvornratanakul, T. (2023). Pixel-level thin crack detection on road surface using convolutional neural network for severely imbalanced data. *Computer-Aided Civil and Infrastructure Engineering*, *38*(16), 2300–2316. https://doi.org/10.1111/mice.13010

Sun, W., Hou, S., Wu, G., Zhang, Y., & Zhao, L. (2024). Two-step rapid inspection of underwater concrete bridge structures combing sonar, camera, and deep learning. Computer-Aided Civil and Infrastructure Engineering. Advance online publication. https://doi.org/10.1111/mice.13401

Sun, L., Yang, Y., Zhou, G., Chen, A., Zhang, Y., Cai, W., & Li, L. (2024). An integration–competition network for bridge crack segmentation under complex scenes. *Computer-Aided Civil and Infrastructure Engineering*, *39*(4), 617–634. https://doi.org/10.1111/mice.13113

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2017). Inception-v4, Inception-ResNet and the impact of residual connections on learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, *31*(1), 1. https://doi.org/10.1609/aaai.v31i1.11231

Torrey, L., & Shavlik, J. (2010). Transfer learning. In E. S. Olivas, J. D. M. Guerrero, M. Martinez-Sober, J. R. Magdalena-Benedito, & A. J. Serrano López (Eds.), *Handbook of research on machine learning applications and trends* (pp. 242–264). IGI Global. https://doi.org/10.4018/978-1-60566-766-9.ch011

Truong-Hong, L., Chen, S., Cao, V. L., & Laefer, D. F. (2018). Automatic bridge deck damage using low cost UAV-based images. *TU1406 Quality Specifications for Roadway Bridges Standardization at a European Level*, Barcelona, Spain. http://archive.nyu.edu/handle/2451/43479

Wang, P., Liu, C., Wang, X., Tian, L., Miao, J., & Liu, Y. (2024). Multicategory fire damage detection of post-fire reinforced concrete structural components. *Computer-Aided Civil and Infrastructure Engineering*, *40*(1), 91–112. https://doi.org/10.1111/mice.13314

Wu, J., Shi, Y., Wang, H., Wen, Y., & Du, Y. (2023). Surface defect detection of Nanjing City wall based on UAV oblique photogrammetry and TLS. *Remote Sensing*, *15*(8), 2089. https://doi.org/10.3390/rs15082089

Yang, X., Li, H., Yu, Y., Luo, X., Huang, T., & Yang, X. (2018). Automatic pixel-level crack detection and measurement using fully convolutional network. *Computer-Aided Civil and Infrastructure Engineering*, *33*, 1090–1109. https://doi.org/10.1111/mice.12412

Yao, H., Liu, Y., Lv, H., Huyan, J., You, Z., & Hou, Y. (2024). Encoder–decoder with pyramid region attention for pixel-level pavement crack recognition. *Computer-Aided Civil and Infrastructure Engineering*, *39*(10), 1490–1506. https://doi.org/10.1111/mice.13128

Zhang, T., Wang, D., Mullins, A., & Lu, Y. (2023). Integrated APC-GAN and AttuNet framework for automated pavement crack pixel-level segmentation: A new solution to small training datasets. *IEEE Transactions on Intelligent Transportation Systems*, *24*(4), 4474–4481. https://doi.org/10.1109/TITS.2023.3236247

Zhou, Q., Ding, S., Feng, Y., Qing, G., & Hu, J. (2022). Corrosion inspection and evaluation of crane metal structure based on UAV vision. *Signal, Image and Video Processing*, *16*(6), 1701–1709. https://doi.org/10.1007/s11760-021-02126-7

Zhou, Z., Zhang, J., Gong, C., & Wu, W. (2023). Automatic tunnel lining crack detection via deep learning with generative adversarial networkbased data augmentation. *Underground Space*, *9*, 140–154. https://doi.org/10.1016/j.undsp.2022.07.003

Zhu, Y., Niu, X., & Tian, J. (2024). A machine vision-based intelligent segmentation method for dam underwater cracks swarm optimization algorithm and deep learning. Computer-Aided Civil and Infrastructure Engineering. Advance online publication. https://doi.org/10.1111/mice.13343

How to cite this article: Savino, P., Graglia, F., Scozza, G., & Di Pietra, V. (2025). Automated corrosion surface quantification in steel transmission towers using UAV photogrammetry and deep convolutional neural networks. *Computer-Aided Civil and Infrastructure Engineering*, 1–21. https://doi.org/10.1111/mice.13434