Deep learning-based framework for vegetation hazard monitoring near powerlines

Nana Ekow Nkwa Sey (nkwa.sey@uenr.edu.gh)
University of Energy & Natural Resources
Mark Amo-Boateng
University of Energy & Natural Resources
Martin Kyereh Domfeh
University of Energy & Natural Resources
Amos T. Kabo-Bah
University of Energy & Natural Resources
Prince Antwi-Agyei
University of Energy & Natural Resources

Research Article

Keywords: Pix2Pix, vegetation encroachment, powerlines, drones, YoLov5, Image-to-Image translation

Posted Date: August 25th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1991473/v1

Abstract

The increasing popularity in the use of drones has also led to their adoption by electric utility companies to monitor intrusive vegetation near powerlines due to their ability to provide reliable and cost-effective inspections, minimising downtime and improving the efficiency of the monitoring operations of such companies. Besides the lines themselves, the monitoring also involves surrounding objects, most specifically vegetation. Despite the importance of trees and shrubs in maintaining a healthy environment, the growth of vegetation around power transmission lines poses a threat to the public and utility infrastructure itself. The study proposes a deep learning-based detection framework compatible with UAVs for monitoring vegetation encroachment near powerlines which estimates vegetation health and detects powerlines. The framework leverages on computing capability of NVIDIA Jetson Nano to integrate the Pix2Pix model for estimation of vegetation indices and YoLov5 for detection of powerlines from RGB images captured from drones. YoLov5 obtained good performance for detecting powerlines in aerial images with precision, recall, mAP @0.5, and mAP@0.5:0.95 values are 0.821, 0.762, 0.798 and 0.563 respectively. The Pix2Pix model generated satisfactory synthetic image translations from RGB to LUT. The proposed vegetation detection framework was able to detect locations of powerlines and generate NDVI estimates represented as LUT maps directly from RGB images captured from aerial images which could serve as a preliminary and affordable alternative to relatively expensive multispectral sensors which are not readily available in developing countries for monitoring and managing the presence and health of trees and dense vegetation within powerline corridors.

1.0 Introduction

The term unmanned aerial vehicle (UAV) has evolved into a variety of names and acronyms. Some of these include quadcopters, aerial robots or commonly drones (Colomina & Molina, 2014). Drone technology offers immense benefits in the field of environmental monitoring and conservation. The ever-increasing number of researchers, students and environmental engineers realising this, are turning to drones in place of traditional ground surveying equipment and costly manned aeroplane services to augment data from lower-resolution satellite imagery. Although high-resolution satellite images are now available, airborne sensors are more capable of providing frequent, flexible and high-resolution overflights for images and data collection making them ideal for several applications (Klemas, 2015). The potential of UAVs has emerged strongly in the field of environmental monitoring, with broad use in several applications such as precision agriculture (Hassler & Baysal-Gurel, 2019; Pérez-Ortiz et al., 2015), water quality assessment (Kageyama et al., 2016), invasive species detection (Gonzalez et al., 2016), endangered land and ocean species monitoring (Gonzalez et al., 2016; Hodgson et al., 2013; Klemas, 2015).

The increasing popularity of drones has also led to their adoption by electric utility companies to monitor intrusive vegetation near powerlines due to their ability to provide reliable and cost-effective inspections, minimizing downtime and improving the efficiency of the monitoring operations of such companies (Sun et al., 2006). Besides the lines themselves, the monitoring also involves surrounding objects, most specifically vegetation. This is done to ensure that the electrical transmission grids are operating safely. Monitoring the vegetation along the corridors of powerlines is important to prevent trees that are close by from damaging the equipment causing short circuits (Ahmad et al., 2013) and blackouts (Ahmad et al., 2015). Also, in dry conditions, the presence of too much-scorched vegetation may result in forest fires destroying powerlines (Mills et al., 2010). During storms, trees can fall across power lines and cause outages. These can also lead to severe damage to the equipment. Foot patrols are among the traditional methods used by utility companies to manage the vegetation around powerlines. One of the main disadvantages of foot patrols is that they are very infrequent, slow to deploy, offer low accuracy and are expensive (X. Li et al., 2021). An evaluation of an unmanned aerial vehicle revealed that it could be faster than traditional foot patrols to inspect and generate models of high-voltage power lines (Ahmad et al., 2015; Guangjian Yan et al., 2007; Luque-Vega et al., 2014). In addition to reducing the overall

cost of maintaining the power lines, the proper management of the vegetation can also help in maintaining the electricity supply.

Various studies have been presented about the importance of monitoring power lines. Some of these include those by (Aggarwal et al., 2000; Ahmad et al., 2013, 2015; Katrasnik et al., 2010). Several studies related to the usage of drones for vegetation encroachment near powerlines have focused on the segmentation, classification and extraction of trees (Matikainen et al., 2016). Classification of tree species can be beneficial as it allows for the identification of undesirable species which are most likely to cause problems in the future while encouraging the planting of low-growing trees as healthy competition (Z. Li et al., 2012). Multicopter platforms equipped with various miniaturized sensors, usually employ remote sensing techniques to gather data for vegetation health determination, detection and classification. Multispectral (MSI) sensors are preferred to colour or Red-Green-Blue (RGB) imagers for certain applications because they provide more spectral information which is not visible to the human eye. Studies from (Z. Li et al., 2009) and (Mills et al., 2010) used multispectral images to automatically extract and identify trees near powerlines. Both studies based their methodology on the ratio of red band to NIR reflectance as inputs into a pulse-couple neural network. The results of the study revealed that the system was able to improve its segmentation output by performing various morphological processing. Their detection rates were higher than the segmentation accuracy of tree crowns. The paper by (Eng L et al., 2018) aimed to verify and identify the algorithm that is most suitable for the processing of vegetation images. The study employed a consumer camera attached to a drone to collect data from various areas near the powerline and analysed them using the Visible Atmospherically Resistant Index (VARI) achieving good detection results for vegetative and non-vegetative areas.

Although colour models can provide the necessary information, they may not be able to identify all the vegetation hazards that can be found within a power line track hence the need for multi- or hyperspectral sensors which provide more spectral information which is not visible to the human eye (Di Gennaro et al., 2022). However, high-end remote sensing equipment such as multispectral sensors, thermal sensors, and hyperspectral cameras are costly and not readily available to the general public (Deng et al., 2018). Deep learning models have exhibited promising results in performing visual applications like object detection, image classification and segmentation (Haq et al., 2021) and reconstruction of multispectral images from RGB images (Arad & Ben-Shahar, 2016; Fu et al., 2019; Rangnekar et al., 2017; Yan et al., 2018). Recently studies by (Arad & Ben-Shahar, 2016; Can & Timofte, 2018; Fu et al., 2019; Huang et al., 2016; Zeng et al., 2021) proposed deep learning-based models to reconstruct multispectral images from coloured images. The availability of low-cost sensors and electronic devices has the potential to stimulate new applications and discoveries in the field of participatory research to potentially provide frequent and effective monitoring of the environment (Chan et al., 2021). Hence an intelligent AI-based framework that leverages the existing knowledge in deep learning and miniaturized high-performance computers has become necessary to promote frequent and sustainable integration of remote sensing solutions in continuous monitoring and management of vegetation encroaching on powerlines. The study, therefore, proposes a deep learning-based detection framework compatible with UAVs for monitoring vegetation encroachment near powerlines. The approach also estimates vegetation health and detects powerlines. The framework leverages on computing capability of NVIDIA's Jetson Nano and/or Graphics Processing Unit (GPU)-enabled machines to integrate generative adversarial networks (Pix2Pix) for estimation of vegetative indices and YoLov5 for detection of powerlines from RGB images captured with drones.

2.0 Materials And Methods

2.1 Data Collection and Processing

2.1.1 Drone Surveys

The flight-planning stage is the preliminary phase and most critical step to ensure safe flights and minimise post-processing times and costs. The unique vegetation characteristics of captured vegetation imagery during the mapping process make it strenuous to match these images. To minimize this issue the drone was programmed to capture images with very high overlap values of at least 75% along flight directions in both the longitudinal and transversal overlaps. Meteorological conditions are also considered during the surveys. All flights were performed under favourable weather conditions during the central hours of the day when the sun was close to its zenith to minimise shadows and bidirectional reflectance distribution function (BRDF) effects. As part of the pre-flight preparations, multispectral images of a reflectance calibration target were captured to perform radiometric calibration. A calibration target is used to convert the values expressed as Digital Numbers (DN) into real reflectance values recommended in vegetation analysis to allow for the comparison of images acquired at different epochs.

2.1.2 Data

The data collection process commenced through aerial image acquisition using a drone and a commercial-grade multispectral camera over several random areas within the Sunyani Municipality of the Bono Region of Ghana. Data samples available from the manufacturers of the multispectral camera used were also included in the study to augment data captured locally. A DJI Mavic2 drone equipped with a multispectral camera was used. Images were captured without ground control points (GCPs); however, all images were geotagged in EXIF format with embedded information on latitude, longitude and altitude. The multispectral images were captured using a 12 Megapixels MAPIR Survey3W OCN camera (MAPIR, n.d.). The images were saved in JPG format with the RGB and multispectral images having resolutions of 5472 × 3078 pixels & 4000 × 3000 pixels respectively. A single UAV flight took between 5–10 mins to complete.

2.1.2.1 Data Pre-processing

The quality of the RGB images collected was estimated using the image quality tool provided by the Agisoft Metashape software (Agisoft Metashape, 2022) based on each image's sharpness. For 3D reconstruction purposes, only images with a quality score above 0.5 were chosen for potential use (Cunliffe et al., 2016). The images captured with the multispectral camera were calibrated using the MAPIR Camera Control application. The next step involved using photogrammetry techniques to create an orthomosaic. The various stages and settings used for the RGB and MSI orthomosaic generation are shown in Table 1.

Table 1
Settings used for RGB and MSI orthomosaic generation

Stages	Description	
	RGB Workflow	**MSI Workflow**
Add Photos	Geotagged colour images added to chunk in jpg format	Geotagged colour images added to chunk in jpg format
"Camera Calibration"	-	Camera Type: Frame Pixel size (mm): 0.00155 x 0.00155 Focal Length (mm): 3.37 Fixed parameters: All
Align Photos	Accuracy: High Generic preselection: Yes Reference preselection: source Key points limit: 150,000 Tie point limit: 70, 000 Apply mask to: None Exclude stationary tie points: No Guided image matching: No Adaptive camera model fitting: Yes	Accuracy: High Generic preselection: No Reference preselection: source Key points limit: 150,000 Tie point limit: 70, 000
Optimize Alignment	Adaptive camera model fitting: Yes Estimate the point covariance: Yes	Estimate the point covariance: Yes
Build Dense Cloud	Quality: Medium Depth filtering: Mild Calculate point colours: Yes Calculate point confidence: Yes	Quality: Medium Depth filtering: Mild Calculate point colours: Yes Calculate point confidence: Yes
Build Mesh	Source data: Sparse cloud Surface Type: Arbitrary Depth maps quality: Medium Face count: Medium Interpolation: Enabled (default) Calculate vertex colours: Yes	Surface Type: Arbitrary (3D) Source data: Sparse cloud Face count: Medium Interpolation: Enabled (default) Calculate vertex colours: Yes

Stages	Description	
	RGB Workflow	MSI Workflow
Build Texture	Texture type: Diffuse map	Texture type: Diffuse map
	Source data: Images	Source data: Images
	Mapping mode: Adaptive Orthophoto	Mapping mode: Adaptive Orthophoto
	Blending mode: Mosaic (default)	Blending mode: Mosaic (default)
	Texture size/count: 4096×1	Texture size/count: 4096×1
	Enable hole filing: No	Enable hole filing: No
	Enable ghost filter: No	Enable ghost filter: No
Build DEM	Projection: Geographic: WGS 84 (EPSG:4326)	Projection: Geographic: WGS 84 (EPSG:4326)
	Source data: Dense cloud	Source data: Dense cloud
	Interpolation: enabled (default)	Interpolation: enabled (default)
Build Orthomosaic	Projection: Geographic: WGS 84 (EPSG:4326)	Projection: Geographic: WGS 84 (EPSG:4326)
	Surface: DEM	Surface: DEM
	Blending mode: Mosaic (default)	Blending mode: Mosaic (default)
	Hole filing: Yes	
Export Orthomosaic	Coordinate System: WGS 84 (EPSG:4326)	Coordinate System: WGS 84 (EPSG:4326)
	Save as type: jpg	Save as type: jpg

The Vegetative Index used was the Normalized Difference Vegetation Index (NDVI) which is a simple measure of plants' photosynthetically active biomass and can be used to determine, visualize and define vegetated areas on the map allowing users to monitor the growth process and identify areas of concern. The NDVI is expressed as shown in Eq. 1.

$$NDVI = \frac{NIR - Red}{NIR + Red} = \frac{Y - X}{Y + X} \quad \text{Eqn. 1}$$

where Y is the NIR light @Band 1 (Red Channel) and X is the orange light @Band 3 (Blue Channel). The images taken with the Survey3 camera were then processed to produce an index image and a coloured Look-Up-Table (LUT) is applied to show the contrast between the different vegetation. To normalise all the generated LUT images the min and max values were set to -1 and 1 respectively. To generate the training, validation and test datasets for the GAN models, the RGB and LUT images (see example in Fig. 1) were aligned and cropped into equal squares with a resolution of 256 × 256 pixels and 1024 × 1024 pixels. The LUT - RGB image pairs were then split into train, validation and test using a ratio of 0.6:0.3:0.1. The dataset contains a total of 3,656 image pairs.

During the drone surveys, several images of powerlines were taken with the drone camera pointing 90° down at different heights. An open-source image annotation tool, LabelMe (Wada, 2016) was used to annotate the region of powerlines in each picture in COCO format. Figure 2 shows the interface of the annotation tool.

A set of training and validation datasets was then generated for the YOLOv5 model. Several augmentation steps (i.e., flipping, rotations, blurring, exposure and noise) were applied to the annotated datasets to increase the training sample

size. Out of 2,859 annotated images, 2,500 (88%), 239 (8%) and 120 (4%) were used as train, validation and test datasets respectively. The images were also resized to a uniform size of 415 × 415.

3.0 Model Architecture And Evaluation

3.1 YOLO Model

The YOLOv5 algorithm is based on the detection architecture of the Yolo model and uses several optimization strategies widely used in CNN. The YoLov5 algorithm features four main parts: the input terminal, the backbone, the output, and the neck (See network structure in Fig. 3).

The input terminal is mainly used to pre-process the data. The YoLov5 algorithm can adapt to different datasets by calculating the initial anchor frame size automatically. This feature can be useful for performing various tasks such as auto-learning and data augmentation. The backbone network mainly uses a combination of spatial pyramid pooling (SPP) and cross-stage partial network (CSPNet) to extract feature maps from an image's input. The advantage of this strategy is that it allows the network to perform various tasks at the same time while reducing the amount of computation. The bottleneck CSP is also useful to increase the speed of the detection by reducing the amount of computation. The spatial pyramid pooling structure is additionally used to generate three-scale feature maps. The neck network utilizes the feature pyramid structures of PAN and FPN to generate feature maps from an image's input. The FPN (Liu et al., 2015) structure provides strong semantic features to the top feature maps while the PAN (W. Wang et al., 2019) structure delivers strong localization features to the lower feature maps. The two feature pyramid structures of the neck network contribute to the improved performance of the detection by providing a strong representation of the various network layers in backbone fusion. As the final step in the process, the head output of the network is used to predict the different sizes of the feature maps. The YoLov5 is composed of four different architecture groups, namely YoLov5s, YoLov5m, YoLov5l, and YoLov5x with the main difference being the number of convolution kernel and feature extraction modules at specific locations within the network.

3.1.1 Evaluation of model

The Yolov5 model's performance is evaluated in two phases. The first phase focuses on the detection performance while the second phase is on object classification. The metric used to evaluate the system is the mean Average Precision (mAP) which measures the accuracy of an object detection algorithm. It is calculated by taking into account the various IoU, recall, and AP parameters. A higher mAP score indicates better model performance. The average and mean precision are the most popular metrics used to evaluate the various algorithms in the field of object detection. They were additionally used to evaluate submissions in various competitions such as the PASCAL VOC and the COCO challenges (Nguyen et al., 2020). Additionally, the model was also evaluated to determine the powerlines' location using the Intersection of Union (IoU) metrics which determines the correctness of the model to predict bounding boxes of categories of objects correctly. A detection is mostly termed as successful if there is an overlap of more than 60% (Nguyen et al., 2020). The model’s false positives (FN), false negatives (FN) and true positive (TP) were analysed specifying a confidence score (conf_sc) according to Table 2 to compute the precision and recall values as shown in Equations 2 and 3.

Table 2
Definitions of terms

Definition	*Ground truth* compared to *predicted*	Confusion score, *conf_sc*
TP (correct detection)	IoU > 0.6 (powerline present)	*conf_sc* ≥ *threshold (*powerline detected)
FP (invalid detection)	IoU ≤ 0.6 (no powerline present)	
FN (Missed *ground truth*)	Missed powerline, but powerline present	*conf_sc* < *threshold*

$$Precision = \frac{TP}{(TP+FP)} = \frac{TP}{AllDetections} \quad \text{Eqn. 2}$$

$$Recall = \frac{TP}{(TP+FN)} = \frac{TP}{AllGroundTruths} \quad \text{Eqn. 3}$$

3.2 Pix2Pix Model

The GAN algorithm, which is based on a generative modelling approach, utilizes a CNN proposed by (Goodfellow et al., 2014). Generative modelling is a type of unsupervised learning task that involves learning to identify and extract patterns and regularities from vast sets of data to generate new examples that are plausible (Goodfellow et al., 2014). GANs are generative models used to map the output images from a random noise vector (z) into a representation of the image while conditional GANs generate new outputs (y) from both random noise (z) and an observed image (x). The goal is to train a generator model (G) to produce outputs that closely resemble real images and then pass these “fakes” through an adversarial discriminator model (D) which learns to detect the fakes produced by the generator. Figure 4 shows the general training process of GANs. The GAN model known as Pix2Pix is designed for general-purpose image-to-image translation which was presented by (Isola et al., 2016).

The model is built on the conditional GAN framework, which allows it to perform various image-to-image translation tasks. It utilizes a discriminator known as PatchGAN (Isola et al., 2016) and a generator known as U-Net (Ronneberger et al., 2015). The U-Net generates a low-level representation of the input image, which is then passed to the PatchGAN. The resulting statistics are then analysed to learn various feature representations. The loss function of the model is:

$$\min_{G}\max_{D} V(D, G) = E_{x,y}[loglogD(y)] + E_{x,z}[loglog(1 - D(G(x)))] + E_{x,z}\left[\|loglog(y - D(G(z)))\|_1\right] \quad \text{Eqn. 4}$$

This GAN model learns to map the noise (z) and input image \left(\mathcal{x}\right) to an output imagery. The discriminator model then tries to maximize its loss function while the generator tries to do the opposite. The study adopted the GAN Compression method proposed by (M. Li et al., 2020), which is a general-purpose method that can reduce the computational cost and time associated with the development of conditional GANs and pix2pixHD architecture proposed by (T. C. Wang et al., 2018) which presents an improved pix2pix framework that uses a coarse-to-fine generator, a multi-scale discriminator architecture, and a robust adversarial learning objective function.

3.2.1 Evaluation of model

Generative adversarial networks are commonly used to generate high-quality synthetic images. They have been shown to perform remarkably well in various problem domains. The generator models are trained by the discriminator model, which learns to distinguish between fake and real images. This method eliminates the possibility of having an objective measure or function for the generator model (Salimans et al., 2016). Although various measures are being introduced to improve the performance of the generator models, there is currently no consensus as to which measure should be used

to compare the strengths and limitations of the models (Borji, 2018). Developing GAN models can be very complex. Manual inspection of output figures can greatly minimize the lengthy process of testing, refining and implementing model configurations. The evaluation of GAN generator models is usually performed in the context of a target problem domain. The quality of the images that the models generate is taken into account to determine the performance of the model. The generator model was saved iteratively over many epochs during training. Each saved model was used to generate some synthetic images to allow for post-hoc evaluation to select an acceptable model for use. Quantitatively, the Fréchet Inception Distance (FID) score (Heusel et al., 2017) was used to evaluate the GAN model by summarizing the quality of images generated by GAN models. The FID score takes into account the distance between the feature vectors that are calculated for real images and those that are generated. It shows how similar the two groups are when it comes to the statistics on the computer vision features of raw images. Lower scores indicate that the two groups of images have more similar statistics, while a perfect score of 0.0 indicates that the two groups are the same.

4.0 Results And Discussions

Powerline detection model

The evaluation of the model was done on 239 test images which were alien to the trained model with an IoU threshold of 0.6. The model obtained good performance for detecting powerlines in aerial images. The precision (P), recall (R), mAP @0.5, mAP@0.5:0.95 values are 0.821, 0.762, 0.798 and 0.563 respectively. Figure 5 shows the graph of each metric considered per training step.

As shown in Figure 6, the various loss function values are seen to approach 0 with increasing training steps. Since the data contains one class (powerline), the model is not prone to misidentifications. The classification error (cls_loss) is therefore constantly zero. The bounding box regression losses (box_loss) and the objectness losses (obj_loss) for both training and validation decrease steadily and show a minimal gap at the final value indicating a good fit learning curve.

Further metrics describing the performance of the model are presented in Figure 7.

The true positive (TP), true negative (TN), false positive (FP) and false negative (FN) are the components that make up the confusion matrix (Figure 7(a)) defined on the IoU threshold of 0.6 The value of the f1 curve (Figure 7 (b)) is used to measure the balance between the accuracy and recall of the given object. High values of the metric imply that both the recall and precision are high. A lower f1 score means that the model's accuracy and recall are lower.

The precision-recall curve (Figure 8) shows the relationship between the number of positive samples and the accuracy of the model. The precision values decrease with increasing recall as the number of samples increases, and the model's accuracy in classifying correctly each sample decreases. This is expected since the model is more prone to failure when there are many samples. A mosaic of the ground truth against predicted results on the test samples is shown in Figure 9.

RGB to LUT image translation model

Some visual results of the original and compressed GAN are shown in Figure 10 which presents the input RGB image, the ground truth LUT image, the generated LUT of the full GAN and the generated LUT image of the GAN compression model.

Some visual results of the pix2pixHD model are shown in Figure 11 which presents the input RGB image, the ground truth and the generated LUT.

Quantitatively, the FID scores (the lower the better) recorded were 61.488 and 208.562 for the original and compressed GAN respectively. Also, it can be observed that the full GAN model generated satisfactory-looking LUTs compared to the compressed version which shows some persistent artefacts bordering the output on the top and left sides. Table 3 shows the results of the experiment compared to other datasets on the same model.

Table 3: Qualitative evaluation of model performance

Model	Dataset	Method	#Parameters	MACs	FID Score
Pix2Pix	RGB → LUT	original	11.3 M	56.8G	61.488
		compressed	0.75 M	4.68G	208.562
	edges → shoes	original	11.3 M	56.8G	24.18
		compressed	0.70 M	4.81G	26.60
	cityscapes	original	11.3 M	56.8G	-
		compressed	0.71 M	5.66G	-
	Map → aerial photo	original	11.3 M	56.8G	47.76
		compressed	0.75 M	4.68G	48.02

The comparatively higher FID score could be because of some irregularities and misalignments in the generation of the RGB – LUT image pairs borne out of the two different lenses of the RGB and MSI cameras and also from the image manipulation software used.

Outline of the proposed framework

The proposed deep learning-based detection framework compatible with UAVs for monitoring vegetation encroachment near the powerline is illustrated in Figure 12.

The hardware of the sensor platform consists of (1) NVIDIA Jetson Nano (2) UPS Power Module and (3) GNSS Module. The components are shown in Fig 13 (a).

The Jetson Nano was announced by NVIDIA in 2019 as a development system kit. This small but powerful computer is capable of delivering 475 GFLOPs for running modern AI algorithms in multiple neural networks in parallel for image classification, object detection and AI applications and processes using as little as 5 watts. The UPS power module provides a 5V uninterruptible power supply to the sensor. The 4G/GNSS Module provides GPS positioning information which is tagged to images captured. As shown in Figure 14, the sensor platform can be either used in three configurations; Standalone, where batches of images captured with a drone can be processed post flights to assess vegetation encroachment automatically using (1) the sensor platform or (2) a GPU-enabled workstation and (3) In flight, where the sensor platform is attached to a drone as vegetation hazards are monitored in real time via a streaming platform. Images that contain potential vegetation hazards detected by the platform are tagged for further visual

inspection and assessment. Figure 14 shows the flowchart of the proposed hazard detection framework which involves three steps; (1) detection of powerlines, (2) Conversion of RGB images to VI maps and (3) vegetation hazard determination.

4.0 Conclusions

The detection framework for vegetation encroachment near powerlines was developed using a couple of deep learning methods. The study employed the use of a Pix2Pix GAN for image-to-image translations from RGB to LUT maps representing NDVI and YoLov5 model for detecting powerlines from images captured from a UAV. The study revealed that:

1. the use of a CNN model, i.e., YoLov5 was able to successfully detect the presence of powerlines of various sizes from aerial images. The model obtained good performance for detecting powerlines in aerial images with precision (P), recall (R), mAP @0.5, and mAP@0.5:0.95 values are 0.821, 0.762, 0.798 and 0.563 respectively.

2. the original pix2pix GAN model generated satisfactory synthetic image translations from RGB to LUT compared to the compressed version which visibly showed several artefacts in the generated LUT images and was not close to the ground truth. Quantitatively, the FID scores (the lower the better) recorded were 61.488 and 208.562 for the original and compressed GAN respectively corroborating the visual results. The pix2pixHD model also generated very good synthetic LUT images.

3. the deep learning models, i.e., YOLOv5 and Pix2Pix GAN were capable of running on the Jetson Nano. The compressed version of the Pix2Pix had a much better computing performance but unsatisfactory-looking synthetic images.

4. the proposed vegetation detection framework was able to generate NDVI estimates represented as LUT maps directly from RGB images captured from aerial images which could serve as a preliminary and affordable alternative to relatively expensive multispectral sensors which are not readily available in developing countries for monitoring and managing the presence and health of trees and dense vegetation within powerline corridors.

Declarations

Acknowledgement

The authors are grateful for the immense assistance received from the Earth Observation Research and Innovation Centre (EORIC), University of Energy and Natural Resources.

Funding body

This research was funded by the Regional Centre for Energy & Environmental Sustainability (RCEES), University of Energy and Natural Resources (UENR).

Conflicts of interest

The authors declare no conflict of interest.

References

Aggarwal, R. K., Johns, A. T., Jayasinghe, J. A. S. B., & Su, W. (2000). An overview of the condition monitoring of overhead lines. *Fuel and Energy Abstracts*, *41*(3), 157. https://doi.org/10.1016/S0140-6701(00)93130-8

Agisoft Metashape. (2022). *Agisoft Metashape Professional (Version 1.7.1) (Software)*. https://www.agisoft.com/downloads/installer/

Ahmad, J., Malik, A. S., Abdullah, M. F., Kamel, N., & Xia, L. (2015). A novel method for vegetation encroachment monitoring of transmission lines using a single 2D camera. *Pattern Analysis and Applications*, *18*(2), 419–440. https://doi.org/10.1007/s10044-014-0391-9

Ahmad, J., Malik, A. S., Xia, L., & Ashikin, N. (2013). Vegetation encroachment monitoring for transmission lines right-of-ways: A survey. *Electric Power Systems Research*, *95*, 339–352. https://doi.org/10.1016/j.epsr.2012.07.015

Arad, B., & Ben-Shahar, O. (2016). Sparse Recovery of Hyperspectral Signal from Natural RGB Images. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 9911 LNCS* (pp. 19–34). Springer, Cham. https://doi.org/10.1007/978-3-319-46478-7_2

Borji, A. (2018). Pros and Cons of GAN Evaluation Measures. *Computer Vision and Image Understanding*, *179*, 41–65. https://doi.org/10.48550/arxiv.1802.03446

Can, Y. B., & Timofte, R. (2018). *An efficient CNN for spectral reconstruction from RGB images*. https://arxiv.org/abs/1804.04647v1

Chan, K., Schillereff, D. N., Baas, A. C. W., Chadwick, M. A., Main, B., Mulligan, M., O'Shea, F. T., Pearce, R., Smith, T. E. L., van Soesbergen, A., Tebbs, E., & Thompson, J. (2021). Low-cost electronic sensors for environmental research: Pitfalls and opportunities. *Progress in Physical Geography: Earth and Environment*, *45*(3), 305–338. https://doi.org/10.1177/0309133320956567

Colomina, I., & Molina, P. (2014). Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *92*, 79–97. https://doi.org/10.1016/j.isprsjprs.2014.02.013

Cunliffe, A. M., Brazier, R. E., & Anderson, K. (2016). Ultra-fine grain landscape-scale quantification of dryland vegetation structure with drone-acquired structure-from-motion photogrammetry. *Remote Sensing of Environment*, *183*, 129–143. https://doi.org/10.1016/j.rse.2016.05.019

Deng, L., Mao, Z., Li, X., Hu, Z., Duan, F., & Yan, Y. (2018). UAV-based multispectral remote sensing for precision agriculture: A comparison between different cameras. *ISPRS Journal of Photogrammetry and Remote Sensing*, *146*, 124–136. https://doi.org/10.1016/j.isprsjprs.2018.09.008

Di Gennaro, S. F., Toscano, P., Gatti, M., Poni, S., Berton, A., & Matese, A. (2022). Spectral Comparison of UAV-Based Hyper and Multispectral Cameras for Precision Viticulture. *Remote Sensing*, *14*(3), 449. https://doi.org/10.3390/rs14030449

Eng L, S., Ismail, R., Hashim, W., Mohamed, R. R., & Baharum, A. (2018). Vegetation Monitoring Using UAV : a Preliminary Study. *International Journal of Engineering & Technology*, *7*(4.35), 223. https://doi.org/10.14419/ijet.v7i4.35.22736

Fu, Y., Zhang, T., Zheng, Y., Zhang, D., & Huang, H. (2019). Hyperspectral Image Super-Resolution With Optimized RGB Guidance. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, *2019-June*, 11653–11662. https://doi.org/10.1109/CVPR.2019.01193

Gonzalez, L., Montes, G., Puig, E., Johnson, S., Mengersen, K., & Gaston, K. (2016). Unmanned Aerial Vehicles (UAVs) and Artificial Intelligence Revolutionizing Wildlife Monitoring and Conservation. *Sensors*, *16*(1), 97. https://doi.org/10.3390/s16010097

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Networks. *Communications of the ACM*, *63*(11), 139–144. https://doi.org/10.1145/3422622

Guangjian Yan, Junfa Wang, Qiang Liu, Lin Su, Pengxin Wang, Junming Liu, Wuming Zhang, & Zhiqiang Xiao. (2007). An airborne multi-angle power line inspection system. *2007 IEEE International Geoscience and Remote Sensing Symposium*, 2913–2915. https://doi.org/10.1109/IGARSS.2007.4423453

Haq, M. A., Rahaman, G., Baral, P., & Ghosh, A. (2021). Deep Learning Based Supervised Image Classification Using UAV Images for Forest Areas Classification. *Journal of the Indian Society of Remote Sensing*, *49*(3), 601–606. https://doi.org/10.1007/s12524-020-01231-3

Hassler, S. C., & Baysal-Gurel, F. (2019). Unmanned Aircraft System (UAS) Technology and Applications in Agriculture. *Agronomy*, *9*(10), 618. https://doi.org/10.3390/agronomy9100618

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2017). GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. *Advances in Neural Information Processing Systems*, *2017-Decem*, 6627–6638. http://arxiv.org/abs/1706.08500

Hodgson, A., Kelly, N., & Peel, D. (2013). Unmanned Aerial Vehicles (UAVs) for Surveying Marine Fauna: A Dugong Case Study. *PLoS ONE*, *8*(11), e79556. https://doi.org/10.1371/journal.pone.0079556

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2016). Densely Connected Convolutional Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, *2017-Janua*, 2261–2269. https://doi.org/10.1109/CVPR.2017.243

Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2016). Image-to-Image Translation with Conditional Adversarial Networks. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, *2017-Janua*, 5967–5976. https://doi.org/10.1109/CVPR.2017.632

Kageyama, Y., Takahashi, J., Nishida, M., Kobori, B., & Nagamoto, D. (2016). Analysis of water quality in Miharu dam reservoir, Japan, using UAV data. *IEEJ Transactions on Electrical and Electronic Engineering*, *11*, S183–S185. https://doi.org/10.1002/tee.22253

Katrasnik, J., Pernus, F., & Likar, B. (2010). A Survey of Mobile Robots for Distribution Power Line Inspection. *IEEE Transactions on Power Delivery*, *25*(1), 485–493. https://doi.org/10.1109/TPWRD.2009.2035427

Klemas, V. V. (2015). Coastal and Environmental Remote Sensing from Unmanned Aerial Vehicles: An Overview. *Journal of Coastal Research*, *315*, 1260–1267. https://doi.org/10.2112/JCOASTRES-D-15-00005.1

Li, M., Lin, J., Ding, Y., Liu, Z., Zhu, J.-Y., & Han, S. (2020). GAN Compression: Efficient Architectures for Interactive Conditional GANs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PP*, 1–1. https://doi.org/10.1109/TPAMI.2021.3126742

Li, X., Li, Z., Wang, H., & Li, W. (2021). Unmanned Aerial Vehicle for Transmission Line Inspection: Status, Standardization, and Perspectives. *Frontiers in Energy Research*, *9*, 336. https://doi.org/10.3389/fenrg.2021.713634

Li, Z., Bruggemann, T. S., Ford, J. J., Mejias, L., & Liu, Y. (2012). Toward automated power line corridor monitoring using advanced aircraft control and multisource feature fusion. *Journal of Field Robotics*, *29*(1), 4–24. https://doi.org/10.1002/rob.20424

Li, Z., Hayward, R., Zhang, J., Liu, Y., & Walker, R. (2009). Towards automatic tree crown detection and delineation in spectral feature space using PCNN and morphological reconstruction. *2009 16th IEEE International Conference on Image Processing (ICIP)*, 1705–1708. https://doi.org/10.1109/ICIP.2009.5413642

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2015). SSD: Single Shot MultiBox Detector. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9905 LNCS*, 21–37. https://doi.org/10.1007/978-3-319-46448-0_2

Luque-Vega, L. F., Castillo-Toledo, B., Loukianov, A., & Gonzalez-Jimenez, L. E. (2014). Power line inspection via an unmanned aerial system based on the quadrotor helicopter. *Proceedings of the Mediterranean Electrotechnical Conference - MELECON*, 393–397. https://doi.org/10.1109/MELCON.2014.6820566

MAPIR. (n.d.). *Survey3W Camera - Orange+Cyan+NIR (OCN, NDVI) - MAPIR CAMERA*. Retrieved August 2, 2022, from https://www.mapir.camera/en-gb/products/survey3w-camera-orange-cyan-nir-ocn-ndvi

Matikainen, L., Lehtomäki, M., Ahokas, E., Hyyppä, J., Karjalainen, M., Jaakkola, A., Kukko, A., & Heinonen, T. (2016). Remote sensing methods for power line corridor surveys. *ISPRS Journal of Photogrammetry and Remote Sensing*, *119*, 10–31. https://doi.org/10.1016/j.isprsjprs.2016.04.011

Mills, S. J., Gerardo Castro, M. P., Li, Z., Cai, J., Hayward, R., Mejias, L., & Walker, R. A. (2010). Evaluation of Aerial Remote Sensing Techniques for Vegetation Management in Power-Line Corridors. *IEEE Transactions on Geoscience and Remote Sensing*, *48*(9), 3379–3390. https://doi.org/10.1109/TGRS.2010.2046905

Nguyen, N. D., Do, T., Ngo, T. D., & Le, D. D. (2020). An Evaluation of Deep Learning Methods for Small Object Detection. *Journal of Electrical and Computer Engineering*, *2020*, 1–18. https://doi.org/10.1155/2020/3189691

Pérez-Ortiz, M., Peña, J. M., Gutiérrez, P. A., Torres-Sánchez, J., Hervás-Martínez, C., & López-Granados, F. (2015). A semi-supervised system for weed mapping in sunflower crops using unmanned aerial vehicles and a crop row detection method. *Applied Soft Computing*, *37*, 533–544. https://doi.org/10.1016/j.asoc.2015.08.027

Rangnekar, A., Mokashi, N., Ientilucci, E., Kanan, C., & Hoffman, M. (2017). *Aerial Spectral Super-Resolution using Conditional Adversarial Networks*. http://arxiv.org/abs/1712.08690

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 9351, pp. 234–241). https://doi.org/10.1007/978-3-319-24574-4_28

Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., & Chen, X. (2016). Improved Techniques for Training GANs. *Advances in Neural Information Processing Systems*, 2234–2242. https://doi.org/10.48550/arxiv.1606.03498

Sun, C., Jones, R., Talbot, H., Wu, X., Cheong, K., Beare, R., Buckley, M., & Berman, M. (2006). Measuring the distance of vegetation from powerlines using stereo vision. *ISPRS Journal of Photogrammetry and Remote Sensing*, *60*(4), 269–283. https://doi.org/10.1016/j.isprsjprs.2006.03.004

Wada, K. (2016). *Labelme: Image Polygonal Annotation with Python*. https://github.com/wkentaro/labelme

Wang, T. C., Liu, M. Y., Zhu, J. Y., Tao, A., Kautz, J., & Catanzaro, B. (2018). High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 8798–8807. https://doi.org/10.1109/CVPR.2018.00917

Wang, W., Xie, E., Song, X., Zang, Y., Wang, W., Lu, T., Yu, G., & Shen, C. (2019). Efficient and Accurate Arbitrary-Shaped Text Detection with Pixel Aggregation Network. *Proceedings of the IEEE International Conference on Computer Vision*, *2019-Octob*, 8439–8448. https://doi.org/10.1109/ICCV.2019.00853

Yan, Y., Zhang, L., Li, J., Wei, W., & Zhang, Y. (2018). Accurate Spectral Super-Resolution from Single RGB Image Using Multi-scale CNN. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 11257 LNCS* (pp. 206–217). Springer Verlag. https://doi.org/10.1007/978-3-030-03335-4_18

Zeng, T., Diao, C., & Lu, D. (2021). U-Net-Based Multispectral Image Generation from an RGB Image. *IEEE Access*, *9*(2), 43387–43396. https://doi.org/10.1109/ACCESS.2021.3066472

Figures

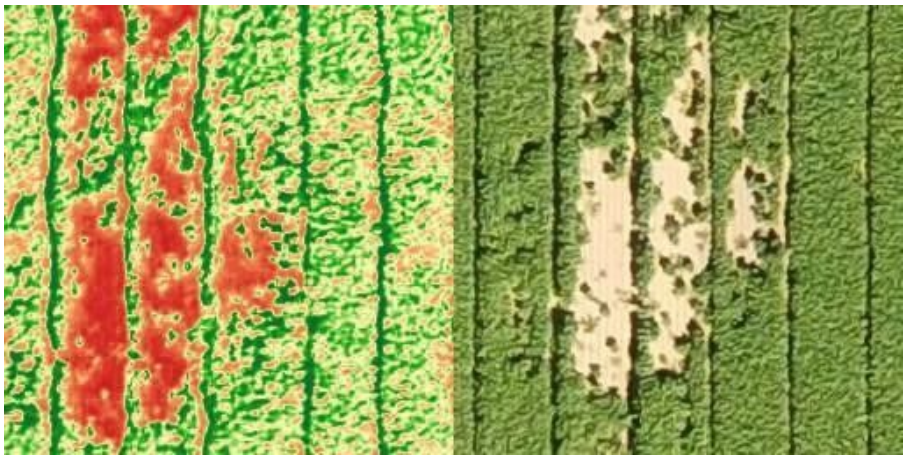

Figure 1

Sample of LUT-RGB dataset

Figure 2

Annotated powerlines in LabelMe

Figure 3

YoLov5 architecture

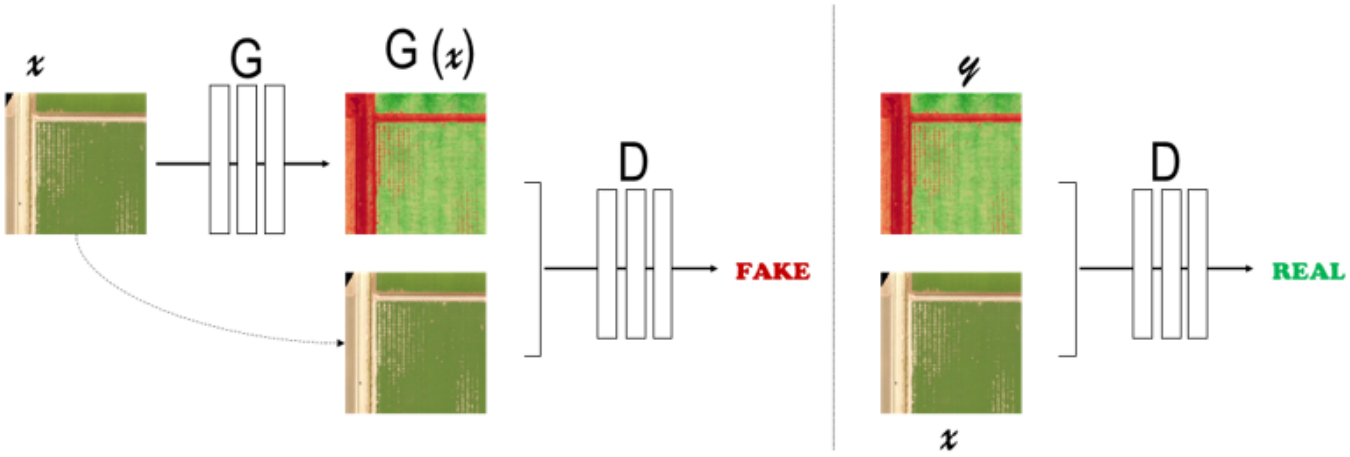

Figure 4

GAN training process

Figure 5

Metrics of model

Figure 6

Learning curves for loss functions

Figure 7

Metric charts (a) Confusion matrix (b) F1 Curve (c) P-Curve (d) R-Curve

Figure 8

PR Curve

Figure 9

Mosaic of ground truth against predicted results on the test samples

Figure 10

Comparison of qualitative results of full and compressed GAN models

Figure 11

Visual results of pix2pixHD

Figure 12

Vegetation hazard detection framework

Figure 13

(a) Sensor Components (b) Sensor platform

Figure 14

Proposed detection framework workflow