

WP1 D2 - Visual Data Assessment Report

Version 1.0

2025-08-04

© 2025 CGI Inc.

Document Revision

Date	Issue	Author	Status
04/08/2025	1.0	I.Psarra/T.Wilson/A.Halim	Approved

Table of contents

1 Introduction 1

1.1 Project Overview 1

1.2 Document Purpose 1

2 Data Overview 2

2.1 Image Dataset Overview 2

2.1.1 Data Source and Equipment 2

2.1.2 Methodology and Data Review 2

2.1.3 Data Storage and Access 3

2.1.4 Fugro System and Data Concept 3

2.2 Metadata Requirements for Image Analysis 3

2.3 Image Quality Assessment Approach 4

2.3.1 Manual Review and Scoring Criteria 4

2.3.2 Automated Image Evaluation Using Machine Vision 9

2.3.3 Network Data and Visual Data Assessment 11

2.4 Use of LiDAR Data to Support Phase Identification 14

2.5 Role of Data Availability in Trial Network Selection 15

3 Findings and Observations 16

4 Conclusions 17

1 Introduction

1.1 Project Overview

The HV Pilot aims to identify the phases to which two-wire sections of HV overhead line are connected in order to be able to model, and ultimately minimise, phase unbalance. Two techniques will be used. Smart meter voltage data will be used, building on the learning of the SMITN project. The other technique will use imagery and LiDAR data to build a virtual model of the network to enable the connection of wires to be physically traced. These techniques will be developed and tested to determine accuracy and cost effectiveness compared to site inspections.

1.2 Document Purpose

The purpose of this document is to evaluate the imagery and associated data collected during the helicopter survey to determine its suitability for supporting phase identification for single-phase transformers through machine vision techniques. The focus is on assessing the visibility, quality, and structure of the image data and particularly the ability to detect and trace single-phase connections from overhead lines. The document also outlines the metadata available to support the analysis and identifies key factors that affect image usability. This assessment forms the basis for selecting a test area where image coverage, clarity, and alignment with network attributes are strong enough to enable reliable analysis and model development.

2 Data Overview

2.1 Image Dataset Overview

The dataset used for this report and project has been sourced directly from aerial surveys conducted by the National Grid Electricity Distribution Helicopter Unit. This dedicated team operates a fleet of helicopters equipped with advanced imaging technology to carry out routine and targeted inspections of overhead lines and associated infrastructure.

These aerial surveys capture high-resolution imagery and spatial data that are critical for network analysis, asset management, and identifying trial areas. The Helicopter Unit plays a vital role in supporting NGED's commitment to network reliability and safety by enabling rapid identification of potential faults, asset deterioration, and environmental risks.

To support the identification of single-phase lines using machine vision techniques, a subset of this dataset consisting of high-resolution aerial photographs have been reviewed with accompanying metadata and LiDAR data captured during a helicopter survey of the distribution network.

2.1.1 Data Source and Equipment

The imagery for this project was captured by a helicopter system configured to take photographs in multiple directions: to the right side, directly downward (nadir), and to the left side. This multi-perspective approach ensures comprehensive coverage of the network infrastructure. A review of the imagery collection setup has indicated that this approach provides broad and effective coverage, allowing for clear identification of relevant network features.

This setup was reviewed as part of the assessment and has been found to provide broad and effective coverage of the network infrastructure. By capturing multiple perspectives - left, right, and downward - the imagery offers enhanced visibility of the infrastructure from different angles. In several cases, we observed that the left-facing camera provided the clearest view of the connection between three-phase lines and single-phase lines, while in other instances the right or nadir views offered better context depending on terrain, line orientation, or obstructions. The availability of these complementary viewpoints increases the likelihood of capturing the relevant infrastructure details necessary for phase identification.



Figure 1: Left Image



Figure 2: Nadir Image

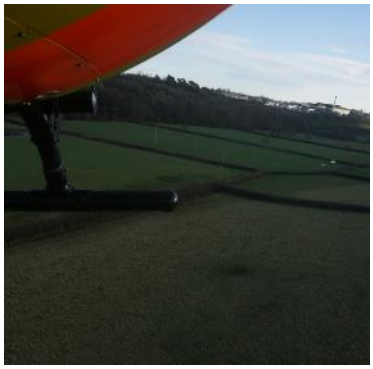


Figure 3: Right Image

2.1.2 Methodology and Data Review

A manual visual assessment was carried out by the team, systematically navigating through the high-resolution images to identify instances where the connections between single-phase and three-phase lines were clearly visible. To assist in this analysis, Fugro provided a small sample of imagery

alongside associated metadata, enabling the team to understand the structure and type of metadata linked to each image.

In addition to the imagery, a subset of LiDAR data was successfully extracted from the Roames platform, allowing for a preliminary evaluation of how this data could support the phase identification process. Both manual and internal tool-based analyses were conducted during the initial feasibility analysis, which was carried out after the data was received directly from Fugro and stored on an internal server.

2.1.3 Data Storage and Access

The data from Fugro is securely stored, with controlled access, ensuring that only the project team has permission to view and analyse it. For the initial analysis, the data was stored on an internal server. Going forward, Machine Vision AI analysis will be carried out in a cloud-based environment designed for the project, enabling the processing of large volumes of images and metadata.

Further agreements are required to finalise the data access protocol. Specifically, it remains to be confirmed whether the project will receive data extracts directly from the Roames platform or the raw imagery and LiDAR files from the helicopter survey team.

2.1.4 Fugro System and Data Concept

The Fugro system used in this project is based on the Roames platform, a comprehensive solution for processing and analysing geospatial data, including aerial imagery and LiDAR data. Roames is designed to handle large volumes of data captured via helicopter or other aerial platforms, facilitating the extraction of valuable insights from large-scale infrastructure surveys.

The data concept behind the Roames platform involves the integration of several data types: high-resolution imagery, LiDAR data, and metadata. LiDAR provides precise 3D measurements of the environment, offering detailed information about the location and orientation of infrastructure elements. The accompanying imagery provides visual context to these measurements, which is essential for understanding the network structure. Metadata describes the characteristics of the imagery and LiDAR data, such as the time of capture, sensor parameters, and geospatial positioning. This combination of data types enhances the ability to perform accurate assessments of the distribution network and supports the identification of key infrastructure features, such as the connection between single-phase and three-phase lines.

By leveraging Fugro's system, the project aims to optimise the identification of single-phase lines through machine vision and advanced data analysis techniques, ultimately contributing to a more efficient and accurate network management process.

2.2 Metadata Requirements for Image Analysis

To enable accurate image analysis and ensure spatial traceability between frames, certain key metadata elements are assumed to be available for each image. These include:

- GPS location of the camera when the image was captured
- Bearing or direction the helicopter (or camera) was facing
- Camera orientation, to understand the viewing angle of the image

This information is typically embedded in the EXIF block of each image file. From the sample images reviewed, we were able to successfully extract the following data of interest:

- Image dimensions (e.g. resolution and aspect ratio)
- Camera orientation
- Estimated field of view.

This metadata plays a central role in allowing the Machine Vision model to interpret infrastructure layout, identify starting points of single-phase lines, and trace connections across a sequence of images. To support this, we aim to build a framework that enables the model to track and follow lines, from one frame to another as the helicopter follows the network.

2.3 Image Quality Assessment Approach

2.3.1 Manual Review and Scoring Criteria

The following outlines the initial strategy used to evaluate the suitability of the aerial imagery. The assessment is structured around three key criteria:

1. **Line Visibility** - whether overhead lines and connection points are clearly visible
2. **Image Quality** - evaluation of focus, contrast, and overall clarity
3. **Presence of Issues** - identification of any obstructions, distortions, or inconsistencies that could affect processing

This preliminary assessment provides an initial filter to determine which images are likely to be usable. A more detailed, secondary evaluation was carried out by the AI model, using automated scoring techniques. To ensure the methodology is tested using the most reliable inputs, we will prioritise areas covered by Cycle 6 and Cycle 7 of the helicopter survey, as these have been highlighted by the survey team for their better image resolution and data quality. Please see Appendix A for a further breakdown of cycle coverage for Cycles 1 to 7.

(A) Line Visibility

- All three input wires are clearly visible.
- All two output wires are clearly visible.
- Connection points are fully visible.
- Wires are distinguishable from each other.
- Wire paths are traceable in both directions from the pole.

(B) Image Quality

- Image is sharp, no significant motion blur.
- Image is well-lit, no harsh shadows or overexposure.
- No visual obstructions (e.g. trees, birds, glare).
- Camera angle provides a clear view of geometry (not tilted/oblique).
- Image cantered on transition (not cut off or framed poorly).

(C) Presence of Issues

- One or more wires are cut off or out of frame.
- Connection point is partially blocked or unclear.
- Wire colours/positions are blending with background.
- Image has strong glare, blur, or shadow over key parts.
- Only the poles are visible.

Scoring Guide

Total A + B Checks	Section C Checks	Suggested Score
9–10	0	10 – Perfect
7–8	≤1	9 – Excellent
6–7	≤2	8 – Very Good
5–6	≤3	7 – Good
4–5	3–4	6 – Usable w/ Effort
3–4	≥4	5 – Borderline
≤3	≥5	1–4 – Unusable

2.3.1.1 Priority Substations

Priority substations are a focused subset of primary substations selected from a broader candidate list to support the targeted review of helicopter survey data. From those substations, NGED HV planning teams have identified a list of priority HV feeders with possible concern for phase imbalance. These substations represent the most suitable sites for initial testing of the helicopter imagery. For more detail, see Section 4 of the WP1 D1 Network Data Assessment Report by Loughborough University.

Primary	Feeder	Done	Cycle No.	No. Images	Assessment (/10)
350050	0022	☒	Only in Cycle 4		
	0023	☒			
	0024	☒			
310011	0017	☒	Only in Cycle 5		
	0018	☒			
310016	0019	☒	6	0	NA
	0021	☒	Only in Cycle 4		
	0024	☒	Not present in any Cycle		
580031	0097	☒	Only in Cycle 5		
	0098	☒			
340012	0017	☒	Only in Cycle 5		
	0018	☒			
600017	0031	☒	6	20	8.4
	0034	☒		29	7.3
350015	4066	☒	6	23	3.6
	4067	☒		18	5.1

670114	0005	☒	Not present in any Cycle
	0006	☒	
750003	0001	☒	Not present in any Cycle
	0007	☒	

2.3.1.2 Additional Substations

Loughborough University provided a further list of primary substations based on the criteria outlined in Section 6 of WP1 D1 Network Data Assessment Report. This list is currently being refined, and the final version along with the selection criteria and full visual assessment will be included in the WP1 D3 - Area Selection report.

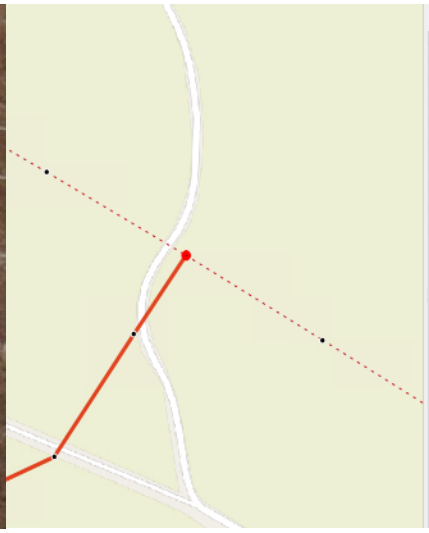
Primary	Feeder	Done	Cycle No.	No. Images	Assessment (/10)
250300	0002	☒	Not present in any Cycle		
	0003	☒			
310032	0017	☒	Not present in any Cycle		
	0018	☒			
350011	0017	☒	Not present in any Cycle		
	0023	☒	6	0	NA
350012	4046	☒	Not present in any Cycle		
	4047	☒			
350014	4055	☒	Only in Cycle 5		
	4057	☒	Not present in any Cycle		
417140	0042	☒	Potential option (no network grid lines but lots of images)		
	0046	☒	6	0	NA
417560	0041	☒	6	20	7.0
	0042	☒	6	12	5.6
437380	0045	☒	Only in Cycle 4		
	0046	☒	6	0	NA
457290	0041	☒	6	0	NA
	0044	☒	6	14	4.7
600085	0006	☒	Only in Cycle 5		
	0009	☒	6	8	3.2
670042	0001	☒	Only in Cycle 1,2		
	0004	☒			
670067	0002	☒	Only in Cycle 2		
	0011	☒			
670090	0002	☒	Only in Cycle 1,2		
	0005	☒			
670096	0005	☒	Not present in any Cycle		
	0006	☒			
	0008	☒			
	0033	☒			
670108	0003	☒	Not present in any Cycle		
	0006	☒			
	0008	☒			
670109	0003	☒	Not present in any Cycle		
	0004	☒			

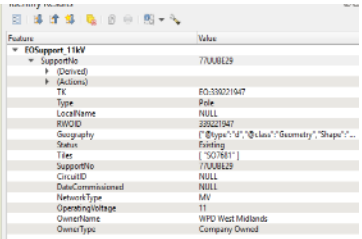
670115	0013	☒	Only in Cycle 2		
	0015	☒			
670126	0001	☒	Only in Cycle 1		
	0002	☒			
	0007	☒			
670129	0002	☒	Not present in any Cycle		
	0007	☒			
	0011	☒			
670143	0001	☒	Not present in any Cycle		
	0006	☒			
750004	0001	☒	Only in Cycle 1		
	0005	☒			
	0006	☒			
770003	0001	☒	6	8/8	8.125
	0009	☒	6	19/26	6.57

This initial manual assessment provides a structured approach for evaluating image suitability, ensuring that the imagery is clear, traceable, and relevant for phase identification. A similar evaluation will be carried out across additional feeders, allowing comparison of image quality and visibility across different parts of the network. The goal is to identify a test area where the data consistently meets the required standards, providing the best conditions for developing and validating the AI model.

Example 1: Feeder 770003_0001







Latitude	52.43412763449536
Longitude	-2.344302115508406
Assessment	9 – Excellent

Example 2: Feeder 770003-0009



Latitude	52.487100828045186
Longitude	-2.409040959619637
Assessment	1– Unusable

Although the pole itself is visible, the overall image quality makes it difficult to use for analysis. The lighting is quite poor, resulting in a dark image where the wires are hard to see. The photo appears to have been taken from a high altitude and at a slight angle, which makes it harder to pick out finer details like wire connections or directions.

Further examples and detailed evaluations of individual images are provided in Appendix B.

2.3.2 Automated Image Evaluation Using Machine Vision

To support the selection of high-quality imagery for phase identification, we use a Machine Vision model that automatically evaluates each image against a range of visual criteria. The model assesses key factors such as focus, contrast, and overall clarity, along with the image's ability to reveal structural elements like power lines and their connections. The model also calculates normalised and percentage variance scores to allow consistent comparisons across large datasets.

The model is specifically trained to determine whether power lines are present and whether their connectivity can be clearly interpreted. Each image is assigned a normalised quality score between 0 and 1, where a higher score indicates that the model can confidently detect and trace the power lines, while a lower score suggests that visibility is insufficient for reliable analysis.

The scoring is based on the model's internal logic, which has been trained to recognise features associated with good-quality imagery and penalise issues such as blur, obstructions, or poor lighting. While the exact formula is embedded within the model's internal logic, it reflects the model's ability to evaluate how well an image can support accurate phase identification. While an image may have multiple visual issues, the final score ultimately reflects whether the model is able to understand and trace the power lines. Even if some metrics are poor, a higher score can still be achieved if the lines and their connectivity are clearly detectable.

Based on our previous experience:

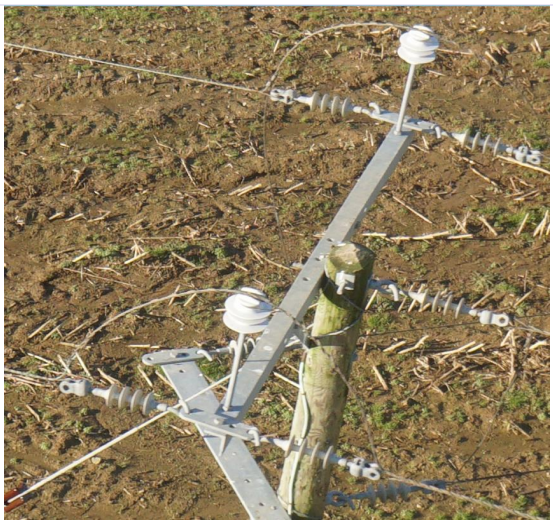
- Scores below 12% (0.12) are considered unusable and are excluded from further processing.
- The top 25% (0.25) of scores are categorised as high quality and suitable for detailed analysis.
- Mid-range scores are considered usable with caution and may require additional validation.

In cases where low-quality images are identified, we may request additional captures from the helicopter team to ensure adequate coverage. The examples below show how the scores are applied to guide data selection and prioritise feeders with the highest-quality imagery.

Example 3: Feeder 600017_0031	**Example 4: Feeder 600017_0034**

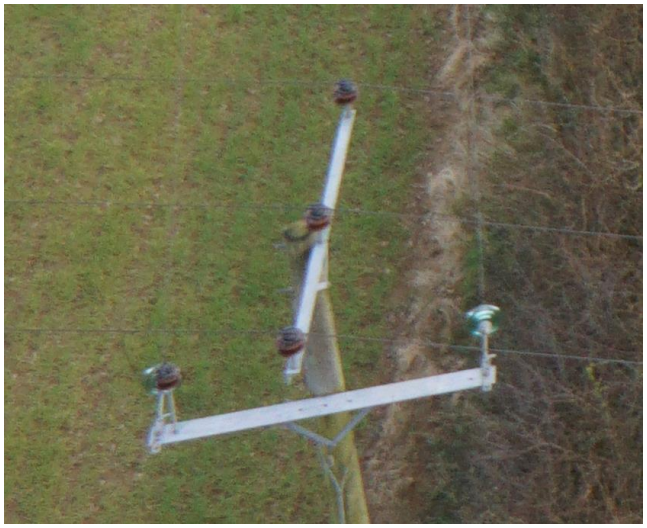


Model Assessment	0.85/1
Manual Assessment	9/10



Model Assessment	0.95/1
Manual Assessment	10/10

Example 5: Feeder 350015_4066



Model Assessment (nvariance)	0.11/1
Manual Assessment	2/10

Example 6: Feeder 350015_4066



Model Assessment (nvariance)	0.13/1
Manual Assessment	3/10

Below is a summary of the image quality scores generated by the model for four feeders.

Primary	Feeder	Done	Cycle No.	No. Images	Assessment (/10)	Model(/1)
600017	0031	☒	6	20	7.0	0.33
	0034	☒		29	7.3	0.36
350015	4066	☒	6	23	3.6	0.11
	4067	☒		18	5.1	0.17

Further examples and detailed evaluations of individual images are provided in Appendix B. Additional assessments across more feeders will be carried out, and the findings will be presented in the WP1-D3 Area Selection Report.

2.3.3 Network Data and Visual Data Assessment

The network data describes the topology of the HV network, including cable types and impedance values for each feeder. This information is sourced from NGED's Electric Office GIS database, which serves as the master reference for geographic network data across the organisation. It provides details on the layout of underground cables, overhead lines, and the connectivity between substations, including whether each wire is single-phase or three-phase.

During the visual data assessment, we observed several instances where the network data classification did not match the physical infrastructure seen in the imagery and LiDAR data. In some areas, lines labelled as single-phase in the GIS appeared to be three-phase when reviewed visually, while in other cases, lines marked as three-phase were clearly single-phase. These inconsistencies highlight the value of using visual and LiDAR data to validate and improve the accuracy of existing network records. In Images 4 and 5, you can see examples of inconsistencies between the visual data and the network data.

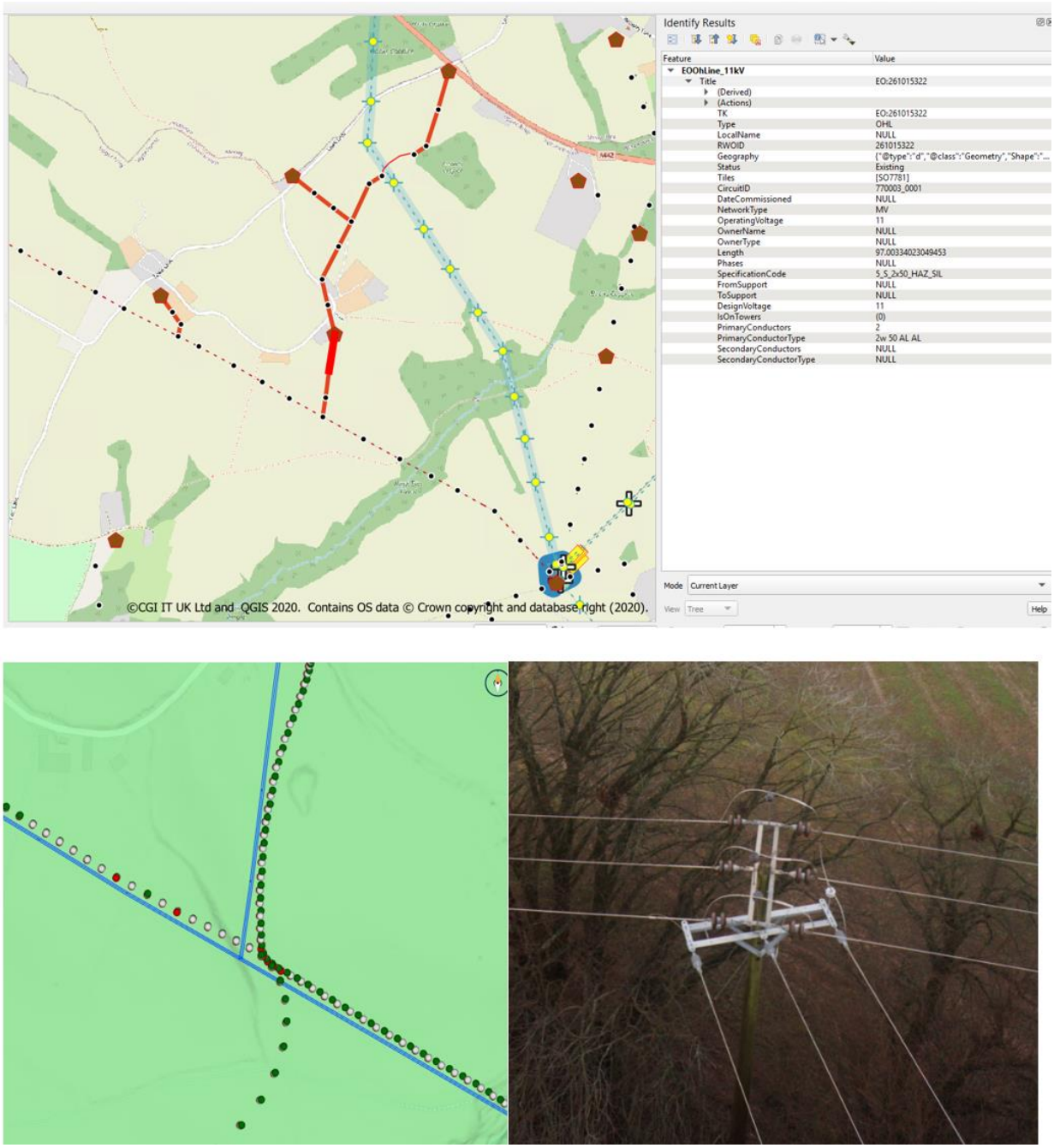


Figure 4: Network data indicates two primary conductors, but the aerial imagery shows three distinct conductors.

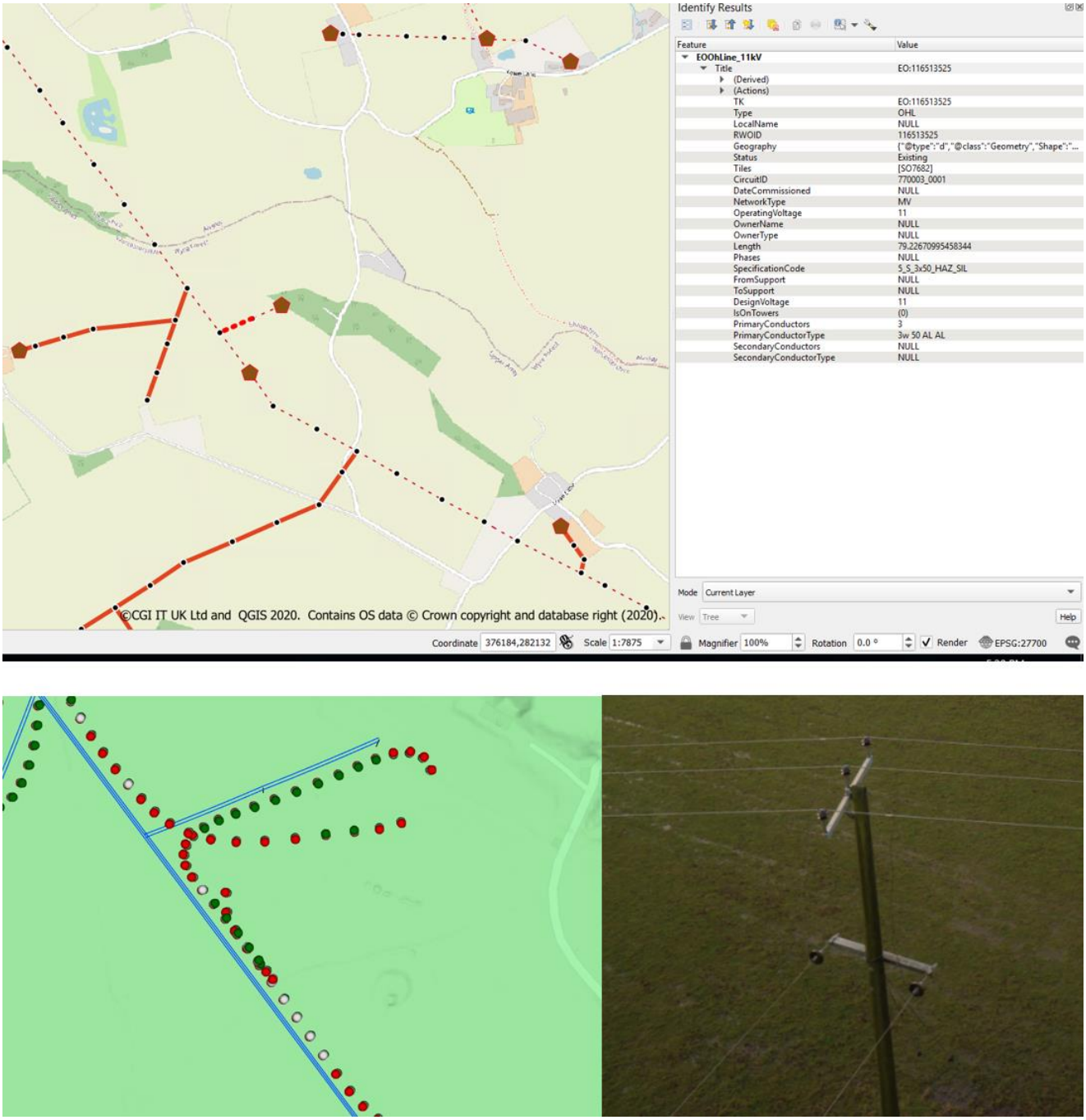


Figure 5: Network data indicates three primary conductors, but the aerial imagery shows single-phase line.

Note that the network data used in this assessment are the INM data, as the most recent extracts from the GIS have not yet been received.

2.4 Use of LiDAR Data to Support Phase Identification

The LiDAR data and associated files can also be processed to support the phase identification workflow. From the available datasets, it has been possible to extract detailed information from DXF files representing lines and poles, as well as from the LAZ point cloud and terrain GeoTIFFs. These sources provide high-resolution 3D spatial data, enabling accurate reconstruction of the physical network layout.

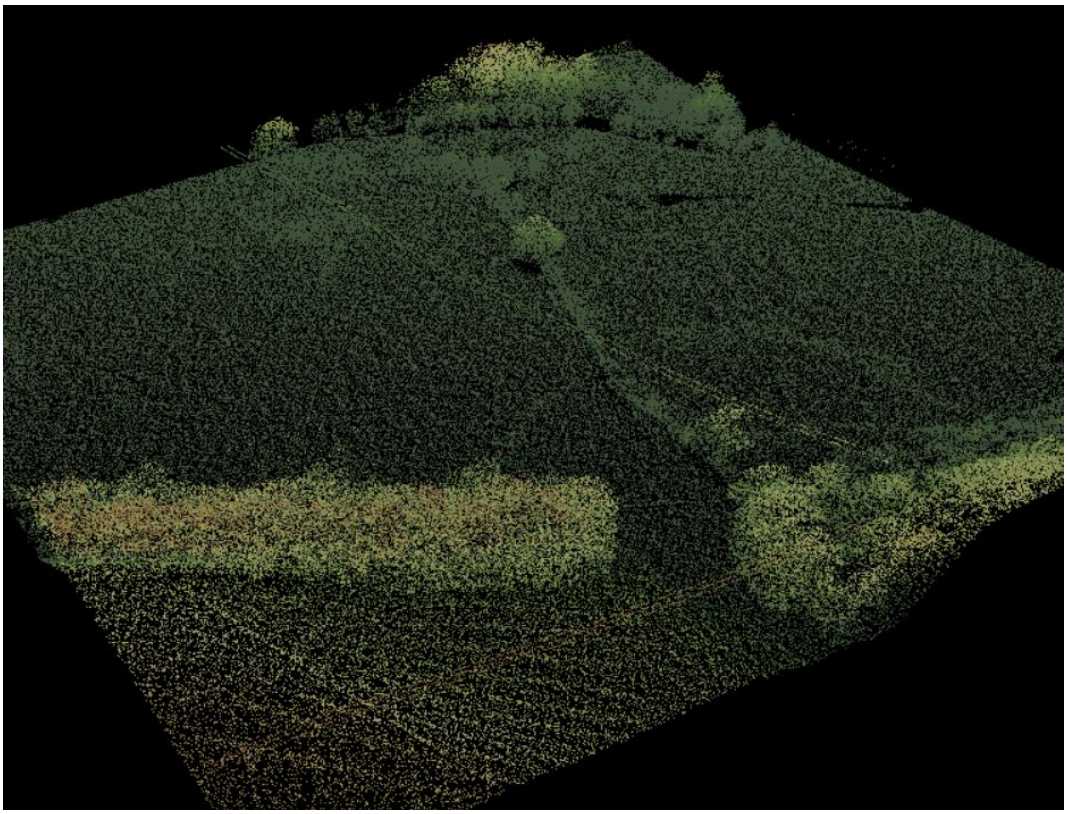


Figure 6: LiDAR point cloud data captured at a location where a single-phase line starts.

LiDAR excels in providing accurate, dense, and georeferenced 3D dimensional information. The LiDAR data can be combined with images to improve the accuracy of interpreting details such as power lines and poles. The LAZ point cloud data, which is a compressed format for storing LiDAR data, provides a wealth of information about the physical layout of the network. By blending this dimensional information with images, our platform can more accurately identify and trace the connections between power lines and poles.

This integration allows for a more comprehensive analysis, as the 3D data from LiDAR can fill in gaps or provide additional context that might not be visible in the images alone. For example, if a power line is partially obscured in an image, the LiDAR data can help confirm its location and orientation, ensuring that the analysis remains accurate and reliable.

2.5 Role of Data Availability in Trial Network Selection

The availability and completeness of helicopter survey data is a key factor in the selection of the trial area. To ensure the methodology is tested on the most suitable inputs, areas with high-quality data coverage will be prioritised, specifically from Cycles 6 and 7, which have been identified by the helicopter survey team as offering the best image resolution and overall data integrity. This is to ensure that the trial is based on reliable inputs for asset identification and modelling.

During the initial assessment, several limitations were identified in the data that could impact the effectiveness of phase identification:

- Incomplete capture of single-phase start points from three-phase lines, which affects the ability to trace single-phase lines from their three-phase connections.
- Limited visibility of physical line connections due to the high altitude of the helicopter during survey flights.
- Obstructed imagery or background noise, which occasionally masks critical features such as lines and connection points.
- Gaps in coverage resulting from no-fly zones, where data could not be captured due to airspace restrictions.

In selecting the trial areas, we will aim to avoid regions impacted by these issues and instead focus on locations where the imagery and LiDAR data are most complete and clear, increasing the chances of accurately identifying phase connections. However, it is recognised that some limitations may still be present, and these will be carefully documented and considered in the interpretation of results.

3 Findings and Observations

The initial feasibility analysis has provided several key insights into the quality and structure of the helicopter survey data and its potential to support machine vision-based phase identification for single-phase transformers.

- **Image Quality and Usability Vary Across Feeders**
  Manual and automated assessments showed significant variation in image quality across different feeders. Some feeders displayed high visibility, clear line paths, and good lighting conditions, while others suffered from poor resolution, glare, and high-altitude capture angles, making them unsuitable for analysis.

- **Metadata Availability**
  The sample images came with metadata that included details like image dimensions, camera orientation, and estimated field of view. This information helps us understand the direction the that the helicopter was flying, which is important for tracing connections from one image to the next.

- **LiDAR Data Enhances Phase Identification**
  LiDAR data could support phase identification by helping to locate poles and trace single-phase lines, especially where imagery is unclear or obstructed.

- **Multiple Camera Views Improve Asset Capture**
  Having three images for each location (right, left, and nadir views) gave us more options to choose from, increasing the chances of finding a clear view of the phase connections.

- **Data Limitations Could Affect Trial Accuracy**
  The analysis highlighted several limitations in the helicopter data: incomplete capture of network single-phase start points, reduced visibility due to flight altitude, image obstructions, and gaps in no-fly zones. These factors can reduce the effectiveness of machine vision models.

- **Data Accuracy and Validation**
  Discrepancies between the network data and visual data such as mismatches in the number of conductors, highlight that errors in the network data need to be investigated and validated to ensure accurate phase identification.

4 Conclusions

This assessment shows that the helicopter survey data from Cycles 6 and 7, hosted on the Roames platform, provides a promising starting point for phase identification using machine vision.

While the data has clear potential, there are still some challenges to address. Differences in image quality and occasional obstructions from terrain or lighting can deteriorate the results. The inclusion of LiDAR data could help to fill gaps and offer additional spatial context where images alone may not be enough.

Looking ahead, the selection of the trial area will focus on areas with the best available data and clear imagery. Although some limitations may remain, prioritising feeders with better quality data will help reduce risks and make the most of the model development phase.

To further streamline the use of the Roames dataset and accelerate future imagery analysis, enhancements to data accessibility should be considered. Enabling programmatic access through APIs within the Roames platform and standardised imagery extraction would greatly support automated processing workflows. Establishing such capabilities would reduce manual effort, improve repeatability, and help scale the analysis across larger sections of the network.

In summary, the findings from the visual data assessment support moving ahead to the next phase, which involves developing and testing machine vision techniques to identify the phase of single-phase transformers.



Appendix A

Cycle	Spans	Network km's
1	57,181	8,205.00
2	207,325	18,721.90
3	246,947	24,326.59
4	81,719	8,323.38
5	138,322	15,531.51
6	149,134	13,683.31
7	51,570	6,824.95
Total	**932,198**	**95,616.64**

Cycle 7 is currently ongoing, and the above data is expected to change.

Appendix B

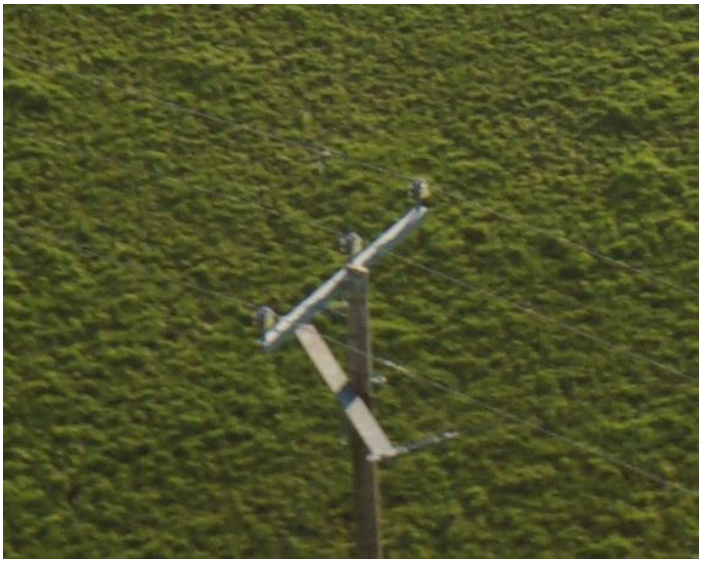
600017

0031



Manual Assessment**:** 7

Model Assessment: 0.39



Manual Assessment**:** 5

Model Assessment: 0.07



Manual Assessment**:** 7

Model Assessment: 0.11



Manual Assessment**:** 6

Model Assessment:0.15



Manual Assessment**:** 8

Model Assessment: 0.38



Manual Assessment: 8

Model Assessment: 0.54



Manual Assessment: 5

Model Assessment: 0.26



Manual Assessment: 8

Model Assessment: 0.33



Manual Assessment: 9

Model Assessment: 0.69



Manual Assessment**:** 6

Model Assessment: 0.26



Manual Assessment**:** 6

Model Assessment: 0.15



Manual Assessment**:** 9

Model Assessment: 0.85



Manual Assessment**:** 6

Model Assessment: 0.26



Manual Assessment**:** 7

Model Assessment: 0.04



Manual Assessment**:** 4

Model Assessment: 0.12



Manual Assessment: 8

Model Assessment: 0.27



Manual Assessment: 6

Model Assessment: 0.65



Manual Assessment: 6

Model Assessment: 0.17



Manual Assessment: 9

Model Assessment: 0.49



Manual Assessment: 7

Model Assessment: 0.46



Manual Assessment**:** 9

Model Assessment: 0.3

600017

0034



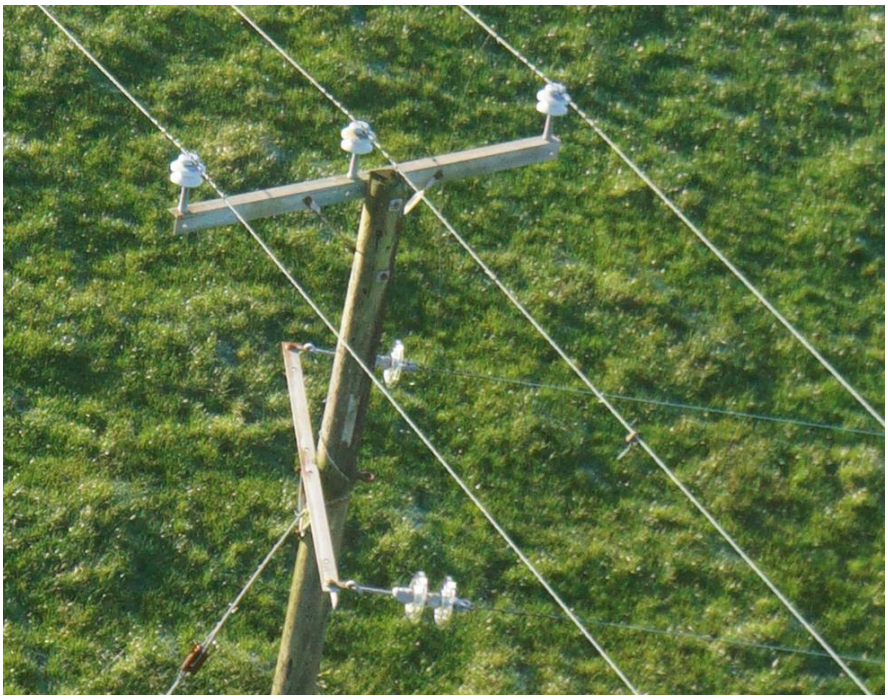
Manual Assessment**:** 9

Model Assessment: 0.3



Manual Assessment**:** 9

Model Assessment: 0.3



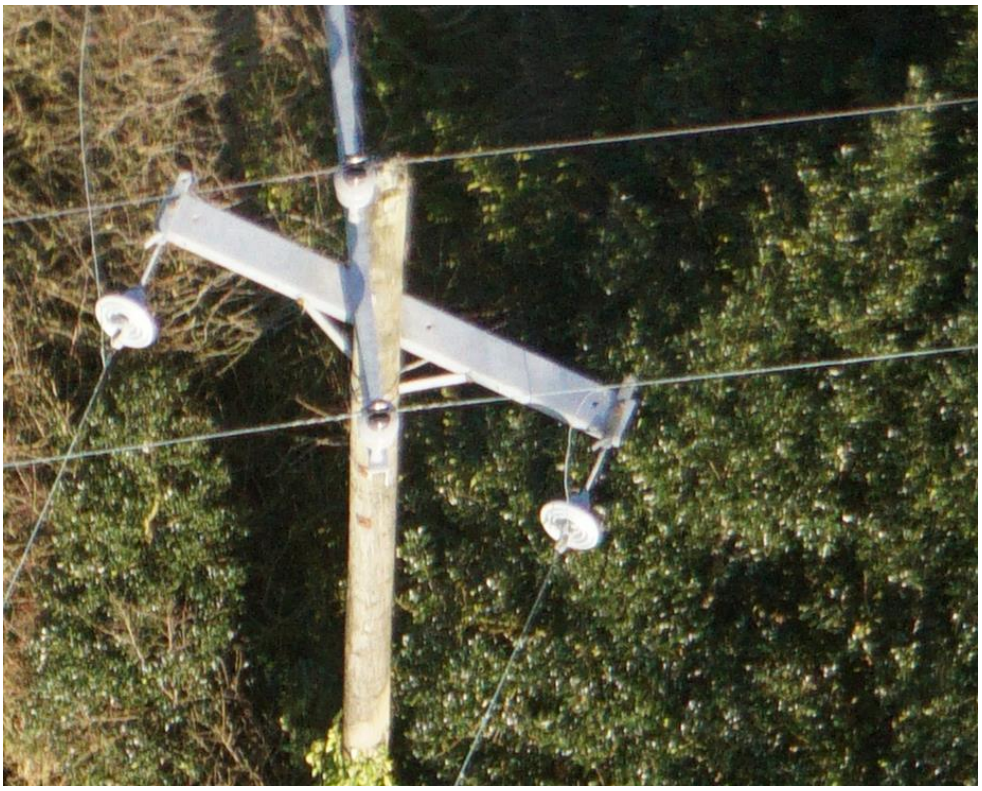
Manual Assessment**:** 9

Model Assessment: 0.3



Manual Assessment**:** 9

Model Assessment: 0.3



Manual Assessment**:** 9

Model Assessment: 0.3



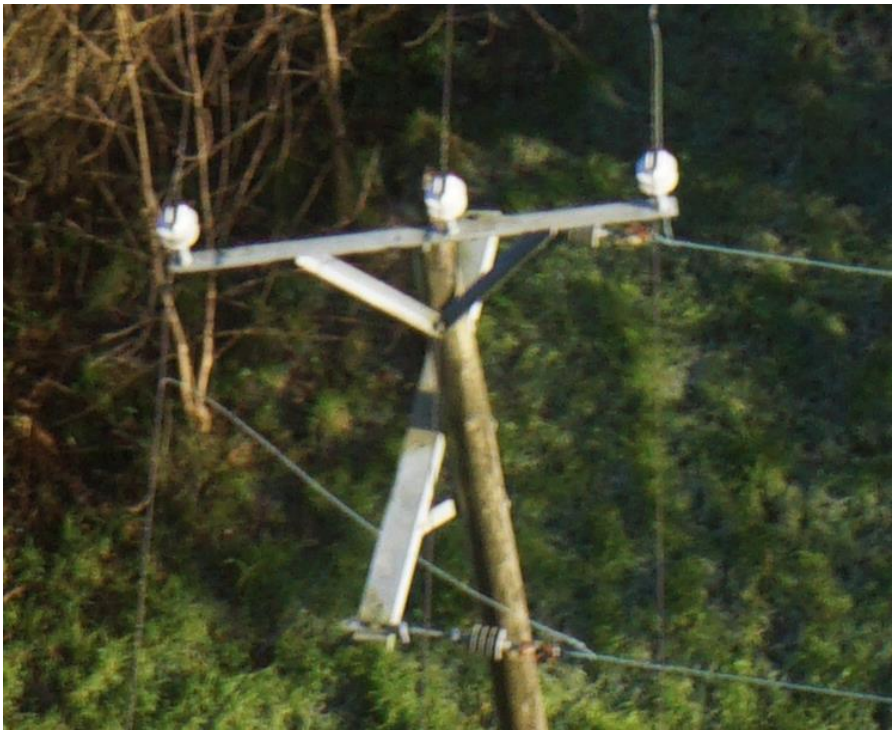
Manual Assessment**:** 9

Model Assessment: 0.3



Manual Assessment**:** 9

Model Assessment: 0.3



Manual Assessment**:** 9

Model Assessment: 0.3



Manual Assessment**:** 9

Model Assessment: 0.3



Manual Assessment**:** 9

Model Assessment: 0.3



Manual Assessment**:** 9

Model Assessment: 0.3



Manual Assessment: 9

Model Assessment: 0.3



Manual Assessment: 9

Model Assessment: 0.3



Manual Assessment**:** 9

Model Assessment: 0.3



Manual Assessment**:** 9

Model Assessment: 0.3



Manual Assessment**:** 9

Model Assessment: 0.3



Manual Assessment**:** 9

Model Assessment: 0.3



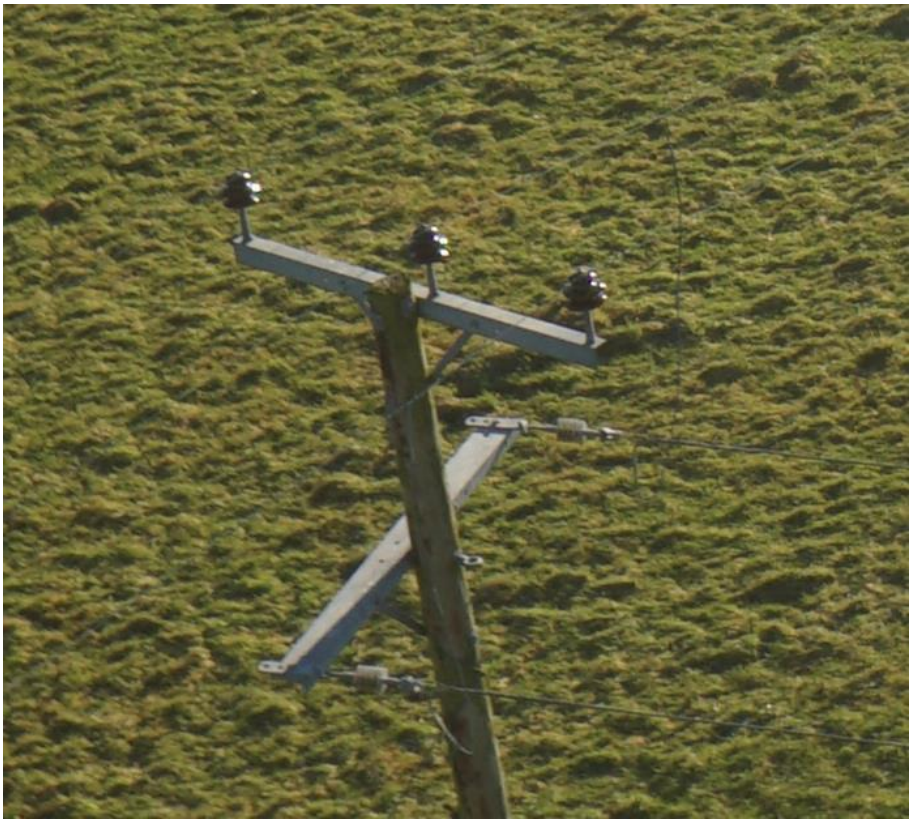
Manual Assessment**:** 9

Model Assessment: 0.3



Manual Assessment**:** 9

Model Assessment: 0.3



Manual Assessment**:** 9

Model Assessment: 0.3



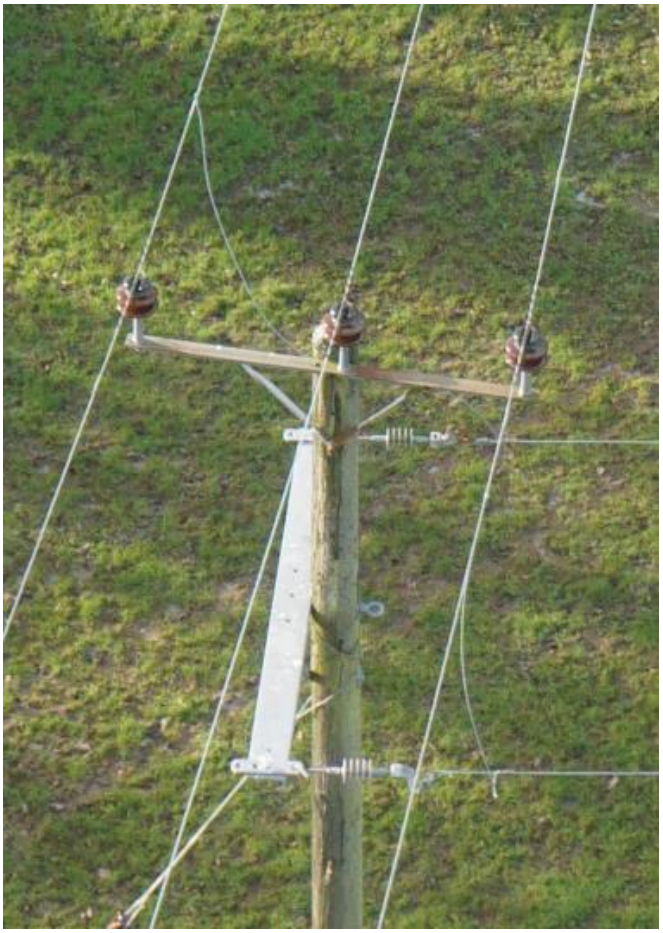
Manual Assessment**:** 9

Model Assessment: 0.3



Manual Assessment**:** 9

Model Assessment: 0.3



Manual Assessment**:** 9

Model Assessment: 0.3



Manual Assessment: 9

Model Assessment: 0.3



Manual Assessment: 9

Model Assessment: 0.3



Manual Assessment**:** 9

Model Assessment: 0.3



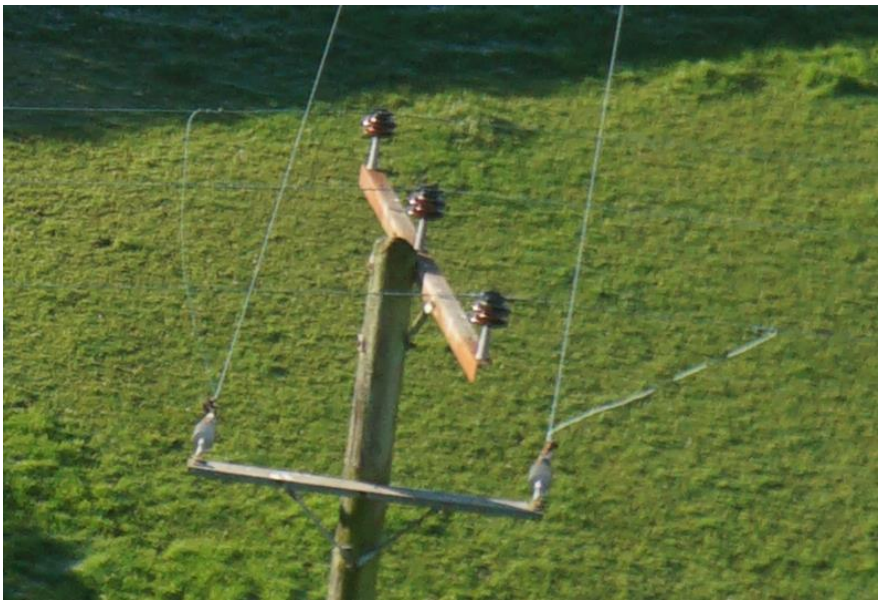
Manual Assessment**:** 9

Model Assessment: 0.3



Manual Assessment: 9

Model Assessment: 0.3



Manual Assessment: 9

Model Assessment: 0.3

