2020 International Conference on Unmanned Aircraft Systems (ICUAS)
Athens, Greece. September 1-4, 2020

# Automatic Condition Monitoring of Railway Overhead Lines from Close-Range Aerial Images and Video Data

Franz Andert[1], Nils Kornfeld[1], Florian Nikodem[2], Haiyan Li[3], Stefan Kluckner[3], Laura Gruber[3], Christian Kaiser[4]

***Abstract*— This paper is about automated condition monitoring of critical railway infrastructure using unmanned aircraft systems as flying sensors. As far as possible, automation shall include flight guidance and management as well as automated processing of large sensor data sets. Since a commercial solution must consider the regulatory framework on remotely piloted aircraft systems, the paper discusses legal issues to make allowance for flights beyond visual line of sight. The work described here is focused on Europe and Germany, however, the major principles are likely to be adaptable to other countries. Next to that, the paper presents a strategy for automated image and video data processing. It consists of a super-resolution approach where onboard video camera data from typical off-the-shelf drones can replace higher-resolution still imagery and thus avoid the necessity to use special flight systems, and a deep-learning approach where specific elements are to be detected in the images. With data from flight tests over railway overhead lines, the paper shows an automated detection of rod insulators. Moreover, it presents resolution improvements from video data so that off-the-shelf camera drones can be qualified for the detection of small defects.**

## I. Introduction

Periodical inspection and maintenance of railway infrastructure is highly important, however, current manual inspections are rather slow and subject to high personal workload. Typically, longer inspection intervals reduce cost, however, safety-critical components are then to be replaced at early stages of abrasion and long before the end of their regular lifetime. Higher material costs are obvious, especially for components which are hard to observe. To cope with that, *predictive health management* methods are developed for optimization, aimed at frequent but cheap inspections and precise prediction of optimal hardware repair and replacement schedules.

Cameras onboard small unmanned aircraft are one fast, flexible, and cheap way to inspect hardly reachable targets. Semi-automatic vehicles (e.g. multicopters with controlled hovering and other easy handling modes) with onboard cameras are state-of-the-art and allow observations by many users and without extensive flight training. In the future, fully automatic flights close to infrastructure and beyond the pilot's visual line of sight will further aid such inspection missions. In this context, this paper discusses some legal issues that arise with current and future operations.

Next to the vehicle, automation of sensor data processing is highly beneficial since typical flight campaigns will easily acquire thousands of images or many hours of video data. Therefore, it is necessary to filter raw data by their content and to identify relevant elements including their status information. For pattern recognition in image and video data, state of research are various deep learning methods together with software frameworks making them ready to use. This paper is about finding rod insulators at railway overhead lines which enables an automatic process to identify potential damages on them. The goal is to detect the insulators with highest accuracy and to achieve the ability of e.g. detecting sub-centimeter cracks while the aircraft is flown at safe distance over the overhead line. To prove these functions, flight experiments (Fig. 1) are performed with multicopters and different sensor setups.

Fig. 1. Test vehicle with onboard camera flown over railway power lines.

## II. Related Work

Among other inspection tasks, e.g. on bridges [1], [2] or wind turbines [3], unmanned aircraft are put into operation for rail and road traffic infrastructure inspection [4]. In the rail domain, typical tasks are monitoring of track geometry, near vegetation and defects on the track [5]. Next to the capabilities, the report in [6] lists some general limitations and risks, such as flying over large distances and beyond visual line of sight, dependency on command and control link and satellite navigation, ground and air collision risks,

*This work is supported by the mFUND initiative of the German Federal Ministry of Transport and Digital Infrastructure (BMVI) within the project *Smart MAintenance of Rail infrastructure by using Analytical Georeferenced Data services* (SMARAGD), Grant Agreement No. 19F2055B

[1]Institute of Transportation Systems, German Aerospace Center (DLR), 12489 Berlin, Germany Franz.Andert@dlr.de
[2]Institute of Flight Systems, DLR, 38108 Braunschweig, Germany
[3]Siemens Mobility GmbH, Germany
[4]Copting GmbH, 38104 Braunschweig, Germany

and it provides some information about regulation in the international context.

With respect to existing solutions, the paper evaluates the regulatory framework and provides a solution towards high-precision imagery. The goal is to detect and to localize sub-centimeter damages from safe flying distances and with off-the-shelf hardware.

## III. Legal Issues

For commercial flights and also research flights in the proximity of infrastructure, vehicle operation is regulated. Further, image data recording, storage, or access can be limited, depending on the existence of individual-related contents and personal interests. As the flights in this paper are conducted in Germany, the paper will present an example case on European drone regulations according to the proposed amendment by EASA [7] together with conditions under the terms of General Data Protection Regulation (GDPR) in the EU [8].

### A. Operational Risk

The paper is focused on flight operation within the EASA 'specific' category [7] which includes regular commercial flight operation, based on JARUS guidelines on Specific Operations Risk Assessment (SORA) [9]. A brief overview of the SORA methodology is described in [10]. Based on the scenario definition which includes e.g. vehicle type, flight area including safety buffers, and operational procedures, the process evaluates the typical harms, hazards and generic threats and provides systematic steps to classify ground risk and air risk to third parties in order to determine a Specific Assurance and Integrity Level (SAIL). This level further defines requirements on flight system performance and robustness, risk mitigation and operational safety objectives. The paper presents two operations for risk assessment:

*1) Experimental Flights:* This concept of operations includes a small number of research flights which will be conducted on a closed railway track on private area. The operation will be carried out in daylight and dry conditions, and in visual line of sight (VLOS) of a safety pilot. The vehicle size (see sec. IV) is below 3m, with an expected kinetic energy below 34 kJ. The operation concept considers an Emergency Response Plan (ERP), a trained operator and a self-tested vehicle but no other harm barriers as parachutes or tethers.

Following the SORA process, a Ground Risk Class **GRC 2**, Air Risk Class **ARC-a**, and resulting **SAIL I** can be determined. Based on the Operational Safety Objectives with at least "low" robustness, the following requirements are relevant:

- Defined and documented maintenance and inspection procedures, including (self-declared) crew training,
- Documented (self-declared) crew training for normal and abnormal flight situations,
- Preparation and communication of Emergency Response Plan before the flight campaign,
- Definition of operational volume, including safety distances and assurance of flight below 150ft above ground level (AGL),
- Documented pre-flight check, including inspection of the flight system and external conditions like weather,
- Normal, Abnormal and Emergency Procedures are compiled in an Operation Manual,
- Defined roles for the crew members,
- A design and installation appraisal is available for Safe Design,
- Crew members do confirm physical and mental fitness before and during flight campaign, stress / fatigue reduction with rest periods,
- Vehicle size and weight with additional payload is checked against limits,
- Fail safe procedures are implemented for autopilot control loss, data link loss, hardware, power supply, and engine failures,
- Vehicle status (health, operational volume) is monitored during the flight.

*2) Commercial Application:* For a possible regular commercial application along railway tracks also beyond visual line of sight, the JARUS standard scenario *STS-01* for aerial work applications [11] can be taken into account. This document provides a standard SORA procedure for BVLOS low-level flights of small aircraft. In this standard case with **GRC 3**, **ARC-b** and resulting **SAIL II**, higher criteria for the Operational Safety Objectives are to be met in comparison to the experimental flight example above. The standard scenario provides specific risk mitigation strategies along with technical and operational aspects required to achieve flight permission. The objectives to be met are generally similar to the list above. Some points require a higher robustness class.

### B. Data Protection and Security

In the EU, General Data Protection Regulation [8] defines general principles relating to processing of personal data and also terms in which data recording and processing is lawful. To cope with that, data collectors and/or drone operators must provide a concept which identifies the relevant legal norms, answers the resulting data protection questions and defines technical and organizational protection measures.

In the presented use case, personal data are not of interest but subject to be present by an oversight. Significant importance is given to onboard cameras as the image data may contain personal information from third parties. Along railway tracks, aerial images can contain e.g. adjacent properties, or identifiable people and cars together with geo-location and timestamp. This becomes even more important for images taken in or in the proximity of railway stations.

Organizational measures are proposed as follows:

- At first stage, it is checked whether data protection rules have to be applied. In the easiest case, the next organizational and technical measures might be disregarded e.g. for closed railways apart from third party properties.
- Possible appearance of personal information is evaluated. Depending on the type of information, people and

property owners must be informed. Permissions may be required. Here, Sec. 3 GDPR requires transparency on data recording and information where personal data may be collected from the data subject.
- A typical method to inform people is to set up of information signs providing information about data collection. Art. 13/14 GDPR provide details about the necessary contents. This is a suitable option where uninvolved people are present.

The following technical measures are proposed:

- Raw data / image access is limited as it may contain personal data. It will be freed from limited access only if no personal information is present.
- For images with potential personal information, image regions apart from the target area (i.e. the railway track and overhead lines) are filtered out. This will lower access limitations. If possible, filtering should be done with automated data and image processing.
- Published image data is checked for contents, and questionable image regions with personal data are blurred.

Apart from possible content in raw data, personal data is not identified and not processed. Thus, data protection rules are not required for processed data and meta data.

## IV. Data Acquisition over Railways

For algorithm development and data evaluation, data recording flights are performed over multiple railway tracks. The following section describes the vehicle and sensor configuration as well as the overall data recording methodology.

### A. Vehicle and Sensor Setup

Data recording is performed with two vehicles (Fig. 2). The first vehicle is custom aircraft made by *Copting* with a gimbaled high-end photo camera (see Tab. I) for research in the direction of maximal image quality. The second is an off-the-shelf drone (see Tab. II) in order to have a look into the possibilities of mass-market drones which are usually equipped with video cameras. Here, it will be evaluated whether data quality drawbacks of these systems can be compensated in some way.

Fig. 2. Test vehicles. Copting drone (left) and small DJI Mavic (right).

Both vehicles come up with cameras at stabilized gimbals and video down-link, and some flight automation which includes hovering modes and stabilized flight. The *Copting* drone can be flown with automatic waypoint navigation (including BVLOS) using a ground control station software, and the images are timestamped and geo-referenced via GPS.

TABLE I: Copting Drone with High-End Image Camera

| | |
|---|---|
| **Vehicle** | **Copting TransformerUAV Hexacopter** |
| Take-off mass | 8 kg |
| Positioning | RTK GPS |
| Flight time | 20–25 min |
| **Camera** | **Sony Alpha 7 RII** |
| Image resolution | 7952 x 5304 (42 MPixel) |
| Frame rate | 1–3 fps |
| Lens | 50 mm fixed (40 x 27 deg FOV) |
| Data format | JPEG and raw images |

TABLE II: Mass-Market Drone with 4K Video Camera

| | |
|---|---|
| **Vehicle** | **DJI Mavic Pro** |
| Take-off mass | 0.8 kg |
| Positioning | Single GPS |
| Flight time | 20 min |
| **Camera** | **Integrated camera** |
| Image resolution | 4096 x 2160 (8 MPixel) (used mode) |
| Frame rate | 24 fps |
| Lens | 8 mm fixed (78 x 50 deg FOV) |
| Data format | MPEG-4 AVC/H.264 video |

### B. Data Recording Flights

Image and video data recording is performed over different railway tracks. Generally, the flights are performed under good weather and daylight conditions and with both vehicles to get image and video data from rather the same perspective. Flight altitude is about 15 m to 25 m to ground level, with flights directly above the track and also at oblique visual perspectives. Image exposure time is kept low ($<2$ ms) to avoid motion blur. An example image is shown in Fig. 3.

Fig. 3. Sample image from Copting drone. Full image (left), 150 x 100 pixel cutout (right). The insulator distance is about 40 m, resulting pixel resolution about 2 mm. In comparison, video pixel resolution is four times lower.

## V. Automated Data Processing

Part of the motivation of this work is to decrease the amount of manual work that is to be conducted in the maintenance of overhead catenary. Thus, we present a system that uses technology to automate the work that is needed to fulfill this task. In this section we will focus on automated data processing, applied in our work.

To automatically identify assets of interest in images (video or still images), an object detection system, based on convolutional neural networks (CNN) is used. The condition monitoring is then performed by specialized CNNs on regions of interest, extracted from the original high resolution image. These image patches are tightly cropped around the asset, detected in the original high resolution image. In the current state of the project, high resolution still images are used for condition monitoring. Because readily available

consumer multicopters usually do not offer high resolution still images, but lower resolution videos, we evaluate a super-resolution approach to create higher resolution still images from video streams. A necessary focus is on applying methods that are robust in the presence of vibrations and other influences caused by the data collection with multicopters.

### A. Still Image versus Video Data

To retrieve images of the highest possible quality for the task at hand, still image recordings are used. This leads to some shortcomings, compared to using images from video streams. To create usable still images, the camera mounted on the UAV has to be targeted exactly at the assets that are supposed to be monitored. For the duration of exposure, the motion of the multicopter has to be limited, as well as vibrations affecting the camera. When video recordings are used, images in which the assets to be monitored are visible can be selected in a post processing step. The resolution of the frames is very limited, compared to the resolutions of still images.

In contrast to special aircraft systems, off-the-shelf UAVs are typically equipped with video cameras, which deliver e.g. FullHD or 4K recordings with a lossy video compression codec. Using such equipment can severely decrease the costs of a multicopter system. The following section evaluates whether it is possible to recover higher resolution from multiple image frames. In the ideal case, the processed video output is comparable to the output of a still image camera.

### B. Super-Resolution Images from Multiple Video Frames

To create higher resolution images from videos, super-resolution methods are applied. Because we want the information present in the upscaled images to be only dependent on data from the current recording, we decided against the usage of deep-learning based approaches [12], [13], [14]. In these approaches, the additional information that is used to create higher resolution images is strongly dependent on training data. So the gaps in the information from lower resolution images are filled with features common in already recorded data. Although there are methods incorporating temporally adjacent frames as additional input to the artificial neural networks [15], the results stay dependent on previously recorded training data. In the condition monitoring task at hand it is important to detect anomalies. These anomalies could get obfuscated by such a super-resolution approach. In Fig. 4 the result of a deep-learning based method is compared to the corresponding ground truth image. In the super-resolved image there are structures present in the ornaments worn by the character, which are not present in the ground truth sample.

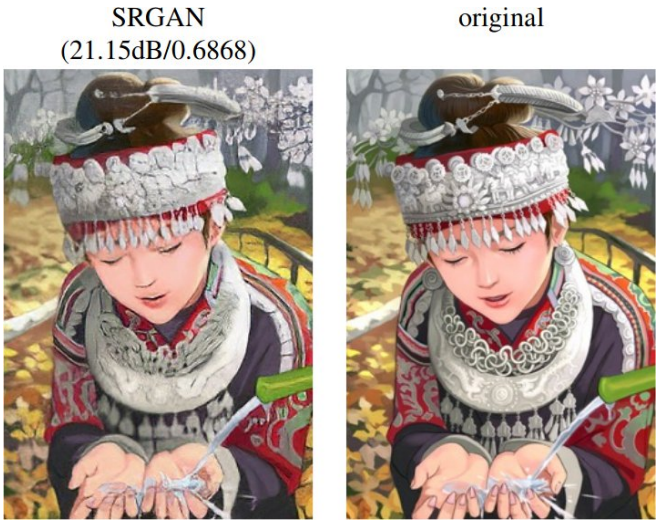


Fig. 4. Exemplary super-resolution result which shows high resolution structures in the ornaments, not present in the ground truth image. From [14].

Instead, we use a super-resolution approach that uses image registration techniques to use the additional information present in temporally adjacent frames in a video stream. Fig. 5 illustrates the general idea. The overall algorithm was proposed in [16]. To decrease motion blur we use $L_1$ norm minimization as proposed in [17].

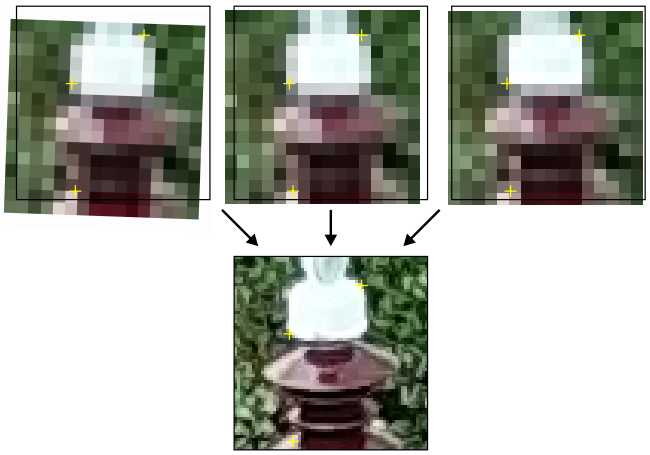

Fig. 5. General idea of super-resolution: Based on homologous features in multiple frames matched with sub-pixel accuracy, images or image regions are aligned at sub-pixel scale. By combining them, a new image with higher resolution can be computed. Obviously, the quality of the new image is highly dependent on feature matching accuracy and the presence of significant image contents.

The TV-$L^1$ optical flow algorithm [18] has shown to cause artifacts in presence of vibrations of the UAV, which are shown as an example in Fig. 6. So we decided to use the Farnebäck method [19] for image registration. This method

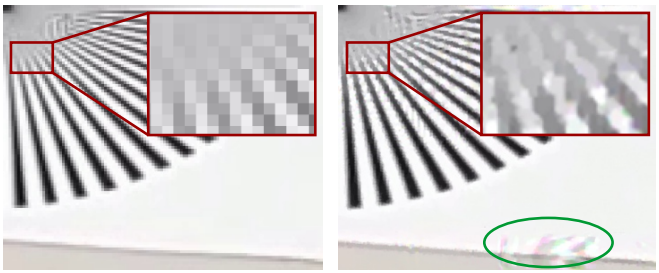

Fig. 6. Left: Sample video frame of a recording of a Siemens star. Right: Image created by TV-$L^1$-super-resolution, it shows better resolution but also artifacts when compressed video is used as input.

was originally designed to be an optical flow algorithm that is robust to the vibrations in videos, recorded from helicopters. To get the results we present in the evaluation, we used the following parameters listed in Tab. III.

TABLE III: PARAMETERS USED TO ENHANCE IMAGE RESOLUTION

| | |
|---|---|
| **Farnebäck algorithm** | |
| Iterations | 3 |
| Number of Levels | 4 |
| Scale of image pyramid | 0.5 |
| Window size | $15 \times 15$ |
| Degree of the polynomial | 7 |
| Smoothing $\sigma$ for derivatives | 1.5 |
| **Super-resolution algorithm** | |
| Iterations | 180 |
| timestep $\tau$ | 0.01 |
| weighting $\lambda$ | 0.5 |
| upscaling factor $s$ | 2 or 3 |
| temporal radius $r_t$ | $s^2$ |

The Farnebäck method [19] uses polynomial expansion for motion estimation. To be more robust against errors in the displacement estimation, a scale pyramid is used. As is shown in Tab. III, a four level pyramid, in which each level is a subsampling by the scaling factor $0.5$ is used. The estimation on sub-sampled images provides a sensible initial estimation for the displacement of pixels on the higher resolution levels [19], [20].

On each level the displacement estimation is finetuned in 3 iterations each. The pixel neighborhood of each polynomial is set to 7. The Gaussian used for the polynomial expansion is set to a size of $15 \times 15$, with standard deviation of $\sigma = 1.5$. These parameter combinations are recommended based on experimental results [20].

To get a higher resolution image, the approach described in [16] was used. The method uses image warping to overlay adjacent video frames on a common higher resolution grid. The algorithm works iteratively. The number of iterations was set to $180$ as suggested by [20]. The timestep $\tau$, which also serves as a scaling factor for the steepest decent method was set to $0.01$ as suggested in [16], [20]. The smoothing of the high resolution image is determined by the weighting parameter $\lambda$, which is set to $0.5$, according to the recommendation in [20]. The images were scaled by the factor $s$. In the experiments we did not process the whole video frames, but regions of interest, cropped around a detected asset. Depending on the size of the cropped ROI it was possible to scale by a factor of $s = 3$ at max, because of the limited memory of the GPU (*NVidia GeForce GTX 1050 ti*) used in the experiments. With larger ROIs, only a scaling factor of $s = 2$ could be used.

It is obvious that if significantly longer processing times are tolerable, swapping data from graphics card to RAM or hard disk will allow even much higher scaling factors. As shown later that the achieved image resolution increase is less than doubling (see Fig. 10), swapping data is not implemented at this stage of testing. Scaling factors of 3 or higher would be only effective for image datasets providing enough salient information at sub-pixel level.

The temporal radius $r_t$ of adjacent frames used in the computation of one higher resolution frame was chosen to be exactly $r_t = s^2$ to avoid problems introduced by an under-determined problem [17]. As a cost function, also for the super resolution method a $L_1$ loss was used [16], [17].

The increased level of detail is visible in Fig. 7. In the detailed view on the right hand side the spikes mounted on top of the insulator's surface are more distinguishable from noisy artifacts in the image. In addition to that, the state of the surface can be assessed based on a smoother representation. The effect of the image scaling on the resolution of the pixel grid is especially visible in the enlarged detailed view in the lower part of Fig. 7.

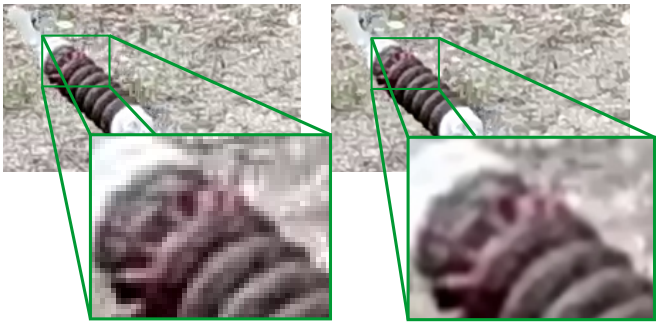

Fig. 7. A detailed view for comparison of an original shot of an insulator and the super-resolution result. In the left image the original image pixels are easily visible to the naked eye. Image data from DJI Mavic video.

### C. Object Detection with Deep Learning

Nowadays, in the computer vision community state of the art approaches for generic object detection are based on deep-learning technology. These methods show the best performance on object detection benchmarks like *COCO* and *Pascal VOC*. An overview on performances achieved by state-of-the-art methods are given in [21]. In order to solve our asset detection challenge we apply a Single Shot Detector (SSD)[22] due to low time and memory consumptions at high detection rates. During runtime the trained models are applied to the frames (video frames or still images) to reliably identify the location of assets. Having available the locations of the assets/objects in the images (i.e. axis-aligned bounding box), camera pose information may be used to estimate the exact position of the asset in the 3D world by using e.g. Structure from Motion [23].

### D. Geo-Localization of Detected Assets

Our flight management aims at collecting redundant image material so that relevant assets or objects are mapped from different viewpoints. Capturing assets from multiple viewpoints enables a more robust analysis, but also reduces the environmental influences into the analytic steps. In addition, multi-view information can be used to compute geo-locations of the assets in a 3D world. We assume availability of the GPS information for all captured photos. Having overlapping image material a 3D reconstruction can be computed using standard Structure from Motion workflows like the one proposed in [23]. The GPS signal is augmented to the

local 3D reconstruction using a robust registration procedure. The geo-referenced space is traversed using a static grid with a constant spacing adapted to the corresponding asset dimension. In our case we use a spacing of around 100 mm. Each 3D grid point is projected into the 2D photo using the recovered image pose information. If the back-projection intersects a 2D detection rectangle in two or more images, we assume that we have estimated the geo-location of an asset in 3D space. The established correspondences between 3D point and 2D detection are used for an improved user guidance within a web application e.g. for each asset multi-view photo information can be presented to an operator in combination of analysis results. Qualitative experimentation shows feasibility, however the GPS accuracy has major influence on localization accuracy of the assets.

## VI. Flight Test Data Evaluation

### A. Still Images

In order to detect railway assets like insulators in high resolution images, we apply SSD [22]. The SSD detector is well-known for its high computation speed. In some comparisons by researchers, it outperforms the other state of the art detectors which are based on R-CNN family of algorithms for larger size of objects [22]. The bounding boxes of insulators, which are clearly visible, in high resolution images have a minimum area of 20,000 pixels, which gives good reason for choosing SSD.

The training of the detector using original high resolution images is conducted by using a standardized deep learning based training and evaluation workflow system. This system provides a *Tensorflow API* for object detection and COCO evaluation after a training session. Two SSD models (ssd_mobilenet_v1_fpn_coco and ssd_mobilenet_v1_coco [22], [21], [24]) are chosen for training to achieve the best accuracy with the available original data. They are composed of 395 high resolution images containing totally 1062 annotated insulators and 752 high resolution images showing relevant scenes without assets. In order to handle the high resolution photos efficiently, we apply an overlapping sliding window approach to detect the assets.

To increase the variety of training data, we heavily use data augmentation techniques such as geometric flipping, rotation and augmentation of various background scenes. Therefore, we segmented representative samples of insulators and augmented random backgrounds. Thus the training dataset contains synthetic images and real images. About 44,000 images are finally generated, which are images rendered with two segmented sample images and images from COCO dataset as background. Before rendering, the sample images are modified with geometrical transformation for rotation, resizing and translation. In addition, randomly selected rendered images are augmented with Gaussian filter for blurring and point transform for adjusting brightness. This synthetic data is capable for learning to identify the body structure of insulators when there is no strong perspective distortion, as shown in Fig. 8. The real data is cropped images in size of 600 x 600 pixels from the high resolution images captured with drones.

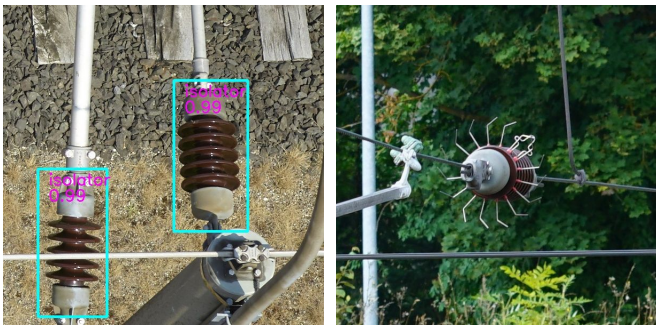


Fig. 8. Result images using the model trained with only synthetic augmented data. It detects insulators with high confidence when they are clearly visible and with small perspective angles. It does not detect the insulator shown in the right image, even by setting a conservative confidence threshold 0.6. The input images are recorded with the Copting drone.

Besides the above two methods as the first level augmentation, random flipping and SSD random cropping from the *Tensorflow API* are applied to each of base data in order to reach higher performance. For training, one type of insulators 6000 cropped images from high resolution images is used. The numeric evaluation metric used is the COCO metric for object detection. The test dataset is composed of 5000 cropped images in size of 600 x 600 pixels. Tab. IV summaries the evaluation results. In our case we set conservative thresholds to ensure that the insulators are detected. Thus, we can accept a higher rate of false positive detections as we use redundancy based on overlapping still images to verify consistency of the detection in various view points. The following Fig. 9 illustrates two detection results for overlapping test images.

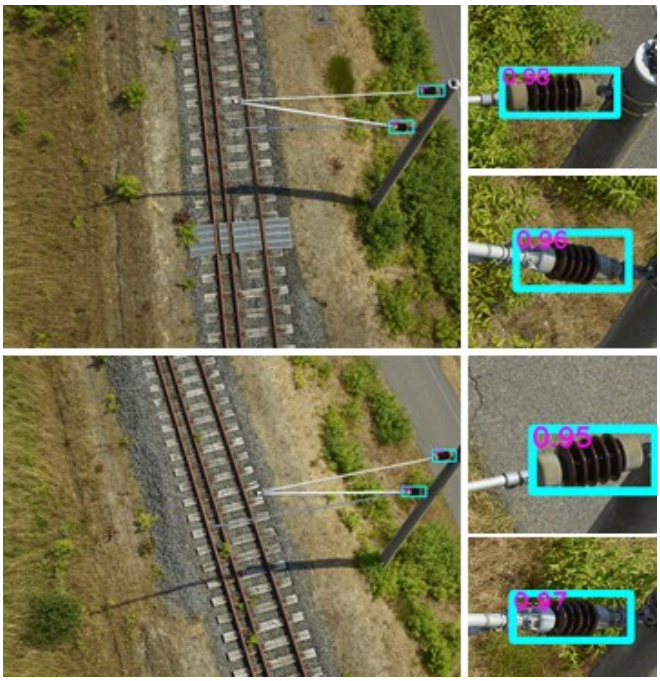


Fig. 9. Detection results achieved for overlapping test images (Copting drone). Insulators are detected from different viewpoints. Together with photo pose information multi-view detections are used to estimate the position of the objects in 3D space

TABLE IV: PRECISION AND RECALL FOR TEST DATA

| Number of images | Image type | mAP0.5 | Average recall |
|---|---|---|---|
| 5000 | cropped images | 0.98 | 0.72 |
| 243 | full resolution | 0.968 | 0.7 |

### B. Compressed Video Data

Eventually, object detection at images from super-resolution will work in the same way as with the still images. Differences in detection rate and quality are due to image resolution. With that, we will analyze whether a comparable resolution is achieved with the super-resolution strategy.

At preliminary flight experiments Siemens stars were added to a prop with an insulator to enable the quantification of the resolution of still images recorded during the flight. In this sections, these records are used to quantify the increase of the resolution caused by the application of the presented super-resolution method.

As can be seen in Fig. 10 the resolution is strongly dependent on the direction in the image coordinate system. This dependency is caused by the skewed pose of the test pattern in relation to the image plane, as well as the motion of the recording vehicle in conjunction with artifacts induce by the applied video compression codec (*h264*). That is, why in Fig. 10 ellipses instead of circles indicate the blurred center region of the test patterns.

To compare the difference in the resolution before and after the super-resolution step, the semi-axes of the marker ellipses around the center of the Siemens star are measured approximately. The values given in Fig. 10 are in pixel units in the common grid, scaled to be equal for both images.

While the length of the semi-major was decreased by a factor of approximately 0.6, the length of the semi-minor stayed roughly constant. This uneven increase in resolution is explained by the direction of the vehicle's motion, which moves in parallel to the ellipses semi-major. This motion leads to a direction-dependent increase in resolution.

Although the size of the blurred area on the test pattern was decreased, there are still periodical artifacts in the region where the marker ellipse was present in the original recording.

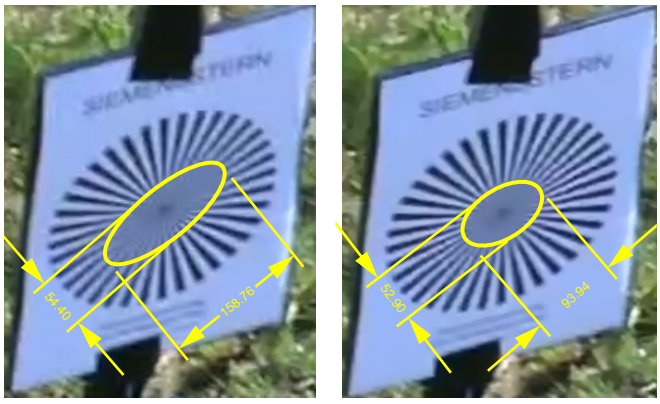


Fig. 10. Images taken of a Siemens star, to make the achieved higher resolution more obvious. In addition to the higher resolution, the processed image shows less artifacts introduced through the motion of the recording vehicle.

### C. Comparison of the Results

By a simple comparison of the resolutions described in Tab. I and Tab. II it is apparent that a super-resolution with scaling factors 2 and 3 can theoretically produce a new image with sufficient pixel resolution comparable to the high quality still images.

The actual resolution of structures on the assets of interest also depends on the focal length of the cameras or respectively the resulting field of view. The FOV of the consumer camera in Tab. II is about twice as large as the field of view (FOV) of the camera in Tab. I. This leads to a doubled angular resolution per pixel for the Sony camera. Taking this into account, the images created with video super-resolution still do not recreate the quality of the still images. However, resolution increase is present as shown in Fig. 10 which makes this approach promising if higher scales are applied.

Since the images produced using video based super-resolution are visually more prone to artifacts, further improvements are subject to improved input quality, e.g. uncompressed video and camera movements in all directions, and the capability to calculate higher resolution, e.g. with more memory and improved algorithms with lower system requirements.

## VII. CONCLUSION AND FUTURE WORK

In the context of predictive health management of railway infrastructure, this paper evaluates how camera drones can be used for automatic monitoring of catenary and other targets which are hard to reach otherwise. Detections are focused on rod insulators at the overhead lines, however, this process can be easily adapted to other assets. As an initial step, the discussed aspects on flight and data protection regulations with the result that legal issues are generally manageable under some constraints (especially risk mitigation). Core of the work is automatic identification in images using deep learning strategies. Based on flight tests with unmanned aircraft, it is shown that insulator detection is successful if a special vehicle with high-end camera system is used. To enable this strategy also for cheaper mass-market systems, the paper investigates how typical video data with lower resolution can be also used for that task. At current stage, super-resolution of successive video frames comes with some drawbacks, especially the requirement of specific movements in all visual directions and a good video input quality to avoid artifacts. Nonetheless, the practical consequence can be the suitability of rather cheap hardware for inspection tasks where raw video resolution drawbacks are compensated by software and adjusted flight profiles designed to produce optimal super-resolution results.

Future work will further investigate into smaller structures (e.g. defects on the insulators) and improvements on super-resolution e.g. with uncompressed video and improved algorithms which can manage higher up-scaling factors. To improve the overall asset detection performance we may further exploit detections generated in multi-view photo material, in particular this is required for defect classification of assets. Beyond that, improvements will include safe and

precise flight control in order hit the assets with the camera also at closer distances.

## REFERENCES


[1] J. Zink and B. Lovelace, "Unmanned aerial vehicle bridge inspection demonstration project," Minnesota Department of Transportation, Tech. Rep. MN/RC 2015-40, 2015. [Online]. Available: https://www.lrrb.org/pdf/201540.pdf

[2] S. Dorafshan, et al., "Unmanned aerial vehicle augmented bridge inspection feasibility study," Department of Civil and Environmental Engineering, Utah State Univ., Tech. Rep. CAIT-UTC-NC31, 2017.

[3] C. Stout and D. Thompson, "UAV approaches to wind turbine inspection," Offshore Renewable Energy Catapult, Tech. Rep., 2019. [Online]. Available: https://www.arpas.uk/wp-content/uploads/2019/04/Cyberhawks-Approach-to-UAV-Inspection-Craig-Stout-ORE-Catapult.pdf

[4] M. Hoffmann, T. Dolleschal, and A. Strauss, "RISKMON: Anlageninspektion und Risk-Monitoring mit Hochleistungsdrohnen (UAS/UAV) (in German)," Austrian Federal Ministry for Transport, Innovation and Technology, Tech. Rep., 2019.

[5] A. K. Singh, A. Swarup, A. Agarwal, and D. Singh, "Vision based rail track extraction and monitoring through drone imagery," *ICT Express*, vol. 5, no. 2, pp. 250–255, 2019.

[6] Working Group on the Trans-Asian Railway Network , "Inspection and monitoring of railway infrastructure using aerial drones," Economic and Social Commission for Asia and the Pacific, Tech. Rep. ESCAP/TARN/WG/2019/4, 2019. [Online]. Available: https://www.unescap.org/sites/default/files/TARWG_4E_Inspection%20and%20monitoring.pdf

[7] European Aviation Safety Agency, "Introduction of a regulatory frameworkfor the operation of drones," A-NPA 2015-10, 2015.

[8] Regulation (EU) 2016/679 of the European Parliament and of the Council, "General data protection regulation," 2016. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2016/679/oj

[9] L. Murzilli, *JARUS guidelines on Specific Operations Risk Assessment (SORA) 2.0*. JAR-DEL-WG6-D.04, 2019. [Online]. Available: http://jarus-rpas.org/sites/jarus-rpas.org/files/jar_doc_06_jarus_sora_v2.0.pdf

[10] Joint Authorities for Rulemaking on Unmanned Systems (JARUS), "Jar doc 06 sora (package), executive summary," http://jarus-rpas.org/content/jar-doc-06-sora-package, 2019.

[11] G. Faur, et al., "JARUS guidelines on SORA: Standard scenario for aerial work operations," JARUS-STS-01, 2019. [Online]. Available: http://jarus-rpas.org/sites/jarus-rpas.org/files/jar_doc_6_sora_sts_01_edition1.1.pdf

[12] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European Conf. on Computer Vision*, 2016.

[13] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1637–1645.

[14] C. Ledig, L. Theis, F. Huszar, et al., "Photo-realistic single image super-resolution using a generative adversarial network," *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[15] W. Shi, J. Caballero, F. Huszár, et al., "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1874–1883.

[16] D. Mitzel, T. Pock, T. Schoenemann, and D. Cremers, "Video super resolution using duality based tv-l 1 optical flow," in *DAGM 2009 Joint Pattern Recognition Symposium*, 2009, pp. 432–441.

[17] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust super-resolution," *International Conf. on Image Processing (Cat. No.03CH37429)*, vol. 2, pp. II–291, 2003.

[18] C. Zach, T. Pock, and H. Bischof, "A duality based approach for realtime tv-l1 optical flow," in *Proceedings of the 29th DAGM Conf. on Pattern Recognition*. Springer, 2007, pp. 214–223.

[19] G. Farnebäck, "Two-frame motion estimation based on polynomial expansion," in *Image Analysis*, J. Bigun and T. Gustavsson, Eds. Springer, 2003, pp. 363–370.

[20] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[21] J. Huang, V. Rathod, C. Sun, et al., "Speed/accuracy trade-offs for modern convolutional object detectors," *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[22] W. Liu, D. Anguelov, D. Erhan, et al., "SSD: Single shot multibox detector," *Lecture Notes in Computer Science*, p. 21–37, 2016. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-46448-0_2

[23] S. Agarwal, N. Snavely, I. Simon, et al., "Building rome in a day," in *Intl. Conf. on Computer Vision (ICCV)*, pp. 72–79, 2009.

[24] A. G. Howard, M. Zhu, B. Chen, et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," *arXiv:1704.04861 [cs.CV]*, 2017.