DEVELOPMENT OF A MACHINE VISION SYSTEM FOR DAMAGE AND OBJECT DETECTION IN TUNNELS USING CONVOLUTIONAL NEURAL NETWORKS

F. Alidoost [1,2] *, M. Hahn[1], G. Austen[1]

[1] Faculty of Geomatics, Computer Science and Mathematics, University of Applied Sciences (HfT) Stuttgart, Germany – (fatemeh.alidoost, michael.hahn, gerrit.austen)@hft-stuttgart.de
[2] vigram GmbH, Freiburg, Germany – fatemeh.alidoost@vigram.com

KEY WORDS: Inspection, 3D Models, Photogrammetry, Crack Segmentation, Deep Learning, Convolutional Neural Networks.

ABSTRACT:

Tunnel inspection, i.e. detection of damages and defects on concrete surfaces, is essential for monitoring structural reliability and health conditions of transport facilities, thus providing safe and sustainable urban transportation infrastructures. In this study, an innovative visual-based system is developed for damage and object detection tasks in roadway tunnels based on deep learning techniques. The main components of the developed Machine Vision System such as industrial cameras, flash-based light sources, controller, the synchronization unit and corresponding software programs are designed to collect high-resolution images with sufficient quality from dimly lit tunnel environments in normal traffic flows with an operating speed of 30-50 km/h. Unlike recent studies, the training data includes multiple types of damage such as cracks, spalling, rust, delamination and other surface changes. Furthermore, 10 classes of common tunnel objects including traffic signs, traffic cameras, traffic lights, ventilation ducts, various sensors and cables are labeled for object detection. As state-of-the-art Convolutional Neural Networks, DeepLab and U-Net are trained and evaluated using accuracy metrics for image segmentation. The results highlight the most important parameters of the discussed Machine Vision System as well as the performance of DeepLab and U-Net for object and damage detection.

1. INTRODUCTION

Underground transportation structures such as highways and railway tunnels are needed to alleviate and optimize traffic on certain routes. They provide a direct and uninterrupted connection between two locations without disrupting traffic on the surface (e.g. France and Great Britain tunnel). This can be particularly beneficial in densely populated areas and cities with high traffic volumes. Aggressive environmental conditions, general aging due to material wear and fatigue and the traffic volume of heavy vehicles (e.g. buses, trucks and trailers) influence the reliability and durability of these civil structures, leading to a loss of stability, safety and functionality for users.

The operation and maintenance of large tunnel structures are challenging needs for urban managers and planners. In most industrialized countries tunnel inspection and inventory are regulated by construction laws and must be carried out at regular intervals. Tunnel inspection involves monitoring and mapping the tunnel surfaces to detect different types of anomalies and damages on the concrete or even steel such as cracking, distortion, spalling and exposed reinforcement. In addition, tunnel elements such as technical equipment are localized to update the inventories for Building Information Modeling (BIM) applications. Conventional and routine manual inspections of tunnels are conducted by trained engineers at regular intervals to visually evaluate damages and defects. Apart from the challenging work conditions, this procedure is extremely time-consuming and costly and usually requires complete or partial tunnel blocking for installing special measuring equipment or manual data collection by trained inspectors. Moreover, the accuracy of the evaluations depends heavily on the person's experience and qualifications and looking for areas, that are difficult to access, such as ventilation ducts, axial fans and ceiling areas, needs special equipment.

On the other hand, significant advances in high-speed imaging technologies offer new solutions for automating the inspection and mapping process, particularly for highway tunnels with low illumination conditions and high-speed traffic flows. To overcome the shortcomings of visual inspections and to improve the speed and automation level of tunnel inspection, techniques of Artificial Intelligence (AI) and Machine Learning (ML), and more recently Deep Learning (DL), can be used to automatically detect damages or objects. Deep Convolutional Neural Networks (DCNNs) have shown promising results in image segmentation and pattern recognition. A Convolutional Neural Network (CNN) is based on hierarchical learning representations of data using a deep structure composed of different hidden layers. The learning strategy requires a sufficient amount of training data as well as a large amount of memory, and the network optimization procedure takes a long time to achieve acceptable accuracy. So far, several CNN-based methods have been developed for damage detection on road asphalts (Liu et al. 2019), bridges (Qiao et al. 2021), and concrete (Kumar et al. 2021; Rajadurai and Kang, 2021); however, they mostly focus on cracks. Therefore, their training data have been prepared for binary classification to detect cracks and non-cracked structures. However, damages and defects in real-world environments like tunnels or bridges are not only made of cracks but also of spalling, rust, delamination and efflorescence on surfaces, making a multi-class damage segmentation desirable for practical applications. Furthermore, the preparation of proper training data including sufficient annotations (e.g. labels) for different classes of interest is highly challenging and needs way too much manual labeling effort. By selecting the right CNN and customizing it to be able to learn with fewer data, as well as employing efficient data augmentation techniques, it is possible to obtain adequate accuracy for tunnel inspection tasks.

* Corresponding author

In this study, we present a cost-effective system for capturing images suitable for detailed visual inspection of tunnels. The main differences between the proposed system and comparable technologies are high-quality images in color mode without motion blur and extraneous light which allows the detection of defects on the tunnel surface. Additionally, more than 10 km/h operation speed allows data acquisition in normal traffic situations in tunnels, at least outside rush hours. The main advantage of the proposed system is the accurate synchronization between machine vision cameras and flash light which provides sufficiently-illuminated images in the dark tunnels. Two state-of-the-art CNNs (DeepLab and U-Net) are explored for image segmentation under challenging tunnel conditions such as low illumination, motion and blur effects and noises. The main contribution of this work is to provide multi-damage and multi-object recognition CNNs that are best suited to detect various types of damages and objects from images, captured in different tunnel environments. The proposed configurations for selected architectures improve the performance and accuracy of the CNNs in detecting several forms of damages due to intra-class variability (e.g. different shapes of traffic signs) and inter-class similarity (e.g. spalling and delamination classes). Moreover, the empirical experiments and challenges for data preparation and training of models using different hyperparameters are described.

2. RELATED WORK

To date, various techniques and methods have been developed for monitoring damages in different types of tunnels (e.g. metro, railway, road, etc.) using various data collection technologies such as total stations, cameras, and laser scanners. The latter can be categorized into conventional, laser, photogrammetric, and hybrid techniques.

2.1 Conventional Techniques

Conventional methods of data collection, such as visual surveys in tunnels are still operated manually by experts using total stations (Luo et al. 2016). To improve the automation of data collection and measurements, different systems have been developed including multi-cameras systems and laser scanners.

2.2 Laser Scanners

Sun et al. (2020) developed a scanning system on a robot vehicle to monitor deformations in railway tunnels. The maximum operating speed of the system is 4.5 km/h and the dislocation measurement accuracy is about 3 mm. Guo et al. (2020) utilized terrestrial laser scanners to monitor deformations in tunnels with a 1-2 mm accuracy. Zhou et al. (2017) developed a mobile laser scanner to monitor rail-based tunnels. However, the GNSS-denied environments of tunnels limit the use of laser scanners for deep tunnels and the accuracy and robustness are degraded for tunnels longer than 100 m. On the other hand, damages with no positional deformations or displacements such as small cracks and spalling might not be visible in laser-based point clouds.

2.3 Photogrammetric Techniques

In photogrammetric inspection techniques, more than one camera is usually employed to capture the tunnel surface. There are a few studies that developed camera-based systems for roadway or railway tunnels. Jiang et al. (2019) developed a system with 7 line-scan cameras and 60 LEDs on a car for crack detection in tunnels. The maximum speed of the vehicle is 100 km/h. Chapman et al. (2016) developed a mobile mapping system equipped with 16 cameras and 34 light sources to capture images from GNSS-denied environments. The average positioning error of this system is about 0.34 m. Zhan et al. (2015) developed a multi-camera system including 7 line-scan cameras and structured-light projectors to measure railway tunnels. Panella et al. (2020) compared the usability of the Go-Pro cameras and terrestrial laser scanning for tunnel inspection. Their assessments show that photogrammetry is a valid alternative to laser scanning for the visual inspection of tunnels. However, the final accuracy of photogrammetric inspection techniques extremely depends on the camera resolution, the vehicle speed and the illumination conditions.

2.4 Hybrid Techniques

To increase the performance of automatic inspection techniques, multiple sensors (such as cameras and lasers) are synchronized to capture data from tunnel surfaces. The company Dibit (Mett et al. 2019) developed a hybrid system including multi-cameras and laser scanners for monitoring roadway tunnels. The maximum measurement speed of Dibit's systems is about 80 km/h. Menendez et al. (2018) developed an autonomous robotic arm to carry several sensors including a laser scanner and two cameras. The maximum error of the system is 110 mm.
In hybrid inspection techniques, the platform can be a train or a rail-based automatic robot for railway tunnels that can move on the rail to collect the data. In this case, the system needs to be continuously and carefully monitored by an operator and the speed of the movement is less than 10 km/h. Despite being time-consuming and cumbersome, these approaches require quite expensive equipment for robot movements on the railway.

2.5 AI-Image Segmentation for Tunnels

There is extensive literature on pattern recognition based on DL in the field of computer vision and/or photogrammetry. Various applications such as building extraction, traffic signs detection and land cover classification are addressed. Recently, several image-based inspection methods using CNNs have been developed for damage detection on concrete surfaces such as roads, bridges and tunnels. For instance, Qiao et al. (2021) developed a CNN with an Expected Maximum Attention (EMA) module for the bridge damage extraction. Rajadurai et al. (2021) trained Alex-Net to classify images into two classes of cracks and no-cracks with a prediction accuracy of 99.9%. Kumar et al. (2021) utilized a Mask R-CNN to detect multi-classes of cracks in different orientations on the concrete. A similar method has been developed by Kim et al. (2020) to optimize Mask R-CNN for detecting cracks, efflorescence, rebar exposure, and spalling with a precision of 87.24%. Shin et al. (2020) developed a CNN involving multi-attention-based modules to detect different types of concrete damages such as cracks, rebar exposure and delamination with an accuracy of 98.9%. Li et al. (2020) developed a new version of U-Net (U-CliqueNet) for binary crack classification from tunnel images with an average IoU of 86.96%.
CNN models require several modifications and improvements to achieve acceptable accuracies in damage detection. Compared to a multitude of computer vision objects, damages are structured in totally random shapes and various patterns which demands a robust CNN architecture to detect them efficiently. Furthermore, the focus of the past studies is on crack detection on concrete surfaces and not on multi-class damages or objects in tunnels with different materials (e.g. concrete, asphalt, and metal). This is due to various challenges and difficulties in collecting proper datasets to train the CNN.

3. PROPOSED METHOD

The main purpose of this study is to provide an automatic solution to monitor and detect damages and objects based on RGB images, captured by a Machine Vision System (MVS). As shown in Figure 1, our MVS utilizes high-speed 5 MP machine vision cameras, high-performance flash lights (flash LEDs), and a synchronization unit installed on a car. The MVS is able to capture images with a very short exposure time, so that the problematic effects such as motion blur, extraneous light and stray light are minimized. The captured images are then employed to train the CNN and evaluate the capability of the trained network for automatic damage detection in road tunnels. The following subsections provide a summary of the individual steps and main components.

3.1 Imaging Sensor

The performance of the cameras is vital for collecting high-quality data from the tunnel surfaces. According to our investigation of the digital sensor markets (Alidoost et al. 2022), machine vision cameras are much more suitable for this type of application, than consumer DSLR cameras. Therefore, we selected the Grasshopper camera, offered by FLIR Company. The sensor is a 5 MP area-based camera with CMOS technology and a global shutter. Compared to the line-based sensors, the advantage of area-based sensors is no linear effects and no distortions when combining lines into area images, less power illuminator, and the interface and setup are standardized. To reduce the motion blur in images that occur in the tunnel at high speed, the global shutter is a better choice than a rolling shutter. With a global shutter camera, the scene will be frozen at a certain point in time and there is no motion blur.
The Field of View (***FOV***) of the camera follows simple geometric considerations according to Equations 1 and 2.

$$FOV_w\ [deg] = 2 \times arctan\left(\frac{w[mm]}{2 \times f[mm]}\right) \times 180\ [deg]/\pi \quad (1)$$

$$FOV_h\ [deg] = 2 \times arctan\left(\frac{h[mm]}{2 \times f[mm]}\right) \times 180\ [deg]/\pi \quad (2)$$

The image size is given by ***w*** and ***h*** in mm and the parameter ***f*** is the camera's focal length. The ***FOV*** must be taken into account when arranging the cameras on the vehicle so that the tunnel surface is imaged seamlessly.
To transfer the data from the sensor buffer memory to the storage unit in real-time during operation, USB 3.0 is used for the camera's interface. Among the different interfaces, USB 3.0 is the fastest interface (e.g. with a maximum bandwidth of about 400 MB/s for USB3.0, compared to 100 MB/s for the GigE interface) and also with the easiest setup. The required bandwidth (***BW***) is calculated based on the required frame rate (***FPS***), the number of cameras (***N***) and the pixel format (8-bit for the mono mode and 24-bit for the color mode), given by Equation 3.

$$BW\ [\mathrm{Mb/s}] = FPS\ [Hz] \times W\ [px] \times H[px] \times BPP \times \mathrm{N} \times 10^{-6} \quad (3)$$

where ***W*** and ***H*** are the image size in pixels, and the ***BPP*** is Bytes per pixel which is 1 for an 8-bit image (in the mono mode) and 3 for a 24-bit image (in the color mode). The bandwidth (***BW***) must also be taken into account for the read/write speed of the hard drive.

3.2 Illumination

In tunnels, it is often dark and there is not enough light to capture bright images. Therefore, lighting plays a major role in industrial machine vision applications. The most widely used lighting source is LED light which offers high performance, stability, high intensity, as well as cost-efficiency. Flashing (or pulsing) a LED light is a powerful technique that can be beneficial for machine vision systems as it increases the light power for larger distances, extends the lifespan of the LED, as well as solves the problem of ambient light. In this study, a meta-bright area-based LED offered by Metaphase Company is used. This spotlight is extremely bright and emits white light at 600,000 Lux. The advantage of white/visible light (e.g. 380-780 nm), compared to the IR wavelengths (e.g. 850-940 nm), is to capture color images which are important for applications such as automatic object recognition and scene understanding. Since cameras and LEDs are directed toward the tunnel walls and ceilings, therefore there is no danger to the human eye, and the visible spectrum is not confusing or dangerous for road users or operators, unlike opinions on the light market.
The working distance of the employed LEDs is 3-8 m, which is suitable for large-area illumination applications like tunnels. With the LED controller, the light power can be controlled by changing the current, the pulse width and the light mode such as flashing or constant light. The LED can provide more than 50 kHz flash strobe with a small pulse width in the range of 20 to 60,000 µs. The optimum pulse width corresponds to the required camera's exposure time (***ET***) to avoid motion blur (***B***) in driving with a speed of ***V***, given by Equation 4.

Figure 1. Main components of the MVS. a: machine vision camera, b: flash LEDs, c: self-made synchronization unit, and d: setup on our van.

$$ET\ [s] = \frac{(D[m] \times B\ [px] \times PS\ [mm])/f[mm]}{V[m/s]} \quad (4)$$

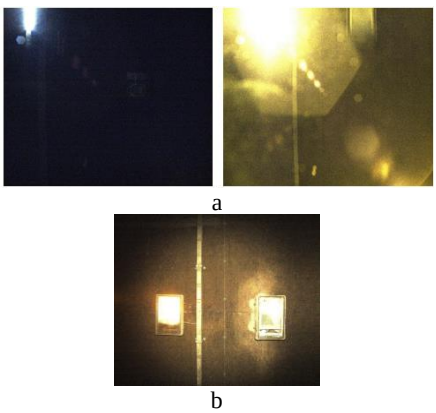
where ***D*** is the working distance and ***PS*** is the CCD's size. The parameter f is the camera's focal length.
Depending on the focusing distance and illuminated area, lenses of different sizes can be used to efficiently project light and to perfectly illuminate rectangular areas. The larger lens generates less light and power. The relation between the working distance (***R***) and LED irradiance (***E***) is given in Equation 5. In this study, we employed lenses with 30° and 50° for larger and smaller ranges, respectively.

$$E\ [W/m^2] = E_o[W/m^2] \times (\frac{R_o[m]}{R[m]})^2 \quad (5)$$

where ***Eo*** is the irradiance of the LED at the distance of ***Ro***, provided by the manufacturer.

3.3 Synchronization Unit

To avoid motion blur as well as other illumination artefacts in the images, the flash should be triggered exactly when the camera's aperture is opened to capture the scene. Therefore, the flash time and the exposure time need to be synchronized precisely. Otherwise, the images are either dark because of the lack of light or too bright with motion blur and ambient light, as shown in Figure 2, a. In this study, we used a synchronization method based on the General Purpose Input/Output (GPIO) connector in the camera. In this method, one camera is considered as a "primary" camera which is used to trigger one or more "secondary" cameras, using the primary camera's strobe. This ensures that the frame rates of the secondary cameras are the same as the primary camera's frame rate. Additionally, the strobe of the primary camera is also connected to the LED controller to trigger the LEDs simultaneously.
In this study, we developed a hand-made synchronization unit to connect the output pin of the "primary" camera to the input pins of "secondary" cameras as well as the input pins of the LEDs. The result of the successful synchronization is clearly visible in Figure 2, b. The synchronization parameters (the frame rate and exposure time of the "primary" camera) have been calculated by Equations 3 and 4.



a

b

Figure 2. Differences between non-synchronized (a) and synchronized (b) cameras and flashing lights.

3.4 Data Capturing Strategy

Figure 3 shows a schematic diagram that illustrates the modular design of the proposed system. The cameras, LEDs, synchronization unit, power source, workstation and storage unit are all integrated into a single lightweight capture unit which is installed on a portable platform. There is a simple interface between the main MVS components that allows for easy upgrades or independent modifications. The modular design allows for the addition or removal of LEDs and cameras to the MVS.
In this study, we employed three cameras and three LEDs to build a low-cost mobile system for mapping roadway tunnels with different shapes like circular or rectangular tunnels. The cameras are connected to a consumer-grade laptop via USB 3.0 cables and to the synchronization unit via GPIO cables. The LEDs are connected to the LED controller using the supported cable. The LED controller can be controlled by the operator using a keypad or an Ethernet connection. The whole system is supported by DC power supplies.

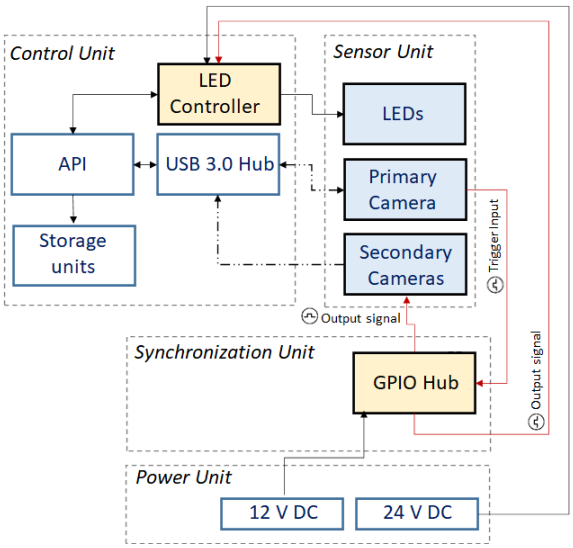


Figure 3. A schematic diagram of the proposed MVS.

According to the viewing directions and working distances for the tunnel sections as well as sufficient overlap in the vertical direction for a second capture in the reverse direction of travel, three cameras are installed with a Relative Angle (R.A.) of 45° to cover the ceiling, the corner and the wall in a one-way trip, as illustrated in Figure 4. In an outbound and return scenario, the tunnel surface is fully captured. The overlap between two consecutive images along the moving direction (we call it the horizontal overlap) and the overlap between two adjacent cameras (we call it the vertical overlap) are determined by Equations 6 and 7.

$$VerticalOverlap\ [\%] = \left(1 - \frac{\alpha[deg]}{FOV_w[deg]}\right) \times 100 \quad (6)$$

$$HorizontalOverlap\ [\%] = \left(1 - \frac{V[m/s] \times f[mm]}{FPS[Hz] \times h[mm] \times D[m]}\right) \times 100 \quad (7)$$

where ***V*** is the speed of the vehicle and ***D*** is the working distance.
The capturing scenario is controlled by an open-source API, developed by Teledyne FLIR (formerly FLIR Company) for machine vision cameras. By proper settings for the start and end times of the capturing, the frame rate, image format, and illumination gain for the primary and secondary cameras in the API, data automation can be achieved on every drive.

Therefore, only one driver and one operator are necessary to go to the field for data acquisition which reduces the overall cost.

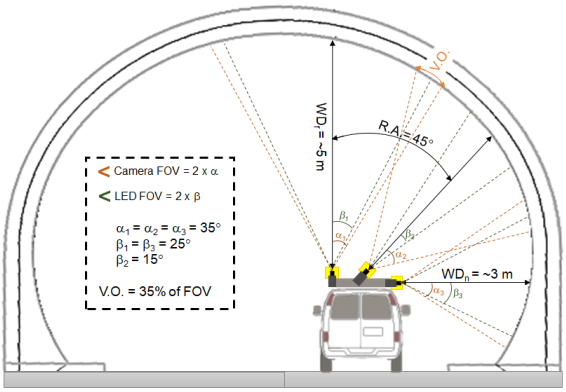


Figure 4. The field of view of cameras in the MVS.

4. IMPLEMENTATION

In order to evaluate the performance of the proposed system, the Wagenburg tunnel in Stuttgart, Germany is selected which is an 824 m long single-tube two-way traffic tunnel with a maximum driving speed of 40 km/h. This paper investigates two widely-used CNNs in image segmentation for multi-object/damage detection tasks from tunnel images. The U-Net (Ronneberger et al. 2015) is a fully convolutional neural network consisting of a contracting path to capture context and a symmetrically expanding path that enables precise localization. The contracting path reduces the spatial resolution of the input images and combines the high-resolution features with the highly localized patterns to enable rapid and precise segmentation of images. The expanding path restores the spatial resolution of the output segmentation map, while also preserving spatial context. U-Net can train effectively on small datasets and has outperformed many other image segmentation methods. The second CNN is DeepLab-v3 (Chen et al. 2017) developed by the Google AI team. DeepLab consists of up-sampling layers (e.g. atrous), instead of max-pooling layers, as well as densely connected conditional random fields (CRF). These features help the CNN to increase the spatial resolution of extracted features and improve the fine details in outputs.
The object detection task is less challenging than the damage detection task as object attributes such as geometric shape, color, and texture are more stable due to changes in the image capturing conditions in tunnels.

4.1 Proposed MVS Configurations

Figure 1, d shows the MVS system, which is portable and adaptable to various types of vehicles and can be mounted manually on the roof of a car without special equipment. The system components are installed on an aluminium carrier system, which is modular and allows adding or removing components as well as changing the directional views depending on the project requirements. Unlike steel platforms, the aluminium construction is lightweight and also possesses the necessary stability to driving vibrations. The main settings of the proposed system are summarized in Table 1.
During the test, the driving speed is set to 36 km/h and the average working distance is about 5 m. A 6 mm lens is used for the cameras, and the exposure time of approximately 250 µs limits the motion blur to less than 1 pixel. The required bandwidth to capture 5 Hz for color images or 15 Hz for grayscale images is about 250 MB/s which fits the USB 3.0 interface. However, apart from the camera interface, the final data rate depends on the USB host controller card and the image storing data rate of the hard drive (here an SSD drive). Ideally, each camera should have a USB 3.0 bus to have a full frame rate (e.g. 75 Hz for the mono mode and 25 Hz for the color mode). In this study, we used an ordinary laptop with a 500 GB SSD drive and one USB 3.0 port, and therefore sharing the bus with other cameras reduces the actual frame rate significantly to 2 Hz for color images and 4 Hz for grayscale images.

Settings	Parameters	Values
Camera	Sensor	CMOS, Area Scan
	Resolution	5 MP
	Pixel size	3.45 µm
	Sensor size	2448 x 2048 pixels
	Focal length (lens)	6 mm
	Shutter	Global
	Required frame rate	5 Hz and 15 Hz
	Interface	USB 3.0 (400 MB/s)
	Pixel Bit Depth	8 bits
	Chroma	color and mono
	Exposure Time	250 µs
	Price range	$1500
LED	LED shape	Area based
	Operation mode	Flash
	Wavelength	White
	Flash Pulse width	250 µs
	Flash rate	5 Hz and 10 Hz
	Irradiance at 2m	878 W/m2
	Price range	$1000
Car	The driving speed	36 km/h
	Working distance	5 m
	Installation angle	45°
Operations	Irradiance at 5m	140 W/m2
	FOVw x FOVh	70° x 60°
	Vertical overlap	~ 35 % of FOVw
	Horizontal overlap @ 2Hz	~ 23 % of FOVh
	Actual frame rate	2 Hz and 4 Hz
	Exposure Time	~ 250 µs

Table 1. The setup used for the project.

4.2 Training Data Preparation

Two training datasets are prepared separately for damage and object detection tasks. Both datasets contain RGB images captured by the MVS in the Wagenburg and Heslacher tunnels in Stuttgart, Germany. As shown in Figure 5, the lighting conditions in the tunnel differ significantly. The distances of the cameras to the wall, corner, and ceiling vary between 1 and 5 m, and LEDs with beam angles of 50° and 30° were used for illumination. The size of damages varies from 1 cm to 1-2 m. Table 2 summarizes how many images are used for training and testing the CNNs for damage detection and object detection. A total of 220 images are manually labelled for damages, of which 200 images are used for training and the rest for testing. A total of 522 images with depicted objects were selected, of which 473 images are used for training and the rest for testing.

Dataset	All	Training	Testing	Augmented data
Damages	220	200	20	23751
Objects	522	473	49	33583

Table 2. Prepared data for training the networks.

a b c

Figure 5. a: Camera 1 towards the roof, illuminated by a 50° LED; b: Camera 2 towards the corner, illuminated by a 30° LED; c: Camera 3 towards the wall, illuminated by a 50° LED.

To prepare the training dataset including RGB images and corresponding annotations, a labeling team manually digitized images for different classes of damages and objects. As a standard, it has been established that there are 8 classes of different damage types (crack, spalling, rust, rust flag, delamination, reinforcing steel, efflorescence and moss) and 10 classes of different objects (traffic signs, lights, sensors, axial fans, traffic lights, cables, alarm speakers, air vents, electronic box, and cable boxes), as illustrated in Figure 6.

Before training a CNN, additional processing steps should be applied to the images such as resizing and data augmentation. The input size or receptive field is an important aspect of the CNN architecture. In this study, all images are resized to 256×256 pixels preserving the global information of the object shapes and avoiding that cropping could affect the random patterns of damages. In addition, various data augmentation techniques such as scaling, rotating, flipping, and color-changing are applied to images to increase the number of training samples.

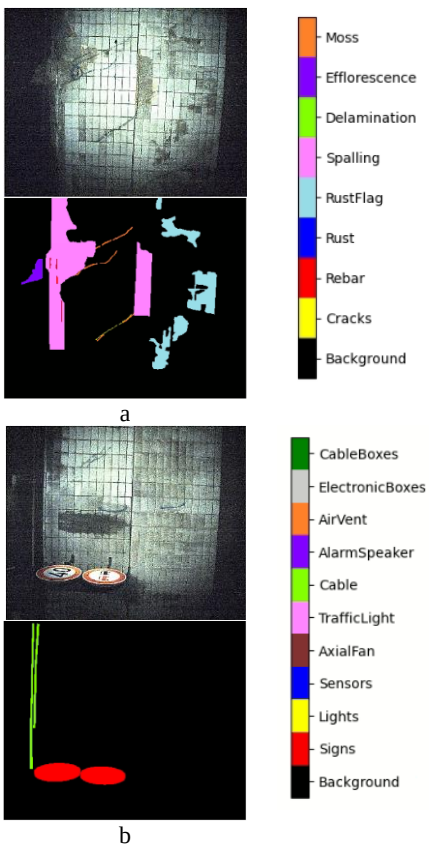


Figure 6. Sample images and corresponding labels for a: damages and b: objects in different datasets.

5. RESULTS AND DISCUSSIONS

The capturing of images in the tunnel with the MVS was designed in such a way that point clouds could also be generated from them. In the chosen setup, the images overlap by approximately 23% in the direction of travel (horizontal overlap) and slightly more at around 35% in the transverse direction (vertical overlap), so that point clouds cannot be generated with this setup. The point cloud shown in Figure 7 demonstrates the potential of a stereo recording, but it was captured during a different survey and will not be further discussed in the following.

The fact that the cameras and LEDs available in the project are not adequately matched in terms of their fields of view (the image size of 8.4 mm x 7.1 mm results in horizontal and vertical FOVs of approximately 70° and 60°, while the LEDs have lenses of only 50° and 30°) leads to significant differences in illumination in the overlap regions. For the investigations with the CNNs, we, therefore, focus on the central areas of the images, where the tunnel objects and damages are adequately illuminated.

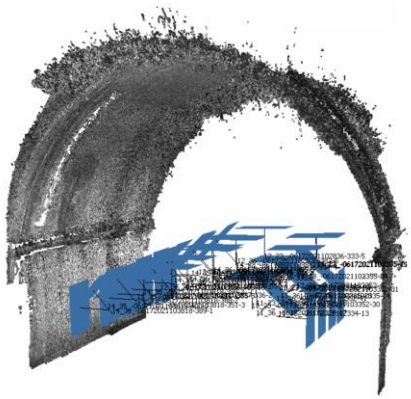


Figure 7. The point cloud of one section.

For object and damage detection, U-Net and DeepLab networks are trained on an NVIDIA GeForce RTX 2080 Ti with 8 GB GPU memory using an augmented training dataset for 10 object and 8 damage classes. The number of iterations to update the weights is fixed at 200, the batch size is 10 images. The low learning rate of about 0.0001 aims to improve the accuracy of the CNN by reducing the risk of overfitting. ADAM is used as the optimizer, and the sparse categorical cross entropy is chosen as the loss function. We used ResNet50 as the backbone model and ImageNet for initializing the weights. Table 3 shows the Intersection over Union (IoU) metrics for U-Net and DeepLab networks. The best results for object detection are achieved using U-Net with an IoU of about 78.6%. The predicted labels are shown in Table 4. The average detection rates for all types

of damages are about 49.3% and 48.5% for U-Net and DeepLab networks, respectively. The predicted labels are shown in Table 5. The accuracy of U-Net and DeepLab networks on the damage detection task is significantly less than the object detection task since the geometry and definitions of damaged areas are more complicated. Therefore, more labeled data are needed to improve the performance of the networks for segmenting damages. However, since damage detection tasks often require the detection of more subtle changes in image features that often not have a clear and well-defined boundary, the performance difference in comparison to object detection tasks, even with more extensive training, will be difficult to overcome.

Model	IoU on object detection (%)	IoU on damage detection (%)
U-Net	78.6	49.3
DeepLab	75.9	48.5

Table 3. The performance of CNNs for object and damage detection in tunnel images.

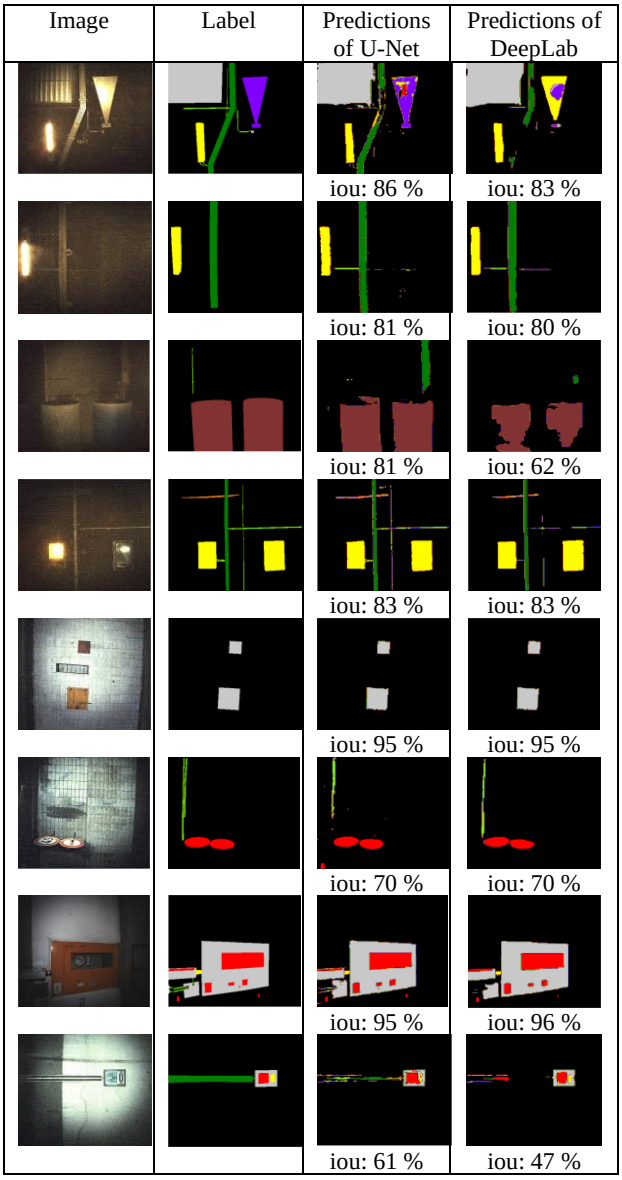


Table 5. The results of CNNs for the object detection task.

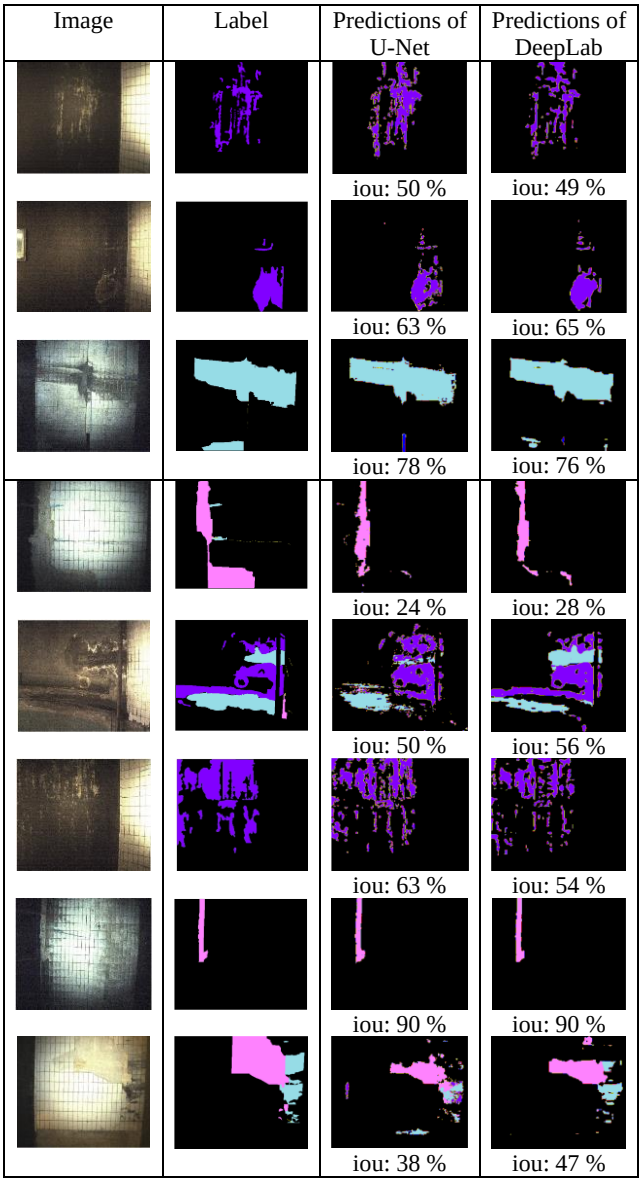


Table 6. The results of CNNs for the damage detection task.

6. CONCLUSION

This study proposes a data automation approach for object/damage detection in tunnels using a combination of machine vision and machine learning techniques. The key parameters of a machine vision system are the synchronization unit for cameras and active lighting and the geometric design of the system so that well-illuminated images with sufficient coverage of the tunnel surfaces are captured. With the investigations limited to images without the inclusion of generated point clouds, the collected data is used to train two CNNs to extract objects and damages from RGB images. The results show that U-Net outperformed the DeepLab model with IoUs of 78.6% and 49.3% for object and damage detection tasks, respectively.

ACKNOWLEDGEMENTS

This research is funded by the Federal Ministry for Economic Affairs and Energy (BMWi) as part of the ZiM project ABOUT. The project partner is Viscan Solutions GmbH (www.viscan.de).

REFERENCES

Alidoost, F., Austen, G., Hahn, M., 2022: A multi-camera mobile system for tunnel inspection. In: *iCity. Transformative Research for the Livable, Intelligent, and Sustainable City*. Coors, V., Pietruschka, D., Zeitler, B., Cham: Springer International Publishing. doi.org/10.1007/978-3-030-92096-8_13.

Chapman, M.A., Min, C., Zhang, D., 2016. Continuous mapping of tunnel walls in a GNSS-denied environment. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLI-B3, 481–485. doi.org/10.5194/isprsarchives-XLI-B3-481-2016.

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

Guo, J.F., Zong, X., Xie X.Y., Wang L., Zhai J.L., 2020. Deformation monitoring of noncircular tunnels based on 3D laser scanning. *IOP Conf. Ser.: Earth Environ. Sci.*, 570(2020)042003. doi:10.1088/1755-1315/570/4/042003.

Jiang, Y., Zhang, X., Taniguchi, T., 2019. Quantitative condition inspection and assessment of tunnel lining. *Automation in Construction*, 102, 258–269. doi.org/10.1016/j.autcon.2019.03.001.

Kim, B., Cho, S., 2020. Automated multiple concrete damage detection using instance segmentation deep learning model. *Appl. Sci.*, 10(22), 8008. doi:10.3390/app10228008.

Kumar, P., Sharma, A., Kota, S.R., 2021. Automatic multiclass instance segmentation of concrete damage using deep learning model. *IEEE Access*, 9, 90330-90345. doi:10.1109/access.2021.3090961.

Li, G., Ma, B., He, S., Ren, X., Liu, Q., 2020. Automatic tunnel crack detection based on u-net and a convolutional neural network with alternately updated clique. *Sensors*, 20(3), 717. doi.org/10.3390/s20030717.

Liu, Y., Yao, J., Lu, X., Xie, R., Li, L., 2019. DeepCrack: a deep hierarchical feature learning architecture for crack segmentation. *Neurocomputing*, 338, 139-153. doi.org/10.1016/j.neucom.2019.01.036.

Luo, Y., Chen, J., Xi, W., Zhao, P., Qiao, X., Deng, X., Qin, L., 2016. Analysis of tunnel displacement accuracy with total station. *Measurement*, 83, 29-37. doi.org/10.1016/j.measurement.2016.01.025.

Menendez, E., Victores, J.G., Montero, R., Martínez, S., Balaguer, C., 2018. Tunnel structural inspection and assessment using an autonomous robotic system. *Automation in Construction*, 87, 117–126. doi.org/10.1016/j.autcon.2017.12.001.

Mett, M., Kontrus, H., Müller, N., Eder, S., 2019. 3D tunnel inspection with photogrammetric and hybrid systems. *Int. Eng. Conf. Shotcrete for Underground Support XIV*, Pattaya, Thailand.

Panella, F., Roecklinger, N., Vojnovic, L., Loo, Y., Boehm, J., 2020. Cost-benefit analysis of rail tunnel inspection for photogrammetry and laser scanning. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLIII-B2-2020, 1137-1144. doi.org/10.5194/isprs-archives-XLIII-B2-2020-1137-2020.

Qiao, W., Ma, B., Liu, Q., Wu, X., Li, G., 2021. Computer vision-based bridge damage detection using deep convolutional networks with expectation maximum attention module. *Sensors*, 21(3), 824. doi.org/10.3390/s21030824.

Rajadurai, R.S., Kang, S.T., 2021. Automated vision-based crack detection on concrete surfaces using deep learning. *Appl. Sci.*, 11(11), 5229. doi.org/10.3390/app11115229.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: convolutional networks for biomedical image segmentation. *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015)*, 234-241. doi.org/10.1007/978-3-319-24574-4_28.

Shin, H.K., Ahn, Y.H., Lee S.H., Kim H.Y., 2020. Automatic concrete damage recognition using multi-level attention convolutional neural network. *Materials* 13(23), 5549. doi:10.3390/ma13235549.

Sun, H., Xu, Z., Yao, L., Zhong, R., Du, L., Wu, H., 2020. Tunnel monitoring and measuring system using mobile laser scanning: design and deployment. *Remote Sens.*, 12(4), 730. doi.org/10.3390/rs12040730.

Zhan, D., Yu, L., Xiao, J., Chen, T., 2015. Multi-camera and structured-light vision system (MSVS) for dynamic high-accuracy 3D measurements of railway tunnels. *Sensors*, 15(4), 8664–8684. doi.org/10.3390/s150408664.

Zhou, Y., Wang, S., Mei, X., Yin, W., Lin, C., Hu, Q., Mao, Q., 2017. Railway tunnel clearance inspection method based on 3d point cloud from mobile laser scanning. *Sensors*, 17(9), 2055. doi.org/10.3390/s17092055.