



# CNN-Based Deep Architecture for Health Monitoring of Civil and Industrial Structures Using UAVs <sup>†</sup>

Thomas Harweg <sup>1</sup>, Annika Peters <sup>2</sup>, Daniel Bachmann <sup>1</sup> and Frank Weichert <sup>1,\*</sup>

<sup>1</sup> Department of Computer Science, TU Dortmund University, 44227 Dortmund, Germany; thomas.harweg@tu-dortmund.de (T.H.); daniel.bachmann@tu-dortmund.de (D.B.)

<sup>2</sup> Faculty of Mechanical Engineering, TU Dortmund University, 44227 Dortmund, Germany; annika.peters@tu-dortmund.de

\* Correspondence: frank.weichert@tu-dortmund.de; Tel.: +49-231-755-6122

<sup>†</sup> Presented at the 6th International Electronic Conference on Sensors and Applications, 15–30 November 2019; Available online: <https://ecsa-6.sciforum.net/>.

Published: 14 November 2019



**Abstract:** Health monitoring of civil and industrial structures has been gaining importance since the collapse of the bridge in Genoa (Italy). It is vital for the creation and maintenance of reliable infrastructure. Traditional manual inspections for this task are crucial but time consuming. We present a novel approach for combining Unmanned Aerial Vehicles (UAVs) and artificial intelligence to tackle the above-mentioned challenges. Modern architectures in Convolutional Neural Networks (CNNs) were adapted to the special characteristics of data streams gathered from UAV visual sensors. The approach allows for automated detection and localization of various damages to steel structures, coatings, and fasteners, e.g., cracks or corrosion, under uncertain and real-life environments. The proposed model is based on a multi-stage cascaded classifier to account for the variety of detail level from the optical sensor captured during an UAV flight. This allows for reconciliation of the characteristics of gathered image data and crucial aspects from a steel engineer's point of view. To improve performance of the system and minimize manual data annotation, we use transfer learning based on the well-known COCO dataset combined with field inspection images. This approach provides a solid data basis for object localization and classification with state-of-the-art CNN architectures.

**Keywords:** health monitoring; civil and industrial structures; convolutional neural networks; unmanned aerial vehicles

## 1. Introduction

As of 2019, there are 39,671 bridges in road traffic in Germany. In the field of steel and steel composite bridges, this amounts to a total bridge area of 3,999,000 m<sup>2</sup>, 65% of these are much older than 20 years. About 46% of these bridges are in the lower half of the rating scale regarding their status of structure [1]. These figures only concern road bridges. Rail traffic, and, accordingly, the rail bridges are not included. There are further 25,677 (December, 2018) bridges, 4899 of them are steel constructions [2]. The *Federal Ministry of Transport and Digital Infrastructure's* traffic forecast for 2030 states that there will be an increase of 43% in rail traffic concerning tone-kilometers. The passenger service on rails should increase by almost 39% [3].

These figures illustrate the importance of a substantial healthy and reliable infrastructure. For ensuring the continuously increasing road and rail traffic, there is a regulation for periodic monitoring of bridges [4–7].

DIN 1076 “Civil Engineering Structures for roads—Monitoring and Examination” [5] regulates the inspection and control of all bridges concerning road traffic. For rail bridges, Ril 804.80 [4,6] is an analog.

Both rules mandate a building examination every six years. This examination has to be carried out visually, at a very close distance to the structure, allowing tactile feedback [5–7]. DIN 1076 calls this examination *general inspection* (H). Inspection is conducted by qualified specialists. To guarantee complete access for the specialist, technical devices are necessary. The use of these big units make (partial) road closure inevitable and are thus time consuming [5].

In addition to the described *general inspection*, DIN 1076 requires a *simple inspection* (E) every six years. This should be an extended visual inspection without tools, which is three years shifted to the *general inspection* (H). The defects found in the *general inspection* (H) should be found and compared in the *simple inspection* (E). A change of the existing defects is the focal point of this inspection (E). In those years when there is neither a *general inspection* (H) nor a *simple inspection* (E), an *observation* (B) is required. Further, a *special investigation* (S) is mandatory by order, or after general storms, fires, or accidents. These considerations demonstrate that currently a bridge inspection in Germany is tied to large expenses in personnel, equipment, and money [5,7].

Based on these challenges, the project *InÜDosS* (InÜDosS—“Inspektion, Überwachung und Dokumentation von stahlbaulichen Strukturen”—Inspection, monitoring and documentation of steel constructions) was initiated to automate monitoring, inspection, and documentation of steel constructions. Figure 1 shows the interconnection of the different stages in our automatic monitoring approach. The scope of the project *InÜDosS* covers different aspects that arise in the process of monitoring steel structures with UAVs. This includes data acquisition and 3D model reconstruction, automatic path planning, and image analysis. In this paper, our focus was on the aspect of automatic image analysis (outlined in red).

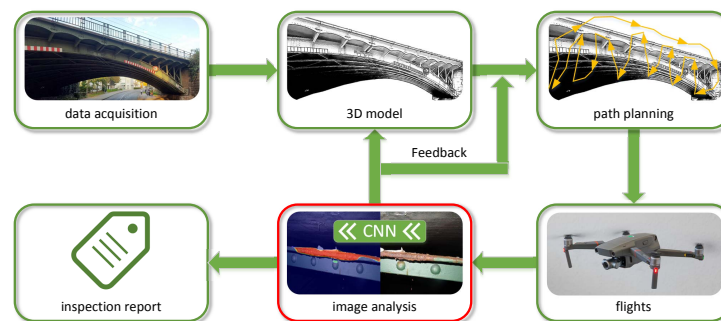


Figure 1. Connections between the different stages in automated bridge inspection.

There is much literature on the subject of (semi-)automatic detection of damages on civil and industrial structures. Koch et al. [8] gave a comprehensive overview of current practice in visual inspection and damage types of civil structures including associated severity scales. The authors stated that listed defect classes are detectable with computer vision-based approaches. Further, the state-of-the-art in computer vision defect detection and assessment in the pre-DCNN era is presented. As one of the main open challenges for automatic defect retrieval, an automated process of image and video collection is identified. Recently, Zhang et al. [9] proposed vision-based damage detection and classification using an adapted YOLOv3 network and Huethwohl et al. [10] used a multi-classifier for reinforced concrete bridge defect detection. Furthermore, automatic detection of cracks is discussed [11–14].

The novelty of the presented work, besides the definition of a hierarchy of damage classes, is that compliance with the rules and regulations of DIN 1076 is discussed. Especially the *observations* (B) and the *simple inspection* (E) should both be performed only visually and without the use of tools (cf. [5,7]).

We discuss a vision based approach in (semi-)automatic damage detection in bridges and steel constructions using UAVs. The main contributions of this work are:

- Development of a semantic classification scheme for bridge inspection in public space, which is compatible to the specifications of DIN 1076
- Devising a hierarchy of defect classes
- Implementation and evaluation of a corresponding cascaded classifier based on state-of-the-art CNN object detection

## 2. Materials and Methods

The special nature of the image data gathered demands a custom approach to tackle the problem. Field inspection images at hand show great variety concerning characteristics of the structures being investigated (cf. Figure 2), as well as the level of detail. Depending on the distance between the UAV and the examined object, the different types of damages only add to the complexity of the problem. Our approach is based on a hierarchy on a conceptual level, as displayed in Figure 3, which we directly transfer to our computational method. The proposed hierarchy consists of two stages. At the first stage, we aim to identify relevant structures. Roughly speaking, this comprises all (coated) supporting steel structures. If such a structure is identified, it will then be analyzed for damages. At the second stage, we consider two main categories of potential damage: coating damages and faulty fasteners. Coating damages are further subdivided into *blistering*, *cracks*, *missing coating*, and *corrosion*. Fasteners can be *intact*, *defective* (e.g., loose nuts), *tilted*, *missing*, or *corroded*.



Figure 2. Example images from the dataset (All images courtesy of PSP—Prof. Sedlacek & Partner, Planung und Entwicklung im Bauwesen, Büro Dortmund).

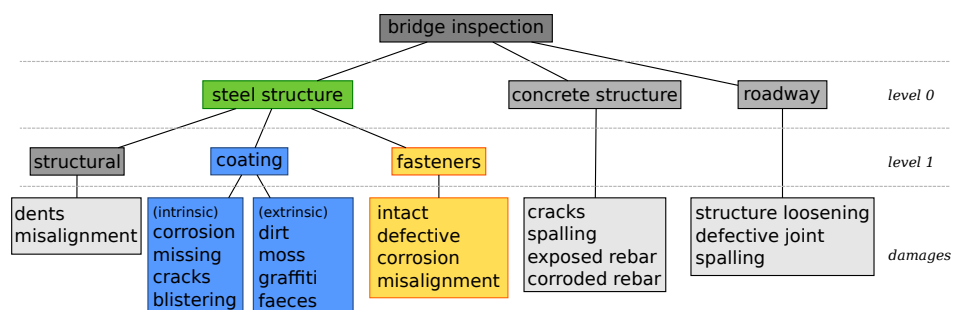
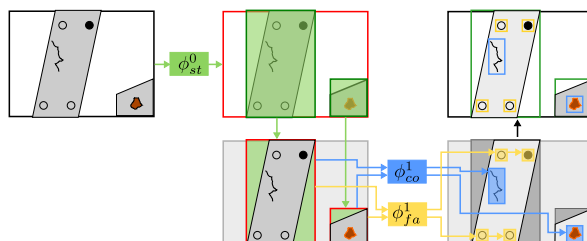


Figure 3. Aspects of bridge inspection and proposed semantic hierarchy. Components relevant to this paper are highlighted in color.

We transfer this concept directly to the computational domain by using a two-stage convolutional neural network (CNN) approach, based on *Mask-RCNN* [15]. We use three separate instances of the Mask-RCNN (two in the second stage). All three networks use the ResNet50-backbone [16] and *Feature Pyramid Networks* (FPN) [17]. Our implementation is based on the implementation by *Facebook Research* [18]. The proposed two-stage architecture is shown in Figure 4.



**Figure 4.** Two-stage CNN pipeline: (coated) steel structures and their positions are identified and errors are classified on the associated bounding boxes.

The first network ( $\phi_{st}^0$ ) is trained to recognize coated steel structures, resulting in bounding boxes of recognized structures. These bounding boxes are used as regions of interest (ROI) to restrict networks of the second stage. In the second stage, we run two separate networks, one for coating damages ( $\phi_{co}^1$ ) and the other for fastener damages ( $\phi_{fa}^1$ ).

More formally, for an input image  $I^0$ , the first-stage detector  $\phi_{st}^0(I^0)$  determines a set  $R_{st}^0 = \{B_{0,0}^0, \dots, B_n^0\}$  of bounding boxes  $B_i^0$  which define the areas of relevant structures. For each bounding box  $B_i^0$ , the corresponding sub-image  $I_i^1$  is cut from the full image  $I^0$ , resulting in the set of images  $I^1 = \{I_0^1, \dots, I_n^1\}$ .

Second-stage detectors  $\phi_{co}^1$  and  $\phi_{fa}^1$  are then applied to each sub-image  $I_i^1$ , again resulting in sets of bounding boxes  $R_{co}^1 = \{\{B_{0,0}^{1,co}, \dots, B_{0,m_0}^{1,co}\}, \dots, \{B_{n,0}^{1,co}, \dots, B_{n,m_n}^{1,co}\}\}$  and  $R_{fa}^1 = \{\{B_{0,0}^{1,fa}, \dots, B_{0,k_0}^{1,fa}\}, \dots, \{B_{0,n}^{1,fa}, \dots, B_{n,k_n}^{1,fa}\}\}$ . All results contained in  $R^1 = R_{co}^1 \cup R_{fa}^1$  are then projected back onto the input image  $I^0$ , giving the final prediction results  $R^0$ .

As a data basis, we used approximately 1000 images of steel constructions, consisting mostly of bridges and supporting steel constructions, many of them showing at least one type of damage we are considering. These images are divided into three categories: *steel structures*, *coating damages*, and *fasteners*. Note that the images of the three datasets are not necessarily disjoint, but the corresponding annotations are. To reduce computational costs while increasing accuracy, we used transfer learning. We initialized all three networks with network weights pre-trained on the COCO (Common Objects in COntext) dataset [19]. We trained each network for 80 epochs, at a learning rate of  $2 \times 10^{-3}$  and a batch size of 2. Further, an initial warm-up phase of 500 iterations (batches) was used to overcome difficulties which may arise in early phases of optimization [20]. During training, validation tests were performed after each epoch. Data augmentation included horizontal and vertical flip, as well as variation in brightness, contrast, and saturation. Datasets were split into 11 parts of (near-)equal size. One part was used as a test set and the other ten parts for cross-validation (10-fold split).

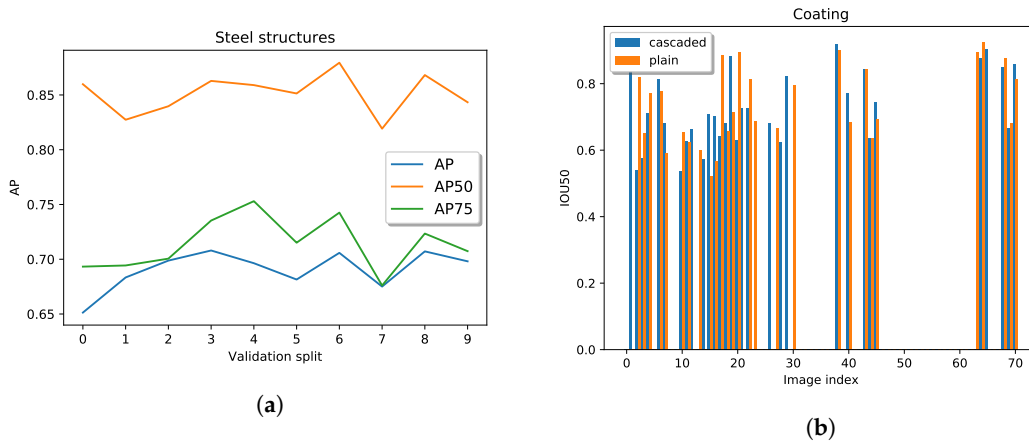
### 3. Results

Figure 2 shows examples of the images used in this study. For each validation test, we determined the model(s) with the best AP, AP50, and AP75, respectively. Figure 5a shows the average precision on the validation set of the ten-fold split (cf. Section 2). This image shows the evaluation for the first stage

(steel constructions). To make results of the cascaded and plain version comparable, we first applied the trained detector for supporting structures and then applied the same model as in the simple step on all areas suggested by the first detector. The results of this step were then projected back on the original (full) image. This allowed for directly comparing the results of the plain and the cascaded detector. We selected models according to the best performance regarding average precision.

To test our proposed two stage method, we exemplarily performed tests regarding the second-tier category *coating* in a simple version and a cascaded version. For the second level coating, we additionally trained networks on the same data but with reduced number of classes, in order to test how well the network is able to only recognize and spot damages without further classifying them. In these tests, we used the classes *intact* and *defective*.

For the simple version, we applied the trained model of the given category on the corresponding test dataset as is common practice. In this case, the detector was simply applied to the whole image. Figure 5b shows IOUs on image level for a test set of 70 images. The cascaded version shows an improvement of 9% compared to the plain version. Figure 6 shows detection results for coatings and fasteners. Red boxes indicate detection results for the first stage (steel constructions), while green boxes and outlines show results of the second stage (coating or fasteners).



**Figure 5.** (a) Average precision on validation sets (Stage 1). (b) IOUs of simple and cascaded variants for category coating.



**Figure 6.** Cascaded prediction examples: red boxes indicate results from the first stage, while green boxes and outlines show results from the second stage

#### 4. Discussion and Conclusions

Detection of steel structures works very well, while detection and especially classification of damages is more difficult. This is mainly due to the fact that the data and the semantic categories are inherently difficult. Some damages are very subtle and hard to spot from images, even for an expert, while others

depend on a significant amount of context information and interpretation. Furthermore, classes are not always clearly distinguishable, and objects may adhere to several classes at once (fasteners may be defective and corroded, for example). Overall spotting works well, and classification may be sensible, even when differing from ground truth annotation. This applies when classes overlap, or classification by a human expert is fuzzy by itself.

An automatic visual detection of defects on steel bridges complies the requirements of DIN 1076 and Ril 804.80. Required inspections are, with the exception of the *main inspection* (H), only based on visual monitoring. Numerous areas of the bridge cannot be reached and inspected without big units such as aerial work platforms or units to inspect bridges from below. These tools are not intended for *observations* (B) and *simple inspections* (E). Thus, a periodic flyover with a UAV could improve and specify both inspections (B and E). Further, existing defects are observed and checked regularly and small changes would be detected immediately [5–7].

The presented work shows the potential of the developed approach in designing a DIN 1076 compliant visual inspection pipeline for automated damage detection. Due to more precise annual inspections by UAV, which would replace the superficial and visual inspections without tools (B and E), a process of change of a defect would strike immediately and maintenance measures can be taken. The structure and its durability could be improved and maximised. This presents a further step towards smart city infrastructures.

**Author Contributions:** All authors contributed substantially to the presented work.

**Funding:** This work was funded by the Forschungsvereinigung Stahlanwendung (FOSTA e.V.)—project number P1326/17/2018 “Inspektion, Überwachung und Dokumentation von stahlbaulichen Strukturen”.

**Acknowledgments:** Some of the images used are courtesy of PSP—Professor Sedlacek & Partner, Planung und Entwicklung im Bauwesen, Büro Dortmund.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Bundesanstalt für Straßenwesen BASt, Bergisch Gladbach. Brücken an Bundesfernstraßen, Brückenstatistik. 2017. Available online: [https://www.bast.de/BASt\\_2017/DE/Statistik/Bruecken/Brueckenstatistik.pdf](https://www.bast.de/BASt_2017/DE/Statistik/Bruecken/Brueckenstatistik.pdf) (accessed on 14 October 2019).
2. DB Netz AG. Brückenkarte. Available online: <https://bruecken.deutschebahn.com/brueckenkarte> (accessed on 14 October 2019).
3. Bundesregierung. Verkehrsprognose 2030 . 2014. Available online: <https://www.bundesregierung.de/breg-de/aktuelles/zunehmender-verkehr-erfordert-investitionen-440698> (accessed on 14 October 2019).
4. DB Netz AG. *DB Richtlinie 804, Modul 8002: Ingenieurbauwerke planen, bauen u. instand halten—Inspektion von Ingenieurbauwerken—Eisenbahnüberführungen*; DB Netz AG: Frankfurt, Germany, 2008.
5. Deutsches Institut für Normung e.V. *DIN 1076: Ingenieurbauwerke im Zuge von Straßen und Wegen – Überwachung und Prüfung*; Deutsches Institut für Normung e.V: Berlin, Germany, 1999.
6. DB Netz AG. *DB Richtlinie 804, Modul 8001: Ingenieurbauwerke planen, bauen u. instand halten—Inspektion von Ingenieurbauwerken—Allgemeine Grundsätze*; DB Netz AG: Frankfurt, Germany, 2008.
7. Bundesministerium für Verkehr und digitale Infrastruktur (BMVI). *RI-EBW-PRÜF: Richtlinie zur einheitlichen Erfassung, Bewertung, Aufzeichnung und Auswertung von Ergebnissen der Bauwerksprüfungen nach DIN 1076*; Bundesministerium für Verkehr und digitale Infrastruktur (BMVI): Berlin, Germany, 2017.
8. Koch, C.; Georgieva, K.; Kasireddy, V.; Akinci, B.; Fieguth, P.W. A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure. *Adv. Eng. Inform.* **2015**, *29*, 196–210.
9. Zhang, C.; Chang, C.; Jamshidi, M. Bridge Damage Detection using Single-Stage Detector and Field Inspection Images. *arXiv* **2018**, arXiv:1812.10590.

10. Huethwohl, P.; Lu, R.; Brilakis, I. Multi-classifier for reinforced concrete bridge defects. *Autom. Constr.* **2019**, *105*, doi:10.1016/j.autcon.2019.04.019.
11. Yang, X.; Li, H.; Yu, Y.; Luo, X.; Huang, T.; Yang, X. Automatic Pixel-Level Crack Detection and Measurement Using Fully Convolutional Network. *Comput.-Aided Civ. Infrastruct. Eng.* **2018**, *33*, 1090–1109, doi:10.1111/mice.12412.
12. Dorafshan, S.; Thomas, R.J.; Maguire, M. Comparison of deep convolutional neural networks and edge detectors for image-based crack detection in concrete. *Constr. Build. Mater.* **2018**, *186*, 1031–1045, doi:10.1016/j.conbuildmat.2018.08.011.
13. Turkan, Y.; Hong, J.; Laflamme, S.; Puri, N. Adaptive wavelet neural network for terrestrial laser scanner-based crack detection. *Autom. Constr.* **2018**, *94*, 191–202, doi:10.1016/j.autcon.2018.06.017.
14. Ni, F.; Zhang, J.; Chen, Z. Zernike-moment measurement of thin-crack width in images enabled by dual-scale deep learning. *Comput.-Aided Civ. Infrastruct. Eng.* **2019**, *34*, 367–384, doi:10.1111/mice.12421.
15. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In Proceedings of the International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017.
16. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
17. Lin, T.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 936–944, doi:10.1109/CVPR.2017.106.
18. Massa, F.; Girshick, R. maskrcnn-benchmark: Fast, modular reference implementation of Instance Segmentation and Object Detection algorithms in PyTorch. 2018. Available online: <https://github.com/facebookresearch/maskrcnn-benchmark> (accessed on 14 October 2019).
19. Lin, T.Y.; Maire, M.; Belongie, S.; Bourdev, L.; Girshick, R.; Hays, J.; Perona, P.; Ramanan, D.; Zitnick, C.L.; Dollár, P. *Microsoft COCO: Common Objects in Context*; Springer: Cham, Switzerland, 2014.
20. Goyal, P.; Dollár, P.; Girshick, R.; Noordhuis, P.; Wesolowski, L.; Kyrola, A.; Tulloch, A.; Jia, Y.; He, K. Accurate, Large Minibatch SGD: Training ImageNet in 1 Hour. *arXiv* **2017**, arXiv:1706.02677.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).