ENHANCING THE QUALITY OF VISUAL ROAD CONDITION ASSESSMENT BY DEEP LEARNING

M. EISENBACH [1], R. STRICKER [1], M. SESSELMANN [2],
D. SEICHTER [1] & H.-M. GROSS [1*]
[1] Neuroinformatics and Cognitive Robotics Lab, TU Ilmenau, Germany
Markus.Eisenbach@tu-ilmenau.de
[2] LEHMANN + PARTNER GmbH, Germany
Sesselmann@lehmann-partner.de

ABSTRACT

Aging public roads need frequent inspections to analyze their condition and guarantee their permanent availability. Unfortunately, condition assessment in terms of distress detection and classification requires manual visual inspection. This manual labor is not only expensive but very time-consuming if millions of high-resolution road images of a country's whole road network must be analyzed. The annotation quality suffers from people getting fatigue during the tedious process, and the results can vary a lot between different operators. Therefore, we developed a deep-learning based road image analysis system that partially automates the image inspection process. Each image is analyzed by a deep neural network that is trained to detect pavement distress at a very high quality. Potential pavement distress is highlighted, helping the operator to focus on regions most critical for the assessment. This speeds up the inspection process and enhances the operators' concentration since they have to process only relevant information. If desired, the deep neural network can also be applied to pre-annotate images and making suggestions for the type of distress following the German standard for assessment of road surface characteristics. To get a deep neural network capable of doing these analyses at human-like performance, we did extensive studies. First, we created a high-quality dataset using a data acquisition system (fast driving mobile mapping system) certified by the German Federal Highway Research Institute (BASt). The dataset covers many variations of distress, road surfaces, and all kind of built-in components, like gullies, manholes, cobblestones, and curbs. Then we evaluated several state-of-the-art deep neural networks for visual image analysis, by training them on our road dataset. Lastly, we derived our network architecture, which is both fast and accurate and outperforms classical machine learning and computer vision methods by far. The deep neural network has been integrated into our analysis tool that follows the German standard for road condition assessment as proposed by the German Road and Transportation Research Association (FGSV). Experiments on inspections of reference tracks show that using this new tool the operators were able to analyze images up to twice as fast and with consistently higher quality. Additionally, 14% of federal highways and inner-city roads images and 30% of freeway images could be held out from manual inspection since they could be ensured to be distress free by the deep neural network at high confidence.

1. INTRODUCTION

Public infrastructures are suffering from aging and therefore need frequent inspection. Distress detection and solid management for maintenance are the keys to guarantee their permanent availability. Thus, condition acquisition and assessment must be applied to the whole road network of a country frequently.

* This work has received funding from the German Federal Ministry of Education and Research as part of the ASINVOS project under grant agreement no. 01IS15036

Following German federal regulations, the surface characteristics must be evaluated i.a. regarding substance condition. The substance condition is captured with camera systems and has to be assessed by visual inspection of the recorded images. Current evaluation is done manually and therefore requires excessive manual labor. This includes tasks like finding very thin cracks that appear only in a few pixels of the image. Thus, the period between the actual inspection and the final evaluation may be up to several months. In the meantime, small damages, like cracks, can lead to substantial downtimes with a high impact on the population.

In the research project ASINVOS*, we aim to automate this process to a high degree by applying machine learning techniques. The basic idea is to train a self-learning system with manually annotated data from previous inspections so that the system learns to recognize underlying patterns of distress. Once the system can identify intact infrastructure robustly, it can reduce the human amount of work by presenting only distress candidates to the operator. This helps to speed up the inspection process significantly and simultaneously reduces costs. Furthermore, inspection intervals can be reduced, which helps to remedy deficiencies in time.

To achieve the goal of automated condition assessment, we utilize a deep neural network. In order to train it, we created a large dataset containing high-quality standardized images with detailed annotations [1] (see Section 3). Due to our extensive analyses and neural network architecture optimizations (see Section 4), this deep-learning based road image analysis system can detect distress of the surface at a very high quality.

In this paper, we show how we embedded the detections results in the analysis process. The basic idea is to highlight potential pavement distress such that the operator can focus on regions most critical for the assessment. This speeds up the inspection process and enhances the operators' concentration since they must process only relevant information. If desired, the deep neural network can also be applied to pre-annotate all images and making suggestions for the type of distress.

2. RELATED WORK

With first attempts published in the early nineties, e.g. [2], automating the distress detection process has already been addressed by researchers for almost three decades. Therefore, a wide variety of different approaches have been presented ranging from traditional image processing techniques to deep learning approaches more recently. Since the typical pavement assessment processes are carried out using 2D image recordings, the related work section is focused on 2D image processing approaches. However, some authors have shown the effectiveness of using 3D data for distress detection [3]. We refer to [4] for a more detailed list of depth-based 3D approaches.

The algorithms developed for 2D image-based evaluation of the pavement surface can be divided into two major categories:

- computer vision algorithms designed explicitly for crack detection, mostly by applying image value thresholding and
- algorithms for general distress detection that use implicit or explicit local feature extraction.

2.1. Crack Detection

The first group of algorithms uses image processing methods to detect road distress structures that can be extracted by thresholding afterward. Therefore, preprocessing

* Assistierendes und Interaktiv lernfähiges Videoinspektiossystem für Oberflächenstrukturen am Beispiel von Straßenbelägen und Rohrleitungen (Interactive machine learning based monitoring system for pavement surface analysis)

algorithms are applied in order to reduce illumination artifacts. Under the assumption that crack structures can be identified as local intensity minima, this group of algorithms usually uses thresholding in the image space to find crack candidates. The resulting crack image is further refined by morphological image operations and by searching for connected components. Approaches belonging to this category are presented in [5], [6], as well as in [7], where the closed source but publicly available CrackIT toolbox is presented. Other variants of this category for example use minimal-path-based [8] or graph-based [9] crack candidate analysis for further refinement, that is also used by the well-known CrackTree approach [10].

2.2. Feature-based Distress Classification

The algorithms of the second category apply classifiers to local regions of the image to extract crack or distress regions. Traditional image processing approaches mostly rely on explicit feature extraction. Using a Support vector machine (SVM) is very common among these conventional approaches. For example, this classifier is applied to Histogram of Oriented Gradient (HOG) features [11] or Local Binary Patterns (LBP) [12], [13]. A recently proposed approach has been applied to frontal-facing images by integrating an image clustering step to extract the street surface first [14].

More recent approaches tend to use implicit feature-extraction by using Convolutional Neural Networks (CNNs). Methods based on CNNs mainly differ regarding network architecture, predicted distress classes and whether downward or frontal-facing input images are processed.

One of the first attempts to apply CNNs in the domain of pavement distress detection is presented in [15]. The network architecture shares some similarities with LeNet-5 [16] but utilizes ReLU activation and four convolution/pooling layers for detection pavement cracks on downward facing road images. Also applied to top-down facing road images, but using VGG-based CNNs [17] are the approaches presented in [18] and [1]. While the former method is focused on crack detection, the latter addresses all distress types addressed by the federal German pavement assessment process.

With the tremendous distress detection improvements over traditional image processing techniques archived by CNNs [10], [1], distress detection in frontal-facing images it getting more common. This kind of image is often preprocessed to constrain distress detection to the pavement area. This step can be carried out using traditional image segmentation techniques like graph-based hierarchical clustering [19] or using CNNs like SegNet [20]. The network architectures used for processing frontal-facing images are based on state-of-the-art image processing networks. [21] for instance compares InceptionV2 and MobileNet for the detection of eight different distress classes while [20] applies Squeeze-Net for distress detection. [22] present a Feature Pyramid and Hierarchical Boosting Network which is used for crack detection in frontal-facing images.

In this paper, we focus on distress detection in downward facing road images. Frontal-facing images do not provide the resolution necessary to detect minimal damages. This can be seen in many approaches trained on frontal-facing images since they often miss tiny cracks, that are still identified by methods applied to downward facing road images.

3. DATASET

The GAPs dataset* is the most extensive dataset in the pavement distress domain that provides standardized, high-quality images. Due to many requests, we provide 500 additional images from an additional Federal highway to enlarge the dataset even more.

* The GAPs dataset is available at http://www.tu-ilmenau.de/neurob/data-sets-code/gaps/

Now, it consists of 2468 HD road surface images. We also provide a sub-sampled dataset containing 50k images to allow for fast training and comparison to the state of the art in an MNIST- or CIFAR-like fashion. Most importantly, we improved annotations. The level of detail is increased by labeling distress by several small bounding boxes that enclose the distress tightly. Additionally, all annotations were checked for correctness by several experts. In the following, we refer to this dataset as GAPs v2.

3.1. Standardized Data Acquisition

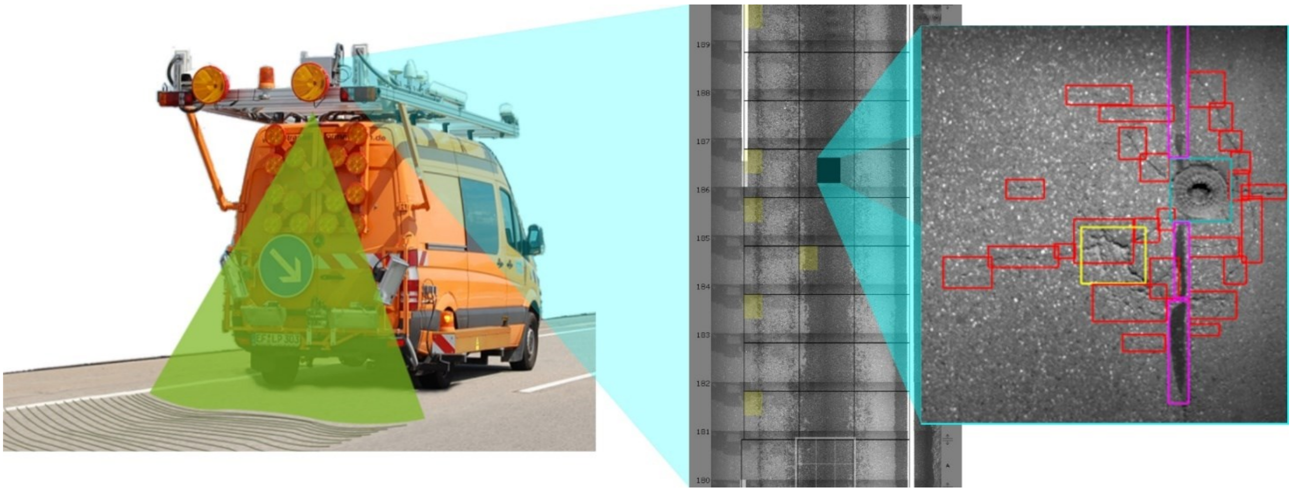

Figure 1 - Mobile mapping system S.T.I.E.R. middle: Labeling as expected by German FGSV-regulation. right: fine labeling of different distress types using bounding boxes.

The data acquisition is based on the specification by the German Road and Transportation Research Association (FGSV) – the so-called Road Monitoring and Assessment (RMA) [23]. The RMA process standardizes data acquisition on a systematic basis and provides nationwide uniform parameters to ensure objective analyses of surface conditions as well as a high degree of quality. The key aspects are longitudinal and transversal evenness, skid resistance and surface distress. Mobile mapping systems, equipped with high-resolution cameras and laser-based sensors, are the state of the art in the RMA context. The image data of the GAPs dataset have been captured by the mobile mapping system S.T.I.E.R (Figure 1), that is certified annually by the German Federal Highway Research Institute (BASt) since 2012. This vehicle is equipped with several high-resolution cameras, i.a. two slightly overlapping bird-eye-view photogrammetrically calibrated monochrome cameras mounted left and right at the rear of S.T.I.E.R's roof rack. These two cameras capture the pavement's surface in detail with images at a resolution of 1920x1080 pixels, which means each pixel covers 1.2 mm x 1.2 mm of the surface. The surface camera system is synchronized with a high-performance lighting unit. This allows continuous capturing of road surface images even at high velocities (ca 80 km/h) and independent of the natural lighting situation. Within the scope of the standard RMA workflow, a sequence of left and right surface camera images is stitched together in driving direction. The result is a continuous sequence of surface images that represent 10 meters of the entire driven traffic lane. According to the FGSV-regulation, the surface damage detection and analysis process is based on these images. For this, an inspection grid is applied to each 10-meter-image. A single grid cell has a longitudinal length of 1m and a transversal length of 1/3 of the lane width. If a grid cell contains relevant surface damage, the whole cell is assigned to this damage type. Once the damage detection and classification is done, the measured raw-data is used to calculate condition variables and finally condition grades ranging from "very good" to "very poor" using a weighting scheme defined by the FGSV. The presented conventional labeling approach is sufficient for indicating the level of safety and comfort for road users, but due to the lack of the precise damage location labels in terms of pixel coordinates, this labeling is not appropriate to train a classifier and therefore, must be improved in detail as shown in Figure 1 right.

For more details regarding the data acquisition process and the measurement vehicle, we refer to [1]. Following the German FGSV-regulations, the surface defect classes shown in Figure 2 must be detected.

3.2. Improvement of the GAPs dataset

The GAPs dataset [1] have been improved in several ways:

3.2.1 *More data*

The GAPs v2 dataset includes a total of 2,468 gray valued images (8bit), partitioned into 1,417 training images, 51 validation images, 500 validation-test images, and 500 test images, following the partitioning suggestions of [24]. Using these images, 692,377 patches of surface defects and 6,035,404 patches of intact road are extracted. Table 1 shows the unbalanced class distribution of the full dataset. The pictured surface material now contains pavement of four different German federal roads.

Table 1 - Class distribution of GAPs dataset.

Class	Full dataset	50k dataset
Intact road	89.71 %	60 %
Cracks	7.28 %	20 %
Applied patches	1.72 %	10 %
Inlaid patches	0.75 %	5 %
Potholes	0.30 %	3 %
Open joints	0.24 %	2 %

3.2.2 *Refined annotations*

The images have been annotated manually by multiple trained operators at a high-resolution scale such that a bounding box encloses actual damage and the non-damage space within a bounding box has a size of lower than 64x64 pixels. All annotations of the first version of the GAPs dataset have been refined, such that the non-damage space within a bounding box is even smaller than that restriction. Conflicting annotations have been resolved (GAPs v1 had only one annotator per image).

3.2.3 *More context*

While GAPs v1 offered only patches of size 64x64 extracted within the annotated regions and the intact surface regions, GAPs v2 offers several patch sizes showing more context (see Figure 3). The defect region is still ensured to be within the 64x64 center region of each patch, but the surroundings may help to make correct decisions. In Section 4.2 we

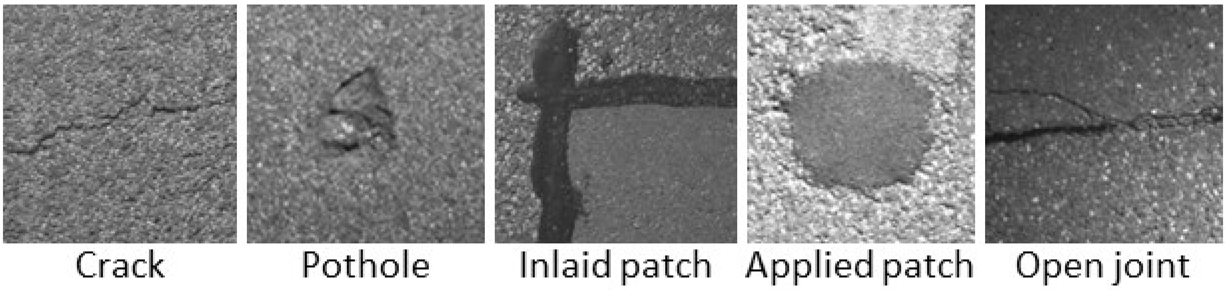

Figure 2 - Surface defect classes in GAPs dataset. The class cracks comprises all sorts of cracks like single/multiple cracking, longitudinal/transversal cracking, alligator cracking, and sealed/filled cracks.

briefly outline the benefits of context information.

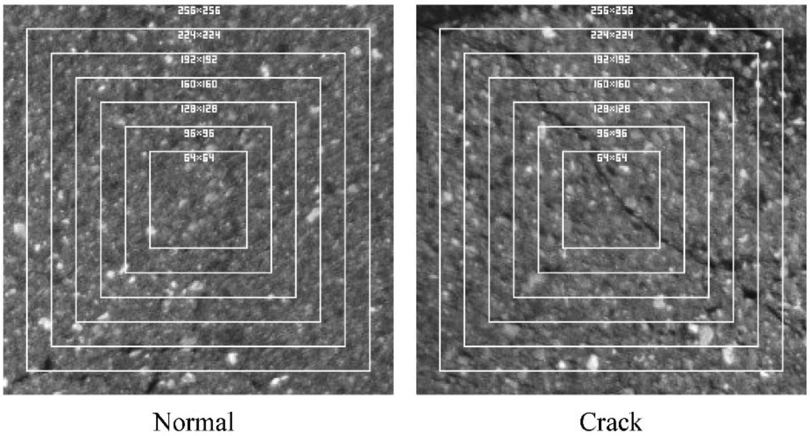

Figure 3 - Visualization of context captured by the different patch sizes

3.2.4 *50k subset available*

Since deep learning benefited most from small size real-world datasets, we also created a smaller subset for fast experiments. Inspired by the MNIST and CIFAR datasets, we created a training set of 50,000 samples. Additionally, the validation set, validation-test set, and test set contain 10,000 samples each. Table 1 shows the chosen class distribution of the 50k subset. The samples for each class were randomly selected until the desired number of samples was reached. The classes are left unbalanced, but the dominant classes of intact road and cracks are not that dominant as in the full dataset. The relative fraction of the non-dominant classes among themselves is similar to the original dataset.
We have chosen this distribution in order to focus more on the distress than on the intact road. The experiments in Section 4.2 confirm that this approach is an excellent choice since observations for classifiers trained on the small subset transfer to identical classifiers trained on the full dataset.

4. SYSTEM OVERVIEW

During the RMA data acquisition process, 10-meter images are generated. Each of these images must be reviewed, and distress regions have to be tagged within the 10x3 RMA grid. To automate this process, first, the correct grid position must be found. Therefore, the lane must be detected. Next, the distress detector has to be applied to the image to find potential distress regions. Finally, the RMA grid and the detection results have to be combined. The whole processing chain is outlined in Figure 4.

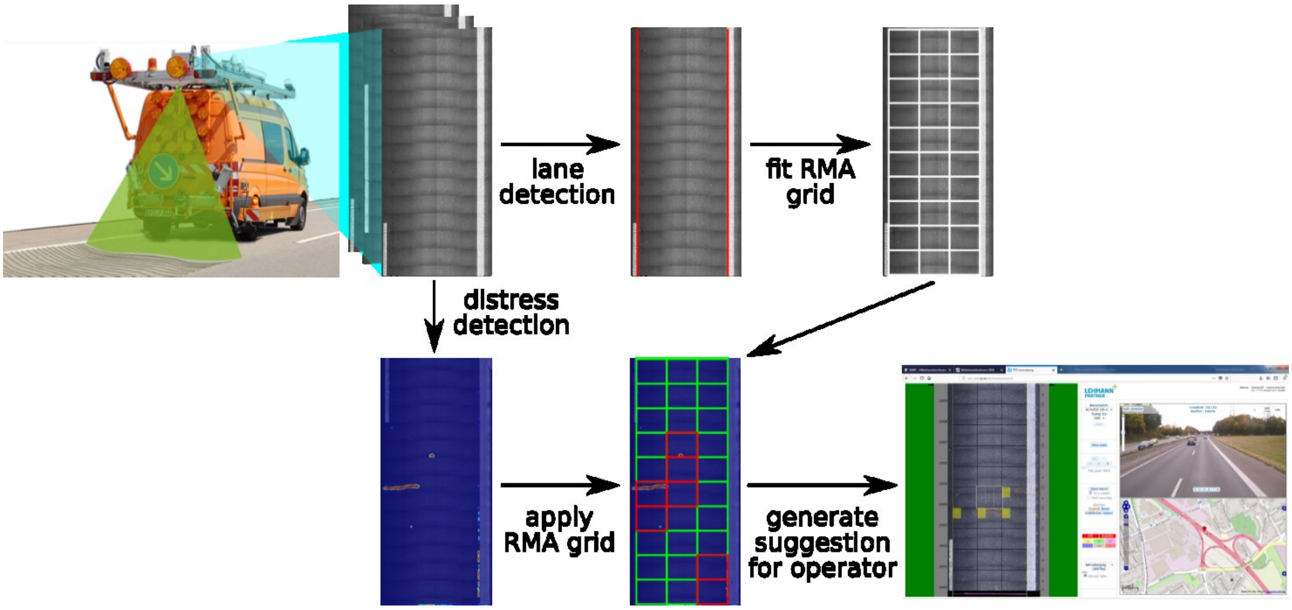

Figure 4 - Outline of the whole process chain. First the lane is extracted, and the RMA grid is fitted to the image. Afterwards, the distress detection is combined with the RMA grid in order to generate distress candidates that need to be reviewed by the operator.

4.1. Lane Detection

To detect the lane, we need to extract information about limiting structures. For example, road markings, curbs, sidewalks, bikeways, and other limiting elements are of relevance. Since these structuring elements occur in various appearances, we decided in favor of a neural network to robustly detect them. We adapted the neural network for distress detection by changing only the output coding. This convolutional neural network can distinguish between road markings, roadsides, and regular road (Figure 5). It was trained on the GAPs dataset with additional annotations for road markings and roadsides. After training, it was converted to a fully convolutional network for image segmentation.

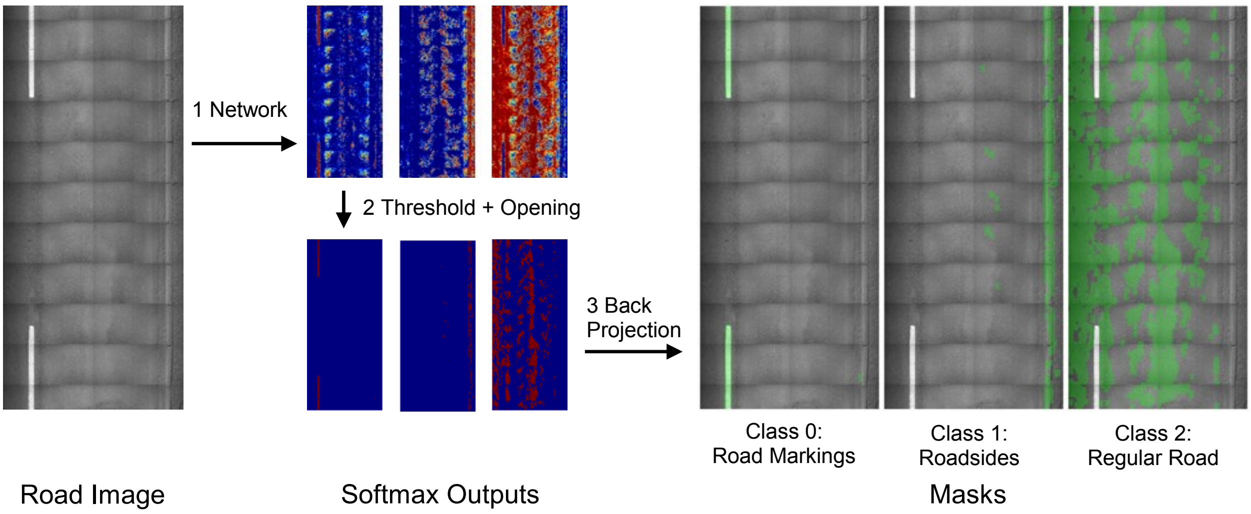

Figure 5 - Workflow to distinguish regular road from road markings and roadsides. First, an adapted ASINVOSnet is applied. Subsequently, to get an accurate segmentation, the network's output is postprocessed and upsampled considering the network's properties.

To detect the limits of the lane, we follow the workflow shown in Figure 5: First, we apply the neural network to the image. Then, a threshold is used to get binary masks. Finally, these masks are upsampled considering both the input shape and the overall stride of the neural network. Since the mask representing regular road is not sufficient to detect the entire lane, we follow a fusion strategy as shown in Figure 6 to get a single mask of lane limiting structures.

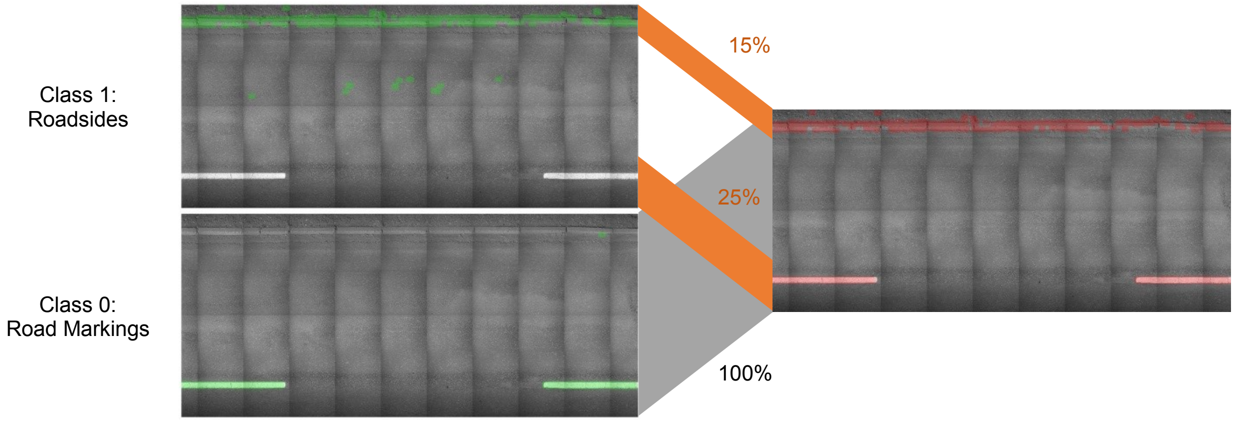

Figure 6 - Workflow for combing the masks representing road markings and roadsides. While roadsides are supposed to appear at the outer borders only, road markings can appear everywhere due to road narrows on both sides.

Given the fused mask containing all lane limiting elements, we can estimate the lane's position. First, we sum up the elements along the driving direction to get the cross-section of the road. Next, the cross-section is divided into several segments using a threshold. Since road markings, like arrows, may lead to disconnected parts, we start from the largest central segment and follow a rule-based fusion strategy incorporating the specifications for legitimate roadway widths to finally detect the lane.
For evaluating the lane detection, we used 1,637 manually annotated high-resolution road images, showing ten meters of a lane each. The images are taken from several German federal highways. Figure 7 shows exemplary results for lane detection. In 1,541 of the 1,637 test images, the lane could be detected correctly (94.14%). Road markings in the middle of the lane can be handled correctly. Errors were mainly caused by faded road markings leading to missed lane limits.

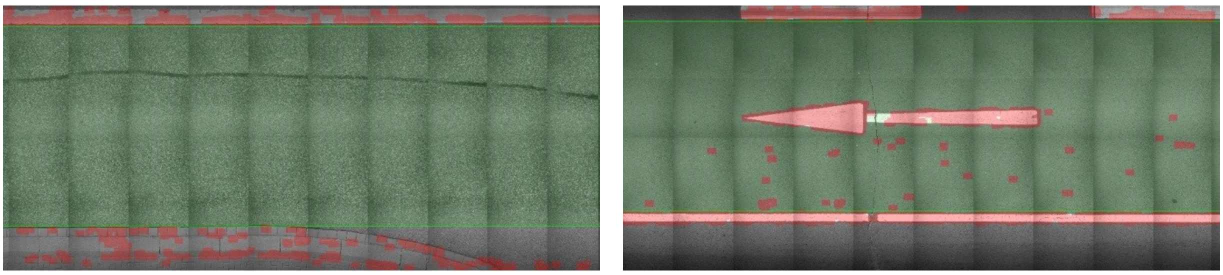

Figure 7 - Results for lane grid detection (green). Road markings in the middle of the lane (right) can be handled correctly. The mask showing potential lane limits is highlighted red.

4.2. Training of the distress detector

We have trained several neural networks with different network architectures and adapted training methods to obtain a capable distress classifier that is both accurate and reasonably fast. We have trained classic VGG-based deep neural networks that were also used in [1] as well as Residual Neural Networks (ResNets) [25] that represent a more modern network architecture. Furthermore, we have trained a traditional machine learning approach that extracts HOG features and Local Binary Patterns from the image patches and uses an SVM for classification.
For comparison, the classifiers were trained on the binary decision problem (distress vs. no distress) and the 50k subset and additional data augmentation has been applied. The training was carried out on the 64x64 pixels training dataset only. The best classifier was selected on the valid-test dataset, and the final classification results were obtained on the

test dataset. We computed the F1 score, which considers both precision and recall, and have found to be most useful to measure performance on the GAPs dataset [1]. The results that are given in Figure 8 clearly show the benefit of the deep-learning-based detectors. The deep networks can learn a representation of distress, that performs much better than the hand-crafted features used in the classic machine learning approach.

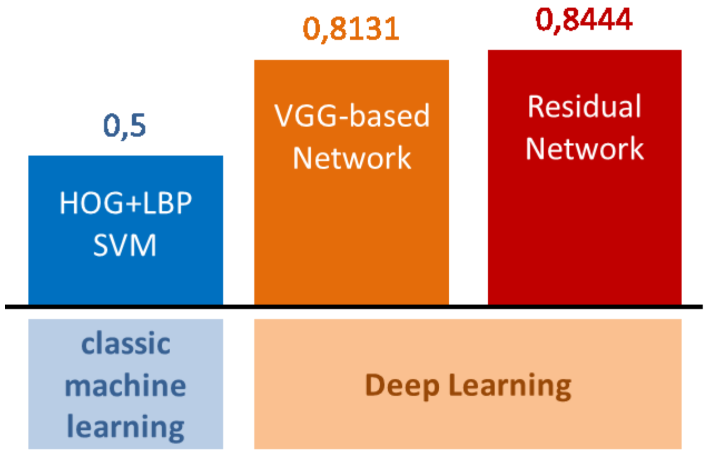

Figure 8 - F1-scores on test dataset for different distress classifiers. Patch size of 64x64 pixels is used.

To further improve the performance, we tested if the detection can be improved by adding more context. Therefore, we trained the VGG-based network and the ResNet on different patch sizes of the GAPs dataset. Figure 9 reveals that both networks greatly benefit from additional context information. However, it also becomes evident that the ResNet is more suitable for distress detection. Since it also requires less computational power during training and inference, we opted for that network architecture and a patch size of 160 pixels for the final distress detection system. Although the performance on the test dataset improves slightly for patch sizes above 160 pixels the additional performance comes at the cost of additional computational complexity.

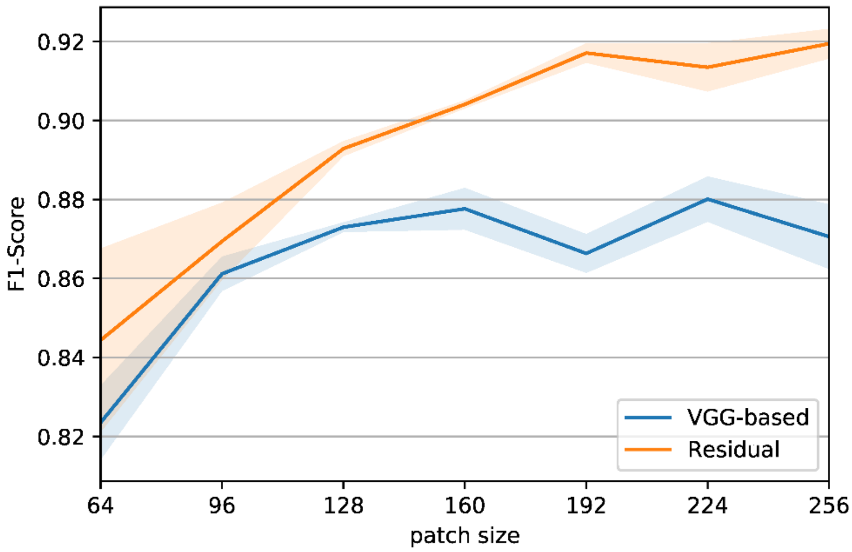

Figure 9 - Comparison between F1 scores on the test dataset for VGG-based network and Resnet for different patch sizes.

All experiments presented so far are carried out on the 50k subset. Therefore, we also evaluated the classifier performance if training is performed on the full dataset. It turns out that the performance is almost equal for both datasets if no data augmentation is applied (F1 score of 0.8573 on the 50k dataset; F1 score of 0.8534 on the full dataset). Therefore,

the needed variance to train a versatile detector seems to be captured by the 50k dataset already.
In a final step, we reduced the computational complexity of the classifier by reducing the number of network layers to the required minimum. While all other experiments have been carried out using 34-layer ResNets, we were able to reduce the number of layers to 18 using stochastic depth and dropout.
Exemplary results of the ResNet based distress detector are shown in Figure 10.

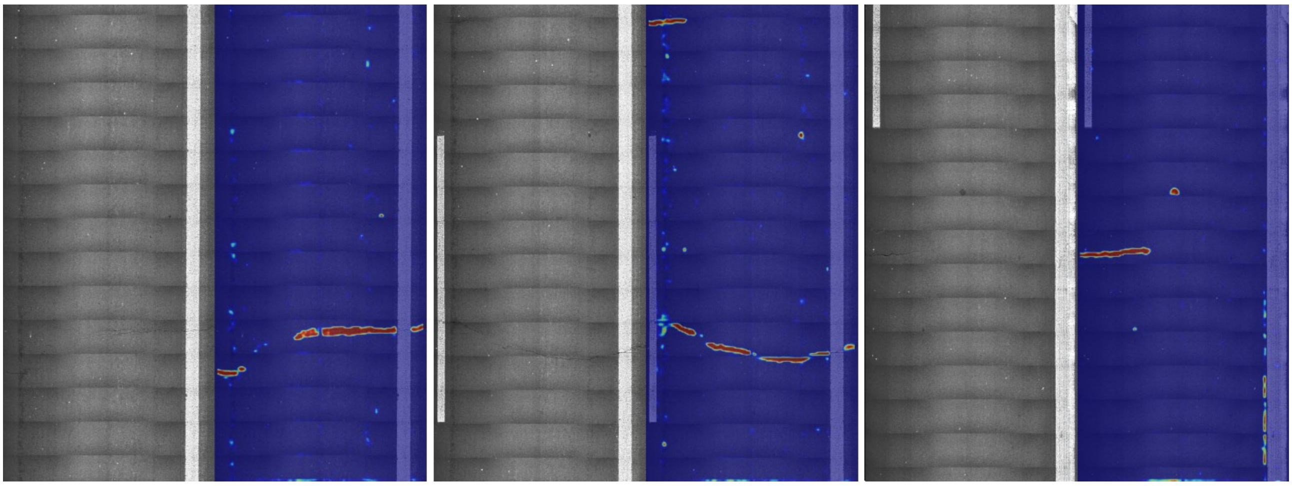

Figure 10 - Examples of distress detection. The detection results are colored using a blue to red color map with blue indicating no distress.

4.3. Embedding detection results into the RMA process

The detection results generated by our system are promising already. However, as we have shown in [26] the network output does not always resemble probability values, especially if the data differs significantly from the training data. Therefore, the decisions taken by the neural network must be reviewed by a human operator to guarantee high-quality standards. Operators can see the overlaid detection results as shown in Figure 10. Therefore, they can focus on regions where the network is probably uncertain, leading to a significant reduction in workload. Additionally, the detection results are transferred to the RMA grid based on thresholding (see Figure 11) such that operator just needs to change erroneous detection. Thus, the evaluation process is sped up.

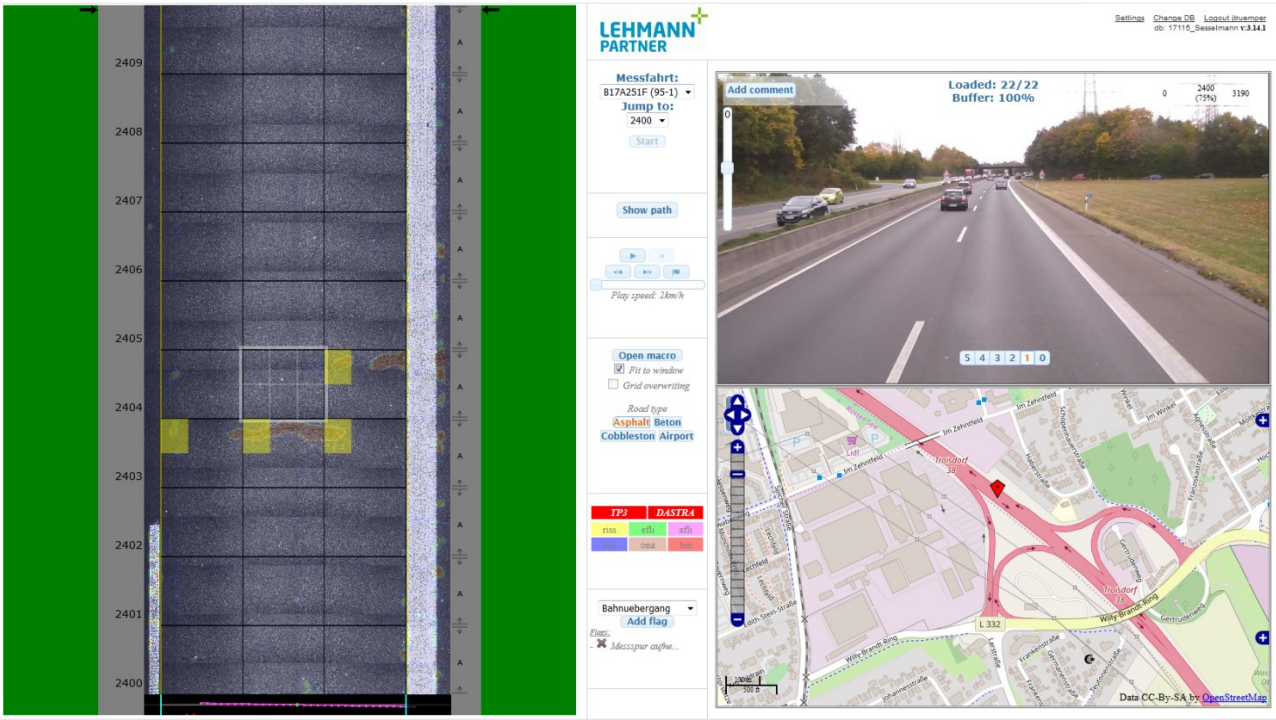

Figure 11 – Embedding of the distress detection result into the RMA process. The operator needs to review the proposed distress detections only. As usual, the operator has also access to the frontal image and the GPS position.

5. COMPARISON WITH HUMAN PERFORMANCE

Validation of the system is carried out by comparing the results of our system (not reviewed by an operator) with the results of an RMA process carried out by a skilled human operator. The test data comprises 2.5 kilometers of a single lane on a German freeway (reference track 1) and 5 kilometers of city roads (reference track 2).
Table 2 shows the high compliance of automated and human results. This shows that the human operator must intervene only rarely. Additionally, we observed that intact road is detected very securely. Taking the average number of 10-meter images with no distress at all into account, the number of images that can be skipped by a human operator is 30% for freeways and 14% for federal roads. Another speed-up is achieved by focusing on the highlighted distress regions only. In sum, the analysis process might be sped up by factor up to 2 based on limited observations that far.

Table 2 - Compliance of automated distress detection with human results measured by accuracy.

	reference track 1	reference track 2
cracks	93%	91%
inlaid patches	95%	96%
applied patches	99%	96%
potholes	98%	98%
open joints	96%	100%

6. CONCLUSION

Frequent road inspections and a timely assessment of the recorded data are the keys for efficient road maintenance. The proposed system can largely improve the time-consuming process of distress detection on recorded road images. To get a deep neural network capable of doing these analyses at human-like performance, we did extensive studies. First, we created a high-quality dataset. Then we evaluated several state-of-the-art deep neural networks for visual image analysis, by training them on our road dataset. Lastly, we derived our network architecture, which is both fast and accurate and outperforms classical machine learning and computer vision methods by far. Trials with operators have shown that a high-quality analysis process with humans in the loop can be sped up by factor up to 2. This does not only save human operation time but also improves the labeling quality by providing consistent detections as the automatic system does not suffer during the tedious process.

REFERENCES

[1] Eisenbach, M., Stricker, R., Seicher, D., Amende, K., Debes, K., Sesselmann, M., Ebersbach, D., Stöckert, U., Gross, H.-M. (2017). How to Get Pavement Distress Detection Ready for Deep Learning? A Systematic Approach. Int. Joint Conf. on Neural Networks (IJCNN), pp 2039-2047

[2] Lee, H. (1991). Application of machine vision techniques for the evaluation of highway pavements in unstructured environments. Proc. Fifth International Conference on Advanced Robotics' Robots in Unstructured Environments, pp 1425-1428

[3] Garbowski, T., Gajewski, T. (2017). Semi-automatic Inspection Tool of Pavement Condition from Three-dimensional Profile Scans. Procedia Engineering, Vol 172, pp 310-318

[4] Gopalakrishnan, K. (2018). Deep Learning in Data-Driven Pavement Image Analysis and Automated Distress Detection: A Review. Data, Vol 3, p 28

[5] Peng, L., Chao, W., Shuangmiao, L., Baocai, F. (2015). Research on Crack Detection Method of Airport Runway Based on Twice-Threshold Segmentation. 2015 Fifth International Conference on Instrumentation and Measurement, Computer, Communication and Control (IMCCC), pp 1716-1720

[6] Chambon, S., Moliard, J. M. (2011). Automatic road pavement assessment with image processing: Review and comparison. International Journal of Geophysics,

[7] Oliveira, H., Correia, P. L. (2014). CrackIT - An image processing toolbox for crack detection and characterization. IEEE International Conference on Image Processing (ICIP), pp 798-802

[8] Kaddah, W., Elbouz, M., Ouerhani, Y., Baltazart, V., Desthieux, M., Alfalou, A. (2018). Optimized minimal path selection (OMPS) method for automatic and unsupervised crack segmentation within two-dimensional pavement images. The Visual Computer, pp 1-17

[9] Fernandes, K., Ciobanu, L. (2014). Pavement pathologies classification using graph-based features. IEEE International Conference on Image Processing, pp 793-797

[10] Zou, Q., Cao, Y., Li, Q., Mao, Q., Wang, S. (2012). CrackTree: Automatic crack detection from pavement images. Pattern Recognition Letters, Vol 3, pp 227-238

[11] Kapela, R., Sniatala, P., Bloch, A., Atrem, S. A. (2015). Asphalt Surfaced Pavement Cracks Detection Based on Histograms of Oriented Gradients.

[12] Quintana, M., Torres, J., Menendez, J. M. (2016). A Simplified Computer Vision System for Road Surface Inspection and Maintenance. IEEE Transactions on Intelligent Transportation Systems, Vol 17, pp 608-619

[13] Varadharajan, S., Jose, S., Sharma, K., Wander, L., Mertz, C. (2014). Vision for road inspection. IEEE Winter Conference on Applications of Computer Vision, WACV 2014, pp 115-122

[14] Chatterjee, S., Saeedfar, P., Tofangchi, S., Kolbe, L. (2018). Intelligent Road Maintenance: A Machine Learning Approach for Surface Defect Detection. Proceedings of the 26th European Conference onn Information Systems (ECIS), pp 1-16

[15] Zhang, L., Yang, F., Zhang, Y. D., Zhu, Y. J. (2016). Road crack detection using deep convolutional neural network. IEEE International Conference on Image Processing (ICIP), pp 3708-3712

[16] Lecun, Y., Bottou, L., Bengio, Y., Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, Vol 86, pp 2278-2324

[17] Simonyan, K., Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. CoRR, Vol abs/1409.1556,

[18] Gopalakrishnan, K., Khaitan, S. K., Choudhary, A., Agrawal, A. (2017). Deep Convolutional Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection. Construction and Building Materials, Vol 157, pp 322-330

[19] Chatterjee, S. a. B. A. B. a. L. S. (2018). Smart Infrastructure Monitoring: Development of a Decision

Support System for Vision-Based Road Crack Detection. Proceedings of the International Conference on Information Systems - Bridging the Internet of People, Data, and Things, (ICIS),

[20] Anand, S., Gupta, S., Darbari, V., Kohli, S. (2018). Crack-pot: Autonomous Road Crack and Pothole Detection. arXiv preprint arXiv:1810.05107,

[21] Maeda, H., Sekimoto, Y., Seto, T., Kashiyama, T., Omata, H. (2018). Road damage detection and classification using deep neural networks with smartphone images. Computer-Aided Civil and Infrastructure Engineering, Vol 33, pp 1127-1141

[22] Yang, F. (2019). Feature Pyramid and Hierarchical Boosting Network for Pavement Crack Detection.

[23] Verkehrswesen, F. f. S.-. u. (2006). ZTV ZEB-StB - Zusätzliche Technische Vertragsbedingungen und Richtlinien zur Zustandserfassung und -bewertung von Straßen [FGSV-Nr. 489]. ISBN 3-939715-03-4, FGSV Verlag,

[24] Ng, A. (2016). Nuts and bolts of building AI applications using Deep Learning. NIPS,

[25] He, K., Zhang, X., Ren, S., Sun, J. (2016). Identity mappings in deep residual networks. European conference on computer vision, pp 630-645

[26] Seichter, D., Eisenbach, M., Stricker, R., Gross, H. M. (2018). How to Improve Deep Learning based Pavement Distress Detection while Minimizing Human Effort. IEEE Int. Conf. on Automation Science and Engineering (CASE), pp 63-68