ROAD CONDITION ASSESSMENT FROM AERIAL IMAGERY USING DEEP LEARNING

N. Merkle*, C. Henry, S. M. Azimi, F. Kurz

Remote Sensing Technology Institute, German Aerospace Center (DLR), Oberpfaffenhofen, Germany - nina.merkle@dlr.de; corentin.henry@dlr.de; seyedmajid.azimi@dlr.de; franz.kurz@dlr.de

Commission II WG 6

KEY WORDS: Aerial Images, Deep Learning, Road Condition Assessment, Image Segmentation, Crack Detection

ABSTRACT:

Terrestrial sensors are commonly used to inspect and document the condition of roads at regular intervals and according to defined rules. For example in Germany, extensive data and information is obtained, which is stored in the Federal Road Information System and made available in particular for deriving necessary decisions. Transverse and longitudinal evenness, for example, are recorded by vehicles using laser techniques. To detect damage to the road surface, images are captured and recorded using area or line scan cameras. All these methods provide very accurate information about the condition of the road, but are time-consuming and costly. Aerial imagery (e.g. multi- or hyperspectral, SAR) provide an additional possibility for the acquisition of the specific parameters describing the condition of roads, yet a direct transfer from objects extractable from aerial imagery to the required objects or parameters, which determine the condition of the road is difficult and in some cases impossible. In this work, we investigate the transferability of objects commonly used for the terrestrial-based assessment of road surfaces to an aerial image-based assessment. In addition, we generated a suitable dataset and developed a deep learning based image segmentation method capable of extracting two relevant road condition parameters from high-resolution multispectral aerial imagery, namely cracks and working seams. The obtained results show that our models are able to extraction these thin features from aerial images, indicating the possibility of using more automated approaches for road surface condition assessment in the future.

1. INTRODUCTION

Road condition assessment is carried out in several countries to maintain an acceptable level of quality for the transportation system. For instance, since 1990, the German Federal Ministry of Transport and Digital Infrastructure has carried out a condition survey and assessment (ZEB) for the road surfaces of all federal roads in Germany[1]. To describe and document the condition of the roads, a set of relevant parameters has been defined and depending on the construction method of the road (asphalt or concrete) different sets of parameter are used. To describe the condition of asphalted roads, the following parameters are used and acquired by the ZEB (all definitions are taken from here[2]):

- **Mesh cracks, crack clusters and single cracks:** Longitudinal and transverse cracks are fine to gaping fractures in the asphalt slabs that do not occur exclusively in the immediate vicinity of the slab corners or edges. Open and cast/sealed cracks are considered equally.
- **Working seams:** Open gaps form along transverse or longitudinal seams (also along inlaid patches). Open/repaired working seams are characterized by their linear course.
- **Inlaid patches:** Inlaid patches can be recognized by the neatly cut or milled edges and the dark traces of edge sealing which is included in the patch. In addition to the mostly manually applied and sometimes unevenly shaped repair patches, there are also machine-applied, mostly rectangular patches and the sealants or grip-improving surface treatments, some of which are applied over large areas in the patching process.

Figure 1. Image samples of the chosen asphalted road parameters. Class one ”cracks ”(first row) and class two ”working seams” (last row).

- **Applied patches:** Applied patches are laid on the existing road surface. When the patches have been laid, a part of the road surface is first removed and then filled with asphalt flush with the old road surface.
- **Excavation/pothole:** Excavations mark punctiform or extensive areas in which parts of the surface course, binder course or base course have broken off or become detached.
- **Binder enrichment:** In the case of unfavorable mixing ratios (too much bitumen or too little binding aggregate) or unsuitable materials, segregation can occur, resulting in binder leaking from the surface of the pavement.

Commonly, terrestrial sensors such as laser scanners or line scan cameras are used for the inspection of the roads and for the detection of the above listed parameters. These conventional techniques provide very accurate information about the condition of the road and even millimeter-sized changes in the road surface or objects can be detected. A drawback of terrestrial sensors is that they are time-consuming and costly. In this

* Corresponding author

[1] https://www.bast.de/BASt_2017/DE/Strassenbau/Fachthemen/gs4-zeb.html

[2] https://itzeb.heller-ig.de/leitfaden/index.html

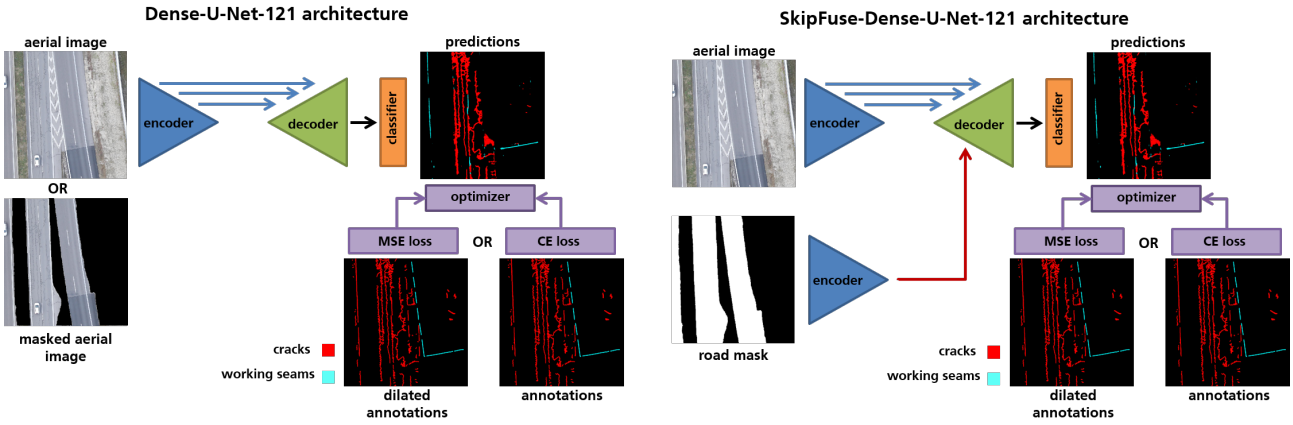


Figure 2. Simplification of the Dense-U-Net-121 (left side) and SkipFuse-Dense-U-Net-121 architecture (right side).

work, we aim to investigate the applicability of aerial imagery in combination with deep learning methods for an automatic and large-scale assessment of the road condition. Since not every parameter captured by terrestrial sensors is visible in aerial images, we focus on two of the above described parameters in this paper, more precisely on the classes "mesh cracks, crack clusters, single cracks" and "working seams". Since we are working on images with a ground sampling distance (GSD) of up to 10 cm, the distinction between sealed or unsealed cracks or working seams is sometimes not possible (see samples in Figure 1 for illustration). Therefore, we adjust the above listed parameters and divide them into the following two classes: **1) cracks**: open and sealed mesh cracks, crack clusters and single cracks and **2) working seams**: open and sealed working seams.

The extraction of thin structures with complex topologies has been the focus of two research tracks from the fields of remote sensing and medical imagery. Whether they be roads (Zhang et al., 2018, Mosinska et al., 2020, Mosinska et al., 2018, Henry et al., 2021a), pavement cracks (Mosinska et al., 2018), retinal blood vessels (Gu et al., 2019, Zhang et al., 2021) or cell membranes (Ronneberger et al., 2015, Mosinska et al., 2018, Gu et al., 2019), some objects require carefully designed segmentation architecture to be accurately extracted (i.e. completely and without spurious predictions) due to their size, ambiguous or noisy neighborhood and partial to total occlusion by other nearby objects. Since the introduction of the first fully-convolutional network (FCN) in (Long et al., 2015), researchers have mostly converged towards the U-Net architecture (Ronneberger et al., 2015) which presents important advantages in the context of the above-mentioned tasks: 1) it is lightweight compared to most models used in common imagery and is therefore fitting for training on smaller datasets, 2) it achieves remarkable performance (Zhou et al., 2018) with simple backbones like ResNet (He et al., 2016) or DenseNet (Huang et al., 2017), and 3) its overall architecture can be easily modified to suit the needs of specific tasks by densifying its skip connections (Yu et al., 2018) or bridge layers (Zhou et al., 2018).

Following this trend, we adopt two U-Net-based networks for the extraction of cracks and working seams, namely Dense-U-Net (Henry et al., 2021a) and SkipFuse-Dense-U-Net (Henry et al., 2021b), which already proved to be effective on aerial imagery analysis tasks like road and parking area segmentation. We adopt these networks for the task of extraction thin structures and investigate the influence of targeted training by using road masks. As most of the common approaches are utilizing mobile mapping systems (Stricker et al., 2019), smart phones (Maeda et al., 2018, Varadharajan et al., 2014) or very high resolution drone data (Pan et al., 2018), we generated a new training data set consisting of 132 labeled high resolution aerial images acquired over Germany. By training the networks on our dataset, we show that the extraction of cracks and working seams from aerial images with a ground sampling distance (GSD) of up to 10cm with this dataset is feasible. The presented results show the good performance of our methods especially with the existing diversity in the image locations, sensors and the varying illumination conditions in our dataset.

2. METHODS

We implement two fully-convolutional architectures for the automatic extraction of cracks and working seams from areal images, called Dense-U-Net (Henry et al., 2021a) and SkipFuse-Dense-U-Net (Henry et al., 2021b). These networks proved to be effective on aerial imagery analysis tasks like road and parking area segmentation. Since our study focuses on the extraction of very thin objects (compared to roads and parking areas) and with a known location (on top of the road surface), we realized some adjustments to the original methods.

Segmentation networks and input masking: As aerial images exhibit visual features outside from the road surface that could naively be assimilated to crack-like or seam-like features, we investigate a strategy to prevent the prediction of false positives on undesired locations. We therefore experiment with networks taking in masked and unmasked input images so that we can systematically filter the regions of interest. To obtain these binary filter for trafficable surfaces, we apply the SkipFuse-Dense-U-Net model from (Henry et al., 2021b) trained on the parking area detection dataset, presented in the same paper, with auxiliary OSM input and retain only the segmented roads. Such masks, while a good starting point for searching for signs of road cracks and seams, are however not perfect, we must expect some of our objects of interest to lie outside their boundaries. One solution to overcome this problem is to use the mask not as a filter, rather as a secondary input to a network capable of data fusion like SkipFuse-Dense-U-Net, to both make the extraction easier on already detected road areas while not arbitrarily preventing predictions in all other areas. This leads us to proposing three approaches to our task at hand: (1) Dense-U-Net-121 on masked images, (2) Dense-U-Net-121 on unmasked images, (3) SkipFuse-Dense-U-Net-121 on unmasked

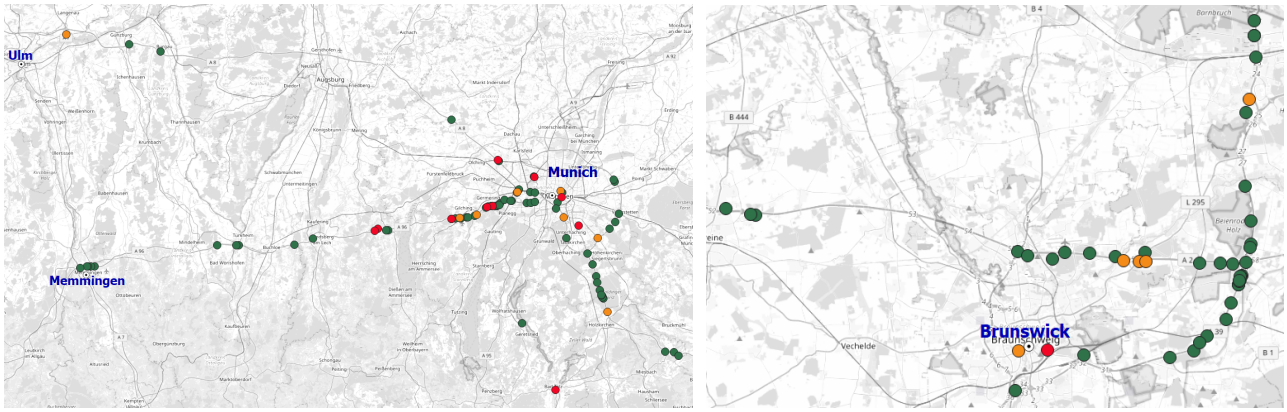


Figure 3. Illustration of the distribution of the aerial images within the training (■), validation (■) and test (■) sets around the German cities of Munich, Memmingen, Ulm (left) and Brunswick (right).

images but with road masks as additional input (cf. Fig. 2 for an overview). More details about the architectures of the two networks can be found in their respective papers.

Re-balancing the loss: Cracks and seams are thin objects that often leads either to an increased difficulty in training a segmentation model, or to low probability scores in the output, necessitating careful post-processing steps. To overcome this, we follow the suggestion from (Homayounfar et al., 2017) and (Henry et al., 2018) to create a smooth ground truth instead of a binary one: for each class separately, a Gaussian kernel is applied as a 2D convolution to dilate a binary mask and convert its values from $0, 1$ to a smooth gradient in $[0, 1]$ around its borders, with value closer to 1 as pixels are closer to the boundary and conversely, with object values preserved as 1. Given a smoothing radius R in pixels, we chose to use a Gaussian kernel of size $[2 * R + 1, 2 * R + 1]$ with a standard deviation of $\frac{R}{3}$. After the convolution, the values greater than 1 are thresholded to 1. To train the model on this new regression task, we use an Mean Squared Error loss (MSE) on the cracks and seams classes but not the background class, weighted with a coefficient of 20 to counter-balance the low frequency of both classes compared to the background's.

3. DATASET

As there is no public dataset available to train our segmentation segmentation model, we created a new image dataset, which contains 132 labeled aerial images acquired by the 3K (Kurz et al., 2012), 4K (Kurz et al., 2014) and MACS (Brauchle et al., 2019) camera system of the German Aerospace Center (DLR) over Germany. Care was taken in the selection of the images to create variation in terms of sun position, cloud cover, weather (dry/wet), GSD, season, road categories, and scenery. The selected images have a GSD between 2-10 cm and are acquired between the years 2015 to 2020 during different seasons with varying sun angles. Some statistical characteristics of the dataset are provided in Figure 4. The image sizes vary between 4864×3232 px and 5616×3744 px. Most images were acquired outside urban areas and contain motorways and main roads, but urban and ancillary roads are also covered. An overview of the image locations around the German cities of Munich, Memmingen, Ulm and Brunswick is provided in Figure 3.

As described in previous sections, we defined the classes "cracks" (sealed and unsealed) and "working seams" (sealed and unsealed). For each image a manual annotation was carried out, where only visible areas were annotated. For this process, an annotation policy was generated where the characteristics of each class was defined. In most cases cracks and working seams can be differentiated by their shape (working seams are commonly straight) but the main criteria we applied was the following: the feature "workings seams" stands more for quality problems in the manufacturing process rather than for a weakness of the substance as in the case of cracks; Working seams are man-made whereas cracks appear over time due to the weakness of the structure. In the end, a multi-level quality check was performed by experts to ensure the quality of the dataset.

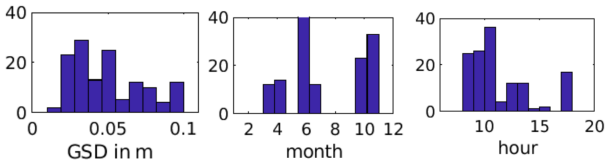


Figure 4. Overview of statistical characteristics of the data set: (a) distribution of GSD, (b) distribution of images through the year and (c) distribution of images within the day.

In order to train, validate and test our neural networks, we divided the 132 images into three disjoint sets: 1) the training set consisting of 105 images, 2) the validation set consisting of 15 images and 3) the test set consisting of 12 images. Some examples of the annotations for the training dataset is displayed in Figure 5. The spatial distribution of each image within the three sets is illustrated in Figure 3. Note that the images from the three sets do not overlap even though the location of images appear to be quite close in the figure.

4. RESULTS & DISCUSSION

The various networks described in Section 2 are trained for 50 epochs, with a patch size of 512×512 px, an Adam optimizer and a learning rate of 10^{-4}. We performed experiments using two different loss functions: 1) a cross-entropy (CE) loss when dealing with the original GT and 2) a smooth MSE loss when using the soft labels (see Section 2). The MSE loss is used together with the soft labels as it typically shows a better performance for regression tasks compared to the CE loss. For the

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-V-2-2022-283-2022 | © Author(s) 2022. CC BY 4.0 License.

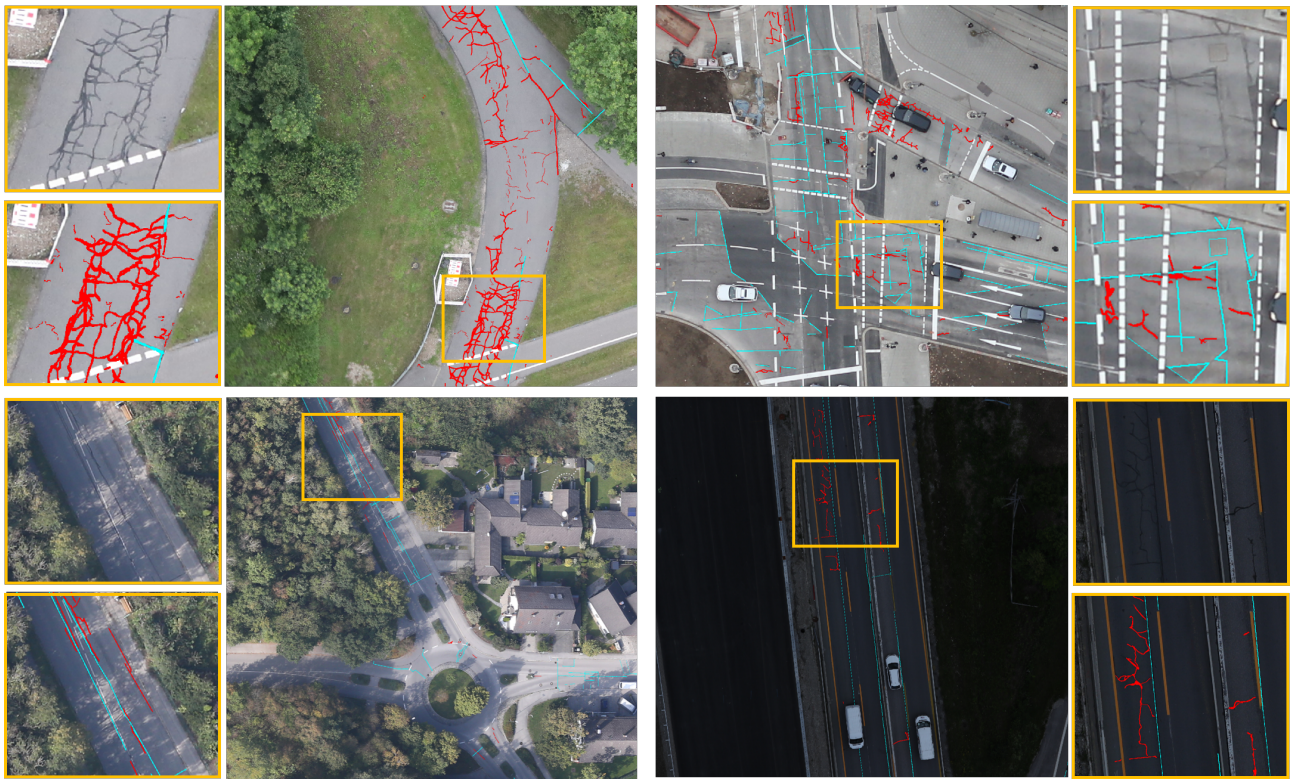


Figure 5. Illustration of the dataset with overlaid annotations for the two semantic classes: ■ cracks and ■ working seams.

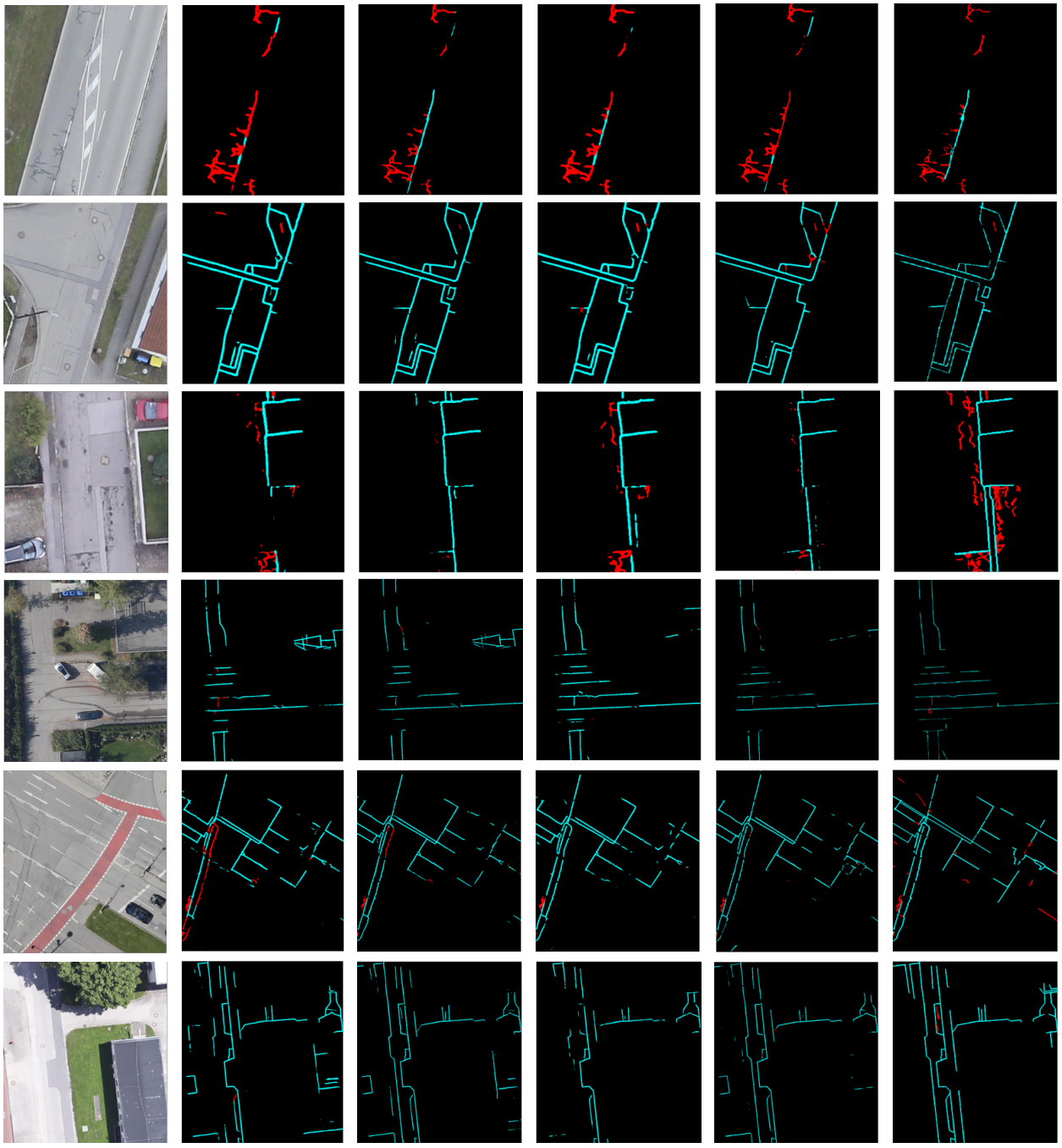
model #	network architecture	road mask?	small training set?	loss function	width dilation	IoU [%] mean	IoU [%] cracks	IoU [%] seams	average [%] recall	average [%] precision
1	Dense-U-Net	-	-	CE	-	32.27	39.87	24.66	65.83	32.52
2	Dense-U-Net	✓	✓	CE	-	28.50	30.94	26.06	**67.60**	32.57
3	Dense-U-Net	✓	✓	MSE	3	33.15	36.02	30.28	59.38	43.54
4	Dense-U-Net	-	-	MSE	3	26.05	41.65	30.44	53.67	51.89
5	Dense-U-Net	-	✓	MSE	3	36.45	42.38	30.52	59.74	47.88
6	Dense-U-Net	-	✓	MSE	2	38.73	45.61	**31.85**	52.26	59.28
7	Dense-U-Net	-	✓	MSE	1	26.95	38.10	15.80	29.72	69.84
8	SkipFuse-Dense-U-Net	✓	-	CE	-	28.51	35.51	21.51	66.66	33.48
9	SkipFuse-Dense-U-Net	✓	-	MSE	3	34.07	40.70	27.44	48.46	53.34
10	SkipFuse-Dense-U-Net	✓	✓	MSE	3	37.19	45.31	29.06	53.37	52.69
11	SkipFuse-Dense-U-Net	✓	✓	MSE	2	**39.00**	**46.82**	31.17	48.57	65.18
12	SkipFuse-Dense-U-Net	✓	✓	MSE	1	28.84	39.50	18.18	30.49	**76.36**

Table 1. Quantitative comparison of the Dense-U-Net and the SkipFuse-Dense-U-Net architecture using different input data (with road masks vs. without road masks), different training datasets (usage of full training images vs. only patches close to roads) and different loss functions (CE vs. MSE with different dilation radius).

smooth MSE loss we considered a radius between 1-3 pixels for the Gaussian kernel (width of the dilation). When training the Dense-U-Net models a batch size of 12, while for the SkipFuse-Dense-U-Net a batch size of 9 was used. The weights of the RGB encoders are initialized with weights pre-trained on ImageNet, whereas the weights of all other layer are initialized randomly with a Xavier uniform distribution and their biases are initialized to 0. The training of all models was performed over the training dataset (see Section 3). As our training images contain large areas without any crack or working seam, we investigated the influence of two setups: 1) training over the whole image data and 2) training over a smaller training dataset where all patches with no overlap with the corresponding roads masks are discarded.

Table 1 provides on overview of the results obtained by applying all trained models (from epoch 50) on the test set images. Overall, we achieved the best results (in terms of IoU) for the SkipFuse-Dense-U-Net with a MSE loss and a dilation width of 2. For the class "working seams" slightly better results can be achieved with the Dense-U-Net architecture using the same loss and dilation width. The usage of the road masks together with the Dense-U-Net architecture does not lead to performance improvements. A possible explanation is that trough the usage of the masked RGB images as input data, the network is lacking a bit of context compared to the other setups. However, using the road masks to compute the smaller training dataset (here all patches without any overlap with the road mask have been discarded) lead to a performance boost for the Dense-U-Net and SkipFuse-Dense-U-Net models. By comparing the loss functions, it can be observed that the smooth MSE loss has some advantages over the CE loss. This could be explained by the thickness of our objects. Since cracks and working seams are relatively thin objects, the use of a smooth loss can support the

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-V-2-2022-283-2022 | © Author(s) 2022. CC BY 4.0 License.



(a) aerial image (b) model #5 (c) model #6 (d) model #10 (e) model #11 (f) ground truth

Figure 6. Qualitative of the two best performing Dense-U-Net and the two best performing SkipFuse-Dense-U-Net architectures. Details about the different models can be found via the model number in Table1. Categories colors are: ■ cracks and ■ working seams.

learning process of the network by penalizing correct predictions that have a small offset from ground truth less severely.

Qualitative results of the two best performing Dense-U-Net and the two best performing SkipFuse-Dense-U-Net architectures are shown in Figure 6. These image samples show the good performance of all models for various scenes but also reveal advantages and disadvantages of the different training setups. Using a smooth loss with dilation width of 3 pixel lead in general to thicker predictions (see second and fourth column of Figure 6), but on the other hand to more complete predictions. The usage of the road mask as an addition input branch in the SkipFuse-Dense-U-Net architectures has the advantage of less incorrect prediction outside the road areas. This can be seen in the fourth and sixth row of Figure 6, where both Dense-U-Net models (trained without the road masks) show several incorrect predictions on a sport field and a roof top. Overall, the models learned to differentiate well between the classes ”cracks” and ”workings seams” and between these classes and similar looking features such as curbstones.

As working seams and cracks are relatively thin objects we investigated in another experiment the influence of a tolerant IoU on the performance of our models. The idea of this tolerant IoU is to define tolerance or buffer areas of certain sizes around the labels of each class. In these buffered areas, all predictions that

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-V-2-2022-283-2022 | © Author(s) 2022. CC BY 4.0 License.

model #	network architecture	road mask?	width dilation	IoU [%] cracks	IoU [%] seams	1px-IoU [%] cracks	1px-IoU [%] seams	2px-IoU [%] cracks	2px-IoU [%] seams	3px-IoU [%] cracks	3px-IoU [%] seams
5	Dense-U-Net	-	3	42.38	30.52	55.81	**39.97**	61.93	**44.16**	**63.98**	**45.58**
6	Dense-U-Net	-	2	45.61	**31.85**	52.65	37.81	54.62	39.58	55.35	40.11
10	SkipFuse-Dense-U-Net	✓	3	45.31	29.06	**57.73**	37.58	**62.28**	40.85	63.55	41.85
11	SkipFuse-Dense-U-Net	✓	2	**46.82**	31.17	54.41	36.04	56.28	37.33	57.01	37.71

Table 2. Quantitative comparison of the two best performing Dense-U-Net and the two best performing SkipFuse-Dense-U-Net architectures using a tolerant IoU. All models are trained with a smooth MSE loss and on the smaller training dataset containing only patches that overlap with the road mask.

are actually correct, but whose location differs from the ground truth by a few pixels, are still considered as correct. For example, if we have predictions for the class "crack", which are lying within the buffer areas of the labels for the class "crack", they will be counted as true positives. If we have predictions for the class "working seams" or the background class within the buffer areas of the class "cracks", they will be counted as true positives as well. The advantages of such a tolerant IoU is that we obtain a better understanding of the location of the incorrect predictions. Otherwise it is difficult to understand from the numbers, if the incorrect predictions area only a couple of pixel next to the ground truth or at completely different locations within the image.

The implementation of the tolerant IoU was realized through a buffered version of the ground truth. The larger the buffer around the ground truth, the more tolerant the metric is to correct predictions close to the ground truth. To create a buffered version of the ground truth, we apply a similar dilation process as for dilating the labels during the training, but to allow for more control on the resulting values, we apply a different algorithm. Based on the original mask, we apply successively three morphological dilation operations with a 3×3 kernel and a cross pattern (the so-called 4-connectivity pattern). The four masks are merged by summation, i.e. the original masks holds values 4, the pixels at an L_1-distance of 1 have values 2, at an L_1-distance of 2 values 1, and from L_1-distance of 3 or more values 0. The computation of the buffered ground truth for a given tolerance radius R is done by thresholding the values greater or equal to $(3 - R)$ to 1, and setting the rest of the values to 0.

Quantitative results of the two best performing Dense-U-Net and the two best performing SkipFuse-Dense-U-Net architectures using the described tolerant IoU are provided in Table 2. For the four models, we computed the IoU and compared it to a 1-, 2- and 3-pixel tolerant IoU. For all models a gain of around 10% in the IoU can be achieved by applying the 3-pixel IoU. This shows that our models tend to predict the objects thicker than the actually are. The biggest changes in the numbers can be observed for the models, which were trained with a smooth loss and a dilation width of 3. This can be explained by the fact that these models are less penalized during training if they provide wider predictions.

5. CONCLUSION & FUTURE WORK

In this paper, we explore the possibility of using neural networks for road condition assessment. As a first step towards this direction, we focused on an automatic extraction of two relevant objects for describing the condition of asphalted road surfaces, namely cracks and working seams. To train and test our models, we generated a new dataset consisting of 132 labeled aerial images having a GSD between 2-10 cm. Since our goal is to extract relatively thin objects with a known position (on the road surface), we adapted two network architectures commonly used in remote sensing. More specifically, we tested the use of a smooth mean square error loss and a road mask (previously extracted from the aerial imagery) as additional input. The obtained results show that state-of-the-art segmentation models are capable of extracting thin objects such as cracks (46.82% IoU and 63.98% for our 3-pixel IoU) and working seams (31.85% IoU and 45.58% for our 3-pixel IoU) from aerial images and therefore have a high potential for a fast and large-scale assessment of road surfaces. In the future, the method will be tested on other areas within and outside Germany and extended to include other relevant objects for describing road conditions such as potholes, applied and inlaid patches. Additionally, the influence of the ground sampling distance as well as the acquisition time and weather or seasonal differences on the quality of the results will be investigated.

REFERENCES

Brauchle, J., Bayer, S., Hein, D., Berger, R., Pless, S., 2019. MACS-Mar: A Real-time Remote Sensing System for Maritime Security Applications. *CEAS Space Journal*, 11(1), 35–44.

Gu, Z., Cheng, J., Fu, H., Zhou, K., Hao, H., Zhao, Y., Zhang, T., Gao, S., Liu, J., 2019. CE-Net: Context Encoder Network for 2D Medical Image Segmentation. *IEEE Transactions on Medical Imaging*, 38(10), 2281–2292.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. *CVPR*, 770–778. arXiv: 1512.03385 Citation Key: He2016 container-title: NV.

Henry, C., Azimi, S. M., Merkle, N., 2018. Road Segmentation in SAR Satellite Images with Deep Fully-Convolutional Neural Networks. *IEEE Geoscience and Remote Sensing Letters*, 15, 1867–1871. arXiv: 1802.01445 Citation Key: Henry2018.

Henry, C., Fraundorfer, F., Vig, E., 2021a. Aerial road segmentation in the presence of topological label noise. *ICPR*, 2336–2343.

Henry, C., Hellekes, J., Merkle, N., Azimi, S., Kurz, F., 2021b. Citywide estimation of parking space using aerial imagery and osm data fusion with deep learning and fine-grained annotation. *ISPRS 2021*, XLIII, 479–485.

Homayounfar, N., Fidler, S., Urtasun, R., 2017. Sports field localization via deep structured models. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4012–4020. Citation Key: Homayounfar2017 container-title: HI.

Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K. Q., 2017. Densely connected convolutional networks. *CVPR*, 2261–2269. arXiv: 1608.06993 Citation Key: Huang2017 container-title: HI.

Kurz, F., Rosenbaum, D., Meynberg, O., Mattyus, G., Reinartz, P., 2014. Performance of a real-time sensor and processing system on a helicopter. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 189-193.

Kurz, F., Türmer, S., Meynberg, O., Rosenbaum, D., Runge, H., Reinartz, P., Leitloff, J., 2012. Low-cost Systems for Real-time Mapping Applications. *Photogrammetrie Fernerkundung Geoinformation*, 159–176.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. *CVPR*, 3431–3440. arXiv: 1411.4038 Citation Key: Long2015 container-title: MA.

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*, 33(12), 1127-1141.

Mosinska, A., Koziński, M., Fua, P., 2020. Joint Segmentation and Path Classification of Curvilinear Structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(6), 1515-1521.

Mosinska, A., Marquez-Neila, P., Kozinski, M., Fua, P., 2018. Beyond the pixel-wise loss for topology-aware delineation. *CVPR*, 3136–3145. arXiv: 1712.02190 Citation Key: Mosinska2017 container-title: UT.

Pan, Y., Zhang, X., Cervone, G., Yang, L., 2018. Detection of Asphalt Pavement Potholes and Cracks Based on the Unmanned Aerial Vehicle Multispectral Imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, PP, 1-12.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical. N. Navab, J. Hornegger, W. M. Wells, A. F. Frangi (eds), *MICCAI*, Springer International Publishing, 234–241. Citation Key: Ronneberger2015.

Stricker, R., Eisenbach, M., Sesselmann, M., Debes, K., Gross, H.-M., 2019. Improving visual road condition assessment by extensive experiments on the extended gaps dataset. *2019 International Joint Conference on Neural Networks (IJCNN)*, 1–8.

Varadharajan, S., Jose, S., Sharma, K., Wander, L., Mertz, C., 2014. Vision for road inspection. *IEEE Winter Conference on Applications of Computer Vision*, 115–122.

Yu, F., Wang, D., Shelhamer, E., Darrell, T., 2018. Deep layer aggregation. *CVPR*, 2403–2412. Citation Key: Yu.

Zhang, J., Zhang, Y., Xu, X., 2021. Pyramid u-net for retinal vessel segmentation. *ICASSP*, 1125–1129.

Zhang, Z., Liu, Q., Wang, Y., 2018. Road Extraction by Deep Residual U-Net. *IEEE Geoscience and Remote Sensing Letters*, 15(5), 749–753. Citation Key: Zhang2017.

Zhou, L., Zhang, C., Wu, M., 2018. D-linknet: Linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction. *CVPR Workshops*. Citation Key: Zhou2018.