CITYWIDE ESTIMATION OF PARKING SPACE USING AERIAL IMAGERY AND OSM DATA FUSION WITH DEEP LEARNING AND FINE-GRAINED ANNOTATION

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

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

Commission III

KEY WORDS: Deep Learning, Aerial Imagery, Image Segmentation, Parking Space Management, OpenStreetMap

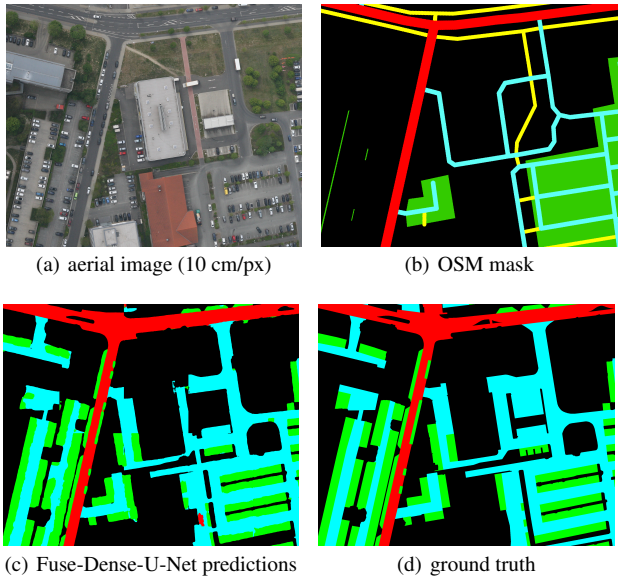
ABSTRACT:

Emerging traffic management technologies, smart parking applications, together with transport researchers and urban planners are interested in fine-grained data on parking space in cities. However, there are no standardized, complete and up-to-date databases for many urban areas. Moreover, manual data collection is expensive and time-consuming. Aerial imagery of entire cities can be used to inventory not only publicly accessible and dedicated parking lots, but also roadside parking areas and those on private property. For a realistic estimation of the total parking space, the observed use of multi-functional traffic areas is taken into account by segmenting not only parking areas but also roads according to their purpose. In this paper, different U-Net based architectures are tested for detecting all these types of visible traffic areas. A new large-scale, high-quality dataset of manual annotations is used in combination with selected contextual information from OpenStreetMap (OSM) to depict the actual use as parking space. Our models achieve a good performance on parking area segmentation, and we show the significant impact of OSM data fusion in deep neural networks on the simultaneous extraction of multiple traffic areas compared to using aerial imagery alone.

1. INTRODUCTION

Accurate information on parking spaces is nowadays relevant for parking guidance systems as well as for traffic management and urban planning. The importance of this data is increasing in the context of intelligent transportation systems, thereby enabled value-added services and autonomous driving. In transportation research, the relevance of parking for citywide traffic is recognized but rarely addressed (Habib et al., 2012). This is mainly caused by the insufficient data basis: the example of Germany shows that there is no standardized, up-to-date and comprehensive database even for the three largest cities (Senate of Hamburg, 2020, State capital of Munich, 2019, Senate of Berlin, 2014). Especially the latter aspect must be emphasized, since despite a considerable share of private parking spaces, only those on public property are covered. Information is partially available for managed parking lots, but this reflects only a fraction. Previous research has focused primarily on determining the occupancy level of designated, large parking areas using deep learning methods on data from surveillance cameras (Amato et al., 2017), drones (Fraunhofer IAO, 2021) or satellites (Drouyer, 2020). Existing datasets with a high ground sampling distance (GSD) deal with the segmentation of parking spaces in addition to a variety of other classes using deep learning (Zhou et al., 2018b, Cheng et al., 2017). One annotation dataset on aerial imagery and the corresponding Fully-Convolutional neural Network (FCN) also separates non-paved parking places (Azimi et al., 2019). A significant category of parking areas has not been considered so far, which requires a new approach for semantic scene understanding: dual-use areas in backyards, on the roadside and on sidewalks that are regularly used for parking, although no markings are visible on aerial imagery.

The state-of-the-art methods for object segmentation in aerial imagery is based on FCNs (Long et al., 2015), and recently more specifically on the U-Net architecture (Ronneberger et al., 2015). This type of networks is particularly adapted to extracting high levels of details in high-resolution imagery, thanks to the skip-connections linking each encoder layer to the corresponding decoder layer, have therefore been successful in a variety of binary object segmentation challenges: for road segmentation (Zhou et al., 2018a, Buslaev, 2017) and building segmentation (Hamaguchi and Hikosaka, 2018, Lindenbaum, 2017) in the respective DeepGlobe18 (Demir et al., 2018) and SpaceNet



(a) aerial image (10 cm/px) (b) OSM mask

(c) Fuse-Dense-U-Net predictions (d) ground truth

Figure 1. Illustration of the dataset and the results: (a-b) input data, (c) predictions of our best model, (d) manually annotated ground truth. Categories colors are: parking area, road, access way and pedestrian/bike way.

* Corresponding author

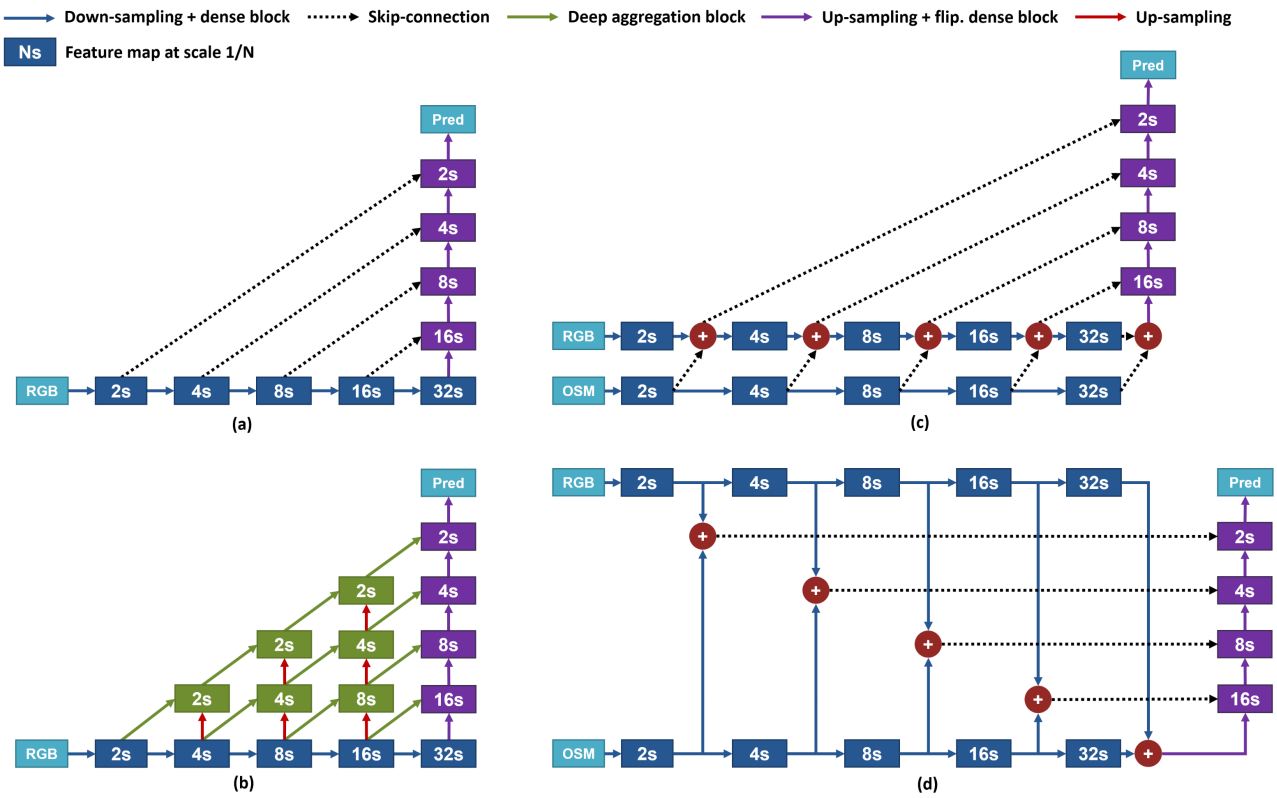


Figure 2. An overview of the different fully-convolutional neural network architectures used in our study: (a) Dense-U-Net, (b) DLA-Dense-U-Net, (c) Fuse-Dense-U-Net, (d) SkipFuse-Dense-U-Net. RGB, OSM and Pred. are described in Figure 1.

challenges (Etten et al., 2019). While complex architectures specialized for a wide variety of classes such as DeepLabV3+ (Chen et al., 2018) are especially effective on ground imagery datasets like the CityScapes (Cordts et al., 2016) or PascalVOC (Everingham et al., 2010), their spatial accuracy is lacking for remote sensing imagery. More advanced architectures were designed for fine-grained semantic extraction in aerial imagery, like SkyScapesNet (Azimi et al., 2019), but they require large amount of memory to run which prevents their conversion into fusion networks.

In addition to the above mentioned approaches, where the input data comes from a single modality, several studies investigated the fusion of multi-modal and multi-temporal data via neural networks (Chlaily et al., 2020, Hong et al., 2020). Thereby, special attention was paid to the way the data was fused within the networks. In (Hong et al., 2020), the impact and differences between shallow and deep models for the multi-model image classification have been investigated. In (Merkle et al., 2019) the benefits of combining RGB, near infrared (NIR) and thermal infrared (TIR) aerial images for the task of semantic vehicle segmentation and the influence of an early or late fusion within the network have been researched. Instead of fusing data from different sensors only, (Audebert et al., 2017) investigated the inclusion of highly processed and semantically richer data. More specifically, they tested different network architectures to explore the utility of OpenStreetMap (OSM) in combination with RGB data for semantic labeling.

In this study we perform parking area extraction using first a base architecture, Dense-U-Net (Henry et al., 2020), then we improve its fine-grained details recovery capability by introducing DLA-Dense-U-Net, following the Deep Layer Aggregation (DLA) technique (Yu et al., 2018). Finally we test and improve the OSM fusion performance of FuseNet (Hazirbas et al., 2016) with our architectures SkipFuse-Dense-U-Net and SkipFuse-DLA-Dense-U-Net. To train these models, we used a new high quality annotation dataset that supports the semantic understanding of both dedicated and regularly used parking areas in contrast to roads and access ways. We show that such fine-grained annotations and extraction methods are yielding excellent results, especially considering the diversity of the urban environment density and the varying illumination conditions in our imagery. Future work should validate the generalization capability such models on imagery from other cities featuring different parking area topologies.

2. METHODS

For the automatic extraction of traffic areas, we implement a fully-convolutional architecture derived from U-Net (Ronneberger et al., 2015), namely Dense-U-Net (Henry et al., 2020) and its DLA variant (Yu et al., 2018). While U-Net is the preferred backbone by the remote sensing community for semantic segmentation tasks thanks to its effectiveness in recovering fine-grained spatial details, its backbone is lacking many features from state-of-the-art architectures and compatible pre-trained weights are rarely available. Therefore we selected the Dense-U-Net architecture with a DenseNet-121 backbone for the following reasons: its densely connected layers allow for a faster training, it has a higher capacity for learning complex semantics, and most deep learning libraries provide weights pre-trained on the ImageNet dataset. Contrary to other U-Net-derived architectures, Dense-U-Net does not use a simple decoder composed only of successive deconvolutions, but rather

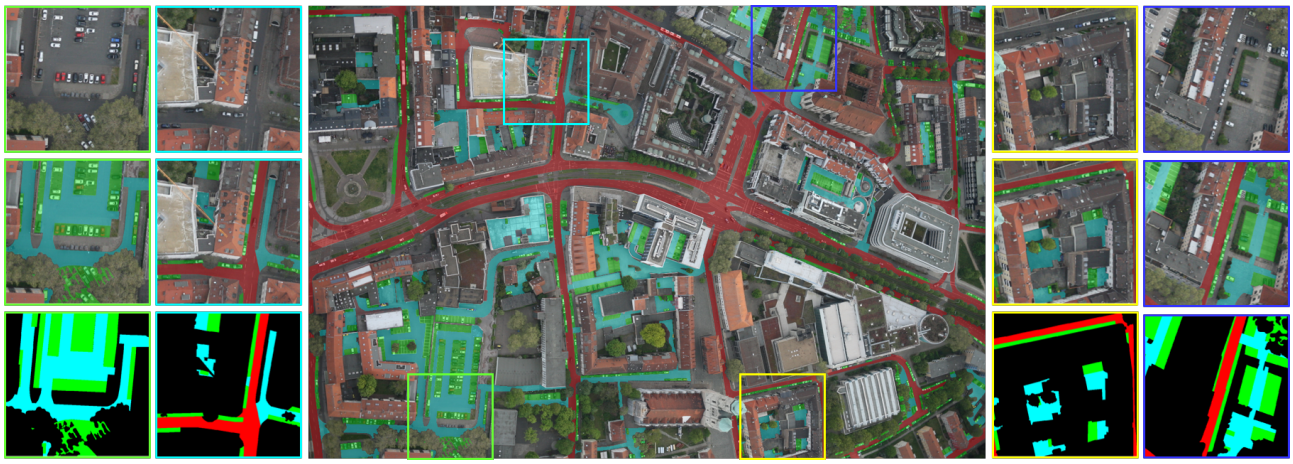


Figure 3. Zoomed-in samples of aerial images, annotations and overlay from our dataset with the three classes: parking area, road and access way.

a flipped DenseNet mirroring the layers from the encoder at the corresponding pooling level (see Figure 2.a). This helps extracting more spatial and semantic information from the low-level and high-level layers respectively. To further improve this aspect, we also bring in the technique called DLA, which densifies the U-Net's skip-connections. In short, it nests U-Nets at each down-sampling level, so that fine-grained details are recovered more progressively than with direct skip-connection (see Figure 2.b). In the following, we are calling this architecture DLA-Dense-U-Net.

Additionally, we intend to leverage as much information as possible from existing data sources like OSM. Although sometimes spatially and semantically inaccurate, these provide a good baseline for distinguishing roads, access ways and parking lots. We extract vector information from 7 traffic-related object categories and rasterize it into classes from 0 to 6: drivable and non-drivable ways, access ways, parking spaces, gas stations, bicycle parking and parking vending machines. We explore three ways of merging the OSM data into the model. In the first one, we apply a naive normalization of the OSM classes into $[0, 1]$ and concatenation to the input data as a fourth channel. In the second one, we implement Fuse-Dense-U-Net and Fuse-DLA-Dense-U-Net following the technique and idea from FuseNet (Hazirbas et al., 2016) and (Audebert et al., 2017) by adding a separate encoder for the normalized OSM raster, whose block-wise output features are added to the aerial image encoder's corresponding output features (see Figure 2.c). In the third one finally, we modify the fusion scheme from FuseNet to keep the aerial RGB encoder features separate from the OSM encoder features. Our intuition is that the fusion of two heterogeneous data types, if done like in FuseNet, will make the optimization of the main encoder harder and lead to lower performance. In our modified architectures, SkipFuse-Dense-U-Net and SkipFuse-DLA-Dense-U-Net, the fusion of RGB and OSM feature maps is done only at the input of the skip-connections, and therefore the two encoders do not interact with one another (see Figure 2.d).

3. DATASET

To train our segmentation networks, we manually annotated a series of aerial images pixel-wise with 3 categories: parking spaces, roads and access ways. The aerial images were acquired from the city of Brunswick (Germany) with the 3K camera system (Kurz et al., 2012) on six days over a period of two years from 2019 till 2020. At each flight, an area of around $40\,\mathrm{km}^2$ was acquired based on five flight strips with a small across track overlap of 10 %. The frame rate was set to 1 Hz, which leads to a 80 % along track overlap of the images. The two full frame $35\,\mathrm{mm}$ cameras aboard capture images of size 5616×3744 pixel. The GSD is derived from the focal length of $50\,\mathrm{mm}$ and ranges between $9.0\,\mathrm{cm}$ and $10.3\,\mathrm{cm}$ depending on the flight altitude, which varies slightly between 650 m and 750 m above ground for the different flight days.

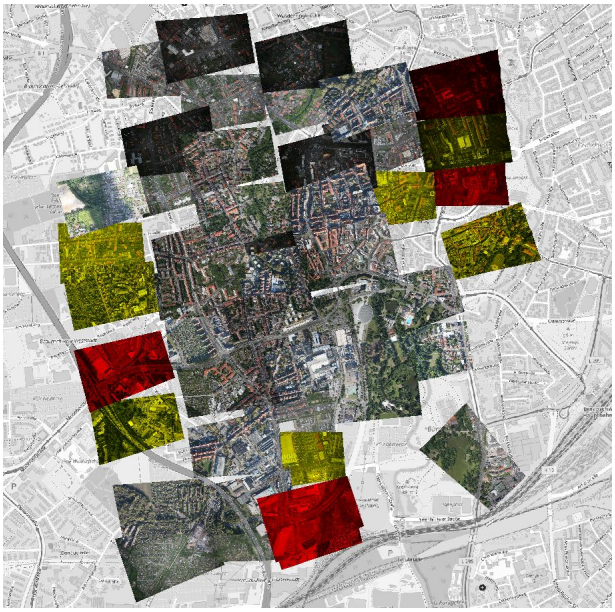


Figure 4. Spatial distribution of training, validation (reddish), and test set (yellowish) images over the city centre of Brunswick (Germany). Note that overlapping areas are only considered in one of the three sets.

Altogether 47 images from the six overflights were selected for annotation. The covered scenery is diverse with respect to buildings and land use, e. g. the historic town center with a large pedestrian area is covered by nine images, the mainly residential areas by 24 images, the industrial areas by ten images, and other areas by four images (see Figure 4). The flight times have been set in such a way that there is a wide variability in terms

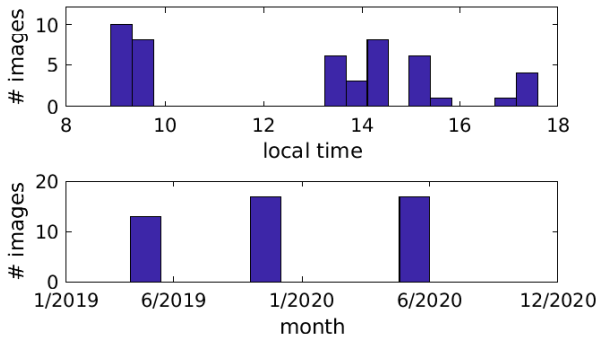


Figure 5. Distribution of selected images across time of day (top) and across months within the years 2019 and 2020 (bottom).

of times, seasons, days of the week, and weather. The distribution of the selected images regarding time and seasons are illustrated in Figure 5. In the selected images various types of car parking are represented, from big parking lots at industrial areas to small along-side parking places in the suburban. Image examples for the different parking types are shown in Figure 6.

For the training, validation and testing of our deep learning based approaches, we divided the 47 images into three disjoint sets: the training set includes 35 images, the validation set contains 5 images and the test set consists of 7 images. Here, we tried to spatially separate the validation and test set as good as possible from the training set while maintaining the same variation of parking types as in the original covered area. The spatial distribution of the three sets is illustrated in Figure 4. Since there was some overlap between images of different sets, we masked out the corresponding areas. More precisely, we excluded all areas contained in the test set from the validation set and all areas contained in the test and validation set from the training set.

The primary goal pursued with the dataset is the most complete detection of parking areas that are both officially dedicated and regularly used as such. The former have various looks: they are of the same or contrasted surfaces compared to roads, have painted or paved markings, different alignments to the direction of travel, and multiple types of exposures to traffic flow. The latter are dual-use areas: most of the day they are taken by moving traffic, and only when parking pressure is high, parking is observed on the roadside or partially on sidewalks. Thus, the fully but not exaggerated detection of parking areas is complex, in which the surroundings also play a role. For this reason, two further classes are defined, which are used to investigate the extent to which the adjacency of road traffic areas can be helpful for the distinction of parking areas. Roads are dedicated to moving motorized traffic and have a connecting function. In contrast to roads, objects of class access way are areas for accessing a destination, e. g. houses, parking garages or backyards. This segmentation is particularly needed in transportation research that builds on this work (Hellekes et al., 2021). Taking the three object classes together, the *background* forms the residual in the images.

Consistent and high quality annotations were achieved on the one hand by refining the annotation policy based on structurally different parts of the test area. On the other hand, manual labelling was performed by experts for German infrastructure and multi-level quality checking. A sample annotation is shown in Figure 3.

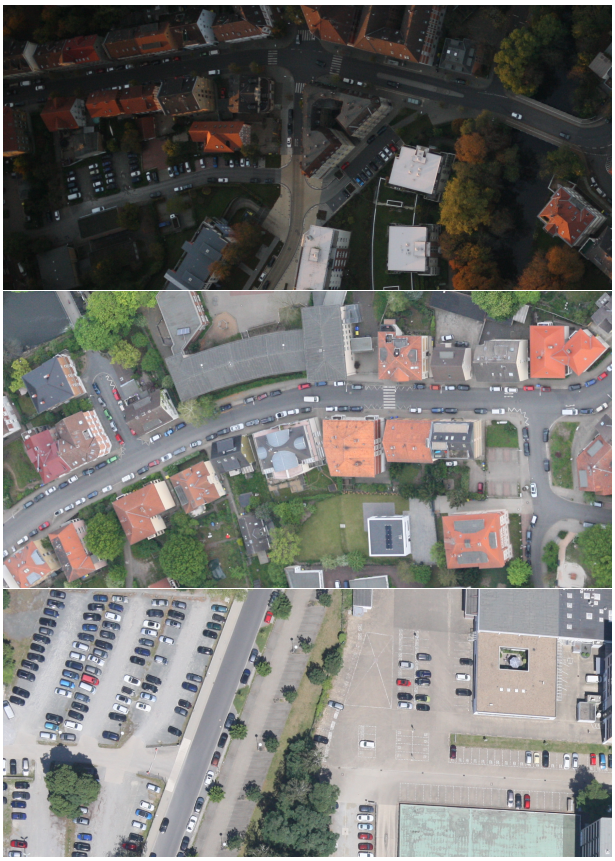


Figure 6. Exemplary image crops (from top to bottom): urban – low illumination, suburban – no sun, industrial – sunny.

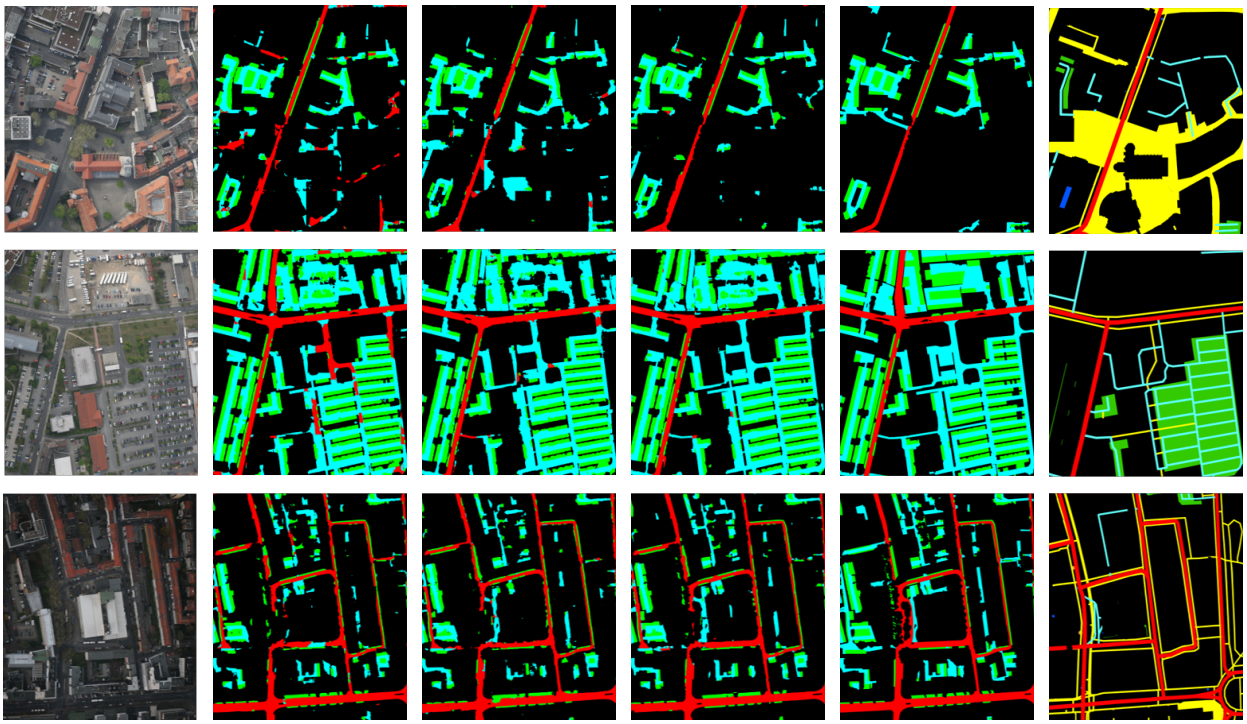
4. RESULTS AND DISCUSSION

All models described in Section 2 are trained over 100 epochs, with a patch size of 512×512 px, with an Adam optimizer and an exponential learning rate schedule initialized at 10^{-4} and decayed at a 0.99 rate after each epoch. When training the Dense-U-Net, the Fuse-Dense-U-Net and the SkipFuse-Dense-U-Net models we used a batch size of 12 and for the training of the DLA-Dense-U-Net, the Fuse-DLA-Dense-U-Net and the SkipFuse-DLA-Dense-U-Net models a batch size of 8. Additionally, we experimented with three OSM data fusion schemes: without OSM (RGB input only), with concatenated OSM as input and the OSM fusion. The training and the validation of the models where performed on sets composed of 3080 patches and 440 patches respectively (for details see Section 3).

The results on the test set images (in total 616 patches) at epoch 100 are presented in Table 1. Overall, we achieved the best results with the SkipFuse-Dense-U-Net. By comparing the different OSM fusion schemes, it can be seen that the concatenation of OSM data into the input features helps extracting roads slightly better, it actually decreases the model's performance for the parking area class. The most likely reason is that this naive approach fuses heterogeneous data through the same encoder, leading to confusion in the feature extraction. In comparison, both Fuse-Dense-U-Net and SkipFuse-Dense-U-Net yield much improved predictions, but with a more significant performance boost for the latter (+2.3 % and +4 % mean IoU respectively, versus Dense-U-Net with no OSM fusion). The DLA-based models however showed unexpected results: without any OSM fusion, DLA-Dense-U-Net reached the best performance on parking area segmentation, by 0.7 % ahead of the second-

backbone network	with OSM	fusion scheme	IoU [%] mean	road	access way	parking	average [%] recall	precision
Dense-U-Net	-	-	72.33	69.30	58.55	68.14	82.16	84.82
Dense-U-Net	✓	concatenation	71.73	72.00	59.51	62.98	81.42	84.64
Dense-U-Net	✓	FuseNet (Hazirbas et al., 2016)	74.62	73.05	63.09	68.49	84.07	85.93
Dense-U-Net	✓	SkipFuseNet	**76.38**	**77.98**	**64.40**	68.77	**84.73**	**87.65**
DLA-Dense-U-Net	-	-	73.49	71.28	59.37	**69.48**	83.47	85.08
DLA-Dense-U-Net	✓	concatenation	72.59	71.91	60.24	65.05	81.84	85.40
DLA-Dense-U-Net	✓	FuseNet (Hazirbas et al., 2016)	72.89	71.80	59.44	67.07	82.64	85.04
DLA-Dense-U-Net	✓	SkipFuseNet	74.60	74.05	61.61	68.70	83.91	86.00

Table 1. Performance comparison of the Dense-U-Net and DLA-Dense-U-Net architectures using different OSM fusion methods.



(a) aerial image (b) no OSM (c) concatenation (d) SkipFuseNet (e) ground truth (f) OSM mask

Figure 7. Qualitative comparison of the impact of OSM and the different fusion schemes on the models with a Dense-U-Net backbone. Categories colors are: parking area, road, access way, pedestrian/bike way and bicycle parking.

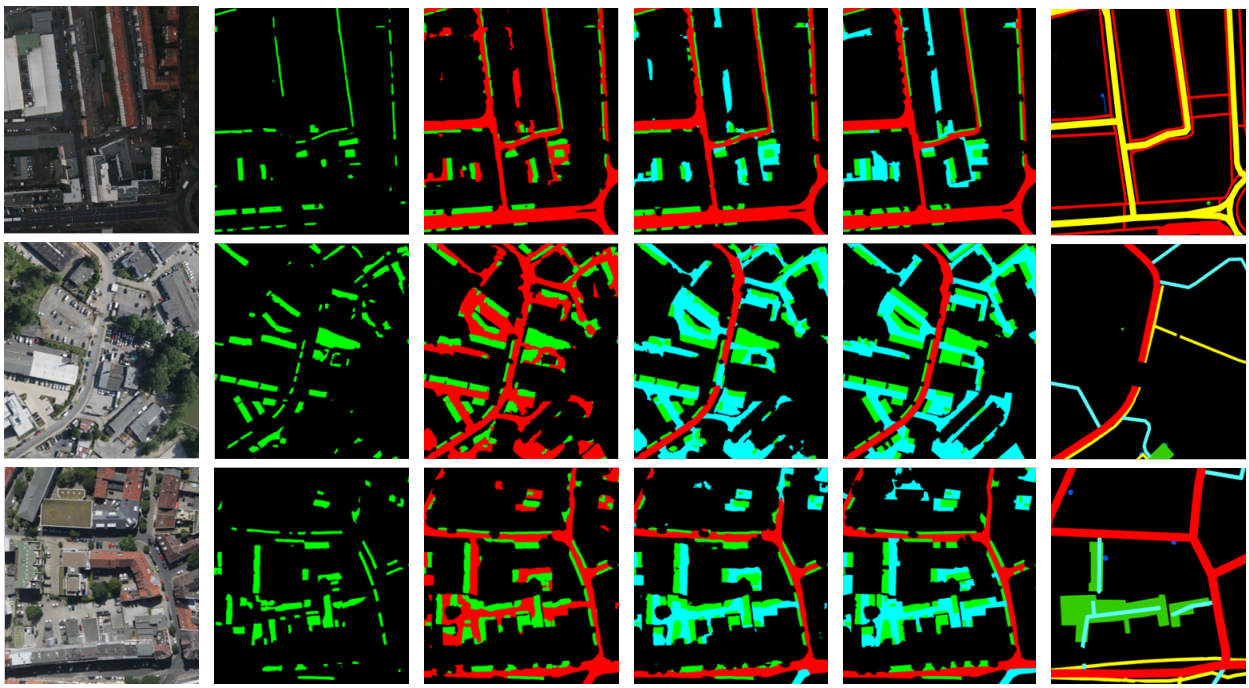
ranked method SkipFuse-Dense-U-Net. And contrary to our expectations, no fusion scheme on the DLA-based model improved this score or that of other classes. A possible explanation could be that the DLA architecture is effective at extracting fine-grained spatial details, but not fine-grained semantic information. Since Dense-U-Net is already providing excellent spatial accuracy in our experiments, little to no performance gain is reasonable to expect from the DLA-based models.

Qualitative results of all models using the Dense-U-Net as backbone are shown in Figure 7. Here, the positive effect of incorporating the OSM data as additional input on the predictions becomes visible. In the first row of the figure, a large pedestrian area is marked in the OSM. When using only the aerial images as input, the network predicts roads and access ways inside the pedestrian zone. In contrast, using the OSM data and the advanced fusion scheme results in almost no wrong prediction in these areas. Furthermore, the OSM data helps the network to better differentiate between the classes access ways and roads. This can be seen in the second row of Figure 7. Here the model in (b) wrongly identifies some access ways as roads, whereas they are consistently predicted as access ways as soon as OSM data is taken into account in (c) and (d).

In another experiment, we investigated the influence of the three classes from our dataset on the predictions of the parking areas. Therefore, we trained our best model, the SkipFuse-Dense-U-Net, on the same training dataset, but 1) with using the class parking area only and 2) with using the class parking area and merging the classes road and access way into one class. All the other hyperparameters were kept the same. A quantitative evaluation of these experiments is provided in Table 2 and image examples in Figure 8. The results in Table 2 show that we gain around 1 % of IoU for the class parking area by training the network only on this class. A possible explanation for this improvement could be that the class parking area is more easily distinguished from other traffic areas than anticipated, and our model can focus more specifically on this class when extracting not other categories of objects. In contrast to the quantitative evaluation, the difference in the predictions' quality for the class parking area is barely visible (see Figure 8). Nevertheless, depending on the use case, it should be considered to use only the class parking area for training if no information about the road network is required later.

backbone network	with OSM	fusion scheme	number of classes	IoU [%] mean	road	access ways	parking	average [%] recall	precision
Dense-U-Net	✓	SkipFuseNet	3	76.38	**77.98**	**64.40**	68.77	84.73	87.65
Dense-U-Net	✓	SkipFuseNet	2	79.73	75.23	-	69.57	87.20	89.56
Dense-U-Net	✓	SkipFuseNet	1	**83.90**	-	-	**69.78**	**89.17**	**92.16**

Table 2. Performance of SkipFuse-Dense-U-Net with all 3 classes, roads and access ways merged, or only parking in the ground truth.



(a) aerial image (b) only parking (c) parking and road (d) three classes (e) ground truth (f) OSM mask

Figure 8. Qualitative comparison of the impact of the number of classes in the training set on the SkipFuse-Dense-U-Net model. Color coding: parking area, road, access way, pedestrian/bike way and bicycle parking.

5. CONCLUSION AND FUTURE WORK

Our experiments have shown that state-of-the-art segmentation models are capable of extracting parking areas accurately, reaching up to 69.78 % IoU in our best setup. Additionally we obtained good results for other traffic related objects, namely roads and access ways, with a best mean IoU score of 76.38 %, which opens opportunities for a future use in transport models. The fusion of OSM brought considerable quantitative and qualitative improvements in both identification of the surface types and smoothness of the region borders. These results were made possible due to the use of an accurately manually annotated dataset, confirming the necessity and value of high-quality labeling for remote sensing tasks. Our dataset reflects well the heterogeneity of Brunswick, Germany, but a large number of additional yet not annotated aerial images is available. This means that in future works, the quality of multiple predictions for the same scene under different conditions could be compared and used to reduce uncertainties of a single prediction through merging. On the generalization side, the layout of traffic areas in Germany is highly standardized but cities have their own characteristics: historical city center, slowly grown structure, centralized urban planning, building density, etc. Future studies should confirm the capability of the proposed method to generalize to new areas and assess the need for fine-tuning on region specific features.

REFERENCES

Amato, G., Carrara, F., Falchi, F., Gennaro, C., Meghini, C., Vairo, C., 2017. Deep learning for decentralized parking lot occupancy detection. *Expert Systems with Applications*, 72, 327–334.

Audebert, N., Le Saux, B., Lefèvre, S., 2017. Joint learning from earth observation and openstreetmap data to get faster better semantic maps. *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1552–1560.

Azimi, S., Henry, C., Sommer, L., Schumann, A., Vig, E., 2019. SkyScapes - Fine-Grained Semantic Understanding of Aerial Scenes. *IEEE International Conference on Computer Vision (ICCV)*.

Buslaev, A., 2017. SpaceNet 3 Road Detection Challenge - 1st Place Solution: `https://github.com/SpaceNetChallenge/RoadDetector/tree/master/albu-solution`.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. *Proceedings of the European Conference on Computer Vision (ECCV)*.

Cheng, G., Han, J., Lu, X., 2017. Remote Sensing Image Scene Classification: Benchmark and State of the Art. *Proceedings of the IEEE*, 105(10), 1865–1883.

Chlaily, S., Mura, M. D., Chanussot, J., Jutten, C., Gamba, P., Marinoni, A., 2020. Capacity and Limits of Multimodal Remote Sensing: Theoretical Aspects and Automatic Information Theory-Based Image Selection. *IEEE Transactions on Geoscience and Remote Sensing*, 1-21.

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B., 2016. The Cityscapes Dataset for Semantic Urban Scene Understanding. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 3213–3223.

Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Basu, S., Hughes, F., Tuia, D., Raskar, R., 2018. Deepglobe 2018: A challenge to parse the earth through satellite images. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

Drouyer, S., 2020. Parking Occupancy Estimation on Planetscope Satellite Images. *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 1098–1101.

Etten, A. V., Lindenbaum, D., Bacastow, T. M., 2019. SpaceNet: A Remote Sensing Dataset and Challenge Series.

Everingham, M., Van Gool, L., Williams, C. K., Winn, J., Zisserman, A., 2010. The pascal visual object classes (VOC) challenge. *International Journal of Computer Vision*, 88(2), 303–338.

Fraunhofer IAO, 2021. Using AI to measure the demand for parking space. Press release: `https://www.iao.fraunhofer.de/en/press-and-media/latest-news/using-ai-to-measure-the-demand-for-parking-space.html`.

Habib, K. M., Morency, C., Trépanier, M., 2012. Integrating parking behaviour in activity-based travel demand modelling: Investigation of the relationship between parking type choice and activity scheduling process. *Transportation Research Part A: Policy and Practice*, 46(1), 154–166.

Hamaguchi, R., Hikosaka, S., 2018. Building detection from satellite imagery using ensemble of size-specific detectors. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 223–2234.

Hazirbas, C., Ma, L., Domokos, C., Cremers, D., 2016. FuseNet: Incorporating Depth into Semantic Segmentation via Fusion-based CNN Architecture. *Asian Conference on Computer Vision (ACCV)*.

Hellekes, J., Merkle, N., Lopez Diaz, M., Henry, C., Heinrichs, M., Azimi, S., Kurz, F., 2021. Assimilation of parking space information derived from remote sensing data into a transport demand model. Under review for ITS World Congress, 10/11/2021 - 10/15/2021, Hamburg, Germany.

Henry, C., Fraundorfer, F., Vig, E., 2020. Aerial Road Segmentation in the Presence of Topological Label Noise. *International Conference on Pattern Recognition (ICPR)*.

Hong, D., Gao, L., Yokoya, N., Yao, J., Chanussot, J., Du, Q., Zhang, B., 2020. More Diverse Means Better: Multimodal Deep Learning Meets Remote-Sensing Imagery Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 1-15.

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.

Lindenbaum, D., 2017. 2nd SpaceNet Competition Winners Code Release: `https://medium.com/the-downlinq/2nd-spacenet-competition-winners-code-release-c7473eea7c11`.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully Convolutional Networks for Semantic Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, 3431–3440.

Merkle, N., S.Azimi, Pless, S., Kurz, F., 2019. Semantic Vehicle Segmentation in Very High Resolution Multispectral Aerial Images Using Deep Neural Networks. *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 5045–5048.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net : Convolutional Networks for Biomedical. N. Navab, J. Hornegger, W. M. Wells, A. F. Frangi (eds), *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Springer International Publishing, Cham, 234–241.

Senate of Berlin, 2014. Datensatz Straßenbefahrung 2014: `https://daten.berlin.de/datensaetze/strassenbefahrung-2014-wms-1`.

Senate of Hamburg, 2020. ParkraumGIS Hamburg: `https://suche.transparenz.hamburg.de/dataset/parkraumgis-hamburg-veraltet2`.

State capital of Munich, 2019. Digitaler Flächennutzungsplan der Landeshauptstadt München: `http://maps.muenchen.de/plan/flaechennutzungsplan`.

Yu, F., Wang, D., Shelhamer, E., Darrell, T., 2018. Deep layer aggregation. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2403–2412.

Zhou, L., Zhang, C., Wu, M., 2018a. D-LinkNet: LinkNet with Pretrained Encoder and Dilated Convolution for High Resolution Satellite Imagery Road Extraction. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

Zhou, W., Newsam, S., Li, C., Shao, Z., 2018b. PatternNet: A benchmark dataset for performance evaluation of remote sensing image retrieval. *ISPRS Journal of Photogrammetry and Remote Sensing*, 145, 197–209.