https://doi.org/10.2352/EI.2023.35.5.IRIACV-324

Terrain Segmentation for Commercial Vehicles and Working Machines

Raimund Edlinger, Ulrich Mitterhuber; University of Applied Sciences Upper Austria, Wels, Austria
Andreas Nüchter; Informatics XVII - Robotics at the Julius-Maximilians University Würzburg, Germany

Abstract

In the field of automated working machines, not only is the general trend towards automation in industry, transport and logistics reflected, but new areas of application and markets are also constantly emerging. In this paper we present a pipeline for terrain classification in off-road environments and in the field of "automated maintenance of slopes", which offers potential for solving numerous socioeconomic needs. Working tasks can be made more efficient, more ergonomic and, in particular, much safer, because mature, automated vehicles are used. At present, however, such tasks can only be carried out remotely or semi-automatically, under the supervision of a trained specialist. This only partially facilitates the work. The real benefit only comes when the supervising person is released from this task and is able to pursue other work. In addition to the development of a safe integrated system and sensor concept for use in public spaces as a basic prerequisite for vehicles licensed in the future, increased situational awareness of mobile systems through machine learning in order to increase their efficiency and flexibility, is also of great importance.

Introduction

Real-time semantic segmentation is a major building block in scene understanding for autonomous robot systems in off-road applications, see Figure 1. The embedded computing platforms employed on autonomous robot systems impose, compute constraints upon the methods used to solve the task. From those constraints, a need for real-time focused semantic segmentation methods did arise. Fortunately, in recent years many new Deep Learning based methods, which can inference in real-time on workstation environments, were proposed. However, those methods were not yet evaluated being applied to an off-road track environment while computing inference on an embedded platform.

The complexity of working tasks in unstructured environments and under changing environmental conditions often poses a challenge even for well-trained human drivers. Nevertheless, automated work equipment has so far had a sufficient level of situational awareness to be able to perform work tasks efficiently and without violating safety requirements.

The aim of this paper is to make environment recognition and localization in dynamic environments more intelligent with the help of adapted machine learning methods. In a concrete application, for example, it should be possible to distinguish between vegetation, people and other obstacles. The creation of a complete dataset with annotation, which includes all possibilities, is time-consuming and costly and not possible within the scope of this project. Therefore, research is being conducted in the areas of "transfer learning" and "domain adaptation". Available datasets from urban and off-road areas, as well as from internal datasets from previous projects, will be aggregated. The aim of "Transfer Learning" is to derive a general visual understanding from these large datasets in order to reduce the data collection effort in the target application.

Figure 1. Metron P48RC is a radio-controlled tool carrier with a true hybrid drive.

Related Work

Modern CV methods are usually bench-marked on public challenges such as the ImageNet Large Scale Visual Recognition Challenge [24], which was the most relevant one for object recognition. Its dataset contains millions of examples of 1000 mutually exclusive classes. The last year in which the challenge was executed in its original form was in 2017. Since then, it is considered solved however, new methods are still evaluated and compared on its dataset. An extensive overview of state-of-the-art technologies of semantic segmentation based on Deep learning can be found in [18].

Semantic Segmentation

Semantic segmentation is a classification problem in which the goal is to label every pixel of an image to one of a set of predefined classes. With its nature of dense pixel labeling it extracts a vast amount of information from its given images and serves therefore as a major building block in all kinds of applications ranging from scene under-

Table 1: State-of-the-art real-time semantic segmentation methods. All metrics listed were obtained by evaluating on the Cityscapes dataset. Only entries marked with superscript '*' were obtained with the Cityscapes validation-set. All GPUs listed are from the brand NVIDIA.

Method	mIoU [%]	FPS $[\frac{1}{s}]$	Input size $[h \times w]$	GPU
AutoRTNet-A [26]	73.9	110	768×1536	Titan XP
BiSeNet (ResNet-18) [31]	74.7	65.5	1024×2048	Titan XP
BiSeNetV2 [30]	72.6	156	1024×2048	GTX 1080Ti
CAS [34]	72.3	108	768×1536	Titan XP
DABNet [9]	70.1	27.7	1024×2048	GTX 1080Ti
DF1-Seg [11]	74.1	106.4	1024×2048	GTX 1080Ti
DFANet A [10]	71.3	100	1024×1024	Titan X
ESNet [29]	70.7	62	512×1024	GTX 1080Ti
FarSee-Net [35]	70.2	68.5	512×1024	Titan X
FasterSeg [4]	71.5	163.9	1024×2048	GTX 1080Ti
FDDWNet [13]	71.5	60	512×1024	RTX 2080Ti
FPENet [14]	70.1	55	768×1536	Titan V
GAS [12]	73.5	108.4	769×1537	Titan XP
GUNet [17]	70.1	33	512×1024	Titan XP
ICNet [36]	70.6	30.3	1024×2048	Titan X
LBN-AA+DASPP+SPN [6]	73.6	51	448×896	Titan X
LEDNet [28]	70.6	71	512×1024	GTX 1080Ti
MSFNet [25]	77.1	41	1024×2048	RTX 2080Ti
RGPNet (ResNet-18) [1]	74.1^{*}	37.4	1024×2048	RTX 2080Ti
ShelfNet18-lw [38]	74.8	59.2	768×1536	GTX 1080Ti
SwiftNetRN-18 [20]	75.5	39.9	1024×2048	GTX 1080Ti

standing in autonomous driving [20] to biomedical image analysis [22]. The arguably most influential Convolutional Neural Network (CNN) architecture in semantic segmentation is the Fully Convolutional Network (FCN) proposed in [15]. Its main idea was to modify existing image recognition CNNs such as GoogleNet or VGG to output a segmentation map. Since FCN started the application of CNNs in semantic segmentation, many new architectures, each proposing different design choices, have been proposed.

An architecture which represents the **encoder-decoder** design paradigm is U-net [22] which was used for biomedical image processing. U-net's left part is doing the down-sampling/feature-extraction while the right part is doing the up-sampling. U-net employs learned up-sampling in the form of up-convolutions while many other architectures use bi-linear or nearest-neighbor interpolation for up-sampling. The skip-connections, depicted as gray horizontal arrows, are used to refine the accuracy of the segmentation by fusing low-level features with high-level ones. An example which represents another design paradigm is the Pyramid Scene Parsing Network (PSP-Net) [37]. It represents **multi-path** architectures which extract and fuse features from different sizes. The extraction of differently sized features is done in its pyramid pooling module. After the module, the extracted features are concatenated with features from earlier layers, which were propagated forward by skip connections.

The methods shown above are no longer state of the art. However, they are well suited for representing some architecture developments of CNNs for semantic segmentation. Current state-of-the-art CNN-based methods are, for example, DeepLabv3+ [3] and OCR [33]. Their high accuracy come however, with big computational burdens which make them impractical for deployment on real-time platforms where computational resources are scarce. Due to this fact, a need for real-time focused methods arose. Fortunately, this need has been addressed by the research community in recent years.

Datasets

Training datasets consist of examples comprised of input features and a target or a label. The mathematical notation of a training dataset for m examples is:

$$\{(x^{(i)}, y^{(i)}); i = 1, ..., m\} \quad (1)$$

Producing densely labeled datasets for semantic segmentation is a very laborious task. Even though there is software to help, the collected images usually have to be labeled largely by hand. Due to this fact, synthetically generated datasets such as SYNTHIA [23] and generator based methods proposed on the game of Grand Theft Auto are starting to gain momentum. However, synthetically generated datasets still have a gap to reality. **Real-world** datasets which are publicly available are here divided into **general purpose** and **driving** environments.

Several **driving** datasets in **urban** environments are Cityscapes [5], ApolloScape [8], Mapilary [19], CamVid [2] and BDD100K [32]. **Off-road** driving datasets are available with the Freiburg Forest [27] and the RoboNav Data Collection [16].

Real-Time Semantic Segmentation

In recent years, many CNN-based architectures for real-time semantic segmentation have been proposed. Their main concern is to find an optimal accuracy-efficiency trade-off. In Table 1, at the time of writing, several of the most relevant methods are listed. All information used in the table was taken from the method's original publications. To reduce the number of methods in Table 1 to the most relevant ones, the methods were chosen when they fulfilled the following two conditions:

- Evaluation metrics on Cityscapes provided.
- Achieving a mIoU on the Cityscapes for test or validation dataset with over 70% while being computed on a workstation platform with above 25 FPS.

Implemented Method

This section is devoted to explaining the implemented real-time semantic segmentation method and the employed data processing including the used datasets. The implemented method is **DABNet** proposed in 2019 in [9]. It was chosen over other methods of related work due to the fact that its training can be done end-to-end, which means no complicated pre-training actions are required, and its original implementation was published alongside its paper in [9]. The implementation of this study is based on the original implementation. The explanations in this section are based on the original publication [9], where further information and a more detailed explanation can be found. The contribution of DABNet, that has to be explained first, is the novel depth-wise asymmetric bottleneck module, short the **DAB module**. It is used to reduce the number of parameters while extracting and combining local and contextual features. The DAB module's main building blocks are standard 3×3 and 1×1 convolutions, dilated convolutions for efficiently broadening the receptive field [7], and depth-wise separable convolutions which are used in many CNNs aimed at compute efficiency. To enable the parallel extraction of local and contextual features, the DAB module employs a two-branch approach.

Experimental Results

To evaluate the performance of dense pixel labeling methods, numerous evaluation metrics have been proposed. The metric nowadays commonly used in semantic segmentation is the **Jaccard Index** also known as **Intersection over Union (IoU)** shown in eqn. 2 [21]:

$$IoU(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \tag{2}$$

It calculates a ratio of how much two sets, which are in semantic segmentation the ground-truth and predicted segmentation maps, overlap. It reaches its maximum value of one when the sets overlap entirely and its minimum value of zero when the sets do not overlap at all. When semantic segmentation with multiple classes is done, usually the **mean Intersection over Union (mIoU)** is used. Other metrics or loss functions used in semantic segmentation such as the **Dice Coefficient** are examined and are more explained in [21].

For training and part of the inference speed evaluation a workstation with two NVIDIA RTX 2080Ti GPUs, 128 GB DDR4 RAM, and an AMD Ryzen 9 3950X CPU was used. It should be noted that, for training, both GPUs were used in parallel to enable larger batch sizes. However, for inference speed evaluation only one GPU was used.

All experiments were conducted with the same software stack on both environments. Python 3.6, CUDA 10.2, CUDNN 8, and PyTorch 1.6 were used. No specific optimization frameworks, such as TensorRT, were used. All experiments were executed with pure Python and PyTorch. During experimentation, the implemented method was trained on four different datasets and then evaluated on a workstation and an embedded-platform to be most relevant to the field of autonomous robot systems. For validating the implemented method, two of the three datasets were the well-studied ones, namely the CamVid and Cityscapes datasets. The third dataset, which explicitly targeted the research objectives, was the less studied Freiburg Forest off-road track dataset. The fourth dataset (Smarter) was done by manual annotation and combined with the other datasets.

To meet the research objectives, a state-of-the-art architecture was implemented. This architecture was then trained and evaluated on four different datasets (three existing online datasets and one self-made dataset). Two of which represented urban-road environments, and one represented off-road track environments. The evaluation was executed on a workstation-platform (NVIDIA RTX 2080Ti), and an embedded-platform (NVIDIA Jetson AGX XAVIER). During evaluation on the off-road track dataset, the implemented method achieved a mean Intersection over Union of 81.5% while computing inference in real-time with 181.5 and 25.3 Frames per Second on the workstation and embedded platform respectively, see Table 2. Based on those results, the research concludes that the current Deep Learning based state-of-the-art real-time semantic segmentation methods are capable of achieving high accuracy on off-road environments while computing inference in real-time on an embedded platform. The Figures 2 - 9 show the validation and training metrics of mIoU and loss of each dataset. The Figures 10 - 15 shows the qualitative evaluation of each on-road and off-road dataset.

The predictions show decent performance in understanding the scene as well as reliable segmentation that can be used for navigation. The field experiments that were carried out with the DABNet implementation are shown in the following Figures 10, 11 and 12. The Freiburg Forest dataset was also tested at the 1^{st} Austrian Alpin Robotic Trial for terrain segmentation of the gravel road with accurate environmental feedback, see Figure 13. For the SMARTER (Slope Maintenance Automation using Real-time Telecommunication and advanced Environ-

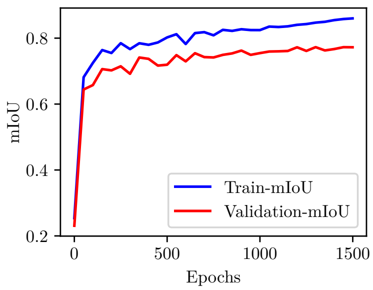

Figure 2. *CamVid training: Validation- and training-mIoU plotted over epochs.*

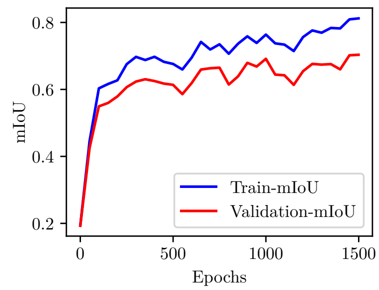

Figure 3. *Cityscapes training: Validation- and training-mIoU plotted over epochs.*

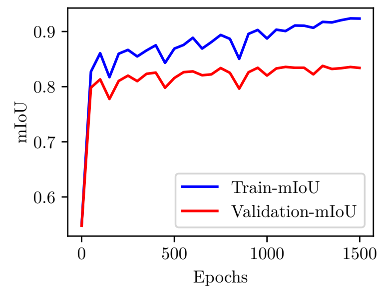

Figure 4. *Freiburg Forest training: Validation- and training-mIoU plotted over epochs.*

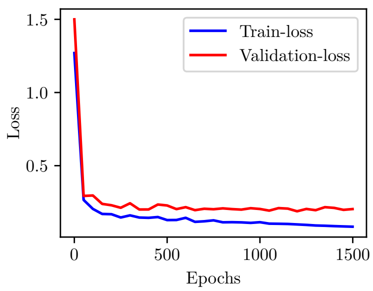

Figure 5. *CamVid training: Validation- and training-loss plotted over epochs.*

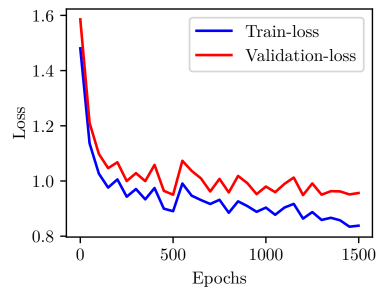

Figure 6. *Cityscapes training: Validation- and training-loss plotted over epochs.*

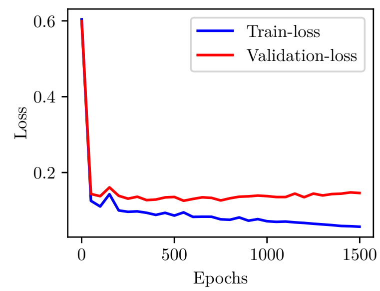

Figure 7. *Freiburg Forest training: Validation- and training-loss plotted over epochs.*

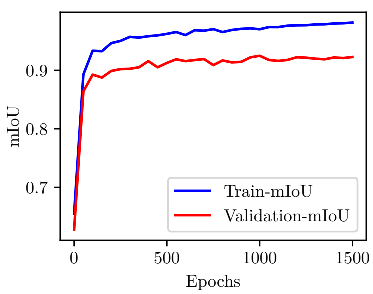

Figure 8. *Smarter training: Validation- and training-mIoU plotted over epochs.*

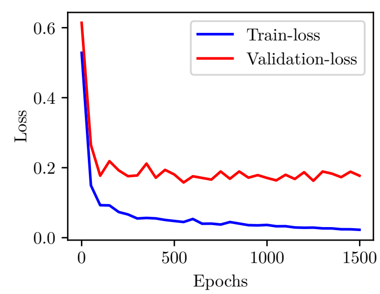

Figure 9. *Smarter training: Validation- and training-loss plotted over epochs.*

Table 2: Quantitative evaluation results of DABNet on different datasets.

Dataset	**Input Size**$[h \times w]$	**mIoU** [%]	**FPS** $[\frac{1}{s}]$	
			RTX 2080Ti	XAVIER
CamVid	360×480	67.2	180.0	28.0
Cityscapes	1024×2048	70.4	39.6	5.0
Cityscapes	512×1024	65.4	162.2	18.4
Cityscapes	256×512	51.5	180.4	27.5
Freiburg Forrest	420×840	81.5	181.5	25.3
Smarter	512×1024	85.4	162.2	18.4

ment Recognition) research project, it is important that the working machine understands not only the learned classes, such as road, meadow, vegetation, people, but also where the meadow has already been mowed (red) and where it still needs to be mowed (blue), see Figure 14 and 15. The figures show quantitatively very good results, considering that we used a total of 250 images for annotating the dataset. This study evaluates the current state-of-the-art real-time semantic segmentation methods applied to the less studied environment of off-road tracks while computing inference on an embedded platform. With the gained knowledge, decisions on the applicability of those methods

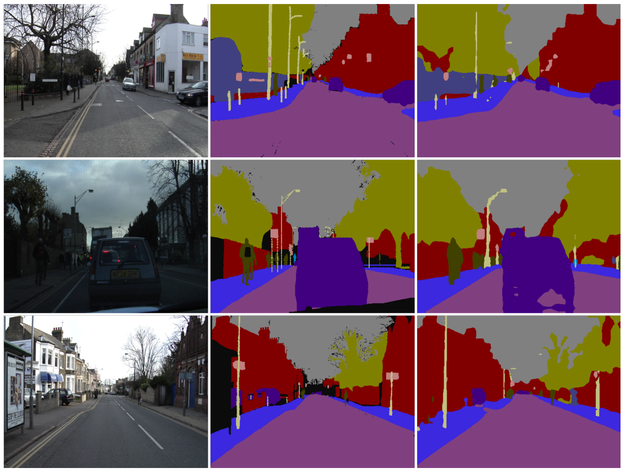

Figure 10. Qualitative evaluation results on CamVid. Columns from left to right: Input image, colorized ground-truth image, and colorized prediction image.

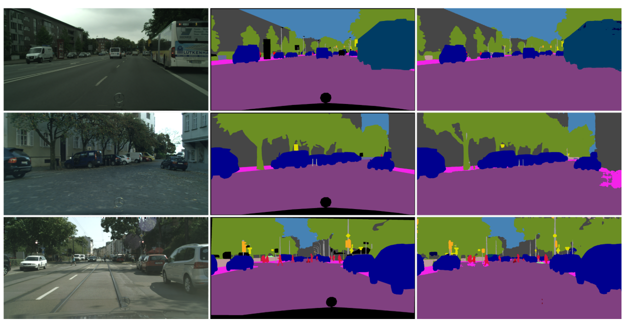

Figure 11. Qualitative evaluation results on Cityscapes. Columns from left to right: Input image, colorized ground-truth image, and colorized prediction image.

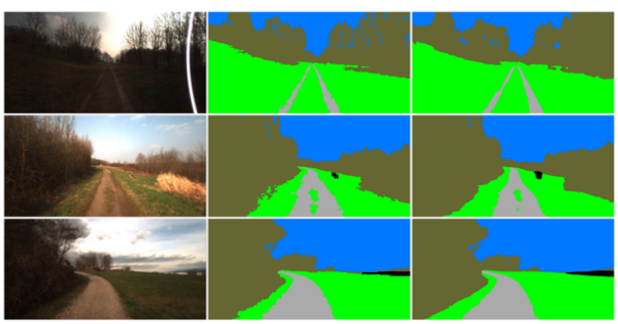

Figure 12. Qualitative evaluation results on Freiburg Forest. Columns from left to right: Input image, colorized ground-truth image, and colorized prediction image.

to other currently unstudied environments such as industrial plants can be made. The evaluation results of the DABNet instances trained on the CamVid and Cityscapes datasets respectively showed better results as in DABNet's original publication [9] documented. This is probably due to the different training hyper-parameters of larger batch size for Cityscapes and a higher number of epochs both for CamVid and Cityscapes. At the evaluation on the CamVid

Figure 13. Freiburg Forest dataset used for AART – 1^{st} Austrian Alpin Robotic Trial.

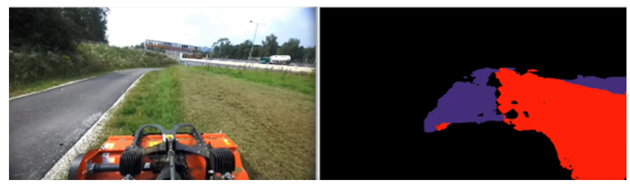

Figure 14. Smarter dataset: Segmentation of mown (red) and unmown (blue) meadows from the view of the tool carrier.

Figure 15. Smarter dataset: Segmentation of mown (red) and unmown (blue) meadows.

test-set a mIoU of 67.2% while computing inference with 180.0 FPS on the RTX 2080Ti platform was achieved. The evaluation on the Cityscapes evaluation-set yielded a mIoU of 70.4% while computing inference with 39.6 FPS on the RTX 2080Ti platform. Interpreting those results, the first research objective, of validating the implemented model on well-known benchmarks, has been met. Further, the evaluation of the instance trained on the Freiburg Forest dataset showed impressive results in terms of mIoU and inference speed. It reached almost the same mIoU while executing at a much faster inference speed as the method documented in the original Freiburg Forest paper [27]. On Freiburg Forest, DABNet achieved a mIoU of 81.5% while computing real-time inference with 25.3 FPS on the XAVIER platform. For the self annotated dataset (150 images training, 50 images for validation and 50 images for testing the dataset) DABnet achieved a mIoU of 85.4% with 18.4 FPS on the XAVIER embedded platform. Contemplating the results of the conducted experiments the current state of the art CNN-based real-time semantic segmentation methods can be applied to off-road environments while computing real-time inference on embedded platforms.

Conclusion and Future Work

In this paper, we propose an overview of state-of-the-art CNN-based semantic segmentation methods and field experiments to develop autonomous robot systems for off-road environments. Current real-time semantic segmentation methods were increasingly developed and applied mainly for on-road applications, because the hype for autonomous cars was much stronger on normal roads and motorways. The experiments and generation of outdoor datasets shows that the DABNet method can achieve high accuracy when applied on off-road environments while computing inference in real-time on an embedded platform.

One such hypothesis is that DABNet should be able to generalize well to other currently unstudied environments such as industrial plants, construction sites, or farmland. To confirm this hypothesis, further research needs to be done and to generate novel datasets of currently unstudied environments.

Acknowledgments

The research of these results has been accomplished within the SMARTER - Slope Maintenance Automation using Real-Time Telecommunication and advanced Environment Recognition project. This work has been funded by the Austrian Research Promotion Agency (FFG) within the program "Mobility of the future" No. 879646.

References

[1] Elahe Arani, Shabbir Marzban, Andrei Pata, and Bahram Zonooz. Rgpnet: A real-time general purpose semantic segmentation. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 3009–3018, 2021.

[2] Gabriel J Brostow, Julien Fauqueur, and Roberto Cipolla. Semantic object classes in video: A high-definition ground truth database. *Pattern Recognition Letters*, 30(2):88–97, 2009.

[3] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018.

[4] Wuyang Chen, Xinyu Gong, Xianming Liu, Qian Zhang, Yuan Li, and Zhangyang Wang. Fasterseg: Searching for faster real-time semantic segmentation. *arXiv preprint arXiv:1912.10917*, 2019.

[5] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016.

[6] Genshun Dong, Yan Yan, Chunhua Shen, and Hanzi Wang. Real-time high-performance semantic image segmentation of urban street scenes. *IEEE Transactions on Intelligent Transportation Systems*, 22(6):3258–3274, 2020.

[7] Matthias Holschneider, Richard Kronland-Martinet, Jean Morlet, and Ph Tchamitchian. A real-time algorithm for signal analysis with the help of the wavelet transform. In *Wavelets*, pages 286–297. Springer, 1990.

[8] Xinyu Huang, Xinjing Cheng, Qichuan Geng, Binbin Cao, Dingfu Zhou, Peng Wang, Yuanqing Lin, and Ruigang Yang. The apolloscape dataset for autonomous driving. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 954–960, 2018.

[9] Gen Li and Joongkyu Kim. Dabnet: Depth-wise asymmetric bottleneck for real-time semantic segmentation. In *British Machine Vision Conference*, 2019.

[10] Hanchao Li, Pengfei Xiong, Haoqiang Fan, and Jian Sun. Dfanet: Deep feature aggregation for real-time semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9522–9531, 2019.

[11] Xin Li, Yiming Zhou, Zheng Pan, and Jiashi Feng. Partial order pruning: for best speed/accuracy trade-off in neural architecture search. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9145–9153, 2019.

[12] Peiwen Lin, Peng Sun, Guangliang Cheng, Sirui Xie, Xi Li, and Jianping Shi. Graph-guided architecture search for real-time semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4203–4212, 2020.

[13] Jia Liu, Quan Zhou, Yong Qiang, Bin Kang, Xiaofu Wu, and Baoyu Zheng. Fddwnet: a lightweight convolutional neural network for real-time semantic segmentation. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2373–2377. IEEE, 2020.

[14] Mengyu Liu and Hujun Yin. Feature pyramid encoding network for real-time semantic segmentation. *arXiv preprint arXiv:1909.08599*, 2019.

[15] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[16] Florian Schöggl Matthias Eder, Raphael Prinz and Gerald Steinbauer-Wagner. Robonav data collection, 2022. `https://robonav.ist.tugraz.at/data/`.

[17] Davide Mazzini. Guided upsampling network for real-time semantic segmentation. *arXiv preprint arXiv:1807.07466*, 2018.

[18] Yujian Mo, Yan Wu, Xinneng Yang, Feilin Liu, and Yujun Liao. Review the state-of-the-art technologies of semantic segmentation based on deep learning. *Neurocomputing*, 493:626–646, 2022.

[19] Gerhard Neuhold, Tobias Ollmann, Samuel Rota Bulo, and Peter Kontschieder. The mapillary vistas dataset for semantic understanding of street scenes. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4990–4999, 2017.

[20] Marin Orsic, Ivan Kreso, Petra Bevandic, and Sinisa Segvic. In defense of pre-trained imagenet architectures for real-time semantic segmentation of road-driving images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 12607–12616, 2019.

[21] Benjamin Planche and Eliot Andres. *Hands-On Computer Vision with TensorFlow 2*. Packt Publishing Ltd, 2019.

[22] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[23] German Ros, Laura Sellart, Joanna Materzynska, David Vazquez, and Antonio M Lopez. The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3234–3243, 2016.

[24] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015.

[25] Haiyang Si, Zhiqiang Zhang, Feifan Lv, Gang Yu, and Feng Lu. Real-time semantic segmentation via multiply spatial fusion network. *arXiv preprint arXiv:1911.07217*, 2019.

[26] Peng Sun, Jiaxiang Wu, Songyuan Li, Peiwen Lin, Junzhou Huang, and Xi Li. Real-time semantic segmentation via auto depth, downsampling joint decision and feature aggregation. *International Journal of Computer Vision*, 129(5):1506–1525, 2021.

[27] Abhinav Valada, Gabriel Oliveira, Thomas Brox, and Wolfram Burgard. Deep multispectral semantic scene understanding of forested environments using multimodal fusion. In *International Symposium on Experimental Robotics (ISER)*, 2016.

[28] Yu Wang, Quan Zhou, Jia Liu, Jian Xiong, Guangwei Gao, Xiaofu Wu, and Longin Jan Latecki. Lednet: A lightweight encoder-decoder network for real-time semantic segmentation. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 1860–1864. IEEE, 2019.

[29] Yu Wang, Quan Zhou, Jian Xiong, Xiaofu Wu, and Xin Jin. Esnet: An efficient symmetric network for real-time semantic segmentation. In *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*, pages 41–52. Springer, 2019.

[30] Changqian Yu, Changxin Gao, Jingbo Wang, Gang Yu, Chunhua Shen, and Nong Sang. Bisenet v2: Bilateral network with guided aggregation for real-time semantic segmentation. *International Journal of Computer Vision*, 129(11):3051–3068, 2021.

[31] Changqian Yu, Jingbo Wang, Chao Peng, Changxin Gao, Gang Yu, and Nong Sang. Bisenet: Bilateral segmentation network for real-time semantic segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 325–341, 2018.

[32] Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, and Trevor Darrell. Bdd100k: A diverse driving dataset for heterogeneous multitask learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2636–2645, 2020.

[33] Yuhui Yuan, Xilin Chen, and Jingdong Wang. Object-contextual representations for semantic segmentation. In *European conference on computer vision*, pages 173–190. Springer, 2020.

[34] Yiheng Zhang, Zhaofan Qiu, Jingen Liu, Ting Yao, Dong Liu, and Tao Mei. Customizable architecture search for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 11641–11650, 2019.

[35] Zhanpeng Zhang and Kaipeng Zhang. Farsee-net: Real-time semantic segmentation by efficient multi-scale context aggregation and feature space super-resolution. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 8411–8417. IEEE, 2020.

[36] Hengshuang Zhao, Xiaojuan Qi, Xiaoyong Shen, Jianping Shi, and Jiaya Jia. Icnet for real-time semantic segmentation on high-resolution images. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 405–420, 2018.

[37] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2881–2890, 2017.

[38] Juntang Zhuang, Junlin Yang, Lin Gu, and Nicha Dvornek. Shelfnet for fast semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 2019.

Author Biography

Raimund Edlinger is assistant professor at the University of Applied Sciences Upper Austria. He received his DI(FH) in sensors and micro-systems (2007) and his MSc. in Automation Engineering (2013) from the University of Applied Sciences Upper Austria. Since 2007 he has worked as a researcher at the R&D at the University in Wels/Austria. His work has focused on the development of mobile robots and sensor systems. He is a IEEE Member and on the board of RoboCup Rescue League as technical member and since 2018 Phd student at Graduate School Science and Technology at University of Würzburg.

Ulrich Mitterhuber was research associate at the University of Applied Sciences Upper Austria were he also completed his DI in robotic systems engineering in the year of 2022 with distinction. As researcher he focused on working and developing computer vision algorithms driven by the latest improvements in the field of Deep Learning. More specifically, he developed software for semantic segmentation, object detection and lane detection deployed on embedded hardware.

Andreas Nüchter is professor of computer science (robotics) at University of Würzburg. He holds a doctorate degree (Dr. rer. nat) from University of Bonn. His thesis was shortlisted for the EURON PhD award. Andreas works on robotics and automation, cognitive systems and artificial intelligence. His main research interests include reliable robot control, 3D environment mapping, 3D vision, and laser scanning technologies, resulting in fast 3D scan matching algorithms that enable robots to perceive and map their environment in 3D representing the pose with 6 degrees of freedom. The capabilities of these robotic SLAM approaches were demonstrated at RoboCup Rescue competitions, ELROB and several other events. He is a member of the GI and the IEEE.