Deep Learning Datasets Challenges For Semantic Segmentation - A Survey

Claire Ponciano[1], Markus Schaffert, Jean-Jacques Ponciano

Abstract: This survey offers a comprehensive analysis of challenges encountered when employing large-scale datasets for deep learning-based semantic segmentation, an area with significant implications for industries such as autonomous driving, precision agriculture, and medical imaging. Through a systematic review of 94 papers from *Papers with Code*, we identified 32 substantial challenges, which we categorized into six key areas: Data Quality and Quantity, Data Preprocessing, Resource Constraints, Data Management and Privacy, Generalization, and Data Compatibility. By identifying and explicating these challenges, our research provides a crucial reference point for future studies aiming to address these issues and enhance the performance of deep learning models for semantic segmentation. Future work will focus on leveraging AI and semantic technologies to provide solutions to these challenges.

Keywords: Deep Learning; Deep Learning challenges; Semantic segmentation; Data quality; Resource constraints; Generalization; Data management; Data privacy; Data compatibility

1 Introduction

Deep learning has revolutionized computer vision, particularly semantic segmentation, which involves pixel or point classification in image or point cloud respectively. Among various emerging deep learning models, our focus is on semantic segmentation due to its versatility across fields like BIM [PPP21], Optimal Recording and Modelling [Po19, PPB21, Po21], precision agriculture [An21], and medical imaging [Is18b]. While these models are effective, they encounter challenges, often discussed in terms of the models themselves [Sh19]. Large-scale datasets, used in deep learning, present their own unique set of hurdles, which curtail their use and model effectiveness, despite their transformational potential [Xi22]. For example, in drone-based semantic segmentation [Ta21], issues like low-resolution images, class imbalance, varying lighting and weather conditions, and the presence of sensitive information pose significant challenges to model accuracy and data privacy. Overcoming these challenges is crucial for enhancing deep learning models' performance in semantic segmentation, which necessitates a thorough understanding of these hurdles and the development of robust, ethical strategies for dataset utilization. This paper examines these challenges by answering the research question: What are the dataset challenges

[1] Mainz University of Applied Sciences, i3mainz – Institute for Spatial Information and Surveying Technology, Lucy-Hillebrand-Straße 2 55128 Mainz, Germany (claire.ponciano,markus.schaffert,jean-jacques.ponciano) @hs-mainz.de

negatively impacting deep learning-based semantic segmentation models? We conduct a survey on 47 datasets (from a pool of 267) and 94 related papers from *Papers with Code*[2]. These selected datasets and the survey of their related papers aim to identify obstacles in large-scale datasets for semantic segmentation based on deep learning, thus promoting the development of more reliable and robust models. This endeavor constitutes a significant step towards understanding and mitigating the challenges linked to deep learning datasets for semantic segmentation, with potential far-reaching benefits across industries. The paper subsequently presents related work in dataset challenge research, outlines our method for identifying challenges, and discusses the 32 identified challenges and their implications. Then, this survey concludes that it unveils a diverse array of challenges in large-scale datasets for deep learning-based semantic segmentation, laying a comprehensive groundwork for future endeavors aimed at harnessing artificial intelligence and semantic technologies to address these issues and advance the field of computer vision.

2 Related Work

Research on constructing high-quality datasets is a nascent field, where works like [Sh19] delve into complexities of deep learning networks, and others such as [Mu19] probe issues linked to these datasets, but bypass semantic segmentation. However, in the paper [Mu19], we identify applicable challenges for creating semantic segmentation datasets, which include hurdles like metadata scarcity and diversity deficiency, which can hamper training and evaluation. In addition, the *heterogeneity in data* signifies a delicate balance between diversity and consistency, pivotal for dataset utility. *Data Quality* is critical as inferior data can degrade model performance, and *Data sources and Distribution* underscores the necessity for varied data sources and balanced classes to avert model bias. These challenges elucidate the diligence required in forming quality semantic segmentation datasets and also highlight the potential for enhanced deep learning outcomes. Despite these challenges identified in related works, our understanding remains limited. Hence, we propose an exhaustive survey to identify dataset-based challenges, leading us to our research question. The subsequent section outlines the method employed to survey these dataset challenges.

3 Method

Semantic segmentation, powered by a myriad of available datasets, is a burgeoning area of research. *Papers with Code* is a hosting platform providing datasets, papers, and code. It provides an invaluable resource for our survey.

[2] https://paperswithcode.com/

Datasets	Selected Papers
2D-3D-S	[Ke17, RFB15]
ADE20K	[Li19, WH21]
ApolloScape	[Ca20, Hu19]
AVE	[Zh22b, Zh23]
BDD100K	[He17, Wa20a]
BraTS 2015	[Is18a, Ka17]
CamVid	[BKC17, Yu21]
Cityscapes	[MWA18a, Ch17c]
COCO	[Ho17, Wa22]
COCO-Stuff	[Ki19, Fa22]
DAVIS	[Ha22, Vo19]
DAVIS2017	[St21, Vo19]
EuroSAT	[GD22, Ne19]
GTA5	[He16, SZ14]
HAM10000	[Ge20, Jh20]
Helen	[De20, Zh22a]
IDD	[CHS22, MV21]
KITTI	[Ca19, MWA18b]
KITTI-360	[Qi17b, Qi17a]
Kvasir	[RFB15, Zh18b]
LabelMe	[Gh15, Zh19]
LIP	[Su19, Zh18a]
Make3D	[Go19, GVZ16]
Mapillary Vistas Dataset	[Bu20, Ch20]
Medical Segmentation Decathlon	[Is18b, Ji22]
NYU Depth V2	[LSD15a, Oq23]
Objects365	[Fa22, Ku22]
Pascal-5i	[LE22, Zh20c]
PascalContext	[Wa20b, Ch17b]
Pascal VOC 2007	[Ch17b, Re16]
Pascal VOC 2012	[Re16, Ch17c]
PartNet	[Gu20, Mo19a]
Promise12	[MNA16, Is19]
ReferItGame	[Yu16, Ac21]
RefCOCO	[Li21, Zo23]
ScanNet	[Pa16, He17]
Semantic3D	[Hu20, Gu20]
SegTrack-v2	[Oh18, Su23]
ShapeNet	[Ch15, Su18]
S3DIS	[Gu17, Qi17a]
SUN3D	[Ch17a, Sa20]
SUNCG	[ESL19, Mo19b]
SUN RGB-D	[LSD15b, BKC17]
Synthia	[Zh20a, Hu18]
VisDA-2017	[Yu20, Zh20b]
Virtual KITTI	[Hu22, Re19]
YouTube-VIS 2019	[Go21, WBP17]

Tab. 1: Selected datasets with surveyed top papers

The survey presented in this paper is based on 47 datasets out of 267. These 47 datasets have been selected according to two criteria showing widespread use and thus enabling us to highlight the relevance of the challenges identified by the survey for practical applications. The first criterion for selecting a dataset is that it should have been used in more than 50 papers[3]. The second criterion is that the dataset has been used by the most *starred* approaches on GitHub. Once these 47 datasets had been identified, the two most cited papers related to these 47 datasets available on *Papers with Code* were studied in order to (i) identify the dataset challenges highlighted by the authors, and (ii) correlate the issues of the approaches related to the same dataset with the characteristics of the dataset in order to deduce the problematic characteristics of a dataset and therefore its challenges. The 47 datasets and the 94 related papers studied for this paper are shown in Table 1. This survey spans a wide range of data types, problem contexts, and application domains. From autonomous driving to precision agriculture to medical imaging, each of them has unique data characteristics and inherent challenges. For example, autonomous driving datasets like KITTI and ApolloScape often present high-resolution urban scenes with many object classes, leading to issues such as class imbalance and extensive computational requirements. Precision agriculture datasets like EuroSAT may comprise multispectral or hyperspectral data, adding complexity to data preprocessing and management. Medical imaging datasets like PROMISE12 and BraTS 2015 are generally marked by class imbalance and stringent privacy constraints, thereby posing a distinct set of challenges. The identified challenges were then grouped together where they related to a similar problem. This method enabled us to identify 32 challenges, which are critical challenges in the development of deep learning models for semantic segmentation. These challenges have been classified into six groups and are described in the following section.

4 Results

The analysis of the comprehensive assortment enabled us to compile an exhaustive array of challenges associated with deep learning-based semantic segmentation as seen in Table 2, organized into six categories: (i) Data Quality and Quantity, (ii) Data Preprocessing, (iii) Resource Constraints, (iv) Data Management and Privacy, (v) Generalization, and (vi) Data Compatibility. In the category of **Data Quality and Quantity**, challenges include ensuring accurate annotations (1), incorporating diverse objects into the dataset (2), achieving class balance (3), accounting for data variability (4), managing data resolution (5), handling class overlap (6), ensuring overall data quality (7), and determining the level of detail in annotations (8). The **Data Preprocessing** category involves challenges such as applying augmentation techniques for data diversification (9), aligning data through registration (10), normalizing data for scale invariance (11), standardizing data for consistency (12), scaling data to desired sizes (13), and resolving ambiguity in class labels (14). Within the **Resource Constraints** category, challenges include dealing with lengthy training times

[3] at the time of writing, i.e., April 2023

(15), optimizing computational resources (16), and managing the cost associated with data annotation (17). In the realm of **Data Management and Privacy**, challenges encompass ethical considerations (18), privacy protection (19), anonymization techniques (20), data security (21), access control (22), and effective data storage and management (23). The **Generalization** category involves challenges related to data bias (24), data distribution (25), outlier detection (26), rare events (27), and long-tail distribution (28).

Category	**Challenges**	**Datasets**
Data Quality and Quantity	Annotation Quality	COCO, ImageNet, Pascal VOC 2012
	Object Diversity	ImageNet, Pascal VOC 2012, Objects365
	Data Imbalance	COCO, ImageNet, Pascal VOC 2012, Cityscapes, BDD100K
	Data Variability	NYU Depth V2, KITTI, SUN RGB-D, SUN3D
	Data Resolution	KITTI, ImageNet, Cityscapes
	Class Overlap	COCO, ImageNet, Pascal VOC 2012, Cityscapes
	Data Quality	KITTI, Cityscapes, BDD100K
	Annotation Granularity	ShapeNet, PartNet
Data Preprocessing	Data Augmentation	All datasets
	Data Registration	KITTI, NYU Depth V2, SUN RGB-D, SUN3D
	Normalization	All datasets
	Standardization	All datasets
	Scaling	ImageNet, COCO, Cityscapes
	Class Ambiguity	COCO, ImageNet, Pascal VOC 2012, Cityscapes
Resource Constraints	Training Time	ImageNet, COCO, BDD100K
	Resource Availability	ImageNet, COCO, BDD100K
	Annotation Cost	COCO, ImageNet, Pascal VOC 2012, Cityscapes
Data Management and Privacy	Ethical Considerations	AVE, YouTube-VIS 2019, BDD100K, IDD
	Data Privacy	AVE, YouTube-VIS 2019, BDD100K, IDD
	Data Anonymization	AVE, YouTube-VIS 2019, BDD100K, IDD
	Encryption	All datasets
	Access Control	All datasets
	Data Storage and Management	ImageNet, COCO, BDD100K
Generalization	Data Bias	ImageNet, COCO, Pascal VOC 2012, Cityscapes
	Data Distribution	ImageNet, COCO, Pascal VOC 2012, Cityscapes
	Presence of Outliers	ImageNet, COCO, Pascal VOC 2012
	Rare Events	BDD100K, Cityscapes, KITTI
	Long-tail Distribution	ImageNet, COCO, Pascal VOC 2012, Cityscapes
Data Compatibility	Data Format	All datasets
	Interoperability	All datasets
	Hardware Compatibility	All datasets
	Data Versioning	All datasets

Tab. 2: Summary of challenges in semantic segmentation deep learning datasets with examples

Finally, challenges within the **Data Compatibility** category include handling varying data formats (29), ensuring interoperability between systems (30), optimizing hardware compatibility (31), and managing data versioning (32). These challenges collectively shape the landscape of data-driven research, prompting scientists and scholars to develop innovative solutions and strategies to overcome them and unlock the full potential of data-driven discoveries. Moreover, Table 2 shows that some challenges were identified across two to five datasets, but others (i.e., Data Augmentation, Normalization, Standardization, Encryption, Access Control, Data Format, Interoperability, Hardware Compatibility, and Data Versioning) were observed in all datasets. The challenges observed in the 47 datasets can therefore be seen as priorities, since they are the most widespread, whatever the application domain. Consequently, Data Compatibility, Data Preprocessing, Data Management and Privacy seem to be the most pressing areas to address in order to improve deep learning-based semantic segmentation.

5 Conclusion

This survey has embarked on a comprehensive journey to identify the key obstacles in employing large-scale datasets for deep learning-based semantic segmentation. From 94 papers available in *Papers With Code* repository, we have identified 32 prominent challenges. These challenges span six categories: Data Quality and Quantity, Data Preprocessing, Resource Constraints, Data Management and Privacy, Generalization, and Data Compatibility. We have delved into each category, shedding light on the nuances and intricacies of the challenges therein. Issues pertaining to the quality, volume, and management of data, as well as the computational resources required, have all been examined. Moreover, we have scrutinized the hurdles that affect the model's ability to generalize to unseen data and the difficulties posed by data compatibility issues. This exploration has enabled us to better understand the landscape of semantic segmentation challenges, and to identify the most pressing challenges to be solved. The identified challenges paves the way for future work aimed at devising effective solutions. Looking ahead, our future work will be geared towards developing a solution leveraging artificial intelligence and semantic technologies to overcome these challenges. By building upon the comprehensive understanding of the challenges that this survey offers, we aim to create a framework that is capable of addressing these issues in a holistic and efficient manner. In summary, the survey has laid a solid foundation for future research in this domain, setting the stage for the development of solutions that can truly harness the power of large-scale datasets in semantic segmentation. This represents a significant stride towards advancing the field of computer vision, with potential implications extending across various industries and domains.

Bibliography

[Ac21] Achlioptas, Panos; Ovsjanikov, Maks; Haydarov, Kilichbek; Elhoseiny, Mohamed; Guibas, Leonidas J: Artemis: Affective language for visual art. In: Proceedings of the

IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11569–11579, 2021.

[An21] Anand, Tanmay; Sinha, Soumendu; Mandal, Murari; Chamola, Vinay; Yu, Fei Richard: AgriSegNet: Deep aerial semantic segmentation framework for IoT-assisted precision agriculture. IEEE Sensors Journal, 21(16):17581–17590, 2021.

[BKC17] Badrinarayanan, Vijay; Kendall, Alex; Cipolla, Roberto: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence, 39(12):2481–2495, 2017.

[Bu20] Buslaev, Alexander; Iglovikov, Vladimir I; Khvedchenya, Eugene; Parinov, Alex; Druzhinin, Mikhail; Kalinin, Alexandr A: Albumentations: fast and flexible image augmentations. Information, 11(2):125, 2020.

[Ca19] Casser, Vincent; Pirk, Soeren; Mahjourian, Reza; Angelova, Anelia: Depth prediction without the sensors: Leveraging structure for unsupervised learning from monocular videos. In: Proceedings of the AAAI conference on artificial intelligence. volume 33, pp. 8001–8008, 2019.

[Ca20] Caesar, Holger; Bankiti, Varun; Lang, Alex H; Vora, Sourabh; Liong, Venice Erin; Xu, Qiang; Krishnan, Anush; Pan, Yu; Baldan, Giancarlo; Beijbom, Oscar: nuscenes: A multimodal dataset for autonomous driving. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11621–11631, 2020.

[Ch15] Chang, Angel X; Funkhouser, Thomas; Guibas, Leonidas; Hanrahan, Pat; Huang, Qixing; Li, Zimo; Savarese, Silvio; Savva, Manolis; Song, Shuran; Su, Hao et al.: Shapenet: An information-rich 3d model repository. arXiv preprint arXiv:1512.03012, 2015.

[Ch17a] Chang, Angel; Dai, Angela; Funkhouser, Thomas; Halber, Maciej; Niessner, Matthias; Savva, Manolis; Song, Shuran; Zeng, Andy; Zhang, Yinda: Matterport3d: Learning from rgb-d data in indoor environments. arXiv preprint arXiv:1709.06158, 2017.

[Ch17b] Chen, Liang-Chieh; Papandreou, George; Kokkinos, Iasonas; Murphy, Kevin; Yuille, Alan L: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence, 40(4):834–848, 2017.

[Ch17c] Chen, Liang-Chieh; Papandreou, George; Schroff, Florian; Adam, Hartwig: Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587, 2017.

[Ch20] Chen, Liang-Chieh; Lopes, Raphael Gontijo; Cheng, Bowen; Collins, Maxwell D; Cubuk, Ekin D; Zoph, Barret; Adam, Hartwig; Shlens, Jonathon: Naive-student: Leveraging semi-supervised learning in video sequences for urban scene segmentation. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16. Springer, pp. 695–714, 2020.

[CHS22] Caron, Mathilde; Houlsby, Neil; Schmid, Cordelia: Location-Aware Self-Supervised Transformers. arXiv preprint arXiv:2212.02400, 2022.

[De20] Deng, Jiankang; Guo, Jia; Ververas, Evangelos; Kotsia, Irene; Zafeiriou, Stefanos: Retinaface: Single-shot multi-level face localisation in the wild. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5203–5212, 2020.

[ESL19] Eysenbach, Ben; Salakhutdinov, Russ R; Levine, Sergey: Search on the replay buffer: Bridging planning and reinforcement learning. Advances in Neural Information Processing Systems, 32, 2019.

[Fa22] Fang, Yuxin; Wang, Wen; Xie, Binhui; Sun, Quan; Wu, Ledell; Wang, Xinggang; Huang, Tiejun; Wang, Xinlong; Cao, Yue: Eva: Exploring the limits of masked visual representation learning at scale. arXiv preprint arXiv:2211.07636, 2022.

[GD22] Gesmundo, Andrea; Dean, Jeff: An evolutionary approach to dynamic introduction of tasks in large-scale multitask learning systems. arXiv preprint arXiv:2205.12755, 2022.

[Ge20] Gessert, Nils; Nielsen, Maximilian; Shaikh, Mohsin; Werner, René; Schlaefer, Alexander: Skin lesion classification using ensembles of multi-resolution EfficientNets with meta data. MethodsX, 7:100864, 2020.

[Gh15] Ghifary, Muhammad; Kleijn, W Bastiaan; Zhang, Mengjie; Balduzzi, David: Domain generalization for object recognition with multi-task autoencoders. In: Proceedings of the IEEE international conference on computer vision. pp. 2551–2559, 2015.

[Go19] Godard, Clément; Mac Aodha, Oisin; Firman, Michael; Brostow, Gabriel J: Digging into self-supervised monocular depth estimation. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 3828–3838, 2019.

[Go21] Gong, Tao; Chen, Kai; Wang, Xinjiang; Chu, Qi; Zhu, Feng; Lin, Dahua; Yu, Nenghai; Feng, Huamin: Temporal ROI align for video object recognition. In: Proceedings of the AAAI Conference on Artificial Intelligence. volume 35, pp. 1442–1450, 2021.

[Gu17] Gupta, Saurabh; Davidson, James; Levine, Sergey; Sukthankar, Rahul; Malik, Jitendra: Cognitive mapping and planning for visual navigation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2616–2625, 2017.

[Gu20] Guo, Yulan; Wang, Hanyun; Hu, Qingyong; Liu, Hao; Liu, Li; Bennamoun, Mohammed: Deep learning for 3d point clouds: A survey. IEEE transactions on pattern analysis and machine intelligence, 43(12):4338–4364, 2020.

[GVZ16] Gupta, Ankush; Vedaldi, Andrea; Zisserman, Andrew: Synthetic data for text localisation in natural images. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2315–2324, 2016.

[Ha22] Hao, Yuying; Liu, Yi; Chen, Yizhou; Han, Lin; Peng, Juncai; Tang, Shiyu; Chen, Guowei; Wu, Zewu; Chen, Zeyu; Lai, Baohua: EISeg: An Efficient Interactive Segmentation Tool based on PaddlePaddle. arXiv e-prints, pp. arXiv–2210, 2022.

[He16] He, Kaiming; Zhang, Xiangyu; Ren, Shaoqing; Sun, Jian: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778, 2016.

[He17] He, Kaiming; Gkioxari, Georgia; Dollár, Piotr; Girshick, Ross: Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp. 2961–2969, 2017.

[Ho17] Howard, Andrew G; Zhu, Menglong; Chen, Bo; Kalenichenko, Dmitry; Wang, Weijun; Weyand, Tobias; Andreetto, Marco; Adam, Hartwig: Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861, 2017.

[Hu18] Huang, Xun; Liu, Ming-Yu; Belongie, Serge; Kautz, Jan: Multimodal unsupervised image-to-image translation. In: Proceedings of the European conference on computer vision (ECCV). pp. 172–189, 2018.

[Hu19] Huang, Xinyu; Wang, Peng; Cheng, Xinjing; Zhou, Dingfu; Geng, Qichuan; Yang, Ruigang: The apolloscape open dataset for autonomous driving and its application. IEEE transactions on pattern analysis and machine intelligence, 42(10):2702–2719, 2019.

[Hu20] Hu, Qingyong; Yang, Bo; Xie, Linhai; Rosa, Stefano; Guo, Yulan; Wang, Zhihua; Trigoni, Niki; Markham, Andrew: Randla-net: Efficient semantic segmentation of large-scale point clouds. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11108–11117, 2020.

[Hu22] Hu, Hou-Ning; Yang, Yung-Hsu; Fischer, Tobias; Darrell, Trevor; Yu, Fisher; Sun, Min: Monocular quasi-dense 3d object tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(2):1992–2008, 2022.

[Is18a] Isensee, Fabian; Kickingereder, Philipp; Wick, Wolfgang; Bendszus, Martin; Maier-Hein, Klaus H: Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: Third International Workshop, BrainLes 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, September 14, 2017, Revised Selected Papers 3. Springer, pp. 287–297, 2018.

[Is18b] Isensee, Fabian; Petersen, Jens; Klein, Andre; Zimmerer, David; Jaeger, Paul F; Kohl, Simon; Wasserthal, Jakob; Koehler, Gregor; Norajitra, Tobias; Wirkert, Sebastian et al.: nnu-net: Self-adapting framework for u-net-based medical image segmentation. arXiv preprint arXiv:1809.10486, 2018.

[Is19] Isensee, Fabian; Jäger, Paul F; Kohl, Simon AA; Petersen, Jens; Maier-Hein, Klaus H: Automated design of deep learning methods for biomedical image segmentation. arXiv preprint arXiv:1904.08128, 2019.

[Jh20] Jha, Debesh; Riegler, Michael A; Johansen, Dag; Halvorsen, Pål; Johansen, Håvard D: Doubleu-net: A deep convolutional neural network for medical image segmentation. In: 2020 IEEE 33rd International symposium on computer-based medical systems (CBMS). IEEE, pp. 558–564, 2020.

[Ji22] Jiao, Rushi; Zhang, Yichi; Ding, Le; Cai, Rong; Zhang, Jicong: Learning with limited annotations: a survey on deep semi-supervised learning for medical image segmentation. arXiv preprint arXiv:2207.14191, 2022.

[Ka17] Kamnitsas, Konstantinos; Ledig, Christian; Newcombe, Virginia FJ; Simpson, Joanna P; Kane, Andrew D; Menon, David K; Rueckert, Daniel; Glocker, Ben: Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Medical image analysis, 36:61–78, 2017.

[Ke17] Ke, Guolin; Meng, Qi; Finley, Thomas; Wang, Taifeng; Chen, Wei; Ma, Weidong; Ye, Qiwei; Liu, Tie-Yan: Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems, 30, 2017.

[Ki19] Kirillov, Alexander; Girshick, Ross; He, Kaiming; Dollár, Piotr: Panoptic feature pyramid networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 6399–6408, 2019.

[Ku22] Kuo, Weicheng; Cui, Yin; Gu, Xiuye; Piergiovanni, AJ; Angelova, Anelia: F-VLM: Open-Vocabulary Object Detection upon Frozen Vision and Language Models. arXiv preprint arXiv:2209.15639, 2022.

[LE22] Lüddecke, Timo; Ecker, Alexander: Image segmentation using text and image prompts. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7086–7096, 2022.

[Li19] Liu, Chenxi; Chen, Liang-Chieh; Schroff, Florian; Adam, Hartwig; Hua, Wei; Yuille, Alan L; Fei-Fei, Li: Auto-deeplab: Hierarchical neural architecture search for semantic image segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 82–92, 2019.

[Li21] Li, Junnan; Selvaraju, Ramprasaath; Gotmare, Akhilesh; Joty, Shafiq; Xiong, Caiming; Hoi, Steven Chu Hong: Align before fuse: Vision and language representation learning with momentum distillation. Advances in neural information processing systems, 34:9694–9705, 2021.

[LSD15a] Long, Jonathan; Shelhamer, Evan; Darrell, Trevor: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3431–3440, 2015.

[LSD15b] Long, Jonathan; Shelhamer, Evan; Darrell, Trevor: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3431–3440, 2015.

[MNA16] Milletari, Fausto; Navab, Nassir; Ahmadi, Seyed-Ahmad: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision (3DV). Ieee, pp. 565–571, 2016.

[Mo19a] Mo, Kaichun; Zhu, Shilin; Chang, Angel X; Yi, Li; Tripathi, Subarna; Guibas, Leonidas J; Su, Hao: Partnet: A large-scale benchmark for fine-grained and hierarchical part-level 3d object understanding. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 909–918, 2019.

[Mo19b] Mousavian, Arsalan; Toshev, Alexander; Fišer, Marek; Košecká, Jana; Wahid, Ayzaan; Davidson, James: Visual representations for semantic target driven navigation. In: 2019 International Conference on Robotics and Automation (ICRA). IEEE, pp. 8846–8852, 2019.

[Mu19] Munappy, Aiswarya; Bosch, Jan; Olsson, Helena Holmström; Arpteg, Anders; Brinne, Björn: Data management challenges for deep learning. In: 2019 45th Euromicro Conference on Software Engineering and Advanced Applications (SEAA). IEEE, pp. 140–147, 2019.

[MV21] Mohan, Rohit; Valada, Abhinav: Efficientps: Efficient panoptic segmentation. International Journal of Computer Vision, 129(5):1551–1579, 2021.

[MWA18a] Mahjourian, Reza; Wicke, Martin; Angelova, Anelia: Unsupervised learning of depth and ego-motion from monocular video using 3d geometric constraints. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 5667–5675, 2018.

[MWA18b] Mahjourian, Reza; Wicke, Martin; Angelova, Anelia: Unsupervised learning of depth and ego-motion from monocular video using 3d geometric constraints. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 5667–5675, 2018.

[Ne19] Neumann, Maxim; Pinto, Andre Susano; Zhai, Xiaohua; Houlsby, Neil: In-domain representation learning for remote sensing. arXiv preprint arXiv:1911.06721, 2019.

[Oh18] Oh, Seoung Wug; Lee, Joon-Young; Sunkavalli, Kalyan; Kim, Seon Joo: Fast video object segmentation by reference-guided mask propagation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7376–7385, 2018.

[Oq23] Oquab, Maxime; Darcet, Timothée; Moutakanni, Théo; Vo, Huy; Szafraniec, Marc; Khalidov, Vasil; Fernandez, Pierre; Haziza, Daniel; Massa, Francisco; El-Nouby, Alaaeldin et al.: DINOv2: Learning Robust Visual Features without Supervision. arXiv preprint arXiv:2304.07193, 2023.

[Pa16] Paszke, Adam; Chaurasia, Abhishek; Kim, Sangpil; Culurciello, Eugenio: Enet: A deep neural network architecture for real-time semantic segmentation. arXiv preprint arXiv:1606.02147, 2016.

[Po19] Ponciano, J-J; Karmacharya, A; Wefers, S; Atorf, P; Boochs, F: Connected Semantic Concepts as a Base for Optimal Recording and Computer-Based Modelling of Cultural Heritage Objects. In: Structural Analysis of Historical Constructions: An Interdisciplinary Approach. Springer, pp. 297–304, 2019.

[Po21] Ponciano, Jean-Jacques; Roetner, Moritz; Reiterer, Alexander; Boochs, Frank: Object Semantic Segmentation in Point Clouds—Comparison of a Deep Learning and a Knowledge-Based Method. ISPRS International Journal of Geo-Information, 10(4):256, 2021.

[PPB21] Ponciano, Jean-Jacques; Prudhomme, Claire; Boochs, Frank: From Acquisition to presentation—The potential of semantics to support the safeguard of cultural heritage. Remote Sensing, 13(11):2226, 2021.

[PPP21] Plaß, B; Prudhomme, C; Ponciano, JJ: Bim on Artificial Intelligence for Decision Support in E-Health. The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 43:207–214, 2021.

[Qi17a] Qi, Charles R; Su, Hao; Mo, Kaichun; Guibas, Leonidas J: Pointnet: Deep learning on point sets for 3d classification and segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 652–660, 2017.

[Qi17b] Qi, Charles Ruizhongtai; Yi, Li; Su, Hao; Guibas, Leonidas J: Pointnet++: Deep hierarchical feature learning on point sets in a metric space. Advances in neural information processing systems, 30, 2017.

[Re16] Redmon, Joseph; Divvala, Santosh; Girshick, Ross; Farhadi, Ali: You only look once: Unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 779–788, 2016.

[Re19] Ren, Zhile; Gallo, Orazio; Sun, Deqing; Yang, Ming-Hsuan; Sudderth, Erik B; Kautz, Jan: A fusion approach for multi-frame optical flow estimation. In: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, pp. 2077–2086, 2019.

[RFB15] Ronneberger, Olaf; Fischer, Philipp; Brox, Thomas: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer, pp. 234–241, 2015.

[Sa20] Sarlin, Paul-Edouard; DeTone, Daniel; Malisiewicz, Tomasz; Rabinovich, Andrew: Superglue: Learning feature matching with graph neural networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4938–4947, 2020.

[Sh19] Sharma, Ochin: Deep Challenges Associated with Deep Learning. In: 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon). pp. 72–75, 2019.

[St21] Stone, Austin; Ramirez, Oscar; Konolige, Kurt; Jonschkowski, Rico: The Distracting Control Suite–A Challenging Benchmark for Reinforcement Learning from Pixels. arXiv preprint arXiv:2101.02722, 2021.

[Su18] Suwajanakorn, Supasorn; Snavely, Noah; Tompson, Jonathan J; Norouzi, Mohammad: Discovery of latent 3d keypoints via end-to-end geometric reasoning. Advances in neural information processing systems, 31, 2018.

[Su19] Sun, Ke; Zhao, Y; Jiang, B; Cheng, T; Xiao, B; Liu, D; Mu, Y; Wang, X; Liu, W; Wang, J: High-resolution representations for labeling pixels and regions. arXiv 2019. arXiv preprint arXiv:1904.04514, 2019.

[Su23] Su, Yukun; Deng, Jingliang; Sun, Ruizhou; Lin, Guosheng; Su, Hanjing; Wu, Qingyao: A unified transformer framework for group-based segmentation: Co-segmentation, co-saliency detection and video salient object detection. IEEE Transactions on Multimedia, 2023.

[SZ14] Simonyan, Karen; Zisserman, Andrew: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

[Ta21] Tan, Haobin; Chen, Chang; Luo, Xinyu; Zhang, Jiaming; Seibold, Constantin; Yang, Kailun; Stiefelhagen, Rainer: Flying guide dog: Walkable path discovery for the visually impaired utilizing drones and transformer-based semantic segmentation. In: 2021 IEEE International Conference on Robotics and Biomimetics (ROBIO). IEEE, pp. 1123–1128, 2021.

[Vo19] Voigtlaender, Paul; Chai, Yuning; Schroff, Florian; Adam, Hartwig; Leibe, Bastian; Chen, Liang-Chieh: Feelvos: Fast end-to-end embedding learning for video object segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9481–9490, 2019.

[Wa20a] Wang, Jingdong; Sun, Ke; Cheng, Tianheng; Jiang, Borui; Deng, Chaorui; Zhao, Yang; Liu, Dong; Mu, Yadong; Tan, Mingkui; Wang, Xinggang et al.: Deep high-resolution representation learning for visual recognition. IEEE transactions on pattern analysis and machine intelligence, 43(10):3349–3364, 2020.

[Wa20b] Wang, Jingdong; Sun, Ke; Cheng, Tianheng; Jiang, Borui; Deng, Chaorui; Zhao, Yang; Liu, Dong; Mu, Yadong; Tan, Mingkui; Wang, Xinggang et al.: Deep high-resolution representation learning for visual recognition. IEEE transactions on pattern analysis and machine intelligence, 43(10):3349–3364, 2020.

[Wa22] Wang, Jianfeng; Yang, Zhengyuan; Hu, Xiaowei; Li, Linjie; Lin, Kevin; Gan, Zhe; Liu, Zicheng; Liu, Ce; Wang, Lijuan: Git: A generative image-to-text transformer for vision and language. arXiv preprint arXiv:2205.14100, 2022.

[WBP17] Wojke, Nicolai; Bewley, Alex; Paulus, Dietrich: Simple online and realtime tracking with a deep association metric. In: 2017 IEEE international conference on image processing (ICIP). IEEE, pp. 3645–3649, 2017.

[WH21] Wang, Weijun; Howard, Andrew: MOSAIC: Mobile Segmentation via decoding Aggregated Information and encoded Context. arXiv preprint arXiv:2112.11623, 2021.

[Xi22] Xiao, Aoran; Huang, Jiaxing; Guan, Dayan; Zhan, Fangneng; Lu, Shijian: Transfer learning from synthetic to real lidar point cloud for semantic segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence. volume 36, pp. 2795–2803, 2022.

[Yu16] Yu, Licheng; Poirson, Patrick; Yang, Shan; Berg, Alexander C; Berg, Tamara L: Modeling context in referring expressions. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14. Springer, pp. 69–85, 2016.

[Yu20] Yu, Chaohui; Wang, Jindong; Liu, Chang; Qin, Tao; Xu, Renjun; Feng, Wenjie; Chen, Yiqiang; Liu, Tie-Yan: Learning to match distributions for domain adaptation. arXiv preprint arXiv:2007.10791, 2020.

[Yu21] Yu, Changqian; Gao, Changxin; Wang, Jingbo; Yu, Gang; Shen, Chunhua; Sang, Nong: Bisenet v2: Bilateral network with guided aggregation for real-time semantic segmentation. International Journal of Computer Vision, 129:3051–3068, 2021.

[Zh18a] Zhao, Jian; Li, Jianshu; Cheng, Yu; Sim, Terence; Yan, Shuicheng; Feng, Jiashi: Understanding humans in crowded scenes: Deep nested adversarial learning and a new benchmark for multi-human parsing. In: Proceedings of the 26th ACM international conference on Multimedia. pp. 792–800, 2018.

[Zh18b] Zhou, Zongwei; Rahman Siddiquee, Md Mahfuzur; Tajbakhsh, Nima; Liang, Jianming: Unet++: A nested u-net architecture for medical image segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4. Springer, pp. 3–11, 2018.

[Zh19] Zhou, Bolei; Zhao, Hang; Puig, Xavier; Xiao, Tete; Fidler, Sanja; Barriuso, Adela; Torralba, Antonio: Semantic understanding of scenes through the ade20k dataset. International Journal of Computer Vision, 127:302–321, 2019.

[Zh20a] Zhao, Sicheng; Yue, Xiangyu; Zhang, Shanghang; Li, Bo; Zhao, Han; Wu, Bichen; Krishna, Ravi; Gonzalez, Joseph E; Sangiovanni-Vincentelli, Alberto L; Seshia, Sanjit A et al.: A review of single-source deep unsupervised visual domain adaptation. IEEE Transactions on Neural Networks and Learning Systems, 33(2):473–493, 2020.

[Zh20b] Zhao, Sicheng; Yue, Xiangyu; Zhang, Shanghang; Li, Bo; Zhao, Han; Wu, Bichen; Krishna, Ravi; Gonzalez, Joseph E; Sangiovanni-Vincentelli, Alberto L; Seshia, Sanjit A et al.: A review of single-source deep unsupervised visual domain adaptation. IEEE Transactions on Neural Networks and Learning Systems, 33(2):473–493, 2020.

[Zh20c] Zhao, Yinan; Price, Brian; Cohen, Scott; Gurari, Danna: Objectness-aware few-shot semantic segmentation. arXiv preprint arXiv:2004.02945, 2020.

[Zh22a] Zheng, Qingping; Deng, Jiankang; Zhu, Zheng; Li, Ying; Zafeiriou, Stefanos: Decoupled Multi-task Learning with Cyclical Self-Regulation for Face Parsing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4156–4165, 2022.

[Zh22b] Zhou, Jinxing; Wang, Jianyuan; Zhang, Jiayi; Sun, Weixuan; Zhang, Jing; Birchfield, Stan; Guo, Dan; Kong, Lingpeng; Wang, Meng; Zhong, Yiran: Audio–Visual Segmentation. In: Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXVII. Springer, pp. 386–403, 2022.

[Zh23] Zhou, Jinxing; Shen, Xuyang; Wang, Jianyuan; Zhang, Jiayi; Sun, Weixuan; Zhang, Jing; Birchfield, Stan; Guo, Dan; Kong, Lingpeng; Wang, Meng et al.: Audio-Visual Segmentation with Semantics. arXiv preprint arXiv:2301.13190, 2023.

[Zo23] Zou, Xueyan; Yang, Jianwei; Zhang, Hao; Li, Feng; Li, Linjie; Gao, Jianfeng; Lee, Yong Jae: Segment everything everywhere all at once. arXiv preprint arXiv:2304.06718, 2023.