1. A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, “A review on deep learning techniques applied to semantic segmentation,” 2017. [Online]. Available: <https://arxiv.org/abs/1704.06857>
2. H. Yu, Z. Yang, L. Tan, Y. Wang, W. Sun, M. Sun, and Y. Tang, “Methods and datasets on semantic segmentation: A review,” Neurocomput., vol. 304, no. C, p. 82–103, aug 2018. [Online]. Available: [https://doi.org/10.1016/j.neucom.2018. 03.037](https://doi.org/10.1016/j.neucom.2018.%2003.037)
3. B. Emek Soylu, M. S. Guzel, G. E. Bostanci, F. Ekinci, T. Asuroglu, and K. Acici, “Deep-learning-based approaches for semantic segmentation of natural scene images: A review,” Electronics, vol. 12, no. 12, 2023. [Online]. Available: <https://www.mdpi.com/2079-9292/12/12/2730>
4. Y. Guo, Y. Liu, T. Georgiou, and M. Lew, “A review of semantic segmentation using deep neural networks,” International Journal of Multimedia Information Retrieval, vol. 7, 06 2018.
5. M. Oberweger, P. Wohlhart, and V. Lepetit, “Hands deep in deep learning for hand pose estimation,” 2016. [Online]. Available: <https://arxiv.org/abs/1502.06807>
6. M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, “The cityscapes dataset for semantic urban scene understanding,” 2016. [Online]. Available: https://arxiv.org/abs/1604.01685.
7. A. Geiger, P. Lenz, and R. Urtasun, “Are we ready for autonomous driving? the kitti vision benchmark suite,” in 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 3354–3361.
8. A. Ess, T. Mueller, H. Grabner, and L. Van Gool, “Segmentation-based urban traffic scene understanding.” in BMVC, vol. 1. Citeseer, 2009, p. 2.
9. Y. Yoon, H.-G. Jeon, D. Yoo, J.-Y. Lee, and I. So Kweon, “Learning a deep convolutional network for light-field image super-resolution,” in Proceedings of the IEEE international conference on computer vision workshops, 2015, pp. 24–32.
10. D. G. Lowe, “Object recognition from local scale-invariant features,” Proceedings of the Seventh IEEE International Conference on Computer Vision, vol. 2, pp. 1150–1157 vol.2, 1999. [Online]. Available: https://api.semanticscholar.org/ CorpusID:5258236
11. N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05), vol. 1, 2005, pp. 886–893 vol. 1.
12. N. Silberman and R. Fergus, “Indoor scene segmentation using a structured light sensor,” 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), pp. 601–608, 2011. [Online]. Available: https://api.semanticscholar.org/CorpusID:13993169
13. S. Gupta, P. Arbelaez, and J. Malik, “Perceptual organization and recognition ´ of indoor scenes from rgb-d images,” 2013 IEEE Conference on Computer Vision and Pattern Recognition, pp. 564–571, 2013. [Online]. Available: https://api.semanticscholar.org/CorpusID:12061055
14. A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” Commun. ACM, vol. 60, no. 6, p. 84–90, may 2017. [Online]. Available: https://doi.org/10.1145/3065386
15. K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2015. [Online]. Available: <https://arxiv.org/abs/1409.1556>
16. G. E. Hinton, S. Osindero, and Y.-W. Teh, “A fast learning algorithm for deep belief nets,” Neural Comput., vol. 18, no. 7, p. 1527–1554, jul 2006. [Online]. Available: https://doi.org/10.1162/neco.2006.18.7.1527
17. Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” nature, vol. 521, no. 7553, pp. 436–444, 2015.
18. C. Farabet, C. Couprie, L. Najman, and Y. Lecun, “Learning hierarchical features for scene labeling,” IEEE transactions on pattern analysis and machine intelligence, vol. 35, pp. 1915–1929, 08 2013.
19. L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “Semantic image segmentation with deep convolutional nets and fully connected crfs,” 2016. [Online]. Available: https://arxiv.org/abs/1412.7062
20. R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” 2014. [Online]. Available: https://arxiv.org/abs/1311.2524
21. D. Eigen and R. Fergus, “Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture,” 2015. [Online]. Available: https://arxiv.org/abs/1411.4734
22. Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” Proceedings of the IEEE, vol. 86, no. 11, pp. 2278– 2324, 1998.
23. J. Shotton, J. Winn, C. Rother, and A. Criminisi, “Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation,” 07 2006, pp. 1–15.
24. P. Kohli, L. Ladicky, and P. Torr, “Robust higher order potentials for enforcing ´ label consistency,” International Journal of Computer Vision, vol. 82, pp. 302– 324, 05 2009.
25. L. Ladicky, C. Russell, P. Kohli, and P. Torr, “Associative hierarchical crfs for object class image segmentation,” 11 2009, pp. 739 – 746.
26. J. Verbeek and B. Triggs, “Region classification with markov field aspect models,” 06 2007.
27. M. Oquab, L. Bottou, I. Laptev, and J. Sivic, “Learning and transferring mid-level image representations using convolutional neural networks,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2014.
28. J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, “How transferable are features in deep neural networks?” 2014. [Online]. Available: https: //arxiv.org/abs/1411.1792
29. S. C. Wong, A. Gatt, V. Stamatescu, and M. D. McDonnell, “Understanding data augmentation for classification: when to warp?” 2016. [Online]. Available: <https://arxiv.org/abs/1609.08764>
30. O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” 2015. [Online]. Available: https: //arxiv.org/abs/1505.04597
31. J. Huang, G. Liu, and B. Wang, “Semantic Segmentation under a Complex Background for Machine Vision Detection Based on Modified UPerNet with Component Analysis Modules,” Mathematical Problems in Engineering, vol. 2020, pp. 1–13, September 2020. [Online]. Available: https://ideas.repec.org/a/ hin/jnlmpe/6903130.html
32. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.
33. J. Pont-Tuset, P. Arbelaez, J. T.Barron, F. Marques, and J. Malik, “Multiscale combinatorial grouping for image segmentation and object proposal generation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 1, p. 128–140, Jan. 2017. [Online]. Available: [http://dx.doi.org/10.1109/TPAMI. 2016.2537320](http://dx.doi.org/10.1109/TPAMI.%202016.2537320)
34. K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2015. [Online]. Available: <https://arxiv.org/abs/1409.1556>
35. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” 2014. [Online]. Available: <https://arxiv.org/abs/1409.4842>
36. B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso, and A. Torralba, “Scene parsing through ade20k dataset,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017.
37. H. Caesar, J. Uijlings, and V. Ferrari, “Coco-stuff: Thing and stuff classes in context,” 2018. [Online]. Available: https://arxiv.org/abs/1612.03716
38. M. Everingham, S. A. Eslami, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, “The pascal visual object classes challenge: A retrospective,” International journal of computer vision, vol. 111, pp. 98–136, 2015.
39. R. Mottaghi, X. Chen, X. Liu, N.-G. Cho, S.-W. Lee, S. Fidler, R. Urtasun, and A. Yuille, “The role of context for object detection and semantic segmentation in the wild,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 891–898.
40. X. Chen, R. Mottaghi, X. Liu, S. Fidler, R. Urtasun, and A. Yuille, “Detect what you can: Detecting and representing objects using holistic models and body parts,” 2014. [Online]. Available: https://arxiv.org/abs/1406.2031
41. N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, “Indoor segmentation and support inference from rgbd images,” in European Conference on Computer Vision, 2012. [Online]. Available: https://api.semanticscholar.org/CorpusID: 545361
42. S. Song, S. P. Lichtenberg, and J. Xiao, “Sun rgb-d: A rgb-d scene understanding benchmark suite,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2015.
43. J. Xiao, A. Owens, and A. Torralba, “Sun3d: A database of big spaces reconstructed using sfm and object labels,” in Computer Vision (ICCV), 2013 IEEE International Conference on. IEEE, 2013, pp. 1625–1632.
44. B. Hariharan, P. Arbelaez, L. Bourdev, S. Maji, and J. Malik, “Semantic contours ´ from inverse detectors,” in 2011 International Conference on Computer Vision, 2011, pp. 991–998.
45. G. Ros, L. Sellart, J. Materzynska, D. Vazquez, and A. 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 (CVPR), June 2016.
46. F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell, “Bdd100k: A diverse driving dataset for heterogeneous multitask learning,” 2020. [Online]. Available: https://arxiv.org/abs/1805.04687
47. G. J. Brostow, J. Fauqueur, and R. Cipolla, “Semantic object classes in video: A high-definition ground truth database,” Pattern Recognition Letters, vol. 30, no. 2, pp. 88–97, Jan. 2009.
48. M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, “The cityscapes dataset for semantic urban scene understanding,” in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 3213–3223.
49. A. Prest, C. Leistner, J. Civera, C. Schmid, and V. Ferrari, “Learning object class detectors from weakly annotated video,” in 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 3282–3289.
50. S. Bell, P. Upchurch, N. Snavely, and K. Bala, “Material recognition in the wild with the materials in context database,” 2015. [Online]. Available: https://arxiv.org/abs/1412.0623
51. A. Geiger, P. Lenz, and R. Urtasun, “Are we ready for autonomous driving? the kitti vision benchmark suite,” in 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 3354–3361.
52. A. Kapitanov, K. Kvanchiani, and K. Sofia, “Easyportrait - face parsing and portrait segmentation dataset,” arXiv preprint arXiv:2304.13509, 2023.
53. F. Perazzi, J. Pont-Tuset, B. McWilliams, L. Van Gool, M. Gross, and A. SorkineHornung, “A benchmark dataset and evaluation methodology for video object segmentation,” in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 724–732.
54. C. Liu, J. Yuen, and A. Torralba, “Sift flow: Dense correspondence across scenes and its applications,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 5, pp. 978–994, 2011.
55. A. Richtsfeld, “The object segmentation database(osd),” 2012.
56. S. Gould, R. Fulton, and D. Koller, “Decomposing a scene into geometric and semantically consistent regions,” 2009 IEEE 12th International Conference on Computer Vision, pp. 1–8, 2009. [Online]. Available: https://api.semanticscholar. org/CorpusID:17448963
57. K. Lai, L. Bo, X. Ren, and D. Fox, “A large-scale hierarchical multi-view rgb-d object dataset,” in 2011 IEEE International Conference on Robotics and Automation, 2011, pp. 1817–1824.
58. T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, “Focal loss for dense object ´ detection,” 2018.
59. O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” 2015.