Reference

1. Y. Lecun, L. Bottou, Y. Bengio, et al., “Gradient-Based Learning Applied to Document Recognition”, *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-3324, 1998.
2. A. Krizhevsky, I. Sutskever, G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks”, *Advances in Neural Information Processing Systems*, vol. 25, pp. 1106-1114, 2012.
3. K. Simonyan, A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition”, *3rd International Conference on Learning Representations (ICLR 2015)*, 2015.
4. K. He, X. Zhang, S. Ren, et al., “Deep Residual Learning for Image Recognition”, *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016*), pp. 770-778, 2016.
5. J. Long, E. Shelhamer, T. Darrell, “Fully convolutional networks for semantic segmentation”, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2015)*, pp. 3431-3440, 2015.
6. O. Ronneberger, P. Fischer, T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation”, *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015 - 18th International Conference Munich*, vol. 9351, pp. 234-241, 2015.
7. F. Milletari, N. Navab, S. A. Ahmadi, “V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation”, *Fourth International Conference on 3D Vision (3DV 2016)*, pp. 565-571, 2016.
8. B. Hariharan, P. A. Arbeláez, R. B. Girshick, et al., “Simultaneous Detection and Segmentation”, *Computer Vision - ECCV 2014 - 13th European Conference*, vol. 8695, pp. 297-312, 2014.
9. K. He, G. Gkioxari, P. Dollár, et al., “Mask R-CNN”, *IEEE International Conference on Computer Vision* *(ICCV 2017)*, pp. 2980-2988, 2017.
10. T. Y. Lin, M. Maire, S. J. Belongie, et al., “Microsoft COCO: Common Objects in Context”, *Computer Vision - ECCV 2014 - 13th European Conference*, pp. 740-755, 2014.
11. D. Bolya, C. Zhou, F. Xiao, et al., “YOLACT: Real-Time Instance Segmentation”, *2019 IEEE/CVF International Conference on Computer Vision (ICCV 2019*), pp. 9156-9165, 2019.
12. A. Kirillov, K. He, R. B. Girshick, et al., “Panoptic Segmentation”, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2019)*, pp. 9404-9413, 2019.
13. A. Vaswani, N. Shazeer, N. Parmar, et al., “Attention is All You Need”, *Advances in Neural Information Processing Systems: Vol. 30*, pp. 5998-6008, 2017.
14. A. Dosovitskiy, L. Beyer, A. Kolesnikov, et al., “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale”, *9th International Conference on Learning Representations (ICLR 2021)*, 2021.
15. N. Carion, F. Massa, G. Synnaeve, et al., “End-to-End Object Detection with Transformers”, *Computer Vision - ECCV 2020 - 16th European Conference*, vol. 12346, pp. 213-229, 2020.
16. S. Zheng, J. Lu, H. Zhao, et al., “Rethinking Semantic Segmentation From a Sequence-to-Sequence Perspective With Transformers”, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2021*), pp. 6881-6890, 2021.
17. B. Cheng, A. G. Schwing, A. Kirillov, “Per-Pixel Classification is Not All You Need for Semantic Segmentation”, *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021 (NeurIPS 2021)*, pp. 17864-17875, 2021.