Proceedings of the 8th World Congress on Electrical Engineering and Computer Systems and Sciences (EECSS'22)
Prague, Czech Republic - July 28- 30, 2022
Paper No. MVML 101
DOI: 10.11159/mvml22.101

Applying Deep Learning for Image Segmentation: A Survey

Md Jamiul Alam Khan[1], Jing Ren[1], Hossam A. Gabbar [2]
[1]Faculty of Engineering and Applied Science, Ontario Tech University
2000 Simcoe Street North, Oshawa, Ontario, Canada
mdjamiul.khan@ontariotechu.net; jing.ren@ontariotechu.net
[2]Faculty of Energy Systems and Nuclear Science, Ontario Tech University
2000 Simcoe Street North, Oshawa, Ontario, Canada
hossam.gaber@ontariotechu.net

Abstract **-** Image segmentation is one of the most important branches of image processing. But it comes with various challenges and problems to be solved. Researchers are always working on improving the accuracy, quality and performance of image segmentation techniques. As in modern days, deep learning being involved in almost all problem solving, it is being used in image segmentation too. In this paper, we discussed few image segmentation techniques developed using deep learning, some implementation of these techniques to applications. And lastly, we addressed some limitations, challenges and research scopes for future.

Keywords**:** image segmentation, deep learning, semantic segmentation, instance segmentation, panoptic segmentation

1. Introduction

In Image processing, the image segmentation refers to partitioning an image or its pixels into different classes or segments according to their properties and features [1]. It is the very first step in many image and visual processing based applications. Use of Image segmentation can be seen in medical science, autonomous vehicles, augmented reality, video surveillance etc. [2].

Various techniques for image segmentation have been developed over the years. Thresholding [3], region-growing [4], watershed [5], clustering [6], graph-cuts [7], active contours [8] and many others. But different images and applications comes with difference challenges, so the performance of each algorithm varies in different conditions. As a result, this field still has plenty of scope for research. As in recent years deep learning has emerged as solution to almost any problem, image segmentation is no different. There are many deep learning-based approaches are developed and researched for image segmentation, often resulting in better performances compared to the traditional methods.

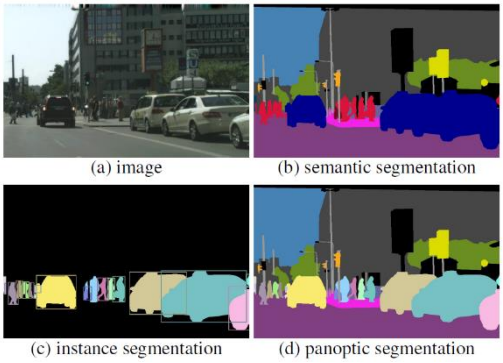

Fig. 1: Different types of Image segmentation [9].

Image segmentation problem can mainly be divided into 2 different approaches. First, Semantic segmentation, where all the pixels of an image is segmented such that all pixels that belongs to a particular class are represented as one segment. Fig. 1 (a) shows an image of a road that has multiple person and cars in it. In the next image (b) the semantic segmentation

is applied. All the cars are shown in blue colour indicating them as same class, although there are multiple cars. Next approach of image segmentation is instance segmentation, where different object of the same class, divided as segments. In Fig. 1. (c), instance segmentation is applied to the image, the individual cars now are shown in different colour, means they are labelled as different classes. There is a third type of labelling proposed recently, Panoptic segmentation [9], which combines both semantic and instance segmentation together. In this approach all the pixels belonging to a single class are identified and then the instance of objects belonging to a class is also identified.

In section 2, some techniques of image segmentation are discussed briefly. The techniques are classified according to their segmentation types.

In section 3, some examples of application of deep learning-based techniques to real world problems are mentioned.

Section 4 addresses few limitations and challenges including computation expense, memory efficiency, dataset etc. regarding image segmentation techniques.

2. Image Segmentation Techniques

Several deep learning-based techniques of image segmentations are classified based on their respective resulting segmentation. Few techniques are discussed in the following subsections.

2.1. Semantic Segmentation

Fully convolutional Networks (FCNs) for segmentation [10], proposed by Long et al. is a milestone of applying deep learning to image segmentation. The key feature of this architecture is that is takes arbitrary sized images as input and results in same size image as output. Where the architecture existing during that time used to take input fixed size images as they have a fully connected layer that caused the limitations. The architecture also consists of up sampling steps with skip connections which fuses the feature maps and earlier layers from convolutions to generate the segments.

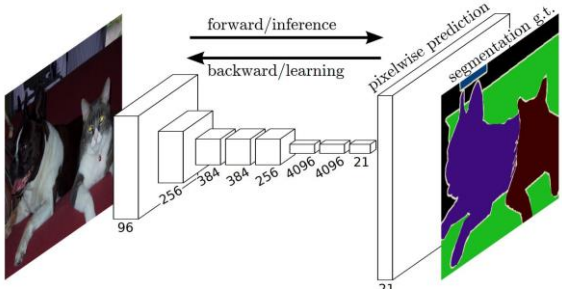

Fig. 2: Image segmentation using FCNs [10].

In traditional CNN architecture after the steps of convolutions, Relu and Pooling, the feature maps are then flattened in fully connected layer and then linked to the final classes. In the approach of FCNs, traditional CNN such as VGG16 and GoogLeNet are modified by removing the fully connected layer so that the architecture returns a spatial segmentation map instead of classification class. Then the spatial segmentation map is resized back to the original size using the upsampling and skip connection techniques. Here the feature map from the final layer of the model is upsampled and then combined with the feature map of the earlier layers to get an accurate segmentation.

Fig. 3 shows how the layers are combined and the prediction is generated. The first row of the image shows upsampling in a single step where the other rows show combining predictions from previous pooling layers. Combining information from previous layers results in higher precision in segmentation. The proposed algorithm was applied to several datasets and showed relative improvement in results compared to other algorithms.

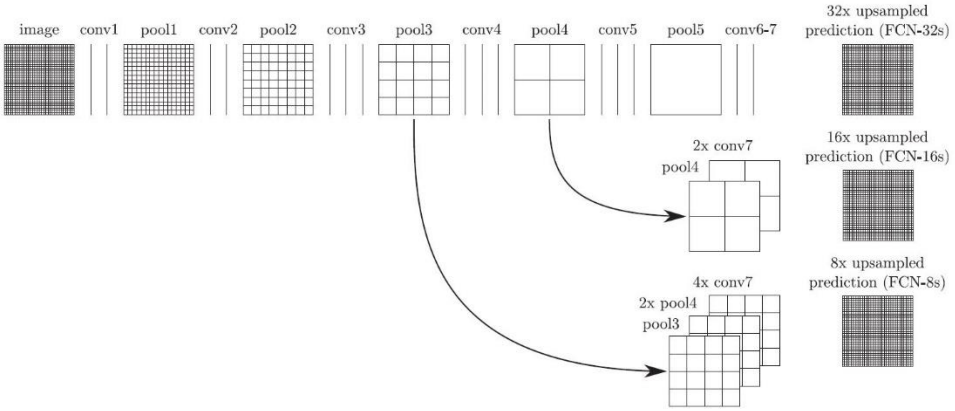

Fig. 3: Upsampling and skip connection used for segmentation prediction [10].

Keeping this FCNs as base step, there are many solutions to segmentation problems has been proposed in various fields such as in medical image processing [11] [12]. Another model developed primarily for biomedical images is U-Net [13]. The Fig.4 shows the architecture to be a symmetrical architecture with two parts. Part on the left is encoder path and the right is the decoder path. The encoder path gives the object class after series of convolution and max pooling layers. Then the decoder path does the upsampling and results in segmentation map. In the decoder path layers of up convolution takes place and data saved from each layer of convolution steps are copied to the respective deconvolution steps and concatenated.

One of the goals of this U-Net architecture is to be able to generate good results with fewer training images. So, data augmentation was applied heavily on the training sets, so the model learns about different variances. The experiments have shown very good results with few labelled data and lower training time.

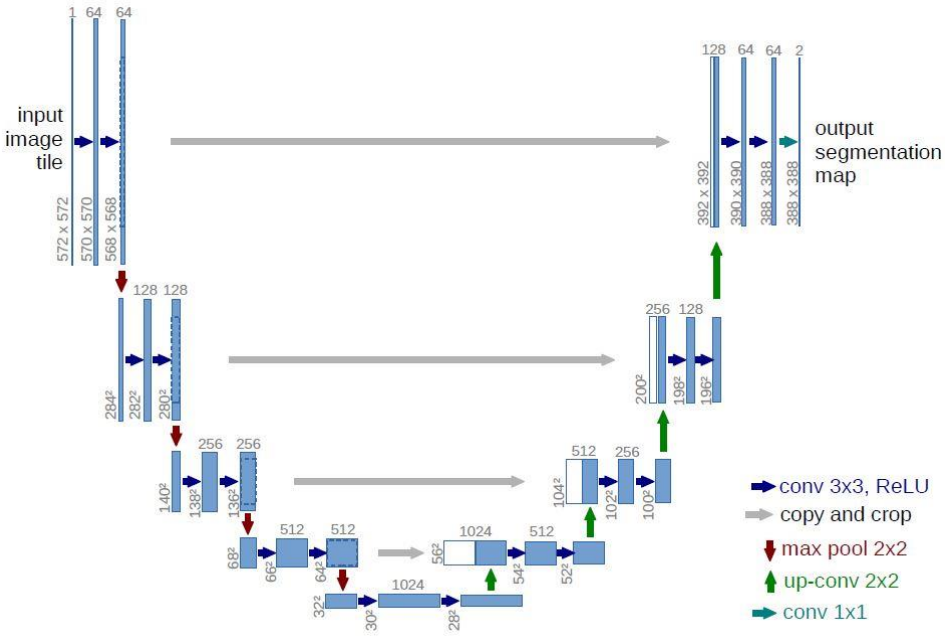

Fig. 4: U-Net model architecture [13]

Another approach is to use dilated convolutional also known as atrous convolution to counter the problem of low-resolution results. DeepLab model proposed by Chen et al [14] in one such model. Atrous convolution helps in reducing the degree of downsampling caused by max-pooling and striding. The kernel in atrous convolution is inflated by introducing holes into it and as a result, after upsampling there are less data lost in the image. For handling of images at multiple scales, the rate can be changed for atrous convolution according to the image scales. After the model training is done, to handle the reduced localization accuracy, some post processing is also proposed. They proposed the use of fully connected conditional random field (CRF). CRF is applied after getting the prediction from the model and getting the upsampled image. Fully connected CRF refines the output of the model. After CRF is applied, the final output is generated.

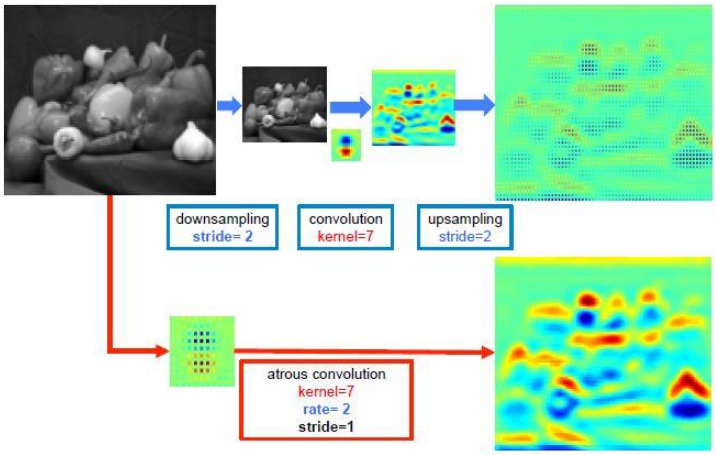

Fig. 5: Atrous convolution in 2-D [14].

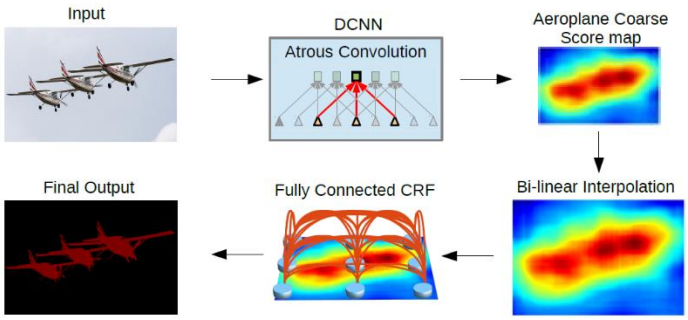

Fig. 6: Illustration of DeepLab model [14].

2.2. Instance Segmentation

Getting inspiration from the success in object detection using Regional CNN (R-CNN) [15] [16] and semantic segmentation algorithm such as FCN, few algorithms of instance segmentation were proposed combining these two approaches. Mask R-CNN [17] proposed by He et al. is one such algorithm. They modified the Faster R-CNN [16] to perform instance segmentation. Faster R-CNN gives 2 outputs, class label and bounding box. They added new branch to predict the object mask in parallel with the bounding box prediction. When the proposed method was compared with other algorithms like FCIS [18]. Mask R-CNN showed better performance.

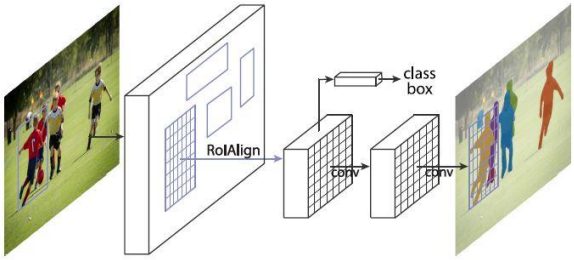

Fig. 7: Mask R-CNN framework [17].

Another proposed architecture built on the Faster R-CNN is the MaskLab [19]. The modifies faster R-CNN here gives three outputs: bounding box, semantic segmentation logits and direction prediction logits. Region of Interest is first generated and then both foreground and background segmentation are carried out in the regions. This is done by combining both semantic segmentation and direction prediction. Semantic segmentation separates objects from background and then the direction prediction does the separation of instances from classes by estimating the direction of each pixel towards its center. Experiments showed MaskLab performing better than other algorithms such as FCIS and Mask R-CNN.

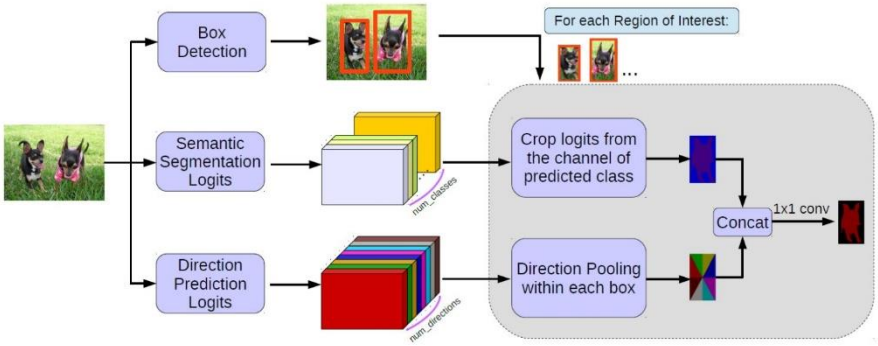

Fig. 8: MaskLab Concept [19].

2.3. Panoptic Segmentation

Panoptic segmentation [20] has appeared in research very recently. Since then, few techniques and developments have been proposed and growing its popularity. Hou et al. [21] proposed a method for real time panoptic segmentation from dense detection. As previously mentioned, panoptic segmentation is combination of both semantic and instance segmentation. Instance segmentation part of panoptic segmentation causes the bottleneck in panoptic segmentation. They tried to overcome the bottleneck and implement a real time solution.

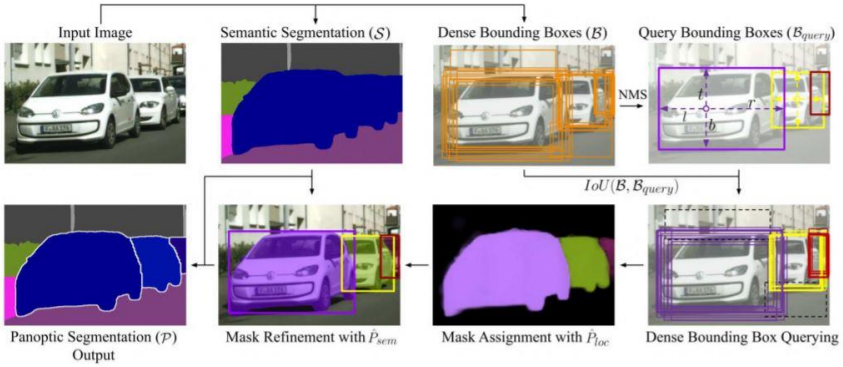

Fig. 9: Real-time panoptic segmentation workflow [21].

Fig.9. shows the workflow of the proposed method. First the semantic segmentation and dense bounding boxes are predicted using single stage convolutional networks. Then using standard Non Maximum suppression (NMS) non duplicate boxes are identified (query boxes). In self attention manner IOU if each query box and dense boxes are calculated. Pixels with higher IOU values are considered to be the instance foreground of the query box. And finally using the semantic segmentation results the results are refined more. For the implementation of the algorithm ResNet50 was used as the backbone and feature paranoid network (FPN) was used to generate the multiscale feature maps. Then the feature maps were fed into CNN model to predict the dense bounding boxes.

3. Image Segmentation Applications

Applying the above discussed techniques and other techniques many applications related to image segmentation has been proposed. Particularly in medical image processing, autonomous vehicles, and others.

In medical images most of the research are based on CT Image processing. Tong et al. [22] used this U-net architecture idea and combined it with a shape representation model (SRM) to have a multi-organ segmentation system. Their target was to segment multiple target organs from a single 3D CT volume of a human head and neck. They trained the system in two parts. First training the model to learn the shape characteristics of the target organs and then training the fully convolutional neural network for segmentation. During the training of the FCNN the pre-trained SRM model is used as a regularizer to generate more accurate segmented regions. Fig. 10 shows the outcome of the trained model, different organs from the 3D CT data are segmented into different classes.

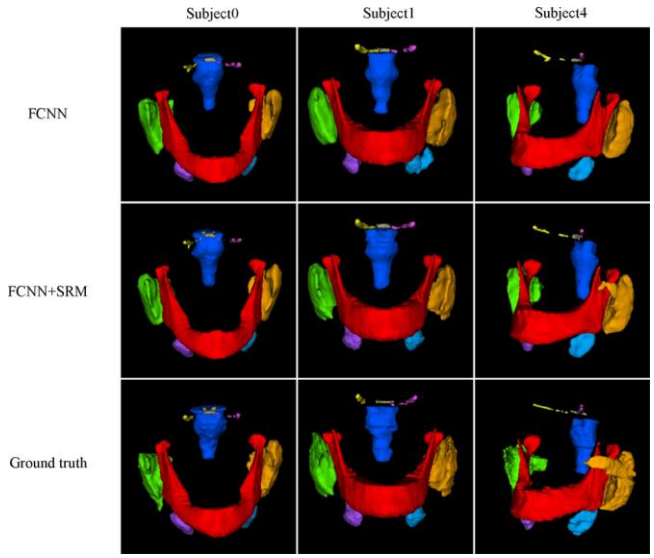

Fig. 10: 3D visualization of organs segmented using SRM and FCNN combined [22].

The idea of FCNs also been applied to other segmentation problems in medical image processing such as organs and blood vessels segmentation [23], brain tumor segmentation [24] etc. Idea of Mask R-CNN was applied to application such as segmentation and measurement of wounds [25] and Skin lesion segmentation [26] from images.

Ghosh et al [27] proposed SegFast-V2 algorithm targeting semantic segmentation for autonomous driving. The algorithm was developed by combination of ideas from SqueezeNet [28], U-Net architecture and depth-wise separable convolution [29] to reduce the cost of computation. Other applications such as road segmentation [30], Human gesture recognition for vehicles [31] etc are also proposed using deep learning.

4. Limitations and Challenges

Although several techniques of deep learning for image segmentation are being proposed, most of them are computationally very expensive for real time use in application. When it comes to 3D volume images, the cost

increases more. Some research are also done for real-time predictions but still it requires more works. Also, while many research are being done focusing on the accuracy, not much have been done to reduce the memory usage. Most modern algorithms require very expensive and powerful machines to run.

Another challenge of using deep learning in applications is managing quality dataset. Generating data for training with ground truth value for certain application are typically difficult job. Most of the algorithms developed are tested with existing datasets but when they are needed to be trained for real world scenario, then data management is a large issue to be considered. Researchers are getting more interested in research for semi-supervised, unsupervised, transfer learning and self-supervised learning techniques for image segmentation.

Another field where more research can be done is combining traditional image processing-based segmentation techniques such as thresholding, graph cuts etc. with deep learning-based techniques. This could lead to advancement of the performance of segmentation.

4. Conclusion

To summarize, few deep learning-based image segmentation techniques were discussed. It is seen that deep learning could solve many challenges of image segmentation and generate accurate predictions. But still they have some limitations and challenges to overcome. With more research and implementation, deep learning could be the ultimate choice for all image processing-based applications.

References

[1] R. Vitale, J. M. Prats-Montalbán, F. López-García, J. Blasco and A. Ferrer, "Segmentation techniques in image analysis: A comparative study," *Journal of Chemometrics*, vol. 30, no. 12, pp. 749-759, 2016.

[2] D. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*. Prentice Hall, 2002.

[3] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[4] R. Nock and F. Nielsen, "Statistical region merging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 11, pp. 1452–1458, 2004.

[5] L. Najman and M. Schmitt, "Watershed of a continuous function," *Signal Processing*, vol. 38, no. 1, pp. 99–112, 1994.

[6] N. Dhanachandra, K. Manglem, and Y. J. Chanu, "Image segmentation using K-means clustering algorithm and subtractive clustering algorithm," *Procedia Computer Science*, vol. 54, pp. 764–771, 2015.

[7] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222–1239, 2001.

[8] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.

[9] A. Kirillov, K. He, R. Girshick, C. Rother, and P. Doll´ar, "Panoptic segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 9404–9413.

[10] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.

[11] G. Wang, W. Li, S. Ourselin, and T. Vercauteren, "Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks," in *International MICCAI Brainlesion Workshop*. Springer, 2017, pp. 178–190.

[12] Y. Yuan, M. Chao, and Y.-C. Lo, "Automatic skin lesion segmentation using deep fully convolutional networks with Jaccard distance," *IEEE Transactions on Medical Imaging*, vol. 36, no. 9, pp. 1876–1886, 2017.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer*, 2015, pp. 234–241.

[14] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2017.

[15] R. Girshick, "Fast R-CNN," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.

[16] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015.

[17] K. He, G. Gkioxari, P. Doll´ar, and R. Girshick, "Mask R-CNN," in *IEEE International Conference on Computer Vision*, 2017, pp. 2961–2969.

[18] Y. Li, H. Qi, J. Dai, X. Ji, and Y. Wei, "Fully convolutional instance aware semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4438–4446.

[19] L.-C. Chen, A. Hermans, G. Papandreou, F. Schroff, P. Wang, and H. Adam, "Masklab: Instance segmentation by refining object detection with semantic and direction features," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4013–4022.

[20] A. Kirillov, K. He, R. Girshick, C. Rother, and P. Doll´ar, "Panoptic segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 9404–9413.

[21] R. Hou, J. Li, A. Bhargava, A. Raventos, V. Guizilini, C. Fang, J. Lynch and A. Gaidon, "Real-time panoptic segmentation from dense detections," In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 8523-8532.

[22] N. Tong, S. Gou, S. Yang, D. Ruan, and K. Sheng, "Fully Automatic Multi-Organ Segmentation for Head and Neck Cancer Radiotherapy Using Shape Representation Model Constrained Fully Convolutional Neural Networks," *Medical Physics*, vol. 45, no. 10, pp. 4558-4567, 2018.

[23] L. Bi, D. Feng, and J. Kim, "Dual-Path Adversarial Learning for Fully Convolutional Network (FCN)-Based Medical Image Segmentation," *The Visual Computer*, vol. 34, no. 6, pp. 1043–1052, 2018.

[24] G. Wang, W. Li, S. Ourselin, and T. Vercauteren, "Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks," in *International MICCAI Brainlesion Workshop*. Springer, 2017, pp. 178–190.

[25] M. Privalov, N. Beisemann, J.E. Barbari, E. Mandelka, M. Müller, H. Syrek, P.A. Grützner and S.Y. Vetter, "Software-Based Method for Automated Segmentation and Measurement of Wounds on Photographs Using Mask R-CNN: a Validation Study," *Journal of Digital Imaging*, vol. 34, no. 4, pp. 788-797, 2021.

[26] F. Bagheri, M.J. Tarokh, and M. Ziaratban, "Skin lesion segmentation from dermoscopic images by using Mask R-CNN, Retina-Deeplab, and graph-based methods," *Biomedical Signal Processing and Control*, vol. 67, pp. 102533, 2021.

[27] S. Ghosh, A. Pal, S. Jaiswal, K.C. Santosh, N. Das, and M. Nasipuri, "Segfast-v2: Semantic image segmentation with less parameters in deep learning for autonomous driving," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 11, pp. 3145-3154, 2019

[28] F.N. Iandola,S. Han, M.W. Moskewicz, K. Ashraf, W.J. Dally and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size," *arXiv preprint arXiv:1602.07360*, 2016

[29] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251-1258.

[30] M. Junaid, M. Ghafoor, A. Hassan, S. Khalid, S.A. Tariq, G. Ahmed, and T. Zia, "Multi-Feature View-Based Shallow Convolutional Neural Network for Road Segmentation," *IEEE Access*, vol. 8, pp. 36612-36623, 2020.

[31] K. Geng and G. Yin, "Using Deep Learning in Infrared Images to Enable Human Gesture Recognition for Autonomous Vehicles," *IEEE Access*, vol. 8, pp. 88227-88240, 2020.