engineering proceedings

Proceeding Paper

Semantic Segmentation for Various Applications: Research Contribution and Comprehensive Review †

Madiha Mazhar *, Saba Fakhar and Yawar Rehman

Department of Electronic Engineering, NED University of Engineering and Technology, Karachi 75270, Pakistan; sabahmed@neduet.edu.pk (S.F.); yawar@neduet.edu.pk (Y.R.)
* Correspondence: madiha@neduet.edu.pk

† Presented at the 2nd International Conference on Emerging Trends in Electronic and Telecommunication Engineering, Karachi, Pakistan, 15–16 March 2023.

Abstract: Semantic image segmentation is used to analyse visual content and carry out real-time decision-making. This narrative literature analysis evaluates the multiple innovations and advancements in the semantic algorithm-based architecture by presenting an overview of the algorithms used in medical image analysis, lane detection, and face recognition. Numerous groundbreaking works are examined from a variety of angles (e.g., network structures, algorithms, and the problems addressed). A review of the recent development in semantic segmentation networks, such as U-Net, ResNet, SegNet, LCSegnet, FLSNet, and GNet, is presented with evaluation metrics across a range of applications to facilitate new research in this field.

Keywords: semantic segmentation; encoder decoder; applications; medical imaging; face recognition; lane detection

check for updates

Citation: Mazhar, M.; Fakhar, S.; Rehman, Y. Semantic Segmentation for Various Applications: Research Contribution and Comprehensive Review. *Eng. Proc.* **2023**, *32*, 21. https://doi.org/10.3390/engproc2023032021

Academic Editors: Muhammad Faizan Shirazi, Saba Javed, Sundus Ali and Muhammad Imran Aslam

Published: 5 May 2023

1. Introduction

Convolutional neural networks (CNNs) have achieved amazing success in semantic segmentation in recent years. Semantic segmentation is the labelling of pixels of an image into different labels, such as cars, pedestrians, and trees. Nowadays, most techniques for generating pixel-by-pixel segmentation prediction use an encoder–decoder architecture. The decoder recovers feature map resolution, while the encoder is used to extract the feature maps.

Due to the significant improvement in diagnostic efficiency and accuracy, medical image segmentation frequently plays a crucial part in computer-aided diagnosis and smart medicine. Liver and liver tumor segmentation [1,2], as well as brain and brain tumor segmentation [3,4], are common medical image segmentation tasks. Moreover, the segmentation of the optic disc [5,6] and cell segmentation [7], lung segmentation, pulmonary nodules [8,9], and heart image segmentation [10,11] are commonly used techniques. The early methods of segmenting medical images frequently rely on edge detection, machine learning, template matching methods, statistical shape models, active contours, and statistical shape models. Convolutional Neural Networks—CNNs—(Deep learning models) have recently proven to be useful for a variety of image segmentation tasks.

2. Applications of Semantic Segmentation

Semantic segmentations have found a variety of applications in many areas, such as medical diagnostics and scanning, face recognition, scene understanding, autonomous driving, handwriting recognition, etc. This literature survey covers three broad applications of semantic segmentation to facilitate researchers in applying the network architectures of one application to the other application.

2.1. Semantic Segmentation in Medical Imaging

One of the most well-known CNN designs for semantic segmentation is the U-Net architecture, which has achieved outstanding results in a wide range of medical image segmentation applications. A novel Dens-Res-Inception Net (DRINet) is proposed in [12] to address this challenging problem by learning distinctive features and has found applications in brain CT, brain tumor, and abdominal CT images. [13] proposed a brand-new high-resolution multi-scale encoder–decoder network (HMEDN), in which dense multi-scale connections are provided to allow the encoder–decoder structure to precisely use all of the available semantic data. Skip connections are added, as well as extra extensively trained high-resolution pathways (made up of densely connected dilated convolutions) to gather high-resolution semantic data for precise border localization, which were successfully validated on pelvic CT images and a multi-modal brain tumor dataset. In [14], an assessment of the prediction uncertainty in FCNs for segmentation was investigated by systematically comparing cross-entropy loss with Dice loss in terms of segmentation quality and uncertainty estimation and model ensemble for confidence calibration of the FCNs trained with batch normalization and Dice loss and tested on applications that included the prostate, heart, and brain. For an accurate diagnosis of interstitial lung diseases (ILDs), [15] proposed an FCN-based semantic segmentation of ILD pattern recognition to avoid sliding window model limitations. Training complexities are addressed in [16] by decomposing a single task into three sub-tasks, such as pixel-wise segmentation, prediction, and classification of an image and a novel sync-regularization was proposed to penalize the nonconformity between the outputs.

To overcome the drawbacks of feature fusion methods, INet was proposed in [17] that used two overlapping max-pooling to extract the sharp features and contributed positively to applications such as biomedical MRI, X-Ray, CT, and endoscopic imaging. The automatic identification of BAC in mammograms is not yet possible with any currently used methods. In [18], the UNet model with dense connectivity is proposed that aids in reusing computation and enhances gradient flow, resulting in greater accuracy and simpler model training. A novel architecture [19] Multi-Scale Residual Fusion Network (MSRF-Net) uses a Dual-Scale Dense Fusion (DSDF) block; the proposed MSRF-Net is able to communicate multi-scale features with different receptive fields. Table 1 illustrates network architectures, methods, problem addressed, performance metrics, and the regions of interest/application.

Table 1. Network Architectures implementation in Medical Imaging.

S. No.	Method/CNN	Backbone/Network Architecture	Problem Addressed	Performnace Metric	Applications
1	UNET	DRI-Net	Distinctive features learned	Dice Coefficient, Sensitivity	Medical Imaging
2	Encoder–Decoder	HMEDN	Exploits comprehensive semantic information	Dice Ratio, Memory consumption	region of research is pelvic CT and brain Tumor
3	FCN		Predictive uncertainty estimation addressed	Dice Loss, Cross Entropy loss, uncertainty estimation	Medical Imaging
4	FCN	FCN with Dilated filters	Sliding window model	Accuracy	Medical Imaging (Lungs)
5	FCN	Source image deomposition	Training complexities addressed	Loss Function, Dice, IoU	Medical Imaging

Table 1. *Cont.*

S. No.	Method/CNN	Backbone/Network Architecture	Problem Addressed	Performnace Metric	Applications
6	Encoder–Decoder-Unet	INet and Dense INet compared with Dense Unet and ResDenseUNet	Feature fusion and feature concatination addressed.	Dice ratio, TPR, Specificity, TNR, HD95	Biomedical (MRI, X-Ray, CT, Endoscopic image, Ultrasound)
7		UNet	Re-use of computation	Accuracy, Sensitivity, Specifici	Arterial Calcification in Mammograms
8	CNN	MSRF-Net	efficiently segments objects	Dice Coefficient (DSC)	Skin lesion

2.2. Semantic Segmentation in Face Recognition

In the realm of machine vision, facial analysis has recently emerged as an active study subject. Neural networks are trained to accurately predict age classification, gender, and other things by using the extracted characteristics.

A particular type of semantic segmentation is face labelling. The goal of face labelling is to give each pixel in a picture a specific semantic category, such as an eye, brow, nose, mouth, etc. End-to-end face labelling is proposed in [20] with pyramid FCN while maintaining a small network size. In order to detect each face in the frame regardless of alignment, [21] created a binary face classifier and presented a technique for creating precise face segmentation masks from input images of any size. A method for enhancing the prediction of facial attributes is discussed in [22]. In this study, we suggest using semantic segmentation to enhance the prediction of facial attributes. FaceNet and VGG-face were utilized [23] as the foundation for face semantic segmentation, which solves the issue of exact and pixel-level localization of face regions. A technique for precisely obtaining facial landmarks is presented in [24] to enhance pixel classification performance by altering the imbalance of the number of pixels in accordance with the facial landmark. Table 2 illustrates network architectures, methods, problem addressed, performance metrics, and region of interest/application.

Table 2. Network Architectures implementation in Face Recognition.

S. No.	Method/CNN	Backbone/Network Architecture	Problem addressed	Performance Metric
1	(Pyramid FCN)	End-to-end face labelling	End-to-end manner	Fscore
2	FCN	Binary face classifier	mask generated from arbitrary size input image	Pixel accuracy
3	FCN		improvement in facial attribute prediction	Classification error, average precision
4	FCN	Face Semantic segmentation	added generalization and features local information	P(Pearson correlation), MAE, RMS error
5	FCN	(Facial landmark Net)	improved imbalance pixels	Pixel accuracy, IoU
6	CNN	UNet	Supplemental bypass in the conventional optical character recognition (OCR) process	Recall, precision, F-measure
7	FCN	LCSegNet(Label coding Segmentation net)	recognition of large-scale Chinese characters	character recognition accuracy
8	Deep Learning	UNet	Improve quality of output, digitization	Jaccard index, TN, TP

2.3. Semantic Segmentation in Lane Detection

To increase the road safety of cars and reduce road accidents, Advanced Driver Assistant Systems (ADAS) play a vital role in designing intelligent driving systems (IDSs).

In lane segmentation algorithms, each pixel of an image is labelled into lane and non-lane classes. Some commonly used lane detection algorithms have been reviewed. SUPER, a novel lane detection system, was introduced by [25], which consists of a semantic segmentation network and physics-enhanced multilane parameters with enhanced learning-based and physics-based techniques. To overcome the drawback of convolutional neural networks (CNNs), which relies only on information transfer between layers without using the spatial information within the layers, an attention-based segmentation network SCNN (Spatial CNN) was proposed by [26]. Further improvements in spatial information in CNN layers are introduced by [27]. Airborne imagery is proposed by [28] in Aerial LaneNet, which is based on a Lane-making segmentation network to apprehend bigger areas in a short span of time. The problem of essential information in features, which is overlooked by most of the lane segmentation problems, is resolved by [29] in which an aggregator network based on multiscale features is proposed.

Since pixel-level segmentation is a tedious task and poses a burden on computation, an alternate scheme wherein grid-level semantic segmentation GNET [30] is proposed. Another grid-based segmentation is proposed by [31] for free space and lane-based detection.

Helping blind people in walking and crossing roads is the responsibility of society and it is also society's responsibility to efficiently design devices that help them in crossing roads. For this, a low depth semantic segmentation network is proposed [32] for blind roads and crosswalks. Accurate features are extracted by using the atrous pyramid module.

The dual power of handcrafted features and convolutional neural networks is utilized in [33]. The localization ability is achieved by using hand-crafted features, and the integration of both also predicts a vanishing line. Semantic segmentation utilizing encoder–decoder for detecting multiple lines is proposed by [34]. In this work, the pixel accuracy of weak class objects is improved by depicting a ground truth dataset. Table 3 illustrates network architectures, methods, problem addressed, performance metrics, and region of interest/application.

Table 3. Network Architectures implementation in Lane Detection.

S. No.	Method/CNN	Backbone/Network architecture	Problem addressed	Performance Metric
1	CNN	SUPER	Optimization of Lane parameters	TPR, FPR, Fmax
2	Encoder–Decoder (Spatial-SCNN)	ABSSNet	Spatial information inside the layers	MIoU
3	FCN (Encoder–Decoder)	Aerial LaneNet	Captures Large area in short span of time	Loss function, Dice Coefficient, Forward time
4	Encoder–Decoder	MFIA Lane	Simultaneous handling of multiple perceptual task	Accuracy, F1, PA, IoU
5	Encoder–Decoder	G-Net	Releases the detection burden	Accuracy, FP, FN, FPS
6	CNN		Network can learn the spatial relationship for point of interest	MIoU
7	Encoder–Decoder	Light weight segmentation network	Reduce the number of parameters	Computation time per image
8	CNN		Accuracy of location	Correct rate
9	CNN	Multilane encoder–decoder	accuracy of weak class objects	Speed and accuracy
10	CNN	Spatial-SCNN	Strong Spatial relationship	Accuracy

3. Conclusions

The purpose of the proposed study aims to establish the state of the art as a baseline for researchers to compare their knowledge of various machine learning and deep learning techniques for semantic segmentation. In total, 34 research articles were chosen for this investigation and were gathered from different research databases. It is concluded that convolutional neural networks and encoder–decoder architectures have been used as a backbone for implementing semantic segmentation. However, the detection accuracy of the network depends on the depth of the neural network chosen.

Author Contributions: Conceptualization, M.M., S.F. and Y.R.; literature review, M.M. and S.F.; writing, original draft, M.M. and S.F.; writing, review, Y.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Li, W.; Jia, F.; Hu, Q. Automatic Segmentation of Liver Tumor in CT Images with Deep Convolutional Neural Networks. *J. Comput. Commun.* **2015**, *3*, 146–151. [CrossRef]
2. Vivanti, R.; Ephrat, A.; Joskowicz, L.; Karaaslan, O.A.; Lev-Cohain, N.; Sosna, J. Automatic Liver Tumor Segmentation in Follow-up CT Studies Using Convolutional Neural Networks. Available online: https://www.researchgate.net/profile/Leo-Joskowicz/publication/314063607_Automatic_Liver_Tumor_Segmentation_in_Follow-Up_CT_Scans_Preliminary_Method_and_Results/links/58f84cfd0f7e9bfcf93c1292/Automatic-Liver-Tumor-Segmentation-in-Follow-Up-CT-Scans-Preliminary-Method-and-Results.pdf (accessed on 25 January 2023).
3. Menze, B.H.; Jakab, A.; Bauer, S.; Kalpathy-Cramer, J.; Farahani, K.; Kirby, J.; Burren, Y.; Porz, N.; Slotboom, J.; Wiest, R.; et al. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Trans. Med. Imaging* **2014**, *34*, 1993–2024. [CrossRef]
4. Cherukuri, V.; Ssenyonga, P.; Warf, B.C.; Kulkarni, A.V.; Monga, V.; Schiff, S.J. Learning Based Segmentation of CT Brain Images: Application to Postoperative Hydrocephalic Scans. *IEEE Trans. Biomed. Eng.* **2017**, *65*, 1871–1884. [CrossRef] [PubMed]
5. Cheng, J.; Liu, J.; Xu, Y.; Yin, F.; Wong, D.W.K.; Tan, N.-M.; Tao, D.; Cheng, C.-Y.; Aung, T.; Wong, T.Y. Superpixel Classification Based Optic Disc and Optic Cup Segmentation for Glaucoma Screening. *IEEE Trans. Med. Imaging* **2013**, *32*, 1019–1032. [CrossRef] [PubMed]
6. Fu, H.; Cheng, J.; Xu, Y.; Wong, D.W.K.; Liu, J.; Cao, X. Joint Optic Disc and Cup Segmentation Based on Multi-Label Deep Network and Polar Transformation. *IEEE Trans. Med. Imaging* **2018**, *37*, 1597–1605. [CrossRef] [PubMed]
7. Song, T.-H.; Sanchez, V.; Eidaly, H.; Rajpoot, N.M. Dual-Channel Active Contour Model for Megakaryocytic Cell Segmentation in Bone Marrow Trephine Histology Images. *IEEE Trans. Biomed. Eng.* **2017**, *64*, 2913–2923. [CrossRef]
8. Wang, S.; Zhou, M.; Liu, Z.; Liu, Z.; Gu, D.; Zang, Y.; Dong, D.; Gevaert, O.; Tian, J. Central focused convolutional neural networks: Developing a data-driven model for lung nodule segmentation. *Med. Image Anal.* **2017**, *40*, 172–183. [CrossRef]
9. Onishi, Y.; Teramoto, A.; Tsujimoto, M.; Tsukamoto, T.; Saito, K.; Toyama, H.; Imaizumi, K.; Fujita, H. Multiplanar analysis for pulmonary nodule classification in CT images using deep convolutional neural network and generative adversarial networks. *Int. J. Comput. Assist. Radiol. Surg.* **2019**, *15*, 173–178. [CrossRef]
10. Chen, C.; Qin, C.; Qiu, H.; Tarroni, G.; Duan, J.; Bai, W.; Rueckert, D. Deep Learning for Cardiac Image Segmentation: A Review. *Front. Cardiovasc. Med.* **2020**, *7*, 25. [CrossRef]
11. Wu, F.; Zhuang, X. CF Distance: A New Domain Discrepancy Metric and Application to Explicit Domain Adaptation for Cross-Modality Cardiac Image Segmentation. *IEEE Trans. Med. Imaging* **2020**, *39*, 4274–4285. [CrossRef]
12. Chen, L.; Bentley, P.; Mori, K.; Misawa, K.; Fujiwara, M.; Rueckert, D. DRINet for Medical Image Segmentation. *IEEE Trans. Med. Imaging* **2018**, *37*, 2453–2462. [CrossRef] [PubMed]
13. Zhou, S.; Nie, D.; Adeli, E.; Yin, J.; Lian, J.; Shen, D. High-Resolution Encoder–Decoder Networks for Low-Contrast Medical Image Segmentation. *IEEE Trans. Image Process.* **2019**, *29*, 461–475. [CrossRef] [PubMed]
14. Mehrtash, A.; Wells, W.M.; Tempany, C.M.; Abolmaesumi, P.; Kapur, T. Confidence Calibration and Predictive Uncertainty Estimation for Deep Medical Image Segmentation. *IEEE Trans. Med. Imaging* **2020**, *39*, 3868–3878. [CrossRef] [PubMed]
15. Anthimopoulos, M.M.; Christodoulidis, S.; Ebner, L.; Geiser, T.; Christe, A.; Mougiakakou, S.G. Semantic Segmentation of Pathological Lung Tissue With Dilated Fully Convolutional Networks. *IEEE J. Biomed. Health Inform.* **2018**, *23*, 714–722. [CrossRef]

16. Ren, X.; Ahmad, S.; Zhang, L.; Xiang, L.; Nie, D.; Yang, F.; Wang, Q.; Shen, D. Task Decomposition and Synchronization for Semantic Biomedical Image Segmentation. *IEEE Trans. Image Process.* **2020**, *29*, 7497–7510. [CrossRef]
17. Weng, W.; Zhu, X. INet: Convolutional Networks for Biomedical Image Segmentation. *IEEE Access* **2021**, *9*, 16591–16603. [CrossRef]
18. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention 2015*; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241. [CrossRef]
19. Srivastava, A.; Jha, D.; Chanda, S.; Pal, U.; Johansen, H.; Johansen, D.; Riegler, M.; Ali, S.; Halvorsen, P. MSRF-Net: A Multi-Scale Residual Fusion Network for Biomedical Image Segmentation. *IEEE J. Biomed. Health Inform.* **2021**, *26*, 2252–2263. [CrossRef] [PubMed]
20. Wen, S.; Dong, M.; Yang, Y.; Zhou, P.; Huang, T.; Chen, Y. End-to-End Detection-Segmentation System for Face Labeling. *IEEE Trans. Emerg. Top. Comput. Intell.* **2019**, *5*, 457–467. [CrossRef]
21. Meenpal, T.; Balakrishnan, A.; Verma, A. Facial Mask Detection using Semantic Segmentation. In Proceedings of the 2019 4th International Conference on Computing, Communications and Security, ICCCS, Rome, Italy, 10–12 October 2019. [CrossRef]
22. Kalayeh, M.M.; Shah, M. Improving Facial Attribute Prediction using Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
23. Yousaf, N.; Hussein, S.; Sultani, W. Estimation of BMI from facial images using semantic segmentation based region-aware pooling. *Comput. Biol. Med.* **2021**, *133*, 104392. [CrossRef]
24. Kim, H.; Kim, H.; Rew, J.; Hwang, E. FLSNet: Robust Facial Landmark Semantic Segmentation. *IEEE Access* **2020**, *8*, 116163–116175. [CrossRef]
25. Lu, P.; Cui, C.; Xu, S.; Peng, H.; Wang, F. SUPER: A Novel Lane Detection System. *IEEE Trans. Intell. Veh.* **2021**, *6*, 583–593. [CrossRef]
26. Pan, X.; Shi, J.; Luo, P.; Wang, X.; Tang, X. Spatial as Deep: Spatial CNN for Traffic Scene Understanding. *Proc. Conf. AAAI Artif. Intell.* **2018**, *32*, 7276–7283. [CrossRef]
27. Li, X.; Zhao, Z.; Wang, Q. ABSSNet: Attention-Based Spatial Segmentation Network for Traffic Scene Understanding. *IEEE Trans. Cybern.* **2021**, *52*, 9352–9362. [CrossRef] [PubMed]
28. Azimi, S.M.; Fischer, P.; Korner, M.; Reinartz, P. Aerial LaneNet: Lane-Marking Semantic Segmentation in Aerial Imagery Using Wavelet-Enhanced Cost-Sensitive Symmetric Fully Convolutional Neural Networks. *IEEE Trans. Geosci. Remote Sens.* **2018**, *57*, 2920–2938. [CrossRef]
29. Qiu, Z.; Zhao, J.; Sun, S. MFIALane: Multiscale Feature Information Aggregator Network for Lane Detection. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 24263–24275. [CrossRef]
30. Wang, H.; Liu, B. G-NET: Accurate Lane Detection Model for Autonomous Vehicle. *IEEE Syst. J.* **2022**. *early access*. [CrossRef]
31. Shao, M.-E.; Haq, M.A.; Gao, D.-Q.; Chondro, P.; Ruan, S.-J. Semantic Segmentation for Free Space and Lane Based on Grid-Based Interest Point Detection. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 8498–8512. [CrossRef]
32. Cao, Z.; Xu, X.; Hu, B.; Zhou, M. Rapid Detection of Blind Roads and Crosswalks by Using a Lightweight Semantic Segmentation Network. *IEEE Trans. Intell. Transp. Syst.* **2020**, *22*, 6188–6197. [CrossRef]
33. Wang, Q.; Han, T.; Qin, Z.; Gao, J.; Li, X. Multitask Attention Network for Lane Detection and Fitting. *IEEE Trans. Neural Networks Learn. Syst.* **2020**, *33*, 1066–1078. [CrossRef]
34. Chougule, S.; Ismail, A.; Soni, A.; Kozonek, N.; Narayan, V.; Schulze, M. An efficient encoder-decoder CNN architecture for reliable multilane detection in real time. In Proceedings of the 2018 IEEE Intelligent Vehicles Symposium (IV), Changshu, China, 26–30 June 2018; pp. 1444–1451. [CrossRef]