Image Segmentation using Optimization Algorithm: A Survey

Suja Paulose[1*], Dr. D. Veera Vanitha[2]

[1]Research Scholar, Department of ECE, School of Engineering, Avinashilingam Institute of Home Science and Higher Education for Women, Coimbatore, Tamil Nadu

[2]Associate Professor, Department of ECE, School of Engineering, Avinashilingam Institute of Home Science and Higher Education for Women, Coimbatore, Tamil Nadu

*Corresponding author's e-mail: 21phelp001@avinuty.ac.in

doi: https://doi.org/10.21467/proceedings.160.41

ABSTRACT

Image segmentation has proven to be an important step in the processing of images, computer vision algorithms, etc. It splits an image into different regions. This survey reviews major contributions in the healthcare l field using deep learning, including the common problems published over the last few years, and also explains the basics of deep learning concepts applicable to medical image segmentation. To solve current problems and improve the development of medical image segmentation problems, the Efficient Net Atrous convolutional encoder & and decoder can be used for segmentation in future research. Efficient Nets have much better accuracy & and efficiency than conv-Nets. The advantage of Efficient-Net is that it can balance the model's depth, width, and image resolution through composite coefficients.

Keywords: Segmentation, Atrous Convolutional Encoder& Decoder, Efficient-Net

1 Introduction

Image segmentation is a sub-field of computer vision and digital imaging that creates similar groups or sections of images under their catalogues. In fact, the task is to allocate a mark to each picture element in the image so that the picture elements with the same label have analogous properties and parcels of similar colour and texture. Segmentation plays a part in a large range of operations, including medical image analysis (e.g.: tumour boundary birth and dimension of towel volumes), independent vehicles e.g.: passable face and rambler discovery), videotape surveillance, and stoked reality etc. Traditional styles took an original view of the point in an image and riveted on original differences and slants in pixels.

The two levels of granularity in image segmentation are Semantic segmentation and instance segmentation. Semantic segmentation classifies image pixels into one or more classes that are semantically interpretable, rather, than real-world objects. It can be considered as image captioning. The difference between instance segmentation and semantic segmentation is that semantic segmentation does not classify every pixel. Instance segmentation provides more detailed localized information about the scene.

The main aim of the survey is to identify the challenges in medical image segmentation, identify an appropriate framework and compare the performances of different approaches.

Deep learning models have yielded a new generation of image partitioning models with remarkable performance improvements and achieves a highest accuracy.

2 Overview of Deep-Learning Techniques

Neural networks that perform segmentations typically use an encoder-decoder structure where the encoder is followed by a tailback and a decoder or up sampling layers directly from the tailback. Encoder-Decoder structure for semantic segmentation came popular with the onset of works like Seg-Net [1] in 2015.

Proceedings DOI: 10.21467/proceedings.160; Series: AIJR Proceedings; ISSN: 2582-3922; ISBN: 978-81-965621-9-9

2.1 Seg-Net

Seg-Net proposes the use of a combination of convolutional and down-slice blocks to squeeze information into a bottleneck and create a representation of the input. It is a Fully Convolutional Network. The decoder also reconstructs the input information to form a member chart pressing regions on the input and grouping them under the closest. Finally, the decoder has a sigmoid activation at the end that compresses the output in the range (0,1).

In the decoding process for up sampling the layers, Seg-Net uses max pooling indices at the corresponding encoder layers are recalled. This makes the training process easier since the network need not learn the up sampling weights again. It is computationally feasible. The performance of Seg-Net is high, it is more efficient memory wise and competitive inference time as compared to other architectures.

The disadvantage of Seg-Net is that it tends to lose neighbouring information when un pooling from low-resolution feature maps. For reducing the loss of information in the down-sampling layers of the encoder and decoder, U-Net architecture can be used.

2.2 U-Net

The main function is to complement the usual contracting network with successive layers, where pooling operations are replaced by up-sampling operators. It is proved that such a network can be summed from every few images and found to be the best method on the ISBI challenge to segment neurological structures in electron microscopic bundles [2].

U-Net architecture directly monitors high-level semantic features and uses skip connections in the same position due to its symmetric structure instead of lossy-back propagation. The connections that go from the encoder straight to the decoder without passing through the bottleneck are skip connections. i.e., feature maps at various levels of encoded representations are captured and concatenated to feature maps in the decoder. It also helps in integrating features of different scales to make predictions at multiple scales [3]. Due of this feature U-Net is the choice in medical imaging. For training deeper networks and extracting deep semantic information, RSU-Net combines RRCNN with U-Net [4]. It works well with vascular, lung, and skin datasets. Due to different tasks, the depth requirements of U-Net vary. So, Zhou et al. proposed U- Net ++, in which U-Nets of different scale levels, and depth supervision are added to allow the network to choose the appropriate depth for itself through training [5].

A further challenge in cell partitioning is the splitting of touching objects of the same class. For this, the use of weighted loss is proposed in U-Net where the isolating background labels separating the touching cells provide a large weight in the loss function.

The architecture is very fast. It takes less than a minute for the segmentation of images of size 512x512 on a recent GPU. The drawback of U-Net architectures is that the learning may slow down in the middle layers of deeper models so that the risk of network learning ignoring the layers is high where abstract features are represented. For further improvements atrous convolutions and U-Net architectures are combined, thereby achieving better performance.

2.3 Atrous Convolution

Atrous convolution is the convolution with hollows, which can expand the receptive field with relatively fewer amount parameters [6]. Atrous Convolutions expand the window size by inserting zero values into each convolution kernel rather than incrementing the quantity of weights.

It is an efficient tool that allows controlling the resolution of features computed by deep convolutional neural networks [7]. It also allows reworking the respective ImageNet networks [8] to extract denser feature maps by removing the down-sampling operations from the last few layers and up-sampling the correlated filter kernels, which is the same as introducing holes between the filter weights. An iterative combination of maximal pooling and striding on successive layers of these networks will significantly reduce the resulting feature map spatial resolution. Deconvolutional layers can be used for recovering spatial resolution.

Proceedings DOI: 10.21467/proceedings.160
ISBN: 978-81-965621-9-9

Series: AIJR Proceedings
ISSN: 2582-3922

Two-dimensional signals are considered for each locale I at the output y and the filter w, and the input x feature map is atrously convolved.

$$Y==\sum_k x(i + r.K]w(k),$$

r is the atrous rate corresponding to the stride with which the input signal is selected, which is analogous to convolving the input x with up sampled filters processed by introducing r-1 zeros among the successive filter values towards each spatial dimension. It also allows to compute the responses of functions in fully convolutional networks. ratio of input image spatial resolution to final output resolution is defined as the Output Stride. For better performance at the top of a feature map is an atrous spatial pyramid pool with four parallel atrous convolutions with different atrous rates [9]. To work on more datasets Deep Lab versions can be used which employs atrous convolution with up sampled filters to draw out dense feature maps and to grab long range context.

2.4 Deep Lab Family

The deep Lab family is considered as one of the most popular semantic segmentation methods developed by chen *et al.* composed of DeepLab v1, DepLabV2, and DeeplabV3 [9]. The Deep Lab applies atrous convolution for up sample.

2.4.1 Deeplabv1

The two challenges: faced in Deeplabv1 is the feature resolution reduction and the localization accuracy reduction due to DCNN invariance. The spatial resolution will be reduced due to multiple pooling and downsampling in DCNN. Instead of removing the down-sampling operator from the last few maximum pooling layers of DCNN, they u-sample the filters in subsequent convolutional layers and evaluate feature maps at excessive sampling rates.

One way to capture fine details with a fully connected CRF, the CRF potentials provide smoothness conditions that maximize label matching between identical pixels and can integrate multiple conditions that model relationships between object classes.

In the Deep labv1 model the images are taken as input and it is passed through DCCN layers followed by one or two Atrous layers it will result in a coarse score map.

The map is scaled to its original size image using bilinear interpolation. To improve the segmentation, result fully connected CRF is used.

2.4.2 Deeplabv2

For improving the performance of Deeplabv1, Deeplabv2 can be used. The difficulty we face in Deeplabv1 is the presence of objects at different scales. To reduce the presence of objects at various scales, different atrous convolutions with dissimilar sampling rates are applied to the input feature map and combined together. Because objects of the same class have dissimilar scales in the image, ASPP will help to account for different object scales and improve accuracy.

2.4.3 Deeplabv3

It is a semantic segmentation architecture used to improve the features of Deeplabv2 with some modifications. To handle segmenting objects at multiple scales, Atrous convolution can be used to capture the multiscale context at multiple Atrous rates in series or parallel. [7]. It was difficult to obtain sharp object boundaries by slowly retrieving spatial information. This architecture uses Atrous convolution to educe the features computed randomly by deep convolutional neural networks. For the problem of image classification, the spatial resolution of the latest feature map is set up in such a way that it is 32 times lower than the input image resolution and the output level is 32. This architecture uses a novel endec with Atrous split convolution to capture sharp object boundaries. Encoder-decoder model will help to obtain sharp object boundaries. Normally, the endec networks contain

a. A unit that works as a coding device that progressively brings down the feature maps and grabs greater image information.

b. b A section that works as a decoding device that progressively resumes the spatial information.

A single convolutional filter can also be applied for each input channel i.e., depth-wise convolution to increase computational efficiency.

Another problem found in this type of model is in the scalability of larger/deeper DCNNs due to limited GPU memory.

Works have been done and identified that a carefully balancing network can lead to better performance. [10] A new core system is designed and scaled to produce a family of models called Efficient Nets having greater accuracy and efficiency than other ConvNets.

2.5 Efficient-Net

A convolutional neural network architecture and scaling method that regularly measures all dimensions of an image such as depth/width/resolution using a composite coefficient [10]. For example, if the computing resources are to be incremented by 2N times, then raise the mesh depth by αN the width by βN, and the image size by γN, where α,β,γ are unchanging coefficients determined by searching the fine mesh on the original slight model. Efficient-Net adopts a compound coefficient represented as Φ to scale the width, depth, and resolution of the network in principle equally. It is one of the most systematic models that acquire the state-of-the-art accuracy for image network along with ideal image classification transfer training function [11].

A key building block of the efficient-net architecture is mobile inverted bottleneck convolution (MBConv) with squeeze and stimulus optimization. The number of these MBConv blocks is different in an Efficient NET network [11]. The Efficient-Net model architecture comprises eight models varying from B0 to B7, with each consecutive pattern digit specifying more parameters and higher precision variables. As we move from functional nets B0 to B7, depth, width, resolution, and model size will increase, and accuracy will also improve. Efficient-Net B7, the best-performing model in terms of accuracy on ImageNet, has a better state of art CNN. and is 8.4x lower and 6.1x quicker than the current leading CNN [12]. The network architecture of EfficientNetB7 is as shown in Fig.1. It can be divided into seven blocks according to filter size, striding, and number of channels.

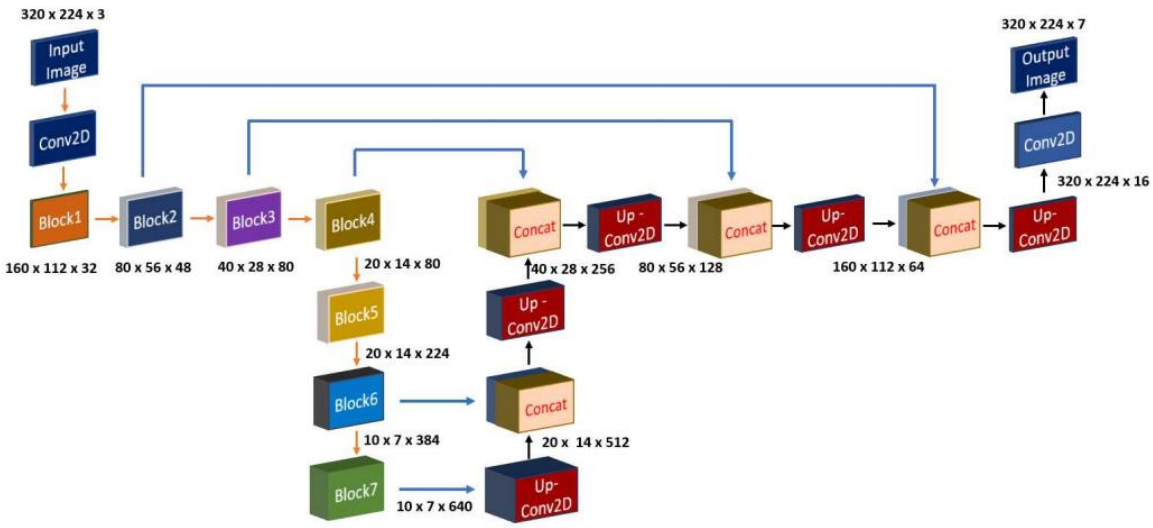

Figure 1: *Network architecture of Efficient – Net B7*

The selection of resolution, depth and width is further controlled by means of various elements.

Proceedings DOI: 10.21467/proceedings.160
ISBN: 978-81-965621-9-9

Series: AIJR Proceedings
ISSN: 2582-3922

Resolution: Resolutions are not divisible by 8,16, etc., causing zero padding at some neighbouring layer extremities, and wasting compute resources. This can be applied to smaller model variants, so the input resolutions of B0 and B1 are preferably 224 and 240, respectively.

Depth and width; Efficient-Net constituents require the channel dimensions to be a multiple of 8.

Resource limit: incrementing the depth and/or width but keeping resolution constant can quietly raise the fulfilment.

This structure takes an input image of shape (224,224,3) and the dataset must be in the range [0,255]. Standardisation is also done on the model.

Efficient-Net –B0 is the standard network developed by AutoML MNAS, and Efficient-Net B1 to B7 are obtained by extending the core network. For the B0 to B7 base models, the datasets are varying, in particularly, efficient-net - B7 achieves a new state-of-the-art accuracy 84.4% top-1/97.1% top5 accuracy, while being 8.4times less than the state of art CNNs. The performance graph is as shown in Fig. 2.

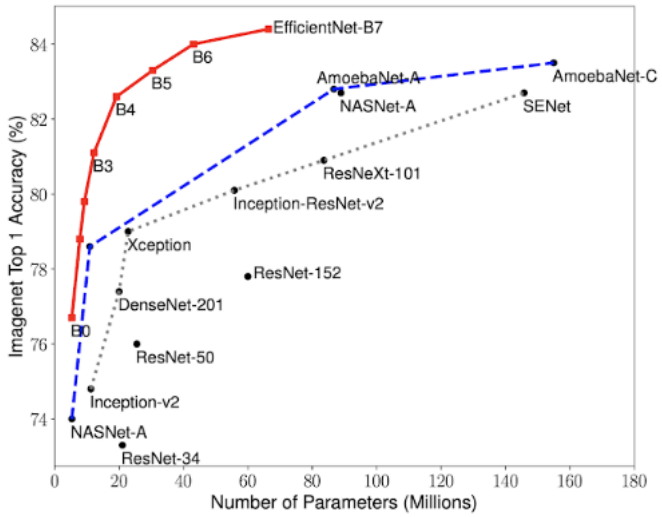

Figure 2: *Efficient-Net Performance*

Efficient Net B4 enhances top-1 efficiency from 76.3% to 83.0% with identical FLOPS.

With the ability to create interdependencies between channels or spatial locations, attention mechanisms can be used in computer vision tasks. Triple Attention uses an efficient attention calculation method that has no information bottlenecks [13]. It also improves the underlying performance of architectures like Res-Net, Efficient Net, etc. for tasks like image classification on Image Net.

3 Conclusion

It is found that there is no absolute technique for image segmentation of the reason that the outcome of image segmentation relies on many aspects, namely pixel color, texture, responsiveness, the resemblances of images, image content, problem domain, etc.

Deep learning is quite a significant tool for image segmentation. The representational capability of neural networks is outstanding, and they are focused on use in computer vision and natural language processing, turning tasks that seem too difficult or even beyond reasonable tasks. Deep learning technology is in its inception. It is possible to develop new models with better accuracy as well as improvements in the diagnostic and modelling domains and future research in the areas of segmentation using efficient net models. Efficient-Net's results outperform all previous architectures on most benchmarking datasets.

4 Publisher's Note

AIJR remains neutral with regard to jurisdictional claims in institutional affiliations.

How to Cite

Paulose & Vanitha (2023). Image Segmentation using Optimization Algorithm: A Survey. *AIJR Proceedings*, 317-322. https://doi.org/10.21467/proceedings.160.41

References

[1] A. Kendall, V. Badrinarayanan, and R. Cipolla, segnet: A deep convolutional encoder-decoder architecture for image segmentation" arXiv preprint arXiv:1511.02680, 2015.

[2] Olaf Ronneberger, Philipp Fischer, Thomas Brox, U-Net – Convolutional Networks for Biomedical Image Segmentation, Springer ,2015

[3] Guoheng Huang, Junwen Zhu, JaijianLi, Zhuowei Wang, Lianglun Cheng, Lizhi Liu, Haojiang Li, Jian Zhou, Channel – Attention U-Net: Channel Attention Mechanism for semantic segmentation of Esophagus and Esophageal Cancer- IEEE Access 2020

[4] M.Zahangir Alom, M. Hasan.C. Yukopcic, T M Taha and V.K Asari" Recurrent residual convolutional neural networks based on U-Net for medical image segmentation" 2018, arXiv:1802.06955

[5] Z. Zhou.M. R Siddique, N. Tajhakhish and J. Liang" U- Net ++: A Nested U- Net architecture for medical image segmentation"in Deep Learning in Medical Image Analysis and Multimodal Learning for clinical Decision support. Springer 2018

[6] Yu. F Koltun, V: Multiscale context aggregation by dilated convolutions. arXiv:1511.07122-2015

[7] Liang-Chieh, Chen George Papandreou Florian Schroff, Harting Adam :Rethinking Atrous Convolution for semantic image segmentation. arXiv:1706.05587v3 – 2017

[8] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. IJCV, 2015.

[9] Chen L C, Papandreou G , Kokkinos I, Murphy K, Yulle A L :DeepLab : Semantic Image Segmentation with deep Convolutional Nets, atrous Convolution and fully Connected crfs.arXiv.1606.00915- 2017

[10] Mingxing Tan and Quoc V. Le. Efficient net: Rethinking model scaling for convolutional neural networks, 2020

[11] Chaurasia and E. Culurciello, "Linknet: Exploiting encoder representations for efficient semantic segmentation," in 2017

[12] A Shamila Ebenezer, S Deepa Kanmani, Mahima Siva Kumar, S Jeba Priya: Effect of image transformation on Efficient Net model for COVID-19 CT image classification -2022

[13] Diganta Misra, Trikay Nalamada, Ajay Uppili, Qibin Hou: Rotate to Attend: Convolutional Triplet Attention Module: arXiv:2010.03045v2,2020