Comprehensive Survey on Skin Segmentation on Dermoscopic Images using SSS-Net

S. A. NEELAVANI[1], DR. K. KUMAR[2]

[1]M.E Student, Department of Computer Science and Engineering, Government College of Technology, Coimbatore.
[2]Associative Professor, Department of Computer Science and Engineering, Government College of Technology, Coimbatore.

---***---

ABSTRACT - *Skin segmentation plays a crucial role in act recognition, video surveillance, hand gesture identification, face detection, human tracking and robotic surgery. The accurate segmentation of the skin is important to acknowledge the act. Skin segmentation on dermoscopic images is proposed supported the Skin Semantic Segmentation Network (SSS-Net). SSS-Net is employed for skin segmentation task for the semantic labelling of pixels during a pixel wise classification framework. SSS-Net is strong to skin detection. A way smaller deep neural network is proposed for skin segmentation that doesn't require additional pre-processing steps. Fully Convolutional Network (FCN) [16] is employed to segment the skin. FCN gives the accurate segmentation, it takes input as dermoscopic images and converting them into a mask with regions of interest highlighted.*

Key Words: Skin Segmentation, SSS-Net, Fully Convolutional Network, Dermoscopic Images, Deep Neural Network.

1. INTRODUCTION

Skin cancers square measure comparatively uncommon malignancies worldwide, not ranking among the primary 10 common cancers. There has been a progressive increase within the incidence of skin cancers, notably that of body covering melanomas over the previous few decades. Three most frequent primary skin cancers square basal cell carcinoma (BCC), squamous cell carcinoma (SCC) and malignant melanoma. Along SCC and BCC square measure said as non-melanomatous skin cancers (NMSC).

In India, skin cancers represent concerning 1-2% of all diagnosed cancers. Basal cell carcinoma disease is that the commonest sort of carcinoma worldwide, however numerous studies from India have systematically reportable SCC because the most prevailing skin malignancy. Though complete information of incidence isn't on the market, numerous cancer registries in India reportable additive incidence of carcinoma varied from 0.5 to 2 of per 1,00,000 population. Though, the incidence of skin cancers in India is lower as compared to the Western world, attributable to an oversized population, absolute range of cases is calculable to be important [17].

1) The task of skin segmentation is sculpturesque as a semantic pixel-wise segmentation downside. For this reason, an SSS-Net with reduced tunable hyper parameters is assumed of. This method features a tendency to believe this work can facilitate bridge the gap between skin segmentation and semantic Segmentation.

2) Low-level semantic information is preserved and thus the preservation of edge info results in robust detection of skin information.

3) The planned methodology is powerful to skin detection.

4) How smaller (in terms of tunable parameters) deep neural network is planned for skin segmentation that doesn't need extra pre-processing steps.

5) Low procedure time overhead throughout reasoning in both train and take a glance at stages.

2. LITERATURE SURVEY

Skin segmentation is the pre-processing step for medical image processing. There are several methods used for skin segmentation. In this Literature survey many methods are discussed. These methods are used to segment the skin. But these methods have some limitations. Table 1 shows the limitations of existing methods. To overcome these limitations Semantic segmentation is used. Semantic Skin Segmentation can give the accurate outputs.

SSS-Net is used for skin segmentation tasks for the Semantic labeling of pixels during a pixel-wise classification framework. Contributions of this work are

Table-1: Comparison of Existing Papers

S. No	Title	Method/ Algorithm	Pros	Cons
1	Accurate Pixel-Wise Skin Segmentation Using Shallow Fully Convolutional Neural Network	Fully Convolutional Neural Network [1].	It gives accurate pixels value.	Datasets contain a smaller number of images.
2	Hand Gesture Recognition Techniques for Human Computer Interaction Using OpenCV	Convex hull [2].	This method increases the freedom of usability. These methods are going to use in different applications in future.	Not gives an accurate output.
3	Skin color segmentation using multi-color space threshold	Multi-color space threshold [3].	Multi color can be included with the dynamic detection of skin.	Thresholding can cause a problem with skin like background pixels.
4	Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation	Semantic Segmentation [4].	This model is faster and stronger.	This model fails to segment heavily occluded objects and objects with rare view.
5	Skin detection by dual maximization of detectors agreement for video monitoring	Color channel method [5].	A convenient way to estimate pixels in the image	Changes in illumination could lead to false result.
6	A weighted skin probability map for skin color segmentation	WSPM [6].	Method is helpful in reducing the error in detection.	Errors related to skin tone can be caused by threshold.
7	Segmentation of hand posture against complex backgrounds based on saliency and skin color detection.	Color information model [7].	Can perform well in indoor and outdoor environments.	Results will be affected by the involvement of different person skin pixels.
8	Improved skin detection based on dynamic threshold using multi-color space.	Multi-colored threshold [8].	Multi color can be included with the dynamic detection of skin.	Thresholding can cause a problem with skin like background pixels.
9	Skin detection based on image color segmentation with histogram and k-means clustering	k-means clustering algorithm [9]	Non-skin pixels detection is reduced by removing the background pixels.	Step needed for the refined detection results
10	Vision based hand gesture recognition using eccentric approach for human computer interaction	Sobel edge detection [10].	A simple image processing technology.	Results for edge detection are not satisfactory for the cluttered background.
11	Comparative study of skin color detection and segmentation in HSV	Histogram-based technology [11].	A simple histogram technology is used for threshold setting.	Threshold needs to be adjusted.

	and YCbCr color space			
12	Robust skin detection in real-world images	SCTGC [12]	It improves the accuracy	Not effective for skin detection in multiple people interaction.
13	Hybrid human skin detection using neural network and K-Means clustering technique	MLP and k-means cluttering algorithm [13].	More reliable approach than hand crafted schemes for feature extraction.	Time consuming as it is based on patch based.
14	Image skin segmentation based on multi-agent learning Bayesian and neural network	Bayesian method Neural network [14].	More efficient than other approaches.	Data is dependent on particular factor.
15	Deep learning-based hand detection in cluttered environment using skin segmentation	Two stage deep learning approach [15].	A very positive result has been seen by hand detection method.	Time consuming as it is based on patch based.

3. NETWORK ARCHITECTURE

Convolution is that the primary layer to extract options from associate input image. Convolution preserves the affiliation between pixels by learning image options mistreatment little squares of computer file. It's a computing that takes two inputs like image matrix and a filter or kernel. Batch normalization could be a technique for coaching terribly deep neural networks that standardizes the inputs to a layer for every mini-batch. This has the result of helpful the training method and dramatically reducing the quantity of coaching epochs needed to coach deep networks.

The ReLU (short for corrected Linear Units) layer usually follows the convolution layer. The addition of the ReLU layer permits the neural network to account for non-linear relationships, i.e., the ReLU layer permits the convnet to account for things within which the connection between the pel price inputs and therefore the convnet output isn't linear. Note that the convolution operation could be a linear one. The output within the feature map is simply the results of multiplying the weights of a given filter by the pel values of the input and adding them up (1):

$$y = W_1X_1 + W_2X_2 + W_3X_3 + \ldots \quad (1)$$

where W could be a weight price and X square measure a pel price.

The ReLU perform takes a price x and returns zero if x is negative and x if x is positive. Max pooling could also be a pooling operation that selects the utmost element from the region of the feature map lined by the filter. Thus, the output once max-pooling layer would be a feature map containing the foremost distinguished choices of the previous feature map.

A building block of a ResNet is termed a residual block or identity block. A residual block is just once the activation of a layer is fast-forwarded to a deeper layer within the neural network. ASPP [4] is employed to get multi-scale context information. The prediction results square measure obtained by up-sampling. within the ASPP network, on high of the feature map extracted from backbone, four parallel Atrous convolutions with totally different Atrous rates square measure applied to handle segmenting the item at different scales. Image-level options are applied to include international context info by applying international average pooling on the last feature map of the backbone. once applying all the operations parallelly, the results of every operation on the channel square measure concatenated and one x one convolution is applied to urge the output.

Transposed Convolutions network unit accustomed upsample the input feature map to a desired output feature map mistreatment some learnable parameters. SoftMax assigns decimal chances to every category in a very multi-class drawback. Those decimal chances should add up to 1.0. This extra constraint helps coaching converge a lot of quickly than it otherwise would. SoftMax is enforced through a neural network layer simply before the output layer. The SoftMax layer should have identical range of nodes because the output layer. Pixel Classification Layer creates a picture component classification output layer for linguistics image segmentation networks.

The layer outputs the particular label for every image element or voxel processed by a CNN. The layer mechanically ignores obscure element labels throughout work.

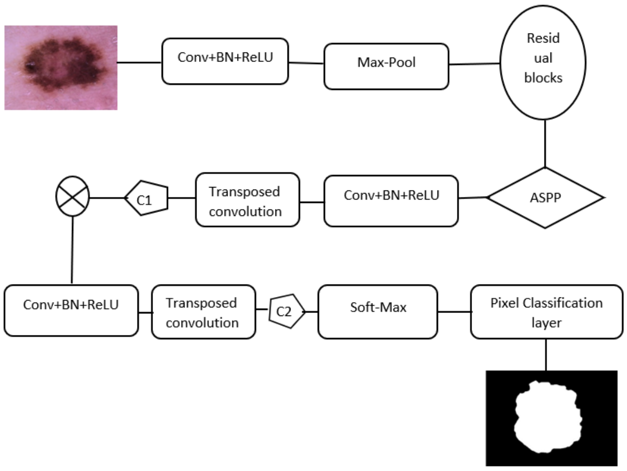

Fig – 1: SSS-NET Architecture

Conv- convolution layer **BN-** Batch Normalization **ASPP-** Atrous Spatial Pyramid Pooling **c1, c2-** Cropping

3.1 SSS-NET ARCHITECTURE

In Skin Semantic Segmentation architecture take dermoscopic image as an input. In the encoder part Input image is sent to the sequence of Convolution layer, BatchNormalization and ReLU Operation. Convolution layer extract the features from the given input. BatchNormalization train the deep networks of given input. ReLU operation is the Operation of Non-Linear operation. It takes a price x and returns zero if x is negative and x if x is positive.

Every sequence of Conv+BN+ReLU has a Max-Pooling layer. Max-Pool layer reduces the size of the image. SSS-NET architecture is shown in the Figure 1. SSS-NET architecture contains a four residual blocks. Each block contains the Conv+BN+ReLU. Every residual block includes of two 3*3 convolutions and to reduce the dimensions of the image every of the block is interpolated with max-pooling operation. A shortcut affiliation is provided to every residual block, which mixes the input with results of residual block before applying ReLU in second convolution of the block. This affiliation allows the previous layers to induce the powerful gradient signal that makes training straightforward for the deeper networks. Encoder with a feature map size of the residual block is shown in Table 2.

In SSS-Net, ASPP captures multi scale discourse information and applies various dilation rates to a sequence of Atrous convolutions. These rates area unit designed to capture the longer context. ASPP integrates image-level options to feature global context info. ASPP contains four operations one 1*1 convolution and three 3*3 convolutions. These convolutions used dilation rates 4,12 and 16.

In decoder part SSS-Net used backward convolution layer to upsample the choices coming back from the encoder part succeeding in high resolution image from a low-resolution image. This is followed by concatenation with the following low-level network choices of an analogous resolution.

Table-2: Encoder with A Feature Map Size of The Residual Block

Blocks in Encoder	Name and Size	No. of Filters	Output feature map size	No. of parameters
Encoder Res block-1	EConv-2a_1**/3*3*64	64	60*80*64	36928
	EConv-2a_b**/3*3*64	64	60*80*64	36928
Encoder Res block-2	EConv-2a_1**/3*3*64	64	60*80*64	36928
	EConv-2a_b**/3*3*64	64	60*80*64	36928
Encoder Res block-3	EConv- 1_1/1*1*128	128	30*40*128	8320
	EConv-2a_1**/3*3*128	128	30*40*128	73856
	EConv-2a_b**/3*3*128	128	30*40*128	147584
Encoder Res block-4	EConv-2a_1**/3*3*128	128	30*40*128	147584
	EConv-2a_b**/3*3*128	128	30*40*128	147584

Decoder with a feature map size of the residual block is shown in Table 3.

Table-3: Decoder with A Feature Map Size of The Residual Block

Blocks in Decoder	Name and Size	No of Filters	Output feature map size	No. of Parameters
Decoder conv. block-1	DConv c1_1**/1*1*256	256	30*40*256	262400
Upsample	Upsample1_1/8*8*256	256	120*160*256	4194560
Decoder conv. block-3	DConv-c3_1**/3*3*256	256	60*80*256	700672
Decoder conv. block-4	DConv-c4_1**/3*3*256	256	60*80*256	590080
Upsample	Upsample2_1/8*8*2	2	240*320*2	258

On these low-level features 1*1 convolution with 256 filters is applied thus as to lessen the quantity of channels, as a result of the resultant low-level features generally have an oversized vary of channels and make the employment of network additional sturdy. A factor of four is applied once concatenation to refine the choices following another straightforward bilinear upsampling.

4. CONCLUSIONS

This paper planned SSS-Net for skin segmentation that's able to capture the multiscale discourse info and provide results with refined edge boundaries. In this paper, dermoscopic images are segmented using SSS-Net. SSS-Net has less variety of layers which ends in reduced variety of parameters i.e., 7.3 M that significantly lower compared to alternative existing networks. Moreover, SSS-Net doesn't require any extra pre-processing steps. The individuality of this network is its ability to capture the multi-scale discourse information. The obtained results show high-quality segmentation results, indicating the effectiveness of SSS-Net for skin segmentation.

Future work can specialize in the extraction of distinctive skin lesion options for the classification of skin cancer skin lesion in dermoscopic pictures mistreatment the PCDS rule for segmentation.

REFERENCES

[1] Komal Minhas and Tariq M. Khan, “Accurate Pixel-Wise Skin Segmentation Using Shallow Fully Convolutional Neural Network” in proc. IEEE journal vol 8., Aug 2020.

[2] M. Panwar and P. Singh Mehra, “Hand gesture recognition for human computer interaction,” in Proc. Int. Conf. Image Inf. Process., Nov. 2011, pp. 367-374.[Online]. Available:http://www.sciencedirect.com/science/article/pii/S1877050917319130.

[3] R. F. Rahmat, T. Chairunnisa, D. Gunawan, and O. S. Sitompul, “Skin color segmentation using multi-color space threshold,'' in Proc. 3rd Int. Conf. Comput. Inf. Sci. (ICCOINS), Aug. 2016, pp. 391-396.

[4] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam,”Encoderdecoder with Atrous separable convolution for semantic image segmentation,” in Proc. Eur. Conf. Comput. Vis. (ECCV), 2018, pp. 801-818.

[5] J. C. SanMiguel and S. Suja, “Skin detection by dual maximization of detectors agreement for video monitoring” Pattern Recognit. Lett., vol. 34, no. 16, pp.2102-2109,Dec.2013.[Online].Available: http://www.sciencedirect.com/science/article/pii/S0167865513002936

[6] B. K. Chakraborty, M. K. Bhuyan, and S. Kumar, “A weighted skin probability map for skin color segmentation,” in Proc. Int. Conf. Wireless Commun., Signal Process. Netw. (WiSPNET), Mar. 2016, pp. 2133-2136.

[7] Q. Zhang, M. Yang, K. Kpalma, Q. Zheng, and X. Zhang, “Segmentation of hand posture against complex backgrounds based on saliency and skin colour detection,” IAENG Int. J. Comput. Sci., vol. 45, pp. 435-444, 08-2018.

[8] M. Z. Osman, M. A. Maarof, and M. F. Rohani, “Improved skin detection based on dynamic threshold using multi-colour space,” in Proc. Int. Symp. Biometrics Secur. Technol. (ISBAST), Aug. 2014, pp. 29-34.

[9] E. Buza, A. Akagic, and S. Omanovic,” Skin detection based on image color segmentation with histogram and k-means clustering,” in Proc. 10th Int. Conf. Electr. Electron. Eng. (ELECO), 2017, pp. 1181-1186.

[10] V. Bhame, R. Sreemathy, and H. Dhumal, ``Vision based hand gesture recognition using eccentric approach for human computer interaction,'' in Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI), Sep. 2014, pp. 949-953.

[11] K. B. Shaik, P. Ganesan, V. Kalist, B. S. Sathish, and J. M. M. Jenitha, “Comparative study of skin color detection and segmentation in HSV and YCbCr color space,” Procedia Comput. Sci., vol. 57, pp. 41-48, 2015.[Online].Available:http://www.sciencedirect.com/science/article/pii/S1877050915018918.

[12] L. Huang, W. Ji, Z. Wei, B.-W. Chen, C. C. Yan, J. Nie, J. Yin, and B. Jiang, “Robust skin detection in real-world images,” J. Vis. Commun. Image Represent., vol.29,pp.147-152,May2015.[Online].Available: http://www.sciencedirect.com/science/article/pii/S1047320315000280

[13] H. K. Al-Mohair, J. Mohamad Saleh, and S. A. Suandi, “Hybrid human skin detection using neural network and K-Means clustering technique,” Appl. SoftComput.,vol.33,pp.337-347,Aug.2015.[Online]. Available:http://www. Sciencedirect.com/science/article/pii/S1568494615002732

[14] Zaidan, N. N. Ahmad, H. Abdul Karim, M. Larbani, B. B. Zaidan, and A. Sali, Image skin segmentation based on multi-agent learning Bayesian and neural network,'' Eng. Appl. Artif. Intell., vol. 32, pp. 136-150,Jun.2014.[Online].Available: http://www.sciencedirect.com/science/article/pii/S0952197614000578

[15] K. Roy, A. Mohanty, and R. R. Sahay, ``Deep learning based hand detection in cluttered environment using skin segmentation,'' in Proc.IEEE Int.Conf. Comput. Vis. Workshops (ICCVW), Oct. 2017, pp. 1-10.

[16] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic

segmentation," in Proc.IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2015, pp. 3431-3440.

[17] https://www.indianjcancer.com/article.asp?issn=0019-509X;year=2005;volume=42;Issue=3;spage=145;epage=150;aulast=Deo.