

# Deep learning image processing framework for automatic segmentation of skin lesions

Joshua Peter Ebenezer<sup>a,\*</sup>, Jagath C. Rajapakse<sup>b</sup>

<sup>a</sup>*Indian Institute of Technology Kharagpur, India*

<sup>b</sup>*Nanyang Technological University, Singapore*

---

## Abstract

Skin cancer is the most common form of cancer in the world and is generally diagnosed visually by dermatologists. This paper presents a fully automated method to accurately segment lesion boundaries from dermoscopic images. The segmentation of skin lesions aids in computer aided diagnosis as well as in accurate surgical efforts. A U-net deep learning network is trained on publicly available data from the International Skin Imaging Collaboration (ISIC). We introduce the use of intensity, color, and texture enhancement operations as pre-processing steps and morphological operations and contour identification as post-processing steps. We empirically show that our deep learning image processing framework gives state-of-the-art results. On the ISIC 2017 challenge's test data, our framework reports a mean Jaccard index of 0.766, which would have placed our method first in the challenge. We also show that our pre-processing model can significantly improve the performance of a network over training with plain color images.

*Keywords:* image enhancement, deep learning, image segmentation, skin cancer, U-net

---

\*Corresponding author

Email address: [joshuapeterebenezer@gmail.com](mailto:joshuapeterebenezer@gmail.com) (Joshua Peter Ebenezer)

## 1. Introduction

Skin cancer is the most common malignancy in the world and accounts for more than 40% of all cases worldwide (American Cancer Society, 2018). Melanoma is the most aggressive form of skin cancer and accounted for 59,800 deaths in 2015 (Global Burden of Disease Study, 2015). Early diagnosis is key to curing skin cancer that can otherwise prove to be fatal. In the U.S., the ratio of dermatologists to the population is 3.2:100,000 (Yoo and Rigel, 2010), which highlights the need and opportunity for computer-aided diagnosis. Considering the large number of cases that dermatologists have to deal with, computer-aided-diagnosis can help ease their workload and also provide a valuable second opinion. In this regard, a number of attempts (Masood and Ali Al-Jumaily, 2013; Ferris et al., 2015; Esteva et al., 2017) have been made to accurately classify skin lesions into various disease categories. The segmentation of the lesion from the dermoscopic image aids this effort by discarding the irrelevant portions of the image and isolating the affected region for further study. Signs of a lesion being malignant such as the asymmetry of the lesion, irregular borders, and the large diameter of the lesion (Schmid-Saugeon, 2000) are dependent on accurate segmentation of the lesion itself and hence segmentation can improve classification accuracies. Segmentation is also important because treatment of skin cancer often involves the surgical excision of the affected region with as little possible damage to the surrounding tissue. An automated and reliable method for this would thus greatly assist dermatologists.

In recent years, many efforts have been made in this regard. These efforts can be broadly divided into three classes: thresholding based, region based, and contour based. Recently, methods based on deep learning have also appeared, that form a class of their own. In the first category, the classic Otsu's thresholding and its adaptations have been extensively explored (Celebi et al., 2007; Abbas et al., 2013). Fuzzy set algorithms have been proposed to find the optimum threshold (Lee and Chen, 2014). Adaptive and iterative thresholding methods have also been proposed in literature (Barata et al., 2014; Rajab et al., 2004). In

the second category, region growing (Iyatomi et al., 2008), statistical, and iterative region merging have been proposed (Wong et al., 2011; Celebi et al., 2008). Contour-based techniques include gradient vector flow based methods (Zhou et al., 2013), adaptive snake (Silveira et al., 2009), and level set (Swierczynski et al., 2018), among others.

Deep learning techniques for segmentation have become very popular over the past few years for medical image segmentation tasks. These have included the U-Net architecture (Ronneberger et al., 2015; Cicek et al., 2016), fully convolutional neural networks (Jiang et al., 2018) and recurrent neural networks (Xie et al., 2016; Andermatt et al., 2016) and their derivatives being used for medical image segmentation tasks. Custom architectures, modified loss functions, novel data augmentation methods, and combinations of various statistical methods have been used to advance the field. Deep learning techniques generally require large labelled datasets for training.

The International Skin Imaging Collaboration (Codella et al., 2018) (ISIC), an international effort to improve melanoma diagnosis, holds a recurring challenge every year for the development of image analysis tools for segmentation of dermoscopic images. ISIC curates the largest publicly available collection of quality controlled dermoscopic images of skin lesions. In particular, they curate the largest dataset for skin lesion segmentation, with corresponding ground truth masks generated using three methods, all vetted by practicing dermatologists. The masks are generated by a (a) fully-automated algorithm, reviewed and accepted by a human expert or (b) a semi-automated floodfill algorithm, with parameters chosen by a human expert, or (c) manual polygon tracing by a human expert. The lesion images are acquired with a variety of dermatoscope types, from all anatomic sites (excluding mucosa and nails), from a historical sample of patients presented for skin cancer screening, from several different institutions. The data released for the ISIC 2017 challenge were used for this study. Submissions to these challenges utilised various methods and in general have been almost wholly based on deep learning networks. Yuan and Lo (Yuan and Lo, 2017) used an ensemble of convolutional-deconvolutional networks with

RGB and hue, saturation and value (HSV) channels as input and won the 2017 challenge, which had 21 participating teams. Berseth (Berseth, 2017) used a U-net architecture (Ronneberger et al., 2015) with elastic distortion as a pre-processing step and won the second place. Bi et al. (Bi et al., 2017) used a fully convolutional network using the ResNet architecture and obtained the third place. Ensembles of encoder-decoder networks have been used by many participants in the segmentation challenge and have obtained commendable results. Most of these approaches have focused on tuning network parameters with RGB color-space input and sometimes with HSV color-space input as well. In the recently concluded 2018 challenge, the winning team used an object detection architecture (He et al., 2017) to find an ROI, and then fed the ROI through a network similar to DeepLab (Chen et al., 2018). The ground truth and metrics for the 2018 ISIC test set have not yet been released for evaluation.

In this paper, we present a deep learning architecture and an image processing pipeline for accurate segmentation of skin cancer lesions from dermoscopic images. Our contributions are using (i) a novel application of intensity, color, and texture enhancement as pre-processing steps, (ii) morphological operations and contour identification as post-processing steps, and (iii) a modified U-net architecture predicting the segmentation mask. Images of lesions generally have poor contrast and boundaries of the lesions are often hard to outline as lesion color is similar to the color of the skin near the edges. In general, FCNs and U-Nets are pushing the boundaries of the field and have already crossed 90% accuracy in the task. Further improvement is likely to come from careful pre-processing so that the networks train faster and better. The pre-processing we propose is able to significantly enhance the information in the images and suppress noise, making the lesion easier to distinguish from the skin. We compare the results from our approach with results of other state-of-the-art methods, evaluated on ISIC 2017 challenge data, as well as results of training the network with other pre-processing methods, and show, by widely-accepted metrics, that the proposed method gives state-of-the-art results, and that the pre-processing framework we introduce gives significantly better results than training the net-

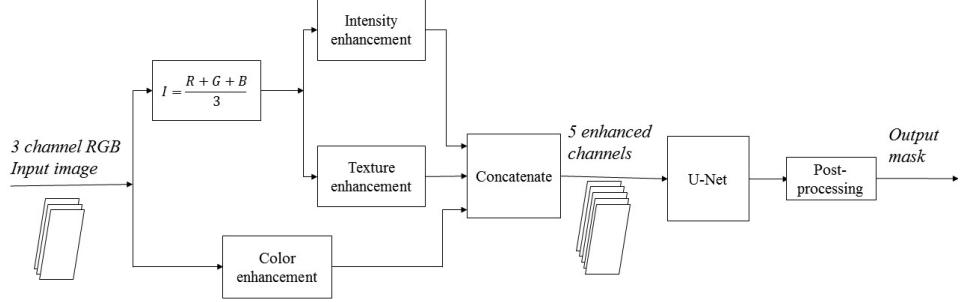


Figure 1: Schematic of the proposed method.

work with RGB data.

## 2. Methodology

We propose a three-fold enhancement pre-processing pipeline that enhances intensity, texture, and color separately. The intensity enhancement and the texture enhancement operations work on the intensity of the input color image and each generate a single channel image, and the color enhancement operation generates 3-channel color enhanced images from the input dermoscopic image. These 5 channels are then concatenated and trained by a network that is similar in architecture to the U-Net. The output segmentation map is post-processed to generate the final result. A schematic of the method is illustrated in Fig. 1.

### 2.1. Intensity contrast enhancement

Skin lesion images are color images. The contrast enhancement algorithm based on the layered-difference representation of 2D histograms proposed by Lee et al. (Lee et al., 2013) is used to enhance the contrast of the intensity of the image, where intensity is defined as the arithmetic mean of the red (R), blue (G), and green (G) channels. If a pair of pixels in the input image have intensities  $k$  and  $k + l$ , they are mapped to the levels  $x_k$  and  $x_{k+l}$  in the output image respectively. The difference  $d_k^l$  is defined as

$$d_k^l = x_{k+l} - x_k \quad (1)$$

for  $0 \leq k \leq 255 - l$ . The 2D histogram  $h(k, k + l)$  represents the number of pairs of adjacent pixels in the input image having pixel intensities  $k$  and  $k + l$ . The histogram is logarithm-attenuated to prevent contrast over-stretching in the output image, as

$$h_k^l = \log(h(k, k + l) + h(k + l, k)) \quad (2)$$

<sup>115</sup> The result is unordered because the order of adjacency is not important. The difference in intensities in any given pair of the output pixels should be proportional to the frequency of their occurrence in the input. This is intuitive because we would like pixel pairs that occur frequently to have larger contrast, and pixel pairs that are infrequent are not very important to the enhancement <sup>120</sup> operation, considering that the intensity differences are constrained by the dynamic range and have to be distributed appropriately. Hence, we impose the relationship

$$d_k^l = \kappa_l h_k^l \quad (3)$$

for  $0 \leq k \leq 255 - l$ . It can be noted that

$$x_k = \sum_{l=0}^{l=k-1} d_k^l \quad (4)$$

and

$$d_k^l = \sum_{i=k}^{i=k+l-1} d_i^1 \quad (5)$$

<sup>125</sup> for  $0 \leq k \leq 255 - l$ . These can be visualized in the form of a layered difference tree, shown in Fig. 1.

Using equations 3 and 5, we can formulate the linear equation

$$\mathbf{A}_l \mathbf{d}_l = \kappa_l \mathbf{h}_l \quad (6)$$

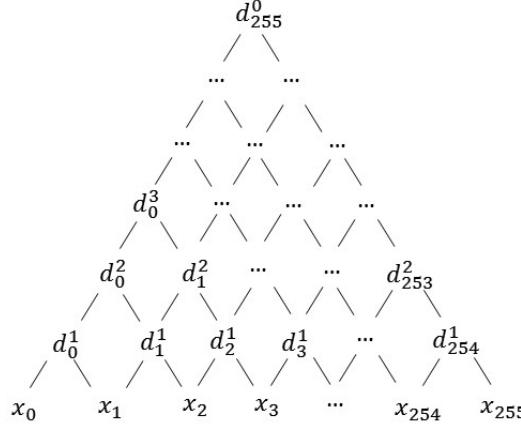


Figure 2: Layered difference representation. The transformation function lies at the lowest level.

for each level  $l$ , where  $\mathbf{A}_l$  is a binary matrix of 0s and 1s that represent equation 5 for all  $k$  for a particular  $l$ . This leads to the constraint optimization problem

$$\mathbf{d}_l = \underset{\hat{\mathbf{d}}_l}{\operatorname{argmin}} \|\mathbf{A}_l \hat{\mathbf{d}}_l - \kappa_l \mathbf{h}_l\|^2 \quad (7)$$

subject to

$$\mathbf{d}_l \succeq 0 \quad (8)$$

and

$$\mathbf{1}^T \mathbf{d}_l = 255 \quad (9)$$

The last two constraints ensure that the output intensities monotonically increase with input intensity and that the full dynamic range of pixel intensities is covered. Incorporating the assumption that the output transfer function must be smooth, it can be shown that the solution to this problem is

$$\mathbf{d}_l = \phi_l^{\max} (\mathbf{B}_l^T \mathbf{B}_l)^{-1} (\mathbf{B}_l^T \mathbf{g}_l - \min(\mathbf{B}_l^T \mathbf{g}_l) \cdot \mathbf{1}) \quad (10)$$

where

$$\phi_l^{\max} = \frac{255}{\mathbf{1}^T (\mathbf{B}_l^T \mathbf{B}_l)^{-1} \mathbf{B}_l^T \mathbf{g}_l - \min(\mathbf{B}_l^T \mathbf{g}_l) \cdot \mathbf{1}^T (\mathbf{B}_l^T \mathbf{B}_l)^{-1} \mathbf{1}} \quad (11)$$

and

$$\mathbf{B}_l = \begin{bmatrix} \mathbf{A}_{l,1} \\ \mathbf{A}_{l,2} \\ \vdots \\ \mathbf{A}_{l,l} \end{bmatrix}, \mathbf{g}_l = \begin{bmatrix} \mathbf{h}_l \\ \mathbf{h}_l \\ \vdots \\ \mathbf{h}_l \end{bmatrix} \quad (12)$$

This directly leads to the transformation function. Two dermoscopic images enhanced by this method are shown in Fig. 2. The edges of the lesion are clearer and details in the interior and in the skin around the lesion are now much more discernible.

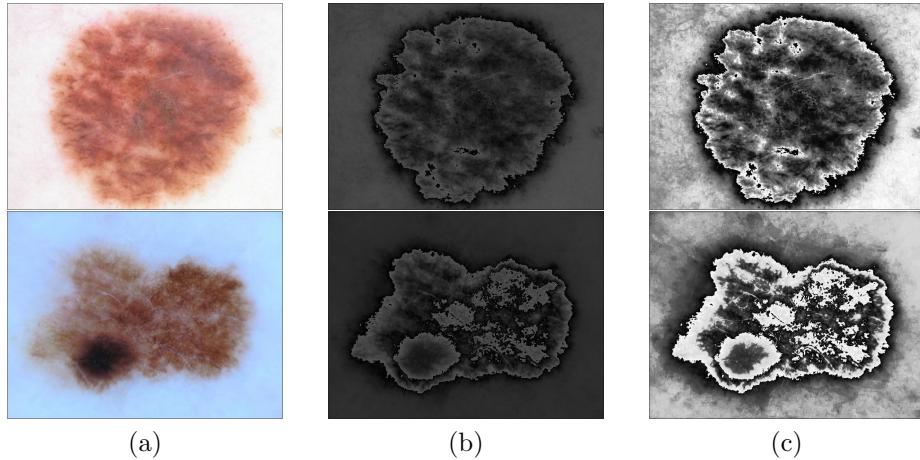


Figure 3: Illustration of samples of skin lesion images: (a) original color image, (b) intensity image (c) contrast-enhanced intensity image.

## 2.2. Hue-preserving color enhancement

Color provides details of skin lesions and following Naik and Murthy (Naik and Murthy, 2003), we apply a hue-preserving color enhancement technique to enhance the contrast of the color images. Hue is the attribute of color according to which an area of an image appears similar to a perceived primary color. It is physically related to the wavelength at which the energy output of the image source is maximum. Mathematically, it can be defined in the three-dimensional

<sup>150</sup> space of  $R$ ,  $G$ , and  $B$ , with respect to the tilted *RGB* cube formed by considering  
 the black-white axis as the  $z$  axis with the edges constrained by the normalized  
 range of  $R, G, B \in [0, 1]$ . The hexagonal projection of the tilted cube on the  
 plane perpendicular to the black-white plane has red, yellow, green, cyan, blue,  
 and magenta as the corners. Hue is the angle of the vector to a point in the  
<sup>155</sup> projection, with red at  $0^\circ$ , and can be written as.

$$h_{rgb} = \text{atan2}(\sqrt{3}(G - B), 2R - G - B) \quad (13)$$

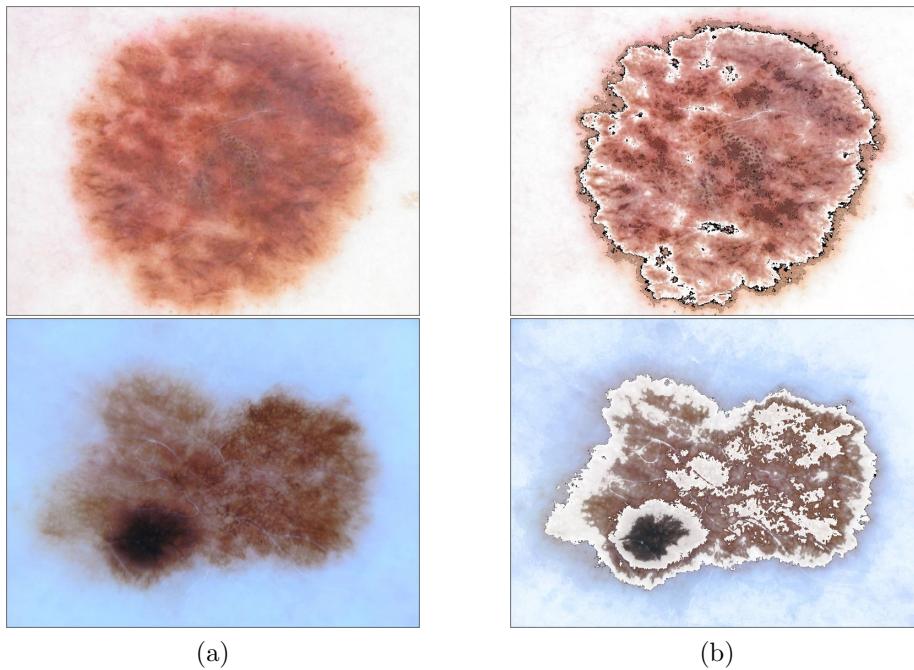


Figure 4: Illustration of samples of skin lesion images: (a) original color image, (b) hue-preserved color-enhanced image.

It is clear from this expression that if  $R, G$  and  $B$  are scaled by the same factor, the hue does not change. Also, the hue is unchanged if scaling is done in the *CMY* color space as well. This follows from the definition above.

<sup>160</sup> Color enhancement techniques (Strickland et al., 1987; Thomas et al., 1997)  
 generally involve non-linear transformations between color spaces, which lead

to the gamut problem when pixel values go out of the acceptable bounds of the RGB space during conversion, which causes undesirable changes in the hue. For example, in order to extend intensity enhancement to color, one technique is to convert the RGB space to the YUV space, apply an intensity enhancement technique to Y without modifying U or V, and convert the image back to RGB. These techniques cause the hue to get distorted or cause the values to get clipped. The present hue-preserving color enhancement takes care of the gamut problem and is described in Algorithm 1 where  $R$ ,  $G$ , and  $B$  denote three RGB color images, respectively, and  $C$ ,  $M$ ,  $Y$  denote CMY color images, respectively.  $\tilde{X}$  represents enhanced image of  $X$ . In case the scaling factor  $\alpha$  is more than 1, which could lead to the gamut problem, the processing is done in the CMY space, and then converted back to the RGB space. This ensures that over-saturation and clipping does not occur, and also preserves the hue. A few results are shown in Fig. 3. The resulting dermoscopic images have enhanced color, their hue is preserved, and they have clearer lesions.

---

**Algorithm 1** Hue-preserving color enhancement

---

- 1: Given skin color images  $R$ ,  $G$ , and  $B$
- 2: Compute intensity  $I$  and histogram-equalized  $H$  images
- 3: Compute CMY color images:  $C$ ,  $M$ , and  $Y$
- 4: **for** every pixel  $p$  **do**
- 5:    $\alpha = \frac{H(p)}{I(p)}$
- 6:   **if**  $\alpha > 1.0$  **then**
- 7:      $\bar{\alpha} = \frac{L-H(x,y)}{L-I(x,y)}$
- 8:      $\tilde{C}(p) \leftarrow \bar{\alpha}C(p)$ ,  $\tilde{M}(p) \leftarrow \bar{\alpha}M(p)$ ,  $\tilde{Y}(p) \leftarrow \bar{\alpha}Y(p)$
- 9:      $\{\tilde{R}(p), \tilde{G}(p), \tilde{B}(p)\} \leftarrow \{\tilde{C}(p), \tilde{M}(p), \tilde{Y}(p)\}$
- 10:   **else**
- 11:      $\tilde{R}(p) \leftarrow \alpha R(p)$ ,  $\tilde{G}(p) \leftarrow \alpha G(p)$ ,  $\tilde{B}(p) \leftarrow \alpha B(p)$
- 12:   **end if**
- 13: **end for**

---

### 2.3. Texture enhancement

We use the multiscale texture enhancement algorithm based on fractional differential masks proposed by Pu et al. (Pu et al., 2010). Fractional differential masks are a generalization of the differential operator to fractional numbers in  
180 the Euclidean space. The Sobel operator and the Gauss operator are special cases of the fractional differential masks (first order and second order respectively).

The  $v$ th order Grumwald-Letnikov based fractional differential mask can be written as

$$\begin{aligned} D_{G-L}^v s(x) &= \frac{d^v}{d(x-a)^v} s(x)|_{G-L} \\ &= \lim_{N \rightarrow \infty} \frac{\frac{(x-a)^{-v}}{N}}{\Gamma(-v)} \sum_{k=0}^{N-1} \frac{\Gamma(k-v)}{\Gamma(k+1)} s\left(x - k \frac{x-a}{N}\right) \end{aligned} \quad (14)$$

185 When  $v < 0$ , the operator becomes a fractional integral operator that acts as a lowpass filter, attenuating high-frequency components. When  $v = -1$ , the operator acts as a first-order integrator. When  $v = 1$ , the operator acts as a first-order differential that linearly attenuates low-frequency components, but cannot enhance texture due to its linear nature that cannot produce results  
190 in smooth areas. When  $0 < v < 1$ , the operator non-linearly attenuates low-frequency components, and does so even in smooth areas (which an integer differential mask cannot do) as well as in areas where the intensity changes. The order  $v$  is a parameter that controls the enhancement of high-frequency information, and as  $v$  increases higher-frequency information is retained to a  
195 greater extent in a non-linear way. This non-linear operation enhances texture throughout the image, in smooth areas as well as in areas where the intensity changes.

Pu et al. suggests 6 different discrete approximations of the analytical expression in equation 14. The second set of fractional differential masks proposed  
200 by Pu et al. is used as it has the best precision, convergence, and texture enhancement properties. Applying the Lagrange 3-point interpolation expression,

the discrete approximation of equation 14 can be written to

$$\frac{d^v}{dx^v} s(x)|_{G-L} \cong \frac{x^{-v} N^v}{\Gamma(-v)} \sum_{k=0}^{N-1} \frac{\Gamma(k-v)}{\Gamma(k+1)} \left[ s_k + \frac{v}{4} (s_{k-1} - s_{k+1}) + \frac{v^2}{8} (s_{k-1} - 2s_k - s_{k+1}) \right] \quad (15)$$

where  $s_k = s(x - \frac{kx}{N})$ . These equations are for the 1-D case. The larger the mask is, the more closely the discrete approximation in equation 15 matches the analytical expression in equation 14. The maximum size of the mask would be when  $k_{\max} = N - 1$ , where  $N$  is the size of the image in each dimension, and the summation is carried out over all pixels along each dimension. When  $k_{\max} = n \leq N - 1$ , the anterior  $n + 2$  approximate backward differences of the fractional partial differentials are extended to 2-D images as

$$\begin{aligned} \frac{d^v}{dx^v} s(x, y)|_{G-L} &\cong \left( \frac{v}{4} + \frac{v^2}{8} \right) s(x+1, y) + \left( 1 - \frac{v^2}{2} - \frac{v^2}{8} \right) s(x, y) + \\ &\quad \frac{1}{\Gamma(-v)} \sum_{k=1}^{n-2} \left[ \frac{\Gamma(k-v+1)}{\Gamma(k+1)!} \left( \frac{v}{4} + \frac{v^2}{8} \right) + \frac{\Gamma(k-v)}{k!} \left( 1 - \frac{v^2}{4} \right) \right. \\ &\quad \left. + \frac{\Gamma(k-v-1)}{\Gamma(k-1)!} \right] s(x-k, y) + \left[ \frac{\Gamma(n-v+1)}{\Gamma(n-1)!\Gamma(-v)} \left( 1 - \frac{v}{4} \right) + \right. \\ &\quad \left. \frac{\Gamma(n-v-2)}{(n-2)!\Gamma(-v)} \left( -\frac{v}{4} + \frac{v^2}{8} \right) \right] s(x-n+1, y) + \\ &\quad \frac{\Gamma(n-v-1)}{\Gamma(n-1)!\Gamma(-v)} \left( -\frac{v}{4} + \frac{v^2}{8} \right) s(x-n, y) \end{aligned} \quad (16)$$

and similarly in a symmetric manner for the partial differential with respect to  $y$ . Hence, the fractional differential masks are sparse matrices, whose elements are the  $n + 2$  coefficients in Equation 16. For example, the coefficients for when  $1 \leq k \leq n - 2$  are

$$C_{s_k} = \frac{1}{\Gamma(-v)} \left[ \frac{\Gamma(k-v+1)}{(k+1)!} \left( \frac{v}{4} + \frac{v^2}{8} \right) + \frac{\Gamma(k-v)}{k!} \left( 1 - \frac{v^2}{4} \right) + \frac{\Gamma(k-v-1)}{(k-1)!} \left( -\frac{v}{4} + \frac{v^2}{8} \right) \right] \quad (17)$$

Similarly,  $C_{s_0}$  is the coefficient for the interest pixel, and is given by

$$C_{s_0} = 1 - \frac{v^2}{2} - \frac{v^2}{8} \quad (18)$$

<sup>215</sup> The other three coefficients at the borders would be

$$C_{s_{-1}} = \frac{v}{4} + \frac{v^2}{8} \quad (19)$$

$$C_{s_{n-1}} = \left[ \frac{\Gamma(n-v+1)}{\Gamma(n-1)!\Gamma(-v)} \left(1 - \frac{v}{4}\right) + \frac{\Gamma(n-v-2)}{(n-2)!\Gamma(-v)} \left(-\frac{v}{4} + \frac{v^2}{8}\right) \right] \quad (20)$$

$$C_{s_n} = \frac{\Gamma(n-v-1)}{\Gamma(n-1)!\Gamma(-v)} \left(-\frac{v}{4} + \frac{v^2}{8}\right) \quad (21)$$

8 convolution kernels, one for each direction, are created for rotation invariance as suggested by Pu et al., with fractional order  $v = 0.5$ , and applied to the intensity of the images, using the coefficients given by equation 16. The arrangement of these kernels are shown in Fig. 5. For the chosen values of  $n = 5$   
<sup>220</sup> and  $v = 0.5$ ,  $C_{s_{-1}} \cong 0.15625$ ,  $C_{s_0} \cong 0.85938$ ,  $C_{s_1} \cong -0.58203$ ,  $C_{s_2} \cong 0.080078$ ,  
 $C_{s_3} \cong -0.052979$ ,  $C_{s_4} \cong -0.030762$ , and  $C_{s_5} \cong 0.0036621$ . Note that  $C_{s_0}$  is not  
at the center of the kernel. The anchor of each convolution operation is located  
at the position of  $C_{s_0}$  in each kernel, named  $W_{x+}$ ,  $W_{x-}$ ,  $W_{y+}$ ,  $W_{y-}$ ,  $W_{LDD}$ ,  
 $W_{LUD}$ ,  $W_{RDD}$ , and  $W_{RUD}$ . The magnitude of the response is found using the  
<sup>225</sup> Euclidean norm. Few results are shown in Fig. 6.

#### 2.4. U-net architecture

A modified architecture of U-net (Ronneberger et al., 2015) is used for this study. The U-net builds upon the fully convolutional network (FCN) architecture (Long et al., 2015). In FCNs, a contracting network is supplemented  
<sup>230</sup> by successive layers whose pooling operations are replaced by upsampling operations. These layers increase the resolution of the output and are called the expanding path. High resolution features from the contracting path are combined, generally by concatenation, with features at the corresponding resolution in the expanding path in order to localize features better.

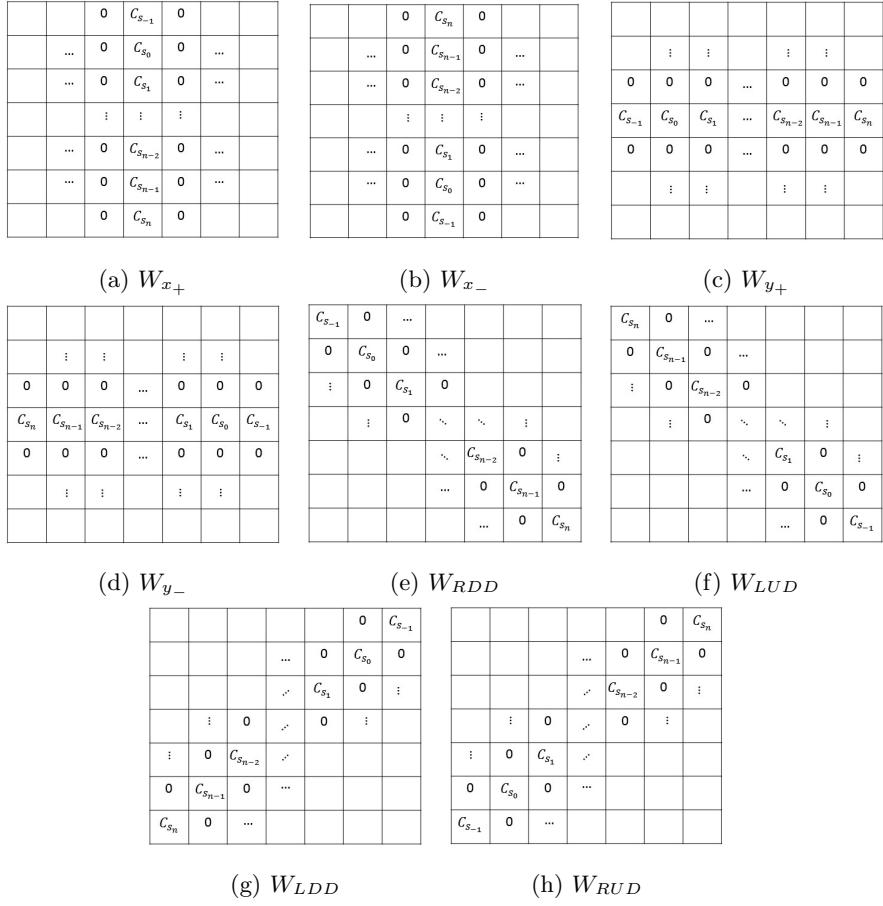


Figure 5: Kernels used for texture enhancement in 8 different directions.

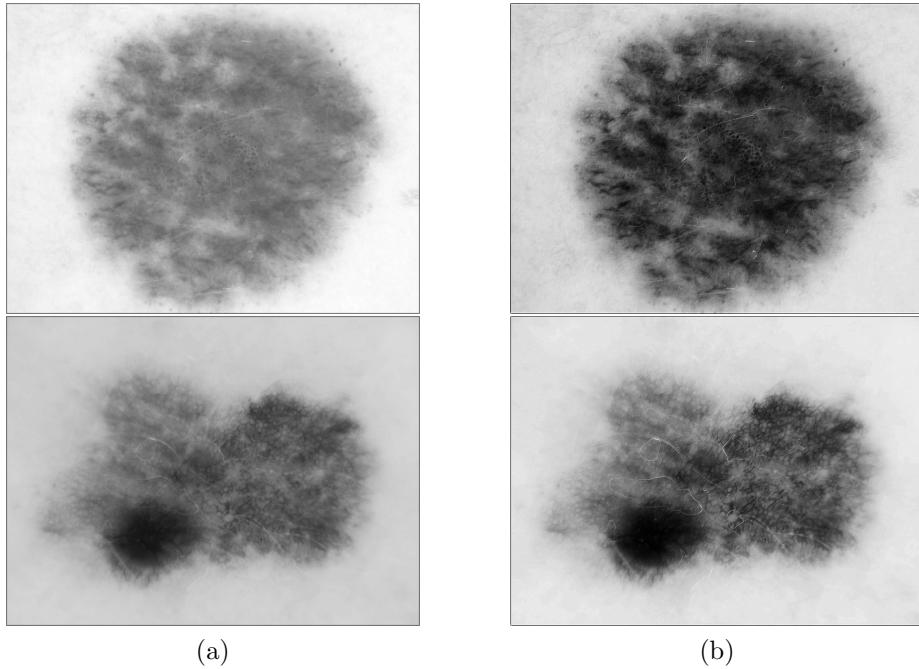


Figure 6: Results of texture enhancement operation (a) Grayscale image (b) Texture enhanced image

In the U-net, the expanding path also has a large number of feature channels, such that the expanding path is almost symmetric to the contracting path. The network does not have any fully connected layers. In the contracting path, there are repeated 3x3 convolutions followed by the rectified linear unit activation (ReLU) function. These are then downsampled by 2x2 max pooling. In the expanding path, transposed convolution (Dumoulin and Visin, 2016) operations (also called backward-strided convolution or up-convolution) are applied repeatedly as the path expands. Feature maps from the contracting path are concatenated (copied) to the feature maps of the corresponding resolution in the expanding path. These are further put through 3x3 convolutions and Re-LUs before reaching the next stage of the expansion. At the last stage, a 1x1 convolution produces the segmentation map.

In the U-net architecture that we propose, the primary difference from the

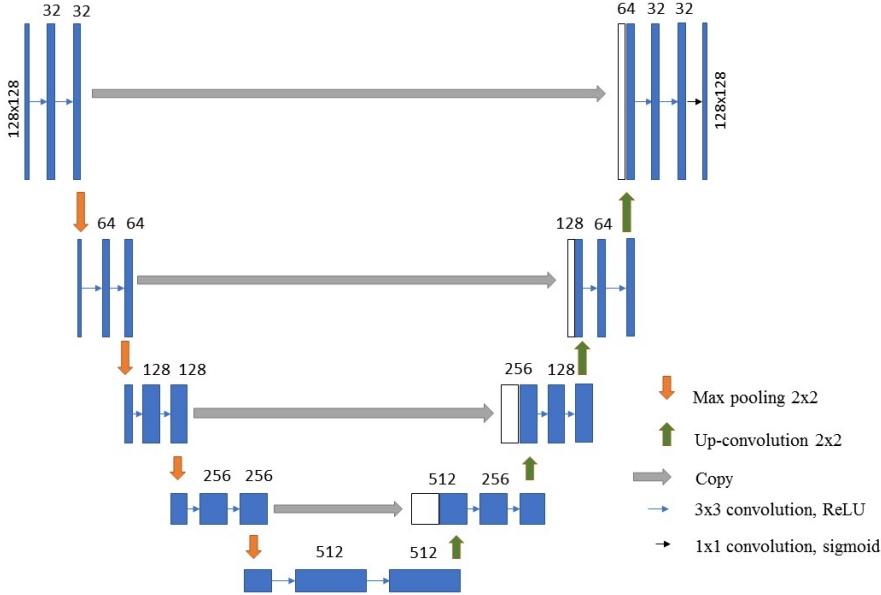


Figure 7: Illustration of the modified U-net architecture. The blue boxes represent the feature maps, and the white boxes represent the copied maps from the contracting path.

original architecture is that the output segmentation map produced is of the same size as the input image, unlike in the original architecture, and the number  
<sup>250</sup> of filters is half of that in the original architecture at each stage, starting from 32 at the first stage. The architecture is shown in Fig. 6. The output map is a single channel image, and are sigmoid probabilities on a 128x128 map. The model takes only about 90 MB of disk space due to the lack of fully connected layers. All input images are resized to 128x128 using bilinear interpolation.  
<sup>255</sup> After the network is trained, the output is converted to a binary labeled image at a threshold of 0.5 applied to logit probabilities.

### 2.5. Post-processing operations

The sigmoid probability map generated by the network does not take into account the connectivity of segmented masks. Consequently, there are holes and  
<sup>260</sup> small blobs in the segmented image. Opening (erosion followed by dilation) and closing (dilation followed by erosion) with a 5x5 square structuring element are

used to remove these. The largest connected contour in the image is found and drawn as the predicted mask. After post-processing, the image is resized to its original shape.

<sup>265</sup> **3. Experiments and results**

*3.1. Skin lesion dataset*

The ISIC 2017 challenge training data was used to train the network. Fig. 2 shows two representative samples of skin lesion images. The dataset consists of 2150 dermoscopic images and 2150 corresponding ground truth masks and the images are of different sizes. ISIC 2017 challenge’s testing data, which has 600 images and corresponding ground truth masks, were used to compare segmentation results with state-of-the-art approaches. The network was trained using all the 2150 images and masks of the 2017 training data, and at the end of each epoch was tested against the 2017 test data, and the best performing weights was saved.

*3.2. Pre-processing and augmentation*

Aggressive on-line data augmentation was performed before contrast enhancement. The images were allowed a random rotation range of  $270^\circ$ , randomly flipped horizontally and vertically, and allowed a random zoom range of 0.8 to 1.2 times the original. The proposed method is to train the network with five channels: three color enhanced channels, a histogram-equalized intensity channel, and a texture-enhanced intensity channel. The results were compared with results from networks trained using just RGB images and networks trained on color, texture and intensity enhanced images separately, all of which underwent the same augmentation operations. Image normalization was performed for the RGB images by subtracting the ImageNet mean from all images. The samples were not normalized for the present method. The code for our method is available at [https://github.com/JoshuaEbenezer/deep\\_segment](https://github.com/JoshuaEbenezer/deep_segment), and is built on a code base made public by the RECOD lab (Menegola et al., 2017).

290    *3.3. U-net training*

The Adam optimizer (Kingma and Ba, 2014) was used for training the U-net with a learning rate of 1e-4, exponential decay rate for the first moment estimate  $\beta_1 = 0.9$  and exponential decay rate for the second moment estimate  $\beta_2 = 0.999$ , and no decay in the learning rate over each update. The loss was defined as the negative of the Dice similarity coefficient. For given sets  $X$  and  $Y$ , the Dice coefficient  $DSC$  is defined as

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|} \quad (22)$$

The Jaccard index, defined as

$$J = \frac{|X \cap Y|}{|X \cup Y|} \quad (23)$$

was monitored as a metric, along with accuracy, sensitivity, and specificity. For 300 the ISIC challenges, the Jaccard index is the only metric used to rank teams. The Jaccard index and the Dice similarity coefficient are related and can be derived from each other. A mini-batch size of 4 was used and the training was run for 200 epochs. Models were saved at checkpoints of best loss on the 2017 test dataset. The Keras 2 library<sup>1</sup> with a Tensorflow backend was used for the 305 training and evaluation and all experiments were run on a 16 GB Nvidia Tesla P100 GPU. Training took less than two hours for the network, including pre-processing and post-processing operations, showing that the proposed method is fast and efficient.

*3.4. Segmentation results*

310    The U-net architecture in our approach receives five image channels: a contrast-enhanced intensity image channel, three hue-preserved color-enhanced R, G, and B channels, and a texture-enhanced intensity image channel. The network was trained using all of the 2017 training data and the best weights

---

<sup>1</sup><https://keras.io>

were saved based on the performance against the 2017 test data at the end of  
 315 each epoch. This network obtained a Jaccard index of 0.7663, which would have placed it first. Segmentations by the network of five representative images are shown in Fig. 3.

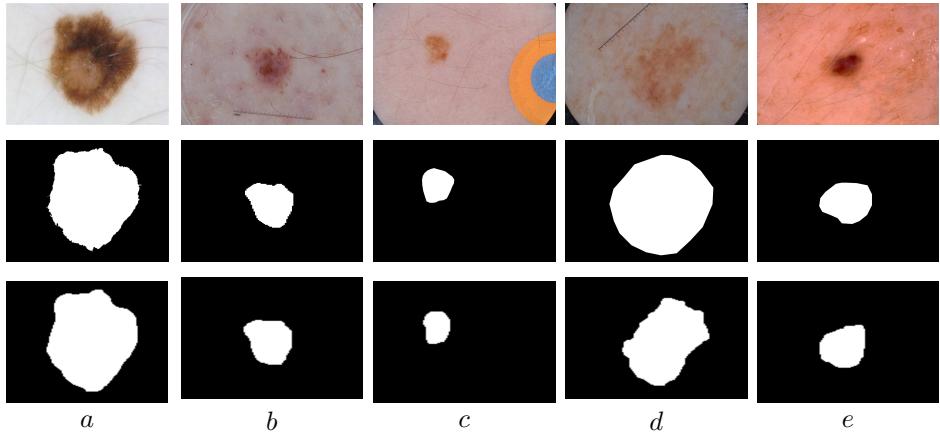


Figure 8: Segmentation results of five skin lesion images *a-e*: top: original color images; middle: ground truth masks; and bottom: segmentations by proposed algorithm

Table 1 shows the results of using different preprocessing techniques and a comparison with previous techniques. The U-Net architecture was fixed for  
 320 comparison with different pre-processing methods but had different numbers of input channels for each. We present the results of the networks whose best weights were saved against 2017 test data at the end of each epoch. It is clearly seen that the proposed method obtains significantly better results than training the network with only RGB data.

We also compared the performances of the present method with previously  
 325 proposed techniques for skin lesion segmentation. The method by Yuan and Lo (Yuan and Lo, 2017), which won the ISIC 2017 challenge, obtained a Jaccard index of 0.765; and the method proposed by Berseth (Berseth, 2017) obtained a Jaccard index of 0.762, and by Bi et al. (Bi et al., 2017) obtained a score of 0.760. The proposed method obtains a higher Jaccard index on the test set than the previously proposed techniques. Only the Jaccard index was used to rank  
 330

Table 1: Results of segmentation on ISIC 2017 test data.

Method	Jaccard index	Dice coef.	Accuracy	Sensitivity	Specificity
RGB channels	0.698	0.791	91.1	70.1	<b>98.6</b>
Enhanced intensity	0.754	0.843	92.7	<b>85.5</b>	93.4
Enhanced color	0.721	0.818	91.7	82.3	95.7
Enhanced texture	0.751	0.841	92.9	84.9	96.1
Proposed method	<b>0.766</b>	<b>0.852</b>	93.2	83.9	97.0
Yuan and Lo (2017)	0.765	0.849	<b>93.4</b>	82.5	97.5
Berseth (2017))	0.762	0.847	93.2	82.0	97.8
Bi et al. (2017)	0.760	0.844	93.4	80.2	98.5

participants in the 2017 ISIC challenge, but we have reported other metrics as well for the sake of completeness.

#### 4. Discussion

We are able to show the results of training our network with the enhanced images are significantly better than training with RGB images. Specifically, there is a significant improvement of 9.7% in Jaccard index over training with RGB data. The enhancement operations that we propose are agnostic to the network being used to infer the segmentation, and hence can be used to boost results for any skin lesion segmentation deep learning network. When combined with the U-Net, it is able to achieve results that are comparable to the state-of-the-art. Even though the dataset has images that have artifacts, the proposed method is able to distinguish between lesions and artifacts quite well (see Fig. 8).

Despite the commendable results, there are shortcomings that can be understood by a careful inspection of Fig. 8. For example, in Fig. 9, the model highlights the exact darker boundary of the lesion, whereas the ground truth (created by manual polygon tracing by an expert) appears to have gone further into an area that cannot be readily distinguished from the skin by the naked eye. This is the case with most predictions that are in error. The model appears

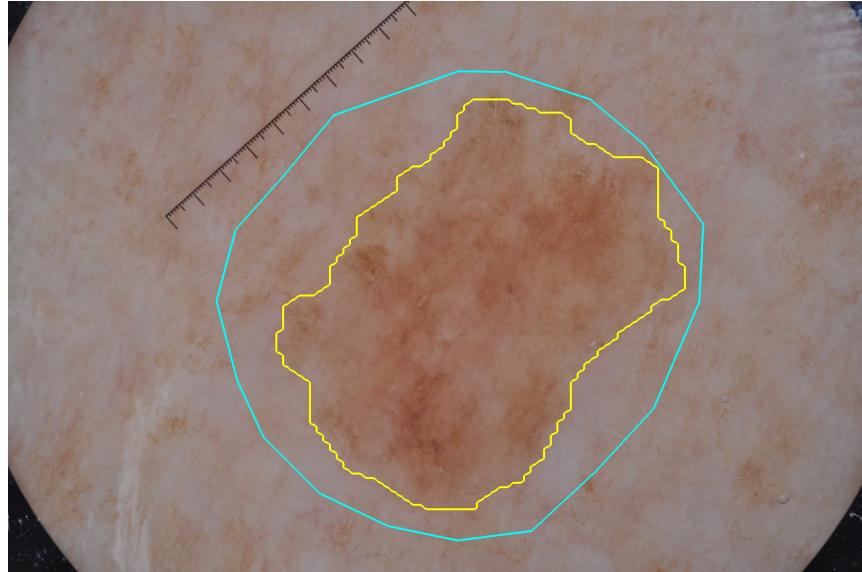


Figure 9: Illustration of a poor prediction. The model prediction is in yellow and the ground truth is in blue.

350 to err on the side of marking a smaller boundary around the discernible regions  
that are clearly distinct from the skin, whereas the ground truth appears to  
have boundaries that extend slightly larger than the distinguishable boundary  
between the skin and lesion.

355 The considerable improvement over training with RGB images should give  
cause for consideration of better methods for pre-processing of dermoscopic im-  
ages before training a network on them. The proposed method can enhance fea-  
tures, suppress noise, and can potentially improve results for any given method  
by its refinement of the input space, but there is scope for further improve-  
ment as well, especially when better techniques for intensity, color, and texture  
360 enhancement are discovered.

## 5. Conclusion

We presented a deep learning image processing framework that achieved state-of-the-art results. We used five channel inputs to train the U-net archi-

tecture: the contrast-enhanced intensity channel, the three hue-preserved color-  
365 enhanced RGB channels, and the texture-enhanced intensity channel; the U-net  
architecture was modified to make both input and output image sizes same with  
different filters at each stage; and a post-processing method that uses morpho-  
logical operations and contour identification was used to arrive at the output  
segmentation. The proposed technique was compared with the results of train-  
370 ing the network with enhanced intensity, enhanced color, and enhanced texture  
separately, as well as results from other state-of-the-art methods. We empiri-  
cally demonstrated that the proposed approach achieves improved segmentation  
performance on ISIC 2017 challenge images as compared to training a network  
with RGB input or other pre-processing methods, and also performs better than  
375 other state-of-the-art methods. The framework is fast, simple, and efficient and  
can be extended to other applications as well.

### **Conflict of interest**

The authors have no potential conflicts of interest to disclose.

### **Acknowledgement**

380 This work is partially supported by AcRF Tier-1 grant RG149/17 by the  
Ministry of Education, Singapore

### **References**

- Abbas, Q., Garcia, I.F., Emre Celebi, M., Ahmad, W., Mushtaq, Q., 2013. A  
perceptually oriented method for contrast enhancement and segmentation of  
385 dermoscopy images. Skin Research and Technology 19, e490–e497.
- American Cancer Society, 2018. American cancer society,  
cancer facts and figures 2018. <https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2018.html>. Accessed: 2018-08-07.

- <sup>390</sup> Andermatt, S., Pezold, S., Cattin, P., 2016. Multi-dimensional gated recurrent units for the segmentation of biomedical 3d-data. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 10008 LNCS, 142–151. doi:10.1007/978-3-319-46976-8\_15.
- <sup>395</sup> Barata, C., Ruela, M., Francisco, M., Mendonça, T., Marques, J.S., 2014. Two systems for the detection of melanomas in dermoscopy images using texture and color features. IEEE Systems Journal 8, 965–979.
- Berseth, M., 2017. Isic 2017-skin lesion analysis towards melanoma detection. arXiv preprint arXiv:1703.00523 .
- <sup>400</sup> Bi, L., Kim, J., Ahn, E., Kumar, A., Fulham, M., Feng, D., 2017. Dermoscopic image segmentation via multi-stage fully convolutional networks. IEEE Trans. Biomed. Eng 64, 2065–2074.
- Celebi, M., Kingravi, H., Iyatomi, H., Aslandogan, Y., Stoecker, W., Moss, R., Malters, J., Grichnik, J., Marghoob, A., Rabinovitz, H., Menzies, S.,  
<sup>405</sup> 2008. Border detection in dermoscopy images using statistical region merging. Skin Research and Technology 14, 347–353. doi:10.1111/j.1600-0846.2008.00301.x.
- Celebi, M.E., Kingravi, H.A., Uddin, B., Iyatomi, H., Aslandogan, Y.A., Stoecker, W.V., Moss, R.H., 2007. A methodological approach to the classification of dermoscopy images. Computerized Medical Imaging and Graphics 31, 362 – 373. doi:<https://doi.org/10.1016/j.compmedimag.2007.01.003>.
- Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2018. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence 40, 834–848.

- Cicek, O., Abdulkadir, A., Lienkamp, S., Brox, T., Ronneberger, O., 2016. 3d u-net: Learning dense volumetric segmentation from sparse annotation. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 9901 LNCS, 424–432. doi:10.1007/978-3-319-46723-8\_49.
- 420 Codella, N.C., Gutman, D., Celebi, M.E., Helba, B., Marchetti, M.A., Dusza, S.W., Kalloo, A., Liopyris, K., Mishra, N., Kittler, H., et al., 2018. Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic), in: Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on, IEEE. pp. 168–172.
- 425 Dumoulin, V., Visin, F., 2016. A guide to convolution arithmetic for deep learning. arXiv preprint arXiv:1603.07285 .
- 430 Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., Thrun, S., 2017. Dermatologist-level classification of skin cancer with deep neural networks. Nature 542, 115.
- Ferris, L.K., Harkes, J.A., Gilbert, B., Winger, D.G., Golubets, K., Akilov, O., Satyanarayanan, M., 2015. Computer-aided classification of melanocytic lesions using dermoscopic images. Journal of the American Academy of Dermatology 73, 769–776.
- 435 Global Burden of Disease Study, 2015. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 19902015: a systematic analysis for the global burden of disease study 2015.
- 440 He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn, in: Computer Vision (ICCV), 2017 IEEE International Conference on, IEEE. pp. 2980–2988.
- Iyatomi, H., Oka, H., Celebi, M., Hashimoto, M., Hagiwara, M., Tanaka, M., Ogawa, K., 2008. An improved internet-based melanoma screening system

- with dermatologist-like tumor area extraction algorithm. Computerized Medical Imaging and Graphics 32, 566 – 579. doi:<https://doi.org/10.1016/j.compmedimag.2008.06.005>.
- Jiang, Z., Zhang, H., Wang, Y., Ko, S.B., 2018. Retinal blood vessel segmentation using fully convolutional network with transfer learning. Computerized Medical Imaging and Graphics 68, 1 – 15. doi:<https://doi.org/10.1016/j.compmedimag.2018.04.005>.
- Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 .
- Lee, C., Lee, C., Kim, C.S., 2013. Contrast enhancement based on layered difference representation of 2d histograms. IEEE transactions on image processing 22, 5372–5384.
- Lee, H., Chen, Y.P.P., 2014. Skin cancer extraction with optimum fuzzy thresholding technique. Applied Intelligence 40, 415–426. doi:[10.1007/s10489-013-0474-0](https://doi.org/10.1007/s10489-013-0474-0).
- Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3431–3440.
- Masood, A., Ali Al-Jumaily, A., 2013. Computer aided diagnostic support system for skin cancer: a review of techniques and algorithms. International journal of biomedical imaging 2013.
- Menegola, A., Tavares, J., Fornaciali, M., Li, L.T., Avila, S., Valle, E., 2017. Recod titans at isic challenge 2017. arXiv preprint arXiv:1703.04819 .
- Naik, S.K., Murthy, C., 2003. Hue-preserving color image enhancement without gamut problem. IEEE Transactions on Image Processing 12, 1591–1598.
- Pu, Y.F., Zhou, J.L., Yuan, X., 2010. Fractional differential mask: a fractional differential-based approach for multiscale texture enhancement. IEEE transactions on image processing 19, 491–511.

- Rajab, M., Woolfson, M., Morgan, S., 2004. Application of region-based segmentation and neural network edge detection to skin lesions. *Computerized Medical Imaging and Graphics* 28, 61–68.
- <sup>475</sup> Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer. pp. 234–241.
- Schmid-Saugeon, P., 2000. Symmetry axis computation for almost-symmetrical and asymmetrical objects: Application to pigmented skin lesions. *Medical Image Analysis* 4, 269–282.
- <sup>480</sup> Silveira, M., Nascimento, J.C., Marques, J.S., Marcal, A.R.S., Mendonca, T., Yamauchi, S., Maeda, J., Rozeira, J., 2009. Comparison of segmentation methods for melanoma diagnosis in dermoscopy images. *IEEE Journal of Selected Topics in Signal Processing* 3, 35–45. doi:10.1109/JSTSP.2008.2011119.
- <sup>485</sup> Strickland, R.N., Kim, C.S., McDonnell, W.F., 1987. Digital color image enhancement based on the saturation component. *Optical Engineering* 26, 267609.
- <sup>490</sup> Swierczynski, P., Papie, B.W., Schnabel, J.A., Macdonald, C., 2018. A level-set approach to joint image segmentation and registration with application to ct lung imaging. *Computerized Medical Imaging and Graphics* 65, 58 – 68. doi:<https://doi.org/10.1016/j.compmedimag.2017.06.003>. advances in Biomedical Image Processing.
- <sup>495</sup> Thomas, B.A., Strickland, R.N., Rodriguez, J.J., 1997. Color image enhancement using spatially adaptive saturation feedback, in: *Image Processing, 1997. Proceedings., International Conference on*, IEEE. pp. 30–33.
- Wong, A., Scharcanski, J., Fieguth, P., 2011. Automatic skin lesion segmentation via iterative stochastic region merging. *IEEE Transactions on Infor-*

mation Technology in Biomedicine 15, 929–936. doi:10.1109/TITB.2011.  
500 2157829.

Xie, Y., Zhang, Z., Sapkota, M., Yang, L., 2016. Spatial clockwork recurrent  
neural network for muscle perimysium segmentation, in: Ourselin, S., Joskow-  
icz, L., Sabuncu, M.R., Unal, G., Wells, W. (Eds.), Medical Image Computing  
and Computer-Assisted Intervention – MICCAI 2016, Springer International  
505 Publishing, Cham. pp. 185–193.

Yoo, J., Rigel, D., 2010. Trends in dermatology: Geographic density of us der-  
matologists. Archives of Dermatology 146, 779. doi:10.1001/archdermatol.  
2010.127.

Yuan, Y., Lo, Y.C., 2017. Improving dermoscopic image segmentation with en-  
510 hanced convolutional-deconvolutional networks. IEEE Journal of Biomedical  
and Health Informatics .

Zhou, H., Li, X., Schaefer, G., Celebi, M.E., Miller, P., 2013. Mean shift based  
gradient vector flow for image segmentation. Computer Vision and Image  
Understanding 117, 1004 – 1016. doi:[https://doi.org/10.1016/j.cviu.  
515 2012.11.015.](https://doi.org/10.1016/j.cviu.2012.11.015)