

Article

Skin Lesion Segmentation in Dermoscopic Images with Combination of YOLO and GrabCut Algorithm

Halil Murat Ünver and Enes Ayan *

Department of Computer Engineering, Kırıkkale University, 71451 Kırıkkale, Turkey

* Correspondence: enesayan@kku.edu.tr

Received: 30 May 2019; Accepted: 8 July 2019; Published: 10 July 2019



Abstract: Skin lesion segmentation has a critical role in the early and accurate diagnosis of skin cancer by computerized systems. However, automatic segmentation of skin lesions in dermoscopic images is a challenging task owing to difficulties including artifacts (hairs, gel bubbles, ruler markers), indistinct boundaries, low contrast and varying sizes and shapes of the lesion images. This paper proposes a novel and effective pipeline for skin lesion segmentation in dermoscopic images combining a deep convolutional neural network named as You Only Look Once (YOLO) and the GrabCut algorithm. This method performs lesion segmentation using a dermoscopic image in four steps: 1. Removal of hairs on the lesion, 2. Detection of the lesion location, 3. Segmentation of the lesion area from the background, 4. Post-processing with morphological operators. The method was evaluated on two publicly well-known datasets, that is the PH2 and the ISBI 2017 (Skin Lesion Analysis Towards Melanoma Detection Challenge Dataset). The proposed pipeline model has achieved a 90% sensitivity rate on the ISBI 2017 dataset, outperforming other deep learning-based methods. The method also obtained close results according to the results obtained from other methods in the literature in terms of metrics of accuracy, specificity, Dice coefficient, and Jaccard index.

Keywords: skin cancer; skin lesion segmentation; melanoma; convolutional neural networks; Yolo; GrabCut

1. Introduction

Skin cancer is one of the most widespread cancer types in over the world [1]. There are different types of skin cancer such as basal cell carcinoma, melanoma, intraepithelial carcinoma, squamous cell carcinoma, etc. [2]. The human skin consists of three tissues called dermis, epidermis, and hypodermis. The epidermis has melanocytes which can produce melanin at a highly unusual rate under some conditions. For instance, long term exposure to the strong ultraviolet radiation from sunshine causes melanin production. The unusual growth of melanocytes causes melanoma, which is a lethal type of skin cancer [3]. Considering the American Cancer Society's annual report for 2019, it is estimated that there will approximately be 96,480 new cases of melanoma and 7230 people will die from the disease [4]. Melanoma is also reported as the most lethal skin cancer with a mortality rate of 1.62% among other skin cancers [5]. Early diagnosis of the melanoma is very important in terms of treatment. If melanoma is diagnosed in early stages, the five-year relative survival rate is 92% [6]. However, visual similarities between benign and malign skin lesions are the main challenge of detecting melanoma. For this reason, diagnosing melanoma could be difficult for even a well-trained specialist. It is a very challenging task to determine the type of lesions with the naked eye. Hence, over the years, different imaging methods have been used and one of them is Dermoscopy. Dermoscopy is a non-invasive imaging technique that allows the visualization of the skin surface by the light magnifying device and immersion fluid [7]. It is one of the most widely used imaging techniques in dermatology and has increased the diagnosis performance of malignant cases by 50% according to the experience of the

observer [8,9]. However, the use of human vision alone for the detection of melanoma in dermoscopic images may be inaccurate, subjective, or irreproducible because it depends on the dermatologist's experience [10]. Diagnostic accuracy of melanoma from the dermoscopic images by an inexperienced specialist is between 75% to 84% [8]. In order to overcome all these difficulties encountered in the diagnosis of melanoma, computer-aided diagnosis (CAD) systems are needed to assist the experts in the diagnosis process. There are four steps in CAD systems for identifying a lesion as melanoma: preprocessing, segmentation, feature extraction, and classification. For a robust identification of melanoma, lesion segmentation is a fundamental step in CAD systems. However, this segmentation step is a troublesome process due to the large differences in color, texture, position, and size of skin lesions in dermoscopic images. Besides, the low contrast of the image prevents the differentiation of the adjacent tissues. In addition, extra factors such as air bubbles, hair, ebony frames, ruler marks, blood vessels, and color illumination cause extra difficulties to the lesion segmentation. Figure 1 shows several dermoscopic image samples from the data set with different artifacts on them.



Figure 1. Various artifact examples in dermoscopic images. First row: hair artifact, ink marker artifact, dark corner artifact. Second row: ruler marker artifact, gel bubble artifact, and illumination artifact from left to right, respectively.

Various methods have been proposed for the segmentation of skin lesions. Particularly in recent years, convolutional neural networks (CNNs), which is one of the deep learning methods, has achieved very successful results in segmentation of skin lesions [11]. However, CNNs accepts low-resolution images for decreasing the number of calculations and parameters in the network [11]. This situation may lead to the loss of some important features in the image. The motivation of this paper is to develop a resolution independent method for skin lesion segmentation in dermoscopic images. For this purpose, a pipeline was proposed combining the GrabCut algorithm [12] and the power of deep learning. The suggested pipeline consists of four steps. The first step is removing hairs on the lesion; the second step is detecting the location of the lesion in the image and drawing a bounding box around the skin lesion; the third step is using the bounding box in the GrabCut algorithm for lesion segmentation; and the last step is removing artifacts in the segmented area using morphological image processing operators. To the best of our knowledge, there is no study using deep convolutional neural networks and the GrabCut algorithm together for lesion segmentation. Our contributions in this paper are:

- Yolov3 [13], which is a deep convolutional neural network that has been trained for the detection of lesion location in the image and it has been used to automate segmentation algorithm GrabCut, which is also known as a semi-automatic algorithm, for segmenting skin lesions for the first time in literature.
- An alternative real-time skin lesion segmentation method has been proposed with the incorporation of different techniques into a pipeline.
- Deep learning-based segmentation methods process fixed-size images and produce low-resolution segmentation images. The present method processes dimension-independent images and produces high-resolutions segmentation results.

The rest of this paper is organized as follows: Section 2 deals with the related works in literature. Section 3 gives detailed information about the dataset and methods used in the study. Section 4 focuses on the result of the introduced method by comparing it with the other methods in the literature. Section 5 is the discussion part. Section 6 is the conclusions.

2. Related Works

Specific and prominent features of lesion images play a critical role in the classification of melanoma. These features can only be obtained by proper segmentation of the skin lesion from surrounding tissue. Segmenting the lesion from the surrounding normal tissue and extracting more representative features are essential for a robust and effective diagnosis [14,15]. There are several segmentation methods developed to segment skin lesions automatically or semi-automatically [16–19]. These segmentation methods can be grouped in to five. Histogram thresholding methods try to identify a threshold value for the segmentation of lesion from the surrounding tissue [20–22]. Unsupervised clustering approaches use the color space properties of RGB dermoscopic images to obtain homogenous regions [23–29]. Edge-based and region-based methods take advantage of the edge operator and different algorithms such as region splitting or merging [28,30,31]. Active contour methods utilize metaheuristic algorithms, genetic algorithms and snake algorithms, etc., for segmentation of lesion area [24,32–34]. The last group is the supervised segmentation methods. These methods segment the skin lesion by training the recognizers, such support vector machines (SVMs), decision trees (DTs), and artificial neural networks (ANNs) [24,35]. More detailed information on these methods can be found in the most comprehensive and current reviews of segmentation techniques used in skin lesions [17,36,37]. All these techniques use low level features that rely on pixel level features. Therefore, these classical segmentation techniques are unable to achieve satisfactory results and cannot overcome difficulties such as fuzzy lesion boundaries, hair artifacts, low contrast, and other artifacts. Nowadays, deep learning-based methods especially CNNs have obtained significant success in image classification, object detection, and segmentation problems [13,38,39]. The main reason behind the success of CNNs is their capability of hierarchical feature learning and extracting more high level and robust features from the raw image data. There are different types of CNN architectures for different purposes such as classification, segmentation, object detection and localization [13,40,41]. In addition to the natural image classification, CNNs also achieved great success in diversified medical problems such as detection of mitosis in histology images [42], brain tumor segmentation in MR images [43], breast cancer detection in mammography images etc. [44]. A detailed review is presented by Litjens et al. [45]. Also, CNNs achieved state-of-the-art results in semantic segmentation. Various deep CNN architectures have been proposed for semantic segmentation such as Fully Convolutional Neural Network (FCN) [44], U-Net [41], SegNet [46], and DeepLab [47]. More about these semantic segmentation methods can be found in a detailed review [48]. New developments in CNN architectures with the capability of semantic segmentation have been used in segmentation of skin lesions by researchers in recent years. For instance, in 2017, Yu et al. presented an end-to-end deep network consisting of two stages called segmentation and classification [49]. They developed a fully convolutional residual network (FCRN) utilizing the power of deep residual networks and the took second place with the accuracy of 94.9% in the segmentation category of the International Symposium on Biomedical Imaging (ISBI) 2016 Challenge [50]. Another type of a deep residual network (DRN) introduced by the same team took first place with the accuracy of 85.5% in the classification category of the same challenge. In another study, Yuan et al., introduced a skin lesion segmentation technique by utilizing FCN [51]. They improved the FCN model by using an unusual loss function named as Jaccard distance. In this way, they solved the imbalance problem between the surrounding skin and lesion pixels. Evaluated on two publicly well-known datasets, the ISBI 2016 and the PH2, it achieved an accuracy of 95.5% and 93.7% respectively. Bi et al. developed a multistage FCN and a parallel integration (PI) method to segment skin lesions in dermoscopic images [10]. The PI method combined with the FCN helps further improve the boundaries of the segmented skin lesions. It was evaluated on two publicly available

datasets, the ISBI 2016 and the PH2 [52], attaining 95.51% and 94.24% of accuracy rates, 91.18% and 90.66% of Dice coefficient indices respectively. In another study in 2017, Goya et al. introduced a deep network for multi-class semantic skin lesion segmentation by means of FCN. Their deep FCN architecture succeeded in segmentation of three class of skin lesions including melanoma, benign nevi, and seborrheic keratoses. They used the ISBI 2017 Challenge dataset for evaluation of the method. This deep FCN architecture attained Dice coefficient indices of 55.7%, 65.3%, and 78.5%, for and seborrheic keratosis, melanoma, and benign lesions, respectively [53]. Lin et al. compared the performance of two skin lesion segmentation approaches. One of them is deep convolutional neural network-based U-Net and another one is C-Means Clustering method [54]. This comparison tested on the ISBI 2017 Challenge dataset [50]. The U-Net achieved 77% Dice coefficient indices while clustering method remained at 61%. The results show that U-net method significantly outperformed the clustering method. Yuan et al. presented a deep neural network architecture consisting of convolutional and deconvolutional layers (CDNN) for skin lesion segmentation in 2017 [55]. They trained their model with the ISBI 2017 dataset using dermoscopic images different color spaces. Their CDNN architecture took first place in the ISBI 2017 Challenge with the 76.5% Jaccard index. In 2018, Al-Masni et al. proposed a novel skin lesion segmentation approach called a full resolution convolutional network (FrCN) [56]. The advantage of this model is to eliminate subsampling layers and use full resolution input in the architecture during the training. In this way, the desired specific features could be obtained from the input image easily. They evaluated their deep model on two publicly available datasets, the PH2 and the ISBI 2017. Test results for sensitivity, specificity, accuracy on the ISBI 2017 dataset were 85.40%, 96.69%, 94.03%, respectively, while on the PH2 dataset, they were 93.72%, 95.65%, 95.08%, respectively. In 2018, Hang Li et al. presented a deep model called dense deconvolutional network (DNN) for the segmentation of skin lesions. Their model consists of dense deconvolutional layers (DDL), chained residual pooling (CRP) and hierarchical supervision (HS) [57]. They trained DDL for maintaining the same resolutions of input and output images without prior knowledge or complicated post-processing procedures. They used CRP for extracting rich contextual information by combining local and global contextual feature combination and used HS to serve as a loss helper as well as to improve the prediction mask. They used the ISBI 2017 dataset for evaluation of their model and obtained the results 0.866%, 0.765%, 0.939%, for Dice coefficient, Jaccard index, accuracy, respectively. In 2018, Peng et al. used a segmentation architecture based on adversarial networks. They utilized generative adversarial network (GAN) to assist segmentation of skin lesions [58]. They used a U-net based network as generator and a CNN network as discriminator to discriminate the ground truth and generated mask. They evaluated their model on the ISBI 2016 dataset and achieved an average segmentation accuracy rate of 0.97% and a 0.94% the Dice coefficient rate. Recently, in 2019, Yuan et al. recommended a segmentation method [59], an improved version of their last study [51]. They developed a deeper network architecture with 29 layers and used small kernel filters for attaining more detailed features and increasing the discrimination capacity of their architecture. They evaluated their method on the ISBI 2017 dataset and achieved 0.76% Jaccard index rate.

3. Materials and Methods

3.1. Datasets

This study was evaluated on two publicly available datasets, the ISBI 2017 and the PH2. The ISBI 2017 dataset was created for a challenge called Skin Lesion Analysis Towards Melanoma Detection. This data set is a small piece of the International Skin Imaging Collaboration (ISIC) archive which is consisting of 23906 dermoscopic images [50] and is available publicly for researchers [60]. The dataset consists of 8-bit RGB dermoscopic images with resolutions varying between 540×722 and 4499×6748 . The images in the dataset have been labelled as seborrheic keratosis, melanoma and benign by specialist. The dataset includes 2000 images for training, 150 images for validation and 600 images for testing. The other dataset the PH2 was provided by a research group of the University of Porto, collected from

dermatology service of Hospital Pedro Hispano, Portugal [52]. The PH2 dataset consists of 80 atypical nevi, 80 common nevi, and 40 melanoma cases. The total size of the dataset is 200 lesion images. Unlike the ISBI 2017 dataset, the PH2 dataset was acquired under the same conditions. All images are 8-bit RGB images with 768×560 pixels resolution and were taken by using a lens with a magnification of $20\times$. Also, both datasets contain lesion images with their segmentation boundaries annotated by an expert dermatologist. Figure 2 shows some samples from the ISBI 2017 and the PH2 datasets as lesion images and their ground truths. Also, Table 1 summarizes the class and data distribution of images in both datasets.

Table 1. Distribution of data and labels for training, validation, and test in the ISBI 2017 and the PH2 datasets.

Dataset.	Training Data				Validation Data				Test Data				Total	
	Label	B	M	SK	Total	B	M	SK	Total	B	M	SK	AT	
ISBI 2017		1375	374	254	2000	78	30	42	150	393	197	90	*	600 2750
PH2		*	*	*	*	*	*	*	*	80	40	*	80	200 200
Total					2000				150				800	2950

B-Benign, M-Melanoma, SK-Seborrheic keratosis, AT-Atypical nevus, * there is no data in this field.

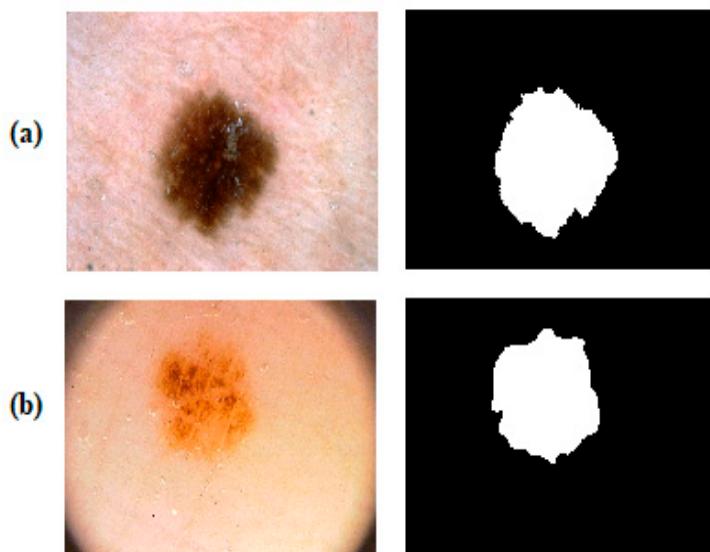


Figure 2. Dermoscopic skin lesion images and their ground truths from datasets. (a) a dermoscopic image sample and its binary ground truth from the ISBI 2017 dataset, (b) a dermoscopic image sample and its binary ground truth from the PH2 dataset.

3.2. Prepossessing and Labelling

This study proposes a real-time skin lesion segmentation method. The proposed method consists of two major and two complement steps. One of the major steps is the detection of the lesion location in the image and the second is use of this information of location for the segmentation of the lesion. The complementary steps are preprocessing and postprocessing operations. Figure 3 illustrates a flow chart of the method.

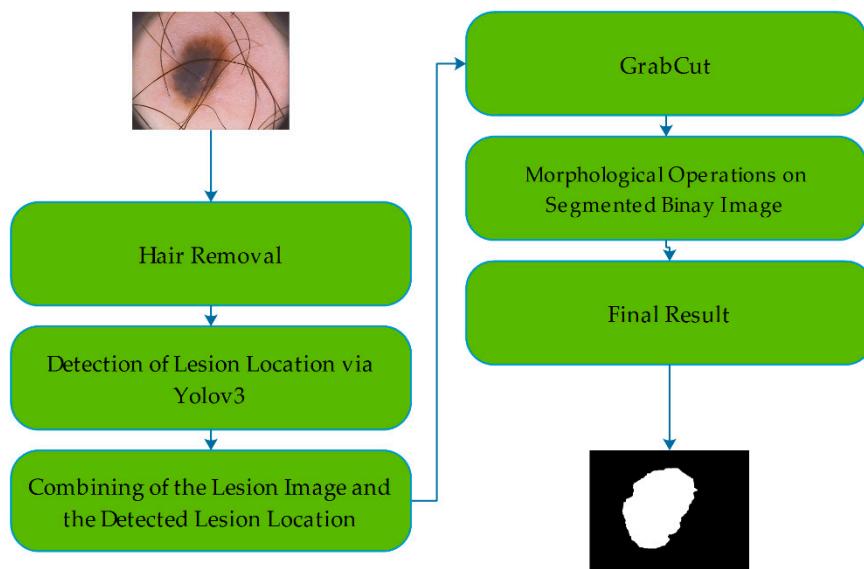


Figure 3. Flow chart of the proposed method including all steps.

All data in the training set were resized to the 512×512 resolution before starting Yolov3 training. Next, each image in the training set was labelled according to the training needs of Yolov3. Yolov3 needs some information about the images as well as images during the training. This information includes middle point coordinates (x, y), width (w) and height (h) values of the bounding box and its class definition of the object to be detected. For this purpose, a bounding box was drawn around the skin lesion using a python script. The upper left corner coordinates (x_1, y_1) and the bottom right corner coordinates (x_2, y_2) of the bounding box were used for the determination of the x and y coordinates, that is the midpoint, and the height (h), width (w) of the bounding box. Detailed calculations are presented in Figure 4.

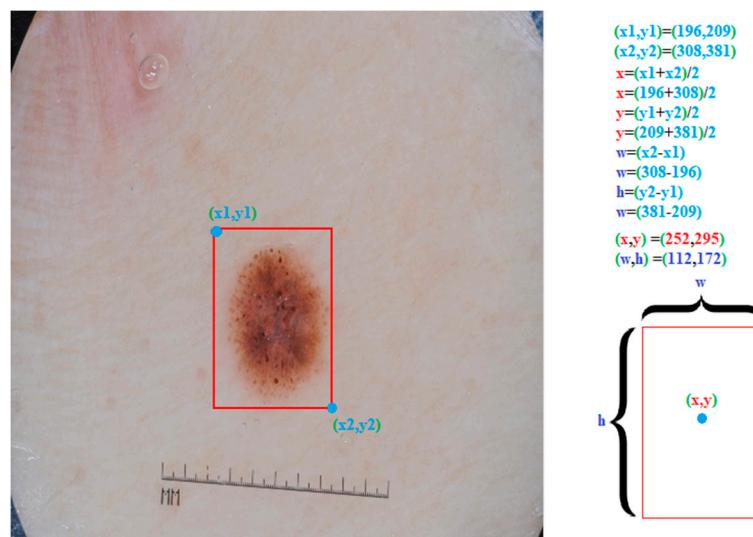


Figure 4. Labeling a skin lesion image for the training of Yolov3. The upper left corner coordinates (x_1, y_1) and the bottom right corner coordinates (x_2, y_2) of bounding box were used for determining of the bounding box x , y , w , and h values.

Also, hairs were removed from all the images during the testing phase by using the DullRazor [61] algorithm for more accurate detection and segmentation. This algorithm removes the hairs over the lesion in three steps. In the first step, it identifies the hair locations by using a grayscale morphological

closing operation. At the second step, it verifies the hair locations by looking at the length and thickness of the detected shapes and then, the verified pixels are replaced by using the bilinear interpolation method. In the final step, it smooths the replaced pixels with the help of an adaptive median filter. Figure 5 demonstrates hair removal results.

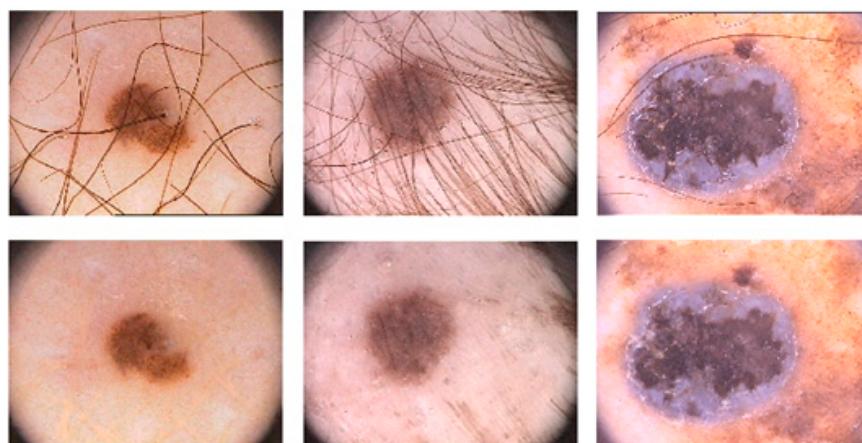


Figure 5. Hair removal results in skin lesion images by using the DullRazor algorithm.

3.3. Yolo Architecture

Classification algorithms try to find the presence of a previously determined object in the image while the object detection algorithms try to describe a bounding box around the object to locate it within the image. There are different object detection and localization algorithms based on deep learning. These algorithms can be divided into two groups. The first group of algorithms consists of two stages. In the first stage, a certain number of potential bounding boxes are created on the image then the CNN based classifiers are run to detect objects in these previously defined boxes. After the classification process, some improvements are made by a post-processing step on the detected bounding boxes such as refining bounding boxes, eliminating duplicate detections, and reordering boxes according to other objects in the scene determined. These complex processes are rather slow, and it is very difficult to optimize each singular component due to the separate training necessary for each one of the components. The most common examples of these algorithms are the region-based convolutional neural network (RCNN) [62] and its more advanced versions Fast-RCNN [63], and Faster-RCNN [64]. The second group algorithms are based on the regression problem. Instead of selecting attractive parts within the image, they try to predict the bounding boxes and classes in the whole image at a single run.

Yolo [65] is one of the best-known examples of these group of algorithms and one of the most powerful and fastest object detection algorithms by using deep learning techniques. The first version of Yolo was presented in 2016 by Redmon et al. [65]. It can detect and classify multiple objects from images in real-time at 45 frames per second. Unlike the other techniques that send to multiple patches of images to the classifiers [62], it sends the whole image to single CNN. The working principle of Yolo is quite simple. A single CNN predicts multiple bounding boxes along with all the class possibilities at the same time (see Figure 6). Therefore, Yolo is more capable of learning generalizable representations of objects than the alternatives. Yolo transforms the detection problem into a regression problem. The predictions are represented as $(S \times S) * B * (5 + C)$ tensor. To explain briefly the underlying idea of Yolo, one must start with the understanding of its prediction tensor. Yolo splits the input image into a non-overlapped $(S \times S)$ grid cells. If the midpoint of an object falls in to a grid cell, that very grid cell oversees the detection of this object. Each grid cell is responsible for the prediction of B possible bounding boxes and the confidence scores for them [65]. The confidence score is the expression of the existence or absence of any object inside in the bounding box.

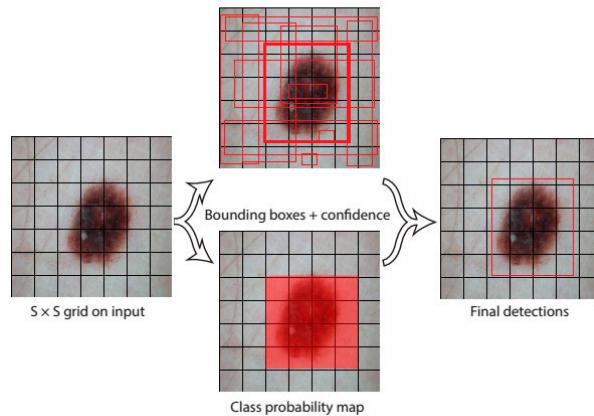


Figure 6. A simple representation of how You Only Look Once (Yolo) detects skin lesion location. Firstly, it divides the image into the ($S \times S$) grids, each grid cell is responsible for creating B possible bounding boxes with their confidence score and C class probabilities.

The confidence is the multiplication of the probability of the existing object and the percentage of the intersection over union (IOU) [64] (see Figure 7) as in following the formula:

$$\text{Confidence} = \Pr(\text{Object}) * \text{IOU}_{\text{predicted}}^{\text{ground truth}} \quad (1)$$

If there is no object in the related cell, the confidence score must be zero. Otherwise, the confidence score is equal to the IOU. Each bounding box has 5 variables such as x , y , w , h , and a confidence score. x and y indicate the coordinates of the central point of the bounding box while w and h represent the width and height values (see Figure 8). Also, there is an additional C variable for the class scores. Every grid cell predicts C conditional class probabilities $\Pr(\text{Class}_i | \text{Object})$ [65]. Yolo calculates class-specific confidence scores for each box by multiplying the conditional class probabilities and individual box confidence predictions at the test time, as shown in formula 2. These scores demonstrate the likelihood of that class emerging in the box and how well the predicted box suits the object.

$$\Pr(\text{Class}_i | \text{Object}) * \Pr(\text{Object}) * \text{IOU}_{\text{predicted}}^{\text{ground truth}} = \Pr(\text{Class}_i) * \text{IOU}_{\text{predicted}}^{\text{ground truth}} \quad (2)$$

The architecture of Yolo was inspired by GoogLeNet [66]. It consists of 24 convolutional layers for feature extraction and two fully connected layers to predict the output of probabilities and coordinates. While Yolo uses a linear activation function for the last layer, the following leaky rectified linear activation (LReLU) function is used for the other layers as seen in the following formula.

$$\beta(x) = \begin{cases} x, & \text{if } x > 0 \\ 0.1 * x, & \text{otherwise} \end{cases} \quad (3)$$

Also, Yolo utilizes sum-squared error as the loss function. The detailed loss function of Yolo is as follows:

$$\begin{aligned} \text{loss} = & \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{\text{obj}} \left[(x_i - \hat{x})^2 + (y_i - \hat{y}_i)^2 \right] \\ & + \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{\text{obj}} \left[(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2 \right] \\ & + \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{\text{obj}} (C_i - \hat{C}_i)^2 + \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{\text{noobj}} (C_i - \hat{C}_i)^2 \\ & + \sum_{i=0}^{S^2} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2 \end{aligned} \quad (4)$$

where x_i and y_i represent the coordinates of the central point of the bounding box while w_i and h_i symbolize width and height. The C_i is the confidence score and $p_i(c)$ classification loss. The λ values are the constants to be used to increase the loss of bounding box coordinate predictions and to decrease the loss of the confidence predictions from boxes that contain no objects. These values are set as $\lambda_{coord} = 5$ and $\lambda_{noobj} = 0.5$. Lastly 1_i^{obj} indicates appearance of an object appears in the cell i and 1_{ij}^{obj} indicates that the j th bounding box predictor in cell i is responsible for that prediction. The first two layers compute the localization loss, while the third layer calculates the confidence loss and the final layer determines the classification loss in the loss formula of Yolo [65].

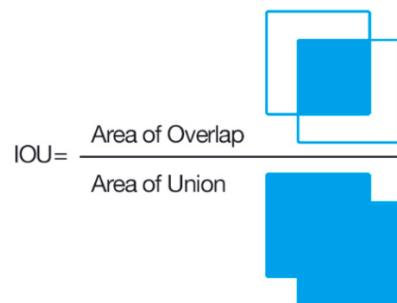


Figure 7. Visual representation of intersection over union (IOU), which is a metric used for measuring object detection performance.

The Yolo is extremely fast when compared to other object detection and classification algorithms since it looks at the image once and does not require complex processes. Also, it makes more accurate predictions as it scans the entire image.

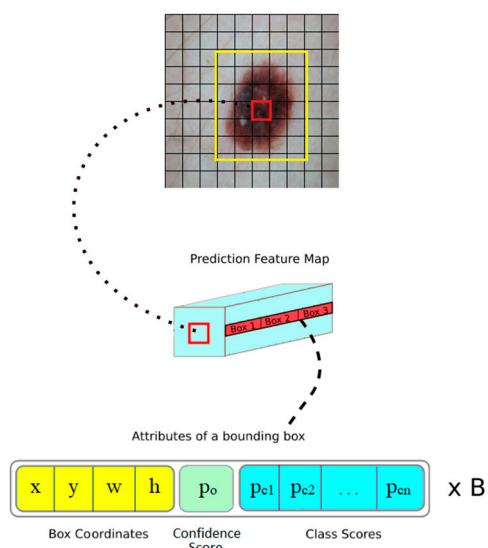


Figure 8. The yellow frame is a bounding box created by the red grid that is responsible for the detection of the lesion. It creates a certain number of bounding boxes. Every bounding box has such parameters as x and y , central coordinates, of the bounding box, w and h , width and height of the bounding box, p_0 confidence score and p_{cn} class probability scores.

This study employed Yolov3 [13] model, the latest and an improved version of previous Yolo networks [65,67]. In line with recent advances in object detection algorithms, few but effective developments in the previous Yolo versions have given way to Yolov3. The major changes in Yolov3 are listed below:

- Yolov3 uses logistic regression as loss function while predicting the confidence scores of bounding boxes.

- Yolov3 utilizes multiple independent logistic classifiers instead of softmax function for the prediction of the class confidence probabilities. This improvement is very important when there are multiple objects in the image.
- Yolov3 benefits from a powerful feature extractor network named DarkNet-53 with 53 convolutional layers and residual blocks. In addition to the DarkNet-53, 53 more layers have been added for the detection task, thus making a 106 layered fully convolutional architecture. Figure 9 illustrates DarkNet-53 architecture.
- Yolov3 employs strided convolution for down sampling instead of max-pooling.
- Yolov3 is made up several convolutional layers in addition to the base feature extractor layers, enhancing its capability of multiscale predictions at three different sizes. This allows the system to make more accurate detections on small objects in the image.
- The last improvement in Yolov3 is cross-layer connections between the prediction layers. The feature maps obtained from the up-sampling operation have been combined with the feature maps of with the previous layers by using concatenation operation. This combination provided more accurate detection performance on small objects.

Type	Filters	Size	Output
1x	Convolutional	32	3 × 3
	Convolutional	64	3 × 3 / 2
	Convolutional	32	1 × 1
2x	Convolutional	64	3 × 3
	Residual		128 × 128
	Convolutional	128	3 × 3 / 2
8x	Convolutional	64	1 × 1
	Convolutional	128	3 × 3
	Residual		64 × 64
8x	Convolutional	256	3 × 3 / 2
	Convolutional	128	1 × 1
	Convolutional	256	3 × 3
4x	Residual		32 × 32
	Convolutional	512	3 × 3 / 2
	Convolutional	256	1 × 1
4x	Convolutional	512	3 × 3
	Residual		16 × 16
	Convolutional	1024	3 × 3 / 2
4x	Convolutional	512	1 × 1
	Convolutional	1024	3 × 3
	Residual		8 × 8
		Avgpool	Global
		Connected	1000
		Softmax	

Figure 9. DarkNet-53 architecture [13].

3.4. GrabCut Algorithm

GrabCut is an iterative semiautomatic image segmentation technique [12]. In this technique, the image to be segmented is represented by a graph. This graph is built by using a minimum cost reduction function to produce the best segmentation of the image. The created graph nodes consist of the image pixels. In other words, each pixel in the image is symbolized by a node in the graph. In addition to these nodes, two extra nodes called sink and source, are added to the graph. Each pixel in the image, namely the nodes in graph, is connected to either of these two nodes. The source node represents the connection point of the foreground pixels while the sink node symbolizes the connection point of the background pixels. A cost function is used for the definition of the edge weights of the graph depending on the region and boundary information in the image. A Min-Cut/Max-Flow technique is used to segment the graph. The GrabCut technique uses Gaussian Mixture Models (GMMs) [68] obtaining the region information by using the color knowledge in the image. The mathematical expression of the algorithm is as follows.

Given an RGB color image as $I, p = (p_1, p_2, \dots, p_N)$ of N pixels where $p_i = (R_i, G_i, B_i), i \in [1, \dots, N]$ in the RGB color space. The segmentation is described as an array as in $s = (s_1, s_2, \dots, s_N), s_i \in \{0, 1\}$ and a label is assigned for each pixel of image describing its relation to the background or foreground.

At the beginning of the algorithm, the user defines a rectangle (R) semi-automatic including the area to be segmented. After the definition of the R by the user, the image is divided into three regions called R_B , R_F , and R_U , representing the initial values of the background, the foreground, and the uncertain pixels. The pixels outside of the R are considered to be R_B while the pixels inside the rectangle are taken as R_U . GrabCut finds out whether the R_U pixels belong to the background or the foreground. For this purpose, it utilizes the color information provided by GMMs. A full covariance GMMs of C components are defined for foreground pixels ($s_i = 1$), and another one for background pixels ($s_i = 0$), parametrized as the following formula:

$$\theta = \{\pi(s, c), \omega(s, c), \Sigma(s, c), s \in \{0, 1\}, c = 1, \dots, C\} \quad (5)$$

where π indicates the weights, ω represents the means of the GMMs and Σ the covariance matrices of the model. The array $c = (c_1, \dots, c_i, \dots, c_N)$, $c_i \in (1, \dots, C)$, $i \in [1, \dots, N]$ is also taken into account representing the component of the foreground or background GMMs (according to s_i) the pixel p_i belongs to. The energy function revealed for segmentation is as follows:

$$E(s, c, \theta, p) = U(s, c, \theta, p) + V(s, p) \quad (6)$$

where U is the likelihood potential, based on the probability distributions $pr(\cdot)$ of the GMMs:

$$U(s, c, \theta, p) = \sum_i -\log pr(p_i | s_i, c_i, \theta) - \log \pi(s_i, c_i) \quad (7)$$

and V is a regularizing prior assuming that segmented regions should be coherent in terms of color, considering a neighborhood C around each pixel.

$$V(s, p) = \gamma \sum_{\{m,n\} \in C} [s_n \neq s_m] \exp(-\beta \|p_m - p_n\|^2) \quad (8)$$

This energy minimization scheme applies to the image with a given initial rectangle. Final segmentation obtained using a minimum cut operation is presented in Figure 10.

Summary of the GrabCut algorithm:

1. The initial step starts with a rectangle surrounding the object of interest that is drawn by the user manually. This step gives information about background and foreground of the interested area. The pixels inside of the rectangle considered as unknown while the pixels outside of the rectangle are considered as background. Based on this information, the algorithm creates a model to determine whether the unknown pixels belong to the foreground or background.
2. An initial segmentation model is created in which the unknown pixels are regarded as the foreground class and all the other except are seen as the background.
3. The initial background and foreground classes are created by using Gaussian Mixture Models, by creating C piece GMM components for two regions.
4. Each pixel in the background class is designated to the most probable Gaussian component in the background GMM. The same process is performed for the foreground pixels designated to the most probable foreground Gaussian component.
5. New GMM are attained by using the sets of pixels created in the previous step.
6. A graph with n nodes is built and the weight values between the connections are determined. After that, a minimum cut algorithm is used to determine foreground and background pixels.
7. Steps of 4–6 are repeated until attaining the final segmentation result.

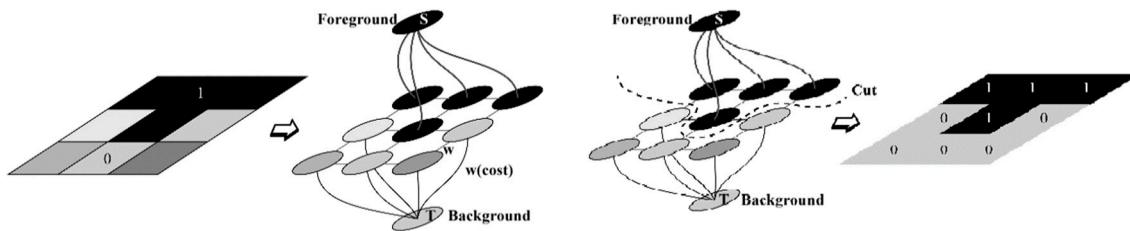


Figure 10. A demonstration of a minimum cut operation in image segmentation [69].

3.5. Proposed System by Using Yolov3 and GrabCut

This paper proposed an automatic GrabCut based segmentation pipeline using Yolov3 deep convolutional neural network. The proposed pipeline for segmentation of the skin lesion consists of four stages. In the first stage, hairs are removed from the image for more accurate detection and segmentation. The second stage is the detection of the lesion location in the image. For this purpose, some modifications were made on Yolov3 to accurately locate the lesion in a dermoscopic image. The original Yolov3 with 1000 class output was trained on the ImageNet data. Therefore, the modified version of Yolov3 used in this study has a single class output. The original Yolov3 gives three different size detections by applying 1×1 detection kernels to the feature maps at 79th, 91th and 106th layers. The shape of detection kernels is $1 \times 1 \times (B \times (5 + C))$. B is the number of bounding boxes of a cell, 5 is the number of the parameters the bounding box has, and C is the number of classes. In this study, there is only one class, so the filter size at 79th, 91th and 109th layers is 18 according to the filter formula, filters = (classes + 5) \times 3 [13] (see Figure 11). After the modifications, Yolov3 model was trained with dermoscopic images and used for detection of the lesion location in the image. Using this location information, a rectangle was drawn around the lesion automatically for creating a start point for GrabCut algorithm. After the location detection, GrabCut segments the image at the third stage. At the final stage, the morphological opening and closing have been applied to the segmented image using a 5×5 kernel for removal of the noise in the segmented binary image. The detailed system is presented in Figure 11.

3.6. Training Yolov3

Yolov3 [70] was trained with the ISBI 2017 dataset as the lesion detection part of our system. The dataset was separated as training and validation sets. Then the final system detection performance was evaluated using two different datasets (the ISBI 2017 and the PH2). There are studies about the effectiveness of transfer learning in deep nets [71,72]. So, in the training phase, we used pretrained weights of ImageNet dataset [73]. Afterwards, Yolov3 was fine-tuned and re-trained with the skin lesion images. The training parameters of Yolov3 are set as follows: batch size = 64, subdivisions = 16, momentum = 0.9, decay = 0.0005, learning rate = 0.001. Yolov3 was trained through 50,000 epochs and the network weights were saved every 10,000 epochs. Test result showed that the weights saved at 10,000th epoch were the most successful at detecting the location of lesion in the image. The whole implementations and computations were performed on a PC with two Intel Xenon processors, 64 GB RAM, NVIDIA GTX 1080Ti GPU and Ubuntu 14.04 operating system. Python and C programming languages, OpenCv image processing framework were used in the development of the system.

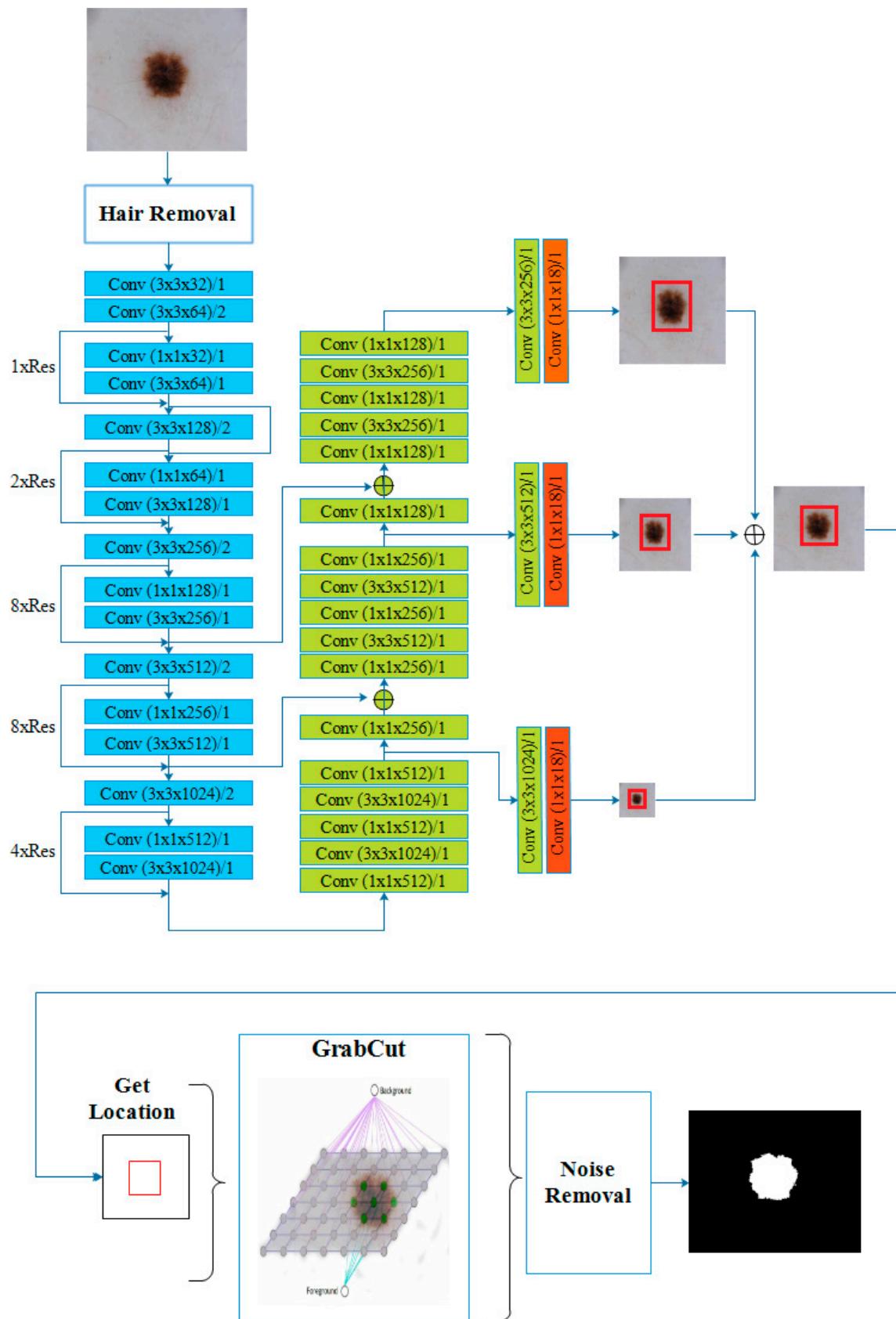


Figure 11. An illustration of the suggested skin lesion segmentation pipeline architecture.

3.7. Performance Evaluation Metrics

The pipeline introduced in this study was evaluated in two stages. In the first stage, the lesion location detection performance of re-trained Yolov3 in skin lesion images, was evaluated. In order to obtain more accurate segmentation results, the location information of the lesion in the image is very critical. Therefore, the detection performance of model was evaluated by using IOU metric. The detected lesion location was regarded as true if the IOU score was greater than 80%. In the second stage, the following performance metrics to further evaluate our method were used: sensitivity (Sen), specificity (Spe), the Dice coefficient (Dic), the Jaccard index (Jac) and accuracy (Acc). Sen represents the amount of correctly segmented lesion pixels while Spe shows the correctly segmented non-lesion areas ratio [53,56]. Dic is an evaluation metric which is used to measure segmented lesions and annotated ground truth similarity [56,74]. Jac is an evaluation metric for the intersection ratio between the obtained segmentation results and ground truths masks [75]. The main difference between Jac and IOU is that Jac is used for the segmentation and segmentation boundaries can be irregular, but IOU is used for the localization and it uses rectangular boundaries. Finally, accuracy shows the overall pixel-wise segmentation performance [56]. All aforementioned evaluation metrics are calculated by the following formulas:

$$\text{IOU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} \quad (9)$$

$$\text{Sen} = \frac{TP}{TP + FN} \quad (10)$$

$$\text{Spe} = \frac{TN}{TN + FP} \quad (11)$$

$$\text{Dic} = \frac{2 * TP}{(2 * TP) + FP + FN} \quad (12)$$

$$\text{Jac} = \frac{TP}{TP + FN + FP} \quad (13)$$

$$\text{Acc} = \frac{TP + TN}{TP + TN + FN + FP} \quad (14)$$

The TP , FP , FN , and TN symbolize the true positive, false positive, false negative and true negative, respectively. The lesion pixels in the image are considered as TPs if they are segmented correctly; otherwise, they are regarded as FNs . On the contrary, the non-lesion pixels in the image are considered as TNs if their prediction is non-lesion pixel; otherwise, they are regarded as FPs .

4. Results

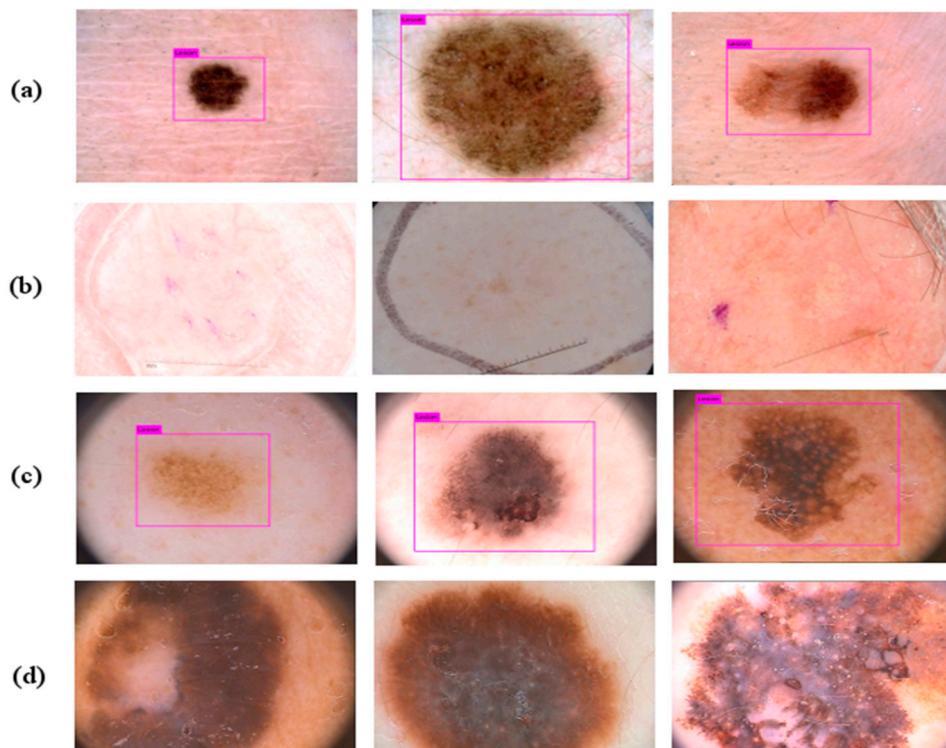
This section focuses on the results of detection and segmentation processes carried out on two different datasets the ISBI 2017 and the PH2, of the pipeline model used in this study.

4.1. Detection Results on the PH2 and the ISBI 2017 Datasets

The detection performance was calculated considering two metrics. The detection performance of the system for the PH2 dataset gave a 94% accuracy. The model was unable to make any predictions on only 6 images out of 200. The model achieved a 90% IOU rate in 194 images. The accuracy performance of model on the ISBI 2017 data set was 96,4% and it was unable detect only 22 images out of 600. The IOU rate of model was 86% in 578 images. Figure 12 shows examples from both datasets detected and undetected. Table 2 represents the detection performance of the model on two datasets.

Table 2. Yolov3 skin lesion location detection performance (%) on the PH2 and the ISBI 2017.

Datasets	Detection Accuracy	IOU	Total Undetectable
PH2	94.40	90	6 images in 200
ISBI 2017	96	86	22 images in 600

**Figure 12.** Results of skin lesion location detection by Yolov3 in dermoscopic images. (a,c) are successful detections on the ISBI 2017 and the PH2 respectively. (b,d) show unsuccessful detections on the ISBI 2017 and the PH2.

4.2. Segmentation Results on the PH2 and the ISBI 2017 Datasets

After evaluation of the detection of the lesion location, the segmentation performance of our method was compared with other new deep learning-based methods by calculating Acc, Sen, Spe, Jac and Dic metrics for both datasets. Table 3 summarizes the segmentation performance of the proposed pipeline method. Also, Figure 13 represents some examples of the segmentation results of the proposed model.

Table 3. Skin lesion segmentation performance (%) of the proposed method on the PH2 and the ISBI 2017.

Datasets	Acc	Sen	Spe	Jac	Dic
PH2	92.99	83.63	94.02	79.54	88.13
ISBI 2017	93.39	90.82	92.68	74.81	84.26

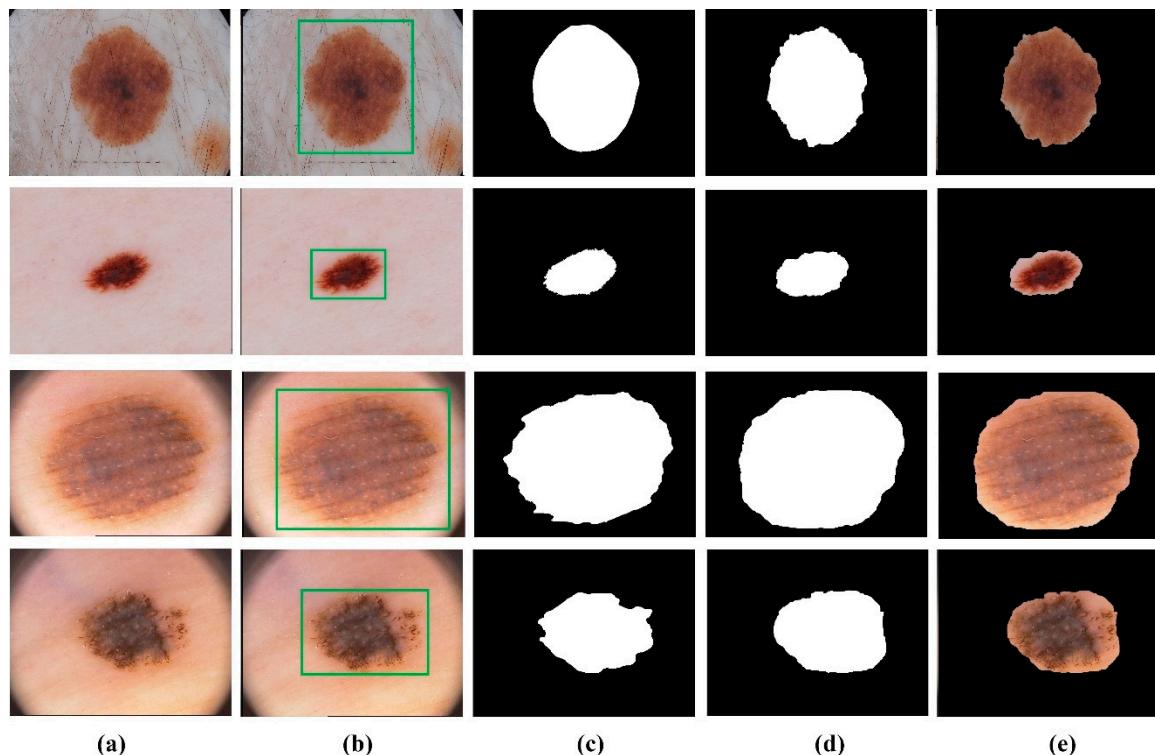


Figure 13. Skin lesion segmentation results of the proposed method. (a) Original image, (b) lesion location detection by Yolov3, (c) ground truth, (d) segmented lesion area by GrabCut and, (e) the final result.

5. Discussion

Automatic segmentation of the skin lesions is a very important step for CAD systems in the classification of skin lesions as melanoma. Especially as deep learning-based methods need a huge amount of data for proper segmentation. No data augmentation methods were employed in this study. Yolov3 was trained only with 2000 images and validated with 150 images. The test results show that the proposed pipeline method attained promising results when compared to other deep learning-based approaches. The main advantage of the proposed method is that this pipeline can segment skin lesions using high spatial resolution images. Therefore, more detailed features can be obtained from the segmented part of the lesion, which increases the classification accuracy. The comparison of the robustness of the proposed method with that of state-of-the art methods is shown in Table 4. This table presents the performance comparison of our method with approaches including U-Net based method [54], the most successful three techniques in the ISBI 2017 Challenge [55,76,77], and a study named FrCN [56].

Table 4. The comparison of the proposed method skin lesion segmentation performance (%) with the latest studies in the literature.

References	Acc	Sen	Spe	Jac	Dic
Yuan et al. (CDNN) [55]	93.40	82.50	97.50	76.50	84.90
Li et al. [76]	93.20	82.00	97.80	76.20	84.70
Bi et al. (ResNets) [77]	93.40	80.20	98.50	76.00	84.40
Lin et al. (U-Net) [54]	-	-	-	62.00	77.00
Al-Masni et. al. [56]	94.03	85.40	96.69	77.11	87.08
Proposed Method	93.39	90.82	92.68	74.81	84.26

According to the results, the proposed method outperformed other methods in sensitivity with 90.82 %. It also attained higher performances at Jac and Dic (74.81% and 84.16%) than the method proposed by Lin et al. When the other results are examined, it is obvious that there is not a significant difference between our proposed method and the others. Our proposed segmentation pipeline method gives a less perfect segmentation of lesion than the other deep learning-based methods. However, this situation is an advantage according to a study [78] claiming that the surrounding border of the skin lesion has beneficial information in the classification of lesions. On the other hand, modified Yolov3 used in this study achieved 90% and 86% IOU rates in the detection of lesion location on the PH2 and the ISBI 2017 datasets respectively. Figure 12 shows the modified Yolov3 does not detect the lesion because of the similarity between the lesion and surrounding tissue in the images with low contrast, and in the images where the lesion occupies the entire image surface. This problem can be overcome with more training data and contrast enhancement. Additionally, the segmentation processing time of the presented pipeline model is almost 7 s, which suggests that the method is feasible for real-time medical practices.

6. Conclusions

This paper deals with a new simultaneous segmentation method taking advantage of GrabCut and deep convolutional neural network Yolov3. Unlike the previous deep learning-based segmentation methods, our approach offers higher resolution and dimension independent segmentation results with the incorporation of different methods to a pipeline. We evaluated our method using two well-known datasets the PH2 and the ISBI 2017. According to the findings obtained from this study, the results are encouraging. It can be an alternative segmentation method for deep learning-based segmentation approaches. Furthermore, the proposed method can be used in different medical segmentation problems. In the featured work, we will add a deep convolutional neural network into the pipeline as a classification step for distinguishing the melanoma using segmented skin lesion images.

Author Contributions: The manuscript was written by E.A. under the supervision of H.M.Ü. The modeling, analysis, and software process was executed by E.A. H.M.Ü. helped in the review.

Funding: This paper was supported by Research Fund (Scientific Research Projects Coordination Unit) of the Kırıkkale University. Project Number:2018/40.

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

References

1. Karimkhani, C.; Green, A.; Nijsten, T.; Weinstock, M.; Dellavalle, R.; Naghavi, M.; Fitzmaurice, C. The global burden of melanoma: results from the Global Burden of Disease Study 2015. *Br. J. Dermatol.* **2017**, *177*, 134–140. [[CrossRef](#)] [[PubMed](#)]
2. Gandhi, S.A.; Kampp, J. Skin Cancer Epidemiology, Detection, and Management. *Med Clin. N. Am.* **2015**, *99*, 1323–1335. [[CrossRef](#)] [[PubMed](#)]
3. Feng, J.; Isern, N.G.; Burton, S.D.; Hu, J.Z. Studies of secondary melanoma on C57BL/6J mouse liver using ¹H NMR metabolomics. *Metabolites* **2013**, *3*, 1011–1035. [[CrossRef](#)] [[PubMed](#)]
4. Jemal, A.; Siegel, R.; Ward, E.; Hao, Y.; Xu, J.; Thun, M.J. Cancer statistics, 2019. *CA Cancer J. Clin.* **2019**, *69*, 7–34.
5. Tarver, T.; American Cancer Society. Cancer facts and figures 2014. *J. Consum. Health Internet* **2012**, *16*, 366–367. [[CrossRef](#)]
6. Siegel, R.; Miller, K.; Jemal, A. Cancer statistics, 2018. *CA Cancer J. Clin.* **2017**, *68*, 7–30. [[CrossRef](#)] [[PubMed](#)]
7. Pellacani, G.; Seidenari, S. Comparison between morphological parameters in pigmented skin lesion images acquired by means of epiluminescence surface microscopy and polarized-light videomicroscopy. *Clin. Dermatol.* **2002**, *20*, 222–227. [[CrossRef](#)]
8. Ali, A.-R.A.; Deserno, T.M. A systematic review of automated melanoma detection in dermatoscopic images and its ground truth data. In *Medical Imaging 2012: Image Perception, Observer Performance, and Technology Assessment*; International Society for Optics and Photonics: Bellingham, WA, USA, 2012; p. 8318.

9. Sinz, C.; Tschandl, P.; Rosendahl, C.; Akay, B.N.; Argenziano, G.; Blum, A.; Braun, R.P.; Cabo, H.; Gourhant, J.-Y.; Kreusch, J.; et al. Accuracy of dermatoscopy for the diagnosis of nonpigmented cancers of the skin. *J. Am. Acad. Dermatol.* **2017**, *77*, 1100–1109. [[CrossRef](#)]
10. Bi, L.; Kim, J.; Ahn, E.; Kumar, A.; Fulham, M.; Feng, D. Dermoscopic image segmentation via multi-stage fully convolutional networks. *IEEE Trans. Biomed. Eng.* **2017**, *64*, 2065–2074. [[CrossRef](#)]
11. Okur, E.; Turkan, M. A survey on automated melanoma detection. *Eng. Appl. Artif. Intell.* **2018**, *73*, 50–67. [[CrossRef](#)]
12. Rother, C.; Kolmogorov, V.; Blake, A. Grabcut: Interactive foreground extraction using iterated graph cuts. In *ACM Transactions on Graphics (TOG)*; ACM: New York, NY, USA, 2004.
13. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
14. Ganster, H.; Pinz, P.; Rohrer, R.; Wildling, E.; Binder, M.; Kittler, H. Automated melanoma recognition. *IEEE Trans. Med Imaging* **2001**, *20*, 233–239. [[CrossRef](#)] [[PubMed](#)]
15. Schaefer, G.; Krawczyk, B.; Celebi, M.E.; Iyatomi, H. An ensemble classification approach for melanoma diagnosis. *Memetic Comput.* **2014**, *6*, 233–240. [[CrossRef](#)]
16. Celebi, M.E.; Iyatomi, H.; Schaefer, G.; Stoecker, W.V. Lesion border detection in dermoscopy images. *Comput. Med. Imaging Graph.* **2009**, *33*, 148–153. [[CrossRef](#)] [[PubMed](#)]
17. Korotkov, K.; Garcia, R. Computerized analysis of pigmented skin lesions: A review. *Artif. Intell. Med.* **2012**, *56*, 69–90. [[CrossRef](#)] [[PubMed](#)]
18. Filho, M.; Ma, Z.; Tavares, J.M.R.S. A Review of the Quantification and Classification of Pigmented Skin Lesions: From Dedicated to Hand-Held Devices. *J. Med. Syst.* **2015**, *39*, 177. [[CrossRef](#)] [[PubMed](#)]
19. Oliveira, R.B.; Filho, M.E.; Ma, Z.; Papa, J.P.; Pereira, A.S.; Tavares, J.M.R. Withdrawn: Computational methods for the image segmentation of pigmented skin lesions: A Review. *Comput. Methods Programs Biomed.* **2016**, *131*, 127–141. [[CrossRef](#)]
20. Emre Celebi, M.; Wen, Q.; Hwang, S.; Iyatomi, H.; Schaefer, G. Lesion border detection in dermoscopy images using ensembles of thresholding methods. *Skin Res. Technol.* **2013**, *19*, e252–e258. [[CrossRef](#)]
21. Yuksel, M.; Borlu, M. Accurate Segmentation of Dermoscopic Images by Image Thresholding Based on Type-2 Fuzzy Logic. *IEEE Trans. Fuzzy Syst.* **2009**, *17*, 976–982. [[CrossRef](#)]
22. Peruch, F.; Bogo, F.; Bonazza, M.; Cappelleri, V.-M.; Peserico, E. Simpler, Faster, More Accurate Melanocytic Lesion Segmentation Through MEDS. *IEEE Trans. Biomed. Eng.* **2014**, *61*, 557–565. [[CrossRef](#)]
23. Möllersen, K.; Kirchesch, H.M.; Schopf, T.G.; Godtliebsen, F. Unsupervised segmentation for digital dermoscopic images. *Ski. Res. Technol.* **2010**, *16*, 401–407. [[CrossRef](#)] [[PubMed](#)]
24. Xie, F.; Bovik, A.C. Automatic segmentation of dermoscopy images using self-generating neural networks seeded by genetic algorithm. *Pattern Recognit.* **2013**, *46*, 1012–1019. [[CrossRef](#)]
25. Zhou, H.; Schaefer, G.; Sadka, A.H.; Celebi, M.E. Anisotropic mean shift based fuzzy c-means segmentation of deroscopy images. *IEEE J. Sel. Top. Signal Process.* **2009**, *3*, 26–34. [[CrossRef](#)]
26. Kockara, S.; Mete, M.; Yip, V.; Lee, B.; Aydin, K. A soft kinetic data structure for lesion border detection. *Bioinformatics* **2010**, *26*, i21–i28. [[CrossRef](#)] [[PubMed](#)]
27. Suer, S.; Kockara, S.; Mete, M. An improved border detection in dermoscopy images for density-based clustering. In *BMC Bioinformatics*; BioMed Central: London, UK, 2011.
28. Abbas, Q.; Celebi, M.; García, I.F. Skin tumor area extraction using an improved dynamic programming approach. *Skin Res. Technol.* **2012**, *18*, 133–142. [[CrossRef](#)] [[PubMed](#)]
29. Ashour, A.S.; Hawas, A.R.; Guo, Y.; Wahba, M.A. A novel optimized neutrosophic k-means using genetic algorithm for skin lesion detection in dermoscopy images. *Signal Image Video Process.* **2018**, *12*, 1311–1318. [[CrossRef](#)]
30. Abbas, Q.; Celebi, M.E.; García, I.F.; Rashid, M. Lesion border detection in dermoscopy images using dynamic programming. *Ski. Res. Technol.* **2011**, *17*, 91–100. [[CrossRef](#)] [[PubMed](#)]
31. Celebi, M.E.; Kingravi, H.A.; Iyatomi, H.; Aslandogan, Y.A.; Stoecker, W.V.; Moss, R.H.; Malters, J.M.; Grichnik, J.M.; Marghoob, A.A.; Rabinovitz, H.S.; et al. Border detection in dermoscopy images using statistical region merging. *Ski. Res. Technol.* **2008**, *14*, 347–353. [[CrossRef](#)] [[PubMed](#)]
32. Silveira, M.; Nascimento, J.C.; Marques, J.S.; Marcal, A.R.S.; Mendonca, T.; Yamauchi, S.; Maeda, J.; Rozeira, J. Comparison of Segmentation Methods for Melanoma Diagnosis in Dermoscopy Images. *IEEE J. Sel. Top. Signal Process.* **2009**, *3*, 35–45. [[CrossRef](#)]

33. Erkol, B.; Moss, R.H.; Stanley, R.J.; Stoecker, W.V.; Hvatum, E. Automatic lesion boundary detection in dermoscopy images using gradient vector flow snakes. *Ski. Res. Technol.* **2005**, *11*, 17–26. [CrossRef]
34. Mete, M.; Sirakov, N.M. Lesion detection in dermoscopy images with novel density-based and active contour approaches. In *BMC Bioinformatics*; BioMed Central: London, UK, 2010.
35. Wang, H.; Moss, R.H.; Chen, X.; Stanley, R.J.; Stoecker, W.V.; Celebi, M.E.; Malters, J.M.; Grichnik, J.M.; Marghoob, A.A.; Rabinovitz, H.S.; et al. Modified watershed technique and post-processing for segmentation of skin lesions in dermoscopy images. *Comput. Med. Imaging Graph.* **2011**, *35*, 116–120. [CrossRef] [PubMed]
36. Celebi, M.; Wen, Q.; Iyatomi, H.; Shimizu, K.; Zhou, H.; Schaefer, G. A State-of-the-Art Survey on Lesion Border Detection in Dermoscopy Images. *Dermoscopy Image Anal.* **2015**, *10*, 97–129.
37. Pathan, S.; Prabhu, K.G.; Siddalingaswamy, P. Techniques and algorithms for computer aided diagnosis of pigmented skin lesions—A review. *Biomed. Signal Process. Control.* **2018**, *39*, 237–262. [CrossRef]
38. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015.
39. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Cambridge, MA, USA, 2012; pp. 1097–1105.
40. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
41. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Berlin, Germany, 2015.
42. Cireşan, D.C.; Giusti, A.; Gambardella, L.M.; Schmidhuber, J. Mitosis detection in breast cancer histology images with deep neural networks. In *Proceedings of the International Conference on Medical Image Computing and Computer-assisted Intervention*; Springer: Berlin, Germany, 2013.
43. Pereira, S.; Pinto, A.; Alves, V.; Silva, C. Brain Tumor Segmentation using Convolutional Neural Networks in MRI Images. *IEEE Trans. Med. Imaging* **2016**, *35*, 1. [CrossRef] [PubMed]
44. Rodriguez-Ruiz, A.; Mordang, J.J.; Karssemeijer, N.; Sechopoulos, I.; Mann, R.M. Can radiologists improve their breast cancer detection in mammography when using a deep learning-based computer system as decision support? In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*; International Society for Optics and Photonics: Bellingham, WA, USA, 2018.
45. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; Van Der Laak, J.A.; Van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [CrossRef] [PubMed]
46. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *arXiv* **2015**, arXiv:1511.00561. [CrossRef]
47. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. [CrossRef]
48. García-García, A.; Orts-Escalano, S.; Oprea, S.; Villena-Martínez, V.; García-Rodríguez, J. A review on deep learning techniques applied to semantic segmentation. *arXiv* **2017**, arXiv:1704.06857.
49. Yu, Z.; Jiang, X.; Zhou, F.; Qin, J.; Ni, D.; Chen, S.; Lei, B.; Wang, T. Melanoma Recognition in Dermoscopy Images via Aggregated Deep Convolutional Features. *IEEE Trans. Biomed. Eng.* **2018**, *66*, 1006–1016. [CrossRef]
50. 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. 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 Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018.
51. Yuan, Y.; Chao, M.; Lo, Y.-C. Automatic Skin Lesion Segmentation Using Deep Fully Convolutional Networks with Jaccard Distance. *IEEE Trans. Med. Imaging* **2017**, *36*, 1876–1886. [CrossRef]
52. Mendonça, T.; Ferreira, P.M.; Marques, J.S.; Marcal, A.R.; Rozeira, J. PH 2-A dermoscopic image database for research and benchmarking. In Proceedings of the 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Osaka, Japan, 3–7 July 2013.

53. Goyal, M.; Yap, M.H. Multi-class semantic segmentation of skin lesions via fully convolutional networks. *arXiv* **2017**, arXiv:1711.10449.
54. Lin, B.S.; Michael, K.; Kalra, S.; Tizhoosh, H.R. Skin lesion segmentation: U-nets versus clustering. In Proceedings of the 2017 IEEE Symposium Series on Computational Intelligence (SSCI), Honolulu, HI, USA, 27 November–1 December 2017.
55. Yuan, Y.; Chao, M.; Lo, Y.-C. Automatic skin lesion segmentation with fully convolutional-deconvolutional networks. *arXiv* **2017**, arXiv:1703.05165.
56. Al-Masni, M.A.; Al-Antari, M.A.; Choi, M.-T.; Han, S.-M.; Kim, T.-S. Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks. *Comput. Methods Programs Biomed.* **2018**, *162*, 221–231. [CrossRef] [PubMed]
57. Li, H.; He, X.; Zhou, F.; Yu, Z.; Ni, D.; Chen, S.; Wang, T.; Lei, B. Dense Deconvolutional Network for Skin Lesion Segmentation. *IEEE J. Biomed. Health Inform.* **2018**, *23*, 527–537. [CrossRef] [PubMed]
58. Peng, Y.; Wang, N.; Wang, Y.; Wang, M. Segmentation of dermoscopy image using adversarial networks. *Multimed. Tools Appl.* **2018**, *78*, 10965–10981. [CrossRef]
59. Yuan, Y.; Lo, Y.C. Improving dermoscopic image segmentation with enhanced convolutional-deconvolutional networks. *IEEE J. Biomed. Health Inform.* **2019**, *23*, 519–526. [CrossRef] [PubMed]
60. ISIC. Skin Lesion Analysis Towards Melanoma Detection 2017. Available online: https://challenge.kitware.com/#challenge/n/ISIC_2017%3A_Skin_Lesion_Analysis_Towards_Melanoma_Detection (accessed on 29 May 2019).
61. Lee, T.; Ng, V.; Gallagher, R.; Coldman, A.; McLean, D. Dullrazor®: A software approach to hair removal from images. *Comput. Biol. Med.* **1997**, *27*, 533–543. [CrossRef]
62. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 24–27 June 2014; pp. 580–587.
63. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 13–16 December 2015.
64. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef]
65. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
66. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015.
67. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. *arXiv* **2017**, arXiv:1612.08242.
68. Zivkovic, Z. Improved adaptive Gaussian mixture model for background subtraction. In Proceedings of the 17th International Conference on ICPR Pattern Recognition, Cambridge, UK, 26–26 August 2004.
69. Wang, D. The Experimental Implementation of GrabCut for Hardcode Subtitle Extraction. In Proceedings of the 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS), Singapore, 6–8 June 2018.
70. Redmon, J. Darknet: Open Source Neural Networks in C. 2013–2019. Available online: <http://pjreddie.com/darknet/> (accessed on 29 May 2019).
71. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [CrossRef]
72. Shin, H.-C.; Roth, H.R.; Gao, M.; Lu, L.; Xu, Z.; Nogues, I.; Yao, J.; Mollura, D.; Summers, R.M.; Hoo-Chang, S. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Trans. Med. Imaging* **2016**, *35*, 1285–1298. [CrossRef] [PubMed]
73. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]
74. Dice, L.R. Measures of the amount of ecologic association between species. *Ecology* **1945**, *26*, 297–302. [CrossRef]
75. Powers, D.M. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *J. Mach. Learn. Technol.* **2011**, *2*, 37–63.

76. Li, Y.; Shen, L. Skin Lesion Analysis towards Melanoma Detection Using Deep Learning Network. *Sensors* **2018**, *18*, 556. [[CrossRef](#)] [[PubMed](#)]
77. Bi, L.; Kim, J.; Ahn, E.; Feng, D. Automatic skin lesion analysis using large-scale dermoscopy images and deep residual networks. *arXiv* **2017**, arXiv:1703.04197.
78. Burdick, J.; Marques, O.; Weinthal, J.; Furht, B. Rethinking skin lesion segmentation in a convolutional classifier. *J. Digit. Imaging* **2018**, *31*, 435–440. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).