



Computer Methods and Programs in Biomedicine Update

journal homepage: www.sciencedirect.com/journal/computer-methods-and-programs-in-biomedicine-update



Automated hair removal in dermoscopy images using shallow and deep learning neural architectures

Konstantinos Delibasis^{a,*}, Konstantinos Moutslos^b, Eleftheria Vorgiazidou^a, Ilias Maglogiannis^b

^a Dept of Computer Science and Biomedical Informatics, University of Thessaly, Lamia Greece

^b Dept. of Digital Systems, University of Piraeus, 80, Karaoli & Dimitriou Str, 18534, Piraeus, Greece

ARTICLE INFO

Keywords:
MLP-ANN
Encoder-decoder architecture
Dermoscopy image
Hair removal

ABSTRACT

Removing hair from digital dermoscopy images is occasionally a necessary step before further analysis is applied to the images. This work considers two machine learning approaches that segment the hair pixels from dermoscopy images. Subsequently, morphological post-processing is applied to refine the segmented hair and an image inpainting algorithm replaces the hair pixels with values based on the surrounding image structures. The first hair segmentation approach combines pixel-wise features extracted using the well-known Gaussian image pyramid with a traditional shallow multilayer perceptron (MLP-ANN), to detect hair pixels in images. The second approach uses a deep neural convolutional Encoder – Decoder (ED) network to segment hair. Both hair segmentation methods (MLP-ANN and ED) are trained with a set of 32 dermoscopy images with manually annotated hair, whereas the MLP-ANN dataset is constructed in a pixel-wise manner.

Both proposed methods underwent three different assessments. First a set of 50 images with a-priori known hair is used for hair segmentation evaluation. Secondly, a set of 13 different dermoscopy images with hair added using a suitably trained Generative Adversarial Network -GAN- are used to assess the quality of hair removal that generates the hair-free image, in terms of several error metrics with respect to the original hair-free image. Finally, both proposed hair segmentation methods (MLP-ANN and ED) are applied on a set of 200 hair and hair-free images, which is used for training an image classifier to recognize melanoma against nevi lesions and the improvement in the image classification accuracy is measured. Comparative results against several other state-of-the-art hair removal techniques are also presented.

Results show that in terms of hair removal, both the proposed hair removal techniques outperform the best performing of the state-of-the-art methods under comparison, in terms of several error metrics. Considering the effect of hair removal on melanoma image classification, the application of both MLP-ANN and ED increases the accuracy of melanoma classification. In all assessments, the ED was consistently the best performer. The statistical significance of the findings is also established.

1. Introduction and related work

Artificial Intelligence (AI) is gradually changing the routine of medical practice, and the level of acceptance by medical personnel is constantly increasing. Recent progress in digital medical data acquisition through medical imaging devices, machine learning and high-performance cloud computing infrastructures, push health-related AI applications into areas that were previously thought to be only the province of human experts. In this context, image analysis tools are being developed and embedded in advanced electronic health records to

automate the extraction of meaningful features from medical images and assist diagnosis [1]. Specifically, in the field of digital dermoscopy, a plethora of works exist in the literature, which aim at the extraction of useful information and the development of machine learning models for assisting skin cancer diagnosis [2]. The existence of hair in dermoscopy images is considered a significant source of noise that may deteriorate the performance of such diagnostic tools. Thus, the removal of hair from digital dermoscopy images may be an important image processing step that will improve accuracy before further analysis is performed.

Several methods have been developed starting with the DullRazor

* Corresponding author.

E-mail address: kdelibasis@gmail.com (K. Delibasis).

[3] that utilized morphological closing operations. Later, Xie *et al.* [4] used morphological closing-based tophat operator followed by PDE-based (partial differential equation) inpainting. A short description of other techniques includes Radon transform with Prewitt filters [5], the difference of Gaussian (DoG) with coherence transport [6], multi-scale matched filters with region growing and linear discriminant analysis (LDA) [7], adaptive canny edge detector with multi-resolution coherence transport [8], tophat operator with patch-based inpainting [9]. More recently, bottom-hat, Laplacian, Laplacian of Gaussian -LoG, and LoG-sobel combined with morphological operations, were proposed in [10], and a Sobel operator with Otsu thresholding in [11]. A method called “Virtual Shaver” was also demonstrated, using multiscale skeleton morphological operations [12]. A closing-based tophat operator with modified morphological operations was also reported [13]. Finally, Attia *et al.* [14] used a hybrid deep ANN with coherence transport and fast marching for inpainting. In a more recent work, Ian Lee *et al.* [15] proposed a hair segmentation from dermoscopy images, based on multiresolution steerable filters and adaptive per-image thresholding, which was shown to outperform other state-of-the-art methods. A similar method has also been proposed by Kasmi *et al.* [16] with very competitive results. Deep learning methods are becoming the current state-of-the-art for hair segmentation. In Lama *et al.* [17], an encoder-decoder neural architecture (“Chimeranet”) has been proposed for hair segmentation that was shown to marginally outperform other deep learning approaches U-Net [18] and ResUNet-a [19].

Evaluating hair segmentation algorithms can be performed using manually annotated hairs, or hairs that have been algorithmically added in dermoscopy images. Except for the procedure of manual annotation of hairs, the major limiting factor for developing, training, and evaluating hair removal methods is the absence of ground truth hair-free image, i.e. the image after the application of hair removal and image filling methods. Towards this end, Attia *et al.* [20] proposed the use of realistic hair synthesis with the use of conditional generative adversarial networks -GANs [21].

Based on an extensive literature review of the existing hair removal techniques, Talavera *et al.* [22] proposed a hair removal benchmark method. More specifically, they compare six (6) state-of-the-art techniques: Tim Lee *et al.* [3], Xie *et al.* [4], Abbas *et al.* [6], Huang *et al.* [7], Toossi *et al.* [8], and Bibiloni *et al.* [13]. These techniques were chosen for comparison due to their availability and use in the literature. The test set consisted of 13 dermoscopic images that were generated using 2 hair simulators: a software called HairSim [23] and a Generative Adversarial Network -GAN [20]. Nine (9) error metrics were utilized for the comparison, namely: Mean Squared Error (MSE), structural similarity index SSIM [19], the well-known Peak Signal-to-Noise Ratio PSNR [24], Root Mean Squared Error (RMSE), VIF [25], Universal Quality Index UQI [26], MSSSIM[27], and two modified PSNRs that correlate well with the human visual system PSNR-HVS-M [28], PSNR-HVS [29]. The error metrics are computed for each image in the test set, and the mean and standard deviation are calculated for each method under comparison. As a final step, the statistical significance of the results is assessed for all pairs of hair removal techniques.

Most of the recent works utilize deep convolutional neural networks for the identification and removal of hair in dermoscopy images [30,31]. However, such methodologies require a large number of annotated images, which are difficult to obtain. Recently, a data-driven unsupervised approach was proposed using GAN [32] which was trained on a dataset containing nearly eight thousand images labeled only by the existence of hairs or not. Li *et al.* [33] created a manually-curated ground-truth hair-mask dataset to train a U-Net segmentation network and a GAN for the inpainting process.

Hairs are linear structures, usually dark, although sometimes brighter than the background, which often are superimposed on other essential similar skin structures, like streaks. The traditional pixel-based image features that are based on linear scale-space, such as the Laplacian pyramid and the eigenvalues of the multi-scale Hessian matrix are good

local image descriptors. The well-established SIFT and SURF image features are based on detecting the spatial and scale-wise extreme response. The eigenvalues of the Hessian matrix also classify image pixels according to the local image structure. Thus, it is expected that an MLP-ANN that receives the normalized LoG image responses, as well as the Hessian eigenvalues would be able to achieve pixel-level classification. Augmenting input of convolutional NNs -CNN- by the response of similar filter-banks has been shown beneficial for the task of image classification [34].

Few similar image processing approaches based on shallow MLP-ANN with pixel-wise input have been tried in the past in other domains with relevant success. In Amiri *et al.* [35], statistical features (including texture-based descriptors) were used as input for an MLP-ANN for white and gray matter segmentation from MR images. In Serrano *et al.* [36], an ANN consisting of few neurons of a new type, trained using the PSO algorithm, was proposed to segment objects from the background image. That ANN was also used successfully for material image segmentation [37]. In Yuqian Zhao *et al.* [38], the smoothed image, the standard deviation and the distance transform of the original image are input into an ANN for segmenting the liver in CT transverse images.

In this work, we propose to combine these well-established local image descriptors with a very simple and fast shallow multi-level perceptron artificial neural network (MLP-ANN), to perform pixel-wise hair segmentation and removal from dermoscopy images and to investigate the performance of different pixel-wise feature vectors.

On the other hand, in the last years deep learning is being increasingly utilized for the recognition and segmentation of objects in Computer Vision. The goal of the segmentation tasks is to associate each pixel of the image with a semantic tag [39]. For the purpose of this work, we selected a deep learning encoder-decoder (ED) neural architecture called DeepLabv3+, proposed by Chen *et al.* [40] to segment the hair from dermoscopy images, as a deep learning alternative to compare with the shallow MLP-ANN.

DeepLabv3+ is an extension of the DeepLabv3 [41] encoder with the simple addition of a decoder unit. The encoder uses atrous convolution [42] to extract backbone network features and encode semantic information, while the decoder part that follows, generates a binary image that corresponds to the segmented object. As explained in [42], atrous convolution allows the explicit control of the image resolution of feature responses and the enlargement of the field of view of filters within the DeepLabv3+, without increasing the number of learnable parameters or the computational complexity.

The output of both the shallow ANN and the DeepLabv3+ are morphologically post-processed and the image inpainting technique of coherent transport [43] is applied to generate the RGB hair-free image. We did not resort to deep learning techniques for inpainting, since the relevant GAN dataset available to us was too limited for the necessary training.

The assessment of the proposed hair removal methods is three-fold.

- First the proposed methods are assessed in terms of hair segmentation, against the method of Ian Lee *et al.* [15] for 50 dermoscopy images with a-priori known hair
- Secondly, they are compared with the method of Ian Lee *et al.* [15], with Xie *et al.* [4], which was the winner hair removal method in Talavera-Martínez [22], as well as with the publicly available Virtual Shaver (VS) application [12] that had not been included in the comparative study of Talavera-Martínez [22]. The comparison is performed in terms of the nine aforementioned error metrics (MSE, SSIM, PSNR, RMSE, VIF, UQI, MSSSIM, PSNR-HVS-M and PSNR-HVS), calculated over the 13 images of the GAN dataset - which contains more realistic hair simulations than hair artificially created by HairSim, as in Talavera-Martínez [22]. The statistical significance of the results is tested utilizing the *t*-test in cases where

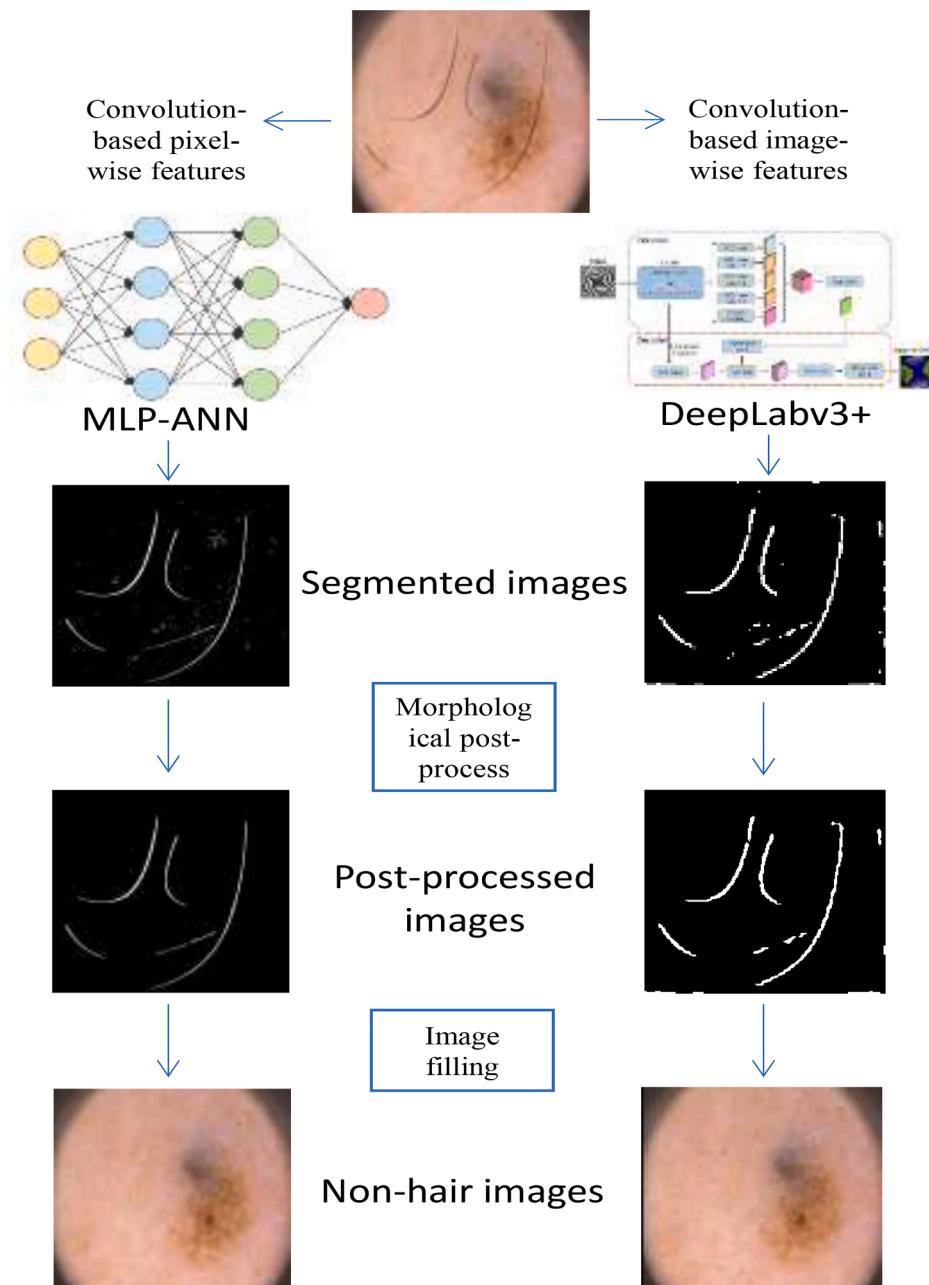


Fig. 1. The proposed overall algorithmic workflow.

the mean errors pass the Shapiro-Wilk normality test, or otherwise the Wilcoxon signed-rank test, both based on a 95% confidence level. • Finally, method comparison is performed in terms of the accuracy of dermoscopy image classification task into two classes: Melanoma and Nevi, using a different set of 200 dermoscopy images, half of which contain hair. The standard ResNet [44] with transfer learning was used for the classification.

The rest of the paper is structured as follows: In [Section 2](#), we present in detail the proposed methodology focusing on the definition of the relevant features and the shallow ANN and ED training. [Section 3](#) includes the qualitative and quantitative evaluation of the proposed methods and [Section 4](#) concludes the paper.

2. The proposed methodology

2.1. Overview of algorithmic workflow

As already mentioned, the Laplacian pyramid is utilized to generate a feature vector for each image pixel that is fed into a traditional shallow MLP-ANN. In contrast to the MLP – ANN's handcrafted features that perform pixel-based segmentation, the ED architecture uses auto-generated image-descriptors. The MLP-ANN and the ED are trained to segment hair pixels. The responses of both architectures are converted to binary and they are further morphologically post-processed to identify the final hair pixels and reject linear structures that resemble hair. As the last step, image inpainting techniques are employed to replace hair pixels with background image values. The overall algorithmic workflow is illustrated in [Fig 1](#).

Table 1

The different pixel-wise feature vectors for the shallow MLP-ANN that were tested in this work.

	F_1	F_2	F_3	F_4
Num. of σ values (n)	8	8	8	8
σ values	1,2,...,8	1,2,...,8	1,2,...,8	1,2,...,8
Gaussian	YES	NO	NO	NO
DoG	NO	YES	YES	YES
Hessian Λ	NO	NO	YES	YES
Hessian λ	NO	NO	NO	YES
Circularity v_c	NO	NO	YES	YES
Feature dimensionality	8	7	10	11

2.2. Hair identification using MLP-ANN and handcrafted feature vectors

Let I be the dermoscopy image, σ_k a set of increasing values for the standard deviation of the Gaussian kernel g_k and the kernel is convolved with image I .

$$G_k = I * g_k \quad (1)$$

The Laplacian pyramid is constructed using the difference of Gaussians DOG approximation. It is well known that the subtraction of any two convolved images with consecutive σ approximates the Laplacian with the smaller σ .

$$G_{k+1} - G_k = I * DoG_k \sim I * (\sigma_{k+1} - \sigma_k) \sigma_k LoG_k \quad (2)$$

Since the scale-normalized $LoG_{k,norm}$ equals $\sigma_k^2 LoG_k$, the above equation can be rewritten:

$$G_{k+1} - G_k = I * \left(\frac{\sigma_{k+1}}{\sigma_k} - 1 \right) LoG_{k,norm} \quad (3)$$

Consequently, the approximated Laplacian response is scale-normalized before it is input into the ANN, by dividing it by the factor $\left(\frac{\sigma_{k+1}}{\sigma_k} - 1 \right)$. We considered that equi-spaced integer values of σ would be advantageous, thus consecutive σ values were used: $\sigma_k = k$. By substituting in the previous equation, we obtain:

$$Log_{k,norm} = k \cdot DoG_k \quad (4)$$

If n different σ values are used, (1) and (2) can generate n and $n-1$ features, respectively, for each pixel (i,j) :

$$\begin{aligned} f &= [G_1, G_2, \dots, G_n] \\ f &= [DoG_1, 2DoG_2, \dots, (n-1)DoG_{n-1}] \end{aligned} \quad (5)$$

(for simplicity, the i,j indices have been dropped). Further, for each value σ_k , the Hessian matrix is calculated using either G_i or DoG_i as indicated in Table 1 (3rd and 4th row), and the maximum and minimum eigenvalues are stored in Λ^k, λ^k for all image pixels (superscripts are not powers).

For each pixel (i,j) of the Λ^k, λ^k series, the element with the maximum absolute value is stored in the final Λ, λ images:

$$\begin{aligned} \Lambda_{ij} &= \Lambda_{ij}^{\Sigma_1}, \lambda_{ij} = \lambda_{ij}^{\Sigma_2} \\ \Sigma_1 &= \text{argmax} \left(|\Lambda_{ij}^k| \right), \\ \Sigma_2 &= \text{argmax} \left(|\lambda_{ij}^k| \right), k = 1, 2, \dots, n \end{aligned} \quad (6)$$

Eq. (6) generates two more features.

Finally, the usefulness of a circularity measure is also explored in this work. Since image pixels inside circular structures exhibit comparable principal curvatures, the absolute values of the two eigenvalues of the Hessian matrix λ_1, λ_2 should be comparable and significantly greater than 0. Accordingly, the circularity measure [10] is utilized as a feature, mainly to reduce the false-positive response of the ANN, defined as follows

$$v_c(\Lambda, \lambda) = \begin{cases} e^{-\left(\frac{0.5}{|\Lambda|+|\lambda|} + 4(|\Lambda|-|\lambda|) \right)}, & \Lambda, \lambda \neq 0 \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

It is easy to confirm that the value of v_c is in [0,1] range and it approaches 1 when $|\Lambda|, |\lambda| > 0$ and $|\Lambda| - |\lambda| \rightarrow 0$.

Fig. 2a depicts the features derived from the Laplacian pyramid resulting from (4), using a common color-scale and Fig. 2b the results of (6),(7) for a patch of a typical dermoscopy image, also using a common color-scale, to facilitate visual comparison.

The Laplacian pyramid and the Hessian matrix contain information about the image structure at different scales. Thus, the constructed

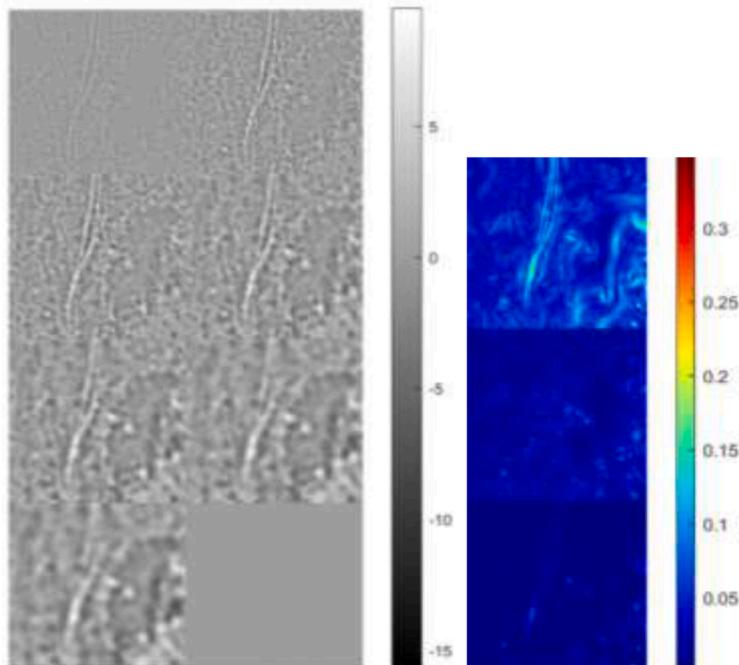


Fig. 2. An example of the features extracted from a typical dermoscopy image. (a) The 7 normalized DoG for $\sigma=1, 2, \dots, 8$ and (b) the Λ, λ and v_c (top to bottom).

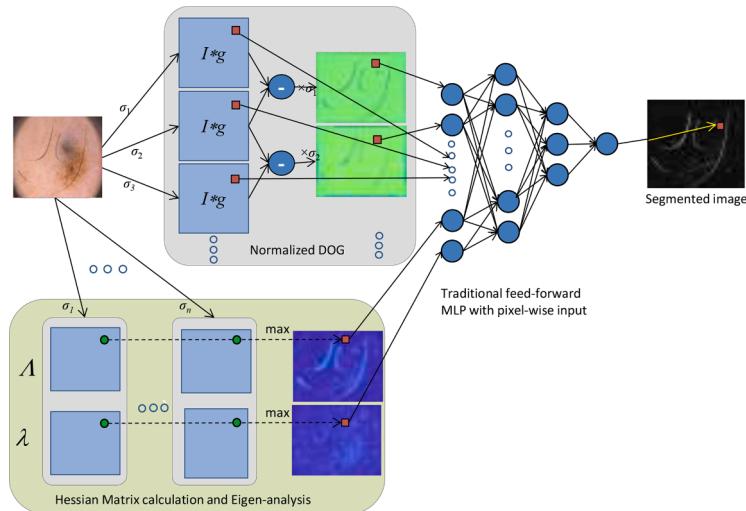


Fig. 3. The proposed architecture for the MLP-ANN pixel-wise classification. The image pyramid of normalized DoG and the Hessian eigen-analysis for different scales are shown, producing the pixel-wise input to the ANN.

feature vector is expected to carry substantial local image information at multiple scales in a hierarchical manner, despite its small dimensionality. Since the feature vector is already the result of several convolution and maximum operators, it may be used in a simple multi-layered perceptron artificial neural network, to classify each pixel in the image. The proposed classification architecture is shown in Fig. 3. The squares depict the pixels that are input into the ANN, whereas the circles mark the pixels that are used to calculate the ANN input. Thus, Fig. 3 shows all

possible features that are considered in this work.

Several feature vectors are explored in this work, as summarized in Table 1. The number and range [1,8] of σ integer values are kept constant, which should provide the ability to detect hair with different width, in images with various pixel resolution. First, the 8 G_k values are utilized (column ' F_1 '). Since this feature vector did not produce useful results, it was replaced by the 7 normalized DoG values (column ' F_2 '). The quantities Λ , λ and v_c are appended to the feature vector, as shown in

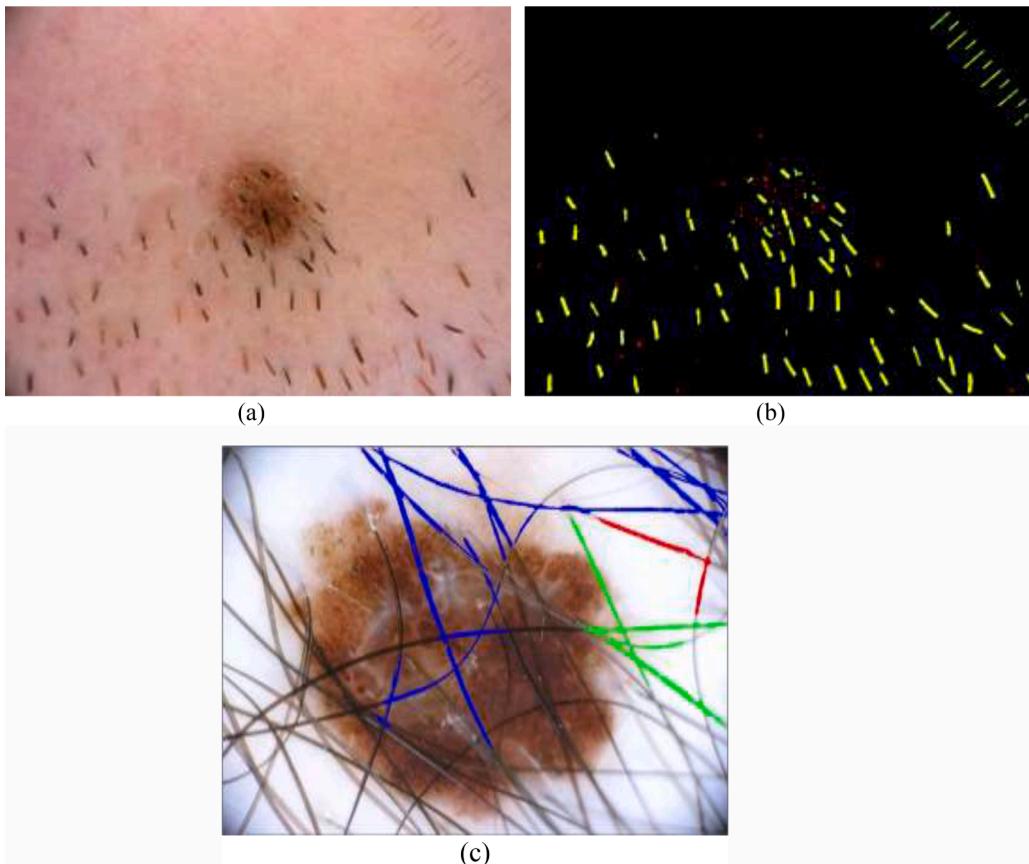


Fig. 4. (a) An image with short hair and (b) the output of the MLP-ANN before (yellow and red pixels) and after (red pixels) the application of the criterion of Eq. (8). In (c) an exemplar dermoscopy image with long hair with three binary connected components, after segmenting the ANN's output, shown in color, indicating the shape's complexity.

columns ‘F₃’ and ‘F₄’ of Table 1. The MLP-ANN architecture that was used comprised of three (3) hidden layers with 18, 8 and 4 neurons, respectively. The hyperbolic tangent was used as the activation function in all layers and the cross-entropy was selected as the error for ANN training.

2.3. Hair identification using ED

For the training and evaluation of an ED neural architecture that performs semantic image segmentation, it is necessary to supply the original images and the labeled ones, so that the ED learns to recognize and detect the desired areas.

Due to the limited available dataset, as described in subsection III.A, it was necessary to use a pre-trained neural model to transfer previously learned image features into our specific task. To detect the hair in the dermatological images we created a DeepLabv3+ convolutional neural network [40] for semantic image segmentation using the pre-trained ResNet18 as network backbone [44]. DeepLabv3+ consists of 100 layers. ResNet18 is a residual network with a total of 71 layers, of which 17 convolutional layers and 1 fully connected layer for classification tasks, that has been pretrained on the ImageNet database [45], which contains over 1 million images. ResNet18 accepts as input 224 by 224 images, thus resize of the initial images was a necessary step in data preprocessing.

After experimentation, the mini-batch size was set to 8 images, the learning rate was set to 0.01 for the first 50 epochs and was reduced to 0.001 for the remaining 50 epochs. Cross-entropy was used as the loss function. Stochastic gradient descent with momentum was selected as the optimization method and the maximum number of training epochs was set to 100. Validation was enabled during training to avoid model’s overfitting. Training took place on a modest GPU (NVIDIA GeForce MX130 with maximum memory 2GB VRAM).

2.4. Post-processing of the results and generation of hair-free images

The response of the output layer of both neural architectures is segmented using hysteresis thresholding with normalized high threshold $T_H=0.1$ and low threshold $T_L=0.02$. The connected components of the resulting binary image are labeled and for each component i , the area A_i and perimeter p_i are calculated. Contrary to initial conception, hair often intersects to create complex segmented shapes, as can be observed in Fig. 4(c), where three different connected components of the segmented hair are shown in red green and blue color, respectively. The components that are very small, or not consisting of elongated parts, equivalently satisfying the following criterion:

$$c_i = \frac{A_i}{p_i^2} < 15 \text{ OR } p_i < 10, \quad (8)$$

are removed from the binary image. This criterion is very fast to apply and very robust to handle the generality of the appearance of the segmented connected components. The value of c_i was determined experimentally and it is expected to be effective on normal and short hairs, as well as independent of the pixelation of the image. Fig. 4 shows a dermoscopy image containing short hair (a), along with the labelled output of the proposed shallow MLP-ANN as non-black pixels in (b). Yellow pixels are the pixels retained as hair after the application of the criterion in Eq. (8), whereas red marks the pixels that were initially classified as hair by the shallow MLP-ANN but subsequently rejected by not satisfying Eq. (8). It is worth noting that ruler markings cannot be rejected, however they are usually placed in non-critical image areas and their removal is desirable, in the process of generating the hair-free image.

The remaining connected labeled components are dilated using a 3 × 3 square, to avoid artifacts during the inpainting process, due to the coincidence of the segmented hair with the hair border [14]. Thus, the

final segmented hair binary image is produced.

The final step consists of replacing the values of hair pixels with values derived from the neighboring image pixels, a task often called inpainting. In this work we utilized two methods for this task: a) the formulation as an isotropic diffusion partial derivative problem, by discretizing the Euler’s equation and b) the more complex method of coherent transport [43]. The latter method proved more accurate; thus, it was selected to generate the reported results. The response of the trained shallow ANN and deep ED for hair detection for a typical dermoscopy image is shown in Fig. 5, along with the post-process hair segmented images and the resulting inpainted image.

3. Results

3.1. Dataset description

3.1.1. Dataset for MPL-ANN and ED training

The training of the MLP-ANN and the ED was performed using 32 dermoscopy images with hand-annotated hair, randomly selected from

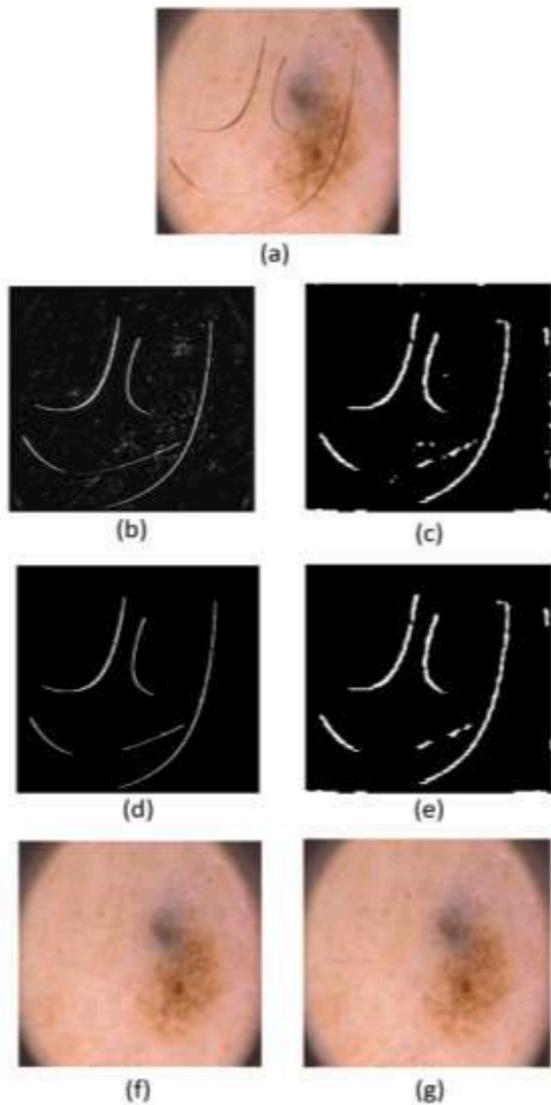


Fig. 5. (a) A typical dermoscopy image with hair, (b), (c), the responses of the trained MLP-ANN and the trained ED, (d), (e), the post-process of the MLP-ANN and ED response respectively and (f), (g), the resulting hair free images from both models.

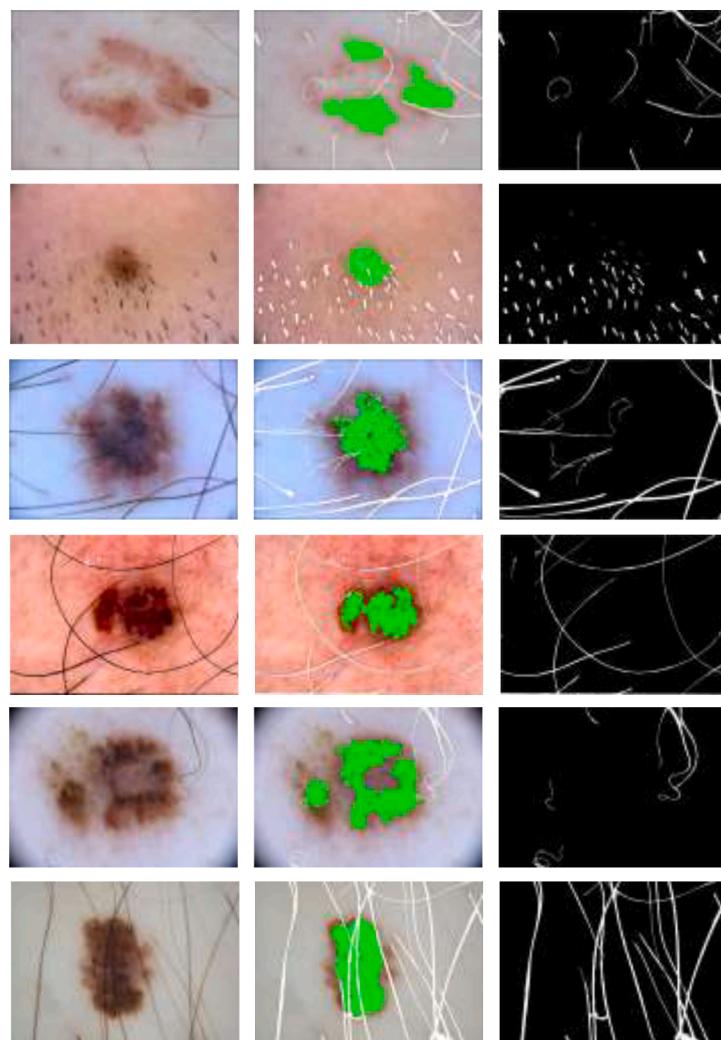


Fig. 6. Typical images of the training dataset: 1st column: the original images from the ISIC datasets, 2nd column: shallow MLP - ANN training set, (white pixels is “hair” pixels and green “no-hair”), 3rd column: ED training set (white areas for hair and black for no-hair).

Table 2
The composition of the 32-image dataset for training the shallow MLP-ANN.

Class	Num. of pixels
Hair (white pixels)	1,275,624
No-hair (green pixels)	1,662,590
Total	2,938,214

the ISIC dermoscopy dataset [46,47,48] using a standard image editor. Fig. 6 shows few typical images of the dataset. The first column depicts the images obtained from the ISIC datasets. The images of the MLP-ANN training set are displayed in the second column, with the color-encoded pixels that constitute the two classes: white for hair pixels and green for no-hair. The 3rd column depicts the corresponding hand-annotated images for the training of the ED.

It has to be mentioned that the proposed shallow MLP-ANN operates in a pixel-wise manner. We used the white color to mark the hair pixels and the green color to mark the non-hair pixels. Pixels with any other color are ignored and excluded from the shallow MLP-ANN dataset. Thus, not all hair pixels need to be labeled and not all of the remaining pixels were selected as the “no-hair” class for the MLP-ANN dataset. Specifically, for the “no hair” pixels we selected mainly pixels on the lesion, since these are more likely to generate false positive responses. In

this way, we minimize the distortion caused by removing hair pixels, while keeping the imbalance between the two classes less severe [49, 50].

Table 2 summarizes the resulting dataset for the shallow MLP-ANN, in pixels. For the training of the ED the images were resized to 224 rows and 224 columns. 19 out of the 32 images were randomly selected as the training subset, which were increased to 38 images with data augmentation techniques (random rotation $\pm 10^\circ$). The validation and test subset were set equal to 6 and 7 images respectively. Data augmentation was not applied to the validation and test dataset because these data should only contain the original images for unbiased evaluation.

3.1.2. Evaluation datasets

Both the proposed shallow MLP-ANN and deep ED architectures were evaluated against other state-of-the-art methods in three different manners:

- First they are assessed in terms of the accuracy of hair segmentation for 50 dermoscopy images, with a-priori known hairs. 13 of these hair dermoscopy images were generated by hair free images using a Generative Adversarial Network (GAN) [20] (see next bullet point) and in 37 images the hairs were manually segmented for evaluation purposes.
- Secondly they are assessed in terms of the *quality of hair removal* for 13 dermoscopy images that contain hair, generated using Generative

Table 3

The composition of the 200-image melanoma classification dataset (Number of images).

Class	Hair	No-hair
Melanoma	50	50
NEVUS	25	25
Atypical NEVUS	25	25

Table 4

Hair segmentation results for the 50 images with known hair.

Method	Accuracy	wIoU	Sensitivity	Specificity
ED (proposed)	98.07 ± 0.59	96.98	86.60	98.87
ED (ResNet50 [44])	97.69 ± 0.72	96.58	89.06	97.87
ED (MobileNetv2 [52])	97.59 ± 0.89	96.46	90.71	97.73
MLP-ANN (proposed)	97.83 ± 1.60	96.42	73.42	98.56
Ian Lee [15]	97.26 ± 1.49	96.03	73.01	97.91
Kasmi [16]	97.08 ± 3.39	95.14	68.47	98.79

Adversarial Networks (GAN), [20], called “GAN-test dataset”. The nine aforementioned error metrics (MSE, SSIM, PSNR, RMSE, VIF, UQI, MSSSIM, PSNR-HVS-M and PSNR-HVS), were utilized.

- Finally they are assessed in terms of the *accuracy of a dermoscopy image classification* task into two classes: “Melanoma” and “Nevi”, using a different set of 200 dermoscopy images from the PH2 dataset [51] and the HAM10000 dataset [46], 100 of which contain hair, as described in detail in **Table 3**. This dataset is herein called “melanoma classification dataset”. The standard ResNet [44] with transfer learning was used for the classification.

Thus, the training and the assessment of both the shallow and the deep ED architectures is performed on different datasets, as well as on different tasks.

3.2. Quantitative results

In this subsection we present results that include the accuracy of hair segmentation, quality metrics for the hair free images (images generated by the proposed algorithm after removing the detecting hair and filling the missing image pixel values, as well as the effect of hair free images on melanoma classification.

3.2.1. Hair segmentation

The accuracy of hair segmentation is studied in this paragraph. We considered the proposed ED and MLP-ANN method, as well as the methods of Ian Lee et al. [15] and Kasmi et al. [16], which was implemented from the corresponding paper, since no source code was available by the authors (the method of Xie [4] and VS [12] were not included since no source code was available). Hair segmentation is evaluated in terms of Accuracy, weighted Intersection over Union (wIoU), sensitivity and specificity. **Table 4** presents the results, calculated over a number of 50 images with known hairs (including the 13 GAN generated images). The proposed ED was further assessed with different backbone networks (ResNet18 [44], ResNet50 [44] and MobileNetv2 [52]) achieving very similar accuracy. Based on wIoU, we proceeded using the ResNet18 [44].

3.2.2. Hair removal

The proposed system was applied to the 13 images of the GAN-test dataset as described in subsection III.A. The VS method [12], the method of Xie et al. [4] (the winner method of the Talavera-Martínez comparison [22]), as well as the method of Ian Lee et al. [15], were also included in the comparison. Nine different error metrics, as already mentioned in the introduction, between the filtered and the ideal (hair-free) image, were used to assess the methods under comparison: MSE, SSIM [24], PSNR [24], RMSE, VIF [25], UQI [26], MSSSIM [27], PSNR-HVS-M [28], PSNR-HVS [29].

Table 5 presents the quantitative results of the proposed methods: the shallow MLP-ANN using the feature vector F4 (**Table 1**) as well as the ED deep learning architecture, calculated for the GAN-test dataset. The feature vectors F1, F2, F3 achieved error metric values that were consistently worse than the MLP-ANN F4 as well as the Xie et al. method [4], thus they were not included in further comparison. The average (μ) and standard deviation (σ) of the 9 error-metrics are calculated over all the images in GAN-generated test set. The best performing method for each error metric is depicted in bold and the second-best in gray color.

Table 6 presents the statistical comparisons between selected hair removal methods from **Table 5**. The symbols “✓” and “✓✓” indicate better performance, statistically not significant and statistically significant for the first method in comparison, in terms of the respective metric. Likewise, the symbols “✗” and “✗✗” indicate the same (e.g., not significantly better and significantly better) for the second method in comparison. Statistical significance is checked using the Student’s *t*-test and the Wilcoxon signed-rank test and (“t”, “w”) in the case of normally and non-normally distributed data, respectively.

Table 5

Comparing Xie [4], VS [12], Ian Lee [15], the proposed mlp-ann hair-removal method and the proposed ed method, for the GAN-test dataset. The best and second-best performing method is shown in bold and gray respectively.

Error metric	State of the art			Proposed	
	Xie [4]	VS [12]	Ian Lee [15]	MLP-ANN F4	ED
MSE ↓	μ	21.54	29.05	32.33	30.88
	σ	11.74	26.43	11.74	21.45
SSIM ↑	μ	0.927	0.92	0.928	0.93
	σ	0.015	0.016	0.015	0.013
PSNR ↑	μ	35.28	34.50	32.79	34.02
	σ	2.035	2.765	2.035	2.627
RMSE ↓	μ	4.508	5.060	5.156	5.303
	σ	1.147	1.932	1.147	1.729
VIF ↑	μ	0.588	0.558	0.555	0.596
	σ	0.038	0.046	0.038	0.033
UQI ↑	μ	0.999	0.999	0.999	0.999
	σ	0.001	0.001	0.001	0.001
MSSSIM ↑	μ	0.975	0.975	0.977	0.982
	σ	0.018	0.01	0.018	0.005
PSNR-HVS-M ↑	μ	37.35	36.33	34.09	35.64
	σ	3.043	3.355	3.043	3.813
PSNR-HVS ↑	μ	35.67	34.73	32.634	34.176
	σ	2.634	3.103	2.634	3.464

↑: higher values better, ↓ lower values better.

Table 6

. Statistical comparisons between the hair removal methods from [Table 5](#).

	Stat. test	MSE	SSIM	PSNR	RMSE	VIF	UQI	MSSSIM	PSNR-HVS-M	PSNR-HVS
F4 vs VS	p-val.	0.83	2e-4	0.63	0.73	2e-4	0.032	2.4e-4	0.685	0.735
	Meth.	w	w	w	w	w	w	w	w	w
	Signif.	x	✓✓	x	x	✓✓	✓✓	✓✓	x	x
F4 vs Xie	p-val.	0.14	0.057	0.11	0.14	0.017	0.73	0.004	0.22	0.19
	Meth.	w	w	w	w	w	w	w	w	w
	Signif.	x	✓	x	x	✓✓	✓	✓✓	x	x
F4 vs Ian Lee	p-val.	4.6e-3	4.9e-4	2.4e-3	3.4e-3	2.4e-4	0.013	2.4e-4	0.013	6.1e-3
	Meth.	w	w	w	w	w	t	w	w	w
	Signif.	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓
VS vs Xie	p-val.	0.19	0.002	0.127	0.14	2e-4	0.146	0.454	0.27	0.24
	Meth.	w	w	w	w	w	w	w	w	w
	Signif.	x	xx	x	x	xx	x	x	x	x
ED vs F4	p-val.	0.07	4e-5	0.032	0.047	0.003	0.99	0.243	0.23	0.12
	Meth.	w	t	w	w	w	w	t	t	t
	Signif.	✓	✓✓	✓✓	✓✓	✓✓	✓	✓	✓	✓
ED vs Xie	p-val.	0.02	1.5e-5	0.032	0.027	0.002	0.19	0.006	0.89	0.68
	Meth.	w	t	w	w	w	w	w	w	w
	Signif.	✓✓	✓✓	✓✓	✓✓	✓✓	✓	✓✓	x	✓
ED vs Ian Lee	p-val.	2.4e-4	1.4e-6	2.4e-4	2.4e-4	2.4e-4	0.109	2.4e-4	0.017	6.1e-3
	Meth.	w	t	w	w	w	w	w	w	w
	Signif.	✓✓	✓✓	✓✓	✓✓	✓✓	✓	✓✓	✓✓	✓✓

‘✓’ and ‘✓✓’: better performance for the first method, not statistically significant and statistically significant, respectively.

‘x’ and ‘xx’ indicate the same for the second method in comparison.

‘w’, ‘t’: Wilcoxon signed-rank test and student’s t-test.

As it can be observed from [Table 5](#) and [6](#), the proposed ED is the best performing method in terms of all 9 error metrics, except for PSNR-HVS-M, in which it is the second-best performer, after Xie *et al.*

According to [Table 5](#) and [6](#), the proposed ED method appears to be

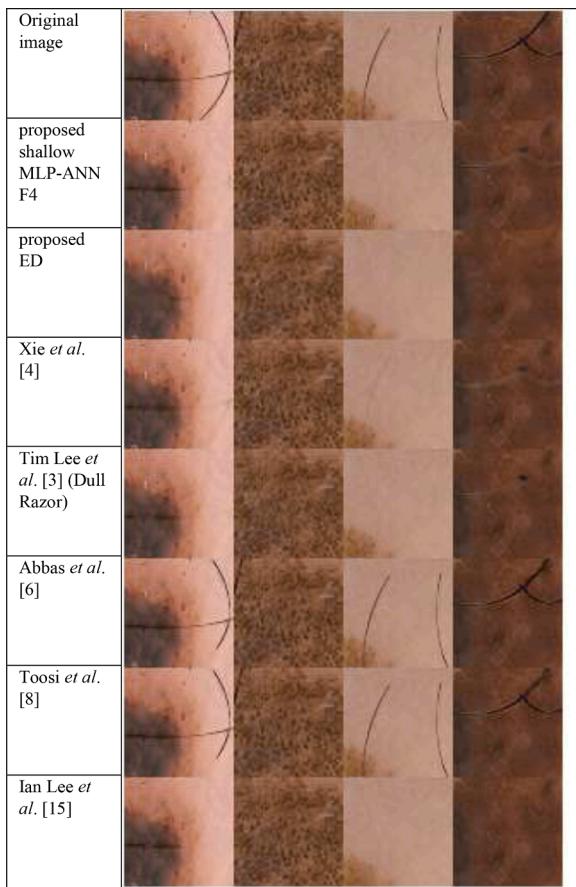


Fig. 7. Details from four (4) original images of the GAN-test dataset (1st row) and the result of the application of some of the methods under comparison (from 2nd to bottom row): proposed shallow MLP-ANN F4, proposed ED, Xie *et al.* [4], Tim Lee *et al.* [3], Abbas *et al.* [6], Toosi *et al.* [8], Ian Lee *et al.* [15].

the best performer in 5 out of 9 error metrics with statistical significance. In the remaining 4 metrics it is the best performer in 3 and the second-best performer only in terms of PSNR-HVS-M (after the Xie *et al.* method), although without statistical significance. The MLP-ANN with the F4 feature vector achieves better values than the other state-of-the-art methods in 4 out of 9 metrics, 2 of which with statistical significance (VIF and MSSSIM), but it is consistently outperformed in terms of all error metrics by the proposed ED method, in 4 of which with statistical significance (MSE, SSIM, PSNR, RMSE). The method of Ian Lee [15] is consistently inferior to the proposed methods, whereas it outperforms VS [12] and Xie [4] in terms of SSIM, VIF, UQI and MSSSIM.

[Fig. 7](#) shows details of four (4) original images of the GAN-test dataset (1st row) and the result of the application of some of the methods under comparison (from 2nd to bottom row): proposed MLP-ANN F4, proposed ED, Xie *et al.* [4], Tim Lee *et al.* [3], Ian Lee *et al.* [15], Abbas *et al.* [6], Toosi *et al.* [8]. It becomes evident that the proposed methods preserve image detail and generate a more natural-looking image. This is consistent with the numeric results of [Table 5](#), where the proposed ED outperforms all methods in every metric, with the exception of the PSNR-HVS-M against Xie *et al.*, although without statistical significance. The proposed MLP-ANN outperforms the state-of-the-art in terms of SSIM, MSSSIM and VIF. These error metrics are reported to correlate well with the human visual system and the assessment of human observers, better than other error metrics such as MSE, RMSE and PSNR.

3.2.3. Effect on melanoma image classification performance

Next, we consider the effect of the dermoscopic image hair removal method on the performance of melanoma image classification. To this end, we used the 200-image melanoma classification dataset as described in subsection III.A to perform image classification into two classes: “melanoma” and “nevus”. We applied both the proposed hair removal algorithms (shallow MLP-ANN and deep learning ED) to all 200 images of the melanoma classification dataset and generated the hair-removed dataset.

The total of 200 images were split by a 60/10/30 ratio with 120 images set for training, 20 images for validation and 60 images for testing and the EfficientNet-B0 [53] was selected as the image classifier. Transfer Learning from ImageNet was utilized due to the small size of the dataset. All layers of the neural architecture were allowed to train, whereas in the last densely connected layer, we increased 10-fold the

Table 7

Statistical significance for the improvement of melanoma image classification achieved by the two proposed hair removal methods with respect to the original images.

Comparison	Stat	p-value
Original – Hair free, shallow MLP - ANN	t-test	0.0380
Original – Hair free, Deep ED	Wilcoxon	9.84e-04

weight on the learning rate. Cross-entropy was used as the loss function. The training was allowed up to a maximum of 50 epochs. The same procedure was repeated on the hair-removed datasets by the MLP-ANN and ED methods.

To account for the statistical variations, the training for each dataset was repeated 10 times. The average accuracy achieved using the original dataset was 82.83%, while hair-removed images using the MLP-ANN and ED model achieved increased average accuracy of 87% and 90%

respectively. Fig. 10 provides the box plot to compare the aforementioned image classification accuracies. The Anderson-Darling test was applied to the 10 validation-accuracy values and confirmed the hypothesis that the values came from a population with a normal distribution for the Original and MLP-ANN case. It is worth noting that we chose to evaluate the impact of hair removal in the case of training with a rather limited dataset containing a significant number of hairy dermoscopy images, using transfer learning, to make the effect of hair removal more apparent, rather than including the full dataset which contains a large proportion of hair-free images. This explains the rather low measured accuracy values.

Finally, we applied the two-sample t-test on the image classification accuracies achieved using the original images, versus the accuracy using the hair-removed images using the proposed shallow network, as a decision test for the null hypothesis that the two sets come from normal distributions with equal means without assuming that they have equal variances. For the comparison between accuracies from the original

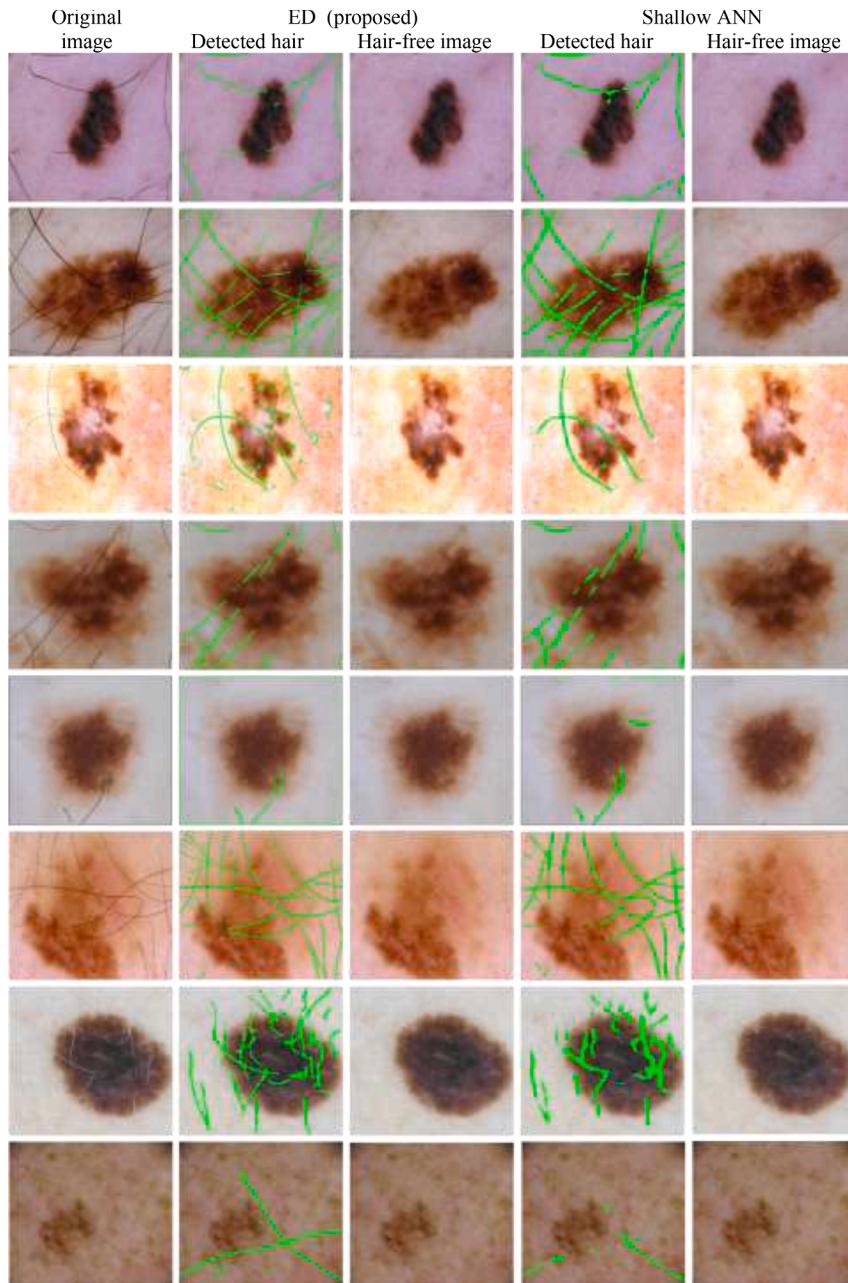


Fig. 8. Typical examples of the resulting hair removal from dermoscopic images using the proposed approaches.

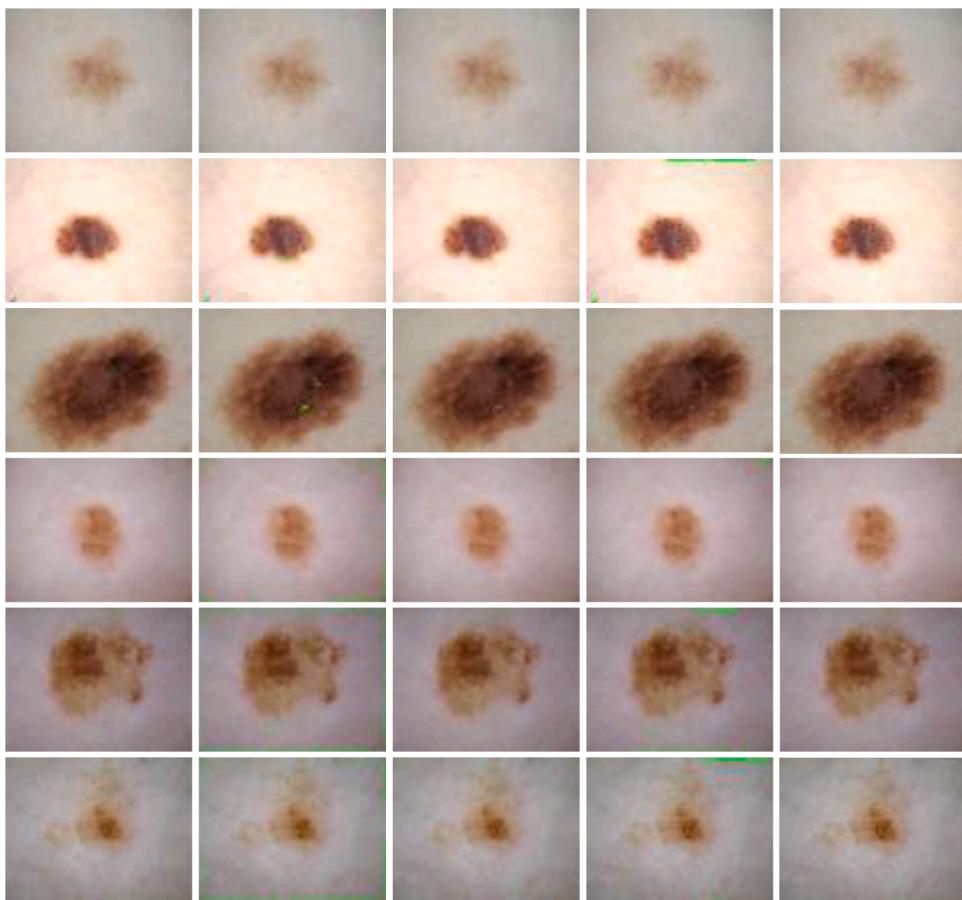


Fig. 9. Typical results of the application of the MLP-ANN on dermoscopic images with few hair pixels.

images and the proposed ED the left-tail Wilcoxon rank sum test was used. The resulting p-values for the null hypothesis at the 5% significance level are given in Table 7. The very low p-value achieved for the comparison between the original and the ED hair-removed dataset, definitely rejects the null hypothesis, thus proving that the application of the ED hair removal is beneficial for this task. Similarly, the comparison between the melanoma classification accuracy by the original dataset and the MLP-ANN hair-removed dataset yields p-value that marginally also rejects the null hypothesis.

3.3. Qualitative results

The proposed algorithms (MLP-ANN with F4 and ED) were applied to a few real dermoscopic images randomly chosen from the ISIC dataset, typical examples of which are shown in Fig. 8. The last two rows contain images with white hair. The 2nd and 4th column display the original images with the hair detected by the proposed MLP-ANN and the proposed ED respectively. Green pixels were finally accepted as hair by the post-process step. The 3rd and 5th column show the final RGB image, after hair removal by both approaches.

To further test the robustness of the proposed methods in terms of false positive responses, they were also applied on dermoscopy images with few, or no hair. As it can be observed in Fig. 9, both methods activate their output for very few pixels, even for images with an extensive pigment network that resembles hairs. Thus, the generated hair-free image is minimally altered.

4. Conclusions

The combination of image features extracted using the linear scale-space with the shallow MLP-ANN has proven very accurate and

effective in identifying hair in digital dermoscopy images. Conceptually, the calculation of a feature vector for each pixel that consists of the response of several convolution operators and maximum of eigen-analysis operators and its combination with traditional ANN carries some of the ingredients of modern deep learning architectures, although on a greatly reduced scale. The small size and easy application of the proposed MLP-ANN make it an attractive approach for fast and accurate

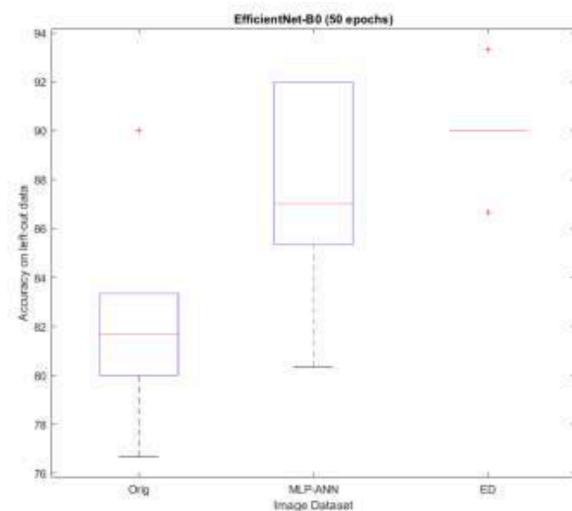


Fig. 10. Comparison of the image classification accuracy achieved using the original images and the hair-free images using the proposed shallow and deep ED architecture.

hair segmentation in dermoscopy images. On the other hand, the deep learning DeepLabv3+ Encoder-Decoder (ED) was also trained with the same dataset to segment hair from dermoscopy images using transfer learning. The proposed methods produce visually plausible results when removing hair from dermoscopy images, as well as a markedly low number of false-positive responses in hair-free images. The MLP-ANN based hair removal outperforms with statistical significance the other state-of-the-art methods in terms of error metrics that correlate well with image quality and human perception of images, namely SSIM, VIF, MSSSIM. This fact can also be visually confirmed in Fig. 7. In terms of less qualitative error metrics, such as RMSE, or PSNR, the proposed method scores 2nd behind Xie's method [4], although, without statistical significance. Apart from the single error metric of PSNR-HVS-M, the proposed ED based hair removal outperformed all methods in the GAN-test dataset.

Considering the melanoma image classification assessment, the application of the proposed MLP-ANN and ED removal methods increase the baseline average accuracy from 82.83% (on the original dataset) to 87% and 90% respectively. Statistical analysis showed strong statistical significance for the ED and marginal statistical significance for the MLP-ANN hair removal method.

Future work includes the investigation of other image features, such as image response to Gabor, or Wavelets filter-banks, that may enhance further the achieved performance.

Funding

This research has been co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (project code: Transition - T1EDK-01385).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to thank Talavera-Martínez L. from the SCOPIA Research Group, Universitat de les Illes Balears, Palma, Spain, for the provision of the hair removal benchmark & code applied in this work along with the fruitful discussions. Also, we thank Mohamed H. Attia from the Biomedical Engineering Department, Medical Research Institute, Alexandria University, Egypt, for providing the GAN-based simulated hair images.

References

- [1] T. Goudas, I. Maglogiannis, An advanced image analysis tool for the quantification and characterization of breast cancer in microscopy images, *J. Med. Syst.* 39 (3) (Mar. 2015) 1–13, <https://doi.org/10.1007/s10916-015-0225-3>.
- [2] I. Maglogiannis, C.N. Doukas, Overview of advanced computer vision systems for skin lesions characterization, *IEEE Trans. Inf. Technol. Biomed. Publ. IEEE Eng. Med. Biol. Soc.* 13 (5) (Sep. 2009) 721–733, <https://doi.org/10.1109/TITB.2009.2017529>.
- [3] T. Lee, V. Ng, R. Gallagher, A. Coldman, D. McLean, DullRazor: a software approach to hair removal from images, *Comput. Biol. Med.* 27 (6) (Nov. 1997) 533–543, [https://doi.org/10.1016/s0010-4825\(97\)00020-6](https://doi.org/10.1016/s0010-4825(97)00020-6).
- [4] F.Y. Xie, S.Y. Qin, Z.G. Jiang, R.S. Meng, PDE-based unsupervised repair of hair-occluded information in dermoscopy images of melanoma, *Comput. Med. Imaging Graph. Off. J. Comput. Med. Imaging Soc.* 33 (4) (Jun. 2009) 275–282, <https://doi.org/10.1016/j.compmedimag.2009.01.003>.
- [5] K. Kiani, A.R. Sharafat, E-shaver: an improved DullRazor(®) for digitally removing dark and light-colored hairs in dermoscopic images, *Comput. Biol. Med.* 41 (3) (Mar. 2011) 139–145, <https://doi.org/10.1016/j.combiomed.2011.01.003>.
- [6] Q. Abbas, M.E. Celebi, I.F. García, Hair removal methods: a comparative study for dermoscopy images, *Biomed. Signal Process. Control* 6 (4) (Oct. 2011) 395–404, <https://doi.org/10.1016/j.bspc.2011.01.003>.
- [7] A. Huang, S.Y. Kwan, W.Y. Chang, M.Y. Liu, M.H. Chi, G.S. Chen, A robust hair segmentation and removal approach for clinical images of skin lesions, *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.* 2013 (2013) 3315–3318, <https://doi.org/10.1109/EMBC.2013.6610250>.
- [8] M.T.B. Toossi, H.R. Pourreza, H. Zare, M.H. Sigari, P. Layegh, A. Azimi, An effective hair removal algorithm for dermoscopy images, *Skin Res. Technol. Off. J. Int. Soc. Bioeng. Skin SBS Int. Soc. Digit. Imaging Skin ISDBC Int. Soc. Skin Imaging ISSI* 19 (3) (Aug. 2013) 230–235, <https://doi.org/10.1111/srt.12015>.
- [9] A. Sultana, I. Dumitracihe, M. Vocurek, M. Ciuc, Removal of artifacts from dermatoscopic images, in: 2014 10th Int. Conf. Commun. COMM, 2014, <https://doi.org/10.1109/ICCOMM.2014.6866757>.
- [10] I. Maglogiannis, K. Delibasis, Hair removal on dermoscopy images, in: 2015 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBC, 2015, <https://doi.org/10.1109/EMBC.2015.7319013>.
- [11] C.V. Niño Rondón, D. Andrés Castellano Carvajal, B.M. Delgado, S. Alexander Castro Casadiego, D.G. Ibarra, Body hair noise suppression in skin lesions by differential operator and maximum variance threshold between classes, in: 2021 IEEE 2nd International Congress of Biomedical Engineering and Bioengineering (CI-BI BI), 2021, pp. 1–4, <https://doi.org/10.1109/CI-BIBI54220.2021.9626111>.
- [12] J. Koehoorn et al., “Automated digital hair removal by threshold decomposition and morphological analysis,” 2015. doi: 10.1007/978-3-319-18720-4_2.
- [13] P. Bibiloni, M. González-Hidalgo, S. Massanet, Skin hair removal in dermoscopic images using soft color morphology, in: Artificial Intelligence in Med. Cham. (2017) 322–326, https://doi.org/10.1007/978-3-319-59758-4_37.
- [14] M. Attia, M. Hossny, H. Zhou, S. Nahavandi, H. Asadi, A. Yazdabadi, Digital hair segmentation using hybrid convolutional and recurrent neural networks architecture, *Comput. Methods Programs Biomed.* (2019), <https://doi.org/10.1016/J.CMPB.2019.05.010>.
- [15] I. Lee, X. Du, B. Anthony, Hair segmentation using adaptive threshold from edge and branch length measures, *Comput. Biol. Med.* 89 (2017) 314–324.
- [16] R. Kasmi, J. Hagerty, R. Young, N. Lama, J. Nepal, J. Miinch, R.J. Stanley, SharpRazor: automatic removal of hair and ruler marks from dermoscopy images, *Skin Res. Technol.* 29 (4) (2023) e13203.
- [17] N. Lama, R. Kasmi, J.R. Hagerty, R.J. Stanley, R. Young, J. Miinch, W.V. Stoecker, ChimeraNet: U-net for hair detection in dermoscopic skin lesion images, *J. Digit. Imaging* (2022) 1–10.
- [18] O. Ronneberger, P. Fischer, and T. Brox, U-net: convolutional networks for biomedical image segmentation. [Online]. Available: <https://lmb.informatik.uni-freiburg.de/>.
- [19] F.I. Diakogiannis, F. Waldner, P. Caccetta, C. Wu, ResUNet-a: a deep learning framework for semantic segmentation of remotely sensed data, *ISPRS J. Photogramm. Remote Sens.* 162 (2019) 94–114, <https://doi.org/10.1016/j.isprsjprs.2019.02.001>.
- [20] M.H. Attia, M. Hossny, H. Zhou, A. Yazdabadi, H. Asadi, and S. Nahavandi, “Realistic hair simulator for skin lesion images using conditional generative adversarial network,” 2018. doi: 10.20944/PRPRINTS201810.0756.V1.
- [21] P. Isola, J.Y. Zhu, T. Zhou, and A.A. Efros, “Image-to-image translation with conditional adversarial networks,” in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, Jul. 2017, pp. 5967–5976. doi: 10.1109/CVPR.2017.632.
- [22] L. Talavera Martínez, P. Bibiloni, M. González Hidalgo, Comparative study of dermoscopic hair removal methods, in: ViPIMAGE 2019, Cham. (2019) 12–21, https://doi.org/10.1007/978-3-03-32040-9_2.
- [23] H. Mirzaalian, T. Lee, G. Hamarneh, Hair enhancement in dermoscopic images using dual-channel quaternion tubularness filters and MRF-based multilabel optimization, *IEEE Trans. Image Process.* 23 (12) (2014), <https://doi.org/10.1109/TIP.2014.2362054>.
- [24] Zhou Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, *IEEE Trans. Image Process.* 13 (4) (2004) 600–612, <https://doi.org/10.1109/TIP.2003.819861>. Apr.
- [25] H.R. Sheikh, A.C. Bovik, Image information and visual quality, *IEEE Trans. Image Process.* 15 (2) (2006) 430–444, <https://doi.org/10.1109/TIP.2005.859378>. Feb.
- [26] Zhou Wang, A.C. Bovik, A universal image quality index, *IEEE Signal Process. Lett.* 9 (3) (2002) 81–84, <https://doi.org/10.1109/97.995823>. Mar.
- [27] Z. Wang, E.P. Simoncelli, A.C. Bovik, Multiscale structural similarity for image quality assessment, The Thirty-Seventh Asilomar Conference on Signals, Syst. Comput. 2 (2003) 1398–1402, <https://doi.org/10.1109/ACSSC.2003.1292216>. Nov. 2003 Vol.2.
- [28] N.N. Ponomarenko, F. Silvestri, K.O. Egiazarian, M. Carli, J. Astola, and V. Lukin, “On between-coefficient contrast masking of DCT basis functions,” in Proceedings of the Third International Workshop on Video Processing and Quality Metrics for Consumer Electronics, VPQM 2007, Scottsdale, Arizona, 2007, p. 4. Accessed: Jul. 29, 2020. [Online]. Available: [https://tutcris.tut.fi/portal/en/publications/on-betweencoefficient-contrast-masking-of-dct-basis-functions-\(821cf0f4-6376-49d3-9aa9-34616cd1128\)/export.html](https://tutcris.tut.fi/portal/en/publications/on-betweencoefficient-contrast-masking-of-dct-basis-functions-(821cf0f4-6376-49d3-9aa9-34616cd1128)/export.html).
- [29] K.O. Egiazarian, J. Astola, N.N. Ponomarenko, V. Lukin, F. Battisti, and M. Carli, “Two new full-reference quality metrics based on HVS,” in Proceedings of the Second International Workshop on Video Processing and Quality Metrics for Consumer Electronics, VPQM 2006 , Scottsdale, Arizona, USA, 2006, p. 4. Accessed: Jul. 29, 2020. [Online]. Available: [https://tutcris.tut.fi/portal/en/publications/two-new-fullreference-quality-metrics-based-on-hvs-\(8866e92c-5ea6-4293-b502-44c8ab331699\).html](https://tutcris.tut.fi/portal/en/publications/two-new-fullreference-quality-metrics-based-on-hvs-(8866e92c-5ea6-4293-b502-44c8ab331699).html).
- [30] L. Talavera Martínez, P. Bibiloni, and M. González Hidalgo, “An encoder-decoder CNN for hair removal in dermoscopic images,” ArXiv201005013 Cs, Oct. 2020, Accessed: Apr. 19, 2022. [Online]. Available: <https://arxiv.org/abs/2010.05013>.

- [31] L. Talavera-Martínez, P. Bibiloni, M. González-Hidalgo, Hair segmentation and removal in dermoscopic images using deep learning, *IEEE Access* 9 (2021) 2694–2704, <https://doi.org/10.1109/ACCESS.2020.3047258>.
- [32] D. Kim, B.W. Hong, Unsupervised feature elimination via generative adversarial networks: application to hair removal in melanoma classification, *IEEE Access* 9 (2021) 42610–42620, <https://doi.org/10.1109/ACCESS.2021.3065701>.
- [33] W. Li, A.N. Joseph Raj, T. Tjahjadi, Z. Zhuang, Digital hair removal by deep learning for skin lesion segmentation, *Pattern Recognit.* 117 (2021), 107994, <https://doi.org/10.1016/j.patcog.2021.107994>. Sep.
- [34] S.V. Georgakopoulos, K. Kottari, K. Delibasis, V.P. Plagianakos, I. Maglogiannis, Improving the performance of convolutional neural network for skin image classification using the response of image analysis filters, *Neural Comput. App.* 31 (6) (2019) 1805–1822.
- [35] S. Amiri, M.M. Movahedi, K. Kazemi, H. Parsaei, An automated MR image segmentation system using multi-layer perceptron neural network, *J. Biomed. Phys. Eng.* 3 (4) (2013) 115–122. Dec.
- [36] J.P. Serrano, A. Hernández, R. Herrera, “Color image segmentation with a hyper-conic multilayer perceptron,” in *Image and Video Technology*, Berlin, Heidelberg, 2014, pp. 360–371. doi: 10.1007/978-3-642-53842-1_31.
- [37] V.H.C. Albuquerque, A.R. de Alexandria, P.C. Cortez, J.M.R.S. Tavares, Evaluation of multilayer perceptron and self-organizing map neural network topologies applied on microstructure segmentation from metallographic images, *NDT E Int.* 42 (7) (2009) 644–651, <https://doi.org/10.1016/j.ndteint.2009.05.002>. Oct.
- [38] Yiqian Zhao, Yunlong Zan, Xiaofang Wang, Guiyuan Li, Fuzzy C-means clustering-based multilayer perceptron neural network for liver CT images automatic segmentation, in: 2010 Chinese Control and Decision Conference, 2010, pp. 3423–3427, <https://doi.org/10.1109/CCDC.2010.5498558>. May.
- [39] R. Mottaghi, et al., The role of context for object detection and semantic segmentation in the wild, in: 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 891–898, <https://doi.org/10.1109/CVPR.2014.119>. Jun.
- [40] L.C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, “Encoder-decoder with atrous separable convolution for semantic image segmentation,” ArXiv180202611 Cs, Aug. 2018, Accessed: Jan. 05, 2021. [Online]. Available: <https://arxiv.org/abs/1802.02611>.
- [41] L.C. Chen, G. Papandreou, F. Schroff, and H. Adam, “Rethinking atrous convolution for semantic image segmentation,” arXiv, arXiv:1706.05587, Dec. 2017. doi: 10.48550/arXiv.1706.05587.
- [42] L.C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs, *IEEE Trans. Pattern Anal. Mach. Intell.* 40 (4) (2018) 834–848, <https://doi.org/10.1109/TPAMI.2017.2699184>. Apr.
- [43] F. Bornemann, T. März, Fast image inpainting based on coherence transport, *J. Math. Imaging Vis.* 28 (3) (2007) 259–278, <https://doi.org/10.1007/s10851-007-0017-6>. Jul.
- [44] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778, <https://doi.org/10.1109/CVPR.2016.90>. Jun.
- [45] J. Deng, W. Dong, R. Socher, L.J. Li, K. Li, L. Fei-Fei, ImageNet: a large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248–255, <https://doi.org/10.1109/CVPR.2009.5206848>. Jun.
- [46] P. Tschanli, C. Rosendahl, H. Kittler, The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions, *Sci. Data* 5 (1) (2018), <https://doi.org/10.1038/sdata.2018.161>. Art. no. 1, Aug.
- [47] N.C.F. Codella et al., “Skin lesion analysis toward melanoma detection: a challenge at the 2017 international symposium on biomedical imaging (ISBI),” ArXiv171005006 Cs, Oct. 2017, [Online]. Available: <https://arxiv.org/abs/1710.05006>.
- [48] M. Combalia, et al., BCN20000: dermoscopic lesions in the wild, *ArXiv E-Prints* 1908 (2019) arXiv:1908.02288Aug.
- [49] K.M. Hasib, N.A. Towhid, M.R. Islam, Hsdlm: a hybrid sampling with deep learning method for imbalanced data classification, *Int. J. Cloud App. Comput. (IJCAC)* 11 (4) (2021) 1–13.
- [50] Hasib, K.M., Iqbal, M., Shah, F.M., Mahmud, J.A., Popel, M.H., Showrov, M., ... & Rahman, O. (2020). A survey of methods for managing the classification and solution of data imbalance problem. arXiv preprint arXiv:2012.11870.
- [51] T. Mendonça, P.M. Ferreira, J.S. Marques, A.R.S. Marcal, J. Rozeira, PH2 - A dermoscopic image database for research and benchmarking, in: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2013, pp. 5437–5440, <https://doi.org/10.1109/EMBC.2013.6610779>. Jul.
- [52] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.C. Chen, Mobilenetv2: inverted residuals and linear bottlenecks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 4510–4520.
- [53] M. Tan and Q.V. Le, “EfficientNet: rethinking model scaling for convolutional neural networks,” ArXiv190511946 Cs Stat, Sep. 2020, Accessed: Oct. 08, 2021. [Online]. Available: <https://arxiv.org/abs/1905.11946>.