IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 23, 2020, accepted December 11, 2020, date of publication December 24, 2020,
date of current version January 6, 2021.

*Digital Object Identifier 10.1109/ACCESS.2020.3047258*

# Hair Segmentation and Removal in Dermoscopic Images Using Deep Learning

**LIDIA TALAVERA-MARTÍNEZ, PEDRO BIBILONI, AND MANUEL GONZÁLEZ-HIDALGO, (Member, IEEE)**

Soft Computing, Image Processing and Aggregation (SCOPIA), Department of Mathematics and Computer Science, University of the Balearic Islands, E-07122 Palma, Spain
Health Research Institute of the Balearic Islands (IdISBa), E-07010 Palma, Spain

Corresponding author: Lidia Talavera-Martínez (l.talavera@uib.es)

This work was supported by the Spanish Grant FEDER/Ministerio de Economía, Industria y Competitividad under Grant AEI/TIN2016-75404-P. The work of Lidia Talavera-Martínez supported by the Ministry of Economy, Industry and Competitiveness through a program co-financed by the European Social Fund under Grant BES-2017-081264.

**ABSTRACT** Melanoma and non-melanoma skin cancers have shown a rapidly increasing incidence rate, pointing to skin cancer as a major problem for public health. When analyzing these lesions in dermoscopic images, the hairs and their shadows on the skin may occlude relevant information about the lesion at the time of diagnosis, reducing the ability of automated classification and diagnosis systems. In this work, we present a new approach for the task of hair removal on dermoscopic images based on deep learning techniques. Our proposed model relies on an encoder-decoder architecture, with convolutional neural networks, for the detection and posterior restoration of hair's pixels from the images. Moreover, we introduce a new combined loss function in the network's training phase that combines the $L_1$ distance, the total variation loss, and a loss function based on the structural similarity index metric. Currently, there are no datasets that contain the same images with and without hair, which is necessary to quantitatively evaluate our model. Thus, we simulate the presence of hair in hairless images extracted from publicly known datasets. We compare our results with six state-of-the-art algorithms based on traditional computer vision techniques by means of similarity measures that compare the reference hairless image and the one with simulated hair. Finally, the Wilcoxon signed-rank test is used to compare the methods. The results, both qualitatively and quantitatively, demonstrate the effectiveness of our model and how our loss function improves the restoration ability of the proposed model.

**INDEX TERMS** Deep neural networks, dermoscopy, hair removal, image processing, inpainting, skin lesion.

## I. INTRODUCTION

Melanoma is the most aggressive, metastatic and deadliest type of skin cancers, turning this disease into a major problem for public health. In Europe, it accounts for 1–2% of all malignant tumors [1], and its estimated mortality in 2018 was 3.8 per 100.000 men and women per year [2]. Although melanoma is still incurable, its early diagnosis is of great importance. Its early detection can prevent malignancy and increase the survival rate and the effectiveness of the treatment. Nowadays, practitioners rely on the dermoscopic evaluation for completing the clinical analysis and the diagnosis of melanoma. This practice improves the diagnostic accuracy up to 10–30% [3] compared to simple clinical observation. This in-vivo, non-invasive skin imaging technique enables the visualization of specific subsurface structures, forms and colors that are not discernible by a simple visual inspection. The diagnosis of skin lesions is mainly based on their morphological characteristics, such as an irregular shape, asymmetry and a variety of colors, along with a history of changes in size, shape, color and/or texture. However, their evaluation might be altered by the individual judgment of the observers, which depends on their experience and subjectivity [4]. Thus, in order to help physicians to obtain an early, objective, and reproducible diagnosis of skin lesions, sophisticated Computer-Aided Diagnosis (CAD) software are developed. These computational tools are designed based on clinical protocols [5]–[8] and focus mainly on image acquisition, artifact removal (hairs, bubbles, etc.), segmentation of the lesion, extraction and selection of features, and final classification of the lesion.

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Jiang.

In the pre-processing stage during the analysis of the lesions, hair removal is one of the key steps. The presence of hair in dermoscopic images usually occludes significant patterns reducing the accuracy of the system. Once the hairs are detected and removed, the next step is to estimate and restore the underlying information (i.e., color and texture patterns) of the skin pixels underneath the hair regions. Extensive previous research has been done addressing the hair removal process in dermoscopic images. To the best of our knowledge, previous works presented approaches based on traditional computer vision techniques. Thus, tackling the problem with classical generative and discriminative models, which rely on hand-crafted features. However, hand-crafted features needed to be defined and their impact is typically tested with small datasets [9]. In recent years, deep learning has shown to be a powerful tool for image analysis. More specifically, deep learning techniques have achieved a higher performance with respect to traditional approaches for the majority of applications within the medical field [10]. Deep Convolutional Neural Networks (CNNs) allow to automatically learn features of different complexity directly from data through a set of trainable filters. Moreover, it has shown to be a powerful tool when working with large datasets.

In the preprocessing step, hair removal stands out as one of the most useful and used methods. However, traditional approaches are still used for this task in more advanced systems in which the main model uses deep learning techniques [11], [12]. Thus, we face the task of developing a deep learning model for the detection and removal of hairs. Such model could be integrated into a complete CAD system based entirely on deep learning. Our objective is threefold. First, to design a novel model based on deep CNNs for the removal of skin hair in dermoscopic images and the subsequent restoration of the affected pixels. Second, to qualitative and quantitatively assess its performance. Third, to compare the results with other hair removal strategies using the same database, which would provide an objective comparison of the strengths and weaknesses, and therefore of the quality of the method presented.

The contributions of this work are three-fold:

- To the best of our knowledge, we are the first to use deep learning techniques for hair removal in dermoscopic images.
- We introduce a loss function for the detection and posterior restoration of hair's pixels based on the combination of loss functions that focus on different aspects, which complement each other towards a more robust reconstruction.
- We extend the dataset presented in [13] by creating more synthetic hair on dermoscopic images. This eases the quantitative evaluation of hair removal approaches. We will make the data available for further research in the field.

The rest of the document is structured as follows. First, in Section II, we review the related work on hair removal methods in dermoscopic images. Then, in Section III, we present our method based on CNNs and detail the loss function that we have used to train the network. Next, in Section IV, we describe the database and the implementation details, as well as presenting the results obtained, and performing an ablation study of the loss terms and some architecture's aspects. Finally, in Section V, we discuss the previous results and study the strengths and limitations of this study.

## II. RELATED WORK

In this section, we describe previous works that addressed the task of hair removal in dermoscopic images. To the best of our knowledge, only traditional computer vision approaches have been used to address this task. In addition to describing these methods, we will use them, in Section IV, to evaluate the performance of our model.

Several traditional computer vision approaches have been used for hair removal in dermoscopic images. Here we highlight six state-of-the-art algorithms, based on their availability and wide use in the literature. These are the ones proposed by Lee *et al.* [14], Xie *et al.* [15], Abbas *et al.* [16], Huang *et al.* [17], Toossi *et al.* [18] and Bibiloni *et al.* [19]. A summary of the approaches considered by each of them can be seen in Table 1.

Nowadays, deep learning techniques have shown their potential when addressing computer vision tasks and have shown to be the state-of-the-art for many problems. Specifically, deep learning-based image restoration techniques have been used in other fields for image inpainting [20], image deblurring and image denoising [21], among others. These methods learn the parameters of the network to reconstruct images directly from training data, that is composed by pairs of clean and corrupted images. This is usually more effective in real-world images. For instance, Xie *et al.* [20] proposed an approach for image denoising and blind inpainting that combines sparse coding with pre-trained deep networks, achieving good results in both tasks. Vincent *et al.* [22] presented a stack of denoising autoencoders for image denoising that is applied not only to the input, but also recursively to intermediate representations, to initialize the deep neural network. Also, Cui *et al.* [23] proposed a cascade of multiple stacked collaborative local autoencoders for image super-resolution. Their method searches in each layer non-local self-similarity to enhance high frequency texture details from the image patches to suppress the noise and combine the overlapping patches. In [24], Mao *et al.* proposed an encoding-decoding framework for image denoising and super-resolution combining convolutions and deconvolution layers linked symmetrically by skip connections, which helps improving the training process and the network's performance. Finally, Jain and Seung [25] and Dong *et al.* [26] proposed a fully convolutional CNN for image denoising and image super-resolution, respectively. The authors shown that their methods achieve comparable results to traditional computer vision techniques.

**TABLE 1. Detection and inpainting techniques employed in the literature to remove hair from dermoscopic images.**

| Year | Method | Hair segmentation | Inpainting method | Color space |
|---|---|---|---|---|
| 1997 | Lee *et al.* [14] | Grayscale closing | Bilinear interpolation | RGB |
| 2009 | Xie *et al.* [15] | Grayscale top-hat | Non-linear PDE | RGB |
| 2011 | Abbas *et al.* [16] | Derivative of Gaussians | Coherence transport | CIELab |
| 2013 | Huang *et al.* [17] | Conventional matched filters | Region growing algorithms and linear discriminant analysis | RGB |
| 2013 | Toossi *et al.* [18] | Canny edge detector | Coherence transport | RGB |
| 2017 | Bibiloni *et al.* [19] | Color top-hat | Morphological inpainting | CIELab |

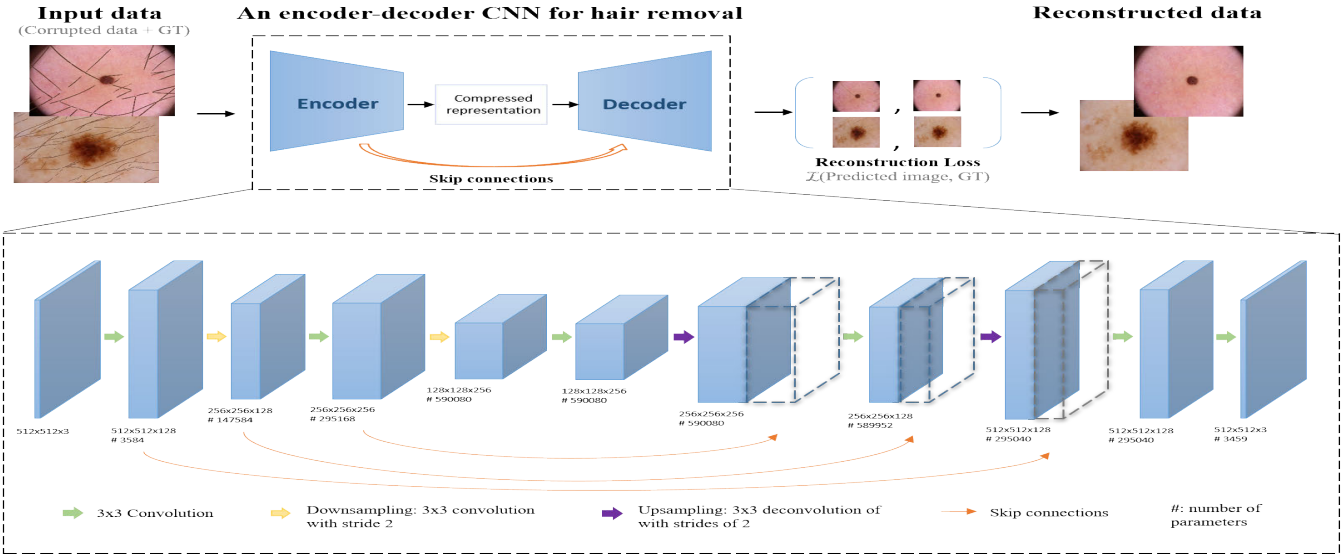


**FIGURE 1. Architecture of our proposed network. The pairs of reference hairless images (GT) and its corrupted (hair simulated) images are passed through the encoder to extract complex features. The decoder, connected to the encoder with skip-connections, reconstructs the image.**

When focusing on the hair removal process on dermoscopy images and deep learning techniques, Attia *et al.* [27] used them to simulate hair. The authors used Generative-Adversarial-Networks to generate fake hair and add it to the image. In this work, we address the inverse problem, including the location and reconstruction of the hair regions in dermoscopic images using CNNs. Given the promising results achieved by deep learning models for other computer vision related tasks and the need for robust models for hair removal in dermoscopic images, we present a novel model that relies on an autoencoder to address this task.

## III. METHODOLOGY

In this section, we describe our proposed deep learning model for hair removal in dermoscopic images and depict the introduced reconstruction loss function.

We designed and propose a convolutional encoder-decoder architecture for the task of hair removal in dermoscopic images. Our model, which is detailed in Figure 1, is composed of 12 layers. To train our model, we use pairs of images formed by the reference image without hair and its corresponding image with simulated hair. The output is the reconstructed dermoscopic without hairs on it. During the learning process, the network relies on our proposed loss function that evaluates the goodness of the output image in comparison to the hairless reference image. Next, we describe the proposed model and loss functions for this relevant task.

### A. ARCHITECTURE STRUCTURE

The first module of our model is an encoder network. The input is of size $512 \times 512 \times 3$ and its output is a hidden representation of high-level features of the input image. When looking for such features, the encoder tends to ignore noise. In our case, our hypothesis is that the network will treat the hair as noise and will be ignored, having as output the hairless skin image. The second layer of the model corresponds to the decoder, which aims to recover the missing information from the high-level feature representation. Its output is a $512 \times 512 \times 3$ cleaned version, without skin hair, of the input image. Both the encoder and the decoder have two blocks. Each block of the encoder consists of one $3 \times 3$ convolution, of 128 filters in the first block and 256 filters in the second one, followed by a down-sampling operation, which is applied by a two-strided $3 \times 3$ convolution to reduce the spatial resolution. On the other side, each block of the decoder consists of an up-sampling of the feature map by a deconvolution of $3 \times 3$ with strides of two in each dimension. A skip connection follows, which concatenates the up-sampled output with the corresponding feature map

from the layer of equal resolution of the encoder. This enables the decoder to recover image details, and therefore improves the restoration performance. Next, a $3 \times 3$ convolution is applied over the merged data. Finally, in the last block, an additional $3 \times 3$ convolution is added to reduce the feature map to the number of output channels.

The choice of this architecture resides in the aim of evaluating the suitability of autoencoders to tackle this task as a denoising problem. In terms of its size, we believe that a rather small network is more suitable to correctly learn the task due to the amount of data we have available.

### B. RECONSTRUCTION LOSS FUNCTION

The loss function guides how the network learns by serving as a criterion that numerically reflects the errors of the model. It is computed between the network output and its corresponding hairless reference image, also known as Ground Truth (GT). There are several loss functions that have been used in image restoration tasks. Some widespread losses are the Mean Square Error (MSE) or the Mean average Error (MAE). These measurements exclusively depend on the difference between the corresponding pixels of the two images. Therefore, the results might have poor quality in terms of human perception, since the noise of a pixel should not be considered independently of the error of its neighbouring pixels. To overcome these limitations, other loss metrics have been proposed, such as the Structural Similarity Metric (SSIM) or the Multiscale Structural Similarity Metric (MSSSIM), which depend on local luminance, contrast and structure [28].

To achieve results appealing to a human observer, and inspired by the results obtained by Liu *et al.* in [29], we propose to capture the best characteristics of the loss functions that measure statistical features locally along with other per-pixel losses.

Thus, our reconstruction loss is defined as follows:

$$\mathcal{L}_{\text{rec}} = \alpha L_1^{\text{foreground}} + \beta L_1^{\text{background}} + \gamma L_2^{\text{composed}} + \delta \mathcal{L}_{\text{SSIM}} + \lambda \mathcal{L}_{\text{tv}}, \quad (1)$$

where $\alpha, \beta, \gamma, \delta$ and $\lambda$ are the weights of each term of the linear combination that define the reconstruction loss function. We opted to perform a random hyperparameter search as there are many parameters of which we have to find the optimal value, and a grid search would require a higher computational cost. Specifically, we performed 10 runs of our model, assigning in each case a random value between 0 and 10 to each of the weights. Afterwards, a statistical test indicates which are the best set of weights according to the measurements explained in Section IV-D.

- The term $L_1^{\text{foreground}}$ is the $L_1$ distance between the original hairless image and the prediction of the network only between those pixels belonging to the hair areas.
- Next, $L_1^{\text{background}}$ estimates the $L_1$ distance between the original hairless image and the network's prediction only among the background pixels, which in our context accounts for the non-hair regions.
- Then, $L_2^{\text{composed}}$ computes the $L_2$ restricted to the hair regions, but normalizing over all pixels rather than over the amount of hair pixels.
- The term $\mathcal{L}_{\text{SSIM}}$ calculates the loss function based on the SSIM metric over the whole image.
- Finally, we use a total variation loss, $\mathcal{L}_{\text{tv}}$, as a regularizer to smooth the transition of the predicted values for the regions corresponding to hair, according to their surrounding context. A more detailed description of this term can be found in [30].

## IV. EXPERIMENTAL FRAMEWORK AND RESULTS

In this section, we first establish the experimental framework by describing the database used and the implementation details of our method. Then, we analyze the results obtained by our method and compare them, qualitative and quantitatively, to the six traditional hair removal methods presented in Section II. To obtain the numerical results, we rely on several performance measures. We determine which method outperforms the rest by means of a statistical test. Finally, we conducted an ablation study of the loss terms, as well as of some aspects of the model's architecture.

### A. DATASET DESCRIPTION

In order to train the CNN and to quantitatively validate, in an effective way, the performance of our method, we need a dataset. It must contain pairs of images: images with hair, used as the algorithm input, along with their corresponding "clean" version, in this case the same image without hair. If we only had the image with hairs, we could only do a qualitative evaluation.

Finding this type of data is a challenging task, the same dermoscopic image can not be captured with and without hair. To tackle this problem, we decided to simulate the presence of skin hair in hairless images extracted from five publicly available datasets, *i.e.* PH2 [31], dermquest,[1] dermis,[2] EDRA2002 [32] and from the ISIC Data Archive.[3] We have avoided selecting images with other artifacts (eg. ruler, bandages etc) that are not hairs. Three different hair simulation methods have been used. The first is the one, presented by Attia *et al.* [27], is based on generative adversarial networks. The second one was implemented by Mirzaalian *et al.* [33], whose software "HairSim" is publicly available in [34]. Finally, the last approach we have used involved extracting hair masks by an automated method, proposed by Xie *et al.* [15], and superimpose them on hairless images.

We constructed a dataset with 618 images, which consists of 322 images from the EDRA2002 dataset, 239 images from the PH2 dataset, 46 images from the ISIC Data

[1] Was deactivated on December 31, 2019
[2] www.dermis.net
[3] www.isic-archive.com

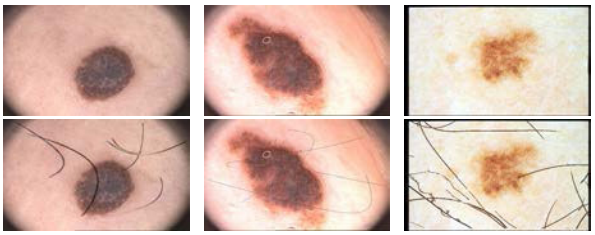

**FIGURE 2.** Original images (top) and simulated hair images (bottom), respectively by a deep neural network [27] (left), "HairSim" [34] (middle), superimposing a hair mask [15] (right).

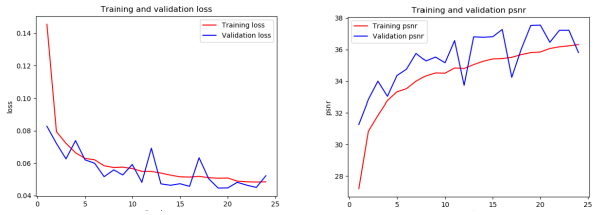


**FIGURE 3.** Plots of the Loss (left), and PSNR performance measure (right), of the training of the proposed model.

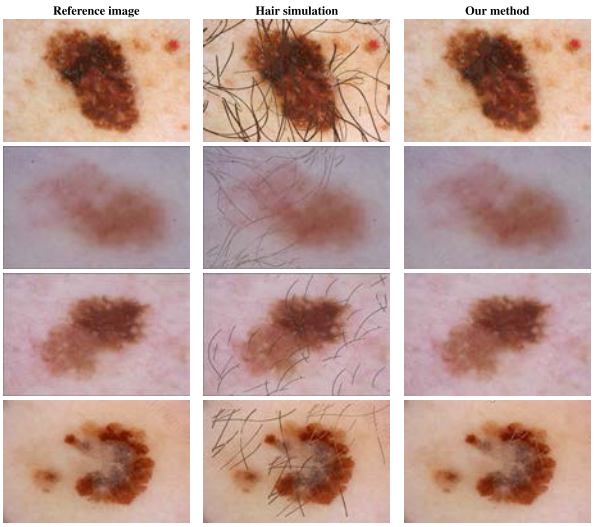


**FIGURE 4.** Example of the hair removal results obtained by our method.

Archive, 6 images from the dermis dataset, and 5 from the dermquest dataset. During the experimentation we divide it into 70% for training and 30% for the test, which gives us 433 and 185 images, respectively. It is composed of images with diverse hair that present a variety of hair thickness, density and color, ranging from coarse to more realistic. This diversity guarantees that we consider hairs with very different characteristics when training the network. We also introduced images without hair to teach the network that some images must not be modified and their textures must be maintained. In Figure 2, we show examples of original hairless images from the dataset, together with the simulations obtained using each of the hair simulation approaches. As we can see, with the first and third method we achieve a much more natural morphology, quantity and distribution that resembles real hair compared to the second. We consider a reduced number of testing images to maximize the training samples and, thus, to help the network learn and generalize well.

### B. EXPERIMENTAL SETUP

We implemented the proposed architecture using Keras [35]. The network was trained from scratch with randomly initialized weights and using the Adam [36] optimizer with a learning rate experimentally set to $10^{-4}$. The coefficients for the different terms of the reconstruction loss function with which the network has been trained were experimentally found to be: $\alpha = 2.626$, $\beta = 3.892$, $\gamma = 0.309$, $\delta = 0.398$ and $\lambda = 0.597$. The network was trained on a single NVIDIA GeForce GTX 1070 with a batch size of 4.

In figure 3, we can see how our model trained over almost 25 epochs (reaching an early stopping policy based on monitoring the validation loss), and how the Peak Signal-to-Noise Ratio (PSNR) metric evolves satisfactory during the training.

### C. QUALITATIVE RESULTS

We conducted a qualitative and quantitative analysis of the results obtained. In Figure 4, we can see that our proposed method attains visually appealing results. In addition, we present a visual comparison of our results with the with the ones obtained by other hair removal methods presented in Section II. Let us remark that we compare with these methods as they are the ones that have been applied to dermoscopic images.

As can be seen in Figure 5, not all the methods are successful in both the hair removal and the subsequent process of inpainting. For instance, Abbas *et al.*'s and Toossi *et al.*'s methods are not capable of segment the hairs as it seems they are not able to detect them properly. In contrast, Huang *et al.*'s and Xie *et al.*'s methods are capable of detecting much of the hair. However, their inpainting process seems to leave a trail of them. Finally, our results and the ones of Bibiloni *et al.*'s and Lee *et al.*'s methods seem to adjust to the reference image at first sight. Although, the last two leave traces, while the new method does not. However, it may be the case that some of them introduce some alterations, to a greater or lesser extent, that blur the lesion's features, such as streaks or reticular textures.

Both in Figure 4 and 5, we have seen that our method reaches good visual results when evaluated on synthetic images. In Figure 6, we show its effectiveness and its generalization ability in dermoscopic images with real hair. We show images from the 5 databases, to demonstrate that although the data is not balanced, the network has not suffered database-specific overfitting.

### D. QUANTITATIVE RESULTS

It is worth noting that once the hairs are removed there are many possible solutions as to what is the expected inpainting result, always with the aim of preserving the texture of

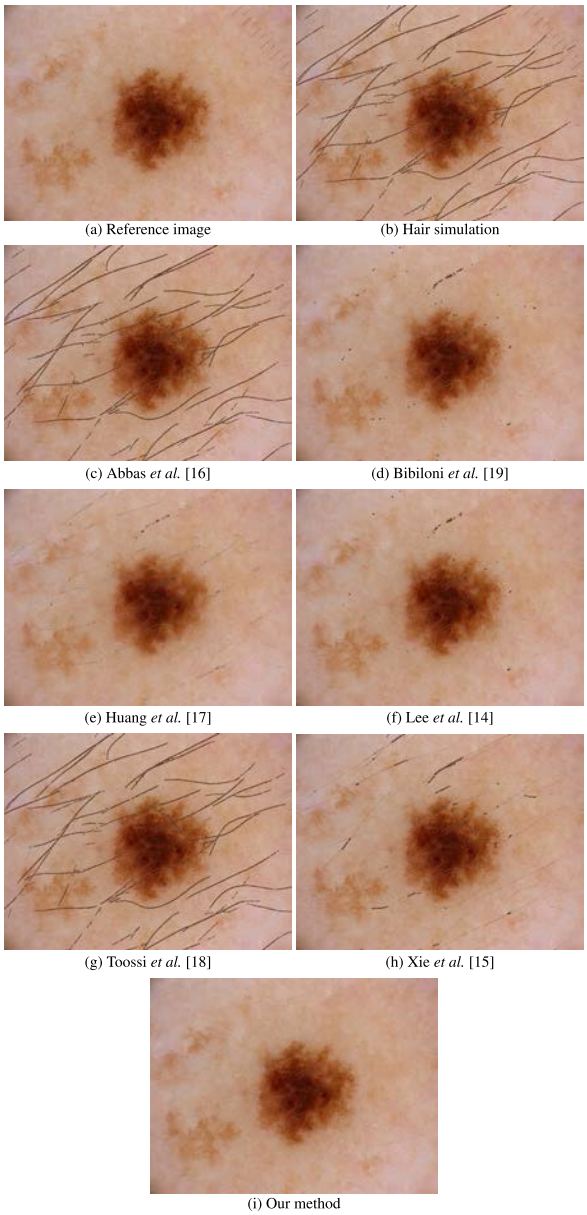


**FIGURE 5.** Given a sample image (a), we simulate hair on it (b), and present the results obtained by several state-of-the-art methods (c)-(h) against our proposed method (i).

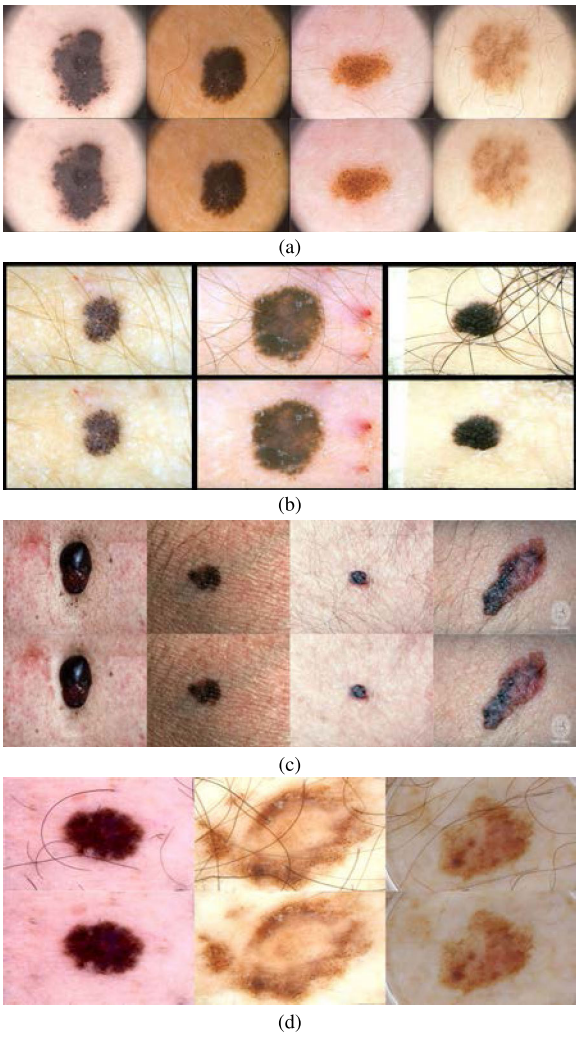


**FIGURE 6.** Example of hair removal results obtained by our method in dermoscopic images with real hair from (a) PH2 dataset, (b) EDRA2002 dataset, (c) Dermis and Dermquest datasets, and (d) ISIC Data Archive. In the first row of each subfigure we find the test sample image, while in the second row we find their corresponding output of our model.

the area involved. Therefore, a qualitative evaluation is not enough to evaluate the quality of the different methods introduced. In the following, we introduce an automatic, objective and comparable performance evaluation system.

We used a set of nine objective error metrics to quantitatively assess the quality of the results obtained by the proposed CNN-based hair removal approach, with respect to the original hairless image. We cluster these measures into three different groups. The first one are the Mean Squared Error (MSE) [37], the Peak Signal-to-Noise Ratio (PSNR) [38], the Root Mean Squared Error (RMSE) [39], and the Structural Similarity Index (SSIM) [38], which are per-pixel metrics. Within the second group we consider the Multi-Scale Structural Similarity Index (MSSSIM) [40] and the Universal Quality Image Index (UQI) [41], which measure statistical features locally and then combine them. Finally, the Visual Information Fidelity (VIF) [42], the PSNR-HVS-M [43] and PSNR-HVS [43], conforming the third group, have been designed to obtain more similar results to those perceived by the Human Visual System (HVS). This set of metrics constitutes a representative selection of the state-of-the-art performance metrics for restoration quality. We must recall that largest values of PSNR, SSIM, MSSSIM, UQI, VIF, PSNR-HVS-M, and PSNR-HVS are indicators of a better quality of reconstructed images. On the other hand, lowest values of MSE and RMSE are indicators of higher similarity.

In Table 2, we show the mean and standard deviation of the results obtained for the 185 images of the test set, and for each of the nine performance measures. In addition, we make a comparison of our results against the ones obtained

**TABLE 2.** Mean ($\mu$) and standard deviation ($\sigma$) of the similarity measures obtained to compare our method with six state of the art hair removal algorithms.

| | | Our method | Abbas [16] | Bibiloni [19] | Huang [17] | Lee [14] | Toossi [18] | Xie [15] |
|---|---|---|---|---|---|---|---|---|
| MSE | $\mu$ | 27.847 | 258.347 | 103.903 | 404.366 | 175.303 | 221.346 | 55.311 |
| | $\sigma$ | 35.261 | 302.243 | 108.762 | 589.341 | 219.633 | 256.319 | 103.553 |
| SSIM | $\mu$ | 0.926 | 0.867 | 0.885 | 0.851 | 0.890 | 0.864 | 0.921 |
| | $\sigma$ | 0.026 | 0.051 | 0.048 | 0.050 | 0.075 | 0.053 | 0.041 |
| PSNR | $\mu$ | 35.137 | 26.080 | 29.785 | 27.001 | 29.158 | 26.570 | 33.096 |
| | $\sigma$ | 3.006 | 4.079 | 4.073 | 6.603 | 5.783 | 3.852 | 3.705 |
| RMSE | $\mu$ | 4.790 | 14.226 | 9.207 | 15.485 | 11.032 | 13.287 | 6.318 |
| | $\sigma$ | 2.220 | 7.501 | 4.386 | 12.864 | 7.340 | 6.711 | 3.935 |
| VIF | $\mu$ | 0.525 | 0.526 | 0.499 | 0.402 | 0.531 | 0.509 | 0.592 |
| | $\sigma$ | 0.099 | 0.180 | 0.176 | 0.130 | 0.186 | 0.178 | 0.186 |
| UQI | $\mu$ | 0.997 | 0.991 | 0.996 | 0.990 | 0.995 | 0.992 | 0.997 |
| | $\sigma$ | 0.004 | 0.009 | 0.005 | 0.013 | 0.006 | 0.008 | 0.004 |
| MSSSIM | $\mu$ | 0.978 | 0.870 | 0.934 | 0.917 | 0.945 | 0.875 | 0.955 |
| | $\sigma$ | 0.011 | 0.073 | 0.039 | 0.042 | 0.051 | 0.069 | 0.037 |
| PSNR-HVS-M | $\mu$ | 36.802 | 25.078 | 29.404 | 26.248 | 28.445 | 25.519 | 33.005 |
| | $\sigma$ | 3.703 | 3.861 | 4.245 | 6.854 | 6.247 | 3.680 | 4.106 |
| PSNR-HVS | $\mu$ | 35.168 | 24.628 | 28.681 | 25.738 | 27.826 | 25.065 | 32.186 |
| | $\sigma$ | 3.465 | 3.823 | 4.158 | 6.634 | 5.991 | 3.639 | 3.896 |

**TABLE 3.** Classification of algorithms according to objective similarity measures. The results are as follows: ✓✓ if the population mean of the first algorithm is better than that of the second algorithm; ✓ if the mean of the first algorithm is better but statistically comparable to that of the second algorithm; ✗ if the mean of the first algorithm is worse but statistically comparable to that of the second algorithm; ✗✗ if the population mean of the first algorithm is worse than that of the second algorithm.

| | | MSE | SSIM | PSNR | RMSE | VIF | UQI | MSSSIM | PSNR-HVS-M | PSNR-HVS |
|---|---|---|---|---|---|---|---|---|---|---|
| Our method *vs.* Abbas | $p$-value | 2.62e-28 | 1.51e-29 | 1.49e-30 | 3.89e-29 | 0.452 | 1.09e-28 | 4.49e-32 | 1.87e-31 | 3.33e-31 |
| | Statistical test | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✗ | ✓✓ | ✓✓ | ✓✓ | ✓✓ |
| Our method *vs.* Bibiloni | $p$-value | 4.44e-30 | 1.29e-25 | 3.38e-31 | 1.27e-30 | 4.07e-04 | 2.05e-10 | 2.68e-30 | 1.06e-31 | 1.49e-31 |
| | Statistical test | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ |
| Our method *vs.* Lee | $p$-value | 1.21e-30 | 2.23e-17 | 4.66e-30 | 1.57e-30 | 0.160 | 8.52e-13 | 3.35e-27 | 1.81e-31 | 3.85e-31 |
| | Statistical test | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✗ | ✓✓ | ✓✓ | ✓✓ | ✓✓ |
| Our method *vs.* Huang | $p$-value | 5.54e-30 | 2.18e-30 | 1.75e-31 | 9.41e-31 | 1.46e-23 | 1.51e-29 | 1.87e-31 | 7.19e-32 | 9.63e-32 |
| | Statistical test | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ |
| Our method *vs.* Toossi | $p$-value | 1.49e-27 | 3.95e-29 | 5.12e-30 | 1.60e-28 | 1.49e-03 | 7.83e-27 | 5.28e-32 | 2.79e-31 | 5.83e-31 |
| | Statistical test | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ |
| Our method *vs.* Xie | $p$-value | 1.86e-14 | 9.59e-04 | 1.14e-14 | 1.32e-14 | 2.31e-10 | 4.05e-05 | 1.36e-25 | 1.72e-19 | 7.22e-18 |
| | Statistical test | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ |
| Abbas *vs.* Bibiloni | $p$-value | 5.24e-23 | 1.27e-12 | 1.40e-23 | 2.43e-23 | 0.012 | 8.04e-29 | 3.07e-27 | 1.39e-24 | 3.42e-24 |
| | Statistical test | ✗✗ | ✗✗ | ✗✗ | ✗✗ | ✓✓ | ✗✗ | ✗✗ | ✗✗ | ✗✗ |
| Abbas *vs.* Lee | $p$-value | 1.36e-11 | 4.70e-16 | 2.48e-15 | 7.87e-13 | 1.03e-10 | 5.69e-23 | 3.96e-26 | 1.63e-15 | 1.22e-15 |
| | Statistical test | ✗✗ | ✗✗ | ✗✗ | ✗✗ | ✗✗ | ✗✗ | ✗✗ | ✗✗ | ✗✗ |
| Abbas *vs.* Huang | $p$-value | 0.839 | 1.22e-03 | 5.1e-02 | 0.488 | 3.95-e21 | 0.118 | 1.49e-17 | 1.61e-03 | 1.56e-03 |
| | Statistical test | ✓ | ✓✓ | ✗✗ | ✓ | ✓✓ | ✓ | ✗✗ | ✗✗ | ✗✗ |
| Abbas *vs.* Toossi | $p$-value | 9.08e-28 | 3.52e-07 | 2.93e-27 | 1.27e-27 | 4.14e-32 | 4.76e-25 | 3.27e-08 | 5.66e-25 | 4.13e-25 |
| | Statistical test | ✗✗ | ✓✓ | ✗✗ | ✗✗ | ✓✓ | ✗✗ | ✗✗ | ✗✗ | ✗✗ |
| Abbas *vs.* Xie | $p$-value | 4.56e-32 | 4.21e-32 | 4.64e-32 | 4.56e-32 | 4.64e-32 | 8.74e-32 | 5.64e-32 | 4.87e-32 | 4.87e-32 |
| | Statistical test | ✗✗ | ✗✗ | ✗✗ | ✗✗ | ✗✗ | ✗✗ | ✗✗ | ✗✗ | ✗✗ |
| Bibiloni *vs.* Lee | $p$-value | 5.07e-05 | 6.25e-13 | 2.51e-02 | 3.71e-04 | 1.34e-17 | 2.05e-10 | 1.36e-15 | 2.42e-03 | 2.36e-03 |
| | Statistical test | ✓✓ | ✗✗ | ✓✓ | ✓✓ | ✗✗ | ✓ | ✗✗ | ✓✓ | ✓✓ |
| Bibiloni *vs.* Huang | $p$-value | 1.53e-13 | 9.88e-24 | 9.07e-15 | 7.50e-14 | 4.51e-17 | 1.72e-17 | 4.58e-14 | 2.30e-16 | 3.38e-16 |
| | Statistical test | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ |
| Bibiloni *vs.* Toossi | $p$-value | 7.98e-21 | 1.59e-14 | 2.37e-21 | 3.29e-21 | 0.298 | 1.56e-27 | 4.32e-27 | 5.77e-23 | 1.22e-22 |
| | Statistical test | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✗ | ✓✓ | ✓✓ | ✓✓ | ✓✓ |
| Bibiloni *vs.* Xie | $p$-value | 1.03e-18 | 4.40e-29 | 2.24e-20 | 1.66e-19 | 4.21e-32 | 5.35e-12 | 8.77e-14 | 1.74e-18 | 1.68e-19 |
| | Statistical test | ✗✗ | ✗✗ | ✗✗ | ✗✗ | ✗✗ | ✗✗ | ✗✗ | ✗✗ | ✗✗ |
| Lee *vs.* Huang | $p$-value | 1.24e-11 | 6.62e-17 | 1.61e-15 | 9.23e-13 | 4.76e-20 | 2.92e-21 | 1.44e-17 | 9.01e-16 | 1.92e-15 |
| | Statistical test | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ |
| Lee *vs.* Toossi | $p$-value | 4.13e-07 | 1.75e-16 | 9.88e-12 | 7.17e-09 | 2.16e-14 | 1.32e-20 | 4.08e-26 | 4.36e-12 | 3.29e-12 |
| | Statistical test | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ | ✓✓ |
| Lee *vs.* Xie | $p$-value | 1.11e-16 | 2.09e-10 | 1.15e-15 | 1.63e-16 | 6.04e-28 | 5.12e-06 | 0.869 | 2.22e-16 | 8.65e-17 |
| | Statistical test | ✗✗ | ✗✗ | ✗✗ | ✗✗ | ✗✗ | ✗✗ | ✗ | ✗✗ | ✗✗ |
| Huang *vs.* Toossi | $p$-value | 0.496 | 7.83e-03 | 0.133 | 0.871 | 1.70e-17 | 0.594 | 5.44e-16 | 0.050 | 5.38e-02 |
| | Statistical test | ✗ | ✗✗ | ✓ | ✗ | ✗✗ | ✗ | ✓✓ | ✓✓ | ✓ |
| Huang *vs.* Xie | $p$-value | 6.91e-25 | 7.80e-32 | 9.36e-27 | 8.70e-26 | 2.82e-28 | 2.40e-24 | 1.08e-25 | 7.17e-27 | 3.89e-27 |
| | Statistical test | ✗✗ | ✗✗ | ✗✗ | ✗✗ | ✗✗ | ✗✗ | ✗✗ | ✗✗ | ✗✗ |
| Toossi *vs.* Xie | $p$-value | 4.87e-32 | 4.27e-32 | 5.11e-32 | 4.87e-32 | 4.14e-32 | 1.35e-31 | 7.55e-32 | 1.72e-19 | 5.11e-32 |
| | Statistical test | ✗✗ | ✗✗ | ✗✗ | ✗✗ | ✗✗ | ✗✗ | ✗✗ | ✗✗ | ✗✗ |

by applying the six state of-the-art hair removal methods, detailed in Section II, to the same 185 images.

The next step in our work, is to study if one algorithm outperforms another one significantly. Given a fixed

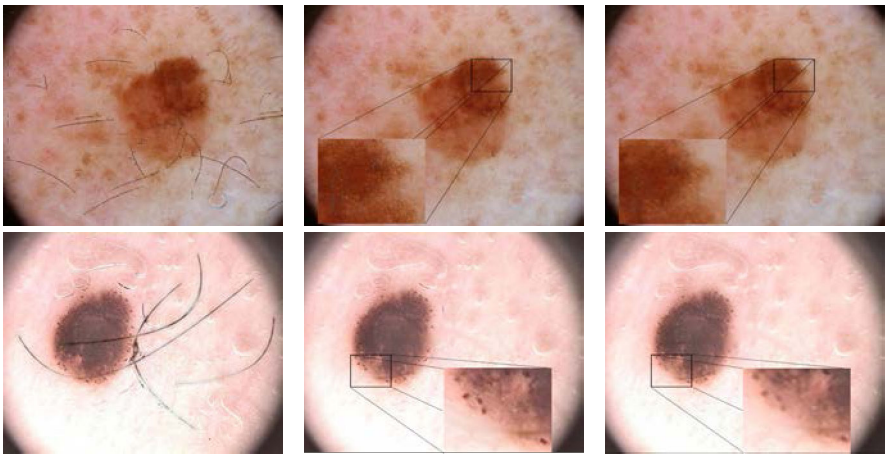

**FIGURE 7.** Hair removal results for two samples (left column), when using skip-connections (middle column), and without using them (right column).

**TABLE 4.** Mean of the similarity measures obtained on the test set images for the ablation study on our model. The second and third row correspond to the skip-connections and pooling layers' ablation study. Then, from the fourth to the last row correspond to the ablation study of the loss terms.

| | MSE | SSIM | PSNR | RMSE | VIF | UQI | MSSSIM | PSNR-HVS-M | PSNR-HVS |
|---|---|---|---|---|---|---|---|---|---|
| Full method | 27.847 | 0.926 | 35.137 | 4.790 | 0.525 | 0.997 | 0.978 | 36.802 | 35.168 |
| Model with Pooling Layers and Skip-Connections | 27.786 | 0.925 | 35.122 | 4.794 | 0.524 | 0.998 | 0.978 | 36.596 | 35.003 |
| Model with DeConv Layers and no Skip-Connections | 32.846 | 0.909 | 33.862 | 5.443 | 0.467 | 0.997 | 0.974 | 35.538 | 33.908 |
| Loss with no $L_1^{\text{foreground}}$ term | 41.197 | 0.881 | 32.719 | 6.138 | 0.437 | 0.997 | 0.954 | 33.946 | 32.717 |
| Loss with no $L_1^{\text{background}}$ term | 26.398 | 0.928 | 35.306 | 4.689 | 0.521 | 0.997 | 0.979 | 36.862 | 35.209 |
| Loss with no $\mathcal{L}_{\text{SSIM}}$ term | 31.220 | 0.926 | 34.488 | 5.119 | 0.523 | 0.997 | 0.979 | 36.802 | 35.154 |
| Loss with no $L_2^{\text{composed}}$ term | 30.881 | 0.926 | 34.463 | 5.113 | 0.523 | 0.997 | 0.978 | 36.744 | 35.143 |
| Loss with no $\mathcal{L}_{\text{tv}}$ term | 30.621 | 0.923 | 34.673 | 5.039 | 0.522 | 0.997 | 0.977 | 36.599 | 34.985 |

similarity measure, we decided to use a statistical test to contrast the means of all pairs of algorithms. Specifically, we have used the $t$-test if the samples pass the Shapiro-Wilk normality test, or the Wilcoxon signed-rank test, otherwise, both considering a significance level of 0.05. According to the statistical test we can determine which method surpasses the others. Table 3 summarizes the results obtained. In it, rows represent all the pairs of algorithms in which the statistical test was applied, and the columns correspond to the measures of similarity. As an example, let us interpret the test comparing *Abbas* vs. *Huang*. According to the SSIM and VIF performance measures, Abbas' algorithm significantly outperforms Huang's algorithm. While according to the PSNR, MSSIM, PSNR-HVS-M and PSNR-HVS measures, Huang's algorithm significantly outperforms Abbas' algorithm. For the rest of measures, both methods obtain statistically comparable results. Let us remark that Abbas' algorithm is non-statistically superior in all of them.

As it can be seen in Table 3, taking into account all the considered performance measures, the proposed method out-stands according to the majority of similarity measures. It is only significantly outperformed on the VIF performance measure compared to Abbas *et al.*' and Lee *et al.*'s algorithms. Among the rest of methods, we can see that Lee *et al.*'s algorithm surpasses statistically Huang *et al.*' and Toossi *et al.*'s algorithms. However, when comparing Lee *et al.*'s algorithm with Bibiloni *et al.*'s, the former is statistically better in very specific settings, namely the SSIM, VIF and MSSSIM measures. It is Xie *et al.*'s algorithm that outperforms Lee *et al.*'s in all performance measures. In the comparison between the algorithms of Bibiloni with Huang and Toossi, it is the first that outperforms statistically the other two in the majority of measures. Finally, the Toossi *et al.*'s algorithm is statistically superior to the Abbas *et al.*'s, except in the SSIM and VIF performance measures. These two algorithms are the ones that provide statistically worse results compared to the rest of the algorithms.

### E. ABLATION STUDY

Some works [24], [29] defend the fact that using skip connections, or convolutions/deconvolutions instead of pooling/unpooling layers may decrease the amount of detail loss and deteriorate the restoration performance. We study how these layers can affect the learning of our model by replacing the pooling layers with convolutions, and introducing skip-connection layers. In Table 4, we can see that the introduction of skip-connections does improve the results numerically, in terms of the similarity measures previously presented. However, these do not vary significantly when pooling layers are used instead of convolutional ones

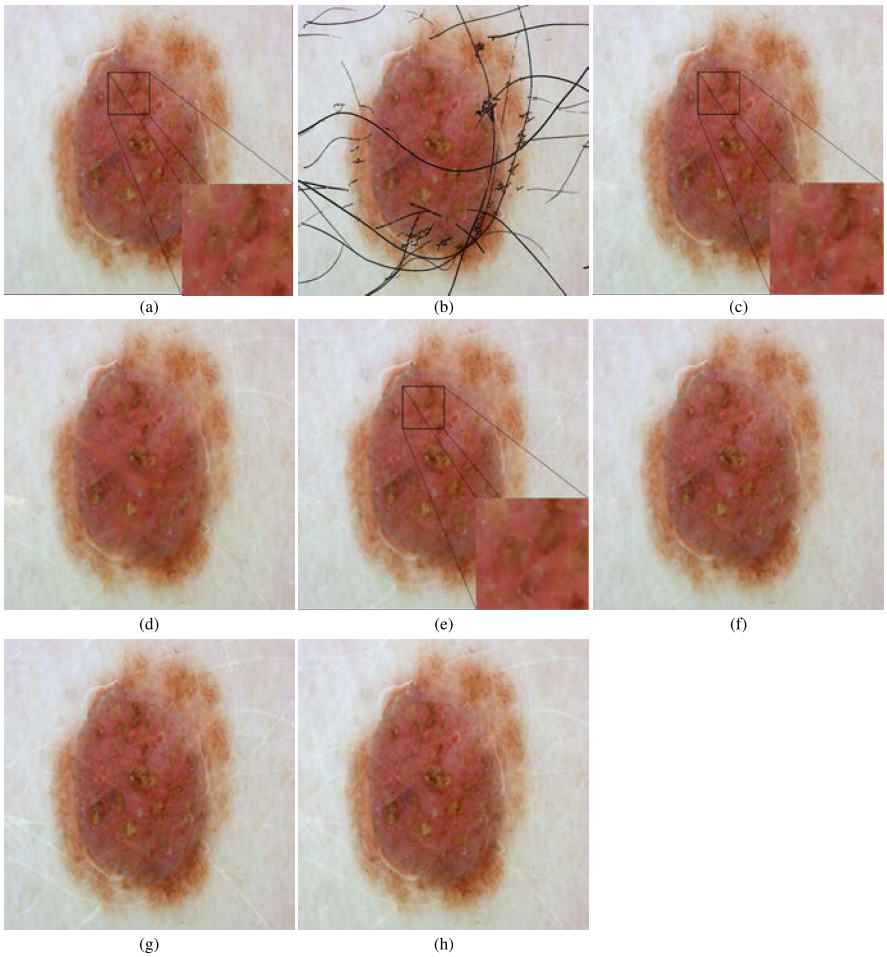


**FIGURE 8.** (a) Original hairless image, (b) Input image with simulated hair, (c) result of our model trained with the complete loss, (d)-(h) results of our model trained by removing (d) $L_1^{\text{foreground}}$, (e) $L_1^{\text{background}}$, (f) $L_2^{\text{composed}}$, (g) $\mathcal{L}_{\text{SSIM}}$, and (h) $\mathcal{L}_{\text{tv}}$.

in our model. In this case, we can visualize its effects in Figure 7, where we compare the results of using or not skip-connections. As can be seen, the network is able to create a more detailed prediction with them, especially when it comes to dermoscopic structures such as streaks or globules.

Another study that we believe is of great importance is the evaluation of the relevance of each term of the loss function. As in the previous case, we show in Table 4 and Figure 8 the quantitative and qualitative results, respectively, of our model trained by removing in each case one of the terms that compose the loss function. As can be seen, most of the performance measures and resulting images are worsen by deleting some of the terms. This is not the case when we stop computing the $L_1$ distance between the GT and the network's prediction only among the background pixels ($L_1^{\text{background}}$). By comparing Figures 8c and 8e, we can see that when we do not use this term, the structures tend to be more blurrier. Such blurrier regions may not be penalized as much when calculating performance measures.

## V. DISCUSSION AND CONCLUSION

In this work, we have presented a novel CNN-based method to the task of hair removal in dermoscopic images. We have built an encoder-decoder architecture, which has shown good results in reconstruction tasks like the one at hand. We highlight an architectural aspect of the network: the use of skip connections helps to retrieve details. The benefits of its use have been demonstrated with an ablation study. In addition, we have analyzed the performance of our method and compared it with six state-of-the-art approaches. To carry out the experiments we created a dataset using different hair simulating strategies over images from publicly available dermoscopic datasets. For the validation of the algorithms, we calculated nine measures of similarity between the hairless reference images and their corresponding image with simulated hair. Finally, we performed a statistical test to objectively study and compare their performance.

The results obtained by means of the statistical tests applied to these measures lead to the conclusion that for eight of the performance measures, our method is statistically

the best algorithm. Except for the VIF measure and when we compare it with Abbas' *et al.* and Lee *et al.*'s methods. As reflected in Figure 5 and in Table 3, Abbas' and Toossi's algorithms produce the least suitable results. This bad behavior may be due to the fact that these algorithms do not seem to distinguish well hairs of greater thickness or dark colors. It is worth mentioning that we have evaluated our model in dermoscopic images with real hair, obtaining good visual results and demonstrating, thus, its effectiveness.

As future work, we aim to use our approach on a more complete skin lesion analysis system, leveraging the knowledge to extract other characteristics. Also, increasing the number of images used in the dataset to train the network might enhance the network's generalization capabilities.

## ACKNOWLEDGMENT

The authors thank Dr. Attia from the Institute For Innovation and Research, Deakin University, Australia, for providing the GAN-based simulated hair images.

## REFERENCES

[1] *Melanoma Molecular Map Project*. Accessed: Apr. 4, 2020. [Online]. Available: http://www.mmmp.org/MMMP/welcome.mmmp

[2] *European Cancer Information System*. Accessed: Apr. 4, 2020. [Online]. Available: https://ecis.jrc.ec.europa.eu/index.php

[3] J. Mayer, "Systematic review of the diagnostic accuracy of dermatoscopy in detecting malignant melanoma," *Med. J. Aust.*, vol. 167, no. 4, pp. 206–210, Aug. 1997.

[4] G. Argenziano, H. P. Soyer, S. Chimenti, R. Talamini, R. Corona, F. Sera, M. Binder, L. Cerroni, G. De Rosa, G. Ferrara, and R. Hofmann-Wellenhof, "Dermoscopy of pigmented skin lesions: Results of a consensus meeting via the Internet," *J. Amer. Acad. Dermatol.*, vol. 48, no. 5, pp. 679–693, May 2003.

[5] G. Argenziano, G. Fabbrocini, P. Carli, V. De Giorgi, E. Sammarco, and M. Delfino, "Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions: Comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis," *J. Amer. Med. Assoc. Dermatol.*, vol. 134, no. 12, pp. 1563–1570, Dec. 1998.

[6] H. Kittler, "Dermatoscopy: Introduction of a new algorithmic method based on pattern analysis for diagnosis of pigmented skin lesions," *Dermatopathol., Practical Conceptual*, vol. 13, no. 1, p. 3, 2007.

[7] S. W. Menzies, "Frequency and morphologic characteristics of invasive melanomas lacking specific surface microscopic features," *Arch. Dermatol.*, vol. 132, no. 10, pp. 1178–1182, Oct. 1996.

[8] W. Stolz, "ABCD rule of dermatoscopy: A new practical method for early recognition of malignant melanoma," *Eur. J. Dermatol.*, vol. 4, no. 7, pp. 521–527, 1994.

[9] N. O'Mahony, S. Campbell, A. Carvalho, S. Harapanahalli, G. V. Hernandez, L. Krpalkova, D. Riordan, and J. Walsh, "Deep learning vs. traditional computer vision," in *Proc. Sci. Inf. Conf.* Cham, Switzerland: Springer, 2019, pp. 128–144.

[10] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[11] J. A. A. Salido, P. De La Salle University, and C. Ruiz, Jr., "Using deep learning for melanoma detection in dermoscopy images," *Int. J. Mach. Learn. Comput.*, vol. 8, no. 1, pp. 61–68, Feb. 2018.

[12] I. Bakkouri and K. Afdel, "Computer-aided diagnosis (CAD) system based on multi-layer feature fusion network for skin lesion recognition in dermoscopy images," *Multimedia Tools Appl.*, vol. 79, nos. 29–30, pp. 20483–20518, Aug. 2020.

[13] L. Talavera-Martínez, P. Bibiloni, and M. González-Hidalgo, "Comparative study of dermoscopic hair removal methods," in *Proc. ECCOMAS Thematic Conf. Comput. Vis. Med. Image Process.* Cham, Switzerland: Springer, 2019, pp. 12–21.

[14] T. Lee, V. Ng, R. Gallagher, A. Coldman, and D. McLean, "Dullrazor: A software approach to hair removal from images," *Comput. Biol. Med.*, vol. 27, no. 6, pp. 533–543, Nov. 1997.

[15] F.-Y. Xie, S.-Y. Qin, Z.-G. Jiang, and R.-S. Meng, "PDE-based unsupervised repair of hair-occluded information in dermoscopy images of melanoma," *Comput. Med. Imag. Graph.*, vol. 33, no. 4, pp. 275–282, Jun. 2009.

[16] Q. Abbas, M. E. Celebi, and I. F. García, "Hair removal methods: A comparative study for dermoscopy images," *Biomed. Signal Process. Control*, vol. 6, no. 4, pp. 395–404, Oct. 2011.

[17] A. Huang, S.-Y. Kwan, W.-Y. Chang, M.-Y. Liu, M.-H. Chi, and G.-S. Chen, "A robust hair segmentation and removal approach for clinical images of skin lesions," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2013, pp. 3315–3318.

[18] M. T. B. Toossi, H. R. Pourreza, H. Zare, M.-H. Sigari, P. Layegh, and A. Azimi, "An effective hair removal algorithm for dermoscopy images," *Skin Res. Technol.*, vol. 19, no. 3, pp. 230–235, Aug. 2013.

[19] P. Bibiloni, M. González-Hidalgo, and S. Massanet, "Skin hair removal in dermoscopic images using soft color morphology," in *Proc. Conf. Artif. Intell. Med. Eur.* Cham, Switzerland: Springer, 2017, pp. 322–326.

[20] J. Xie, L. Xu, and E. Chen, "Image denoising and inpainting with deep neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 341–349.

[21] C. Tian, L. Fei, W. Zheng, Y. Xu, W. Zuo, and C.-W. Lin, "Deep learning on image denoising: An overview," 2019, *arXiv:1912.13171*. [Online]. Available: http://arxiv.org/abs/1912.13171

[22] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proc. 25th Int. Conf. Mach. Learn. (ICML)*, 2008, pp. 1096–1103.

[23] Z. Cui, H. Chang, S. Shan, B. Zhong, and X. Chen, "Deep network cascade for image super-resolution," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2014, pp. 49–64.

[24] X. Mao, C. Shen, and Y.-B. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 2802–2810.

[25] V. Jain and S. Seung, "Natural image denoising with convolutional networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 769–776.

[26] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.

[27] M. Attia, M. Hossny, H. Zhou, S. Nahavandi, H. Asadi, and A. Yazdabadi, "Realistic hair simulator for skin lesion images: A novel benchemarking tool," *Artif. Intell. Med.*, vol. 108, Aug. 2020, Art. no. 101933.

[28] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Trans. Comput. Imag.*, vol. 3, no. 1, pp. 47–57, Mar. 2017.

[29] G. Liu, F. A. Reda, K. J. Shih, T.-C. Wang, A. Tao, and B. Catanzaro, "Image inpainting for irregular holes using partial convolutions," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 85–100.

[30] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D, Nonlinear Phenomena*, vol. 60, nos. 1–4, pp. 259–268, Nov. 1992.

[31] T. Mendonça, M. Celebi, T. Mendonça, and J. Marques, "$PH^2$: A public database for the analysis of dermoscopic images," *Dermoscopy Image Anal.*, 2015.

[32] G. Argenziano, H. Soyer, V. De Giorgi, D. Piccolo, P. Carli, and M. Delfino, *Interactive Atlas of Dermoscopy (Book and CD-ROM)*. EDRA Medical Publishing & New Media, 2000.

[33] H. Mirzaalian, T. K. Lee, and G. Hamarneh, "Hair enhancement in dermoscopic images using dual-channel quaternion tubularness filters and MRF-based multilabel optimization," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5486–5496, Dec. 2014.

[34] H. Mirzaalian. *Hair SIM Software*. Accessed: Mar. 7, 2019. [Online]. Available: http://www2.cs.sfu.ca/~hamarneh/software/hairsim/Welcome.html

[35] F. Chollet, "Keras: The python deep learning library," Astrophys. Source Code Library, Tech. Rep. Rec. ascl:1806.022, 2018.

[36] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980

[37] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98–117, Jan. 2009.

[38] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[39] A. G. Barnston, "Correspondence among the correlation, RMSE, and Heidke forecast verification measures; refinement of the Heidke score," *Weather Forecasting*, vol. 7, no. 4, pp. 699–709, Dec. 1992.
[40] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th Asilomar Conf. Signals, Syst. Comput.*, vol. 2, 2003, pp. 1398–1402.
[41] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.
[42] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006, doi: 10.1109/TIP.2005.859378.
[43] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, and M. Carli, "New full-reference quality metrics based on HVS," in *Proc. 2nd Int. Workshop Video Process. Qual. Metrics*, vol. 4, 2006, pp. 1–4.

**LIDIA TALAVERA-MARTÍNEZ** was born in Jaén, Spain, in 1993. She received the B.Sc. degree in industrial electronics and automation engineering from the University of the Balearic Islands (UIB), in 2016, and the M.Sc. degree in computer vision from the Autonomous University of Barcelona (UAB), in 2017. She is currently pursuing the Ph.D. degree in information and communication technologies with the Soft Computing, Image Processing and Aggregation (SCOPIA) research group. Her research interests include computer vision, deep learning, and medical imaging.

**PEDRO BIBILONI** received the B.S. degree in mathematics and the B.Sc. degree in telecommunications engineering from the Universitat Politècnica de Catalunya, in 2013 and 2015, respectively, the M.Sc. degree in information security from the University College London, in 2014, and the Ph.D. degree from the University of the Balearic Islands, in 2018. He is currently a Lecturer with the University of the Balearic Islands. His research interests include medical image processing and pattern recognition.

**MANUEL GONZÁLEZ-HIDALGO** (Member, IEEE) was born in Leon, Spain, in 1964. He received the degree in mathematics (specialty in general mathematics and guidance mathematical analysis) from the University of Valencia, in 1988, and the Ph.D. degree in computer science from the University of the Balearic Islands (UIB), in 1995. He is currently an Associate Professor with the Department of Mathematics and Computer Science, UIB. He is a member of the Soft Computing, Image Processing and Aggregation (SCOPIA) research group and a regular collaborator with the Computer Graphics and Vision and AI Group (UGiVIA) research group. He is also a member of the Balearic Islands Health Research Institute. This research activity has been reflected in scientific publications in international journals and in papers presented at national and international conferences. He is a referee for several journals of reference in his research fields and has organized several congress and special sessions, and different scientific activities. His research interests include computer vision, image analysis, modeling and animation of deformable objects, analysis and synthesis of human movement, medical imaging and 3D modeling, and recently the study of aggregation operators and their applications to image processing and analysis, focusing on the fuzzy mathematical morphology and its applications. He is currently working in soft computing techniques and its applications to biomedical image analysis.