Skin_Hair dataset: Setting the benchmark for effective hair inpainting methods for improving the image quality of dermoscopic images

Joanna Jaworek-Korjakowska[1,2][0000−0003−0146−8652], Anna Wojcicka[1][0000−0001−8060−2009], Dariusz Kucharski[1][0000−0002−0107−2407], Andrzej Brodzicki[1][0000−0001−7713−526X], Connah Kendrick[3][0000−0002−3623−6598], Bill Cassidy[3][0000−0003−3741−8120], and Moi Hoon Yap[3][0000−0001−7681−4287]

[1] AGH University of Science and Technology, Al Mickiewicza 30, 30-059 Krakow, Poland {jaworek,wojcicka,darekk,brodzicki}@agh.edu.pl

[2] Stanford School of Medicine ,Department of Pathology, 300 Pasteur Drive, Stanford, CA 94305-5324, USA
http://www.springer.com/gp/computer-science/lncs

[3] Manchester Metropolitan University, John Dalton Building, Chester Street, Manchester M1 5GD, UK
{Connah.Kendrick,B.Cassidy,M.Yap}@mmu.ac.uk

Abstract. Dermoscopic images are often contaminated by artifacts including clinical pen markings, immersion fluid air bubbles, dark corners, and most importantly hair, which makes interpreting them more challenging for clinicians and computer-aided diagnostic algorithms. Hence, automated artifact recognition and inpainting systems have the potential to aid the clinical workflow as well as serve as an preprocessing step in the automated classification of dermoscopic images. In this paper, we share the first release of a public dermoscopic image dataset with hair artifacts which can be accessed here https://skin-hairdataset.github.io/SHD/. The Skin_Hair dataset contains over 252 dermoscopic images including artificial hair and will be expanded over time. Furthermore, we present the primary results of applying machine learning algorithms and GAN based architectures to the hair inpainting problem in dermoscopic images. We envision that these results will serve as a benchmark for researchers who might work on the hair detection and reconstruction tasks with this dataset in the future. In this work, we present a skin lesion image dataset based on the ISIC dataset containing dermoscopic images, images containing artificial hairs and the corresponding ground-truth masks. Furthermore, we use four hair inpainting methods including Navier-Stokes, Telea, Hair_SinGAN and R-MNet architectures which we evaluate using image quality assessment metrics MSE, PSNR, UQI and SSIM. The R-MNet architecture achieved the highest SSIM score of 0.960.

Keywords: Melanoma, dermoscopy, hair inpainting, artifacts, hair removal, image quality, GAN

1 Introduction

Removal of artifacts from dermoscopic images is a necessary step in classifying skin lesions since artifacts can lead to severe misinterpretation of the global and local structures both for clinical and computer-aided diagnosis. Automated analysis of dermoscopic images is a challenge task [12], with one of the main difficulties being the existence of a variety of artifacts including clinical pen markings, rulers, immersion fluid air bubbles, size-reference stickers, dark corners, and most commonly - hair (see Fig. 1). These artifacts are strikingly different when compared to the rest of the image in both color, shape, and features. As the unique patterns in human skin are often very subtle, those unwanted artifacts often draw the attention of the deep neural network, leading to a falsified diagnosis. Furthermore, the presence of hair may obscure and distort important areas that could determine the final classification.

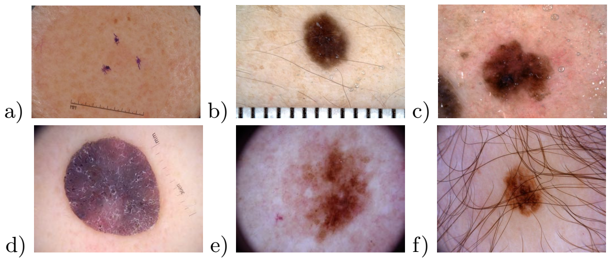


Fig. 1. Sample images from the ISIC database with the following artifacts: a) clinical pen markings, b) rulers, c) immersion fluid air bubbles, d) lens measurement reference, e) dark corners, e) hair

One of the advantages of deep learning methods is the relative lack of preprocessing needed. In most computer vision tasks, including segmentation and classification which are mostly based on CNNs, datasets without any preparation or preprocessing are directly passed to the backbone of the CNN network in order to learn the features. However, prior research [28–30, 38] indicates that, in the case of dermoscopic image analysis, most of the algorithms perform better when the artifacts are removed or inpainted.

The presence of hair in dermoscopic images poses a significant challenge as they may occlude some of the information of the lesion such as its boundary and texture. Hence, the removal of hair is an important preprocessing step which, due to its diverse appearances, causes significant problems. We propose a dermoscopic image dataset which gives the possibility to work in the area of removing artifacts and can be used as a benchmark for researchers working on hair detection and inpainting. The dataset uses images from the ISIC datasets [14–16, 18, 31, 41] and consists of dermoscopic images with artificially added hairs as well as corresponding binary masks.

Based on the proposed dataset, which consist of 252 dermoscopic images, we have trained and evaluated two traditional inpainting methods, Telea [39] and Navier–Stokes [7]). In addition deep learning based methods Hair_SinGAN and R-MNet have been proposed. The hair inpainting algorithms have been evaluate using image quality assessment metrics MSE, PSNR, UQI and SSIM. The R-MNet architecture achieved the highest SSIM score 0.960.

The main novelty of this paper can be summarised as follows:

- We introduce a benchmark dataset with consists of 252 cases including: raw dermoscopic images (reference images), corresponding images with overlaying artificial hairs and binary masks which serve as ground-truth.
- We use state-of-the-art Reverse-Masking networks for the inpainting of hairs in dermoscopic images which applies changes only to the target region.
- We propose the Hair_SinGAN architecture, based on the work of [32] et al. which is trained on a single image.
- We statistically evaluate our hair inpainting methods using image quality assessment metrics MSE, PSNR, UQI and SSIM and suggest the R-MNet architecture to serve as the pre-processing method for dermoscopic images.

2 Related Work

Research focusing on automated skin lesion analysis often observe the occurrence of artifacts, but do not discuss how to circumvent the possible negative effects of their presence [23], or do not investigate the effects of their removal [45]. Early attempts to remove hair from skin lesion images were conducted by Lee et al. [25] who created the Dullrazor software. They used grayscale morphological closing to perform hair segmentation and bilinear interpolation to remove hairs from melanoma images. However, this approach is limited in that it is only effective in removing thick dark hairs from skin lesion images. This method would later be improved by [24] et al. who developed the E-shaver application which used an edge detector with color averaging making it more effective on different types of hairs. However, in their experiments, they tested on only 50 images. In the same year, Fiorese et al. [17] proposed the VirtualShave tool which used partial differential equation inpainting, and claimed performance comparable to human operators removing hair manually, with resulting images being almost indistinguishable from hair-free skin.

Later, Xie et al. [43] used a top-hat operator to segment and anisotropic diffusion to remove hair from skin lesion images. As per previous works, this study was not able to handle all types of hair. Additionally, this method was only tested on a very small dataset of 40 just images. Limited dataset testing is a common theme in many prior research projects in this domain [1,10,17,20,37,40].

Maglogiannis [27] et al. used combinations of Bottom-hat, Laplacian, and Sobel methods to identify and remove hair from dermoscopic images. They observed that the Laplacian of Gaussian and Sobel edge detection methods combined, together with a 3x3 wiener noise reduction filter, provided the best results.

Salido et al. [33] performed hair removal on dermoscopic images using morphological bottom-hat filtering with erosion and dilation by morphological opening. Inpainting was completed using a nonlinear model based on curvature-driven diffusions for nontexture images, originally proposed by [13].

Bardou et al. [6] used a variational autoencoder to remove hair from dermoscopic skin lesion images in the HAM10000 dataset without the need for paired samples. The encoder uses dermoscope images as input and builds a latent distribution which ignores hair as noise, while the decoder reconstructs a hair-free image. Their results show high quality inpainting, the reconstructed images are not identical to the input images as they look blurry and often distort the features of lesions.

In 2019, Talavera et al. [38] identified that there are currently no methods to benchmark the effectiveness of hair removal algorithms. They extracted 13 hairless images from the PH2 dataset and overlaid artificial hairs to test the effectiveness of 6 state-of-the-art algorithms and compared the results.

Li et al. [26] trained a U-Net with ISIC data to obtain hair masks, and propose an inpainting architecture comprising a gated convolution and SN-PatchGAN. They categorised hair in ISIC images as: thin; overlapping; faded; of similar contrast or colour to the underlying skin; and obscuring lesions. They observe that traditional hard-coded threshold-based hair removal methods are ineffective, and can result in over-removal which can cause loss of important lesion details, or under-removal where the hair cannot be removed effectively. They also propose an evaluation method (intra-structural similarity) to analyse the effect of hair removal based on a single dermoscopic image.

Song et al. [36] proposed a novel hair extraction method which utilised maximum variance fuzzy clustering, with a Criminisi algorithm used for repairing image regions where hair had been removed. This method is capable of fast hair extraction and segmentation with reduced computational complexity. The implementation does not require extensive learning based on a large number of parameters and training images, resulting in high execution efficiency.

More recently, Nauta et al. [29] found that CNN classifiers partly based predictions of benign images on the presence of colour calibration patches placed onto the skin during examinations. By artificially inserting colour calibration patches into malignant images, they showed that shortcut learning results in a significant increase in misdiagnoses. This work indicates that other artifact types may present similar issues.

We surveyed 38 state-of-the-art papers from the field of computer aided diagnosis in skin lesion tasks and checked if the authors mention any techniques for hair removal. Although most of the papers describe the difficulties of dealing with artifacts they often state that it is an issue for computer vision processing methods. The deep learning methodology hasn't been explored in this area, yet. Only in 7 papers researchers indicated the artifact removal or enhancement stage [2–5, 8, 9, 44].

3 Skin_Hair Dataset

Fig. 2. Illustration of the dataset creation process: (a) clear image without hair, (b) hair extracted from different image placed over the clear image, and (c) ground truth reference mask.

The main issue in the process of detection, reconstruction and assessment of artifacts is mostly due to the lack of properly prepared datasets which do not include ground truth masks and reference images. Due to the artifact removal evaluation process we propose a novel Skin_Hair dataset that includes the raw dermoscopic images, images containing artificial hairs as well as ground-truth masks for evaluation purpose. This dataset is created by taking raw dermoscopic images without hair artifacts from the ISIC dataset that serve as a reference ground-truth image and for applying manually extracted hairs from other dermoscopic images from the ISIC dataset. The dataset can be obtained from the following repository: https://skin-hairdataset.github.io/SHD/.

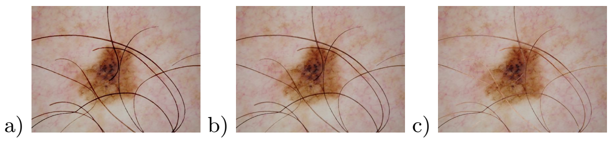
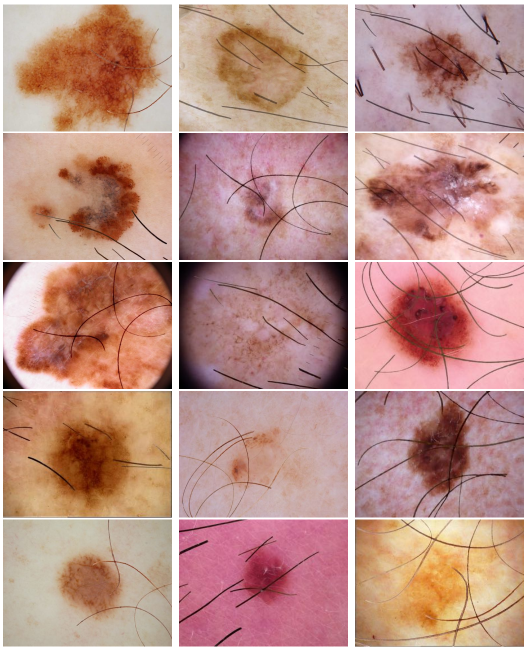


Fig. 3. Illustration of three types of hair colour: (a) dark, (b) brown, and (c) light.

Raw dermoscopic images without hair, as well as hair patterns, were taken from the ISIC database [21], which is the largest publicly available dataset containing dermoscopic images with metadata. To successfully determine the effectiveness of the hair removal methods, we use the balanced ISIC dataset as presented in [11]. Pewton and Yap [30] annotated the dataset for numerous artifacts, including hair. Based on the provided information we divide the dataset into two separate parts containing hair and without hair, respectively. Due to the very large variety and complexity of the hair patterns, we decided to transfer the hair from other dermoscopic images, which allowed us to maintain their natural appearance (Fig. 2). The process of creating a dermoscopic image with hairs consists of the following steps: 1) Choosing a raw image without artifacts from the ISIC dataset, 2) Choosing an image including hairs from the ISIC dataset,

3) Manually marking the hair areas using Photoshop quick mask with alpha channel, soft, round brush with full opacity and size adapted to the size of the marked hair, 4) Cutting out the hair to a new transparent layer and clearing any additional areas of skin visible on this layer, 5) Applying the hair mask to the dermatoscopic image.

The extracted hair patterns have been augmented using the following methods: 1) randomly moving and rotating the mask, 2) modification of the selection with small, medium and large number of hairs; 3) changing the color of the hair into three main categories (light, brown, and dark - defined based on the analysis of the dataset) using brightness, contrast tool and color blending mode; 4) randomly applying different masks onto different clean images, without hair; and 5) for each modified pattern, a reference mask was created using a threshold tool.



a) b) c)

Fig. 4. Illustrations of three types of hair size: (a) small, (b) medium, and (c) large.

The method is repeated for three different hair colours - dark, brown and light as presented in Fig. 3. In total, we used 77 non-hair images as the basis for applying different hair configurations. We augmented the extracted hair by changing the size, amount and colour. In total 252 images were generated with 84 unique masks to cover the different hair types. The Skin_Hair dataset contains: 35 images with small density (each in three colours - light, brown and dark), 27 images with medium density (each in three colours - light, brown and dark) and 22 images with high density (each in three colours - light, brown and dark). The process of mask extraction and hair addition was performed using Adobe Photoshop 23.4.1.

4 Effective hair inpainting algorithms

As the artifact removal process is an obligatory step in image preprocessing we have considered 5 different inpainting techniques in order to compare the traditional computer vision inpainting methods including Navier-Stokes and Telea with two state-of-the art deep learning techniques - SinGAN and R-MNet.

4.1 Navier-Stokes

Bertalmio et. al. found an analogy between the image inpainting problem and the stream function in a two-dimensional (2D) incompressible fluid. An approximate solution to the inpainting problem is obtained by numerically approximating the steady state solution of the 2D NSE (Navier-Stokes Equations) vorticity transport equation, and simultaneously solving the Poisson equation between the vorticity and stream function, in the region to be inpainted [7]. Image intensity is changed via a 'stream function'. Isophote lines (lines of equal brightness intensity) are propagated along the edges from the outside into the region that is being inpainted. Instead of using the vorticity of the fluid, the method uses the laplacian of the intensity. The direction of the flow is a vector field defined by the stream function. The algorithm continues the isophote lines and matches gradient vectors at the boundary of the inpainting region [7]. Results of the Navier-Stokes algorithm are presented in Fig. 5.

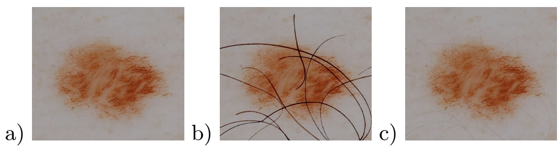


Fig. 5. Illustration of the effects of the Navier-Stokes inpainting method: a) original dermoscopic image, b) dermoscopic image containing artificial hair, and c) Navier-Stokes inpainting outcomes.

4.2 Telea

Telea [39], proposed an inpainting algorithm using a Fast Marching Method (FMM). This method is considered faster and less complex to compute than other typical inpainting methods [39]. The algorithm uses known regions to grow inpainted regions into the target regions, and is enforced by the use of the FMM. The Fast Marching Method itself is a numerical technique used for solving a boundary value problem [34]. Here it is used to ensure that pixels closer to the known neighbours are inpainted before pixels with unknown neighbours. It performs a similar role to a distance transform but has an advantage of maintaining narrow bands - a boundary between known and unknown areas. The algorithm defines three types of pixels: $BAND$: the pixel belongs to the narrow band, $KNOWN$: the pixel is outside the inpainting boundary (known) and $INSIDE$: the pixel is inside the inpainting boundary (unknown). For each pixel, there are two values - T (distance to the edge) and I (grey-level intensity). The algorithm works in the following steps: 1) extract the $BAND$ point with the smallest T, 2) march the boundary inward by adding new points to it, 3) perform the inpainting: iterate over the $KNOWN$ points in the neighborhood of the current point (i, j) and compute $I(i, j)$ and the image gradient gradient is estimated by central differences, 4) propagates the value T of point (i, j) to its neighbors (k, l) by solving the finite difference discretization problem, 5) inserts (k, l) with its new T in the heap. Results of the Telea algorithm are presented in Figure 6.

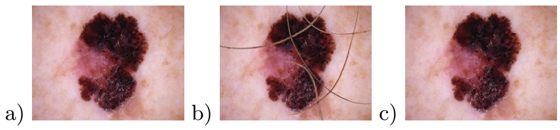


Fig. 6. Illustration of the effects of the Telea inpainting algorithm: a) original dermoscopic image, b) dermoscopic image containing artificial hair, and c) Telea inpainting outcomes.

4.3 Hair_SinGAN architecture

While most of deep learning models require large numbers of examples in order to be trained effectively, we tried to design an approach which works on as few examples as possible. In practice, this is technically difficult as the dataset needs to represent the underlying distribution, and the more examples the dataset consists of, the more accurate the representation is. However, considering a single image as a dataset itself, it can represent its own distribution. The general idea behind our approach is to analyze parts of an image which are not hidden behind artifacts that we want to remove, train the model on those parts, and then use the model in order to reconstruct areas hidden behind (Fig. 7). The algorithm starts with dividing the image into a set of smaller training rectangular regions. Those areas where the GT image mask shows no hair are training regions and a the rest of the image becomes a reconstruction region (see steps T1 and R1 on

Fig. 7). Then, from the proposed image parts, multiple rectangles are extracted, which constitute inputs to a model based on Generative Adversarial Networks (T2 and R2). Additionally, in training phase (T4), those inputs are enriched with artificial artifacts (T3). On the other hand, original snippets serves as model's reconstruction goal (network output). After the training process is finished, snippets prepared from the reconstruction region (R2) are then fed into the network (R3). The outputs, with the hair removed, replaces fragments on the original image.

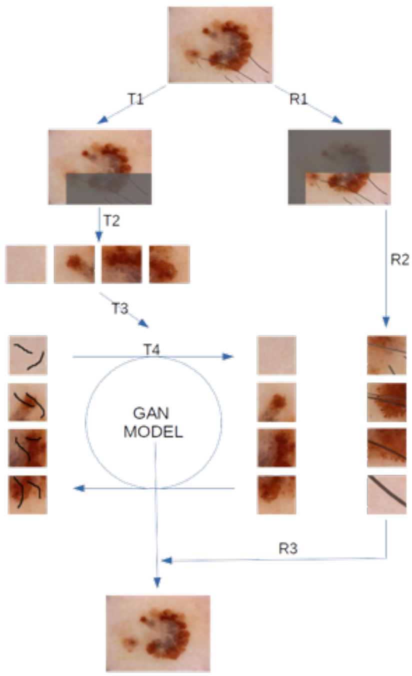


Fig. 7. SinGAN [35] algorithm pipeline, consisting of two independent branches. The T branch represents a model training process on artificially generated examples. Such a model is then used for inpainting on fragments covered with artifacts (showed on the branch R). In the final step, reconstructed fragments are replaced with those on the original image.

Given that this method requires only a single image for reconstruction, the results represent a valid alternative compared to prior traditional methods, especially in the case of poor quality datasets (Table 1). The main drawback of the algorithm is the ratio between the training area and the reconstruction area.

When the ratio is below 1, reconstruction starts to disclose insufficient dataset problem.

4.4 R-MNet method

For hair inpainting we employ the use of Reverse-Masking networks [22]. The advantage of this method is that it is similar to the traditional methods and the GAN focuses only on target regions, making no change to the surrounding regions. The network does this by importing the mask and then feeding into the network as in traditional structures, as illustrated in Fig. 8.

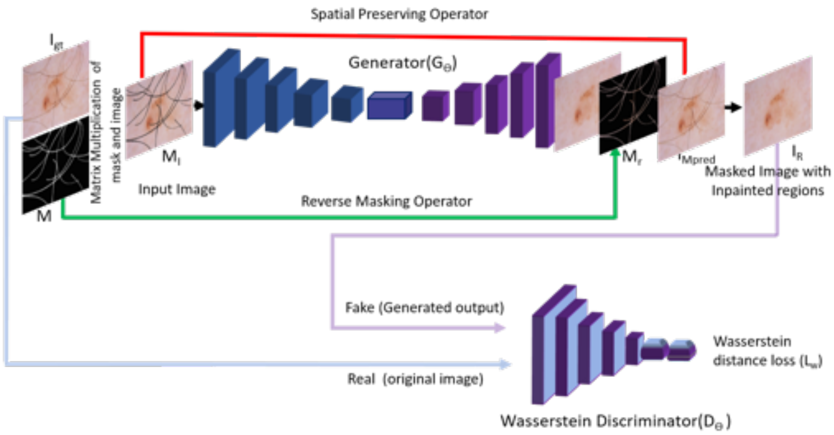


Fig. 8. Illustration of R-MNet [22] on skin lesions inpainting.

The mask is applied internally, using matrix operations which inverse the mask which is used to reapply the undamaged areas to the inpainted images. The network then uses surrounding regions to inpaint the selected areas, as with traditional prior techniques. The network uses an encoder-decoder structure, to focus on non-damaged regions, and reconstructs the inpainted image during the decode stage. However, owing to the network structure a custom loss function is used which focuses on damaged regions to assess reconstruction. The decoder uses a series of 5×5 convolutions of increasing filter depth. Each convolution is followed by a LeakyReLU activation, using an alpha of 0.2, a dropout of 0.5, followed by a max-pooling layer. The LeakyReLU is used to prevent the non activation of the neurons, as instead of the function being zero when $x < 0$, the leakyReLU will return some small negative number instead. The LeakyReLU is used during the encoding stage, allowing a diverse set of features to be captured, the dropout aids the network to deal with less features and the pooling focuses the network onto the core features. The decoder follows a series of up-sampling layers, with transposed convolutions, standard ReLU and batch normalisation, with a final tanh layer. The encoder mixes both 2D up-sampling and transpose convs to resoles the feature back to the original size while having some convolutional functions, the batch normalisation aids the network to generalise during training. To train our network we used the generated hair masks on the hairless images to ensure the network learned skin features only. Owing to this the network requires hairless examples as ground truth to ensure the system only learns to inpaint skin and the skin lesions. When training using R-MNet the network manages to inpaint the hair masks, as illustrated in Fig. 9 which shows that the network manages to successfully inpaint skin regions and parts of the lesions However, limitations, due to the limited umber of input images, are apparent.

Fig. 9. Illustration of the visual appearances on skin lesions: a) original dermoscopic image, b) dermoscopic image containing artificial hair, and c) R-MNet inpainted image.

5 Result analysis

Image Quality Assessment (IQA) is considered as a characteristic property of an image and describes the degradation of the perceived image. Quality of an image can be described technically with statistical metrics as well as objectively to indicate the deviation from the ideal or reference model. In our case the ideal is an image without the hair mask. There are several techniques and metrics available that can be used for objective image quality assessment. Here, we use the full-Reference (FR) approach, as we assess the quality of a test image in comparison with a reference image which is considered to be of perfect quality. We take advantage of image quality techniques to compare the outcomes of our proposed inpainting algorithms for the hair regions such as MSE (Mean Square Error), PSNR (Peak Signal to Noise Ratio), SSIM (Structured Similarity Index Measure), and UQI (Universal Quality Index).

MSE is the most common estimator of image quality measurement and refers to the second moment of error. The error is the difference between the estimator and the estimated outcome. It is a function of risk, considering the expected value of the squared error loss or quadratic loss, with a bias towards large deviation from the ground truth. MSE is a full reference metric with values closer to zero indicating higher similarity. MSE between two images such as $I(x, y)$ and $K(x, y)$ is defined as [42]:

$$MSE = \frac{1}{m\,n}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2. \tag{1}$$

PSNR is used to calculate the ratio between the maximum possible signal power and the power of the distorting noise which affects the quality of its representation. This ratio between two images is computed in decibel form and is usually calculated as the logarithm term of decibel scale where the dynamic range varies between the largest and the smallest possible values which are changeable by their quality. The PSNR is defined as [19]:

$$PSNR = 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right) \tag{2}$$

SSIM is a perception based model where image degradation is considered as the change of perception in structural information [46]. It also takes into consideration other important perception based elements such as luminance masking and contrast masking. The difference of this method when compared to other techniques, such as MSE or PSNR, is that the other approaches estimate absolute errors. Structural information is the idea that the pixels have strong inter-dependencies especially when they are spatially close. The SSIM index is calculated on various windows of an image. The measure between two windows x and y of common size $N \times N$ is defined as [46]:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{3}$$

where L is the dynamic range of the pixel-values (typically this is $2^{\#bits\ per\ pixel} - 1$, $k_1 = 0.01$ and $k_2 = 0.03$ by default).

UQI was the predecessor of SSIM and evaluates quality of an image using loss of correlation, luminance distortion, and contrast distortion. UQI is global rather than being local or specially intended to the images being tested or on the individual observers. The quality index is defined as:

$$Q = \frac{4\sigma_{xy}\overline{x}\overline{y}}{(\sigma_x^2 + \sigma_y^2) + ((\overline{x})^2 + (\overline{y})^2)} \tag{4}$$

where, $\overline{x}$ and $\overline{y}$ are the mean values of the original and distorted images respectively. σ_x^2 and σ_y^2 are the variances. σ_{xy} is the covariance. The range of UQI is $[-1, 1]$ where 1 is achieved when the two images are identical.

Due to the R-MNet algorithm which requires the training set the proposed Skin_Hair dataset has been divided into training and testing sets including 170 and 82 images respectively. After inpainting (reconstructing) skin lesion images, we estimated the quality by using MSE, PSNR, UQI and SSIM metrics. The summary of quality matrices calculations is shown in Table 1. From this table, we observe that all metrics have given almost consistent results. From a representation perspective, SSIM and UQI is normalized, but MSE and PSNR are not. Therefore, SSIM and UQI can be treated as more understandable than MSE

Table 1. Summary of the IQA metrics including MSE, PSNR, SSIM and UQI for hair inpaining methods including Navier-Stokes, Telea, Hair_SinGAN and R-MNet. † = higher value is better; ⊎ = lower value is better.

Hair inpainting method	MSE ⊎	PSNR †	SSIM †	UQI †
Navier-Stokes	7.380	40.305	0.959	0.9984
Telea	**7.114**	**40.558**	0.959	0.9984
Hair_SinGAN	53.735	34.489	0.881	0.9976
R-MNet	23.743	**40.655**	**0.960**	**0.9985**

and PSNR. This is due to MSE and PSNR being absolute errors, however, SSIM provides perception and saliency-based errors. The highest SSIM value has been achieved by the R-MNet GAN based architecture (Fig. 10).

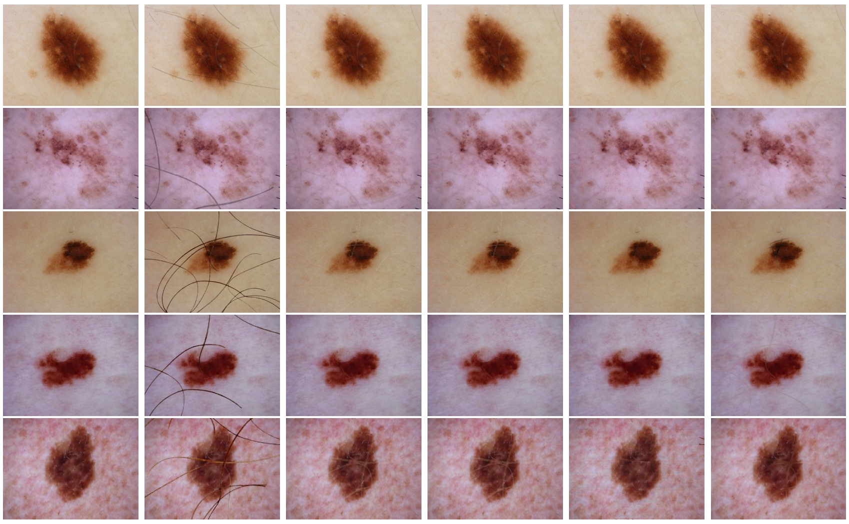


Fig. 10. Visual comparison of the results of different methods: a) original dermoscopic image, b) dermoscopic image containing artificial hair, c) Navier-Stokes inpainted image, d) Telea inpainted image, e) Hair_SinGAN inpainted image, and f) R-MNet inpainted image. The regions inpainted with Hair_SinGAN are the least distinguishable to the human eye. We can also observe, that the regions containing hair overlapping the lesions were the hardest to inpaint.

When artifact levels increase, the recovery quality of the output image is also shown to deteriorate, which we can be observed in Figure 11.

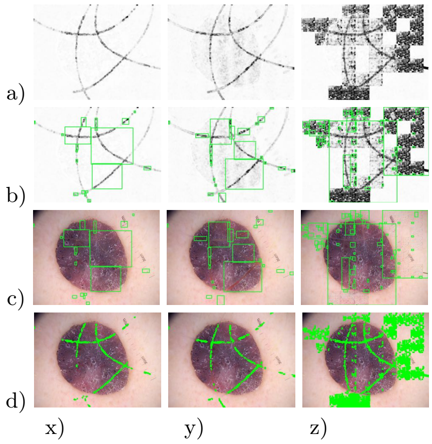


Fig. 11. Illustrations of the SSIM algorithm for three hair inpainting methods: x) Telea, y) R-MNet and z) Hair_SinGAN: a) image differences with darker regions show more disparity, b-c) filter using a minimum threshold area to remove the gray noise, and highlight the differences with a bounding box, d) visualisation of the exact differences.

6 Conclusions

We demonstrate the application of image inpainting onto skin lesions for hair removal, and highlight key issues in this field. Namely the lack of ground truth data, and present a novel hair inpainting dataset for qualitative evaluation of inpainting techniques. The most important contribution is the release of a dataset of skin images with added hair. Although the dataset is limited in size it can provide valuable benchmarking on future hair removal techniques. However, we continue to work on extending the number of images and will release a larger second version at a later date. We hope that the existence of a large collection of corresponding images with their reference ground truths, will be a useful addition to the ISIC database, helpful for researchers wishing to work on skin lesions. Furthermore, we plan to add different artifacts such as measuring tools, such as air bubbles and dark corner. We have used this dataset to test four inpainting algorithms - two classical approaches (Navier-Stokes and Telea) and two of our own implementations based on GANs. Available ground truth images allowed us to statistically evaluate those methods. The lowest MSE was achieved by the Telea method. However, in terms of other metrics the best performing method proved to be R-MNet. We note however that while Hair_SinGAN achieved slightly inferior results, it was only trained on a single image for each example.

Acknowledgments

We gratefully acknowledge the funding support of the research project by the program "Excellence initiative—research university" for the AGH UST and the NAWA Bekker Scholarship for J. Jaworek-Korjakowska.

References

1. Abbas, Q., Celebi, M., García, I.F.: Hair removal methods: A comparative study for dermoscopy images. Biomedical Signal Processing and Control **6**(4), 395–404 (2011). https://doi.org/https://doi.org/10.1016/j.bspc.2011.01.003, https://www.sciencedirect.com/science/article/pii/S1746809411000048
2. Adegun, A.A., Viriri, S.: Deep learning-based system for automatic melanoma detection. IEEE Access **8**, 7160–7172 (2020)
3. Adegun, A.A., Viriri, S.: Deep learning techniques for skin lesion analysis and melanoma cancer detection: a survey of state-of-the-art. Artificial Intelligence Review pp. 1–31 (2020)
4. Almaraz-Damian, J.A., Ponomaryov, V., Sadovnychiy, S., Castillejos-Fernandez, H.: Melanoma and nevus skin lesion classification using handcraft and deep learning feature fusion via mutual information measures. Entropy **22**(4) (2020). https://doi.org/10.3390/e22040484, https://www.mdpi.com/1099-4300/22/4/484
5. Barbosa, J., Baleiras, M.: Melanoma detection using deep learning methods (2019)
6. Bardou, D., Bouaziz, H., Lv, L., Zhang, T.: Hair removal in dermoscopy images using variational autoencoders. Skin Research and Technology **28**(3), 445–454 (2022). https://doi.org/https://doi.org/10.1111/srt.13145, https://onlinelibrary.wiley.com/doi/abs/10.1111/srt.13145
7. Bertalmio, M., Bertozzi, A.L., Sapiro, G.: Navier-stokes, fluid dynamics, and image and video inpainting. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001. vol. 1, pp. I–I. IEEE (2001)
8. Bisla, D., Choromanska, A., Berman, R., Stein, J., Polsky, D.: Towards automated melanoma detection with deep learning: Data purification and augmentation. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) pp. 2720–2728 (2019)
9. Bisla, D., Choromanska, A., Stein, J., Polsky, D., Berman, R.: Skin lesion segmentation and classification with deep learning system. ArXiv **abs/1902.06061** (2019)
10. Borys, D., Kowalska, P., Frackiewicz, M., Ostrowski, Z.: A simple hair removal algorithm from dermoscopic images. In: Ortuño, F., Rojas, I. (eds.) Bioinformatics and Biomedical Engineering. pp. 262–273. Springer International Publishing, Cham (2015)
11. Cassidy, B., Kendrick, C., Brodzicki, A., Jaworek-Korjakowska, J., Yap, M.H.: Analysis of the isic image datasets: usage, benchmarks and recommendations. Medical Image Analysis **75**, 102305 (2022)
12. Celebi, M.E., Barata, C., Halpern, A., Tschandl, P., Combalia, M., Liu, Y.: Guest editorial: Image analysis in dermatology. Medical Image Analysis **79**, 102468 (05 2022). https://doi.org/10.1016/j.media.2022.102468
13. Chan, T.F., Shen, J.: Nontexture inpainting by curvature-driven diffusions. Journal of Visual Communication and Image Representation **12**(4), 436–449 (2001). https://doi.org/https://doi.org/10.1006/jvci.2001.0487, https://www.sciencedirect.com/science/article/pii/S1047320301904870
14. Codella, N., Rotemberg, V., Tschandl, P., Celebi, M.E., Dusza, S., Gutman, D., Helba, B., Kalloo, A., Liopyris, K., Marchetti, M., Kittler, H., Halpern, A.: Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic) (2018)

15. Codella, N.C., Gutman, D., Celebi, M.E., Helba, B., Marchetti, M.A., Dusza, S.W., Kalloo, A., Liopyris, K., Mishra, N., Kittler, H., et al.: Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). In: Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on. pp. 168–172. IEEE (2018)
16. Combalia, M., Codella, N.C.F., Rotemberg, V., Helba, B., Vilaplana, V., Reiter, O., Carrera, C., Barreiro, A., Halpern, A.C., Puig, S., Malvehy, J.: Bcn20000: Dermoscopic lesions in the wild (2019)
17. Fiorese, M., Peserico, E., Silletti, A.: Virtualshave: Automated hair removal from digital dermatoscopic images. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 5145–5148 (2011). https://doi.org/10.1109/IEMBS.2011.6091274
18. Gutman, D., Codella, N.C.F., Celebi, E., Helba, B., Marchetti, M., Mishra, N., Halpern, A.: Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (isbi) 2016, hosted by the international skin imaging collaboration (isic) (2016)
19. Horé, A., Ziou, D.: Image quality metrics: Psnr vs. ssim. In: 2010 20th International Conference on Pattern Recognition. pp. 2366–2369 (2010). https://doi.org/10.1109/ICPR.2010.579
20. Huang, A., Kwan, S.Y., Chang, W.Y., Liu, M.Y., Chi, M.H., Chen, G.S.: A robust hair segmentation and removal approach for clinical images of skin lesions. In: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). pp. 3315–3318 (2013). https://doi.org/10.1109/EMBC.2013.6610250
21. ISIC: Isic archive gallery. Online (Jul 2020), https://www.isic-archive.com
22. Jam, J., Kendrick, C., Drouard, V., Walker, K., Hsu, G.S., Yap, M.H.: R-mnet: A perceptual adversarial network for image inpainting. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 2714–2723 (2021)
23. Kassem, M.A., Hosny, K.M., Fouad, M.M.: Skin lesions classification into eight classes for isic 2019 using deep convolutional neural network and transfer learning. IEEE Access **8**, 114822–114832 (2020). https://doi.org/10.1109/ACCESS.2020.3003890
24. Kiani, K., Sharafat, A.R.: E-shaver: An improved dullrazor® for digitally removing dark and light-colored hairs in dermoscopic images. Computers in Biology and Medicine **41**(3), 139–145 (2011). https://doi.org/https://doi.org/10.1016/j.compbiomed.2011.01.003, https://www.sciencedirect.com/science/article/pii/S0010482511000047
25. Lee, T., Ng, V., Gallagher, R., Coldman, A., McLean, D.: Dullrazor®: A software approach to hair removal from images. Computers in Biology and Medicine **27**(6), 533–543 (1997). https://doi.org/https://doi.org/10.1016/S0010-4825(97)00020-6, https://www.sciencedirect.com/science/article/pii/S0010482597000206
26. Li, W., Joseph Raj, A.N., Tjahjadi, T., Zhuang, Z.: Digital hair removal by deep learning for skin lesion segmentation. Pattern Recognition **117**, 107994 (2021). https://doi.org/https://doi.org/10.1016/j.patcog.2021.107994, https://www.sciencedirect.com/science/article/pii/S0031320321001813
27. Maglogiannis, I., Delibasis, K.: Hair removal on dermoscopy images. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). pp. 2960–2963 (2015). https://doi.org/10.1109/EMBC.2015.7319013

28. Maron, R.C., Hekler, A., Krieghoff-Henning, E., Schmitt, M., Schlager, J.G., Utikal, J.S., Brinker, T.J.: Reducing the impact of confounding factors on skin cancer classification via image segmentation: Technical model study. Journal of Medical Internet Research **23** (2021)
29. Nauta, M., Walsh, R., Dubowski, A., Seifert, C.: Uncovering and correcting shortcut learning in machine learning models for skin cancer diagnosis. Diagnostics **12**(1) (2022). https://doi.org/10.3390/diagnostics12010040, https://www.mdpi.com/2075-4418/12/1/40
30. Pewton, S.W., Yap, M.H.: Dark corner on skin lesion image dataset: Does it matter? In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. pp. 4831–4839 (June 2022)
31. Rotemberg, V., Kurtansky, N., Betz-Stablein, B., Caffery, L., Chousakos, E., Codella, N., Combalia, M., Dusza, S., Guitera, P., Gutman, D., Halpern, A., Helba, B., Kittler, H., Kose, K., Langer, S., Lioprys, K., Malvehy, J., Musthaq, S., Nanda, J., Soyer, P.: A patient-centric dataset of images and metadata for identifying melanomas using clinical context. Scientific Data **8**, 34 (01 2021). https://doi.org/10.1038/s41597-021-00815-z
32. Rott Shaham, T., Dekel, T., Michaeli, T.: Singan: Learning a generative model from a single natural image. In: Computer Vision (ICCV), IEEE International Conference on (2019)
33. Salido, J.A.A., Ruiz, C.: Using morphological operators and inpainting for hair removal in dermoscopic images. In: Proceedings of the Computer Graphics International Conference. CGI '17, Association for Computing Machinery, New York, NY, USA (2017). https://doi.org/10.1145/3095140.3095142, https://doi.org/10.1145/3095140.3095142
34. Sethian, J.A.: A fast marching level set method for monotonically advancing fronts. Proceedings of the National Academy of Sciences **93**(4), 1591–1595 (1996)
35. Shaham, T.R., Dekel, T., Michaeli, T.: Singan: Learning a generative model from a single natural image. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4570–4580 (2019)
36. Song, X., Guo, S., Han, L., Wang, L., Yang, W., Wang, G., Anil Baris, C.: Research on hair removal algorithm of dermatoscopic images based on maximum variance fuzzy clustering and optimization criminisi algorithm. Biomedical Signal Processing and Control **78**, 103967 (2022). https://doi.org/https://doi.org/10.1016/j.bspc.2022.103967, https://www.sciencedirect.com/science/article/pii/S1746809422004669
37. Sultana, A., Dumitrache, I., Vocurek, M., Ciuc, M.: Removal of artifacts from dermatoscopic images. In: 2014 10th International Conference on Communications (COMM). pp. 1–4 (2014). https://doi.org/10.1109/ICComm.2014.6866757
38. Talavera-Martínez, L., Bibiloni, P., González-Hidalgo, M.: Comparative study of dermoscopic hair removal methods. In: Tavares, J.M.R.S., Natal Jorge, R.M. (eds.) VipIMAGE 2019. pp. 12–21. Springer International Publishing, Cham (2019)
39. Telea, A.: An image inpainting technique based on the fast marching method. Journal of graphics tools **9**(1), 23–34 (2004)
40. Toossi, M.T.B., Pourreza, H.R., Zare, H., Sigari, M.H., Layegh, P., Azimi, A.: An effective hair removal algorithm for dermoscopy images. Skin Research and Technology **19** (2013)
41. Tschandl, P.: The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions (2018). https://doi.org/10.7910/DVN/DBW86T, https://doi.org/10.7910/DVN/DBW86T

42. Wang, Z., Bovik, A.C.: Mean squared error: Love it or leave it? a new look at signal fidelity measures. IEEE Signal Processing Magazine **26**(1), 98–117 (2009). https://doi.org/10.1109/MSP.2008.930649
43. Xie, F.Y., Qin, S.Y., Jiang, Z.G., Meng, R.S.: Pde-based unsupervised repair of hair-occluded information in dermoscopy images of melanoma. Computerized Medical Imaging and Graphics **33**(4), 275–282 (2009). https://doi.org/https://doi.org/10.1016/j.compmedimag.2009.01.003, https://www.sciencedirect.com/science/article/pii/S0895611109000056
44. Xie, Y., Zhang, J., Xia, Y., Shen, C.: A mutual bootstrapping model for automated skin lesion segmentation and classification. IEEE Transactions on Medical Imaging **39**, 2482–2493 (2020)
45. Zand, H., Nguyen, N., Zeinali, B., Chang, J.: A new preprocessing approach to improve the performance of cnn-based skin lesion classification. Medical Biological Engineering Computing **59**, 1–9 (04 2021). https://doi.org/10.1007/s11517-021-02355-5
46. Zhou Wang, Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4), 600–612 (2004). https://doi.org/10.1109/TIP.2003.819861