Received: 1 October 2021 | Accepted: 17 January 2022

DOI: 10.1111/srt.13145

ORIGINAL ARTICLE

WILEY

Hair removal in dermoscopy images using variational autoencoders

Dalal Bardou[1] | **Hamida Bouaziz**[2] | **Laishui Lv**[3] | **Ting Zhang**[3]

[1] Department of Computer Science and Mathematics, University of Abbes Laghrour, Khenchela, Algeria

[2] Mécatronique Laboratory, Department of Computer Science, Jijel University, Jijel, Algeria

[3] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China

Correspondence
Dalal Bardou, Department of Computer Science and Mathematics, University of Abbes Laghrour, Khenchela, Algeria.
Email: dalal.bardou@njust.edu.cn

Abstract

Background: In recent years, melanoma is rising at a faster rate compared to other cancers. Although it is the most serious type of skin cancer, the diagnosis at early stages makes it curable. Dermoscopy is a reliable medical technique used to detect melanoma by using a dermoscope to examine the skin. In the last few decades, digital imaging devices have made great progress which allowed capturing and storing high-quality images from these examinations. The stored images are now being standardized and used for the automatic detection of melanoma. However, when the hair covers the skin, this makes the task challenging. Therefore, it is important to eliminate the hair to get accurate results.

Methods: In this paper, we propose a simple yet efficient method for hair removal using a variational autoencoder without the need for paired samples. The encoder takes as input a dermoscopy image and builds a latent distribution that ignores hair as it is considered noise, while the decoder reconstructs a hair-free image. Both encoder and decoder use a decent convolutional neural networks architecture that provides high performance. The construction of our model comprises two stages of training. In the first stage, the model has trained on hair-occluded images to output hair-free images, and in the second stage, it is optimized using hair-free images to preserve the image textures. Although the variational autoencoder produces hair-free images, it does not maintain the quality of the generated images. Thus, we explored the use of three-loss functions including the structural similarity index (SSIM), L1-norm, and L2-norm to improve the visual quality of the generated images.

Results: The evaluation of the hair-free reconstructed images is carried out using t-distributed stochastic neighbor embedding (SNE) feature mapping by visualizing the distribution of the real hair-free images and the synthesized hair-free images. The conducted experiments on the publicly available dataset HAM10000 show that our method is very efficient.

KEYWORDS
dermoscopy images, hair occlusion, hair removal, perceptual loss, variational autoencoders

1 | INTRODUCTION

The skin is the largest organ of the human body.[1] It has for purpose to protect us from infections, and it also acts as a body temperature regulator and sensation detector. Skin cancer incidence has drastically increased in recent years, while melanoma is considered the most serious type that rises at a faster rate in the last 50 years compared to other cancers.[2–7] For example, in 2020, in the United States alone, it was predicted that approximately 100 350 individuals would be diagnosed with melanoma compared to the year 2017 where the estimated number was about 87 110 individuals.[8,9]

The diagnosis of melanoma of skin at early stages makes it curable. Thus, it is important to provide better methods for early detection.[6] However, there is a problem of a dermatology workforce crisis.[10] This could be considered as a threat not only to the current quality of the provided care services but also to the future demand of these services with the growing population. Moreover, the diagnosis is a subjective task where we rely on the expertise of the practitioner.[7,11,12] The advances in technology, precisely imaging techniques, over the last decade allowed acquiring and collecting high-quality skin cancer images.[13] These images are now being standardized, which empowered researchers to conduct studies and develop reliable computer-aided tools to automatize the diagnosis and provide aid to the practitioners.

Different skin lesion artifacts can be present in dermoscopy images such as hair, which is considered the artifact with the heaviest effect on the quality of the diagnosis, ruler markers, air bubbles, ink markings… etc.[14] The presence of hair in dermoscopy images leads to the occlusion of the information of the lesion concerning its texture and boundary.[15] Hence, it is important to eliminate this artifact in the preprocessing step before carrying out lesion segmentation.[15–17] The obvious solution is to exclude the images with hair from the dataset, which affects the diversity of the data.[18–19] Another solution is to use a razor to shave the hair. But, this solution is impractical as it causes bleeding and other problems.[18–20] Besides, it is time-consuming. Thus, there is an excessive need to come up with new automatic methods to remove the hair and keep the quality of the dermoscopic images unaltered.

Several digital hair removal methods were developed to deal with the problem of hair. The first developed ones relied on image processing, and they comprise two tasks: hair segmentation and gap inpainting.[11,15–17] These methods suffer from many limitations: (1) They require a certain amount of expertise to choose a good method for hair segmentation and another one for gap inpainting among many. (2) They also do not allow detecting thin (false negatives), overlapping hairs, and hairs of similar color to underlying skin.[21] (3) Besides, there is the problem of false positives where objects that have hair-like structures can be detected during hair segmentation.

On the other hand, deep learning came with more developed solutions that use an end to end learning and yield better models. Generative models, a class of deep learning that is used to generate realistic-looking samples showed great success in many fields including the problem of hair in dermoscopy images. Autoencoders (AEs) are a family of generative models that were used by researchers for hair removal.[22] However, the proposed models require paired samples for the training[22] that consist of the hair-free image and its correspondent hair-occluded image, which is difficult to acquire due to the unavailability of such datasets.

In this paper, we propose a simple yet efficient method for hair removal in dermoscopy images using deep learning. This is achieved by using a variational AE that considers hair as noise when learning the latent representation and reconstructs hair-free images using an encoder-decoder architecture. The proposed variational AE model is trained in an end-to-end fashion. Our contributions are summarized as follows:

1. We design an encoder-decoder architecture based on a variational AE model that learns efficiently how to eliminate hair.
2. The proposed method does not require paired samples for the training of the model.
3. To provide a better reconstruction of the images, we combine the basic loss with perceptual loss and other loss functions to force the model to reconstruct accurate images.
4. We carry out two-stage training to maintain the quality of the hair-free generated images.

2 | RELATED WORK

2.1 | Traditional methods for hair removal

Hair removal in dermoscopy images is an active research area where new methods are constantly being implemented to tackle this problem. The first attempts for hair removal used image processing operations to eliminate hair (summarized in Table 1). The earliest one was proposed by Lee et al.[23] in 1997 where they proposed an algorithm and developed a software called Dullrazor. The algorithm involves the use of the morphological closing operation to identify dark hair pixels and after replacing them with non-hair pixels by bilinear interpolation.[24,25] This algorithm considers only thick dark hairs. Thus, an improved version of Dullrazor named E-shaver was proposed by Kiani et al.[24] where light-color hairs are considered along with dark hairs. Fiorese et al.[26] also proposed an automatic shaver named VirtualShave. The main idea consists of identifying hair using a top-hat filter and morphological postprocessing and then replacing hair pixels by applying partial differential equations (PDE)-based inpainting. Similarly, Xie et al.[25] processed hair through a morphologic closing-based top-hat operator for image enhancement. Afterward, they used the statistic threshold for segmentation followed by hair extraction using an elongate feature function. To repair the occluded information, the authors applied PDE-based image inpainting. The same authors assessed the degree of hair occlusion in dermoscopy images in Xie et al.[27] depending on the distribution of hair in the image in which they proposed a local adaptive hair detection method that works well for both sparse and dense hair. Abbas et al.[15] proposed an algorithm for hair segmentation and repairing. To detect hair, the authors applied the derivative of Gaussian and refined it using a morphological operation, and to repair the

TABLE 1 The summarized state-of-the-art methods related to hair removal in dermoscopy images

Year	Work	Software name	Hair segmentation	Hair-pixels repair method	Type of hair	Color space	Data set	Type of validation
1997	Lee et al.[23]	Dullrazor	Grayscale morphological closing	Bilinear interpolation	Thick dark hair	RGB	Captured images	Real hair
2009	Xie et al.[25]	/	Top-hat operator	Anisotropic diffusion	Weak and thick hair	RGB	40 images	Real hair, manual hair
2011	Kiani et al.[24]	E-shaver	Edge detector	Color averaging	Thick dark hair and light hair	Grayscale	50 images	Real hair
2011	Fiorese et al.[26]	VirtualShave	Top-hat operator	PDE-based inpainting	/	RGB	20 images	Real hair
2011	Abbas et al.[15]	/	Derivative of Gaussian	Coherence transport	/	CIELab, RGB	100 images	Manually added hair
2013	Toossi et al.[28]	/	Canny edge detector	Multiresolution coherence transport	/	Grayscale	50 images	Real hair
2013	Huang et al.[29]	/	Multiscale matched filters	Median filtering	Fine hair, hair in the shade	RGB	20 images	Real hair
2015	Xie et al.[27]	/	Adaptive threshold	/	sparse and dense hair	RGB	110 images	Real and Simulated hair
2016	Koehoorn et al.[16]	/	Multiscale skeletons/morphological operators	Fast marching method	Long and stubble hair	RGB	300 images	Real hair
2017	Bibiloni et al.[30]	/	Morphological curvilinear object detector	morphological-based inpainting	/	CIELab	PH2 dataset	Real hair
2020	Talavera-Martínez et al.[22]	/	Convolutional encoder-decoder		Long thin thick dark and light hair	RGB	PH2, dermquest, dermis, EDRA2002, ISIC	Simulated hair

Abbreviations: ISIC, The International Skin Imaging Collaboration; Nbr, number; RGB, red, green, blue.

hair pixels, they applied the fast matching inpainting method. Similar to Abbas et al.,[15] Toossi et al.[28] used an adaptive canny edge detector to detect light and dark hairs and then improved it through morphological operators. Huang et al.[29] found out that Dullrazor does not work well on images with fine hair and hair in the shade. Therefore, they proposed a better version that uses thresholding with hysteresis to detect hair, and region growing algorithms to recover the more complicated hair intersection patterns. Bibiloni et al.[30] proposed an efficient hair removal filter that comprises an object detector based on morphological curvilinear and a morphological-based inpainting algorithm, while Koehoorn et al.[16] developed a threshold-set decomposition-based model to detect and remove long and stubble hair.

2.2 | Deep learning-based methods

Contrary to the aforementioned methods, Talavera-Martínez et al.[22] relied on deep learning to address this problem. The authors proposed a convolutional encoder-decoder-based architecture that consists of 12 layers, and it uses pairs of images for training. The model takes the hair-free image and its correspondent image with simulated hair as input. The loss function used to construct and evaluate the model is an explicit linear loss function that combines five weighted loss terms.

3 | MATERIALS AND METHODS

Our proposed hair removal method is based on variational AEs (VAEs), which are a variant of the AEs. The variational AE is a neural network that consists of the encoder and decoder. The encoder compresses the input data into a lower-dimensional representation vector and the decoder does the opposite. It takes the representation vector and reconstructs it back to the original domain. During training, the VAE seeks to find the weights that minimize the computed loss between the input and output images after passing through both the encoder and the decoder. The variational AE can be successfully used to denoise images by letting the encoder learns to ignore the noise and build noise-free positions in the latent space.[31] In this work, we assume that the hair presented in the dermoscopy images is noise, and we design a variational AEs architecture that learns how to eliminate hair and reconstruct noise-free (hair-free) images. The training of our architecture does not require paired images, which usually consist of the reference image and its correspondent hair-free image. It only takes the image with hair as input instead of the hair image along with its correspondent hair-free image. This architecture succeeds to remove the hair but it does not provide a good restoration of the reconstructed images. Thus, we opt for applying a quality loss named perceptual loss to preserve the quality of the reconstructed images. The full loss function can be written as follows: $\text{Loss}_{\text{total}} = \text{Loss}_{\text{VAE}} + \text{Loss}_{\text{Quality}}$. The proposed architecture is depicted in Figure 1. The detailed architecture of the encoder along with the decoder is given in Figure 2. In Subsection 3.1, we will provide a brief introduction about the variational AE and also the detailed architecture of the encoder and decoder, while in Subsection 3.2, we define the loss function. In Subsection 3.3, we give the training parameters used to train the model.

3.1 | Variational AE

A short introduction of the variational AEs is given in this subsection along with the architecture of the encoder and the decoder we used to build our model. Variational AE was proposed by Kingma et al.[32,33] in 2013. It is an advanced variant of the conventional AE, which uses a combination of deep neural networks and Bayesian inference.[34] VAE maps a given input as a multivariate normal distribution in the latent space instead of a single point. The encoder is the inference model and is defined as an approximate posterior distribution $q_\varphi(z|x)$, modeled by a neural network, with variational parameters denoted φ.[33,35] x is the input data; z is the latent variable where $z = \text{encoder}(x) \sim q_\varphi(z|x)$.[36] The decoder decodes z back to the original input $\bar{x}$ and is also modeled by a neural network, where $\bar{x} = \text{decoder}(x) \sim p_\theta(x|z)$ and θ are the parameters of the decoder. The encoder outputs the parameters of the latent distribution used to sample yielding the input to the decoder.[37]

The objective function of the VAE is given below, where φ and θ are the parameters. To optimize the parameters θ and φ, stochastic gradient descent is used.

$$L\ (\theta, \varphi, x) = E_{q_\varphi}\ (z|x)\ [\log p_\theta\ (x|z)] - D_{\text{KL}}\ \left[q_\varphi\ (z|x) p_\theta\ \|(z)\ \right]$$

Where $E_{q_\varphi}(z|x)[\log p_\theta(x|z)]$: is the reconstruction term.

$D_{\text{KL}}[q_\varphi(z|x)\|p_\theta(z)]$ is the Kullback–Leibler divergence and is considered as a regularization term. The D_{KL} is the difference between the approximate posterior and the prior.[38] The standard Gaussian distribution $N \sim (0, 1)$ is usually chosen as the prior.

The architecture of both the encoder and the decoder consists of two convolutional layers. This number of layers was chosen empirically. The number of filters, kernel size, and activation function details are given in Table 2. The latent dimension is set to 50.

3.2 | Loss function

The loss function aims to evaluate how good is the methods we propose for modeling our datasets. During the training phase, the loss function is used to compute the error (the loss) between the ground truth and the output of the trained model. The computed error indicates how good/bad our model is. The basic variational AE loss function consists of two terms as described in Section 3.1. This loss function successfully outputs hair-free images. But, it does not restore well the images; therefore, we investigated the use of L_1 loss, L_2 loss, and perceptual loss together with the VAE loss intending to overcome this problem by finding the best weights in the weights space that not only remove hairs but also output intact reconstructed images.

Mean square error (MSE), also known as L_2, is a loss measure that has been widely adopted to address many problems such as

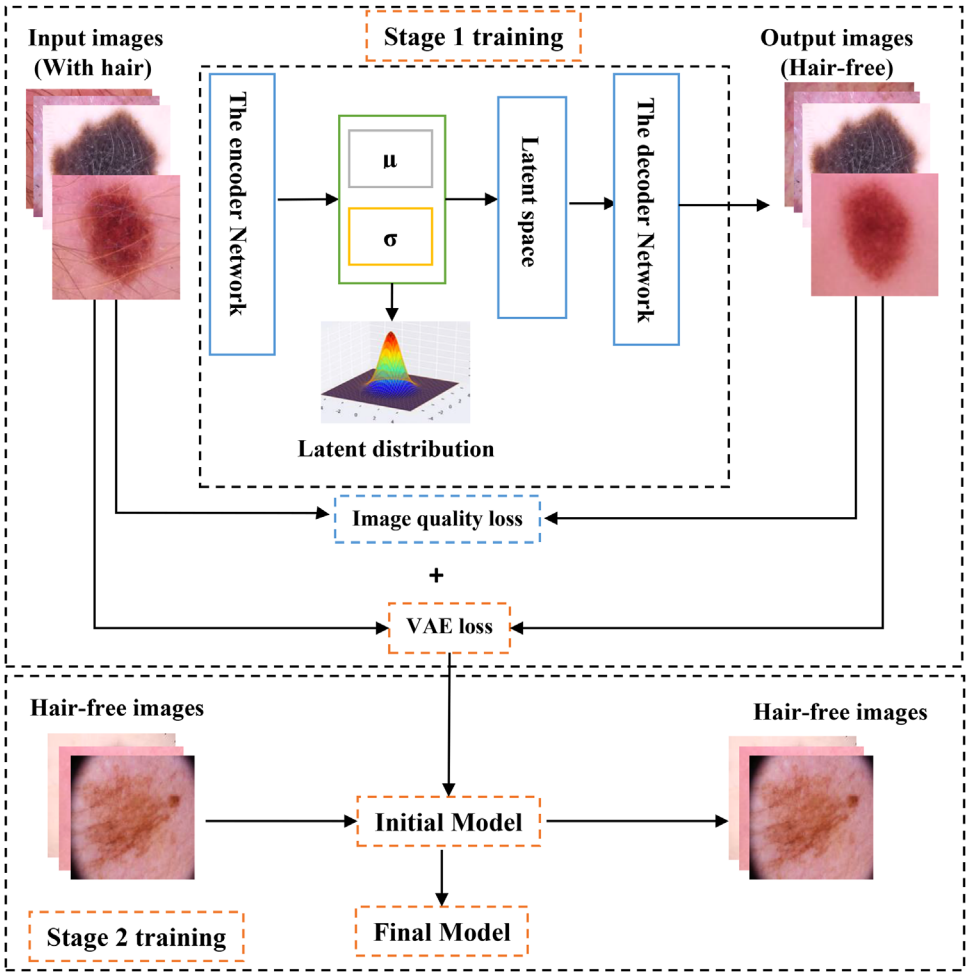

FIGURE 1 The architecture of the proposed model

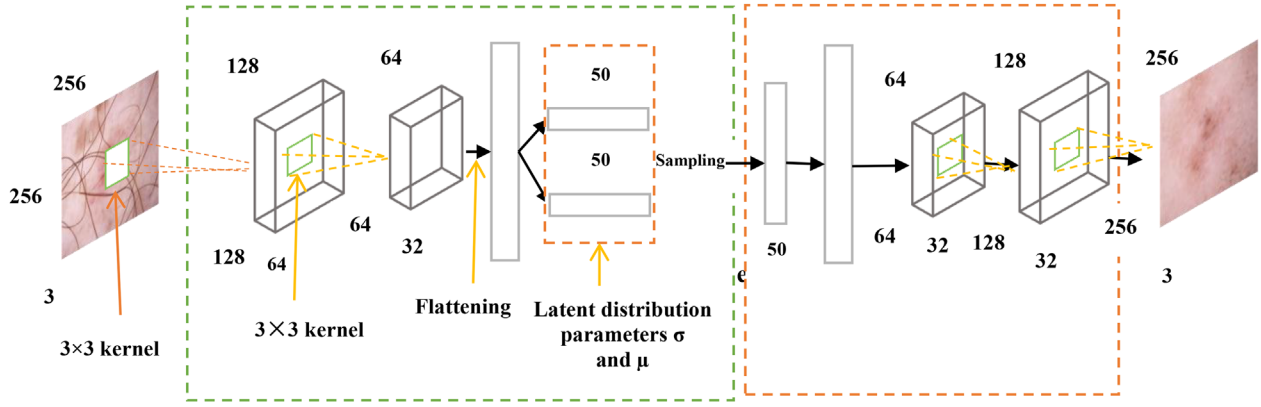

FIGURE 2 The detailed architecture of the encoder and decoder

TABLE 2 The detailed architecture of the encoder and decoder

The encoder architecture					The decoder architecture			
	Kernel size	Activation	Stride	Nbr of filters	Kernel size	Activation	Stride	Nbr of filters
Layer 1	3	Relu	2	64	3	Relu	2	32
Layer 2	3	Relu	2	32	3	Relu	2	64

regression and signal and image processing.[39,40] The MSE is computed as follows:

$$L_2 = \sum_{i=1}^{n} \left(y_{\text{true}} - y_{\text{predicted}}\right)^2 \quad (1)$$

Mean average error, also known as L_1, is another loss measure that is computed as follows:

$$L_1 = \sum_{i=1}^{n} \left|y_{\text{true}} - y_{\text{predicted}}\right| \quad (2)$$

Despite the known advantages of L_2,[40] L_1 was claimed to be better regarding the quality of images in terms of blurring.[41,42] Another drawback of L_2 is its negative effect on the visual quality as it over penalizes the loss during the training.[43]

The generative model VAE uses per-pixel loss to construct the model, but it generates blurry images that do not suit the human visual system compared to input images.[11,44] Therefore, perceptually motivated methods were developed to tackle this problem by using metrics that optimize the reconstruction of the images. Structural similarity index[45,46] (SSIM) is among the proposed metrics that takes into account the human visual system perspective. Since L_1 and L_2 losses rely on the evaluation of the per-pixel loss, SSIM came to overcome their limitations where illumination and color-related information is evaluated, and a similarity index is computed between the input and the output image patches. The SSIM loss function is defined as follows:

$$L_{\text{SSIM}} = 1 - \text{SSIM}\,(x, y) \quad (3)$$

where x is the input image patch, and y is the output image patch, and:

$$\text{SSIM}\ (x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_{x+}^2\mu_y^2 + C_1} \cdot \frac{2\sigma_{xy} + C_2}{\sigma_x^2\sigma_y^2 + C_2} \quad (4)$$

Where μ_x and μ_y are the mean of the image x and y patches, respectively. σ_x^2 and σ_y^2 are the variance within the patches. σ_{xy} is the covariance of the local patches. C_1 and C_2 are two constants with small values used as stabilizing parameters.[47]

L_1, L_2 loss, and perceptual loss functions have already shown a remarkable amelioration in the learning of generative adversarial neural networks.[11,48–51] Thus, in this work, we investigate the use of them making up part of the per-pixel loss function aiming into improving the learning of the VAE.

3.3 | Training parameters

Our implementation is based on variational neural networks. We did choose the best architecture of the decoder and encoder. We tested simple and deep architectures with many layers, but a decent architecture led to the best performances. Batch normalization was also put to test. However, it did not have any effect on the optimization performance. The model was trained using Adam optimizer, and the learning rate was set to 0.0001 in the first stage of training and 0.0005 in the second stage. The batch size is set to 16, while the number of epochs is set to 200 in the first stage of training and 3 in the second stage. In the second stage of training, we have chosen three epochs because, after this number, the color intensities of images are getting affected.

FIGURE 3 Some samples from the HAM10000 dataset. Samples with thin and thick hair, light and dark hair, sparse and dense hair, short and long hair

4 | EXPERIMENTAL EVALUATION AND ANALYSIS

In this section, we will provide the achieved results applying our proposed method and also the data we used to train and validate our model. Moreover, we make a comparison with the state-of-the-art methods.

4.1 | Data

In this work, we have used the HAM10000 ("Human Against Machine with 10 000 training images") dataset, which is a large collection of multi-source dermoscopy images.[52] This dataset is available on the The International Skin Imaging Collaboration (ISIC) archive and contains images of different pigmented skin lesions. It was created to overcome the problem of a small-size dataset. The total number of samples is 10 015 images. We have manually divided the dataset into hair-occluded images and hair-free images. The total number of hair-occluded images is 8876, while the total number of hair-free images is 1139. All the images were resized to 256 × 256. In the training phase, we only used 3000 hair-occluded images. For validation, we used 100 images. We also used 1000 hair-free images in the second stage of training. We tried to vary our training data so it contains images with thin and thick hair, light and dark hair, sparse and dense hair, and short and long hair. Figure 3 shows some samples from our training data.

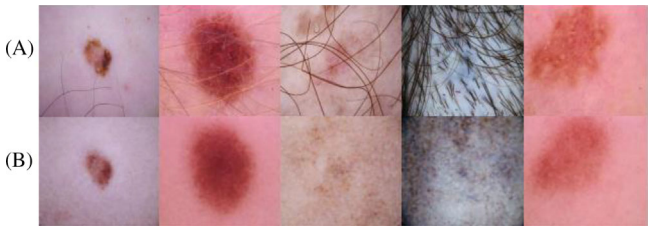

FIGURE 4 Results of the baseline configuration. (A) Images occluded with real hair. (B) Correspondent hair-free images after applying the baseline configuration

4.2 | Baseline configuration

In the baseline configuration, we used the basic VAE without adding any metrics. The trained model has been validated on100 hair-occluded images. In the phase of training, the model was trained for the first 200 epochs on the dataset with hair using Adam optimizer. The training is then again resumed on the dataset with hair-free images for three other epochs using the same optimizer. The additional training on hair-free images outputted improved texture of the images. Figure 4 shows the results after applying our VAE-based method on some image samples. The proposed baseline configuration successfully removed hair from all images. However, the reconstructed images are not 100% identical to the input images, and they look blurry.

4.3 | VAE with L_1 and L_2 losses

In this configuration, we use L_1 and L_2 losses separately and combined, added to the VAE loss to maintain the quality of the hair-free generated images by the VAE. L_1 and L_2 losses are considered as regularization terms. Figure 5 shows the results of L_1, L_2, and their combination, respectively, added to the basic VAE on some validation images. The application of these losses showed an improvement in the reconstruction of the VAE model compared to the baseline configuration for all the samples except for the ones occluded with dense hairs.

4.4 | VAE with perceptual loss

In this configuration, we try on perceptual loss comprising the SSIM loss model on the ImageNet dataset. For SSIM, we used the following parameters: The Gaussian filter is set to 3. The filter sigma is set to 1.5. The combination of the perceptual loss with L_1 and L_2 was also examined. Figure 6 shows the results of applying SSIM with L_1 and L_2. To evaluate the performance of the models quantitatively, we could not use the hair-occluded images. Therefore, we used a hair simulator developed by Attia et al.[11] to superimpose hair on hair-free images and then apply the different models to remove hair (Figure 7). After, we computed SSIM and MSE and compared them to the hair-free images. The results given in Table 3 show that the combination of SSIM with L_2 provided the highest performance compared to the other models.

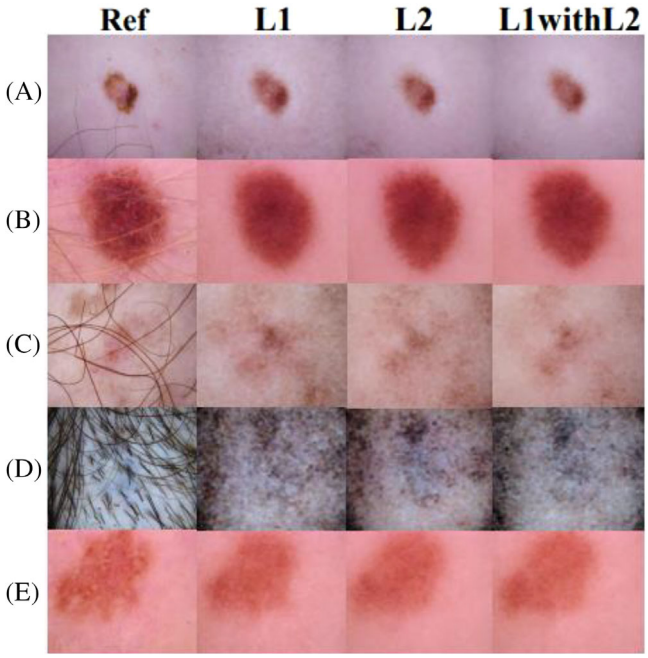

FIGURE 5 Illustration of the output of the modified variational autoencoder (VAE) on real hair occluded images for five examples

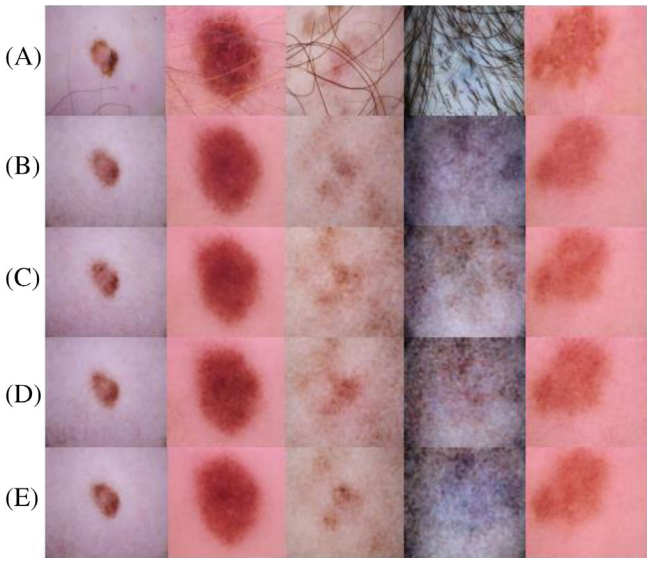

FIGURE 6 The output of the variational autoencoder (VAE) model after adding structural similarity index (SSIM) loss. (A) The reference images. (B) The result after adding SSIM. (C) The result after adding SSIM with L_1. (D) The result after adding SSIM with L_2. (E) The result after adding SSIM with L_1 and L_2

4.5 | Comparison to the state of the art methods

Our proposed method showed its efficiency in removing hair as given in the experiment section. Now, we compare it with state-of-the-art methods. Lee et al.[23] were the first ones who addressed this problem. They developed software called DullRazor that works only on thick

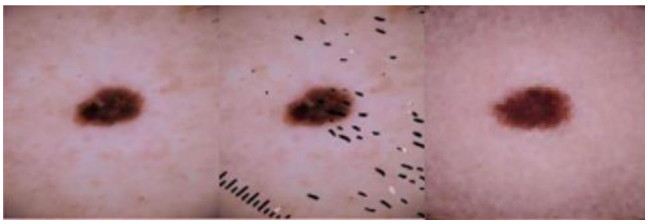

FIGURE 7 The application of variational autoencoder (VAE) loss combined with structural similarity index (SSIM) and L_2 for the elimination of synthetized hair

TABLE 3 The calculation of structural similarity index (SSIM) and mean square error (MSE) for the different models. For SSIM, the higher the better

Loss function	SSIM	MSE
SSIM	0.7700	163.580
SSIM + L_1	0.7775	171.481
SSIM + L_2	0.8053	118.24
SSIM + L_1 + L_2	0.7698	163.58

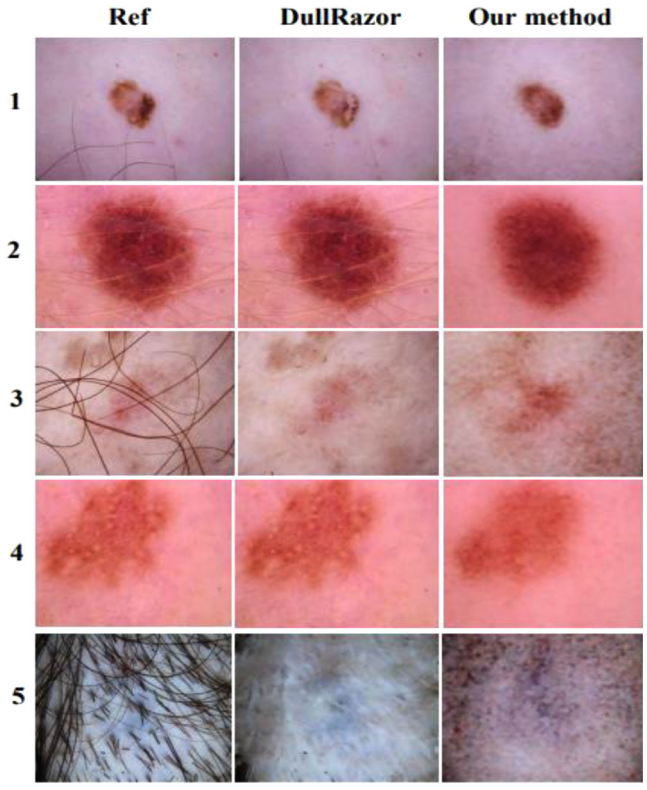

FIGURE 8 Comparison of our model with DullRazor[23]

dark hairs. Figure 8 shows the performance of the DullRazor and our method on some sample images. DullRazor did not remove completely the hair in all the images even for thick dark hair such as in Figures 8 (3), while our method successfully eliminated all kinds of hair.

The method of Lee et al.[23] involves using traditional techniques comprising morphological closing for hair segmentation and bilinear interpolation for hair repair. These techniques obviously will not work well due to the complexity of the task. Although Dullrazor does not remove all the dense hair (Figures 8 (5)), it provides better reconstruction compared to our method. The reason is that DullRazor replaces hair pixels with non-hair pixels by bilinear interpolation. The authors in references[15–16,24–30] also proposed several works as an extension of the work of Lee et al.[23] Different hair segmentation and hair repair techniques were used where each method has its drawbacks. Besides, these studies were validated on small datasets (20, 40, 50 images… etc.). Talavera-Martínez et al.[22] tried out deep learning by proposing a convolutional encoder-decoder architecture that consists of 12 layers. The training of the model was based on paired images. For each hair-free image, a simulated hair is superimposed using a hair simulator. The used loss function was composed of five weighted terms, and the values of the weights were chosen empirically. On the other hand, our method does not require paired images. We only need to provide the hair-occluded images as input to the model. As the work in Talavera–Martínez et al.[22] uses simulated hair, which does not look very realistic, it is hard to judge its efficiency. Moreover, our proposed architecture is simpler than the one proposed by Talavera–Martínez et al.[22] Ours consists only of four layers. The loss function used in Talavera–Martínez et al.[22] is also complex compared to ours. There is also the problem of weight selection as the weight of each term was chosen empirically.

Due to the unavailability of the code for works proposed in Talavera–Martínez et al.,[22] we were not able to make a direct comparison with our proposed method. Therefore, we opted for using t-distributed stochastic neighbor embedding (SNE) mapping to the distribution of the hair-free images generated using our model, hair-free images generated using Dullrazor, and hair-free images taken from the HAM10000 dataset. We used the pretrained model[53] for melanoma detection to extract deep features from the last fully connected layer, which outputs a vector of size 2048*1 and then projected them. Figure 9 shows that the sets of images belong to the same data distribution with the higher closeness between the ground truth and the images generated by our method.

5 | CONCLUSION

In this paper, we have proposed a new method based on generative models for hair removal in dermoscopic images without the need for paired samples contrary to the previous method found in the literature, which used an encoder-decoder architecture. We have used a variational AE that ignores hairs during the reconstruction phase to construct hair-free images. We have also proposed an efficient encoder-decoder architecture with few layers that leads to better performance. The training of the model comprised two training stages. The first stage of training aimed at eliminating hair, while the second stage helped to optimize the model parameters to produce images with the same features as the input images. Although the VAE gives the desired output, it does not maintain the quality of the images. Therefore, we have opted

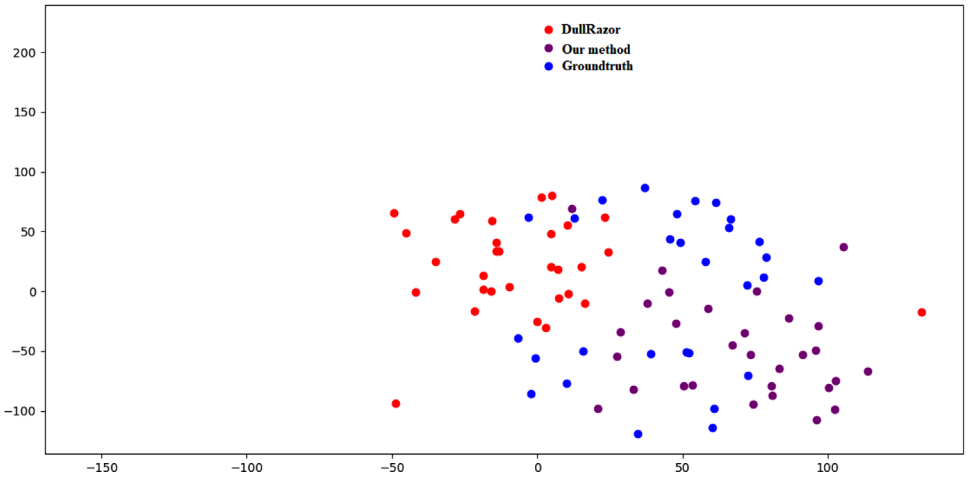

FIGURE 9 Data visualization using t-SNE for real hair-free images, DullRazor, and for images generated using our method with structural similarity index (SSIM) and L_2 loss

for using a combined loss function by applying a variety of metrics that allows not only having hair-free images but also provides a good reconstruction of the images. The perceptual loss added to the basic loss has provided an improvement into generated images by the VAE. Our proposed method can be applied to generate hair-free images that can be used for data augmentation to address the problem of lesion segmentation and skin cancer image classification. To enhance the quality of the generated images, we may try in our future work to rely on disentangle representation to separate hair and skin without impacting skin pixels, which showed great success in many fields such as pose-invariant face recognition and face completion under structured occlusions.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

REFERENCES

1. Hendi A, Martinez JC. Atlas of skin cancers: a practical guide to diagnosis and treatment. Berlin/Heidelberg, Germany: Springer Science & Business Media; 2011.
2. Rigel DS, Carucci JA. Malignant melanoma: prevention, early detection, and treatment in the 21st century. CA Cancer J Clin. 2000;50(4):215–36.
3. Ferlay J. Cancer incidence and mortality worldwide. Lyon, France: IARC; 2012.
4. Kosary CL, Altekruse SF, Ruhl J, Lee R, Dickie L. Clinical and prognostic factors for melanoma of the skin using SEER registries: collaborative stage data collection system, version 1 and version 2. Cancer 2014;120:3807–14.
5. Guy GP Jr, Thomas CC, Thompson T, Watson M, Massetti GM, Richardson LC, et al. Vital signs: melanoma incidence and mortality trends and projections—United States, 1982–2030. Morb Mortal Wkly Rep. 2015;64(21):591.
6. Riker AI, Zea N, Trinh T. The epidemiology, prevention, and detection of melanoma. Ochsner J. 2010;10(2):56–65.
7. Celebi ME, Codella N, Halpern A. Dermoscopy image analysis: overview and future directions. IEEE J Biomed Health Inform. 2019;23(2):474–8.
8. SEER. Melanoma of the Skin - Cancer Stat Facts. https://seer.cancer.gov/statfacts/html/melan.html. Accessed 31 July 2021.
9. Matthews NH, Li W-Q, Qureshi AA, Weinstock MA, Cho E. Epidemiology of melanoma. Exon Publications; 2017: 3–22.
10. Resneck J Jr, Kimball AB. The dermatology workforce shortage. J Am Acad Dermatol. 2004;50(1):50–4.
11. Attia M, Hossny M, Zhou H, Nahavandi S, Asadi H, Yazdabadi A. Realistic hair simulator for skin lesion images: a novel benchmarking tool. Artif Intell Med. 2020;108:101933.
12. Bafounta M-L, Beauchet A, Aegerter P, Saiag P. Is dermoscopy (epiluminescence microscopy) useful for the diagnosis of melanoma? Results of a meta-analysis using techniques adapted to the evaluation of diagnostic tests. Arch Dermatol. 2001;137(10):1343–50.
13. Goyal M, Knackstedt T, Yan S, Hassanpour S. Artificial intelligence-based image classification for diagnosis of skin cancer: challenges and opportunities. Comput Biol Med. 2020;127:104065.
14. Salido JA, Ruiz C Jr. Hair artifact removal and skin lesion segmentation of dermoscopy images. Asian J Pharm Clin Res. 2018;11(3):36.
15. Abbas Q, Emre Celebi M, García IF. Hair removal methods: a comparative study for dermoscopy images. Biomed Signal Process Control. 2011;6(4):395–404.
16. Koehoorn J, Sobiecki AC, Boda D, Diaconeasa A, Doshi S, Paisey S, et al. Automated digital hair removal by threshold decomposition and morphological analysis. International symposium on mathematical morphology and its applications to signal and image processing. Cham: Springer; 2015.
17. Attia M, Hossny M, Zhou H, Nahavandi S, Asadi H, Yazdabadi A. Digital hair segmentation using hybrid convolutional and recurrent neural networks architecture. Comput Methods Programs Biomed. 2019;177:17–30.
18. Kiani K, Sharafat AR. E-shaver: an improved DullRazor® for digitally removing dark and light-colored hairs in dermoscopic images. Comput Biol Med. 2011;41(3):139–45.
19. Emre Celebi M, Kingravi HA, Uddin B, Iyatomi H, Alp Aslandogan Y, Stoecker WV, et al. A methodological approach to the classification of dermoscopy images. Comput Med Imaging Graph. 2007;31(6):362–73.
20. Rubegni P, Burroni M, Perotti R, Fimiani M, Andreassi L, Cevenini G, et al. Digital dermoscopy analysis and artificial neural network for the differentiation of clinically atypical pigmented skin lesions: a retrospective study. J Invest Dermatol. 2002;119(2):471–4.

21. Lee I, Du X, Anthony B. Hair segmentation using adaptive threshold from edge and branch length measures. Comput Biol Med. 2017;89:314–24.
22. Talavera-Martínez L, Bibiloni P, González-Hidalgo M. Hair segmentation and removal in dermoscopic images using deep learning. Piscataway, NJ: IEEE Access; 2020.
23. Lee T, Ng V, Gallagher R, Coldman A, Mclean D. Dullrazor®: a software approach to hair removal from images. Comput Biol Med. 1997;27(6):533–43.
24. Kiani K, Sharafat AR. E-shaver: an improved DullRazor® for digitally removing dark and light-colored hairs in dermoscopic images. Comput Biol Med. 2011;41(3):139–45.
25. 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. Comput Med Imaging Graph. 2009;33(4):275–82.
26. 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; August 30–September 3, 2011; Boston, MA, USA
27. Xie F, Li Y, Meng R, Jiang Z. No-reference hair occlusion assessment for dermoscopy images based on distribution feature. Comput Biol Med. 2015;59:106–15.
28. Toossi MTB, Pourreza HR, Zare H, Sigari M-H, Layegh P, Azimi A. An effective hair removal algorithm for dermoscopy images. Skin Res Technol. 2013;19(3):230–5.
29. 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); July 3–7, 2013; Osaka, Japan.
30. Bibiloni P, González-Hidalgo M, Massanet S. Skin hair removal in dermoscopic images using soft color morphology. Conference on artificial intelligence in medicine in Europe. Cham: Springer; 2017.
31. Foster D. Generative deep learning: teaching machines to paint, write, compose, and play. Newton, MA: O'Reilly Media; 2019.
32. Kingma DP, Welling M. Auto-Encoding Variational Bayes. Stat. 2014;1050:1.
33. Kingma DP, Welling M. An introduction to variational autoencoders. Found Trends Mach Learn. 2019;12(4):307–92.
34. Amirhosain SS, Hammad A, Eshraghi P. Generation of whole building renovation scenarios using variational autoencoders. Energy Build. 2021;230:110520. https://doi.org/10.1016/j.enbuild.2020.110520
35. Zakharov N, Su H, Zhu J, Gläscher J. Towards controllable image descriptions with semi-supervised VAE. J Visual Commun Image Represent. 2019;63:102574.
36. Ishfaq H, Hoogi A, Rubin D. TVAE: Deep metric learning approach for variational autoencoder. Proc. ICLR Workshop. Vancouver, BC, Canada, April 30 - May 3, 2018.
37. Odaibo S Tutorial: Deriving the standard variational autoencoder (vae) loss function. arXiv preprint arXiv:1907.08956. 2019.
38. Zhu Q, Su J, Bi W, Liu X, Ma X, Li X, et al. A Batch Normalized Inference Network Keeps the KL Vanishing Away. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online, July 5–10, 2020.
39. Hang Z, Orazio G, Iuri F, Jan K. Loss functions for image restoration with neural networks. IEEE Trans Comput Imaging. 2017;3(1):47–57. https://doi.org/10.1109/tci.2016.2644865
40. Wang Z, Bovik AC. Mean squared error: love it or leave it? A new look at signal fidelity measures. IEEE Signal Process Mag. 2009;26(1):98–117.
41. Isola P, Zhu J-Y, Zhou T & Efros AA Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; July 21–26, 2017; Honolulu, HI.
42. Pathak D, Krahenbuhl P, Donahue J, Darrell T & Efros AA Context encoders: feature learning by inpainting. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; June 27–30, 2016; Las Vegas, NV.
43. Zhang L, Zhang L, Mou X & Zhang D A comprehensive evaluation of full reference image quality assessment algorithms. In: 2012 19th IEEE International Conference on Image Processing; September 30–October 3, 2012; Orlando, FL.
44. Hou XShen L, Sun K & Qiu G Deep feature consistent variational autoencoder. In: 2017 IEEE Winter Conference on Applications of Computer Vision (WACV); March 24–31, 2017; Santa Rosa, CA.
45. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process. 2004;13(4):600–12.
46. Zhao H, Gallo O, Frosio I & Kautz J Loss functions for neural networks for image processing. arXiv preprint arXiv:1511.08861. 2015.
47. Xiang Y, Syed Zulqarnain G, Hanlin Q, Mian Ajmal. Structural Similarity Loss for Learning to Fuse Multi-Focus Images. Sensors. 2020;20(22):6647. https://doi.org/10.3390/s20226647
48. Rad MS, Bozorgtabar B, Marti U-V, Basler M, Ekenel HK, Thiran J-P. Srobb: targeted perceptual loss for single image superresolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision; October 27–November 2, 2019; Seoul, Korea (South).
49. Galteri L, Seidenari L, Bertini M, Del Bimbo A. Deep generative adversarial compression artifact removal. Proceedings of the IEEE International Conference on Computer Vision. 2017: 4826–35.
50. Sidorov O, Wang C, Cheikh FA. Generative smoke removal. Comput Vis Pattern Recogn. 2019;116:81-92.
51. Agrawal V, Kori A, Anand VK, Krishnamurthi G. Structurally aware bidirectional unpaired image to image translation between CT and MR. arXiv preprint arXiv:2006.03374. 2020.
52. Philipp T, Cliff R, Harald K. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific Data. 2018;5(1). https://doi.org/10.1038/sdata.2018.161
53. Lequan Yu, Hao C, Qi D, Jing Q, Pheng-Ann H. Automated melanoma recognition in dermoscopy images via very deep residual networks. IEEE Trans Med Imaging. 2017;36(4):994–1004. https://doi.org/10.1109/tmi.2016.2642839

How to cite this article: Bardou D, Bouaziz H, Lv L, Zhang T. Hair removal in dermoscopy images using variational autoencoders. Skin Res Technol. 2022;28:445–454. https://doi.org/10.1111/srt.13145