Hair removal in dermoscopic images using deep multitask learning

Lidia Talavera-Martínez Pedro Bibiloni
Manuel González-Hidalgo
SCOPIA Research Group. University of the Balearic Islands
Health Research Institute of the Balearic Islands (IdISBa)
l.talavera, p.bibiloni, manuel.gonzalez@uib.es

Abstract

When analyzing dermoscopic images, the hairs and their shadows on the skin may occlude relevant information about the lesion at the time of diagnosis. In this work, we present a new approach for hair removal on dermoscopic images based on deep learning techniques, as well as study in depth the behavior of the tasks of skin lesion segmentation, hair mask segmentation, and inpainting of those regions, in a multitasking framework to discover how tasks influence each other. Moreover, we describe our built database specifically for the task of hair removal. Qualitative and quantitative results demonstrate the efficacy of our model. Finally, we conclude from the multitasking experimentation that while the inpainting task does not benefit from this type of learning, the rest of the tasks do benefit by improving their performance compared to their corresponding single-task model.

1. Introduction

Malignant skin lesions, whose incidence rate is raising, poses a major problem for public health [7]. The diagnosis of skin lesions is mainly based on their morphological features, such as an irregular shape, presence or absence of some structures, and a variety of colors [3]. However, the existence of hair in dermoscopic images often occludes significant patterns, and introduces uncertainty and imprecision, reducing the accuracy of the lesion assessment. Hence, hair removal stands out as one of the key methods in the preprocessing step during lesions analysis [5]. This process consists of first removing any hairs that may be present, to later recover the underlying texture of these areas by inpainting. So far, only traditional techniques were used for this task, despite the extensive research within the computer vision field [11]. Recently, multitask learning has become one of the most interesting approaches in computer vision applications [10]. This technique intends to solve related tasks simultaneously by improving their generalization ability. In multitask learning, some of the hidden layers of the model are shared, so they learn a joint representation, leveraging both their commonalities and differences [4]. Some of its main advantages are the reduction of the computational time, an improvement in the robustness of the model against overfitting, and a potential improvement in prediction accuracy compared to single-task models.

Here, we present the first deep learning technique used for hair removal in dermoscopic images. Also, we explore in a CNN-based multitask model the behavior of the: 1) segmentation of hairs that can occlude the lesion, 2) inpainting of these hairs regions, and 3) segmentation of the skin lesion, trying to discover how they influence each other.

2. Methodology

Architecture structure In Figure 1 we show the proposed architecture for the multitasking model, part of which is the basis for the hair removal model –an encoder-decoder convolutional model–. Autoencoders are suitable to tackle this task since it is, essentially, a denoising task. We hypothesized that the network will treat the hair as noise and will be ignored, having as output the hairless skin image. Taking into account the other tasks, we extend the proposed model with a low-resolution module for a more complete view of the context of the images. Then, high-level features of hidden representations, obtained with a different resolution by both the low-resolution module and the encoder part, are merged and fed to the decoder. Skip connections concatenate feature maps corresponding to encoder and decoder layers of equal resolution, enabling the decoder to recover image details. In terms of size, we believe that a rather small network is more suitable to correctly learn the task due to the relatively small amount of available data.

Loss function: The proposed model optimizes a weighted average of the loss functions considered for the different tasks. The dice coefficient loss is used to asses the segmentation of the lesion. The weighted binary cross-entropy loss evaluates the segmentation of hairs. For the evaluation of the inpainting of the regions with hairs, a combined loss

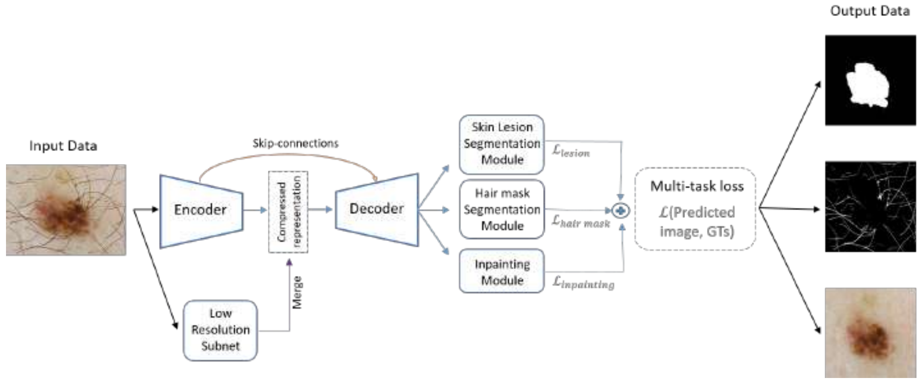

Figure 1. Architecture of our proposed multitask network for skin lesion segmentation, hair mask segmentation, and inpainting.

function is computed which focuses on measuring statistical features locally along with other per-pixel losses.

Learning details The performance of the multitask model depends on the relative weighting between the loss of each task. Thus, we use a GradNorm strategy [6] to automatically balance training by dynamically tuning each task's weight on the loss so that their contributions to the gradient in a certain shared layer are similar. Both approaches are trained following an early stopping policy based on monitoring the validation loss [6]. We apply data augmentation in the training phase to improve the model's ability to generalize (zooms, rotations, shifts and flips).

Dataset There are many public databases that provide the lesion segmentation, given its relevance in dermoscopic image analysis. However, finding images with hair, along with their corresponding hair mask, and their "clean" version, –the same image without hair–, was a challenging task. For all we know, there is no dataset with such expert information, since the same dermoscopic image cannot be captured with and without hair. To address this problem, we used three different hair simulation methods [2, 9] over hairless images extracted from five publicly available dermoscopic datasets [1, 8]. An example of simulated hair can be seen in Figure 2a. To train and evaluate the single-model for hair removal we build a dataset of 618 images, while for the multitasking model we collected a dataset with 1060 images.

3. Results and discussion

Hair removal model In Figure 3, we can see how our hair removal model, despite having been trained with synthetic images, is effective and achieves visually appealing results on dermoscopic images with real hair. Quantitatively, we performed a statistical test of the results, presented in Table 1, to objectively study and compare the performance of our approach with that of six traditional hair removal methods. We considered a set of nine measures of similarity between the hairless reference images and their corresponding model output. Our method outperforms the rest for the majority of similarity measures.

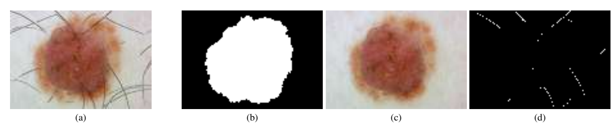

Figure 2. Test image example (a) and its three GTs for skin lesion segmentation (b), inpainting (c), and hair mask segmentation (d).

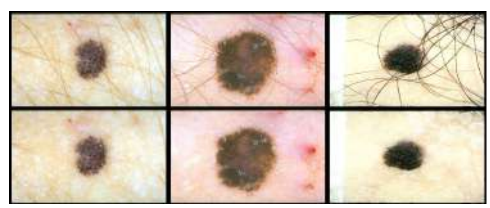

Figure 3. Example of hair removal results (bottom) obtained by our method in dermoscopic images with real hair (top).

Multitask analysis We present, in Figure 4, a visual comparison of the results obtained for each possible combination of tasks in Figure 2a. In addition, we present the relative gains and losses of the performance measures, see Table 2, as well as a bilateral statistical test between the single-model task and the same task when jointly trained with other tasks, to determine whether the latter models significantly outperform the former.

As can be seen in Table 2, in 5 out of 9 cases the results improve in terms of their performance measures. Statistically, we conclude that when we combine the three tasks, the performance metrics for the hair mask and the lesion segmentation tasks have a statistically better indicator when compared to the performance of their individual tasks, although the difference is not significant. As for the pairwise combinations of the tasks, in the case of hair mask segmentation, according to the Balanced Accuracy, both multitask

Table 1. Mean of the similarity measures obtained to compare our method with six state of the art hair removal algorithms.

	Our method	Abbas	Bibiloni	Huang	Lee	Toossi	Xie
MSE	**27.847**	258.347	103.903	404.366	175.303	221.346	55.311
SSIM	**0.926**	0.867	0.885	0.851	0.890	0.864	0.921
PSNR	**35.137**	26.080	29.785	27.001	29.158	26.570	33.096
RMSE	**4.790**	14.226	9.207	15.485	11.032	13.287	6.318
VIF	0.525	0.526	0.499	0.402	0.531	0.509	**0.592**
UQI	**0.997**	0.991	0.996	0.990	0.995	0.992	**0.997**
MSSSIM	**0.978**	0.870	0.934	0.917	0.945	0.875	0.955
PSNR-HVS-M	**36.802**	25.078	29.404	26.248	28.445	25.519	33.005
PSNR-HVS	**35.168**	24.628	28.681	25.738	27.826	25.065	32.186

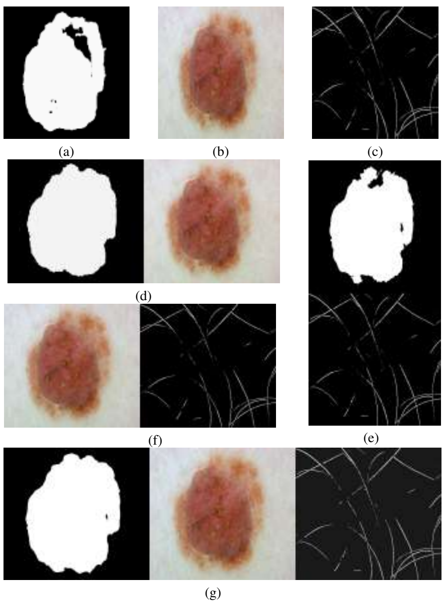

Figure 4. Results on Figure 2a for training a model for (a) lesion segmentation, (b) hair removal, (c) hair mask segmentation. Multitask models for (d) lesion segmentation and hair removal, (e) lesion and hair mask segmentation, (f) hair removal and hair mask segmentation; and a combination of all three tasks (g) lesion and hair mask segmentation, and hair removal.

Table 2. Relative gains and losses (%) over single-task (rows) performance measure when incorporating auxiliary tasks (cols).

	With Skin Lesion Seg.	With Inpainting	With Hair mask Seg.	With the rest of the tasks
Skin Lesion Seg. Dice Coefficient	-	2.77%	0.80%	1.80%
Inpainting SSIM	-0.97%	-	-0.53%	-1.29%
Hair mask Seg. Balanced Acc.	0.24%	-0.17%	-	0.98%

models obtain statistically comparable results to the baseline hair mask segmentation model. However, the multitask model that presents a better indicator is the one resulting from its combination with the lesion segmentation task. Similarly, according to the SSIM measure, the inpainting task in a multitasking framework is statistically inferior to its solo performance. Finally, the segmentation task bene-

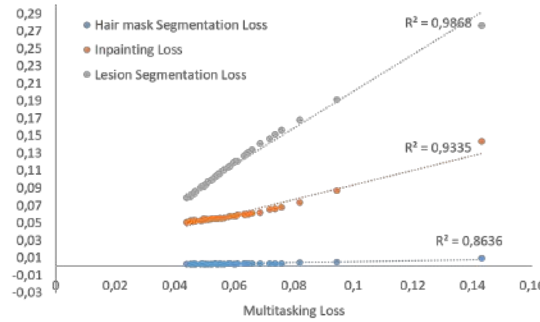

Figure 5. Correlation between each single-task model and the multitasking model for the three tasks.

fits from incorporating either of the other two tasks, as the performance measure of its multitask models statistically outperforms that of the lesion segmentation model alone.

To complete the study, we analyze how each task influences the validation loss function of the multitask model when considering all three tasks. In Figure 5 we can see that the loss function of each task has a high correlation with the averaged one, the model has been able to focus on learning all tasks equally. Even for the correlation with respect to the inpainting task, which has the lowest correlation index of 0.86, we can see that it is due to the influence of the presence of outliers, corresponding to the early epochs, where the training is more unstable. For the loss functions of the hair mask and lesion segmentation tasks, we obtain a correlation index of 0.94 and 0.99, respectively.

4. Conclusions

In this work, we have presented a novel CNN-based model for hair removal in dermoscopic images. The results obtained by analyzing the performance of our method and comparing it with six state-of-the-art approaches, allow us to conclude that our method is the best algorithm in eight of the nine performance measures used. This validates its potential in out-of-sample real hair images and the suitability of deep learning models for this task. Also, from the multitask experiments, we conclude that the lesion segmentation task benefits from the incorporation of either of the other two tasks, as the performance measure of the multitask models statistically outperforms the single segmentation task. In contrast, the inpainting task in a multitasking framework is statistically inferior to its solo performance. We also show that when combining the three tasks, the metrics for the hair mask and lesion segmentation tasks are statistically better than the ones their corresponding base model. As future work, we will integrate to this approach other criteria that make up the diagnostic protocols applied by physicians, such as the classification of the lesions' symmetry, diagnosis or presence of dermoscopic structures .

Acknowledgements

This paper is part of the R+D+i Project PID2020-113870GB-I00- "Desarrollo de herramientas de Soft Computing para la Ayuda al Diagnóstico Clínico y a la Gestión de Emergencias (HESOCODICE)", funded by MCIN/AEI/10.13039/501100011033/. Lidia Talavera-Martínez also benefited from the fellowship BES-2017-081264 conceded by the *Ministerio de Economía, Industria y Competitividad* under a program co-financed by the *European Social Fund*.

References

[1] Giuseppe Argenziano, HP Soyer, V De Giorgi, Domenico Piccolo, Paolo Carli, and Mario Delfino. *Interactive atlas of dermoscopy (Book and CD-ROM)*. EDRA Medical Publishing & New media, 2000. 2

[2] Mohamed Attia, Mohammed Hossny, Hailing Zhou, Saeid Nahavandi, Hamed Asadi, and Anousha Yazdabadi. Realistic hair simulator for skin lesion images: A novel benchemarking tool. *Artificial Intelligence in Medicine*, 108:101933, 2020. 2

[3] Ralph Peter Braun, Harold S Rabinovitz, Margaret Oliviero, Alfred W Kopf, and Jean-Hilaire Saurat. Dermoscopy of pigmented skin lesions. *Journal of the American Academy of Dermatology*, 52(1):109–121, 2005. 1

[4] Rich Caruana. Multitask learning. *Machine learning*, 28(1):41–75, 1997. 1

[5] M Emre Celebi, Hitoshi Iyatomi, Gerald Schaefer, and William V Stoecker. Lesion border detection in dermoscopy images. *Computerized medical imaging and graphics*, 33(2):148–153, 2009. 1

[6] Zhao Chen, Vijay Badrinarayanan, Chen-Yu Lee, and Andrew Rabinovich. Gradnorm: Gradient normalization for adaptive loss balancing in deep multitask networks. In *International Conference on Machine Learning*, pages 794–803. PMLR, 2018. 2

[7] Gery P Guy Jr, Cheryll C Thomas, Trevor Thompson, Meg Watson, Greta M Massetti, and Lisa C Richardson. Vital signs: melanoma incidence and mortality trends and projections—united states, 1982–2030. *MMWR. Morbidity and mortality weekly report*, 64(21):591, 2015. 1

[8] TF Mendonça, ME Celebi, T Mendonça, and JS Marques. PH^2: A public database for the analysis of dermoscopic images. *Dermoscopy Image Analysis*, 2015. 2

[9] Hengameh Mirzaalian, Tim K Lee, and Ghassan Hamarneh. Hair enhancement in dermoscopic images using dual-channel quaternion tubularness filters and MRF-based multi-label optimization. *IEEE Transactions on Image Processing*, 23(12):5486–5496, 2014. 2

[10] Sebastian Ruder. An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*, 2017. 1

[11] Lidia Talavera-Martínez, Pedro Bibiloni, and Manuel González-Hidalgo. Comparative study of dermoscopic hair removal methods. In *ECCOMAS thematic conference on computational vision and medical image processing*, pages 12–21. Springer, 2019. 1