A Systematic Investigation on Deep Architectures for Automatic Skin Lesions Classification

Pierluigi Carcagnì

Introduction

- ✓One of the hot topics in the healthcare and medical fields nowadays concerns the automatic classification of skin lesions.
- ✓Promising works lever Convolutional Neural Networks (CNN).
- ✓Existing pipelines mainly rely on complex data pre-processing.
- ✓There is no systematic investigation about how available deep models can actually reach the accuracy needed for real applications.

In this work, an end-to-end pipeline is introduced and some of the most recent Convolutional Neural Networks (CNNs) architectures are included in it and compared.

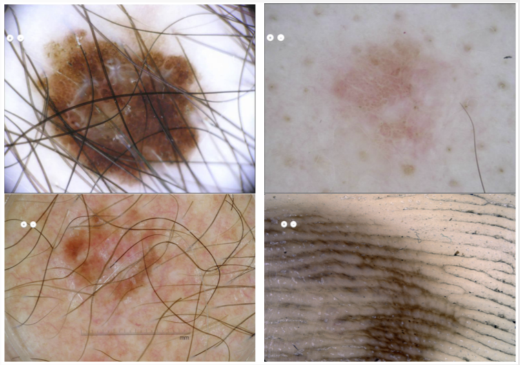
For the first time in this application context, a new network design paradigm, namely RegNet, has been exploited to get the best models among a population of configurations.

Method

An end-to-end pipeline has been employed in order to investigate and compare 5 families of CNN architectures: ResNet [1], DenseNet [2], SENet [3] and EfficientNet [4] and RegNet [5].

The employed approach performs classifications without the necessity of any preprocessing of the input images (such as for example an image segmentation step), but analyzing only color pixel values.

RegNet network design paradigm is for the first time exploited in this application context.



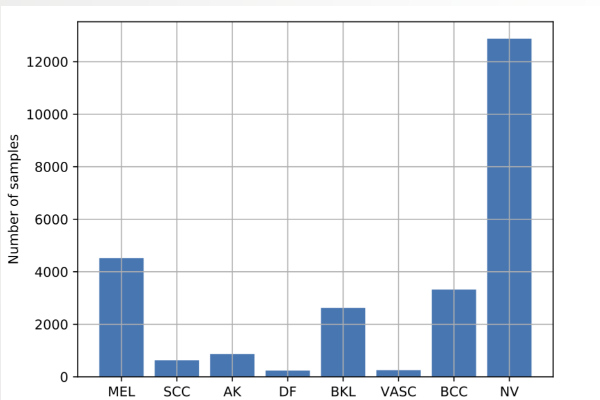
Some challenging situations to be solved while analyzing skin lesions

[1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
[2] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4700–4708.
[3] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7132–7141.
[4] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," arXiv preprint arXiv:1905.11946, 2019.
[5] I. Radosavovic, R. P. Kosaraju, R. Girshick, K. He, and P. Dolla ́r, "Designing network design spaces," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 10 428–10 436.

Dataset employed

The dataset employed is that provided by the **ISIC for the 2019 challenge [6-8]**, consisting of approximately 25331 images covering 8 types of skin lesions

1. **Actinic Keratoses** (AKIEC)
2. **Basal cell carcinoma** (BCC)
3. **Melanocytic nevi** (NV)
4. **Benign keratosis** (BKL)
5. **Vascular skin lesions** (VASC)
6. **Dermatofibroma** (DF)
7. **Melanoma** (MEL)
8. **Squamous cell carcinoma** (SSC)



ISIC 2019 dataset samples distribution

MEL	SCC	AK	DF	BKL	VASC	BCC	NV	Total
4522	628	867	239	2624	253	3323	12875	25331

Number of images for each class

[6] Tschandl P., Rosendahl C. & Kittler H. "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions." Sci. Data 5, 180161 doi.10.1038/sdata.2018.161 (2018)
[7] Noel C. F. Codella, David Gutman, M. Emre Celebi, Brian Helba, Michael A. Marchetti, Stephen W. Dusza, Aadi Kalloo, Konstantinos Liopyris, Nabin Mishra, Harald Kittler, Allan Halpern: "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)", 2017; arXiv:1710.05006.
[8] Marc Combalia, Noel C. F. Codella, Veronica Rotemberg, Brian Helba, Veronica Vilaplana, Ofer Reiter, Allan C. Halpern, Susana Puig, Josep Malvehy: "BCN20000: Dermoscopic Lesions in the Wild", 2019; arXiv:1908.02288.

Training and test

Fine-tuning of the CNNs starting, for all the considered architectures, from same Imagenet dataset [9].

K-fold strategy with k=5 employing the whole dataset.

Data augmentation by means of simple transformations such as rotations, flipping, scale, color contrast, color saturation, random cropping, etc.. [10].

Stocastic Gradient Descent optimizer.

Pytorch framework.

[9] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein et al., "Imagenet large scale visual recognition challenge," International journal of computer vision, vol. 115, no. 3, pp. 211–252, 2015.
[10] F. Perez, C. Vasconcelos, S. Avila, and E. Valle, "Data augmentation for skin lesion analysis," in OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Proce- dures, and Skin Image Analysis. Springer, 2018, pp. 303–311.

Experimental results

Results in terms of **F1-Score, Precision and Recall** of the trained CNNs.

All CNNs achieved f1-scores greater than 75%, and 7 of them exceeded 80%.

In particular, it is worth noting that the best f1-score was obtained by RegNetY16GF with 81.9% (in general RegNet paradigm performed very well with an average f1-score of 80.81%).

Also Densent201 (81.7%) and EfficientNet-B7 (81.4%) collected very oustanding f1-scores.

Model	F1-score	Precision	Recall
DenseNet121	75.0 ± 1.2	77.6 ± 0.9	72.8 ± 1.7
DenseNet169	78.6 ± 0.8	81.1 ± 0.5	76.6 ± 1.2
DenseNet201	81.7 ± 0.6	**83.6** ± **0.4**	80.1 ± 0.9
EfficientNet-B1	75.0 ± 1.3	76.5 ± 1.0	73.8 ± 1.6
EfficientNet-B2	75.9 ± 1.2	76.4 ± 1.0	75.6 ± 1.3
EfficientNet-B3	79.4 ± 0.8	79.1 ± 0.8	79.8 ± 1.0
EfficientNet-B4	77.7 ± 1.0	78.2 ± 0.9	77.4 ± 1.2
EfficientNet-B5	79.3 ± 0.8	80.0 ± 0.7	78.8 ± 0.9
EfficientNet-B6	79.5 ± 0.9	80.1 ± 0.9	79.0 ± 0.9
EfficientNet-B7	81.4 ± 0.7	81.4 ± 0.6	**81.7** ± **0.7**
RegNetY12GF	81.1 ± 0.7	83.2 ± 0.5	79.2 ± 1.1
RegNetY16GF	**81.9** ± **0.6**	83.4 ± 0.4	80.6 ± 0.8
RegNetY32GF	81.7 ± 0.6	83.5 ± 0.5	80.1 ± 0.8
RegNetY4.0GF	78.6 ± 0.8	80.7 ± 0.6	76.8 ± 1.2
RegNetY6.4GF	81.1 ± 0.6	82.7 ± 0.4	79.7 ± 0.9
RegNetY8.0GF	80.5 ± 0.8	82.0 ± 0.6	79.2 ± 1.0
ResNet101	78.4 ± 0.9	80.0 ± 0.6	77.2 ± 1.1
ResNet152	78.8 ± 0.9	81.0 ± 0.7	77.0 ± 1.1
ResNet50	77.7 ± 1.0	79.9 ± 0.7	75.8 ± 1.3
SE-ResNet101	78.9 ± 0.8	81.1 ± 0.6	77.2 ± 1.1
SE-ResNet50	78.7 ± 1.0	80.7 ± 0.7	77.1 ± 1.3

Experimental results

Best performing CNNs according to f1-score

Even if all the architectures seem adequate to tackle the classification task, RegNetY16GF, DenseNet201 and EfficientNet-B7 are undoubtedly the most suited ones.

MEL	DenseNet201, EfficientNet-B7 and RegNetY16GF
AK	RegNetY16GF
BBC	EfficientNet-B7
SCC	RegNetY16GF
DF	DenseNet201
NV	DenseNet201, EfficientNet-B7 and RegNetY16GF
BKL	RegNetY16GF
VASC	RegNetY16GF

Institute of Applied Sciences and Intelligent Systems

Science App

National Research Council of Italy

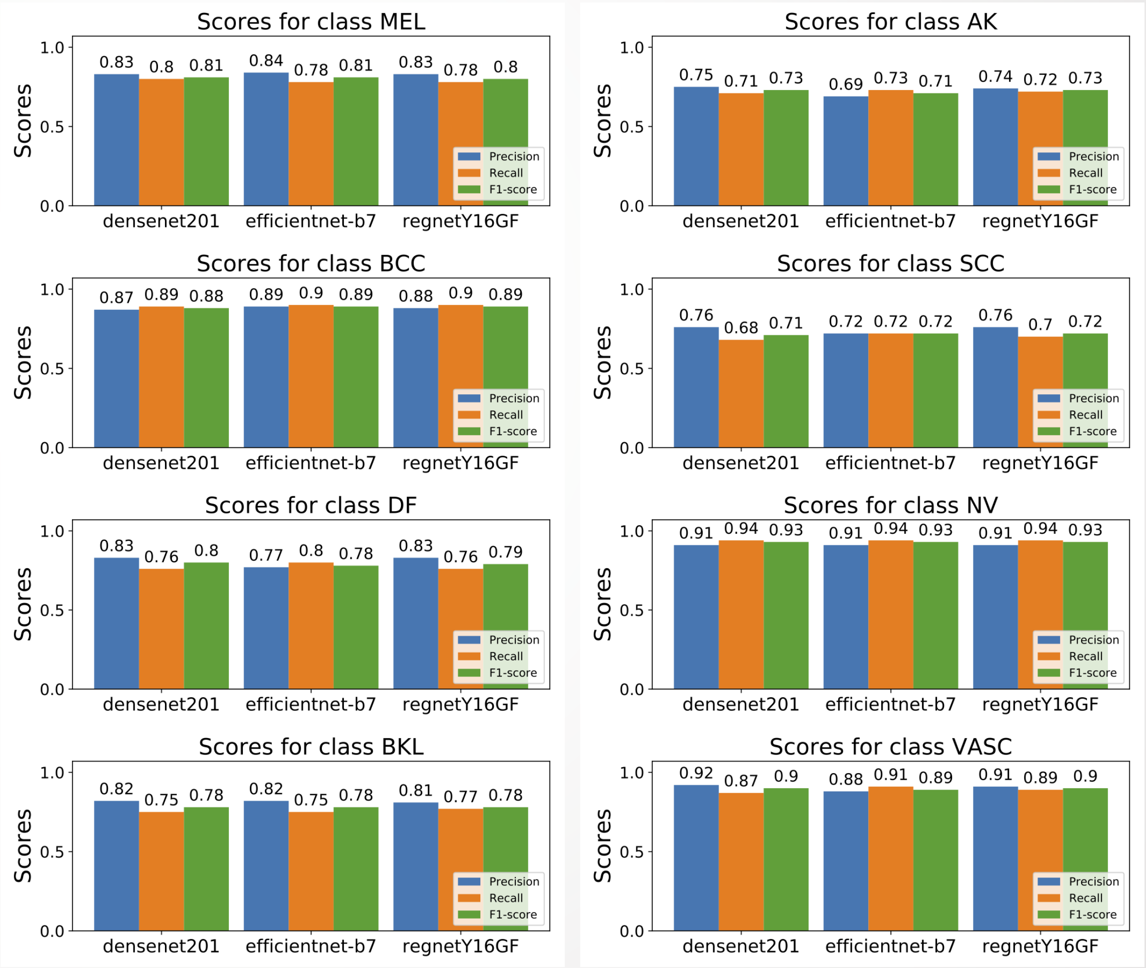
ICPR 2020

Experimental results

Densenet 201, EfficientNet-B7 and RegnetY16GF classes performance comparison.

On the one hand, it possible to conclude that RegNetY16GF is the model that performs at best for the majority of classes.

On the other hand, results make possible to pave the way towards the possibility to effectively introduce ensembling strategies.



Experimental results

Comparison of the best proposed models with leading approaches in the literature using the same benchmark dataset and the same k-fold strategy.

Model	F1-score	Precision	Recall
Proposed			
DenseNet201	81.7 ± 0.6	$\mathbf{83.6 \pm 0.4}$	80.1 ± 0.9
EfficientNet-B7	81.4 ± 0.7	81.4 ± 0.6	$\mathbf{81.7 \pm 0.7}$
RegNetY16GF	$\mathbf{81.9 \pm 0.6}$	83.4 ± 0.4	80.6 ± 0.8
previous			
Optimal Ensemble [11]			72.5 ± 1.7
Dynamic Augmentation [12]	64		65
Two-level ensembling [13]			59.1

[11] N. Gessert, M. Nielsen, M. Shaikh, R. Werner, and A. Schlaefer, “Skin lesion classification using ensembles of multi-resolution efficientnets with meta data,” MethodsX, p. 100864, 2020.
[12] T. A. Putra, S. I. Rufaida, and J.-S. Leu, “Enhanced skin condition prediction through machine learning using dynamic training and testing augmentation,” IEEE Access, vol. 8, pp. 40 536–40 546, 2020.
[13] S. Bagchi, A. Banerjee, and D. R. Bathula, “Learning a meta-ensemble technique for skin lesion classification and novel class detection,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 746–747.

Conclusion and future works

In this work recent machine learning models based on convolutional neural networks have been tested on the challenging ISIC 2019 skin lesions dataset namely ISIC 2019.

Gathered classification results by introduced models were very encouraging and largely outperformed previous approaches on the same dataset.

In future works, ensembling approaches built by training the base learners separately either in parallel or in a cascading manner will be contemplated.

Thank You