

딥러닝 기반 주름 평가

김진숙*, 김용남*, 김두홍*, 박래정**, 백지훈***, 강상구*

*메디사피엔스(주)

**강릉원주대학교

***더마프로

e-mail : jsk@msapiens.com

Rating wrinkled skin using deep learning

Jin-Sook Kim*, Yongnam Kim*, Duhong Kim*, Lae-Jeong Park**, Ji Hwoon Baek***, Sanggoo Kang*

*MedySapiens, Inc.

** Gangneung-Wonju National University

***DERMAPRO

요 약

The paper proposes a new deep network-based model that rates periorbital wrinkles in order to alleviate the shortcomings of the evaluation by human experts as well as to facilitate the automation. Periorbital wrinkles still need to be classified by human experts. Furthermore, the classification results from experts are different from each other in many cases due to the inter-interpretability variability and the absence of quantification criteria. Unlike existing classification methods which classify original images, the proposed model consists of a cascade of two deep networks: U-Net for the enhancement of wrinkles on an input image and VGG16 for final classification based on the wrinkle information. Experiments of the proposed model are made with a data set that consists of 433 images rated by experts, showing the promising performance.

1. 서론

외모에 대한 관심이 증가하면서 세정 및 미용을 위한 기능성 화장품의 시장이 점차 커지고 있다. 이에 따라 관련 기능성 제품의 효과를 객관적으로 평가하기 위한 방법과 절차에 대한 필요성이 대두되었고, 피부 중 주로 얼굴 피부의 노화를 측정하기 위한 여러 시도가 있어 왔다. 얼굴 피부 평가는 주로 객관적으로 평가가 가능한 주름 정보 (깊이, 길이 등)를 기반으로 이루어지며, 이를 위해 전문가, 학습된 모델이나 이미지 처리 기술들이 사용되었다.

관련 연구를 살펴보면, 우선 전문가에 의해 육안으로 분류되어는 경우로서 [1]에서는 미용 제품의 효능을 증명할 주름 평가 기준을 만들고 그 유효성을 검증하기 위해 9 명의 전문가가 주름 피부를 1 차 분류하였다. [2]에서는 분류 기준의 유효성 검증을 위해 5 명의 피부과 전문의가 198 명의 동남아시아 사람들의 주름 피부를 분류하였다. 학습된 모델을 사용하는 경우로 [3]에서는 보다 생동감 있는 아바타를 구성하기 위해 두 가지 학습 모델을 사용하여 주름을 포함한 얼굴의 미세한 특징까지 표현하는 고해상도 displacement map 을 만들어 냈다. [4]에서는 딥러닝 모델을 사용하여 dynamic texture 를 만들어 내며 이를 사용하여 주름을 포함한 얼굴의 미세한 특징들을 추론하였다. [5]에서는 피부 이미지를 normal, spot, wrinkles 로 분류하는 데 있어 딥러닝 모델을 적용하

였으며, GoogLeNet 이 Support Vector Machine 보다 좋은 성능을 보였다. [6]은 이미지 처리 기술을 사용하여 얼굴 주름을 추출하는 경우로서 Gabor 필터와 image morphology 가 사용하였으며, [7]에서는 잡음 제거를 위하여 intensity 값과 위치 정보 값을 함께 사용하는 Bilateral 필터를 사용하였다. 나이 추정을 위해 [8]에서는 국소 영역의 특징으로 Local Binary Patterns(LBP), Gabor wavelets(GW), Local Phase Quantization (LPQ)을 사용하여 주름을 추출하였다.

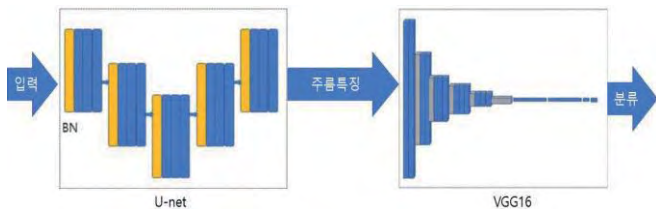
눈가 주름은 피부 노화의 지표로 주로 사용되며 이미지 처리 기술이 발전하였 어도 여전히 전문가의 평가 및 등급 분류가 필요한 대상이며 등급별로 평가 기준이 명시되어 있기는 하지만 정량화된 기준의 부재로 전문가마다 평가 결과가 다른 경우가 종종 있다.

본 논문은 딥러닝 모델을 사용함으로써 전문가의 육안 평가 위주의 눈가 주름의 등급 분류의 단점을 보완하고 그 평가 과정을 자동화하고자 한다. 이를 위해서, 원본 이미지를 분류하는 데 있어 학습 속도나 효율을 높이기 위해 원본 이미지를 그대로 분류 모델의 입력으로 사용하지 않고 U-Net[9]을 사용하여 원본 이미지의 주름을 추출하는 전처리 단계를 채택하였다. 추출된 주름은 VGG16[10]을 사용하여 최종 등급 분류하였으며 VGG16 학습 시, 각 등급의 학습 이미지 수가 등급별로 차이가 심한 경우에 그 차이를 학습에 반영하여 VGG16의 성능을 높였다.

2. 딥러닝 모델 기반 피부 주름 평가

딥러닝 모델에서 U-Net 은 convolutional neural network(CNN)의 일종으로 생체의학 이미지 분할을 위해 개발되었다. U-Net 은 문맥 정보와 위치 정보를 결합하여 이미지를 분할하며 이미지의 수가 적어도 적절히 학습할 수 있다[9]. 본 논문에서는, 원본 이미지를 분류하는 데 있어 원본 이미지를 그대로 분류 모델의 입력으로 사용하지 않고 우선 U-Net 을 사용하여 원본 이미지에서 주름을 추출한다. 눈가 주름은 주름의 깊이에 따라 얇은 주름, 보통 주름, 깊은 주름으로 분류되며 주름 별 주름의 수도 평가 기준이 된다. U-Net 의 목적은 잔털, 점, 기미 등 주름과 유사한 정보가 포함되지 않고 주름 정보에만 기반하여 피부 등급을 구분하기 위해서 등급 분류 전용 딥네트 전에 주름 정보만을 증대(enhancement)하는 것이다. 이로 인해 분류기 모델의 학습에 효과적인 특징만이 입력되므로 학습 속도나 효율을 높일 수 있었다. 원본 이미지에 있는 잔털이나 미세한 피부 결이 주름으로 추출되는 경우를 줄이기 위해 U-Net 의 각 층에 Batch normalization(BN) [11][12]을 추가하여 추출 성능을 개선하였다. BN 은 batch 의 분포를 평균 0, 표준편차 1로 만드는 과정으로 U-Net 의 학습 성능을 높인다.

등급을 분류하는 두 번째 단계는 딥러닝 모델 중 이미지 분류 성능이 좋을 뿐만 아니라, 분류에 사용된 가중치가 공개된 VGG16[10]을 채택하였다. VGG16 학습 시, 분류 그룹별 데이터의 불균형은 학습에 종종 부정적 영향을 주기 때문에 이런 차이를 학습에 반영하였다. (그림 1)이 논문에서 제시한, U-Net 과 VGG16 을 결합한 모델을 보여주고 있다.



(그림 1) 제안하는 주름 등급 평가 모델의 구조

3. 실험 결과

3.1 시스템 사양

<표 1>은 실험에 사용된 시스템 사양을 보여준다

<표 1> 시스템 사양

시스템	버전
Anaconda	4.5.2
Python	2.7.14
Keras	2.2.0
Numpy	1.14.3
Tensorflow-gpu	1.4.0

3.2 영상 데이터 및 전처리

실험에 사용한 이미지는 VISIA CR (Canfield, USA)를

이용하여 cross-polarized mode 로 촬영된 이미지로서 해당 전문가의 육안 평가를 거쳐 0~9 등급으로 분류된 총 413 개 이미지이다. (그림 2)는 3 등급(왼쪽)과 6 등급(오른쪽) 이미지를 보여주고 있다. 원본 영상에서 VGG16 분류기를 사용하여 크롭핑(cropping) 영역에 대한 사전 실험을 수행하였는데, 눈꼬리 영역을 포함한 경우가 그렇지 않은 경우에 비해 현저히 나쁜 성능을 보였다. 이는 눈꼬리의 주름은 전문가의 평가 대상이 아니기 때문이며, 이를 고려하지 않으면 모델의 유효한 특징 추출에 방해가 됨을 확인하였다. 이하 눈꼬리를 제외한 영역을 크롭핑하여 사용하였다.



(그림 2) 3 등급과 6 등급의 영상

3.3 분류 모델

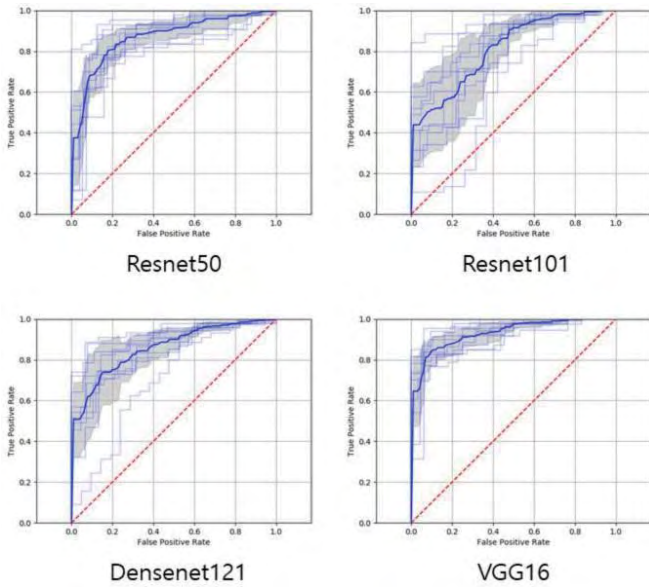
최적의 분류 모델을 선택하기 위해 Resnet50, Resnet101, Densenet121, VGG16 의 성능을 비교해 보았다. (그림 3)은 3 등급과 그 외 등급(4 등급 이상)으로 분류한 경우의 성능을 ROC(receiver operating characteristic) 그래프로 나타냈다(0~2 등급의 이미지는 개수가 충분하지 않아서 3 등급의 영상을 최상급 base 등급으로 간주하였다. 피부의 주름이 적을수록 등급이 낮다.). ROC 평면 상에서는, ROC 곡선이 왼쪽 상단(false positive rate 가 0 이고 true positive rate 가 1 인 경우)에 가까울수록 그 분류기의 성능이 우수하다. <표 2>는 ROC 곡선에 대한 AUC(Area Under the Curve) 값을 보여준다. AUC 값 또한 분류기의 성능을 보여주며 분류 성능이 좋을수록 AUC 값이 높다. (그림 3)과 <표 2>에서 보듯이 VGG16 이 가장 좋은 성능을 보여준다.

3.4 U-Net 을 활용한 특징 추출

(그림 4)는 U-Net 을 전처리 단계로 사용하여 눈가 주름만을 증대한 결과를 나타낸다. 왼쪽이 원본 이미지이며 가운데가 U-Net 을 사용하여 주름을 추출한 이미지이다.

3.4.1 Batch normalization

(그림 4)에서 왼쪽과 가운데 이미지를 자세히 보면 원본 이미지에 있는 털이 주름으로 추출된 것을 볼 수 있다. 털과 미세한 피부 결이 주름으로 추출되어 분류 성능에 영향을 미침으로 이를 개선하기 위해 U-Net 의 각 계층에 Batch normalization 을 추가하였다. (그림 4)의 오른쪽 이미지를 보면 원본 이미지에 있는 털과 미세한 피부 결이 주름으로 추출되는 경우가 줄어들었음을 볼 수 있다.



(그림 3) 모델 성능 비교

<표 2> 모델 별 AUC 값

모델	AUC 값
Resnet50	0.876456
Resnet101	0.838636
Densenet121	0.873464
VGG16	0.932322



(그림 4) 특징 추출과 Batch normalization

3.4.2 Class weight

VGG16 학습 시, 분류 그룹별 데이터의 불균형은 학습에 종종 부정적 영향을 준다. 학습 데이터가 수가 적은 그룹은 학습 시 학습 데이터가 많은 그룹에 비해 학습되어지는 빈도가 적어 분류 시 분류 성능이 떨어지게 된다. 예를 들면 3등급과 그 외 등급으로 분류할 경우 학습 데이터의 수가 약 1:2로 차이가 난다. 3등급 학습 데이터에 그 외 등급보다 더 높게 가중치를 주지 않고 VGG16을 학습시키면 3등급 분류 성능이 그 외 등급에 비해 떨어지게 된다. 학습 데이터 수의 차이를 VGG16의 학습에 반영하지 않은 경우와 반영한 경우를 비교해 보았을 때 반영한 경우가 더 좋은 성능을 보였다.

4. 결론

본 논문은 눈가 주름에 대해 육안 평가의 단점을 보완하기 위해서 2 단계의 딥러닝 모델을 제시하였다.

첫 번째 단계로 주름 정보만을 증대(enhancement)하기 위해 U-Net 을 사용하여 원본 이미지에서 주름 성분을 추출한다. 이 과정에서 Batch normalization 은 잡음을 줄여 주름 추출 성능을 높였다. 두 번째의 등급 분류 단계에서는 딥러닝 모델 중 이미지 분류 성능이 좋을 뿐만 아니라, 분류에 사용된 가중치가 공개된 VGG16 을 사용하였다. 분류 그룹별 데이터의 불균형에 의해 발생하는 학습 과정에서의 부정적 영향을 최소화하기 위해 클래스 간 가중치를 조절 반영하였다. 제시한 모델의 성능 개선을 위해 U-Net 의 주름 추출 성능을 개선하고, 추출한 주름 정보로부터 분류에 보다 효과적인 특징 추출을 위한 연구를 진행 중에 있다.

참고문헌

- [1] G. Lemperele, R. E. Holmes, S. R. Cohen, and S. M. Lemperele, "A classification of facial wrinkles", *Plast Reconstr Surg*, 108, 1735-50, 2001.
- [2] B. Tian, "Wrinkle Severity Grading Scale:A Southeast Asia Study", *Open Access Library Journal*, 5, 2018.
- [3] L. Huynh, W. Chen, S. Saito, J. Xing, K. Nagano, A. Jones, P. Debevec, and H. Li, "Mesoscopic Facial Geometry Inference Using Deep Neural Networks", *IEEE Int. Conf. on CVPR*, 2016.
- [4] K. Olszewski, Z. Li, C. Yang, Y. Zhou, R. Yu, Z. Huangk, S. Xiang, S. Saito, P. Kohli, and H. Li, "Realistic Dynamic Facial Textures from a Single Image Using GANs", *IEEE Int. Conf. on Computer Vision*, 2017.
- [5] J. S. Alarifi, M. Goyal, A. K. Davison, D. Dancey, R. Khan, and M. H. Yap, "Facial Skin Classification Using Convolutional Neural Network", *Int. Conf. on Image Analysis and Recognition*, 2014.
- [6] N. Batool and R. Chellappa, "Fast detection of facial wrinkles based on Gabor Features using Image morphology and geometric constraints", *Pattern Recognition*, vol. 48, pp. 642-658, 2015.
- [7] 조민우, "비접촉식 피부 진단 장비를 위한 디지털 영상처리 기법에 관한 연구", 서울대학교 대학원 학위 논문, 2014.
- [8] J. K. Pontes, A. S. Britto Jr., C. Fookes, and A. L. Koerich, "A flexible hierarchical approach for facial age estimation based on multiple features", *Pattern Recognition*, vol. 54, 34-51, 2016.
- [9] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation", *MICCAI*, 2015.
- [10] K. Simonyan and A. Zisserman, "Very deep convolutional networks for image-scale image recognition", *ICLR*, 2015.
- [11] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal

covariate shift”, ICML, pp. 448–456, 2015.

- [12] S. Santurkar, D. Tsipras, A. Ilyas, and A. Madry. “How does batch normalization help optimization? (no, it is not about internal covariate shift)”, arXiv preprint, arXiv:1805.11604, 2018.