Please cite this article as: Čedomir Vasić. Facial wrinkle categorization using Convolutional Neural Network. Dermatol Rep 2024 [Epub Ahead of Print] doi: 10.4081/dr.2024.10034

Facial wrinkle categorization using Convolutional Neural Network

Čedomir Vasić

Information-Technology-FB&L, University MB, Belgrade, Serbia

Correspondence: Čedomir Vasić, Information-Technology-FB&L, University MB, Teodora Drajzera 27, Belgrade, Serbia.

E-mail: cedav1961@gmail.com

Key words: classification, categorization, skin, wrinkle, CNN.

Abstract

A tool for the detection and classification of wrinkles on the facial skin is always welcomed in the pursuit of tight and beautiful skin. If this tool involves the use of a state-of-the-art neural network and high-quality images, there is a high likelihood that it will be practical. Five thousand and ninety-eight images were categorized into four classes by a trained expert and prepared for neural network training. The task was to determine whether such prepared data could serve as good material for learning and whether they could provide sufficiently high accuracy in prediction. It was found that the answer to this question is positive.

Introduction

The analysis of facial skin images is conducted in various ways and at different locations, ranging from medical studies for therapeutic purposes to cosmetic endeavors dedicated to achieving clear and beautiful skin. Momentum for new research is sometimes driven by the healthcare system and at other times by the cosmetics industry. Wrinkles result from structural changes in the skin and subcutaneous tissue. An effective method for identifying and quantifying the condition of facial skin wrinkles is always welcome. This paper continues the exploration of the potential use of a modern convolutional neural network as an efficient tool for facial skin categorization. The uniqueness of our approach lies in a distinctive dataset created by combining two essential factors: the use of specialized equipment and the expertise of professionals in the field. In our previous work (Vasić, 2023), we examined facial skin categorization in terms of pore conditions, while this paper focuses on categorization based on wrinkles. We opted for categorization based on expert assessments, ranging from very mild to severe wrinkling, introducing a total of 4 descriptive categories. This minimal number of categories is essential for monitoring skin conditions and evaluating the effectiveness of recommended treatments. It is clear that this categorization is specifically tailored for cosmetic practice purposes.

The training data were categorized based on the expert's free assessment rather than image processing with specialized software tools. This setup presented a challenge for the convolutional neural network, as it had the opportunity to learn from both the knowledge and experience of the expert. During decision-making, the expert considers various factors observed during examinations and learned through patient conversations, such as whether the patient is a smoker, an alcoholic, their age, etc. Expert analysis is an advantage we have over conventional datasets, but simultaneously, this advantage may be a weakness in our data. The human factor inherently includes the potential for error. Subjective assessment is one thing, and precise data are another. It is evident that the hidden flaw in our training dataset may be its inconsistency, i.e., the real possibility that an individual expert's assessment may not align with the information about wrinkle conditions contained in the image itself. The fundamental aim of this paper is to address this dilemma. The question we attempt to answer is: despite potential drawbacks, can the

dataset we use serve as a foundation for training a neural network, and can such a learned network be practical in real-world applications?

The relatively small amount of training data poses another challenge, significantly influencing the choice of neural network architecture. The network must not be too large or too deep, as too many parameters with a small amount of training data is a shortcut to the network memorizing data rather than learning from image features.

Previous works

It is an undeniable fact that initiating a discussion on the classification of facial wrinkles poses an immediate challenge. There is no universally accepted classification, not even a consistent terminology regarding the shape, dimensions, position, and nature of wrinkles. Terms such as wrinkles, creases, furrows, grooves, lines, etc., are in use. An example of clinical categorization using ratings from 1 to 5 is provided in Doris J. Day's work [2], where wrinkles are classified into the following classes: 1: none, 2: mild, 3: shallow, 4: shallow but noticeable, and 5: dense network of wrinkles.

Focusing on the microtopography of facial skin and observing furrows reveals that they become sparser and deeper due to intrinsic aging and sun exposure. The microrelief of facial skin dynamically changes with aging, influenced by dominant aging factors such as the passage of time, genetics, ultraviolet and infrared radiation, diet, smoking, diseases, hormonal imbalances, gravitational force, etc. Each of these factors allows for an independent classification, as comprehensively presented in Quatresooza's work [3]. An evaluation of facial skin texture and the gradation of microrelief into one of 4 categories is given in Setaro's work [4]. Here, the assessment is based on primary and secondary lines and their points of intersection: i) clearly visible and equal depth, ii) flattened, iii) primary lines exist while secondary ones do not, and iv) formations of lines and intersection points do not exist.

The challenges of measuring and evaluating wrinkles are analyzed in Gormlev's work [5]. He identified changes in the dynamic distortion, alteration, and misinterpretation of wrinkles.

When classifying facial skin based on wrinkle conditions across the entire face surface, a significant challenge at the outset is face segmentation and the isolation of areas of interest. Illustrative works in this

field include those by Osman [6], Yapa [7], and Jiang [8].

In our study, the area of interest is focused on the part of the skin covering the upper cheek, specifically analyzing periorbital wrinkles. This region was analyzed in Cula's work [9].

In this paper, we opted for a scale assessing the prevalence of wrinkles comprising 4 categories: i) Mild wrinkling, no wrinkles or very mild ones, smooth skin line and very shallow creases; ii) moderate wrinkling, slightly deeper creases; iii) significant wrinkling, coarse irregularities in skin microrelief, pronounced furrows; iv) severe wrinkling, deep furrows.

EfficientNet model, upon which our neural network is based, is elucidated in the work by M. Tan [10

Materials and Methods

Images required for training the neural network were obtained from a database formed during examinations of users subsequently recommended specific cosmetic products. The examination of facial skin and categorization were conducted by a trained expert. The process of preparing the input data set is illustrated in Figure 1.

During patient skin examinations, a mobile device from the MSS (Mobile Smart Scope) device family, model API-100, manufactured by Aram Huvis-Korea, was used. Although this device has several different cameras and generates various types of images, only images intended for wrinkle analysis were utilized. Following the manufacturer's instructions, the same part of the face, visibly marked in Figure 1, was consistently captured. During the assessment of the image's category, the expert also relied on the results of image processing performed using the accompanying "Skin-Solutionist" software (V1.2.49). The drawback of the material obtained in this way is the relatively small skin area analyzed. In the case of the API-100 device, it is approximately 1 cm². However, considering that the analysis is based on capturing the same part of the face, it provides an acceptable basis for comparison and categorization.

For the purpose of training the neural network, a total of 5,098 images were provided, collected over a period of 3 years. These images were categorized into 4 classes, with the number of images in each category as evenly balanced as possible, with minimal deviation from the ideal distribution, as also illustrated in Figure 1.

Neural Network Architecture and input data preparation

In order to facilitate the learning process, the images underwent preprocessing. Initially, the RGB images were transformed into the HSV color space, and then a Sobel filter was applied to the V channel. This idea has been extensively discussed in the work of Bora [11], demonstrating the Sobel filter as a useful tool for contour detection and enhancement.

The Sobel filter is a discrete differential operator that calculates the gradient of the image intensity in the x and y directions, and by combining these gradients, contours are formed. The V channel transformed by the Sobel filter is combined with the G and B channels of the original RGB image. After gamma correction and adjustment of gray level, we obtain an image ready for analysis by the neural network. The entire preprocessing procedure is illustrated in Figure 2. The resolution of the training images was adjusted to the model, reduced from the original 640x480 to 224x224 pixels.

The architecture of the convolutional neural network constructed for this work was influenced by the relatively small set of input data. An advantageous factor was the relatively small number of output categories, allowing for the construction of a functional and efficient model with a few adaptations. The "Keras" library from the TensorFlow2 development environment (v2.13.0) was used for building the neural network, with Python3 (v3.10.12) called from a JupyterLab notebook.

The neural network itself was designed based on the EfficientNet_V2B0 function, whose efficiency is well-explained in M. Tan's work [10]. The "Adam" optimization, learning rate of 0.001, and "sparse-categorical-crossentropy" were employed. The input data were split into training and validation sets in an 80%:20% ratio. A "batch" consisted of 58 images. Due to the need to reduce overfitting, data augmentation was introduced, implemented through functions such as horizontal and vertical image flipping, rotation, and zooming. The architecture of the neural network is depicted in Figure 3.

The network training was conducted on a modest computer with a Core i3-10105/16GB-RAM processor and an Nvidia RTX3060/12GB-RAM graphics card. The hardware ran the Ubuntu 22.04 operating system, enabling computations to be performed on the GPU processor.

Results

Considering the circumstances emphasized in the introductory part of this paper, we can be satisfied with the achieved results. The answer to the fundamental question about the usability of input data posed at the beginning of the study is affirmative. Modern neural network models allow us to train a convolutional neural network with a relatively modest amount of training data and potentially inconsistent data, which can reasonably categorize the presented images.

Figure 4 presents a graph comparing the prediction accuracy of training and validation data. It is evident that the small volume of input data leads to slight overfitting, but it is also clear that the network is consistently learning, sometimes faster and sometimes slower, with constant progress.

Figure 4 presents a graph comparing the prediction accuracy on training and validation data. It is evident that the small volume of input data leads to a slight overfitting, but it is also clear that the network is consistently learning, sometimes faster and sometimes slower, with constant progress.

The oscillations on the validation data graph are consequences of the limited volume and potential errors in the categorization of input data. Various optimization formulas were tested, and the learning rate and batch size were adjusted, but no significant improvement in percentages compared to the shown graph was achieved.

Table 1 displays a table with the best results obtained in the 228-th out of a total of 230 epochs, i.e., at the end of the training process. This resulted in an accuracy of 93.57% on the training data and an accuracy of 81.05% on the validation data.

Finally, model testing was conducted using 300 images not used during the training process. The classification accuracy achieved on the test images was 91.50%.

Discussion and Conclusions

The primary conclusion drawn is that the training data available to us can serve as usable material for training convolutional neural networks. The usability of a network trained in this manner is significant, but only for non-medical purposes. For patient treatment in healthcare institutions, higher accuracy of

categorization is required, achieved through differently structured input data. However, for the recommendation and monitoring of the effectiveness of cosmetic products dedicated to skin hygiene and beauty, the proposed neural networks are entirely usable and can be a valuable tool in everyday practice.

References

[1] Vasic C. Skin pore detection and classification using convolutional neural network. Australasian Journal of Dermatology 2024;65:178-81.

[2] Day DJ, Littler CM, Swift RW, Gottlieb S. The Wrinkle Severity Rating Scale. Am J Clin Dermatol 2004;5:49-52.

[3] Quatresooz P, Thirion L, Pierard-Franchimont C, Pierard GE. The riddle of genuine skin microrelief and wrinkles. International Journal of Cosmetic Science 2006;28:389-95.

[4] Setaro M, Sparavigna A. Irregularity skin index (ISI): a tool to evaluate skin surface texture. Skin Research and Technology 2001;7:159-63.

[5] Gormlev DE, Workman MS. Objective Evaluation of Methods Used to Treat Cutaneous Wrinkles. Clinics in Dermatology 1988;6:15-23.

[6] Osman OF, Elbashir RMI, Abbass IE, et al. Automated Assessment of Facial Wrinkling: a case study on the effect of smoking. 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC) Banff Center, Banff, Canada, October 5-8, 2017.

[7] Yap MH, Batool N, Ng C-C, et al. A Survey on Facial Wrinkles Detection and Inpainting: Datasets, Methods, and Challenges. IEEE Transactions on merging topics in computational intelligence, June.30.2021, doi: 10.1109/TETCI.2021.3075723

[8] Jiang R, Kezele I, Levinshtein A, et al. A new procedure, free from human assessment that automatically grades some facial skin structural signs. Comparison with assessments by experts, using referential atlases of skin ageing. International Journal of Cosmetic Science 2019;41:67-78.

[9] Cula GO, Bargo PR, Kollias N. Assessing Facial Wrinkles: Automatic Detection and Quantification. Photonic Therapeutics and Diagnostics, Proc. of SPIE Vol. 7161 71610J-1, doi: 10.1117/12.811608

[10] Tan M, Le QV. EfficientNetV2: Smaller Models and Faster Training. Proceedings of the 38th International Conference on Machine Learning, PMLR 139, 2021.

[11] Bora DJ. A Novel Approach for Color Image Edge Detection Using Multidirectional Sobel Filter on HSV Color Space. International Journal of Computer Sciences and Engineering 2017;5:2347-693.
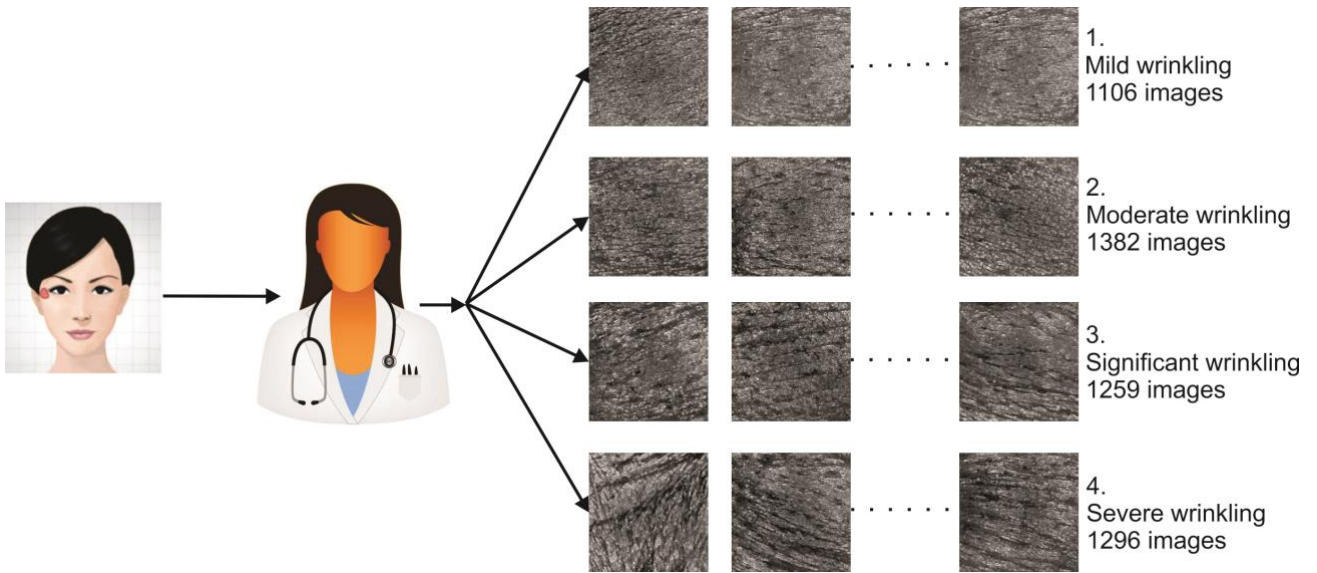
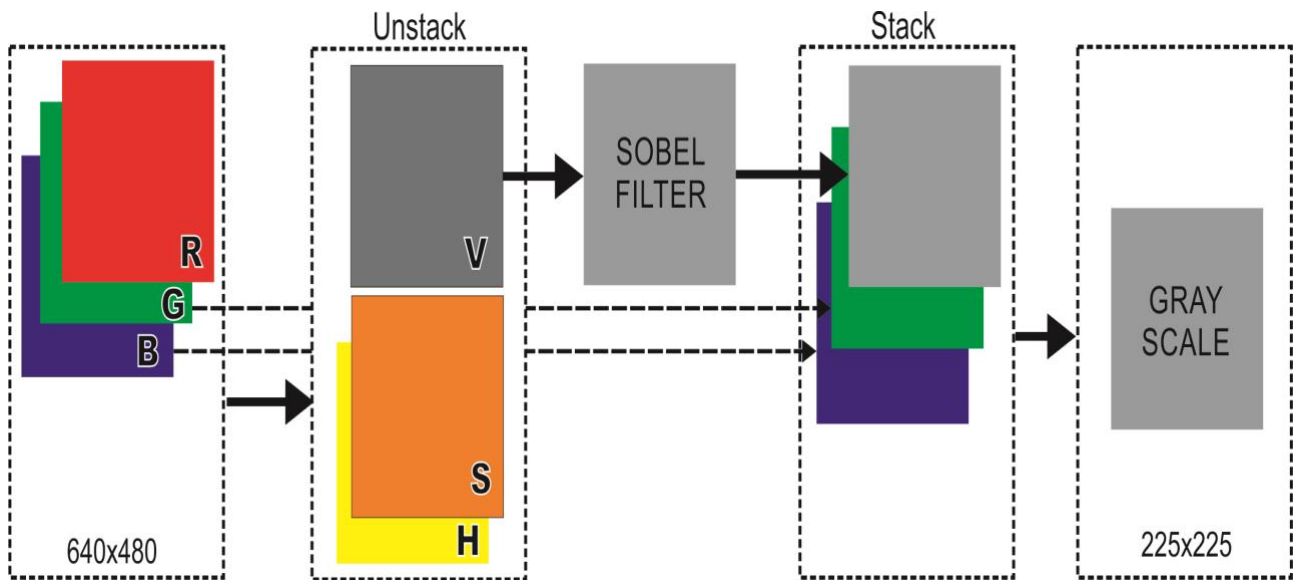**Figure 1.** Data preparation and formation of the dataset.

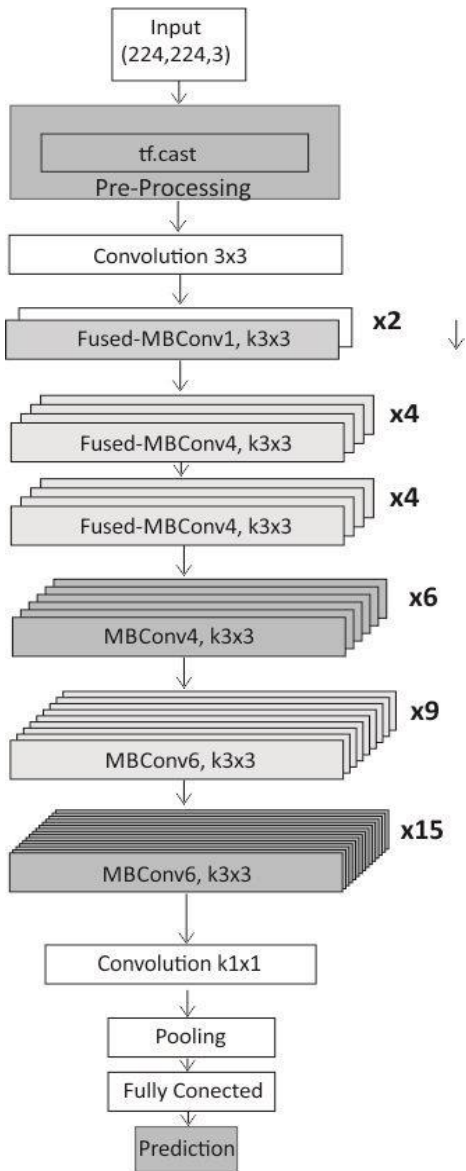**Figure 2.** Stages in the preprocessing of input images.
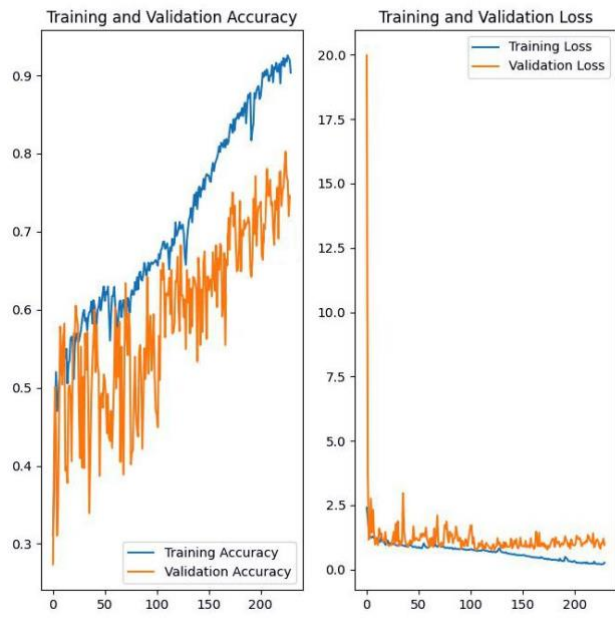
Figure 3. EfficientNet model architecture.

Figure 4. Training and validation - accuracy and loss.

Table 1. Results of the learning and testing process.

Parameter	Value
Max. training accuracy / 230 epochs	93,57%)
Max. validation accuracy / 230 epochs	81,05%
Prediction accuracy / 300 test images	91,05%