Computer-based Approach to Detect Wrinkles and Suggest Facial Fillers

Amal Alrabiah[1], Dr. Mai Alduailij[2]

College of Computer and Information Sciences
Princess Nourah bint Abdulrahman University
Riyadh, Saudi Arabia

Dr. Martin Crane[3]

School of Computing
Dublin City University
Dublin, Ireland

Abstract—Modern medical practice has embraced facial filler injections as part of the innumerable cosmetic procedures that characterize the current age of medicine. This study proposed a novel methodological framework. The Inception model is the core of the framework. By carefully detecting the classification of wrinkles, the model can be built for different applications to aid in the detection of wrinkles that can objectively help in deciding if the forehead area needs to have filler injections. The model achieved an accuracy of 85.3%. To build the Inception model, a database has been prepared containing face forehead images, including both wrinkled and non-wrinkled face foreheads. The face image pre-processing is the first step of the proposed framework, which is important for reliable feature extraction. First, in order to detect the face and facial landmarks in the image, a Multi-task Cascaded Convolutional Networks model has been used. Before feeding the images into the deep learning Inception model for classifying whether the face foreheads have wrinkles or no wrinkles, an image cropping process is required. Given the bounding box and the facial landmarks, face foreheads can be cropped accurately. The last step of the proposed methodology is to retrain an Inception model for the new categories (Wrinkles, No Wrinkles) to predict whether a face forehead has wrinkles or not.

Keywords—*Deep learning; classification; facial fillers; wrinkle detection*

I. INTRODUCTION

Filler injections are a modern cosmetic procedure and have been widely embraced by women and men alike because of their wonderful ability to create fuller cheeks, lips, and other facial features. Filler injections are also used to reduce the effects of wrinkles around the mouth, eyes, and eyebrows and to hide any scars that may be causing an individual to feel self-conscious and unattractive.

Human beings can without much difficulty see what the image represents. As an example, humans can easily see that the image contains a number of objects and can detect faces in an image as well as distinguish between the different features of the face. Computer Systems, on the other hand, can have difficulties. Computers cannot easily see whether the image contains objects or not. Also, cannot easily detect human faces and facial features.

As many researchers are trying to set up computers with artificial intelligence capabilities to be able to serve patients with different health problems such as diabetes, blood pressure, and obesity [1]. Artificial intelligence databases can help in easily identifying whether there are similar cases that have been registered before, and then returns the diagnosis and treatment of these similar cases [1]. There is still a dearth in research studies regarding how computer vision can be deployed in facial filler injections.

This study presents a novel methodological framework. The Inception model is the core of the framework. With carefully detecting the classification of wrinkles, the model can be built for different application to aid the detection of wrinkles to that can objectively help in determining if the forehead area needs to have filler injections.

II. RELATED WORK

Previous studies have also proposed a new algorithm referred to as Hessian Line Tracking (HLT) for detecting wrinkles [2]. The researchers began with a group of seeds that the researchers extracted from the Hessian Matrix's ridge area before proceeding to apply a multi-scale tracking system recursively to all the seeds. After completion, the researchers validated each pixel confidence over the scales with the objective of producing an initial map of wrinkles. The last step involved post-processing in which the researchers carried out a series of mini-steps including median and directional filtering as well as area thresholding in order to reduce noise [2]. In the experimental set-up, the researchers employed the services of three number of coders instructed to manually annotate the wrinkle on 100 cropped images of the forehead– the images were extracted from the Bosphorus dataset [3]. The dataset contains forehead wrinkles of varying sizes. Particularly, the dataset that was employed in the study contained 106 subjects from whom the researchers used an ordinary camera to fetch 2-D facial images under good flat, illuminated conditions. Although the researchers took several images of varying poses and facial expressions of each subject, the experiment only used frontal images. The researchers reported both intra and inter-reliability with regard to the manual annotation process – reliability was 94% and above [2]. Wrinkles typically appear in a wide variation in both images, in pattern, length, and width as well as within the same image. This significantly challenges the generation of an automatic wrinkle detection operator. This justified the development of a multiscale HLT – an approach based on seed extraction by Hybrid Hessian Filter (HHF) as well as multiscale tracking for overcoming the weakness of HHF while also making it possible to capture wrinkle variability in the entire image.

In another study researchers also proposed a new algorithm for automatic tracking of linear, fixed and chaotic forms of transient wrinkles [4]. For the automatic analysis of wrinkles, the researchers came up with two clusters of wrinkles including transient and permanent. While the latter are usually found on the faces of older individuals, the former on the other hand often appear in relatively wider regions in the course of generating an expression. The research was divided into two parts whereby while the first part explored an algorithm for detecting transient wrinkles, the second part dwelled on its application. The proposed wrinkle detector was made up of three steps. The first step involved the Canny edge detector that the authors applied to the input face for detecting pairs of continuous wrinkles. This was followed by applying an active appearance model to locate all candidate wrinkle lines. This generated data for constructing the structure of the wrinkle. In the third step, the researchers defined quantitative metrics which they subsequently used for Support Vector Machine (SVM) classification – this step was critical in helping to discriminate regions of the face with wrinkles from those without. Despite the fact that competitive results are achieved by the proposed transient wrinkle detector model that the researchers propose which is also the case with improved wrinkle mapping, there is a number of areas in which future studies should focus. A case in point regards long wrinkles in the forehead, for which [4] employed the five points in the wrinkle structure–future studies should add on more points which will go a long way towards improving the approximation of accuracy of the wrinkle edge.

Batool and Chellapa [5], in their work presented a quick deterministic algorithm based on image morphology as well as Gabor filters with the aim of improving localization results. The researchers proposed features derived from Gabor filter bank–the aim here was to shed light on the subtle curvilinear discontinuities in the texture of the skin attributed to wrinkles. The researchers then employed image morphology for integrating geometric constraints to localize curvilinear wrinkle shapes at the locations of wrinkles of pronounced Gabor filter responses. Experiments were carried out at two sets of images including those with high and low resolutions before the researchers compared the results to those generated from Marked Point Processes (MPP). Experiments illustrate that the suggested algorithm is not only faster compared to the MPP framework, but also generates the merit of visually satisfactory results.

In the last study analysed in the review, researchers compounded texture orientation fields with Gabor filter responses to detect wrinkles [6]. In the experiment, a bimodal Gaussian Mixture Model (GMM) described the distribution of normal skin verse skin imperfection of Gabor features [6]. The researchers then proceeded to employ a Markov random field model to integrate the spatial relationships for their texture orientations between adjacent pixels as well as GMM distribution. To classify skin versus skin imperfections, the study employed an expectation-maximization algorithm. As opposed to blending or blurring the detected wrinkles, the study removed them completely. The exemplar-based constrained texture synthesis algorithm, as the researchers

conclude is the most ideal tool for in-painting irregularly shaped gaps that are left behind in the form of scars of removed wrinkles [6]. Overall, the experiment illustrates that most skin imperfections and wrinkles are usually detected and in-painted. Nevertheless, there are a few areas on the face that have less contrast to the skin around it that fails to be detected. A case in point includes small parts of the upper parts of an individual's forehead. Moreover, the experiment also reported challenges of the impact of sagging skin and aging, illumination, as well as artefacts resulting from the repetition of patches. Particularly, facial images of subjects with sagging skin that appears alongside wrinkles according to the researchers posed significant challenges. Such is the case since the same patch of wrinkled skin is selected as a source of skin texture resulting in repeated patterns of imperfection.

III. BACKGROUND AND METHODOLOGY

A. Deep Learning

Deep learning is one of the major theories of machine learning that is founded on learning data representation and not task-related algorithms. It was first introduced in 1986 to the community of machine learning by Rina Dechter [7]. Learning using this theory might be unsupervised, supervised or a hybrid of both. As a branch of machine learning, deep learning is inspired by the human brain's main function and structure. The brain constitutes of neural networks whose interconnected neurons plays a crucial role in processing and transmitting signals from one neuron to the other. On the basis of this operation, the founder of deep learning, Geoffrey Hinton [8], made some artificial neural networks comprising of man-made neurons that could easily conduct operation as well as process the required information. The three layers of the neural network in deep learning include the input, hidden and output layers. The input layer is responsible for accepting a variety of input using formats such as audio, picture, number or text. The hidden layer conducts mathematical functions, feature extraction and manipulation of data. The output layer, on the other hand, is essential in getting the desired final output [8].

B. Transfer Learning

In 1993, Lorien Pratt came up with algorithm that was founded on discriminability-based transfer [9]. In so doing, he gave a platform through which the transfer learning theory was born. In this method, a model that has been formulated for an activity is normally reutilized as a beginning point for some other second task. For instance, the knowledge that one gets when learning to distinguish a jet could be useful in recognizing a helicopter or a spacecraft. Therefore, this model concentrates on the storage of acquired knowledge while at the same time striving to be utilized in another similar scenario or problem [10]. This theory is used in deep learning particularly in instances where the pre-trained models are employed as beginning points on the natural language and computer vision. In comparisons to the conventional machine learning, transfer learning utilizes these pre-trained models that were useful in another instance to kick start the process of development of the new problem or task. As stated earlier, this model depends on the task and domain concepts [10].

C. Multi-Task Cascaded Convolutional Neural Network

Multi-Task Cascaded Convolutional Neural Network model or simply MTCCNN is a deep learning algorithm-based model that comprises of three stages that identify faces' and bounding box in a particular image as well as the main five-point landmarks on an individual's face [11]. Back in 2001, researchers attempted to put forward a method of forward-cascade detection founded on the features of AdaBoost and Haar to conduct cascade classifiers [12]. Kaipeng Zhang later introduced the MTCNN theory which in addition to having three phases, helped in the detection of the bounding box, points, and landmarks [11].

For every MTCNN phase that the image goes through, the investigator sees an improvement in quality. The input goes to the CNN which not only give it a particular score but also return a bounding box. The initial stage ensures that the input is scaled downwards. The CNN also ensures the facilitation of the MTCNN in making a pyramid of a picture in question. The next stages involve the extraction of the patches of this picture for every bounding box. It is then resized and resized even more in the third stage. Other than bounding the image onto the box and later assigning some score, this stage also computes the points of the five face landmarks for every bounding box [11].

D. Inception-v3 Model

Inception V3 is one of the most popular image-recognition deep learning models. It has been a culmination of multiple ideas by a pool of researchers in the years. Inception V3 was theorized by Szegedy in a paper that rethought the Inception concept in computer vision [13]. In itself, this model is composed of both asymmetric and symmetric building blocks. They include convolutions, max, and average pooling as well as fully linked layers. Throughout the Inception –v3 model, Batchnorm is extensively utilized in addition to being applied to the activation inputs. On the other hand, Softmax is usually used in computing the loss. The two parts of Inception-v3 model include the feature extraction component and the classification component. The former components are reliant on the convolutional neural networks while the latter is reliant on softmax and fully-linked layers [13].

E. Dataset

The FERET database [14], [15] is used in this study to experiment and evaluate the performance of the proposed methodology. The FERET database was established to support machine learning algorithms in both development and evaluation. The database contains 14,126 images associated with 1,199 people.

IV. EXPERIMENTS AND RESULTS

A. Overview of the Proposed Methodology Framework

As shown in Fig. 1, the core of the framework is the Inception model. To build the Inception model, a database has been prepared containing 618 cleaned face forehead images, including both wrinkled and non-wrinkled face foreheads. The face image pre-processing is the first step of the proposed framework, which is important for reliable feature extraction. First, the face and facial landmarks are detected in the image using a Multi-task Cascaded Convolutional Networks model [11], known for its strong and accurate abilities to detect faces and facial landmarks quickly. Before feeding the images into the deep learning Inception model for classifying whether the face foreheads have wrinkles or no wrinkles, an image cropping process is required. Given the bounding box and the facial landmarks, face foreheads can be cropped accurately. Next, k-means algorithm has been used to separate 1,199-image data set of face foreheads into two clusters; then went through each cluster and separated wrinkled foreheads from non-wrinkled foreheads. Also removed the foreheads covered by hair (581 images), then labelled images Wrinkles (309 images) and No-Wrinkles (309 images). The last step of the proposed methodology is to retrain an Inception model for the new categories (Wrinkles, No Wrinkles) to predict whether a face forehead has wrinkles or not.

B. Dataset and Pre-Processing Details

The FERET database [14], [15] is used in the implementation experiments to evaluate the performance of the proposed methodology. The FERET database was established to support machine learning algorithms in both development and evaluation. The database contains 14,126 images associated with 1,199 people. In the experiments only the frontal face images have been used from the database, which amounted to 1,199 unlabelled images.

To detect faces in an image and discover the location of different facial features a pre-trained Multi-task Cascaded Convolutional Networks has been used [11]. Face alignment and face detection are executed jointly in a multi-task training method, enabling the model to properly detect faces and locate five points of facial landmarks.

By default, the pretrained Multi-task Cascaded Convolutional Networks packaged with a face detection weights model. The detector passes a list of Javascript object notation objects. Each JavaScript object notation object carries three main keys: 'confidence', 'keypoints', and 'box':

- The confidence is how probable it is that a bounding box will match a face.

- The box is put into a format [x, y, W, H] which can produce the bounding box around the face.

- The keypoints are arranged into a Javascript object notation object with the keys 'eyeLeft', 'eyeRight', 'nose', 'leftMouth', and 'rightMouth'. Every keypoint is recognized by the position of the pixel (x, y).

Fig. 2 shows a result of an image with a bounding box around the face and five landmarks located on the eyes, mouth, and nose after it has been detected by the MTCNN model. An image cropping process is required. Given the bounding box and the facial landmarks, face foreheads can be cropped accurately.

As regards the labelling process and cleaning data, the manual labelling of images would be a hurdle, so the k-means unsupervised learning clustering algorithm has been used to help with this process.
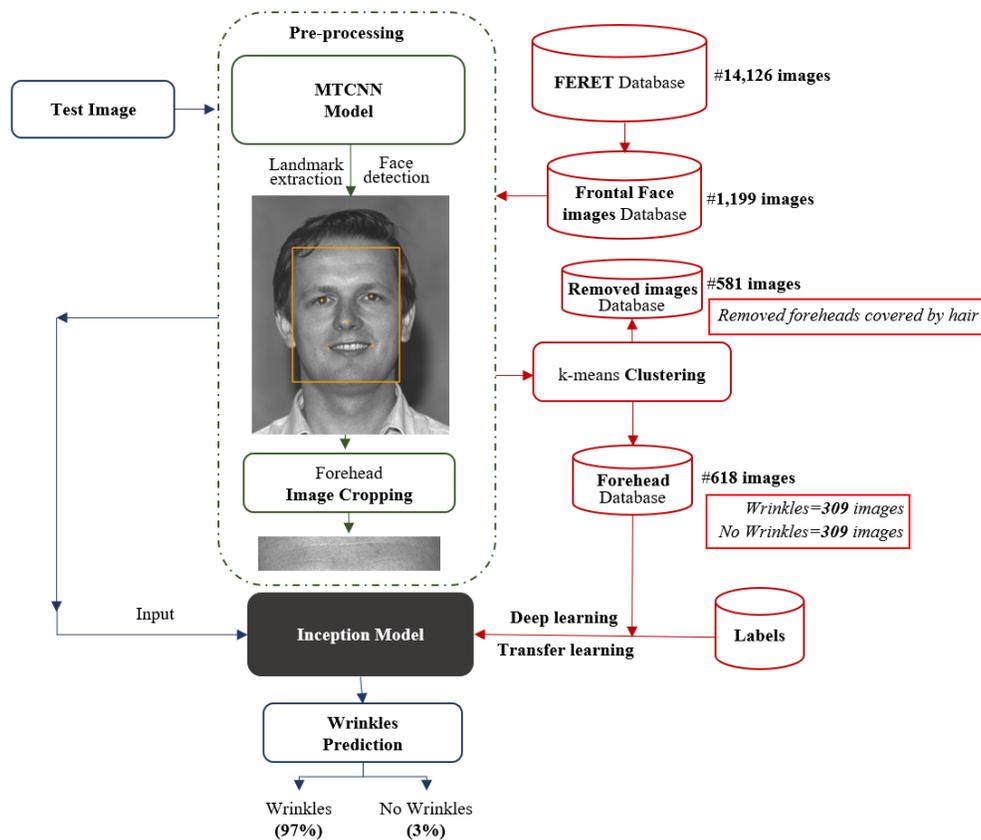
Fig. 1.   Overview of the Proposed Methodology.
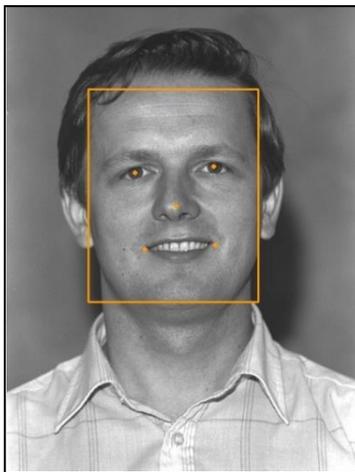
Fig. 2.   The Result of MTCNN Detector.

Data clustering is the process of placing data in similar clusters, which is a branch of data mining. The clustering algorithm divides a data set into several groups since the similarity between points within a particular cluster is greater than the similarity between two points within two different clusters [15]. The idea of data clustering is simple in nature and very close to human in its way of thinking. Whenever people deal with a large amount of data, they tend to summarize the huge amount of data into a few groups or categories in order to facilitate analysis.

So, to label 1,199-image data set of face forehead images into two categories (1- Wrinkles, 2- No Wrinkles), the k-means clustering method has been used to do the job. The k-means algorithm has been used to separate 1,199-image data set of face foreheads into two clusters. After that went through each cluster and separated wrinkled foreheads from non-wrinkled foreheads. Also removed the foreheads covered by hair (581 images), then labelled images into Wrinkles (309 images), and No-Wrinkles (309 images). Finally, an experienced dermatologist reviewed the labelling of the images to ensure they were labelled correctly.

C. Wrinkle Detection

Advanced image classification models have hundreds of parameters. Also, training convolutional neural networks from scratch requires a set of labelled training data and a huge computing power capacity (GPU). Transfer learning is a method that shortcuts many of this by using a model that has previously been trained on a similar task and data and then reusing it in a new model.

In this study, a deep learning image classifier has been retrained for new categories (Wrinkles, No Wrinkles) and made reuse of the capabilities of feature extraction from the powerful image classifier Inception-v3 [13] which was trained on ImageNet and then retrained using a different classification layer on top. The power of transfer learning is that lower layers that become trained to detect among some objects can be reused for many classification tasks without any modification.
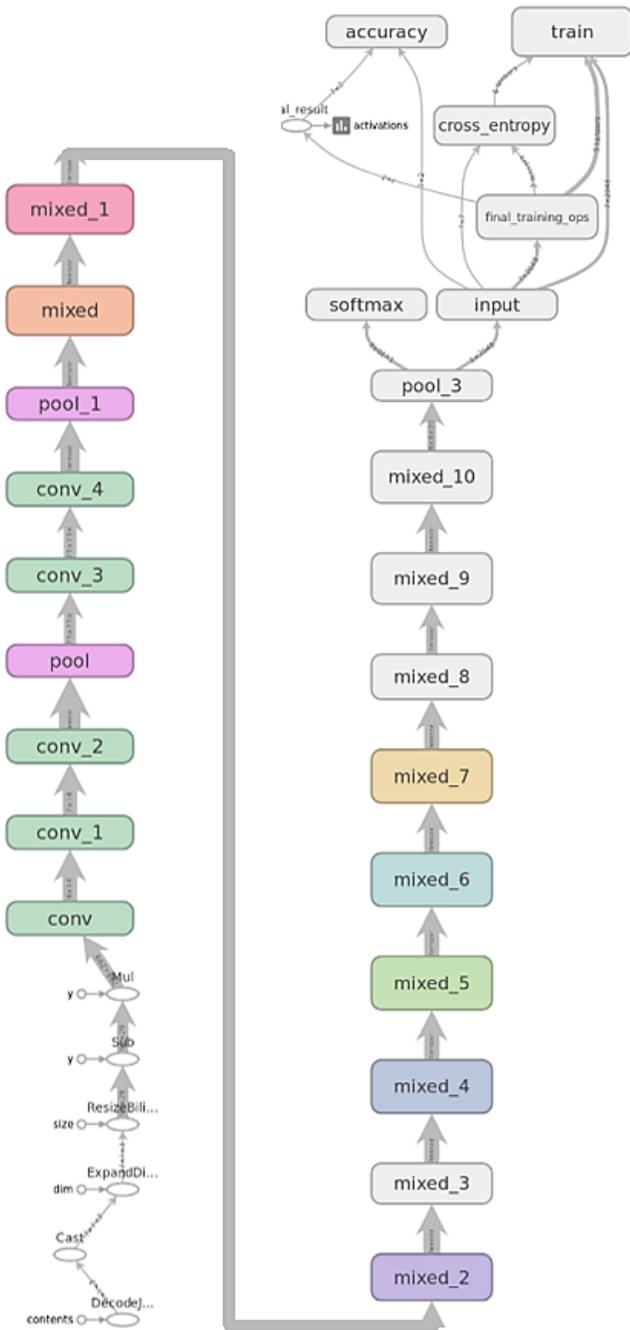
Fig. 3. Retrained Inception Model main Graph.

The aforementioned pretrained model is useful for various applications, works with reasonable training data amounts (hundreds rather than thousands of labelled data), and can be run in as short as ten minutes on a regular laptop without a GPU.

The image data set has been divided up into three separate sets. The largest was the training set, which implies all of the images were fed in the Inception convolutional neural network.

The results were utilized to adjust the weights of the Inception model during training. The first stage analyses all of the images for each class in the folders and then the bottleneck values are calculated and stored for each of them. 'Bottleneck' is a non-formal phrase that is often meant for the layer that carries out the classification before the last output layer. This second-to-last layer has been trained to produce a collection of values that is fair enough to be used by the classifier to discriminate between all of the classes it has been requested to classify [13]. By the time that the bottlenecks values are completed, the real training of the neural network's top layer begins. The learning process is carried out for 500 training steps. Every step takes ten images randomly from the data set for the training, gets their bottlenecks out of the cache, and feeds them inside the last layer to make predictions. Then by the back-propagation process, the last layer's weights get updated. The retrained Inception model graph is shown in Fig. 3.

Some of the image data has been kept outside of the training process so that the Inception model could not memorize them. After that, these data images have been used as an investigation to make sure that overfitting was not happening; if it shows good accuracy on these it's a genuine sign that the network is not overfitting. The data set has been split, with 80% of the images put into the training set and 10% put aside to run as validation regularly during training, ending up with 10% used as a testing set which can predict the real-world performance of the Inception model classifier. Sample results of wrinkle predictions done on test data are shown in Fig. 4.

The final test accuracy = 85.3% with error rate = 14.7%. Test evaluation is the best estimation of how the re-trained Inception model will perform on the classification task.

Input Image (Forehead)	Wrinkles Prediction Score	No Wrinkles Prediction Score
	100%	0%
	20.8%	79.2%
	88.6%	11.4%
	18.5%	81.5%
	98%	2%
	1.5%	98.5%

Fig. 4. Sample Results of Wrinkle Predictions Done on Test Data.
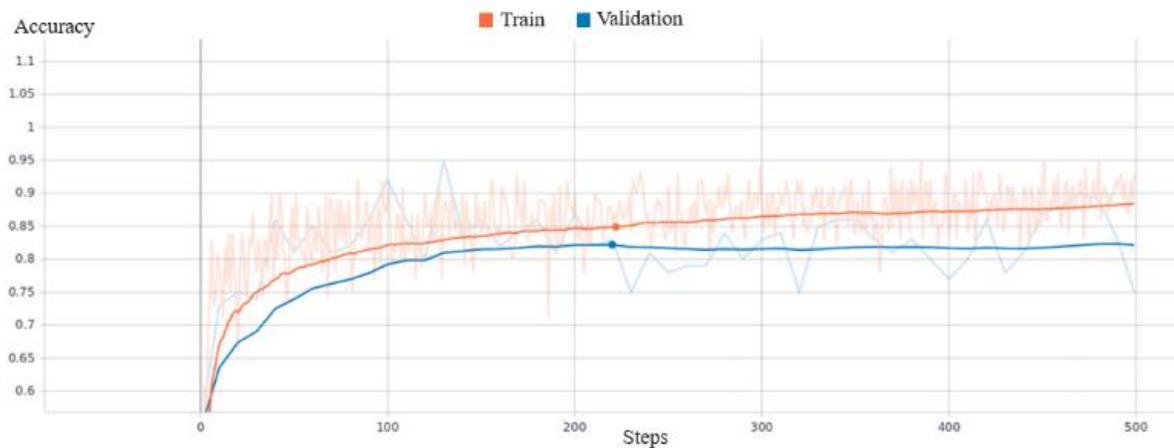
Fig. 5. The Learning Curves for the Training and the Validation of Inception Model.

In order to ensure that overfitting is not occurring in the re-trained Inception model with the new dataset and labels, as shown in Fig. 5, a line plot that displays the learning curves for the training and the validation of the Inception model. The curve for the training set is slightly more accurate than the test set to avoid overfitting.

V. DISCUSSION

One of the main limitations of this study that have been encountered during the investigations of the proposed research is the lack number of previous studies that explore how computer vision methods can help in the facial cosmetic healthcare field.

Furthermore, the resolution quality of FERET images database used in this study is low compared to today's image resolutions also the time constraint allowed only to study one area of the face which is forehead.

VI. CONCLUSION AND FUTURE WORK

Experiments accomplished in the study investigate a proposed novel methodological framework. The Inception model is the core of the framework. By carefully detecting the classification of wrinkles, the model can be built for different applications to aid in the detection of wrinkles that can objectively help in deciding if the forehead area needs to have filler injections. The model achieved an accuracy of 85.3%. To build the Inception model, a database has been prepared containing face forehead images, including both wrinkled and non-wrinkled face foreheads. The face image pre-processing is the first step of the proposed framework, which is important for reliable feature extraction. First, the face and facial landmarks are detected in the image using a Multi-task Cascaded Convolutional Networks model known for its strong and accurate abilities to detect faces and facial landmarks quickly. Before feeding the images into the deep learning Inception model for classifying whether the face foreheads have wrinkles or no wrinkles, an image cropping process is required. Given the bounding box and the facial landmarks, face foreheads can be cropped accurately. The last step of the proposed methodology is to retrain an Inception model for the new categories (Wrinkles, No Wrinkles) to predict whether a face forehead has wrinkles or not.

Since there is a lack in the previous studies there is much to do in the future research studies. For the presented research, there is only one part has been investigated so the other area of the face to be investigated such wrinkles around the eyes and mouth.

Semantic segmentation methods can be used to label each pixel of the wrinkles, and fine lines in the face so that the person can know exactly the location of the areas that need to be filled with filer injection instead of classifying whether an area of the face has wrinkles or not. For more complicated applications it could also be used two frontal face pictures of the same person, one image with a blank expression and the other image with happy facial expression. The model will then compare those two images against each other and then decide if the wrinkles show only in the happy facial expression then it's transient wrinkles and have to be filled with Botox, if the wrinkles appear in the blank facial expression then it needs to be filled with natural substances such as body fat or collagen and artificial substances such as hyaluronic acid. Essentially that would be like an artificial intelligent cosmetic consultant.

REFERENCES

[1] P. Hamet and J. Tremblay, "Artificial intelligence in medicine", Metabolism, vol. 69, pp. S36-S40, 2017. Available: 10.1016/j.metabol.2017.01.011.

[2] C. Ng, M. Yap, N. Costen and B. Li, "Wrinkle Detection Using Hessian Line Tracking", IEEE Access, vol. 3, pp. 1079-1088, 2015. Available: 10.1109/access.2015.2455871.

[3] A. Savran, B. Sankur and M. Taha Bilge, "Regression-based intensity estimation of facial action units", Image and Vision Computing, vol. 30, no. 10, pp. 774-784, 2012. Available: 10.1016/j.imavis.2011.11.008.

[4] W. Xie, L. Shen and J. Jiang, "A Novel Transient Wrinkle Detection Algorithm and Its Application for Expression Synthesis", IEEE Transactions on Multimedia, vol. 19, no. 2, pp. 279-292, 2017. Available: 10.1109/tmm.2016.2614429.

[5] N. Batool and R. Chellappa, "Fast detection of facial wrinkles based on Gabor features using image morphology and geometric constraints",

Pattern Recognition, vol. 48, no. 3, pp. 642-658, 2015. Available: 10.1016/j.patcog.2014.08.003.

[6] N. Batool and R. Chellappa, "Detection and Inpainting of Facial Wrinkles Using Texture Orientation Fields and Markov Random Field Modeling", IEEE Transactions on Image Processing, vol. 23, no. 9, pp. 3773-3788, 2014. Available: 10.1109/tip.2014.2332401.

[7] R. Dechter, "Learning while searching in constraint-satisfaction-problems", Proceeding AAAI'86 Proceedings of the Fifth AAAI National Conference on Artificial Intelligence, pp. 178-183, 1986.

[8] A. Krizhevsky, I. Sutskever and G. Hinton, "ImageNet classification with deep convolutional neural networks", Communications of the ACM, vol. 60, no. 6, pp. 84-90, 2012. Available: 10.1145/3065386.

[9] L. Y. Pratt, "Discriminability-based transfer between neural networks", Proc. Adv. Neural Inf. Process. Syst., pp. 204-211, 1993.

[10] K. Weiss, T. Khoshgoftaar and D. Wang, "A survey of transfer learning", Journal of Big Data, vol. 3, no. 1, 2016. Available: 10.1186/s40537-016-0043-6.

[11] K. Zhang, Z. Zhang, Z. Li and Y. Qiao, "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks", IEEE Signal Processing Letters, vol. 23, no. 10, pp. 1499-1503, 2016. Available: 10.1109/lsp.2016.2603342.

[12] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features", Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, 2001. Available: 10.1109/cvpr.2001.990517.

[13] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision", 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016. Available: 10.1109/cvpr.2016.308.

[14] P. Phillips, H. Wechsler, J. Huang and P. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms", Image and Vision Computing, vol. 16, no. 5, pp. 295-306, 1998. Available: 10.1016/s0262-8856(97)00070-x.

[15] D. MacKay, Information theory, inference, and learning algorithms. Cambridge: Cambridge University Press, 2003, pp. 284–292.