Journal of Imaging

Article

Masked Face Analysis via Multi-task Deep Learning

Vatsa S. Patel[1], Zhongliang Nie[1], Trung-Nghia Le[2] and Tam V. Nguyen[1]

[1]Department of Computer Science, University of Dayton, Dayton, OH 45469 USA
[2]National Institute of Informatics, Tokyo, Japan
Received: XXXX; Accepted: XXXX; Published: XXXX

Abstract: Face recognition with wearable items has been a challenging task in computer vision, which involves human wearing a facial mask. Masked face analysis via multi-task learning could effectively improve the performance in many fields of face analysis. In this paper, we propose a unified framework to predict age, gender, and emotions of people wearing masks. We first construct FGNET-MASK, a masked face dataset for the problem. Then, we propose the multi-task deep learning model to tackle the problem. In particular, the multi-task deep learning model takes the inputs as the data and shares their weight to yield the prediction of age, expression, and gender from the human masked face. Through the extensive experiments, the proposed framework provides a better performance than the existing methods.

Keywords: Multi-task learning, masked face, age, gender, expression, face detection

1. Introduction

Face recognition has been one of the active research problems studied in computer vision for decades due to their practical applications in automotive, security, retail, beautification, and social networks [1-5]. The field of facial analysis has always been a high demand; the so-called facial expression recognition systems are computer programs which aim to automatically translate and understand facial actions from visual information. The processing of facial expression is often confused with emotional interpretation in the field of the machine vision. Due to the high demand, there are lot of developments in this field. Especially, due to covid-19 that make people wear face mask to prevent infection, many previous works meet challenges to analyze the face wearing mask. There are few methods which have been introduced in creating the face mask dataset [6], detecting the face [7], recognizing facial identities [8][9], multitask learning [10, 11, 12], recognizing facial features [13]. The practice/inventions of face detection have been happening through years but due to COVID-19 everything has come to a hold [14], because of the face mask which made the previous methods struggle to analyze the human face. The proposed idea will overcome this problem. This study will assist not only in predicting a person's age [1, 15, 16], but also in predicting his/her gender [2, 17, 18] and mood (expression) [19, 20] with the face mask on. Moreover, we will also release our masked face dataset upon the publication. Multitask learning with different backbones gave a better result in our created dataset than other methods, the output of method is shown in Figure 1.

The main contributions of this paper are three-fold. First, we introduce the simple yet effective mask synthesis method. Second, we build the dataset of masked face with three separate modalities (i.e. age, gender, and expression). Third, we propose the multi-task deep learning framework to tackle the problem. Last but not least, we conduct experiments on the multitask learning model and compare it with the single models [12, 21]. To make this possible, we need to have a good dataset with appropriate labels as an input, which is not available in the market, so we have introduced the dataset with the face mask to make it possible to work. The dataset of the face with labelled age is derived from FG-NET [16], then we manually added labels of gender and expression after that we rendered the faux mask on face [6]. To this end, landmark points on the face was generated, by using the generated landmark points extracted from a landmark point detector [22]. Following the dataset collection, we evaluate different separate models for each label (age, gender, and expression), such as LBP [23] (Local Binary Pattern), Eigenfaces with SVM (Support Vector Machine) classifier [17, 18, 20], deep learning models with two backbones: traditional CNN (Convolutional Neural Networks) [21] and ResNet (Residual Neural Network) and comparing the performance. Finally, multitask deep learning [10, 11, 12] was evaluated, which outperformed single task learning by reducing the effort of constructing different models for each task and the model's result.

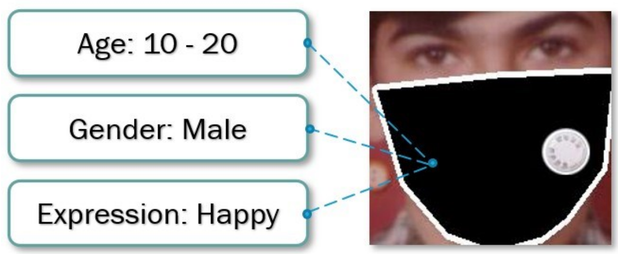


Figure 1. The exemplary input and output of the proposed method.

The rest of the paper is organized as follows. The related works are summarized in Section 2. The dataset and the computational framework are introduced in Section 3. Section 4 presents the experimental results. Finally, Section 5 concludes the paper and paves way to the future work.

2. Related Work

This section explores the current facial datasets. Then, we go through the early studies on facial recognition [24, 25], which is used for feature classification, as well as the various techniques for identifying the face. Finally, we discuss mask face analysis briefly in order to analyze existing work on facial identification with various backbones.

2.1. Face Datasets

Many previous research studies, such as FG-NET [26], LFW (Labelled Faces in the Wild) [27], and Yamaha [28] and many more, have developed databases for facial recognition [8, 9] that are being used in a variety of research projects. The Yamaha dataset [28] only includes Asian faces with no annotation, the large-scale LFW dataset [27] lacks annotation, and the FG-NET dataset [26] contains 926 images including human age annotation. As a result, we are using the FG-NET dataset for our system.

There have been several previous studies that have generated their dataset and conducted various tasks on it. Wang et al. [6] proposed three separate forms of datasets to recognize the individual with the mask, including Masked Face Detection Dataset (MFDD), Real-world Masked Face Recognition Dataset (RMFR), and Simulated Masked Face Recognition Dataset (SMFR). Similarly, many approaches have used datasets and incorporated them into their frameworks, but none of them met our requirements; as a consequence, in our scheme, we use FG-NET [26] as the base and further annotate gender and expression and applying mask on any face photo to construct our dataset.

2.2. Face Recognition

For face recognition, the crucial step is to extract facial features known as a "signature." There are several methods to extract the shape of the lips, eyes, or nose to classify the face based on its scale and distance. Some techniques, which are widely used to extract the face features such as Histograms Oriented Gradient (HOG) [29, 30], Eigenfaces have shown their good performance in terms of system speed and accuracy. Since Eigenfaces method is primarily a dimension reduction method, a system can represent a large number of subjects with a small amount of data. It is also somewhat insensitive to major decreases in image sizing as a face-recognition system; however, it tends to struggle significantly as the difference between the seen images and probe image is large. There are more techniques such as, Independent Component Analysis (ICA), Scale-Invariant Feature Transform (SIFT) [31], Gabor filter, Local Phase Quantization (LPQ), Haar, Local Binary Pattern (LBP) [32, 33]. Here, LBP is a basic but effective textural feature that marks pixels in an image by thresholding each pixel's neighborhood and treating the result as a binary number. Principal component analysis (PCA) [34, 35], which is used in multiple applications and has variety of outcomes, was implemented into our dataset to get the predicted labels. We can derive a wide variety of features from images using CNNs. This feature extraction concept can also be applied to face recognition. For example, a binary classification, when two images of the same person are passed in, the network should return identical outputs (i.e. closer numbers) for both images; while images of two different people are passed in, the network should return somewhat different outputs for both images. The CNN is used to extract the most important data characteristics of the faces, and then the k-nearest neighbors (K-NN) is utilized as a classifier. As the predictive utility of a strong instance value, the K-NN algorithm employs neighborhood classification. An instance-based learning with K-NN [36] is widely used in many applications. In [5], Adjabi *et al.* reviewed facial recognition in both 2D and 3D images. Ulrich *et al.* [37] analyzed the use of RGB-D images for supporting different facial usage scenarios. Bock *et al.* [38] explored low-cost 3D camera in security. Likewise, Ruiqin *et al.* [39] introduced a face recognition access entrance guard system. Dagnes *et al.* [40] investigated the face recognition with eye and mouth occlusions in 3D geometry.

There are few methods which were introduced for emotion recognition [19, 20], and gender recognition [2, 17, 18], age prediction [1, 15, 16], performing separate tasks for each. There are many works implementing multiple tasks with separate model, which is not feasible every time. In multi-task learning for dense prediction, Vandenhende *et al.* [12] review papers on multitasking and variants like hard parameter sharing, soft parameter sharing Encoder-focused model and Decoder-focused model. In our framework we are focusing on hard parameter sharing and sharing the data of age, gender, and expression in such a way.

2.3. Masked Face Analysis

Many prior works have focused on facial recognition of occlusion [41, 42]. The analysis work is conducted in a number of ways, including identification of the face in the wild, twin recognition [43], occluded face detection [41, 42], detecting the face between mask and actual face, and the use of GANs for face modulation and detection, and there are also few detecting the face with mask.

To detect masked faces in the wild with LLE-CNN, Ge *et al.* [9] created a dataset dubbed MAFA. Then, they proposed LLE-CNN method with three modules. The proposal module first combines two pre-trained CNNs to extract candidate facial regions from the input image. Then, the embedding module turns feature descriptors into vectors of weights with respect to the components in pre-trained dictionaries of representative normal faces and non-faces by using locally linear embedding. the verification module takes the weight vectors as input and identifies real facial regions as well as their accurate positions by jointly performing the classification and regression tasks within unified CNNs There also exists research effort [44] to detect a person is with the face mask or without the face mask using OpenCV and Haar Cascade.

We note that the single-model based research on human face recognition has recently achieved state-of-the-art results. There are few of research for facial identification with the face mask achieving high accuracy on recognizing the face [36]. In this work, rather than using a single model for each task, we aim to simultaneously train multi-task for predicting the age, gender, and expression.

3. Data Collection and Proposed Framework

In this section, we introduce the FGNET-MASK dataset and the multi-task deep learning model. The two different methods, namely single and multi-task learning, are shown in Figure 2.

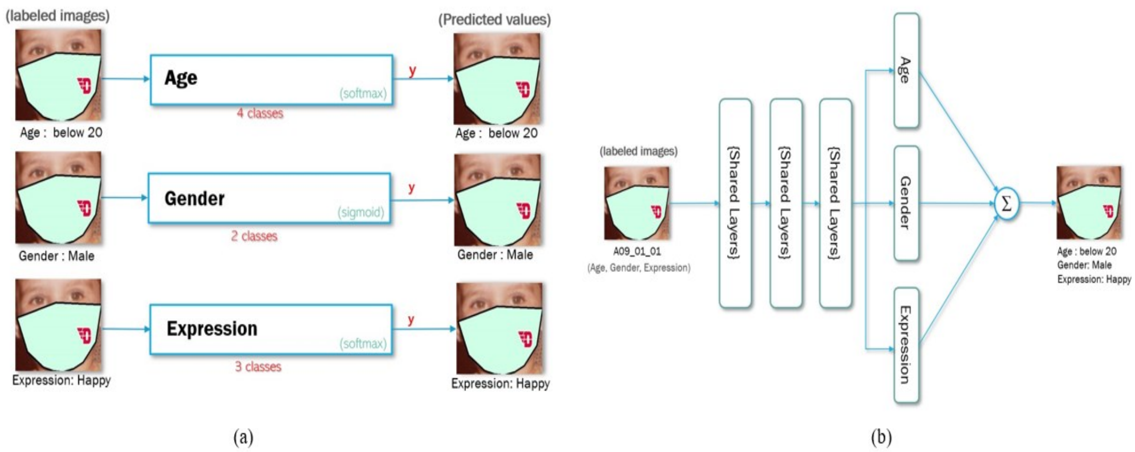


Figure 2. Visualization of deep neural network of (a) single model with single input and output of individual models and (b) multitask neural network with single input and single output with multiple labels.

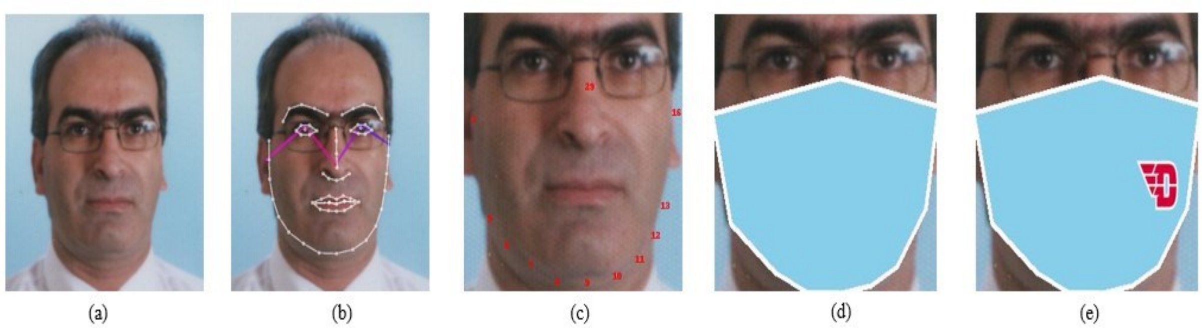


Figure 3. The flowchart of our masked face synthesis (a) is the original image from the FGNET dataset (b) image rendered with 69 facial landmark key points (c) selected landmark points to create mask (d) mask rendered and face cropped by taking min and max values (e) external logo embedded.

3.1. FGNET-MASK dataset collection

The most important step in the framework is the dataset creation. It is extremely difficult to assemble the dataset of individuals of various ages, genders, and expressions with the mask on them, so we render the mask and label them. The construction of FGNET-MASK dataset is detailed as follows.

First, the human face images (without mask) from FGNET [26] were adopted with previously labelled age. Then, the dataset was further manually labelled with people's gender and expression; in total, we have 925 images with three types of labels on each image. Next, the images were run through the OpenPose [22] to detect and generate 2D landmark points on detected faces in the dataset. Eventually, we synthesized the face mask using Pillow package [45] after receiving the landmark points, with a variety of colors and logos. Since the initial dataset only contains 925 images, which is insufficient for a machine to train, we constructed a replica of each masked image x4 with various permutations of mask color, as shown in the Figure 4, as well as changing the undetected landmark points, resulting in the final FGNET-MASK dataset of 3,404 images with rendered face mask, which is sufficient for a machine to be trained. Age, gender, and expression were all branded in the dataset. The dataset contains four age categories: under 10, 10 – 20, 20 – 40, and over 40 to balance the samples for each age group. There are only two genders labelled: Male and Female. Finally, expression labels were classified as Happy, Neutral, or Unhappy.

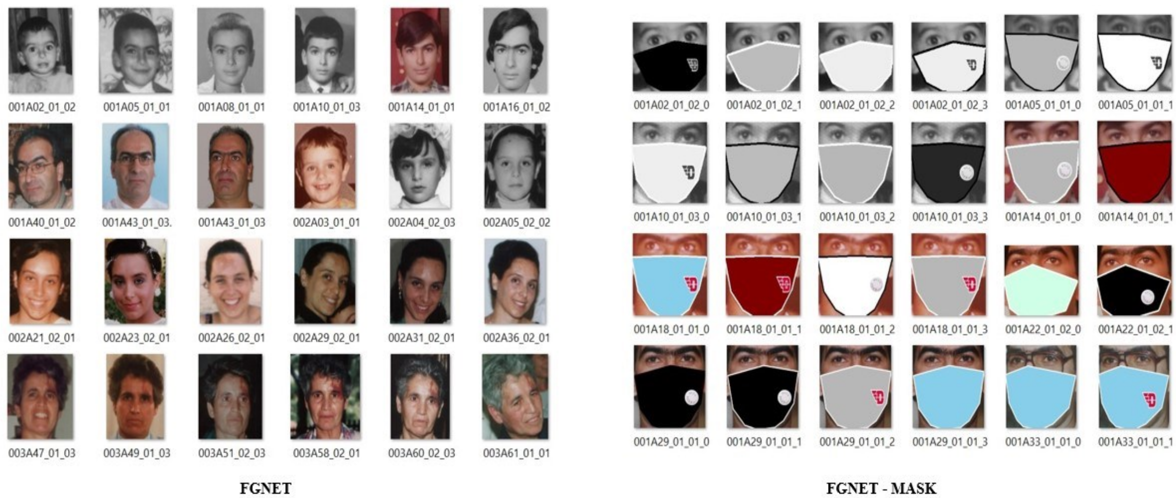


Figure 4. Left: the original images from FGNET, right: the synthesized images of our FGNET-MASK dataset.

The FGNET-MASK dataset was ultimately ready to be used after it was labelled and a variety of masks with logos and valves were applied to the data. The total number of pictures grouped and categorized into their age groups is 1,400 images under the age of 10, 1010 images between the ages of 10 and 20, 720 images between the ages of 20 and 40, and 274 images over the age of 40. There are two gender categories, Male and Female, with 2000 and 1404 images, respectively. There are three types of expressions: happy, neutral, and unhappy. There are 1800 images with happy expressions, 950 images with neutral expressions, and 654 images with unhappy expressions. The outcome of the FGNET-MASK dataset is shown in Figure 4.

3.2. Single models

Following the creation of FGNET-MASK dataset, which includes rendering the mask and labeling the images, the images are fed into three distinct CNNs for Age, Expression, and Gender. The model 'Age' is a multi-class classification with four distinct classes based on criteria of less than 20, 20 – 30, 30 – 40, and greater than 40. The 'Expression' model is also a multi-class classification model with three classes: happy, neutral, and unhappy. The final model, 'Gender,' is a binary classification with Male and Female options.

According to the categories of the respective classes, the single model has three convolutional layers, in which each is followed by a max pooling layer, and dense layers. Convolutional layers use a filter to make a feature map that summarizes the presence of detected features in the input. Max pooling layers is expected to down sample feature detection in feature maps. We used Adam optimizer for training deep networks. Our newly constructed dataset was tested again with a different network with higher complexity of convolutional layers and maximum pooling, namely ResNet152 with 60,430,849 total parameters [45], to compare its accuracy with the traditional CNN model with only 7,654 parameters. The received testing accuracy for age, gender, and expression, structure of both the backbones are shown is Figure 4. ResNet-152 [46] used pre-trained weight on ImageNet as their weights to train the model. The top fully connected layers were excluded, and the model was fine-tuned with 137 layers out of 152.

For the single model, we consider using LBP and Eigen faces with SVM classifier. Local Binary Pattern (LBP) [16, 32, 33] is a simple texture operator that marks the pixels of an image by thresholding the neighborhood of each pixel and treating the result as a binary number. After pre-processing, the dataset was transformed to decimal numbers and fed into the SVM model using linear kernel. All of these data are linearly separated using this kernel which were used to separate models for age, gender, and expression and the results were reported, but deep learning outperformed the LBP - SVM process. Following the LBP implementation, the dataset was further implemented with eigen faces using PCA on SVM models, but the results were worse than the LBP.

3.3. Multi-task Deep Learning

Multi-task deep learning (MTDL) is an inductive transfer learning approach that involves the cooperative training of two or more learning machines. MTDL refers to the mechanism by which a machine learns from one task to the next. The idea is that each task should benefit from the knowledge gained while preparing for other related assignments. Deep multi-task architectures were divided into two types: hard parameter sharing techniques and soft parameter sharing techniques. The parameter set is split into shared and task-specific parameters in hard parameter sharing. In this proposed method we are using hard sharing parameter, MTDL models using hard parameter sharing typically consist of a shared encoder that branches out into task-specific heads.

The most common hard parameter sharing design includes a shared encoder that branches out into task-specific decoding heads. Backpropagation in MTDL is the most efficient method for solving learning distributed representations. For example, in every model, the equation will be the same, if M>2 (i.e. multiclass classification), we calculate a separate loss for each class label per observation and sum the result. For example, L_{age}, the loss function of the age model is computed as:

$$L_{age} = -\sum_{c=1}^{M} y_{o,c} \, log\left(p_{o,c}\right) \qquad (1)$$

Where M is the number of classes (below 10, 10 – 20, 20 – 40, and 40 and above, y is the groundtruth label, p is the predicted probability observation o is of class c. Meanwhile, the total loss function L for the multitask model is computed as follows:

$$L = L_{age} + L_{gender} + L_{\ expression} \qquad (2)$$

The total loss function here solves three optimization problems at the same time: minimization of loss function and making norm of our parameters. Our proposed multi-task learning follows this approach. Following the sharing of the layers with the data, the output was determined in accordance with the specified task (i.e., age with respect to gender and expression, gender with respect to age and expression or expression with respect to age and gender).

4. Experimental Results

In this section, we compare the proposed method for masked face analysis with two implementation variants: basic CNN and ResNet-152. We also compare the single model with the multitask learning model. We include many baselines such as EigenFace [30], LBP [23], TinyImage [47], and VGG Face [48] in the evaluation. All experiments are conducted on the testing set of the collected FGNET-MASK dataset. Regarding the results, we adopt accuracy as the main performance metrics:

$$Accuracy = \frac{\sum_{i=1}^{k} \frac{tp_i + tn_i}{tp_i + tn_i + fp_i + fn_i}}{k} \quad (3)$$

, where tp_i, tn_i, fp_i, fn_i are the true positive, the true negative, the false positive, and the false negative, respectively. Meanwhile, k is the number of classes for each classification task.

4.1. Single Model

Three distinct models were developed in the deep learning system by using two separate backbones (simple CNN and ResNet-152) and three distinct approaches were also used in the SVM method, and the results of the model predicting age, gender, and expression were phenomenal in the deep learning methods than both (LBP and SVM) the methods using SVM method. The testing precision of the single trained models is as follows:

4.1.1. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised machine learning algorithm that can be used to solve classification and regression problems. In the SVM algorithm, each data object is plotted as a point in n-dimensional space (where n is the number of features), with the value of each element being the value of a certain coordinate. Then classification is performed by finding the hyper-plane that differentiates the two or more classes according to the requirement. We use linear SVM as the classifier for LBP, Eigenfaces, TinyImage, and Multi-Block Color-Binarized Statistical Image Features (MB-C-BSIF). As shown in Table 1, LBP obtains the unsatisfactory performance. Using the same process, we have implemented and compared our results of Eigenfaces obtained from PCA (Principal Control Analysis) [2], which is the method of calculating the principal components and using them to modify the basis of the data. The result of Eigenfaces is slightly better than the one of LBP. Regarding the TinyImage, the face image is downsized into 32x32, and the features are extracted by concatenating all image pixels. The extracted TinyImage features are used to train an SVM model, yielding results that were better than eigen face, LBP, and single task CNN. For MB-C-BSIF [49], the extracted features do not perform well, i.e., on par with LBP. One possible reason is that MB-C-BSIF possesses a large dimensionality. That may cause overfitting on the model training. Meanwhile, the features extracted from the VGG Face [48] using a pretrained model, on the other hand, outperform all other feature types.

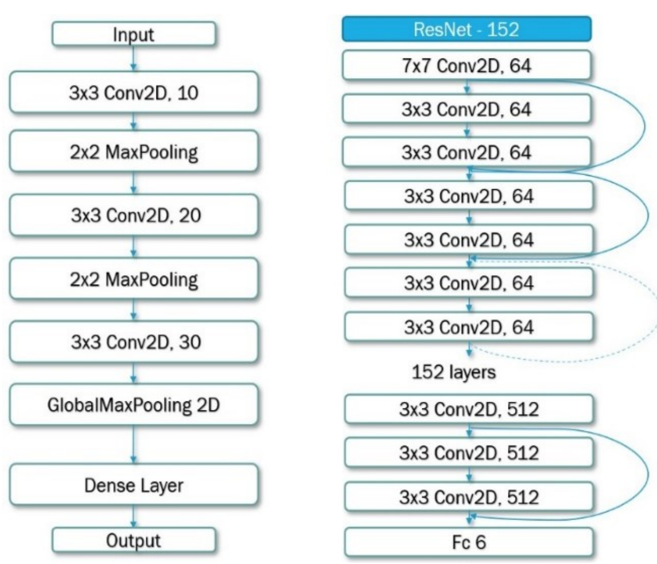


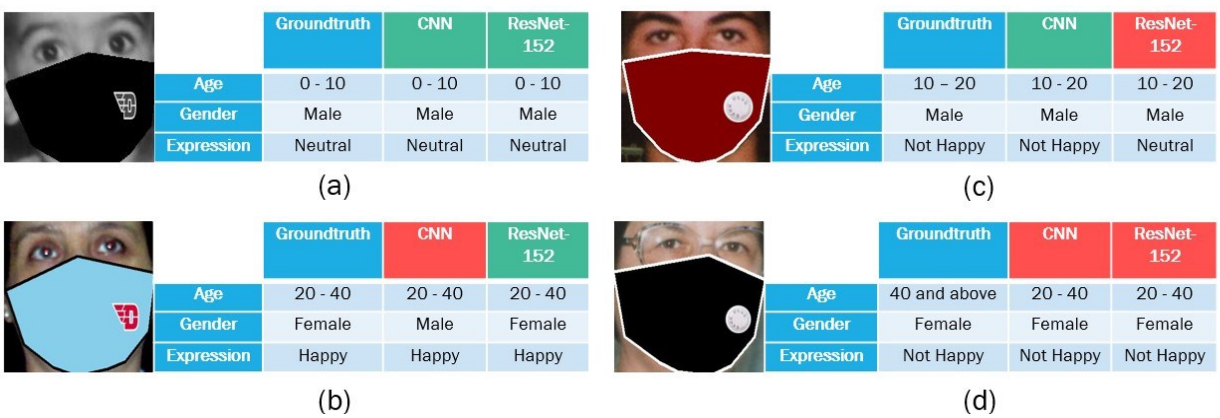
Figure 5. Backbones used in our implementation: (left) simple CNN, (right) ResNet-152.

Table 1. Testing accuracy of the both the models with different backbones. The best performance of each category is marked in boldface.

Accuracy / Method	Age	Gender	Expression
Eigenface	0.59	0.68	0.58
LBP [18]	0.53	0.64	0.55
TinyImage [47]	0.73	0.82	0.70
VGG Face [48]	0.84	0.89	0.75
MB-C-BSIF [49]	0.48	0.64	0.53
Single task (simple CNN)	0.68	0.77	0.60
Single task (ResNet)	0.91	0.95	0.82
MTDL (simple CNN)	0.74	0.83	0.70
MTDL (ResNet)	**0.95**	**0.98**	**0.90**

4.1.2. Simple Convolutional Neural Network (CNN)

In this work, we first try a simple CNN model (as shown in Figure 5 left) for each class, namely, age, gender, and expression. Each model was trained using the same CNN architecture but with different activations functions for binary and multiclass classification. The age model's accuracy on unknown testing data was 0.68, the gender model's testing accuracy was 0.77, and the expression model's accuracy



was 0.60.

Figure 6. The exemplary pictures of age, gender, and expression prediction of a MTDL model from both backbones (CNN and ResNet), with green indicating that the expected values match the ground-truth and red indicating that they do not.

4.1.3. ResNet-152

Furthermore, we try a deeper network, namely, ResNet-152 (as shown in Figure 5 right). By using the deeper model, the result of the age classification task reaches 0.91 whereas gender and expression classification results obtain 0.95 and 0.82, respectively. Here, the accuracy rate obtained from the residual neural network-ResNet-152 is significantly higher than that obtained from the other approaches we used: SVM and Deep Learning (CNN).

4.2. Multitask Deep Learning Model

The method of designing multiple models for multiple labels was exhausting and unconvincing, so the concept of using multitask deep learning was a brilliant way to save time and effort by only creating one model for the requisite multiple labels. The MTDL technique is the best approach to getting better results when compared to single CNN models. Results obtained after comparing the model are far conversing, with respect to age, gender, and expression. Figure 6 showcases the example results. Regarding the simple CNN, the testing accuracy obtained using the CNN backbone for each class are better than the single models. The testing precision obtained after conducting the multitask with respect to age, gender, and expression is 0.74, 0.83, and 0.70, respectively. This evidently outperforms the single models in terms of output. Meanwhile, ResNet-152 model produces a better performance than the simple CNN model. In particular, the results for age, gender, and expression are 0.95, 0.98, and 0.9, respectively. This clearly demonstrates that the deeper backbone tends to obtain the better performance in multi-task deep learning. Note that our work can be adopted in many scenarios such as surveillance systems, person re-identification, targeted advertisement, to name a few.

5. Conclusions and Future Work

In this paper, we investigated the problem of human masked face recognition. We constructed FGNET-MASK, a new masked face dataset with different modalities via face synthesis. We then proposed the multi-task deep learning (MTDL) to give the prediction of the person's (with mask) age, expression, and gender. The experiments show the impressive performance of the proposed method on the testing data. In the future, we would like to collect more data for diversity, we will also explore our work on different datasets like RMFRD [6] in the future. In addition, we will investigate different tasks in masked face analysis such as facial landmark point detection or mask removal.

Acknowledgement: This work was supported by the National Science Foundation (NSF) under Grant 2025234. We also gratefully acknowledge the support of NVIDIA Corporation with the donation of GPU used for this research.

References

1. Guo, G.; Guowang, Mu.; Fu, Y.; Huang, T. S. Human age estimation using bio-inspired features. IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 2009, pp. 112-119, doi: 10.1109/CVPR.2009.5206681.
2. Cao, D.; He, R.; Zhang, M.; Sun, Z.; Tan, T. Real-world gender recognition using multi-order LBP and localized multi-boost learning. IEEE International Conference on Identity, Security and Behavior Analysis (ISBA 2015), Hong Kong, China, 2015, pp. 1-6, doi: 10.1109/ISBA.2015.7126350.
3. Wang, K.; Nguyen, T. V.; Feng, J.; Sepulveda, J. Sense Beyond Expressions: Cuteness. ACM Multimedia, 2015, 1067-1070.
4. Nguyen, T. V.; Liu, L. Smart Mirror: Intelligent Makeup Recommendation and Synthesis. ACM Multimedia, 2017, 1253-1254.
5. Adjabi, I.; Ouahabi, A.; Benzaoui, A.; Taleb-Ahmed, A. Past, Present, and Future of Face Recognition: A Review. MDPI Electronics, vol.9 (8), July 2020, pp.1188.
6. Wang, Z.; Wang, G.; Huang, B.; Xiong, Z.; Hong,Q.; Wu, H.; Yi, P.; Jiang, K.; Wang, N.; Pei, Y.; Chen, H.; Miao, Y.; Huang, Z.; Liang, J. Masked face recognition dataset and application. 2020.
7. Chaudhuri, B.; Vesdapunt, N.; & Wang, B. Joint face detection and facial motion retargeting for multiple faces. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2019, pp. 9719-9728.
8. Kortli, Y.; Jridi, M.; Falou, A.; Atri A & M. Face Recognition Systems: A Survey. Sensors (Basel, Switzerland), 2020, 20(2), 342.
9. Ge, S.; Li, J.; Ye, Q., & Luo, Z. Detecting masked faces in the wild with lle-cnns. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 2682-2690.

10. Cao, J.; Y, Li; Z, Zhang. Partially shared multi-task convolutional neural network with local constraint for face attribute learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 4290-4299.
11. Dmitry, Y.; Tamir, B.; Roman, V. MaskFace: multi-task face and landmark detector. 2020.
12. Vandenhende, S.; Georgoulis, S.; Gansbeke, V.; Proesmans, W.; Dai, M. D., & Van, G.L. Multi-task learning for dense prediction tasks: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.
13. Najibi, M.; Samangouei, P.; Chellappa, R.; & Davis, L. S. Ssh: Single stage headless face detector. In Proceedings of the IEEE international conference on computer vision, 2017, pp. 4875-4884.
14. Mundial, I. Q.; Ul Hassan, M. S.; Tiwana, M. I.; Qureshi, W. S.; Alanazi, E. Towards Facial Recognition Problem in COVID-19 Pandemic. 4rd International Conference on Electrical, Telecommunication and Computer Engineering (ELTICOM), Medan, Indonesia, 2020, pp. 210-214, doi: 10.1109/ELTICOM50775.2020.9230504.
15. Yang, H.; Huang, D.; Wang, Y.; Jain, A. K. Learning face age progression: A pyramid architecture of gans. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 31-39.
16. Nithyashri, J.; Kulanthaivel, G. Classification of human age based on Neural Network using FG-NET Aging database and Wavelets. Fourth International Conference on Advanced Computing (ICoAC), Chennai, India, 2012, pp. 1-5, doi: 10.1109/ICoAC.2012.6416855.
17. Santarcangelo, V.; Farinella, G. M.; Battiato, S. Gender recognition: Methods, datasets and results. IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Turin, Italy, 2015, pp. 1-6.
18. Levi, G.; Hassncer, T. Age and gender classification using convolutional neural networks. IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Boston, MA, USA, 2015, pp. 34-42, doi: 10.1109/CVPRW.2015.7301352.
19. Tivatansakul, S.; Ohkura, M.; Puangpontip, S.; Achalakul, T. Emotional healthcare system: Emotion detection by facial expressions using Japanese database. 6th Computer Science and Electronic Engineering Conference (CEEC), Colchester, UK, 2014, pp. 41-46, doi: 10.1109/CEEC.2014.6958552.
20. Arriaga, O.; Valdenegro-Toro, M.; Ploger, P. Real- ¨time convolutional neural networks for emotion and gender classification. 2017, CoRR abs/1710.07557.
21. Yang, L.; Ma, J.; Lian, J.; Zhang, Y.; Liu, H. Deep representation for partially occluded face verification. EURASIP Journal on Image and Video Processing, 2018, (1), 1-10.
22. Cao, Z.; Hidalgo, G.; Simon, T.; Wei -E, S.; Sheikh, Y. OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. in IEEE Transactions on Pattern Analysis and Machine Intelligence, 1 Jan, 2021, vol. 43, no. 1, pp. 172-186, doi: 10.1109/TPAMI.2019.2929257.
23. Napoléon, T.; Alfalou, A. Local binary patterns preprocessing for face identification/verification using the VanderLugt correlator. In Optical Pattern Recognition XXV; International Society for Optics and Photonics; SPIE, 2014, Volume 9094, p. 909408.
24. Alfalou, A.; Brosseau, C. Understanding Correlation Techniques for Face Recognition: From Basics to Applications. In Face Recognition; Oravec, M., Ed.; IntechOpen: Rijeka.
25. Schroff, F.; Kalenichenko, D.; Philbin, J. Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE conference on computer vision and pattern recognition, Boston, MA, USA, June 2015, pp. 815–823, 7–12.
26. Chen, K.; Gong, S.; Xiang, T.; Chang Loy, C. Cumulative attribute space for age and crowd density estimation. in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013.
27. Huang, B.; Gary, M.; Ramesh, T.; Berg; Learned-Miller, B. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. University of Massachusetts, Amherst, Technical Report 07-49, October, 2007.
28. Fu, Y.; Huang, T. Human Age Estimation With Regression on Discriminative Aging Manifold. IEEE Transactions on Multimedia, 2008, 10(4): pp. 578-584.
29. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In IEEE computer society conference on computer vision and pattern recognition, CVPR'05, 2005, Vol. 1, pp. 886-893.
30. Annalakshmi, M. S.; Roomi, M. M. Naveedh, A. S. A hybrid technique for gender classification with SLBP and HOG features. Cluster Computing, 2019, 22(1), 11-20.

31. Yang, H.; Wang, X.A. Cascade classifier for face detection. J. Algorithms Computational. Technol, 2016, 10, 187–197.
32. Napoléon, T.; Alfalou, A. Pose invariant face recognition: 3D model from single photo. 2017, Opt. Lasers Eng. 89, 150–161.
33. HajiRassouliha, A.; Gamage, T.P.B.; Parker, M.D.; Nash, M.P.; Taberner, A.J.; Nielsen, P.M. FPGA implementation of 2D cross-correlation for real-time 3D tracking of deformable surfaces. In Proceedings of IVCNZ 2013, pp. 352–357.
34. Seo, H.J.; Milanfar, P. Face verification using the lark representation. IEEE Trans. Inf. Forensics Secur. 6, 2011, 1275–1286.
35. Shah, J.H.; Sharif, M.; Raza, M.; Azeem.; A. A Survey: Linear and Nonlinear PCA Based Face Recognition Techniques. Int. Arab J. Inf. Technol, 2013, 10, 536–545.
36. Liu, S. Q.; Lan, X.; Yuen, P. C. Remote photoplethysmography correspondence feature for 3D mask face presentation attack detection. In Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 558-573.
37. Ulrich, L.; Vezzetti, E.; Moos, S.; Marcolin, F. Analysis of RGB-D camera technologies for supporting different facial usage scenarios. Multimedia Tools and Applications, 2020.
38. Bock, R. D. Low-cost 3D security camera. In Autonomous Systems: Sensors, Vehicles, Security, and the Internet of Everything. International Society for Optics and Photonics, May 2018, Vol. 10643, p. 106430E.
39. Ruiqin, L.; Wenan, T.; Zhenyu, C. Design of Face Recognition Access Entrance Guard System with Mask Based on Embedded Development. In Journal of Physics: Conference Series, April 2021, Vol. 1883, No. 1, p. 012156.
40. Dagnes, N.; Marcolin, F.; Nonis, F.; Tornincasa, S.; Vezzetti, E. 3D geometry-based face recognition in presence of eye and mouth occlusions. International Journal on Interactive Design and Manufacturing (IJIDeM), 2019, 13(4), 1617-1635.
41. Shi, Y.; Yu, X.; Sohn, K. M.; Jain, A. K. Towards universal representation learning for deep face recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 6817-6826.
42. Chen, Z.; Xu, T.; Han, Z. Occluded face recognition based on the improved SVM and block weighted LBP. 2011 International Conference on Image Analysis and Signal Processing, Wuhan, China, pp. 118-122, doi: 10.1109/IASP.2011.6109010.
43. Nie, Z.; Mattey, A.; Huang, Z.; Nguyen, T. V. Revisit of Region-Feature Combinations in Facial Analysis. IEEE International Conference on Systems, Man, and Cybernetics (SMC), Miyazaki, Japan, 2018, pp. 2347-2352, doi: 10.1109/SMC.2018.00403.
44. COVID-19 Face Mask Detector with Open CV, Keras, Tensorflow, and Deep Learning, https://www.pyimagesearch.com/2020/05/04/covid-19-face-mask-detector-with-opencv-keras-tensorflow-and-deep-learning/. last retrieved on May 18, 2021.
45. Pillow package, https://pillow.readthedocs.io/en/stable/. last retrieved on May 18, 2021.
46. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778.
47. Torralba, A.; Fergus, R.; Freeman. W. 80 Million Tiny Images: A Large Data Set for Nonparametric Object and Scene Recognition. IEEE Trans. Pattern Anal. Mach. Intell, 2008, 30(11), pp. 1958-1970.
48. Parkhi, O. M.; Vedaldi, A.; Zisserman, A. Deep face recognition. In Proceedings of the British Machine Vision Conference (BMVC) 2015, pp. 41.1-41.12.
49. Adjabi, I.; Ouahabi, A.; Benzaoui, A.; Jaques, S. Multi-Block Color-Binarized Statistical Images for Single-Sample Face Recognition, January 2021, Sensors, vol.21 (3), pp.728, Doi: 10.3390/s21030728.