2020 IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS) | 978-1-7281-6921-7/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICECOCS50124.2020.9314511

Control The COVID-19 Pandemic: Face Mask Detection Using Transfer Learning

Abdellah Oumina
PMIC Laboratory, Faculty of Science and Technology Al-Hoceima, Abdelmalek Essaadi University,
Tetouan, Morocco
abdellah.oumina@gmail.com

Noureddine El Makhfi
PMIC Laboratory, Faculty of Science and Technology Al-Hoceima, Abdelmalek Essaadi University,
Tetouan, Morocco
n.elmakhfi@gmail.com

Mustapha Hamdi
PMIC Laboratory, Faculty of Science and Technology Al-Hoceima, Abdelmalek Essaadi University,
Tetouan, Morocco
hamustapha2000@yahoo.fr

Abstract*- Currently, in the face of the health crisis caused by the Coronavirus COVID-19 which has spread throughout the worldwide. The fight against this pandemic has become an unavoidable reality for many countries. It is now a matter involving many areas of research in the use of new information technologies, particularly those related to artificial intelligence. In this paper, we present a novel contribution to help in the fight against this pandemic. It concerns the detection of people wearing masks because they cannot work or move around as usual without protection against COVID-19. However, there are only a few research studies about face mask detection. In this work, we investigated using different deep Convolutional Neural Networks (CNN) to extract deep features from images of faces. The extracted features are further processed using various machine learning classifiers such as Support Vector Machine (SVM) and K-Nearest Neighbors (K-NN). Were used and examined all different metrics such as accuracy and precision, to compare all model performances. The best classification rate was getting is 97.1%, which was achieved by combining SVM and the MobileNetV2 model. Despite the small dataset used (1376 images), we have obtained very satisfactory results for the detection of masks on the faces.

Keywords*- *Coronavirus COVID-19, Face Mask Detection, Deep Learning, Computer Vision.

I. INTRODUCTION

COVID 19 (Coronavirus) pandemic has affected almost all countries and has made a significant effect on the available healthcare facilities and treatment systems [1]. Public health is considered as the top priority for governments. Therefore, it is a requirement for the introduction of various advanced technologies to tackle multiple problems related to this viral pandemic. Masks could cut the spread of coronavirus. At the moment, the World Health Organization recommends that people should wear face masks if they have respiratory symptoms, or they are taking care of people with symptoms [2]. Furthermore, many public service providers require customers to use the service only if they wear masks. In this paper, we proposed a system to recognize whether a person wearing a mask or not and what is the location of the face based on transfer learning techniques[3], by extracting features from faces using different pre-trained deep learning models (i.e., VGG19, Xception, MobileNetV2, etc.) and combining it with multiple machines learning classifiers (Support Vector Machine, K-Nearest Neighbors, i.e.) for classifying extracted features vectors all this under evaluation metrics.

The rest of this paper is organized as follows: the second section explores related works and similar approaches to object detection and convolutions neural networks. We describe materials and methods used in this experiment, including techniques and the dataset in the third section. In the fourth section, evaluation metrics and results, and conclusion, we review and discuss the experiment results and discuss future work.

II. RELATED WORK

In constructing machine learning classifiers to detect objects, traditional methods were used to extract the features. The most popular is the Viola-Jones detector, which uses Haar features with an integral image method to detect faces[4]. In [5] for human detection, a useful feature extractor called histogram of oriented gradients (HOG) is a feature descriptor where computes the directions and magnitudes of oriented gradients over image cells, scale-invariant feature transform (SIFT) is very useful in practice for image matching and object recognition [6][7], speed up robust feature (SURF) [8], And other, etc. But deep learning-based detector demonstrated excellent performance in the last these years, due to its high feature extraction capability [9], such as CNN, which plays an essential role in pattern recognition tasks related to computer vision, due to its superior spatial feature extraction capability [10]. CNN uses convolution kernels to convolve with the original images or feature maps to extract higher-level features. CNN has been exploited in face recognition and hand-written character recognition [11]. we proposed a CNN model for robust face mask detection.

III. MATERIALS AND METHODS

In the domain of face recognition as a case study for general object recognition with object detection, dataset size can be a decisive factor in model performance due to the limited availability of human-labelled training data. Apart from data in the target domain, related data in a different field can also be included to expand the availability of our prior knowledge about the target future data [12], so we used deep learning models pre-trained on ImageNet [13] dataset (containing 1.2M images of 1000 classes) for our images classification experience. Transfer learning addresses such cross-domain learning problems by extracting useful information from data in a related domain and transferring them for being used in target tasks.

A. Description

The proposed method to detect faces with the mask or without it is based on feature selection using pre-trained deep convolutional neural networks and classification using machine learning algorithms such as K-NN and SVM. The diagram below illustrates in general terms the experiences conducted.

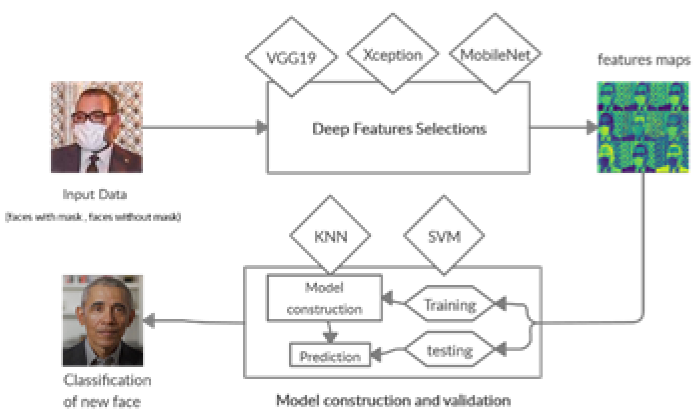

Fig. 1. Proposed method for Proposed method for face mask classification

B. Dataset

The method is tested on a face mask dataset, whose examples can be found in Fig. 2. The dataset covers various faces with masks, faces without masks [14].

Fig. 2. Examples of images in dataset [14].

Since the face mask dataset is a relatively small dataset where features may be hard to extract. For that we used data augmentation, is a strategy to significantly increase the diversity of data available for training models [15], without actually collecting new data. Data augmentation techniques such as cropping, padding, and horizontal flipping [16]. In order, to be able to generalize the results of our experiments on different situations.

Fig. 3. Example of an image by augmentation data

Dataset class distribution is summarized in Table 1.

Table 1 : Classes Of Images And The Number Of Samples

Image Class	Samples count
faces with masks	690
faces without masks	686
Total	1376

This dataset is considered as small in CNNs standards (usually CNNs need thousands of images if the training is started from scratch), but the use of pre-trained models reduces the need for large datasets while maintaining high performances due to transfer learning.

C. Techniques of deep Feature Selection and classifiers

Convolutional neural networks have shown their superior performance compared to other classifiers in image classification tasks [17]. Even though they are not invariant to rotation and geometric distortions [18], pre-trained CNN models trained on huge image datasets (i.e. ImageNet) can extract a deep feature vector invariant to rotation and form changes [20]. Were chose models in this paper because they are publicly available, free, open-source and easy to modify. We compared 3 deep convolutional neural networks to find out the optimal deep learning model as a feature selector for our problem. The concrete structural parameters of all models are provided in Table 2.

Table 2 : Structure Of Used Deep Models

Deep Model	Parameters	Depth	Fully Connected	Default Input size
VGG19	143,667,240	26	4096	224x224
Xception	23,851,784	159	2048	299x299
MobileNetV2	3,538,984	88	1280	224x224

1- VGG19: is a deep convolutional neural network invented by Visual Geometry Group (VGG) from the University of Oxford and was the 1st runner-up of the ImageNet Large Scale Visual Recognition Competition (ILSVRC) 2014 in the classification task and the winner of the localization task. Its simplicity characterizes the network [21]. It uses only 3x3 convolutional layers on top of each other in increasing depth, followed by a max-pooling step to reduce volume size. Two fully connected layers, each with 4,096 nodes are then followed by a softmax activated layer.

2- Xception: is a pre-trained deep learning model for image classification into 1000 classes [22]. it is an extension of the Inception architecture which replaces the standard Inception modules with depthwise separable convolutions , This model was trained on the ImageNet dataset and has achieved an error rate of 3.5% and become the 1st Runner Up for image classification in ImageNet Large Scale Visual Recognition Competition (ILSVRC) 2015 [23].

3- MobileNetV2: is a convolutional neural network architecture that aims to run very efficiently on mobile devices. It can be used as a basic image classifier or as a feature extractor that is part of a larger neural network or in combination with other classifiers [24].

In our case, we will stop just before the last layer of each model and recover a vector of relevant features per image which will be processed in the next step by SVM and K-NN classifiers.

1- Support vector machines (SVMs) are a set of supervised learning methods used to recognize classes of objects, SVMs is one of the most prevailing and exciting supervised learning models with associated learning algorithms that analyze data and recognize patterns. It is used for solving both regression and classification problems. However, it is mostly used in solving classification problems [25].

2- K-Nearest Neighbors (K-NN) is one of the simplest algorithms used in Machine Learning for regression and classification problems. K-NN algorithms use data and classify new data points based on similarity measures (e.g. distance function). Classification is done by a majority vote to its neighbors. The data is assigned to the class which has the nearest neighbors. As you increase the number of nearest neighbors, the value of k, accuracy might increase [26].

IV. EXPERIMENTAL RESULTS AND DISCUSSION

The experiments were conducted in Anaconda environment by using Python language, It uses different Python scientific and open-source libraries for data processing (i.e. Pandas, NumPy, etc.) and for classification algorithms (i.e. TensorFlow, Keras, PyTorch, Scikit learn, etc.). All used classifiers are tested using their default parameters (no special tuning or optimization). The experimental setup computer is an Intel i5-3320M CPU @ 2.60GHz with 8GB RAM.

A. Model Development

The experiment was using all possible combinations of deep models and classification algorithms on our dataset to find the best possible combination for our classification based on the accuracy (i.e. the percentage of instances classified correctly by the classifiers). To validate the constructed model, the dataset was split into a training set and a testing set:

- 80% for the training dataset.
- 20% for the testing dataset.

Table 3 shows the top 4 models based on accuracy.

Table 3 : Best 4 Models Based On Accuracy

Created Model	**Accuracy (%)**
MobileNetV2 – SVM	97.11 %
VGG19 – K-NN	96.65 %
MobileNetV2 – K-NN	94.92 %
Xception - SVM	94.57 %

Based on overall results, we concluded that using transfer learning in detecting the mask on the face is indeed an effective approach, because all models performed well (accuracy over 90%) without any optimization or special tuning. Furthermore, models based on MobileNetV2 and VGG19 surpassed other deep models for this task regardless of the classifier used in the last step. Additionally, SVM was the best classifier with every feature selector and got up to 97.11% of accuracy (with MobileNetV2) followed tightly by K-NN (96.65% with VGG19).

B. Evaluation Metrics

We employed precision and recall as metrics; they are defined as follows.
Recall: Actual true positives over how many times the classifier predicted that class.

$$Recall = \frac{TP}{TP + FN} \tag{1}$$

Precision: Number of correct predictions over how many occurrences of that class were in the test dataset.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

Where TP, FP, and FN denoted the true positive, false positive, and false negative, respectively.

Figure 4 presents the confusion matrix of these top 2 models using test set. The number 0 indicates the faces without a mask and 1 for the face with a mask.

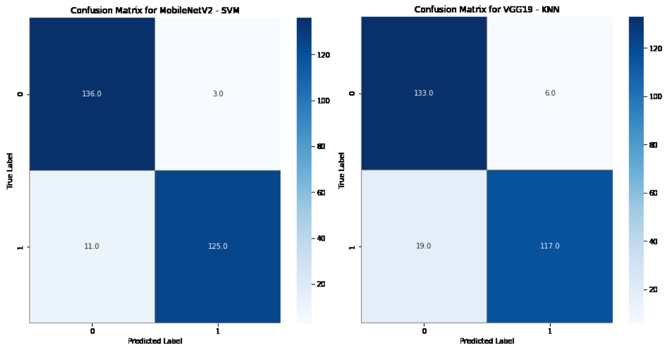

Fig. 4. Confusion matrix for top 2 performing models using test set

The following table presents the precision and recall for top 2 models:

Table 4 : precision and recall for top 2 models

Top 2 Models	Recall	Precision
MobileNetV2 – SVM	94.84%	95.08%
VGG19 – K-NN	90.09%	91.3%

Cross-validation results confirmed the superior performance of SVM with all deep models (over 90%). On the other hand, MobileNet-V2 performed very well regardless of classifier used SVM or K-NN.

Some samples are shown in Fig. 5, including face with a mask, face without a mask and face is masked by hand.

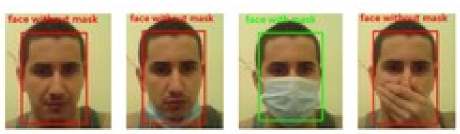

Fig. 5. Examples of detection results

V. CONCLUSION

In this paper, we have proposed a novel face mask detector using deep feature selection with different award-winning pre-trained deep learning models. Features selected are forwarded to different classifiers for classification purposes to 2 categories. The experimental results show that despite a small dataset size (1376 faces), the use of transfer learning is a good approach to classifying faces with masks and faces without masks. Seeing that the combination of MobileNet-V2 model with SVM got an excellent performance (97.11%) with default parameters, we can recommend this approach to be used in practice to detect faces with masks and faces without masks which can possibly contribute to public healthcare, since the main objective of all research is to obtain good results and achieve an effective recognition system.

In our future work, we plan to explore new feature selection techniques and more specialized machine learning algorithms and use a bigger dataset for validation to deal with more complex problems than this one.

REFERENCES

[1] World Health Organization et al. Coronavirus disease 2019 (COVID-19): situation report, 2020.

[2] Shuo Feng, Chen Shen, Nan Xia, Wei Song, Mengzhen Fan, and Benjamin J Cowling. Rational use of face masks in the COVID-19 pandemic. The Lancet Respiratory Medicine, 2020.

[3] Chuanqi Tan, Fuchun Sun, Tao Kong, Wenchang Zhang, Chao Yang, Chunfang Liu. A Survey on Deep Transfer Learning. The 27th International Conference on Artificial Neural Networks, 2018.

[4] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition, 2001.

[5] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. IEEE computer society conference on computer vision and pattern recognition, 2005.

[6] Pedro Felzenszwalb, David McAllester, and Deva Ramanan. A discriminatively trained, multiscale, deformable part model. IEEE Conference on Computer Vision and Pattern Recognition, 2008.

[7] N. El Makhfi and R. Benslimane, ”Feature extraction in segmented words for semi-automatic transcription of handwritten Arabic documents”, J. of Theoretical and Applied Information Technology, v.70, n. 1, 68–75, 2014.

[8] Ebrahim Karami, Siva Prasad, Mohamed Shehata. Image Matching Using SIFT, SURF, BRIEF and ORB: Performance Comparison for Distorted Images. In Proceedings of the 2015 Newfoundland Electrical and Computer Engineering Conference,St. johns, Canada, November, 2015.

[9] Zhengxia Zou, Zhenwei Shi, Yuhong Guo, and Jieping Ye. Object detection in 20 years: A survey. IEEE TPAMI for possible publication, 2019.

[10] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. IEEE conference on computer vision and pattern recognition, 2009.

[11] Musab Coşkun, Ayşegül Uçar, Özal Yildirim, Yakup Demir. Face recognition based on convolutional neural network. International Conference on Modern Electrical and Energy Systems (MEES), 2017.

[12] L. Shao, F. Zhu, and X. Li, “Transfer learning for visual categorization: A survey,” IEEE Transactions on Neural Networks and Learning Systems, vol. 26, May 2015.

[13] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei: ImageNet: A large-scale hierarchical image database, IEEE Conference on Computer Vision and Pattern Recognition, 2010.

[14] https://github.com/prajnasb/observations/tree/master/experiements/data/ Faces dataset, Accessed: June 15, 2020

[15] Agnieszka Mikołajczyk; Michał Grochowski: Data augmentation for improving deep learning in image classification problem. 2018 International Interdisciplinary PhD Workshop (IIPhDW), 9-12 May 2018.

[16] Connor Shorten & Taghi M. Khoshgoftaar: A survey on Image Data Augmentation for Deep Learning, Journal of Big Data volume 6, Article number: 60, 2019.

[17] J. Wang, Y. Yang, J. Mao, Z. Huang, C. Huang, and W. Xu: CNNRNN : A Unified Framework for Multi-label Image Classification, in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, 2016.

[18] Y. Lecun, Y. Bengio, and G. Hinton: Deep learning , 2015.

[19] A. S. Tarawneh, C. Celik, A. B. Hassanat, and D. Chetverikov: Detailed investigation of deep features with sparse representation and dimensionality reduction in cbir: A comparative study, nov 2018.

[20] K. Simonyan and A. Zisserman: Very deep convolutional networks for large-scale image recognition, 3rd International Conference on Learning Representations, 2015.

[21] K. Simonyan and A. Zisserman : Very deep convolutional networks for large-scale image recognition, 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, 2015.

[22] François Chollet: Xception: Deep Learning with Depthwise Separable Convolutions, Computer Vision and Pattern Recognition, Apr 2017.
[23] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna: Rethinking the Inception Architecture for Computer Vision, in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, 2016.
[24] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen: MobileNetV2: Inverted Residuals and Linear Bottlenecks, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Jan 2018.
[25] Gidudu Anthony, Hulley Greg, Marwala Tshilidzi: Classification of Images Using Support Vector Machines, Machine Learning & Artificial Intelligence (cs.AI), 2007.
[26] Yijie Dang, Nan Jiang, Hao Hu, Zhuoxiao Ji, Wenyin Zhang : Image Classification Based on Quantum K-NN Algorithm, omputer Vision and Pattern Recognition, May 2018.