Effective COVID-19 Screening using Chest Radiography Images via Deep Learning

Uboho Victor*, Xishuang Dong*†, Xiangfang Li*†, Pamela Obiomon*†, Lijun Qian*†
*Center of Excellence in Research and Education for Big Military Data Intelligence (CREDIT)
†Center for Computational Systems Biology (CCSB)
Department of Electrical and Computer Engineering, Prairie View A&M University
Prairie View, TX 77446, USA

***Abstract*—Timely and accurate screening/testing is crucial to fighting COVID-19. Compared to commonly used reverse-transcriptase polymerase chain reaction (RT-PCR), chest radiography imaging (X-ray) is also a reliable, practical and rapid method to diagnose and assess COVID-19. In this paper, two types of deep learning models, namely, Convolutional Neural Networks (CNN) and Residual Neural Networks (ResNet) have been designed and tested for accurate diagnosis of COVID-19 with chest X-ray images. Experimental results demonstrate the effectiveness of the proposed approach.**

***Index Terms*—Coronavirus Disease, Residual Neural Network, Convolutional Neural Networks, Chest Radiography Image**

I. Introduction

The COVID-19 outbreak is a human tragedy and has a growing impact on everyone's life and the global economy. While the medical community is working hard to find specific therapeutic drugs and vaccines for COVID-19, it is essential to detect the disease at an early stage and put an infected patient in quarantine immediately. It has been shown that effective and timely screening/testing of COVID-19 plays a major role to reduce the spreading of the virus and reduce the death rate [1].

Two kinds of tests are available for COVID-19: viral tests and antibody tests [2]. A viral test try to detect a current infection, while an antibody test tells you if you had a previous infection. However, for viral tests, the current detection method using reverse-transcriptase polymerase chain reaction (RT-PCR) may not have the desired accuracy [3]. Thus, an alternative diagnostic tool that complementary to RT-PCR is necessary.

Computerized Tomography (CT) scan is a diagnostic tool that physicians use to rule out or confirm the presence of certain abnormalities or diseases. In the case of COVID-19, chest CT imaging is an alternative method to diagnose and assess COVID-19 [3], [4]. There have been promising studies applying machine learning to help the diagnosis of COVID-19 based on CT scans [5], [6]. Despite the success of these studies, the fact remains that COVID-19 is an infection that is likely to be experienced by diverse communities of all sizes. Comparing to CT scan, X-rays are inexpensive and they are more accessible to healthcare providers, thus they are readily available in many communities and especially communities with less medical resources. Furthermore, X-rays are quick to perform, thus they are attractive options when there are a very large number of patients to screen during the COVID-19 pandemic. Therefore, using chest radiography imaging may be a more reliable, practical and rapid method to screen patients for COVID-19 [7], [8].

In this study, two types of deep learning models, namely, Convolutional Neural Networks (CNN) and Residual Neural Networks (ResNet) have been designed and tested for accurate diagnosis of COVID-19 with chest radiography images. Specifically, the proposed deep learning models have been tested on a COVIDx dataset that was recently made public [9]. COVIDx is comprised of 13,800 chest radiography images across 13,725 patient cases from three open access data repositories. The dataset contains three classes of X-ray images: normal, Pneumonia (but not COVID-19), and COVID-19 [9]. Hence, the machine learning task is formulated as a 3-class classification problem. In the first part of the experiments, three designs of small CNNs have been trained and tested on a subset of the COVIDx dataset to carry out a proof-of-concept experiment. Then ResNet with different depths and hyperparameters are trained from scratch and tested using the entire COVIDx dataset. It is observed that ResNet with appropriate depth and hyperparameters performs the best. The experiments demonstrate that deep learning based classifier may provide a promising tool to aid the screening of COVID-19 using chest radiography images.

II. Deep Learning Models

There are two types of deep learning models considered in this study: CNN and ResNet.

A. CNN

Convolutional neural networks (CNN) has significantly promoted developments of visual processing tasks such as image classification [10], object detection [11] and tracking [12], and semantic segmentation [13]. The advancement of CNN benefits from accessible big datasets such as ImageNet [14] and YouTube-BoundingBoxes [15] that can be used to train large-scale models. Its general architecture for image classification is shown in Figure 1. CNN based state-of-the-art deep learning architectures such as AlexNet [10], VGG [16], and GoogleNet [17] are proposed to make rapid progress in image classification. Millions of annotated samples in these big datasets help us to estimate appropriate parameters in these architectures successfully. Further work has advanced CNN by combining CNN with other deep learning models. For example, Wang *et al.*[18] combine CNN with recurrent neural networks (RNN) for multi-label image classification. In addition, combining CNN with autoencoder [19], [20] has been verified to solve tasks such as face rotation and intrinsic transformations for objects. In this study, three small CNNs have been trained and tested on a subset of the COVIDx dataset to carry out a proof-of-concept experiment.

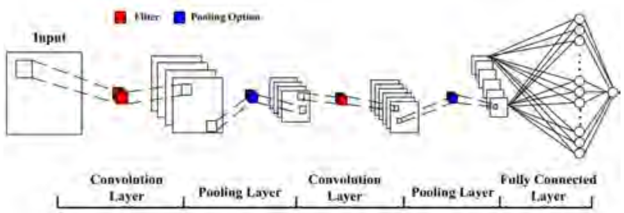


Fig. 1. Convolutional neural network architecture for image classification.

B. ResNet

Residual neural networks (ResNet) [21] is an artificial neural network that builds on constructs known from pyramidal cells in the cerebral cortex. It utilizes skip connections, or shortcuts to jump over some layers. The main idea of ResNet is introducing a so-called "identity shortcut connection" that skips one or more layers. It is based on the observation that stacking layers should not degrade the network performance, because we could simply stack identity mappings upon the current network, and the resulting architecture would perform the same. It indicates that the deeper model should not produce a training error higher than its shallower counterparts. The design is motivated by the intuition that the stacked layers fitting a residual mapping is easier to train than directly fitting the desired underlying mapping. This idea is implemented by the residual block as shown in Figure 2.

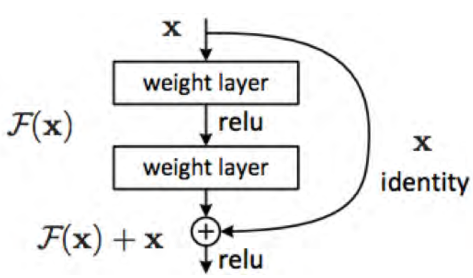


Fig. 2. General architecture of the residual block of ResNet [21].

The most important modification on standard CNN is the "skip connection" for identity mapping. This identity mapping does not have any parameters and is just to add the output from the previous layer to the layer ahead. However, sometimes x and $F(x)$ will not have the same dimension. Recall that a convolution operation typically shrinks the spatial resolution of an image, e.g., a 3×3convolution on a 32×32 image results in a 30×30 image. The identity mapping is multiplied by a linear projection W to expand the channels of shortcut to match the residual. This allows for the input x and $F(x)$ to be combined as input to the next layer. Given the benefit we would obtain from ResNet, we will employ different structures of ResNet for screening COVID-19 using the entire COVIDx dataset.

III. Experiments and Results Analysis

A. Dataset

Currently, COVID-19 pandemic is still continuing to affect the health and well-being of the global population. A critical step to fight against COVID-19 is to test and locate infected patients with effective and efficient tools. One of potential methods is to examine radiological images using chest radiography. From results in the previous studies, patients presenting abnormalities in chest radiography images would have high risk of being infected with COVID-19 [9]. Therefore, we employ a chest radiography dataset in this study. The dataset is COVIDx that is comprised of 13,800 chest radiography images across 13,725 patient cases from three open access data repositories. Example chest radiography images belong to normal, pneumonia, and COVID-19 classes from COVIDx dataset are shown in Figure 3. The detailed information on the dataset are shown in Table I and Table II, which contains sample distribution and patient distribution, respectively. It can be observed in Table I that the sample distribution is extremely unbalanced with very small number of samples in the COVID-19 class. This poses a great challenge for obtaining a classifier with high performance. Similar observation is obtained for patient distribution in Table II.

TABLE I
Sample distribution in different classes for training and testing datasets

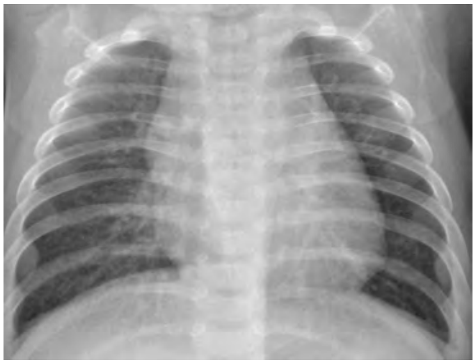
Dataset	Normal	Pneumonia	COVID-19	Total
Training	7,996	5,451	152	13,569
Testing	100	100	31	231

TABLE II
Patient distribution in different classes for training and testing datasets

Dataset	Normal	Pneumonia	COVID-19	Total
Training	7,996	5,440	107	13,513
Testing	100	98	14	212

B. Experimental settings

We performed two groups of experiments. In the first group of experiments, three small CNNs have been trained and tested on a subset of the COVIDx dataset to carry out a proof-of-concept experiment. Then ResNet with different depths and hyperparameters are trained from scratch and tested using the entire COVIDx dataset. Experimental settings for these two groups are shown below.



(a) normal

(b) pneumonia

(c) COVID-19

Fig. 3. Example chest radiography images belong to normal, pneumonia, and COVID-19 classes from COVIDx dataset are shown in (a), (b) and (c), respectively.

1) Small CNNs: We build three small CNNs and the detailed architectures are given as follows.

- CNN structure 1: CONV1(16, (3*3)) - Relu1 - Maxpooling1(2*2) - Dropout1(0.25) - CONV2(32, (3*3)) - Relu2 - Maxpooling2(2*2) - Dropout2(0.25) - Dense(512) - Relu3 - Dropout3(0.5) - Output(Softmax)
- CNN structure 2: CONV1(16, (3*3)) - CONV2(16, (3*3)) - Relu1-Maxpooling1(2*2) - Dropout1(0.25) - CONV3(32, (3*3)) - Relu2 - Maxpooling2(2*2) - Dropout2(0.25) - Dense(512) -Relu3 - Dropout3(0.5) - Output(Softmax)
- CNN structure 3: CONV1(16, (3*3)) - Relu1 - Maxpooling1(2*2) - Dropout2(0.25) - CONV2(16, (3*3)) - CONV3(32, (3*3)) - Relu2 - Maxpooling2(2*2) - Dropout2(0.25) - Dense1(512) - Relu3 - Dropout3(0.5) - Output(Softmax)

2) ResNet: The key hyper parameters of ResNet are: ResNet Models: 18, 34, 50, 101, and 152, Input size: $64 \times 64 \times 3$, Filter size: 3×3, Stride: 1, Number of filters: 32, No of classes: 3 (Normal, Pneumonia, COVID-19), Learning rate: 0.01, Epoch: 82, Batch size: 256, Regularizer: L2. They are determined by trial and error.

C. Evaluation metrics

We applied different evaluation metrics to evaluate the performance of our proposed models. Accuracy is calculated by dividing the number of image identified correctly over the total number of testing images.

$$Accuracy = \frac{N_{correct}}{N_{total}}. \tag{1}$$

$Fscore$ is the weighted average of $Precision$ and $Recall$ scores,

$$Fscore = \frac{2 \times Precision \times Recall}{Precision + Recall}, \tag{2}$$

where $Precision$ indicates precision measurement that defines the capability of a model to represent only correct image classes and $Recall$ (or sensitivity) is the fraction of the total amount of relevant instances that were actually retrieved:

$$Precision = \frac{TP}{TP + FP}. \tag{3}$$

$$Recall = \frac{TP}{TP + FN}. \tag{4}$$

where TP (True Positive) counts total number of images matched with the images in the classes. FP (False Positive) measures the number of recognized classes does not match the annotated images. FN (False Negative) counts the number of images that does not match the predicted images.

D. Experimental results

We evaluated the proposed models from two perspectives. One is to verify if we could recognize COVID-19 cases from small dataset with small CNNs effectively. The other is to figure out whether we could apply ResNet to recognize COVID-19 images from a large dataset *without transfer learning techniques*.

1) Small Dataset: For the case of small dataset, we select 350 images from the original COVIDx dataset for training and testing three small CNNs, namely 300 images for training and 50 images for testing, respectively. The training and testing accuracy of these three models are shown in Figure 4. It is observed that the shallow CNN (CNN1) has the underfitting issue. When we add more layers for extracting more sophisticated features, higher accuracy is obtained with CNN3.

To examine the performance of the three CNNs in detail, we listed the precision, recall, and fscore during testing in Table III. We can observe that when we employ deeper models like CNN2 and CNN3, higher performance is obtained with respect to the Fscore values. It means that higher level features will contribute more significantly to recognize COVID-19 images.

2) COVIDx Dataset: For this experiment, we employ all images from the COVIDx dataset to implement the classifier for COVID-19 recognition. Firstly, we preprocess the data by compressing the images. The original size of the X-ray images from the dataset was $1024 \times 1024 \times 3$. In order to speed up training, we compressed the size of the image to

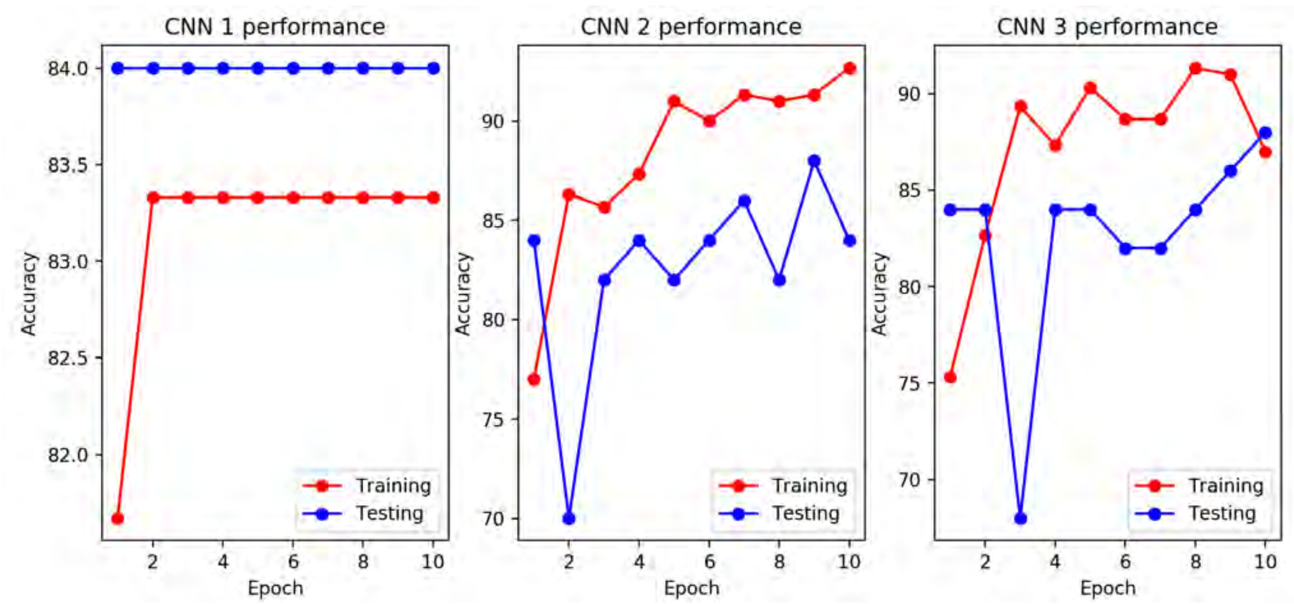


Fig. 4. Performance on training and testing accuracy for three small CNNs.

TABLE III
PERFORMANCE COMPARISON FOR THREE CNNS TO RECOGNIZE COVID-19 IMAGES

Model	Accuracy	Precision	Recall	Fscore
CNN1	83.99%	0.00%	0.00%	0.00%
CNN2	83.99%	50.00%	100.00%	66.66%
CNN3	87.99%	89.7%	92.5%	92.6%

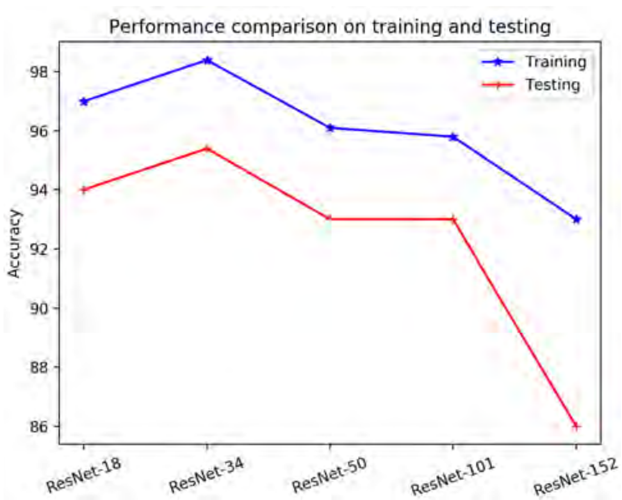


Fig. 5. Performance comparison on training and testing

$64 \times 64 \times 3$. Then we complete the training from scratch on this preprocessed dataset. The performance results of training and testing are shown in Figure 5. It is observed that there exists a sweet spot, ResNet-34 that outperforms other models. After ResNet-34, a steady decline of performance is found as we increased the ResNet layers to 152. This is most likely due to the fact that as the number of layers increased, the models are overfitting and the data is not sufficient to effectively train the model although it took slightly longer time to train the ResNet models as we increased the layers. The results also demonstrate that we are able to obtain good performance by training ResNet from scratch without the help of transfer learning.

TABLE IV
PERFORMANCE COMPARISON FOR DIFFERENT RATIOS OF LABELED IMAGES FOR TRAINING RESNET TO RECOGNIZE COVID-19 IMAGES

Model	10%	20%	30%	40%	50%
ResNet-18	89.3%	91.8%	92.6%	93.40%	93.46%
ResNet-34	90.0%	91.8%	93.1%	93.54%	94.05%
ResNet-50	86.4%	89.7%	92.5%	90.70%	92.16%

To explore detailed influence of the available labeled data to the model performance, we performed the experiments of training ResNet with different percentage of labeled data. To ensure that all classes are represented in the dataset distribution for the percentage labeled training data, we split the training data based on their classes and then selected a percentage of samples from each class. The detailed performance results are given in Table IV. The results showed consistent improvement as we increased the percentage of the labeled data. With as few as 10% labeled data, we achieved the highest testing accuracy of 90% with the ResNet-34 model. Overfitting was observed when we used ResNet-50 to train with the 10% labeled data as the training accuracy was as high as 95% while the testing accuracy was about 86.4%. These results indicate that the performance of ResNet with appropriate depth is rather robust to available labeled images.

IV. Related Work

There are many recent works on applying machine learning techniques to identify COVID-19 using CT scan images, such as [5], [6], just to name a few. Although the results are encouraging, the dataset are not publicly available. Moreover, comparing to CT scan, X-rays are inexpensive and they are more accessible to healthcare providers, and they are available in communities with less medical resources. Furthermore, X-rays are quick to perform, and this will reduce the screening time when there are a very large number of patients to screen during the COVID-19 pandemic.

Deep learning techniques had been applied to screen COVID-19 using chest X-ray images [7], [8]. Pre-trained CNN and ResNet were applied to detect COVID-19 using a small dataset (several hundred images) in [7]. The authors in [8] applied ResNet to identify COVID-19 using chest X-ray images. In both of their works, transfer learning had been applied. In this paper, we trained ResNet from scratch without transfer learning. Moreover, we performed the experiments of training ResNet with different percentage of labeled data, and the results indicate that the performance of ResNet with appropriate depth is rather robust to available labeled images. This may be significant in practice when the number of labeled images are limited.

V. Conclusion and Future Work

Early diagnosis of COVID-19 is crucial for disease treatment and control, and effective screening/testing is the key. Comparing to RT-PCR, using chest radiography imaging may be a more reliable, practical and rapid method to screen patients for COVID-19. In this paper, two types of deep learning models, namely, Convolutional Neural Networks (CNN) and Residual Neural Networks (ResNet) have been designed and tested for accurate diagnosis of COVID-19 with chest radiography images. The experimental results demonstrate that we are able to obtain good performance by training ResNet from scratch without the help of transfer learning. Furthermore, we observe that the performance of ResNet with appropriate depth is rather robust to available labeled images. This is our first attempt to apply deep learning techniques to help in the fight against the COVID-19 pandemic. The deep learning techniques and models are currently in research stage and they are not intended for direct clinical diagnosis. However, we are continuously improving them as new data becomes available. The hope is to provide highly accurate yet practical deep learning solutions for detecting COVID-19 cases using chest radiography images.

Acknowledgment

This research work is supported by the U.S. Office of the Under Secretary of Defense for Research and Engineering (OUSD(R&E)) under agreement number FA8750-15-2-0119. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Office of the Under Secretary of Defense for Research and Engineering (OUSD(R&E)) or the U.S. Government.

References

[1] "NY Times: Germany got testing right, what can we learn?" https://www.nytimes.com/2020/04/28/opinion/coronavirus-testing-united-states.html, accessed: 2020-04-29.

[2] "CDC: Testing for COVID-19," https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/testing.html, accessed: 2020-04-19.

[3] T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, Q. Tao, Z. Sun, and L. Xia, "Correlation of chest ct and rt-pcr testing in coronavirus disease 2019 (covid-19) in china: A report of 1014 cases," *Radiology*, p. 200642, pMID: 32101510. [Online]. Available: https://doi.org/10.1148/radiol.2020200642

[4] J. Zhao, Y. Zhang, X. He, and P. Xie, "Covid-ct-dataset: A ct scan dataset about covid-19," 2020.

[5] Y. Song, et. al., "Deep learning enables accurate diagnosis of novel coronavirus (covid-19) with ct images," *medRxiv*, 2020. [Online]. Available: https://www.medrxiv.org/content/early/2020/02/25/2020.02.23.20026930

[6] F. Shan, Y. Gao, J. Wang, W. Shi, N. Shi, M. Han, Z. Xue, D. Shen, and Y. Shi, "Lung infection quantification of covid-19 in ct images with deep learning," 2020.

[7] L. Hall, D. Goldgof, R. Paul, and G. M. Goldgof, "Finding covid-19 from chest x-rays using deep learning on a small dataset," Apr 2020. [Online]. Available: http://dx.doi.org/10.36227/techrxiv.12083964

[8] M. Farooq and A. Hafeez, "Covid-resnet: A deep learning framework for screening of covid19 from radiographs," 2020.

[9] L. Wang and A. Wong, "Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest radiography images," *arXiv preprint arXiv:2003.09871*, 2020.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, 2012, pp. 1106–1114.

[11] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.

[12] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014.

[13] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.

[14] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *IEEE Conference on Computer Vision and Pattern Recognition, 2009.*, pp. 248–255.

[15] E. Real, J. Shlens, S. Mazzocchi, X. Pan, and V. Vanhoucke, "Youtube-boundingboxes: A large high-precision human-annotated data set for object detection in video," *arXiv preprint arXiv:1702.00824*, 2017.

[16] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[17] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[18] J. Wang, Y. Yang, J. Mao, Z. Huang, C. Huang, and W. Xu, "Cnn-rnn: A unified framework for multi-label image classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2285–2294.

[19] J. Yim, H. Jung, B. Yoo, C. Choi, D. Park, and J. Kim, "Rotating your face using multi-task deep neural network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 676–684.

[20] J. Yang, S. E. Reed, M.-H. Yang, and H. Lee, "Weakly-supervised disentangling with recurrent transformations for 3d view synthesis," in *Advances in Neural Information Processing Systems*, 2015, pp. 1099–1107.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.