A Lightweight Face Recognition Network with Bootstrapping

Jianwen Ding, Jingjing Tang, Xuexin Xu, Lanliang Lin, Jiahui Liu
Department of Software Engineering,School of Information,Xiamen University

Abstract

With the development of edge computing and the use of limited resources of mobile devices, we intends to build a lightweight CNN framework using semantic bootstrap method to process noisy labels in the data set for face recognition. First, we will introduce a special case of Maxout activation named Max-feature-Map (MFM) in each convolutional layer of CNN.MFM is realized by means of competitive relation, which can not only separate noise signal and information signal, but also choose the feature between two feature graphs. In order to make the network prediction more consistent with the noisy labels, the semantic bootstrap method is used to adapt the training data set with noisy labels. With MFM, we designed three lightweight CNN networks. The experimental results show that the computational cost and storage space of Lightweight CNN can be significantly reduced, and the training data set with noisy labels can be treated effectively.

Introduction

Over the past decade, CNNs and deep learning have proven to be powerful tools in a wide range of visual analysis and recognition tasks in the field of computer vision. Many visual tasks, such as image classification, target detection and face recognition, benefit from the robustness and discriminative representation learned by CNN and obtain satisfactory results. This improvement is mainly due to the CNNs from a large number of training data to learn robust face embedding. In recent years, edge computing has gained wide attention thanks to the development of IoT. However, the limited computing resources cannot meet the requirements of the increasingly complex CNN. Therefore, it is necessary to design a lightweight CNN network. In order to achieve the optimal accuracy rate, the size of training data set has been increasing. A number of large face data sets were released, such as CASIAWebFace, CelebFaces+, VGG face, UMDFace, massive Celebrity and VGGFace2.However, accurately recognizing faces remains a challenge for several reasons. Specifically, these large data sets usually contain a large number of noise signals, especially if they are automatically collected by image search engines or movies. In order to solve the problems in face recognition, a variety of studies have been carried out(Hu et al. 2017; Wu et al. 2017). Inspired by (Wu et al. 2018), they introduced a novel architecture, max-feature-map (MFM) operation, which can obtain a compact representation and select feature filter. Our method includes MFM, small convolution filter and network, and was trained on MS-Celeb-1M dataset. The semantic bootstrap method proposed by (Wu et al. 2018) can automatically remark the training data through the pre-trained deep network, and can process the images with noisy labels. Thanks to these two techniques, we can effectively eliminate the impact of noise tags on face images and achieve the most advanced results on popular face benchmarks.

Related Work

Lightweight CNN based on face recognition

Recently, lightweight CNNs based on face recognition has been studied. PCANet(Chan et al. 2015) is a lightweight CNN, which uses PCA instead of BP algorithm to train neural network, with few parameters and calculations and small demand for samples, which is suitable for feature extraction of specific scenes. Based on PCANet, Yu et al proposed 2DPCANet(Li, Wu, and Kittler 2018), using 2DPCA to replace PCA algorithm and retained the structural information in 2-D images. MobileNet(Howard et al. 2017) is specifically used for mobile terminal and embedded deep learning. It is based on streamlined architecture and uses deep separable convolution to build lightweight CNN. Compared with MobileNet V1,MobileNet V3's(Howard et al. 2020) accuracy and speed have been greatly improved. Also, MobileNet V2 is only about 5MB in size, making it ideal for embedded and mobile devices. In addition, there are many lightweight CNNs for face recognition, such as SqueezeNet(Iandola et al. 2016), ShuffleNet(Zhang et al. 2017), which have also attracted wide attention.

Noisy Label Problems

Noisy label is an important issue in deep learning because datasets tend to be large. (Ostyakov et al. 2018) trained an ensemble of classifiers on data with noisy labels using cross-validation and used the predictions of the ensemble as soft labels for training the final classifier. Identification of incorrect labels based on prediction confidence was also shown to be highly effective in extensive experiments on face recognition by (Ding et al. 2018). (Köhler, Autenrieth, and Beluch 2019) proposed an iterative label noise filtering approach based on similar concepts as Rank Pruning, which estimates prediction uncertainty during training and relabels data samples that are likely to have incorrect labels. (Zhou et al. 2017) proposed a GAN for removing label noise from synthetic data generated to train a CNN. GANs were used to generate a training dataset with clean labels from an initial dataset with noisy labels by (Chiaroni et al. 2019). Although some strategies have been studied for noisy label problem, noisy label is still an open issue for deep learning methods.

Architecture

In this section, we first introduce Max-Feature-Map operation to obtain more information for face recognition in a lightweight CNN framework. Then we propose a semantic bootstrapping strategy to address noisy labeled images in large-scale dataset.

Max-Feature-Map

There are a lot of noises in large-scale face dataset. If these noises are not properly handled, training will be biased. The existing method of ReLU can separate noisy labels and informative labels by a threshold to determine the activation of one neuron. If the neuron is not active, its output will be 0 to eliminate the noisy labels(value$<$ 0). But, using a threshold may loss some information especially for the first several convolutional layers because both positive and negative results of these layers are useful.

We propose the Max-Feature-Map operation inspired by the concept of neural inhibition, which means when one neuron fires, the corresponding neuron will be inhibited. MFM is an alternative of ReLU to suppress the activation of a small number of neurons via competitive relationship.

We define two types of MFM operations to obtain competitive feature map. The first one is MFM2/1 operation which combines two feature maps and outputs element-wise maximum one can be written as:

$$\hat{x}_{ij}^{k} = max(x_{ij}^{k}, x_{ij}^{k+N}) \tag{1}$$

Where $x^n \in \mathbb{R}^{H\times W}$, $1 \leq n \leq 2N$, 2N means the channels of the input convolution layer, $1 \leq k \leq N$, $1 \leq i \leq H$, $1 \leq j \leq W$, W and H denote the spatial width and height of feature maps. As is shown in Eq.(1), the output $\hat{x}$ via MFM operation is in $\mathbb{R}^{H\times W\times N}$. The gradient of Eq.(1) takes the following form,

$$\frac{\partial \hat{x}_{ij}^{k}}{\partial x_{ij}^{k}} = \begin{cases} 1, if x_{ij}^{k} \geq x_{ij}^{k+N} \\ 0, \text{otherwise} \end{cases} \tag{2}$$

$$\frac{\partial \hat{x}_{ij}^{k}}{\partial x_{ij}^{k+N}} = \begin{cases} 0, if x_{ij}^{k} \geq x_{ij}^{k+N} \\ 1, \text{otherwise} \end{cases} \tag{3}$$

By using MFM2/1, we generally obtain 50% informative neurons.

The second one is MFM3/2 operation which inputs three feature maps and removes the minimal one element-wise, can be defined as:

$$\begin{cases} \hat{x}_{ij}^{k_1} = max(x_{ij}^{k}, x_{ij}^{k+N}, x_{ij}^{k+2N}) \\ \hat{x}_{ij}^{k_1} = median(x_{ij}^{k}, x_{ij}^{k+N}, x_{ij}^{k+2N}) \end{cases} \tag{4}$$

where $x^n \in \mathbb{R}^{H\times W}$, $1 \leq n \leq 3N$, $1 \leq k \leq N$ and median(.) is the median value input feature maps. The gradient of MFM3/2 is similar to Eq.(2) and Eq.(3), in which the value of gradient is 1 when the feature map x_{ij}^{k} is activated, and it is set to be 0 otherwise. In this way, we select and reserve 2/3 information from input feature maps.

MFM operation can separate the noisy signals and informative signals via inactive and active neurons. It can also suppress the activation of a small number of neurons so that MFM based CNN model are light and robust. Meanwhile, the inhibition of one neuron is free of parameters so that it does not depend on training data extensively.

The Lightweight CNN Framework

In this section, we discuss the lightweight CNN-29 model with MFM. Our model unilizes the idea of residual blocks and contains 29 layers. With the development of residual networks, deep CNNs are widely used and often achieve high performance in various computer vision tasks. The residual block in our model includes two MFM operations without batch normalization.

Compared to the traditional residual networks, our model with MFM operation has some improvements. On the one hand, when testing data domain is different from training domain, the advantages of batch normalization which accelerates the convergence of training and avoids overfitting are not obvious. On the other hand, we employ the fully connected layer instead of global average pooling layer on the top. In our training scheme, input images are all aligned, so that each node for high -level feature maps contains both semantic and spatial information which may be damaged by the global average pooling. The details of our model are presented in Table I.

Semantic Bootstrapping for Noisy Labels

Semantic bootstrapping is proposed to make prediction of the network more consistent with the noisy labels. It, called "self-training", provides an effective and simple method to estimate sample distribution. we use the CASIA-WebFace and MS-Celeb-1M dataset for training in the experiment.

Assume $x \in X$ and t denote data and labels, respectively. The CNN based on softmax loss functions can be represented as a conditional probability $P(t\,|f(x))$, $\sum_i P(t_i\,|f(x)) = 1$. The maximun probability $P(t_i\,|f(x))$ determines the most convincing prediction label. First, we train a Lightweight CNN model on the CASIA-WebFace dataset and then fine-tune it on the MS-Celeb-1M dataset. Second, we employ this model to obtain the conditional probability and label for each sample on the original noisy label MS-Celeb-1M dataset. We change the label of sample whose probability is greater than the threshold to the prediction label. The first bootstrapping is employed to select samples. We accept the sample whose prediction label is the

Table 1

Type	Filter Size /Stride,Pad	Output Size	#Params
Conv1	5×5/1,2	128×128×96	2.4K
MFM1	-	128×128×48	-
Pool1	2×2/2	64×64×48	-
Conv2_x	$\begin{bmatrix}3 \times 3/3,1 \\ 3 \times 3/1,1\end{bmatrix} \times 1$	64×64×48	82K
Conv2a	1×1/1	64×64×96	4.6K
MFM2a	-	64×64×48	-
Conv2	3×3/1,1	64×64×192	165K
MFM2	-	64×64×96	-
Pool2	2×2/2	32×32×96	-
Conv3_x	$\begin{bmatrix}3 \times 3/1,1 \\ 3 \times 3/1,1\end{bmatrix} \times 2$	32×32×96	662K
Conv3a	1×1/1	32×32×192	18K
MFM3a	-	32×32×96	-
Conv3	3×3/1,1	32×32×384	331K
MFM3	-	32×32×192	-
Pool3	2×2/2	16×16×192	-
Conv4_x	$\begin{bmatrix}3 \times 3/1,1 \\ 3 \times 3/1,1\end{bmatrix} \times 3$	16×16×192	3,981K
Conv4a	1×1/1	16×16×384	73K
MFM4a	-	16×16×192	-
Conv4	3×3/1,1	16×16×256	442K
MFM4	-	16×16×128	-
Conv5_x	$\begin{bmatrix}3 \times 3/1,1 \\ 3 \times 3/1,1\end{bmatrix} \times 4$	16×16×128	2,356K
Conv5a	1×1/1	16×16×256	32K
MFM5a	-	16×16×128	-
Conv5	3×3/1,1	16×16×256	294K
MFM5	-	16×16×128	-
Pool4	2×2/2	8×8×128	-
fc1	-	512	4,194K
MFM_fc1	-	256	-
Total	-	-	12,637K

same as the ground truth and whose label is modified to form the MS-Celeb-1M re-labeling dataset, denoted as MS-1. Third, MS-1 is used to retrain the Lightweight CNN model and then we relabel the original noisy label MS-Celeb-1M to re-sample the dataset, denoted as MS-2. It is the second bootstrapping. Finally, we retrain the Lightweight CNN model on MS-2. So far, we can obtain a model that allows large noisy label dataset to share contributions and work well.

Experiment

In this section, we evaluate our Lightweight CNN models on various face recognition tasks. And then compare our MFM operation with different activation functions. Finally, we discuss the effectiveness of the semantic bootstrapping method for selecting training dataset.

Experimental Methods and Preprocessing

The CASIA-WebFace and MS-Celeb-1M datasets are used for training, and use gray-scale face images to alleviate the influence of large illumination discrepancy when training and testing. When training, the face images are aligned to 144x144 by the five landmarks and then randomly cropped to 128x128 as inputs. Besides, each layers use BatchNormalization so that the value of each pixel falls between 0 and 1.

To train the Lightweight CNN, we randomly select one face image from each identity as the validation set and remaining images as the training set. Dropout is used for fully connect layers and the ratio is set to 0.7. The momentum is set to 0.9, and the weight decay is set to 5×10^{-4} for convolution layers and a fully-connected layer except the fc2 layers. Note that fc1 contains face representation that can be used for face verification, fc2 contains large number of parameters, but not used for feature extraction. Thus we increase the weight decay of fc2 layer to 5×10^{-3} to avoid overfitting. The learning rate is set to 1×10^{-3} initially and reduced to 5×10^{-5} gradually. The parameter initialization for convolutional layers and fully-connected layers is Xavier and Gaussian, respectively.

Multi-view Face Recognition

To further demonstrate the effectiveness of the proposed model in different domain face databases. As we know, large pose variations are one of the major factors that significantly reduce the performance of face recognition algorithms. So it is important to evaluate the effectiveness of face recognition model in large pose variations.

We compare our proposed method with multi-view face recognition methods(Kan, Shan, and Chen 2016; Yin and Liu 2017; Zhu et al. 2013, 2014), and pose-aware face image synthesis methods(Tran, Yin, and Liu 2017; Yim et al. 2015) in Multi-PIE databases(Gross et al. 2010). Note that all the compared methods are trained on Multi-PIE, while our posed method are trained on Ms-Celeb-1M (Guo et al. 2016) where the imaging condition is quite different form Multi-PIE.

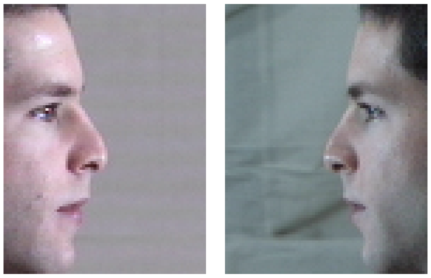
In test protocol, we follow from (Yim et al. 2015), where neural expression images from all four sessions are used. One gallery image is selected for each testing identity from their first appearance. Specifically, we selected a single frontal face image for each subject in the test dataset and treated the selected face images as the gallery set, leaving the remaining face images as the probe or test set. We then extracted the deep features using our proposed face recognition network, The rank-1 recognition accuracy is evaluated by comparing the features from faces in the probe set and those from the real frontal faces in the gallery set. The comparison was performed using the cosine distance metric. The evaluation results are given in Table 2. and compared with the competing methods.

As shown in Table. 2, our method achieves very competitive performance in ±45°, but in large pose variations (e.g. ±75° ,±90°), our face recognition network is collapsed in accuracy performance. As shown in Fig. 1, this is because

Table 2: Comparison of state-of-the-art methods in terms of recognition accuracy (%) on Multi-PIE database.

Method	±15°	±30°	±45°	±60°	±75°	±90°
Zhu et al. (Zhu et al. 2013)	90.7	80.7	64.1	45.9	-	-
Zhu et al. (Zhu et al. 2014)	92.8	83.7	72.9	60.1	-	-
Kan et al. (Kan, Shan, and Chen 2016)	100	100	90.6	85.9	-	-
Yin et al. (Yin and Liu 2017)	99.2	98.0	90.3	92.1	87.8	77.0
CPF (Yim et al. 2015)	95.0	88.5	79.9	61.9	-	-
DR-GAN (Tran, Yin, and Liu 2017)	94.0	90.1	86.2	83.2	-	-
A3FCNN (Zhang et al. 2018)	98.7	98.9	95.8	92.7	-	-
Ours	98.6	97.4	92.1	62.1	24.2	5.5
Ours + TP-GAN (Huang et al. 2017)	98.7	98.1	95.4	87.7	77.4	64.6

there is very little visible information about faces at large poses, the network can not recognize what is the identity of the test face.



(a) 75° (b) 90°

Figure 1: The 75°, 90° faces are sample form Multi-PIE dataset.

To tackle this problem, an extension module of our face recognition network has been added, e.g. (Huang et al. 2017; Luan et al. 2020; Qian, Deng, and Hu 2019; Rong, Zhang, and Lin 2020), one of these methods is TP-GAN(Huang et al. 2017) that it is a face frontalization model that can eliminate pose variations by first synthesizing a frontal view of the face from a given nonfrontal image. It can get abetter preserve the facial texture details by processing the global and local transformations separately. The evaluation results are given in Table. 2. Thanks to this module, our face recognition network can achieves higher accuracy by rotating faces. We improve the rank-1 accuracy from 24.2% to 77.4% (on ±75°) and further improve from 5.5% to 64.6% (on ±90°). The results indicate that our proposed face recognition network can efficiently capture the characteristics of different identities and obtain features invariant to pose and illumination for face recognition.

The experiments suggest that the proposed face recognition network obtain discriminative face representations and have good generalization ability for multi-view face databases.

Network Analysis

MFM operation plays an important role in our Lightweight CNN models. Hence we give a detail analysis of MFM on the Lightweight CNN-9 model in this subsection.

First, we compare the performance of MFM 2 / 1 and MFM 3 / 2 with ReLU, PReLU and ELU on the LFW dataset. We simply change activation functions and it is obvious that the output channels of ReLU, PReLU and ELU for each layer are 2× compared with MFM 2 / 1, and 1.5× compared with MFM 3 / 2. The experimental results of different activation functions are shown in Table 3, our MFM operation generally superior to the other three activation functions.

Table 3: Comparision with different activation functions on LFW verification and identification protocol by the Lightweight CNN-9 model.

Method	Accuracy(%)	Randk-1(%)	DIR@FIR=1%(%)
ReLU	98.30	88.58	67.56
PReLU	98.17	88.30	66.30
ELU	97.70	84.70	62.09
MFM 2/1	98.80	93.80	84.40
MFM 3/2	**98.83**	**94.97**	**88.59**

The reason is that MFM uses a competitive relationship rather than a threshold (or bias) to active a neuron. Since the training and testing sets are from different data sources, MFM has better generalization ability to different sources. Compared with MFM 2 / 1, MFM 3 / 2 can further improve performance, indicating that when using MFM, it would be better to keep only a small number of neurons to be inhibited so that more information can be preserved to the next convolution layer. That is, the ratio between input neurons and output neurons should be set to between 1 and 2. In addition, benefit by the MFM activation function, our Lightweight CNN model is more lightweight and can be adapted to the mobile terminal with a little modification.

Noisy Label Data Bootstrapping

In this subsection, we verify the efficiency of the proposed semantic bootstrapping method on the MS-Celeb-1M dataset. We select Lightweight CNN-9 for semantic bootstrapping. The testing is performed on LFW.

First, we train a Lightweight CNN model on the CASIA-WebFace dataset that contains 10,575 identities in total. Then, we initialize our pre-train model by CASIA-WebFace and fine-tune it on MS-Celeb-1M (contains 99,891 identities). To alleviate the difficulty of CNN convergence, we firstly set the learning rate of all the convolution layers to 0, so that the softmax loss only contributes to the last fully-connected layers to train a classifier. When it is about to converge, the learning rate of all the convolution layers is set to the same, and then the learning rate is gradually decreased from 1×10^{-3} to 1×10^{-5}.

Second, we employ the trained model in the first step to make predictions on the MS-Celeb-1M and obtain the probability $\hat{p_i}$ and label $\hat{t_i}$ for each sample $\hat{x_i} \in X$. We accept the re-labeling samples: 1) The prediction $\hat{t}$ is the same as the ground truth label t; 2) The prediction $\hat{t}$ is different from the ground truth label t, but the probability $\hat{p_i}$ is greater than the threshold. As show in Figure 2, we set threshold p_0 to [0.6, 0.7, 0.8, 0.9] to construct four re-labeling dataset. Obviously, the best performance is obtained when p_0 is set to 0.7. In this way, the MS-Celeb-1M re-labeling dataset, defined as MS-1M-1R, contains 79,077 identities.

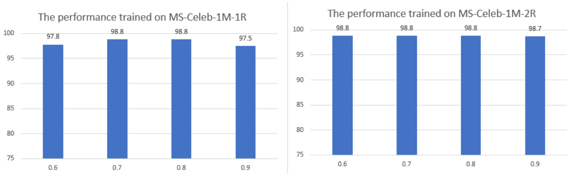


Figure 2: The performance of LFW. The models are trained on the cleaned datasets with different threshold settings to sample images.

Finally, we retrain Lightweight CNN-9 on MS-1M-2R. Table 4 shows experimental results of the CNN models learned on different subsets. We have the following observations: 1)The MS-Celeb-1M database contains massive noisy labels. If the noisy labels are correctly dealt with, the performance on the two testing datasets can be improved. Our semantic bootstrapping method provides a practical way to deal with the noisy labels on the MS-Celeb-1M database. 2) Verification performance benefits from larger datasets. 3) After two bootstrapping steps, the number of identities drops from 99,891 to 79,077 and performance improvement tends to be smaller. These indicate that our semantic bootstrapping method can obtain a purer training dataset that could in turn result in a light CNN with higher performance.

Table 4: The performance on LFW for Lightweight CNN-9 model trained on different datasets.

	Accuracy(%)	FAR=1%	FAR=0.1%
CASIA	98.13	96.73	87.13
MS-Celeb-1M	98.47	98.13	94.97
MS-1M-1R	**98.80**	98.43	95.43
MS-1M-2R	**98.80**	**98.60**	**96.77**

Conclusion

In this paper, we developed a lightweight CNN framework to learn robust facial recognition on datasets with noisy labels. Inspired by neural inhibition and maxout activation, we propose the Max-Feature-Map operation to obtain compact and low-dimensional face recognition. Small kernel sizes of convolution layers, Network in Network layers and Residual Blocks have been implemented to reduce the parameter space and improve performance. The advantages of our framework is that it is faster and smaller than other CNN methods, and contains only 12,637K parameters in the lightweight CNN-29 model. In addition, an effective semantic bootstrapping has been proposed to deal with the noise label problem. The experimental results verify that the proposed lightweight CNN framework has potential value for some real-time facial recognition systems.

References

Chan, T. H.; Jia, K.; Gao, S.; Lu, J.; Zeng, Z.; and Ma, Y. 2015. PCANet: A Simple Deep Learning Baseline for Image Classification? *IEEE Transactions on Image Processing* 24(12): 5017–5032.

Chiaroni, F.; Rahal, M.-C.; Hueber, N.; and Dufaux, F. 2019. Hallucinating A Cleanly Labeled Augmented Dataset from A Noisy Labeled Dataset Using GAN. In *2019 IEEE International Conference on Image Processing (ICIP)*, 3616–3620. IEEE.

Ding, Y.; Wang, L.; Fan, D.; and Gong, B. 2018. A semi-supervised two-stage approach to learning from noisy labels. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 1215–1224. IEEE.

Gross, R.; Matthews, I.; Cohn, J.; Kanade, T.; and Baker, S. 2010. Multi-PIE. *Image & Vision Computing* 28(5): 807–813.

Guo, Y.; Zhang, L.; Hu, Y.; He, X.; and Gao, J. 2016. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In *European conference on computer vision*, 87–102. Springer.

Howard, A.; Sandler, M.; Chen, B.; Wang, W.; Chen, L. C.; Tan, M.; Chu, G.; Vasudevan, V.; Zhu, Y.; and Pang, R. 2020. Searching for MobileNetV3. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV).*

Howard, A. G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; and Adam, H. 2017. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications .

Hu, G.; Yang, H.; Yang, Y.; Zhang, Z.; and Yang, Y. 2017. Attribute-Enhanced Face Recognition with Neural Tensor Fusion Networks. In *2017 IEEE International Conference on Computer Vision (ICCV).*

Huang, R.; Zhang, S.; Li, T.; and He, R. 2017. Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis. In *Proceedings of the IEEE International Conference on Computer Vision*, 2439–2448.

Iandola, F. N.; Han, S.; Moskewicz, M. W.; Ashraf, K.; Dally, W. J.; and Keutzer, K. 2016. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and¡ 0.5 MB model size. *arXiv preprint arXiv:1602.07360* .

Kan, M.; Shan, S.; and Chen, X. 2016. Multi-view deep network for cross-view classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4847–4855.

Köhler, J. M.; Autenrieth, M.; and Beluch, W. H. 2019. Uncertainty Based Detection and Relabeling of Noisy Image Labels. In *CVPR Workshops*, 33–37.

Li, Y. K.; Wu, X. J.; and Kittler, J. 2018. L1-(2D)2PCANet: A Deep Learning Network for Face Recognition. *Multimedia Tools and Applications* 77(4): 1–16.

Luan, X.; Geng, H.; Liu, L.; Li, W.; Zhao, Y.; and Ren, M. 2020. Geometry Structure Preserving based GAN for Multi-Pose Face Frontalization and Recognition. *IEEE Access* .

Ostyakov, P.; Logacheva, E.; Suvorov, R.; Aliev, V.; Sterkin, G.; Khomenko, O.; and Nikolenko, S. I. 2018. Label denoising with large ensembles of heterogeneous neural networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 0–0.

Qian, Y.; Deng, W.; and Hu, J. 2019. Unsupervised Face Normalization with Extreme Pose and Expression in the Wild .

Rong, C.; Zhang, X.; and Lin, Y. 2020. Feature-Improving Generative Adversarial Network for Face Frontalization. *IEEE Access* 8: 68842–68851.

Tran, L.; Yin, X.; and Liu, X. 2017. Disentangled representation learning gan for pose-invariant face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1415–1424.

Wu, X.; He, R.; Sun, Z.; and Tan, T. 2018. A Light CN-N for Deep Face Representation With Noisy Labels. *IEEE Transactions on Information Forensics and Security* .

Wu, X.; Song, L.; He, R.; and Tan, T. 2017. Coupled Deep Learning for Heterogeneous Face Recognition. *CoRR* abs/1704.02450. URL http://arxiv.org/abs/1704.02450.

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

Yin, X.; and Liu, X. 2017. Multi-task convolutional neural network for pose-invariant face recognition. *IEEE Transactions on Image Processing* 27(2): 964–975.

Zhang, X.; Zhou, X.; Lin, M.; and Sun, J. 2017. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices .

Zhang, Z.; Chen, X.; Wang, B.; Hu, G.; Zuo, W.; and Hancock, E. R. 2018. Face Frontalization Using an Appearance-Flow-Based Convolutional Neural Network. *IEEE Transactions on Image Processing* 28(5): 2187–2199.

Zhou, H.; Sun, J.; Yacoob, Y.; and Jacobs, D. W. 2017. Label denoising adversarial network (ldan) for inverse lighting of face images. *arXiv preprint arXiv:1709.01993* .

Zhu, Z.; Luo, P.; Wang, X.; and Tang, X. 2013. Deep learning identity-preserving face space. In *Proceedings of the IEEE International Conference on Computer Vision*, 113–120.

Zhu, Z.; Luo, P.; Wang, X.; and Tang, X. 2014. Multi-view perceptron: a deep model for learning face identity and view representations. *Advances in Neural Information Processing Systems* 27: 217–225.