What have we learned in Face Verification Features?

Anonymous CVPR submission

Paper ID 2467

Abstract

In this paper, we present an in-depth statistical analysis to understand face verification features. In conventional setting of face verification, a deep facial vector is extracted and it is used to compare the facial similarity between different faces via dot product. Ideally, if two faces are from the same person, the two deep facial vectors should be identical. However, due to the different environmental/personal variations, the extracted deep facial vectors are not exactly the same. In our analyses, we plot the variations of numerical values in each dimension of deep facial vectors. We found that, not all dimensions in a deep facial vector are equally important. Thus, by simply augmenting the dot product with weighted measurement, we can boost up the performance of a face verification network. Next, we investigate if there is age/gender dependent factor in a deep facial vector. Not surprisingly, such dimensions exist in a deep facial vector. This suggests that additional attribute information can help to improve face verification. Finally, we investigate if our observations about attribute can be applied to individual instance of a person, and we found that if we can provide identity information of a person, we can further boost up the performance of face verification by aggregating the natural appearances of a person over time. This paper presents the very important analyses to understand face verification features, and our suggested weighted measurement strategy is applicable to several popular face verification networks, not limited to a particular network.

1. Introduction

Face verification in unconstrained conditions has been studied extensively in recent years. Deep learning based approaches such as Deep-ID [1, 2, 3, 4], Triplet Loss [5], Light CNN [6], Center Loss [7], and Sphere Face [8], have demonstrated excellence performance suppressing human performance in face verification. A common strategy among them is to train a deep neural network to extract high level facial feature vectors which will be used to measure similarity between different faces, as illustrated in Fig. 1.

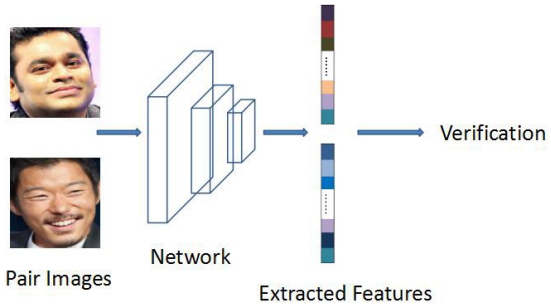

Figure 1: Given a pair of face images, a face verification network extracts deep facial features and compares the similarity between the two features to verify if the two faces are of the same person.

These facial features are trained in an end-to-end manner, and it is a non-trivial task to understand what features have been learned, and how robust these trained features are.

In this paper, we analyze the distributions of variations of numerical values in each dimension of facial feature vector. Through our analyses, we found that the plotted distributions in each dimension are non-uniform, this suggests that some dimensions in a facial feature vector are more reliable than the others. This observation is valid across all different facial feature vectors extracted by different face verification networks [6, 7, 8]. We cluster the face images according to the numerical values in the dimension which has the largest variation, and notice that these variations are mainly caused by lighting and pose variations. In other words, the similarity measured by the extracted facial feature vectors can be affected by lighting and pose variations.

In our second experiment, we cluster the face images according to gender and age, and we perform the same statistical analyses to evaluate the distributions of mean of numerical values in each dimension of facial feature vectors in each group. Ideally, if the extracted facial feature vectors are unaffected by gender and age, the distributions in each small group should be identical. However, we found some dimensions have large mean value across different groups. The average face within each group reveals that these dimensions correspond to gender or age dependent features.

Finally, we cluster the face images according to identity. We notice that each person has different distribution. Thus, the appearance variations across different people are not identical which cause the different distributions.

From our analyses, we notice that environmental-dependent, attribute-dependent, and identity-dependent features appear in the extracted facial feature vectors. These features are encoded in different dimensions of the facial feature vector. Our observations are useful. It not only helps to understand the extracted facial features, but it can also be used to enhance performance of face verification. We present two weighting strategies which give larger weights to the dimensions with smaller variation and difference as these dimensions are more stable than the others. Our experimental results show that our proposed weighted measurement of similarity can consistently improve the performance of face verification networks despite these networks are trained differently using different architectures and objective functions. Finally, we present applications which applies our identity aware weighting strategy to M-to-N face retrieval, and video face clustering.

2. Related Works

Representative works in face verification, and face set recognition are reviewed in this section.

Face Verification In face verification, the LFW dataset [9] has been widely used to evaluate the performance of face verification algorithms. Since the usage of deep learning, DeepFace [10] was the first deep learning based approach which achieves human level performance in face verification. Since then, the Deep-ID series [1, 2, 3, 4], the FaceNet [5], the light CNN [6], and the follow-up works [7, 8, 11, 12, 13, 14, 15, 16] have demonstrated excellent performance in the LFW dataset suppressing human level performance. However, the problem of 1-to-1 face verification is still challenging.

The recently developed datasets such as MegaFace [17], IJBA [18] and MS-Celeb1M [19] show that, even for an algorithm which achieves more than 99% accuracy in the LFW dataset, their performances are still far from satisfactory in practice. This is because the environmental variations, such as lighting, viewing angles, facial expression variations, such as smiling, laughing, age variations, body-status variations, health conditions, as well as facial decoration such as make-up, all of these factors would affect the appearances of a person. Ideally, the face verification networks should be able to extract an identity features which are robust to the aforementioned factors. Some works have been designed for specific task, such as age-invariant [20, 21], illumination [22], pose [23], low-resolution[24]. However, as we will demonstrate in our analyses, none of the state-of-the-art methods can completely separate the identity features from these factors. Thus, instead of re-training a new face verification network, we propose a remedy which can reduce the effects of unstable variations in the extracted facial feature vectors. Consequently, even without re-training a new network, we can boost-up the performance of a face verification network. Our remedy is general, and it is applicable to all existing networks.

Face Set Recognition In face set recognition, instead of the 1-to-1 face verification, the face set recognition considers multiple images of the same persons for recognition. For deep learning based models, Rao *et al.* [25] proposed a aggregation network to directly aggregates raw video frames. Yang *et al.* [26] proposed an attention model to find the weights of features. The insight is that different frames in a video should not contribute equally to the recognition task. Thus, they proposed a method to learn different weights for different frame features. Inspired by [26], Rao *et al.* [27] also tried to learn different weights, and their method learned to discard misleading frames which achieves better results. Some other methods use pairwise frame feature similarity [5, 10] or frame feature pooling [28, 29] for video face recognition. Hayat *et al.* [30] introduced deep reconstruction model to automatically discover the underlying geometric structure. Unlike these methods, our work focuses on understanding face verification features and our proposed remedy is general, and it is also applicable for face set recognition. For face set recognition, our approach learns an identity-aware weight vector which can further boost up the verification accuracy beyond our remedy in the 1-to-1 face verification task.

3. Analyses of Face Verification Features

In this section, we analyze the verification features with global, attribute and instance level variations. We collect several state-of-the-art face verification models, and use them to extract the face verification features and compute the mean value or variations in each dimension of all the images in CACD dataset [31]. The CACD dataset is chosen because it contains multiple faces of the same person at different age, which allows us to perform attribute analyses on age. For the verification features, we choose the Light CNN-4, Light CNN-9, Light CNN-29 [6], Sphereface [8] and Center Loss [5, 32]. These face verification models are chosen because their network architectures and trained model are publicly available. Also, their performance represents the state-of-the-art face verification algorithms.

3.1. Global Feature Variations

In Fig. 2, we show the variations of numerical values in each dimension of the feature vectors of all the face images in the CACD dataset. The facial feature vectors are extracted using the selected face verification networks. Ide-

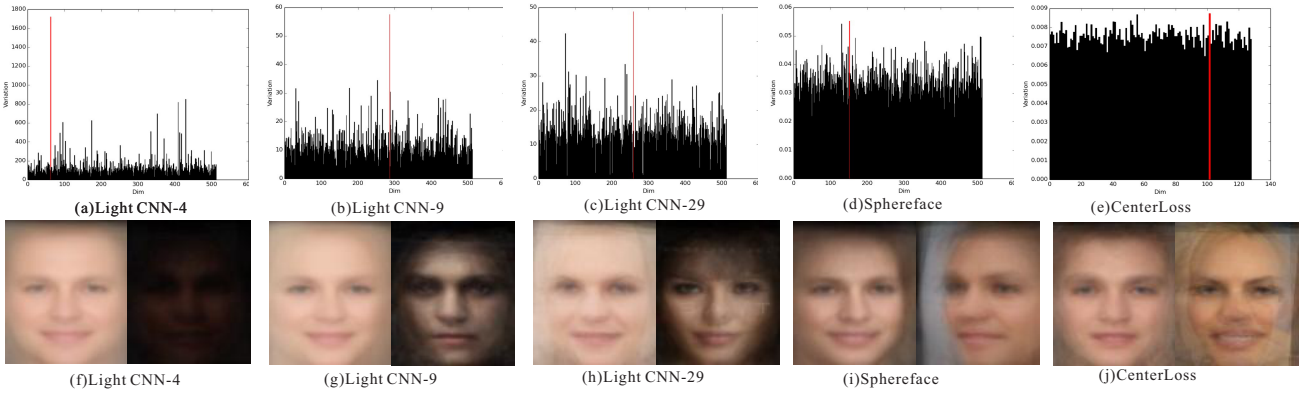

Figure 2: First row shows the plots of variations of each dimension of different face verification models; the second row shows the average faces obtained according to the numerical values in the dimension which has the largest variation (highlighted with red line in the plots): the left average face represents faces which have the largest numerical value in the selected dimension, and the right face represents faces which have the smallest numerical value in the selected dimension respectively.

ally, we hope the plot of variations to be uniformly distributed because we are using equal weight on each dimension when we compare the similarity between two facial vectors. However, as shown in the plots, there are some dimensions that have very large variations. In order to find out what features have been captured in these dimension, we cluster the face images according to the numerical values in the dimension which has the largest variation, and compute the average faces of these clustered faces, as shown at the second row of Fig. 2. From the average faces, we notice that these variations are mainly caused by lighting and pose variations. In particular, for Light CNN-4, Light CNN-9 and Light CNN-29, the variations are caused by lighting variations. For Sphereface, the head pose orientation dominates the variations, and for CenterLoss, the facial expression causes the largest variations. This experiment reveals that face verification features are sensitive to environmental changes, e.g. light CNN are more sensitive to lighting changes, Sphereface is more sensitive to head pose orientation, and CenterLoss is more sensitive to facial expression. Note that these variations are not related to identity.

One interesting observation is that, for the face verification networks that achieve higher accuracy, the distributions of the plots tend to be more uniform. For instance, the Center Loss has achieve 99.28% veification accuracy in LFW, and 76.52% accuracy in MegaFace (VR@FAR10e-6), while the Light CNN-4 achieve 97.77% accuracy in LFW, and 62.34% accuracy in MegaFace. This observation agrees with our expectation that an ideal face verification network should produce a uniform distribution in the above analyses.

3.2. Attribute Dependent Analyses

Next, we try to explore if there are attribute dependent factors in the face verification features. We choose gender and age as our study attributes. We divide the CACD dataset into two separate groups for gender, and another two separate groups for age. Because of page limitation, we report our analyses about gender in Fig. 3, and our analyses about age can be found in the supplementary materials.

We compute the average numerical value of both subgroups in each dimension (first and second row in Fig. 3). Then, we compute the differences between the two plots as shown in the third row in Fig. 3. Similar to our study in global variations about facial feature vector, we cluster the face images according the the numerical value of the selected dimension which has the larges difference between the male and female group. The average faces reveal that the selected dimension corresponds to gender information. It is obvious that one of the average face is a man's face, and the other one is a woman's face.

From Fig. 3, we notice that there are some clear differences between the man and woman features. More importantly, we can find the dimensions which are highly related to gender. Note that, the face verification features were all trained in end-to-end manner. There is no guidance to group the dimension of facial features according to gender, age, or other attributes. From the last row of Fig. 3, we can clearly distinguish the gender of the bottom and top-value average faces. It makes sense because gender is also part of the identity features. However, with the known dimensions that are highly correlated to gender, we can improve the face verification accuracy by altering the similarity measurement if the gender/age of a person is given. In real world applications, the gender/age information can be obtained from

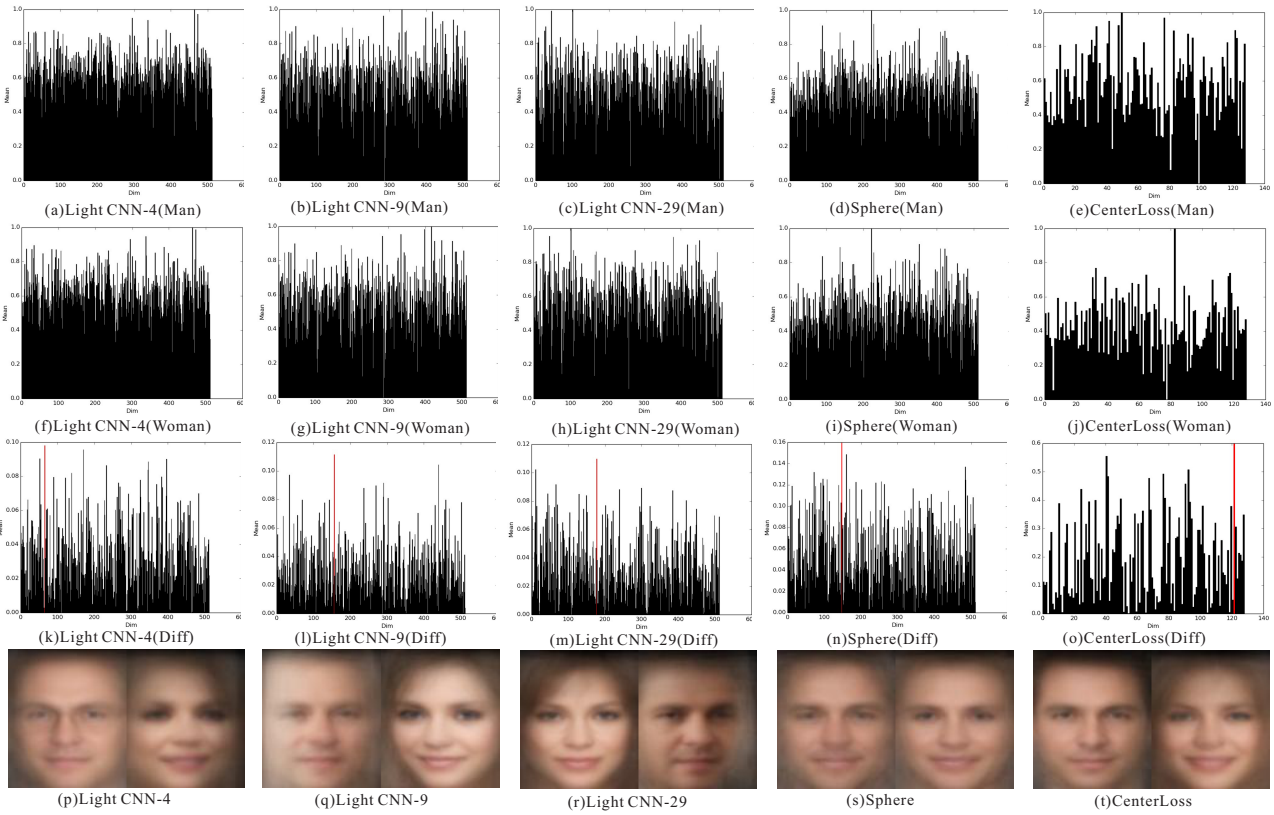

Figure 3: The analyses of features related to gender through five different face verification networks. The first row is the average features of man set, and the second row is for woman. The third row is the difference of features between man and woman set. The fourth row is average faces for top and bottom 15% faces according to largest different dimension.

other channels such as identity card or driving license.

3.3. Feature analyses of different people

In this section, we collect five thousand images of five different people, Barack Obama, Donald Trump, Hillary Clinton, Ivanka Trump and Michelle Obama. These five different people have different age, gender and color. Each one has around one thousand images. We also compute the average features and variations in each dimension of different people. All the features in this section are extracted by the Sphereface network. Not surprisingly, the average features and variations are quite different from each other.

There are several observations in Fig. 4. First, the average features are very different from each other, which is normal since they are different people. However, what is more important is that, the feature variations within the same person are also very different from each other. This implies that the feature vector variations across faces of different people are also different. To better visualize the differences, we cluster the face images according to the dimension which has the largest variations. The differences between the bottom and the top-value average faces are obvious. Notice that, for each pair of average faces, they show different facial expressions. These facial expressions are the most common facial expressions that we can found on the internet. From the above analyses, we can conclude that the inter-person variations are mainly caused by different frequent of facial expressions in their daily life. These observations motivate us to introduce an identity-aware face verification method, which will be described in the next section.

4. Improving Face Verification Accuracy

In this section, we present our weighting strategies to enhance the accuracy of face verification networks. Our weighting strategies are based on statistical approach, and they do not require any re-training or fine-tuning of the networks.

4.1. Our Weighting Strategies

From the above analyses, we notice that the distributions of variations of face verification vectors are far from the ideal distribution. In particular, the lighting environment, face pose, expressions, as well as interpersonal variations,

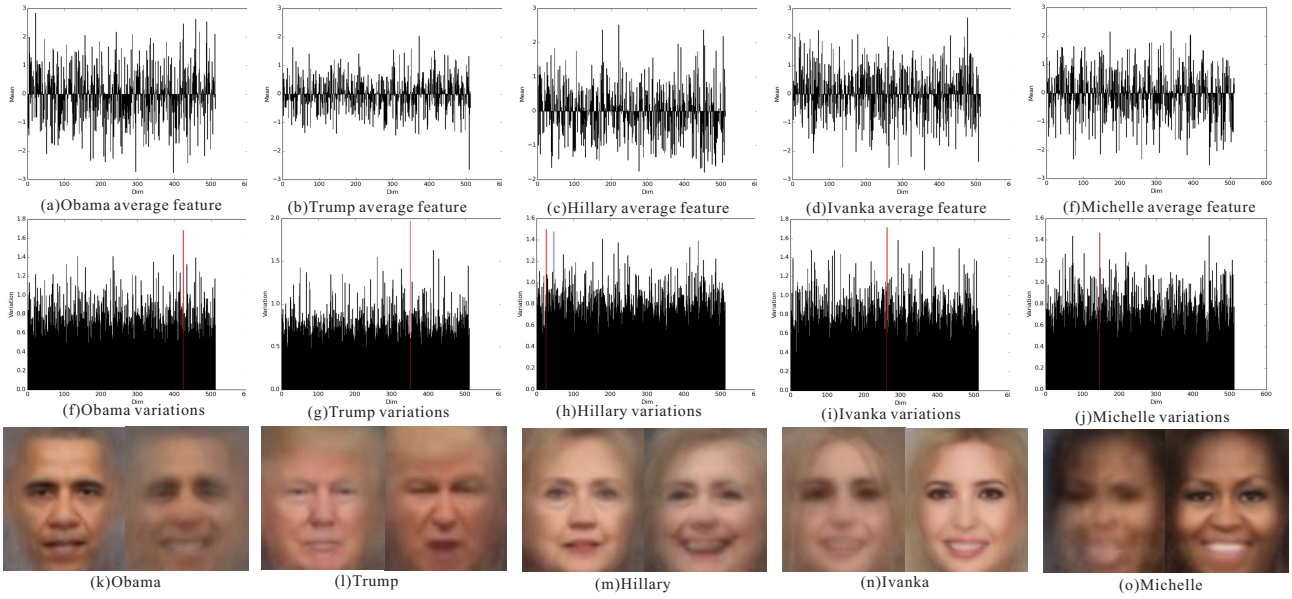

Figure 4: The analyses of features related to identity through five different celebrities. The first row is the average features for five different people, and the second row is variations for them in each dimension. The third row is average faces for top and bottom 15% faces according to largest different dimension.

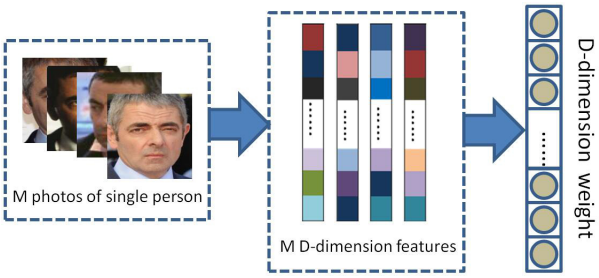

Figure 5: Our weighting strategy for face verification. From a set of faces, we compute a weight vector according to the distribution of extracted facial feature vector. During verification, the weight vector is used to normalize the value of each dimension to achieve weighted measurement. When all face images are from the same person, we obtain identity aware weight vector. The identity aware weight vector can be used in the M-to-N face retrieval problem.

all of these factors affect the face similarity measurements. In order to reduce the effects of "irrelevant" dimensions, we propose the following remedy solutions which compute weighting vectors to normalize the similarity measurement.

The framework of our weighting strategies are presented in Fig. 5. Our network takes a set of face images as input and output a weighting feature mask. The weighted features could be used to enhance face verification accuracy.

Strategy 1: ***Weighting for General Face Verification*** In the above analyses, we notice that the feature dimensions which have large variations are mostly caused by environmental and personal factors that are less related to the identity of a person. Thus, we propose to introduce a weighted measurement which penalizes the feature dimensions which have large variations evaluated by a large validate dataset. Note that this measurement is a global statistics, which is applicable to all face images without knowing the attribute or identity of query face images. Our weighting mask is defined as follow:

$$w_q = \frac{1}{1 + \eta \cdot var_q} \tag{1}$$

where w_q is the weighting mask for the q-th dimension of a facial feature vector, var_q is the variance of numerical values of the q-th dimsnsion, and η is a parameter which controls the decay of w_q with respect to the variance.

In some application scenarios, the attributes or even identity of query face images can be obtained as auxiliary verification information. For example, when comparing the face image of a person captured by a handheld device with the face photo on his/her identity card or driving license, the gender or age information can be retrieved. In face retrieval, the identity of query face image can be a known parameter especially for celebrities. In order to cover with the above application scenarios, we introduce attribute-aware / identity-aware weighting strategies. We modify Eq. 1 as

follow:

$$w_q^g = \frac{1}{1 + \eta \cdot var_q^g} \quad (2)$$

where g is the group index of query face image. g can be $0, 1$ for gender, or $0, 1, ..., n$ for n age group, or even for n face identity group. With our groupwise weighting strategies, the variations within each group could be considered, which is more accurate measurement comparing with the group statistical measurement in Eq. 1.

Strategy 2: ***Weighting for identity-aware face verification*** The weighting strategy in Eq. 2 is based on a heuristic definition. In this second strategy, we aim to estimate the optimal weighting within each identity group by considering pairwise distance between faces within the same group. Mathematically, we want to find the optimal weights which minimizes:

$$\arg\min_{w_q^g} \sum_{(i,j)\in g} \sum_q ||w_q^g \cdot x_{qi}^g - w_q^g \cdot x_{qj}^g||_2^2 + \alpha||w_q^g||_2^2 \quad (3)$$

where $\sum w_q^g = d$, d is the total dimension of the face feature vector, $(i, j) \in g$ represents face images within the same group, and x_{qi}^g indicates the corresponding feature value.

After some rearrangement, we obtain the close-solution:

$$w_q^g = \frac{\lambda_g}{\sum\limits_{(i,j)} (x_{qi}^g - x_{qj}^g)^2 + \alpha} \quad (4)$$

where,

$$\lambda_g = \frac{2d}{\sum\limits_{q=1}^{d} \frac{1}{\sum\limits_{(i,j)} (x_{qi}^g - x_{qj}^g)^2 + \alpha}} \quad (5)$$

From Eq. 4, if the selected pairs within the same group differ a lot in dimension q, the assigned weight on dimension q would be small. This is coincident with our assumption that the large variation dimension should be restrained.

During testing, we assume the attribute group or identity group of the query face image is known, and use the corresponding weighting of the group to measure the similarity between faces.

4.2. Verification results on LFW

In this section, we apply our weighting strategies to several popular face verification methods, and then report verification accuracy on LFW [9]. The baseline models and LFW features are directly picked from their released Github.

Different from conventional 1-to-1 face verification, we utilize the additional global, gender or age information. Note that, getting these information is much easier than training a new model in some specific scenarios. For example, we can get the gender information from identity card or driving license. Also, if we have multiple photos of the same person, we can obtain the identity aware information to estimate the identity weight. During the weight calculating of instance, we exclude all the images used in the same-identity pairs to make the verification results fair.

From Table 1, we can see that our weighting strategies improve the accuracy in every model. The identity-aware face verification achieves the best performance since the weighted measurement better captures the useful information in the face verification features.

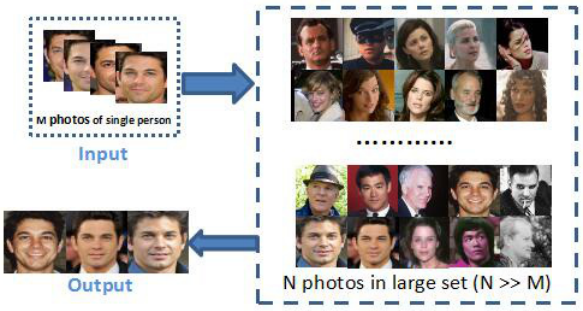

Figure 6: The M-to-N face retrieval task. Given M images of a certain person, we find all the images of this person in large dataset.

5. M-to-N Face Retrieval Application

In this section, we introduce the M-to-N face retrieval task, which is an excellent application of identity-aware face verification. Given M photos of a person, our goal is to retrieve all of the face images in a dataset which contains the target person. This problem is important, and has many applications in face tagging, video face clustering and surveillance. In the conventional 1-to-N face retrieval, there is only one photo of the target person, and the problem is reduced to a 1-to-1 face verification problem by comparing the target face with all of the face images in the dataset. In M-to-N face retrieval, we have additional M face images of the target person. Although the M photos are of the same target person, their face verification features are not strictly identical. Thus, we can apply our weighted face verification features which give larger weights to the feature dimensions that are more stable across the M face images. As demonstrated in our analyses and the verification experiments in the LFW dataset, the more stable is the numerical value in a feature dimension, the more reliable is the feature dimension to represent the target person.

5.1. Retrieval Performance Measures

We experiment the M-to-N face retrieval on LFW, YTF, and CACD face datasets. For each face dataset, we divide it

Model	Light CNN-4	Light CNN-9	Light CNN-29	Sphere (20)	CenterLoss
Baseline	97.77	98.22	98.90	99.30	99.28
Baseline + S1 (global)	97.88	98.25	98.92	99.30	99.32
Baseline + S1 (gender)	97.87	98.22	98.92	99.30	99.30
Baseline + S1 (instance)	98.02	98.47	99.17	99.43	99.38
Baseline + S2 (instance)	98.07	98.45	99.17	99.45	99.38

Table 1: Verification accuracy on LFW. In this paper, we use CenterLoss to stand for CenterLoss + Softmax for simple.

into two mutual exclusion parts: probe set and gallery set. each person in the probe set have more than one photo, and we learn the weight online. We evaluate the retrieval performance using mean Average Precision (mAP) (%). The mAP measure is widely used in image search applications.

For image retrieve task, query with multiple faces should be more accurate than a single face image. For evaluation, we compute the distance between each image in the probe set and the images in the gallery set. And applying the average distances of each image in the probe set for this person to rank the gallery set to computer mAP. In this way, we can reduce the bias caused by random sampling.

5.2. Experimental Setup

In our experiment, a pretrained model of CNN network is downloaded from [32]. The CNN architecture is Inception ResNet v1, under the supervision of Softmax + Center loss. During the training of CNN, the faces are detected and aligned by MTCNN [33]. The input image size is 160x160 and the dimension of output feature is 128. After training the CNN, the network is fixed and we focus on the effectiveness of our weighting strategies to enhance the performance of our trained network for face retrieval. In our evaluation, if a person in the dataset has less than m images, we exclude this person in the evaluation. In our evaluation, we record performances with different value of m. We compute the average mAP of the images in the retrieval set. For reference, we also report the results when we apply the face retrieval using *all* query face images (calculating average distance as described in Sec 5.1). Using *all* query face images could be seen as an voting mechanism, which enhances the performance significantly, but with a major drawback that the retrieval time is multiple times slower than using a single image with weighted feature for face retrieval.

5.3. Experiments On LFW dataset

We evaluate our approach on LFW [34]. The LFW contains 13,233 images of 5,749 people and it is widely used to evaluate the discriminant ability of face features. We apply the face retrieval task on LFW and the results are reported in Table 2.

Our weighting strategies show their effectiveness on all of the image retrieval tasks, including retrieval with single image (e.g. from 97.65% to 98.71%) and retrieval with multiple images (e.g. from 99.48% to 99.70%).

m	CL	CL + S1	CL + S2
5	97.65	98.39	98.71
10	98.10	98.58	98.78
15	98.24	98.72	99.20
5(s)	99.48	99.61	99.70
10(s)	99.73	99.42	99.80
15(s)	99.63	99.78	99.90

Table 2: Face Retrieval in LFW. m represents the number of query images of each person. The first 3 rows is the average mAP (retrieval with only one image each time); the last 3 rows retrieves with all images in the set, and s stands for 'set' (the same as follows). CL, S1 and S2 stands for CenterLoss, Strategy 1 and Strategy 2 respectively.

m	CL	CL + S1	CL + S2
10	83.17	90.13	91.73
20	81.32	90.81	92.52
30	81.16	91.05	92.84
10(s)	86.26	96.63	94.88
20(s)	96.86	97.34	97.44
30(s)	97.09	97.67	97.95

Table 3: Face Retrieval in YTF (mAP). m represents the number of images from each video.

5.4. Experiments On YTF dataset

We test our algorithm on the YouTube Face (YTF) [35]. The YouTube Face is designed for face verification in videos, and it contains 3425 videos of 1595 different people and the video lengths vary from 48 to 6070 frames with an average lengths of 181 frames. The results are reported in Table 3.

From Table 3 we can see that the our strategies improve the performance greatly on both image retrieval with single image (e.g. from 81.16% to 92.84%) and retrieval with multiple images (e.g. from 86.26% to 94.88%). Also we find that the improvement is more obvious in retrieval with

m	CL	CL + S1	CL + S2
10	67.82	69.84	70.21
20	67.64	69.31	69.51
40	67.54	69.03	69.10
10(s)	81.01	81.76	81.38
20(s)	81.10	82.48	82.88
40(s)	81.40	82.36	82.43

Table 4: Face Retrieval in CACD.

m	CL	CL + S1	CL + S2
5	82.58	85.61	85.67
10	85.32	86.74	86.91
20	85.95	86.36	86.37
5(s)	89.63	89.95	89.63
10(s)	90.36	90.36	90.36
20(s)	89.71	89.95	90.01

Table 5: Face Retrieval in Big Bang Theory.

single image task. Compared with larger set, our improvement in smaller set is more impressive. Notice that retrieval with single image is much faster than retrieval with multiple images, our strategies make large contribution to real-time retrieval.

5.5. Experiments On CACD dataset

The Cross-Age Celebrity Dataset (CACD) [31] contains 163,446 images from 2,000 celebrities from the Internet. Each celebrity has faces across 10 years. From Table 4 we can see that cross age retrieval is difficult and our weighting strategies can still improve the performance of face retrieval.

In the experiment, we use the face images of young people to retrieve the images with old age. This task is challenging and valuable, and our strategy also bring some significant improvements (e.g. from 67.82% to 70.21%).

5.6. Face Retrieval in Big Bang Theory

Besides using the face dataset for evaluations, we are also interested in a real world application. In this task, we would like to retrieve the 5 leading characters in the Big Bang Theory. In order to collect our query images, we download multiple images of the 5 leading characters of Big Bang from Internet, and then use these images to retrieve their faces in the Big Bang Theory dataset introduced by [36]. The Big Bang Theory dataset contains 186,717 images of these leading characters. This task is very challenge. Because the flames in the Big Bang Theory Dataset are not aligned, and many of them are with low resolution, large pose or even partial lost.

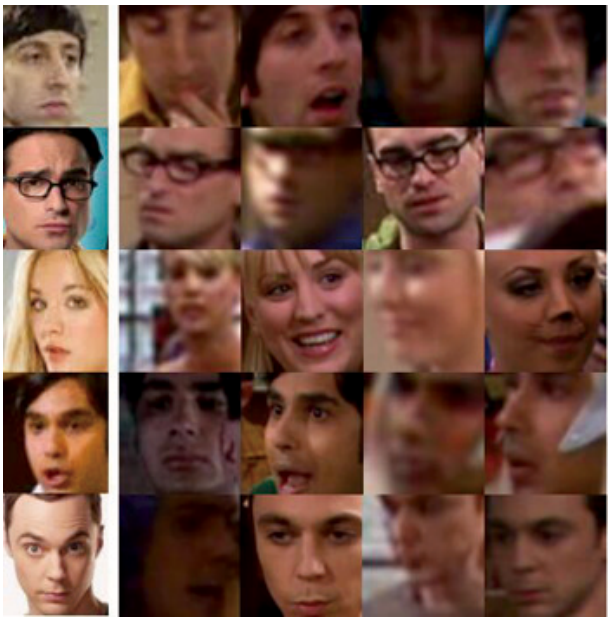

Figure 7: The big-bang retrieval task with our weighting Strategies. The first column is the query image. The second to fifth column are some hard samples that could be easily retrieved with our weighting strategy.

Fig 7 shows some hard samples. These images are hard to retrieve for several reasons, not aligned, large pose and illumination variation or low resolution. Our identity weighting masks proposed in this paper have tried to restrain these kind of variations, and their effectiveness are validated in Big Bang Theory dataset.

6. Conclusion

In this paper, we have analyzed the verification features of several state-of-the-art CNN face verification networks. In our analyses, we notice that some feature dimensions are highly correlated with environmental-dependent factors such as illumination or pose, and attribute-dependent factors such as age or gender. We have also noticed that the variations of face verification vector are identity-aware. In order to enhance the performance of face verification. We propose two weighting strategies to suppress influence of less relevant feature dimensions. In our experiments, we show that if we know the distributions of face verification vectors, we can further improve the verification accuracy of a well-trained network without re-training or fine-tuning. Also, by using different weights for different group, we can achieve attribute-aware, and identity-aware face verification. In the M-to-N face retrieval task, the identity aware weighting strategies have improved the accuracy with a large margin in several famous datasets.

References

[1] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1891–1898, 2014. 1, 2

[2] Y. Sun, Y. Chen, X. Wang, and X. Tang, "Deep learning face representation by joint identification-verification," in *Advances in neural information processing systems*, pp. 1988–1996, 2014. 1, 2

[3] Y. Sun, X. Wang, and X. Tang, "Deeply learned face representations are sparse, selective, and robust," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2892–2900, 2015. 1, 2

[4] Y. Sun, D. Liang, X. Wang, and X. Tang, "Deepid3: Face recognition with very deep neural networks," *arXiv preprint arXiv:1502.00873*, 2015. 1, 2

[5] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 815–823, 2015. 1, 2

[6] X. Wu, R. He, Z. Sun, and T. Tan, "A light cnn for deep face representation with noisy labels," *Computer Science*, 2017. 1, 2

[7] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," in *European Conference on Computer Vision*, pp. 499–515, Springer, 2016. 1, 2

[8] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "Sphereface: Deep hypersphere embedding for face recognition," 2017. 1, 2

[9] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," tech. rep., Technical Report 07-49, University of Massachusetts, Amherst, 2007. 2, 6

[10] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1701–1708, 2014. 2

[11] X. Zhang, Z. Fang, Y. Wen, Z. Li, and Y. Qiao, "Range loss for deep face recognition with long-tail," 2016. 2

[12] W. Liu, Y. Wen, Z. Yu, and M. Yang, "Large-margin softmax loss for convolutional neural networks," in *International Conference on International Conference on Machine Learning*, pp. 507–516, 2016. 2

[13] Y. Zhong, J. Chen, and B. Huang, "Towards end-to-end face recognition through alignment learning," *arXiv preprint arXiv:1701.07174*, 2017. 2

[14] J. Deng, Y. Zhou, and S. Zafeiriou, "Marginal loss for deep face recognition," in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition (CVPRW), Faces in-the-wild Workshop/Challenge*, vol. 4, 2017. 2

[15] X. Yin and X. Liu, "Multi-task convolutional neural network for face recognition," *arXiv preprint arXiv:1702.04710*, 2017. 2

[16] F. Wang, X. Xiang, J. Cheng, and A. L. Yuille, "Normface: L2 hypersphere embedding for face verification," 2017. 2

[17] I. Kemelmacher-Shlizerman, S. M. Seitz, D. Miller, and E. Brossard, "The megaface benchmark: 1 million faces for recognition at scale," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4873–4882, 2016. 2

[18] B. F. Klare, B. Klein, E. Taborsky, A. Blanton, J. Cheney, K. Allen, P. Grother, A. Mah, and A. K. Jain, "Pushing the frontiers of unconstrained face detection and recognition: Iarpa janus benchmark a," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1931–1939, 2015. 2

[19] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, "Ms-celeb-1m: A dataset and benchmark for large-scale face recognition," in *European Conference on Computer Vision*, pp. 87–102, Springer, 2016. 2

[20] Y. Wen, Z. Li, and Y. Qiao, "Latent factor guided convolutional neural networks for age-invariant face recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4893–4901, 2016. 2

[21] L. Du and H. Ling, "Cross-age face verification by coordinating with cross-face age verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2329–2338, 2015. 2

[22] H. Han, S. Shan, X. Chen, and W. Gao, "A comparative study on illumination preprocessing in face recognition," *Pattern Recognition*, vol. 46, no. 6, pp. 1691–1699, 2013. 2

[23] W. AbdAlmageed, Y. Wu, S. Rawls, S. Harel, T. Hassner, I. Masi, J. Choi, J. Lekust, J. Kim, P. Natarajan, *et al.*, "Face recognition using deep multi-pose representations," in *Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on*, pp. 1–9, IEEE, 2016. 2

[24] S. P. Mudunuri and S. Biswas, "Low resolution face recognition across variations in pose and illumination," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 5, pp. 1034–1040, 2016. 2

[25] Y. Rao, J. Lin, J. Lu, and J. Zhou, "Learning discriminative aggregation network for video-based face recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3781–3790, 2017. 2

[26] J. Yang, P. Ren, D. Zhang, D. Chen, F. Wen, H. Li, and G. Hua, "Neural aggregation network for video face recognition," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. 2

[27] Y. Rao, J. Lu, and J. Zhou, "Attention-aware deep reinforcement learning for video face recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3931–3940, 2017. 2

[28] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *British Machine Vision Conference*, pp. 41.1–41.12, 2015. 2

[29] A. Roychowdhury, T. Y. Lin, S. Maji, and E. Learned-Miller, "One-to-many face recognition with bilinear cnns," pp. 1–9, 2015. 2

[30] M. Hayat, M. Bennamoun, and S. An, "Deep reconstruction models for image set classification," *Pattern Analysis Machine Intelligence IEEE Transactions on*, vol. 37, no. 4, p. 713, 2015. 2

[31] B. C. Chen, C. S. Chen, and W. H. Hsu, "Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset," *IEEE Transactions on Multimedia*, vol. 17, no. 6, pp. 804–815, 2015. 2, 8

[32] D. Sandberg, "Face recognition using tensorflow." https://github.com/davidsandberg/facenet/. 2, 7

[33] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, 2016. 7

[34] G. B. H. E. Learned-Miller, "Labeled faces in the wild: Updates and new reporting procedures," Tech. Rep. UM-CS-2014-003, University of Massachusetts, Amherst, May 2014. 7

[35] L. Wolf, T. Hassner, and I. Maoz, "Face recognition in unconstrained videos with matched background similarity," in *Computer Vision and Pattern Recognition*, pp. 529–534, 2011. 7

[36] Y. Hu, J. S. Ren, J. Dai, C. Yuan, L. Xu, and W. Wang, "Deep Multimodal Speaker Naming," in *Proceedings of the 23rd Annual ACM International Conference on Multimedia*, pp. 1107–1110, ACM, 2015. 8