

Locally Rejected Metric Learning Based False Positives Filtering for Face Detection

Nanhai Zhang^(✉), Jiajie Han, Jiani Hu, and Weihong Deng

Beijing University of Posts and Telecommunications, Beijing, China
{nhzhang,dxs,jnhu,whdeng}@bupt.edu.cn
<http://www.bupt.edu.cn/>

Abstract. Face detection in the wild needs to deal with various challenging conditions, which often leads to the situation where intraclass difference of faces exceeds interclass difference between faces and non-faces. Based on this observation, in this paper we propose a locally rejected metric learning (LRML) based false positives filtering method. We firstly learn some prototype faces with affinity propagation clustering algorithm, and then apply locally rejected metric learning to seek a linear transformation to reduce the differences between each face and prototype faces while enlarging the differences between non-faces and prototype faces and preserving the distribution of learned prototype faces with locally rejected term. With the learned transformation, data are mapped into a new domain where face can be exactly detected. Results on FDDB and a self-collected dataset indicate our method is better than Viola-Jones face detectors. And the combination of the two methods shows an improvement in face detection.

Keywords: Locally rejected · Face detection · Prototype faces

1 Introduction

Great success of face detection in constrained environments have been made over past years. However it is still a challenge to detect face in wild environments, due to various variations like lighting, illumination, expression and occlusion.

From the view of metric learning, failure of face detection results from that intraclass distance (between faces) may be larger than interclass distance (between faces and complex background). Figure 1 shows this situation. To deal with various challenging variations, researchers have proposed diverse approaches. For the feature-based methods, researchers try to extract robust feature, e.g. SURF [1], or feature learned by CNN [2]. For the model-based methods, researchers try to model large variations with deformable part-based model [3]. Nevertheless, most of these approaches are time consuming or have expensive computation.

To deal with various challenging variations, borrowing the idea of metric learning, we can map the feature from the origin domain into a new domain where differences of interclass are larger than the differences of intra-class. Therefore,



Fig. 1. Example illustrates the situation where intraclass distance exceeds interclass distance. *Left* is a face image, *middle* is a average face¹ and *right* is a non-face image. The cosine similarity between the face image and the average face is 0.37 which is lower than 0.62 that is the cosine similarity between the non-face image and the average face.

we propose a new false positives filtering approach for face detection based on locally rejected metric learning and affinity propagation clustering algorithm, which can efficiently improve the results of classical Viola-Jones detector [4]. Since the training process can be done off-line and the predication procedure is simple, the whole procedure of face detection can be done fast and efficiently.

In summary, the contributions of this paper are threefold:

- (1) We propose to use affinity propagation clustering algorithm to learn some prototype faces being more representative than average faces.
- (2) We propose a robust framework of false positives filtering for face detection based on locally rejected metric learning by reducing the intraclass differences while enlarging the interclass differences and preserving the distribution of prototype faces with the locally rejected term.
- (3) We evaluate our approach on self-collected dataset and FDDB [5], both of which are collected from unconstrained conditions. Results on the two datasets show an efficient improvement of Viola-Jones detector.

2 Related Work

Clustering: The goal of clustering analysis is to mine the underlying structure of an unlabeled dataset. Most of clustering algorithms can be broadly categorized into three groups: partitioning based clustering, such as K-means [6], graph based clustering, such as spectral clustering [7] and density based clustering, such as mean-shift [8]. K-means proposed in 1955 is still a popular algorithm. Frey [9] proposed a powerful cluster algorithm with the ability of discovering representative faces from a gallery of face images dataset in 2007. With this method, we can learn some prototype faces from real world face datasets.

Face Detection: The seminal work of face detection was done by Viola and Jones [4], which has become a standard paradigm for face detection. Recently, Chen [10] proposes a new framework for cascade face detection with the help of aligned shape indexed feature. Li [2] proposes a CNN cascade based face detection. As far as we know, few papers discuss the post process for face detection

¹ <http://faceresearch.org/>.

except Chen’s work [10]. It introduces a simple SVM classifier for post process to prove facial point based features can improve face detector performance. Different from it, our false positives filtering approach is more universal and faster.

Metric Learning: Since the pioneering work of Xing [11], researchers have proposed many metric learning algorithms. Most metric learning algorithms can be roughly categorized into linear metrics and non-linear metrics. For linear metrics, the mahalanobis distance takes the dominant place. Most classic metric learning methods adopt the mahalanobis distance form including LMNN [12], ITML [13], OASIS [14]. For non-linear metrics, besides the kernel tricks, neural network is also used to map data into a non-linear space [15]. Our approach benefits from these works, especially Hu’s work DDML [15]. Differently, our approach is more suitable this point-to-set metric learning problem while most conventional methods focus on point-to-point metric learning problem.

3 Proposed Approach

3.1 Learning Prototype Faces

To reduce the intraclass difference, a straightforward idea is to minimize the differences between real world faces and average faces. However average faces are usually influenced by age, gender and race. To avoid this problem, we propose to learn some prototype faces from real world face dataset.

For the powerful ability of affinity propagation clustering algorithm [9], we adopt it to discover underlying prototype faces from face datasets. Affinity propagation clustering algorithm has a good performance in clustering face. The key idea of affinity propagation clustering is to take similarity as input and exchange “responsibility” and “availability” messages between data points. In this way, a high-quality set of exemplars and corresponding clusters will gradually emerge. The procedure of messages exchange is as following:

$$r(i, k) \leftarrow s(i, k) - \max_{k' \neq k} \{a(i, k') + s(i, k')\} \quad (1)$$

$$a(i, k) \leftarrow \begin{cases} \min\{0, r(k, k) + \sum_{i' \notin \{i, k\}} \max\{0, r(i', k)\}\} & i \neq k \\ \sum_{i' \notin \{i, k\}} \max\{0, r(i', k)\} & i = k \end{cases} \quad (2)$$

where $s(i, k)$ indicates similarity between i and k , $r(i, k)$ denotes responsibility from i to k and $a(i, k)$ indicates availability sent from k to i . $r(i, k)$ reflects the accumulated probability for how well-suited K is to serve as the exemplar for point i . $a(i, k)$ reflects the accumulated probability for how well-suited for i choosing k as its exemplar. After some iterations, points with high responsibility and availability will be chosen as prototype faces. Details can refer to [9].

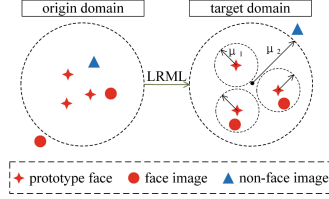


Fig. 2. Schematic illustration of our method. *Left* shows the data distribution in the origin domain, there are some easily misclassified samples. *right* shows the new data distribution in target domain after LRML. Samples can be easily classified in the new domain.

3.2 Locally Rejected Metric Learning

Basic Idea: The goal of metric learning is to seek a transformation that reduces the distance between face images and prototype faces while enlarges distance between non-face images and prototype faces. Figure 2 intuitively illustrates the proposed method. In the origin domain, distance between face image and prototype face is larger than distance between non-face image and prototype face. With the transformation function, data points can be mapped into a new domain where distance between face and prototype face has been reduced to less than threshold μ_1 while distance between non-face and prototype face has been increased to over threshold μ_2 .

Besides, considering that different prototype faces learned by affinity propagation clustering stand for different kinds of faces. To preserve the structure, we believe there should be a gap between inter-prototype-faces, which means a face image is not necessary to keep a small distance to all prototype faces. Therefore, it motivates us to propose a new locally rejected metric learning method that a face image is pushed closer to the nearest prototype face, while keeping a gap with other prototype faces.

LRML: Let $X = \{x_i | i = 1, 2, \dots, N\}$ be the set of N training samples, where $x_i \in \mathbb{R}^d$ denotes the i th training sample, and $Y = \{y_i | i = 1, 2, \dots, K\}$ be the set of K prototype faces, where $y_i \in \mathbb{R}^d$ denotes the i th prototype face. Our approach adopts the Mahalanobis distance form and the distance between a sample from X and a sample from Y can be computed as: $d_M(x_i, y_j) = \|Wx_i - Wy_j\|_2$. Data points parameterized by w are mapped into a new space where the distance between two data points can be computed with squared L2 distances.

As discussed in **Basic Idea**, our approach can be formulated as:

$$\begin{aligned}
 \min_W J(W) &= J_1(W) + J_2(W) + J_3(W) + J_4(W) \\
 &= \frac{1}{NK} \sum_i^N \sum_j^K h_\beta((1 - l_{ij})(\mu_2 - d_M^2(x_i, y_j))) + \frac{1}{N} \sum_i^N h_\beta(l_{iit}(d_M^2(x_i, y_{it}) - \mu_1)) \quad (3) \\
 &+ \frac{1}{N(K-1)} \sum_i^N \sum_{j \neq t}^K h_\beta(l_{ij}(\mu_1 - d_M^2(x_i, y_j))) + \gamma \|W\|_F^2
 \end{aligned}$$

where l_{ij} is the label of image pair (x_i, y_j) , $l_{ij} = 1$ for (x_i, y_j) being intraclass pair (face and prototype face) and $l_{ij} = 0$ for (x_i, y_j) being interclass pair (non-face and prototype face). μ_1 and μ_2 are respectively the threshold for similar pair and dissimilar pair. $h_\beta(z)$ is a generalized logistic loss function to smoothly approximate the hinge loss [16], where $h_\beta(z) = \frac{1}{\beta} \log(1 + \exp(\beta z))$ and β is a sharpness parameter. $h_\beta(z)$ converges to hinge loss as β increases. y_{it} denotes the nearest prototype face to sample face x_i , where $t \in [1, K]$ and l_{iit} is label for (x_i, y_{it}) . γ is a parameter to trade off the regularization term and the hinge loss.

The cost function defined in Eq. (3) consists of four parts. $J_1(W)$ in Eq. (3) is to ensure distance is larger than threshold u_2 if it is a dissimilar pair. $J_2(W)$ in Eq. (3) is to ensure distance is smaller than threshold u_1 if x_1 is a face sample and y_1 is the nearest prototype face to x_i . $J_3(W)$ in Eq. (3) is to preserve the structure of prototype face by keeping a gap between face sample x_i and non-nearest prototype face to x_i . $J_4(W)$ in Eq. (3) is a regularization term.

To solve the objective function defined in Eq. (3), we utilize a batch-stochastic gradient descent scheme. With this scheme, we can obtain a robust solution quickly. Besides, considering that l_{ij} is either 0 or 1, the gradient of cost function can be computed with a classified discussion idea:

$$J(W) = \begin{cases} J_1(W) + J_4(W) & l_{ij} = 0 \\ J_2(W) + J_3(W) + J_4(W) & l_{ij} = 1 \end{cases} \quad (4a)$$

$$(4b)$$

Therefore the batch-stochastic gradient of $J(W)$ can be computed as:

$$\text{For } l_{ij} = 0, \quad \frac{\partial J}{\partial W} = \frac{2}{K} \sum_j^K (h'_\beta(\mu_2 - d_M^2(x_i, y_j))(Wx_i - Wy_j)x_i^T + h'_\beta(\mu_2 - d_M^2(x_i, y_j))(Wy_j - Wx_i)y_j^T) + 2\gamma W \quad (5)$$

$$\text{For } l_{ij} = 1, \quad \frac{\partial J}{\partial W} = 2(h'_\beta(d_M^2(x_i, y_{it}) - \mu_1)W(x_i - y_{it})x_i^T + h'_\beta(d_M^2(x_i, y_{it}) - \mu_1)W(y_{it} - x_i)y_{it}^T) + \frac{2}{K-1} \sum_{j \neq t}^K (h'_\beta(\mu_1 - d_M^2(x_i, y_j))W(x_i - y_j)x_i^T + h'_\beta(\mu_1 - d_M^2(x_i, y_j))W(y_j - x_i)y_j^T) + 2\gamma W \quad (6)$$

For the batch-stochastic gradient descent scheme, batch means a batch of K or $K-1$ prototype faces while stochastic means randomly selecting a sample from dataset x . Then parameter w can be updated by multiplying the batch-stochastic gradient by a learning rate. The main procedure of our method is shown as Algorithm 1.

Score of our method can be computed as:

$$score = \exp(-\min_{j \in [1, K]} (d_M^2(t_m, y_j))). \quad (7)$$

where t_m denotes the m -th test sample.

Combine: Besides, to achieve advanced performance of false positives filtering, we combine the results of our method with Viola-Jones detector. Since the Viola-Jones detector adopts the number of neighbors for filtering false face, we propose following formulation for combining:

$$Score_combine = num_neighbors \times \exp(-\theta \min_{j \in [1, K]} (d_M^2(t_m, y_j))) \quad (8)$$

Algorithm 1. LRML

Input: Training set X and Y , threshold μ_1, μ_2 , regularization parameter γ , learning rate ν , convergence error ε
Output: parameter W
 (training the parameter W):
 Initialization W with diagonal position set 1, otherwise 0
for $iter = 1, 2, 3, \dots$ **do**
 Randomly select a sample x_i from X , y_i from Y
 if $l_{ij} = 0$ **then**
 Compute gradient according to Eq. (5)
 else
 Compute gradient according to Eq. (6)
 end if
 Update W with $W = W - \nu \frac{\partial J}{\partial W}$
 if $|J(W)_{iter} - J(W)_{iter-1}| < \varepsilon$ **then**
 break
 end if
end for

where θ is a parameter to balance the weight of our method and Viola-Jones detector. *num_neighbors* is the confidence score of Viola-Jones detector.

4 Experiments

To evaluate the proposed approach, we test it on the challenging FDDB dataset and our self-collected face dataset. Following describes it in detail.

4.1 Implementation Details

The training set for clustering and metric learning is collected as following: we firstly collect a large set of images containing faces from Flickr. Then we use Viola-Jones detector [4] to detect faces in these images. Due to the limited performance, the results of Viola-Jones detector may contain many non-face images. Lastly, we annotate these detected results and obtain the training set consisting of about 20000 face images and 10000 non-face images. This training set is just used to train a transformation function W .

We extract two kinds of features: histogram of oriented gradients (HOG) and local binary patterns (LBP). Before that we align all images with face alignment algorithm Supervised Descent Method (SDM) [17]. For LBP, we divide each image into 10×10 non-overlapping blocks with size 10×10 . For each block we extract a 59-dimensional LBP feature. Lastly we apply LDA to project the 5900-dimensional feature into a 100-dimensional feature. For HOG, we also divide each image into 10×10 non-overlapping blocks with size 10×10 . For each block, we choose 5×5 cell size and 18 directions. LDA is applied to project get a 100-dimensional feature too. Finally we concatenate the two kinds of feature and get a 200-dimensional feature for each face. Before training, we apply WPCA to the feature. For the parameters of metric learning, we set threshold $\mu_1 = 1, \mu_2 = 10$, regularization parameter $\gamma = 0.01$ and learning rate $\nu = 0.0005$.

4.2 Learned Prototype Faces

We select the true face images from the training set and get a small image dataset only containing faces. Then we apply affinity propagation clustering



Fig. 3. Illustration of 3 kinds of faces learned by affinity propagation clustering algorithm. The learned prototype faces are roughly categorized into: (a) woman face, (b) man face, and (c) children face

algorithm to this dataset to obtain some learned prototype faces. For the affinity propagation clustering algorithm, we set 3 clusters. The clustering algorithm discovers three kinds of face and we roughly category them into woman face, man face and children face, as Fig. 3 shows. Every category contains expression and pose variations, as the left image of each pair in Fig. 3 shows. The reason for pose variation not taking the dominant position is that we have aligned images before extracting feature.

4.3 Experiments on Our Self-collected Dataset

We collect an unconstrained face detection dataset in the wild to evaluate our approach. This dataset is from personal photo album. Different from FDDB, our face dataset contains various poses pictures, scenery and multi-person pictures. It altogether contains 225 images with totally 630 faces. Following the discrete evaluation protocols as FDDB, we count the correct detections according to intersection ratio. We compare our method with original Viola-Jones detector

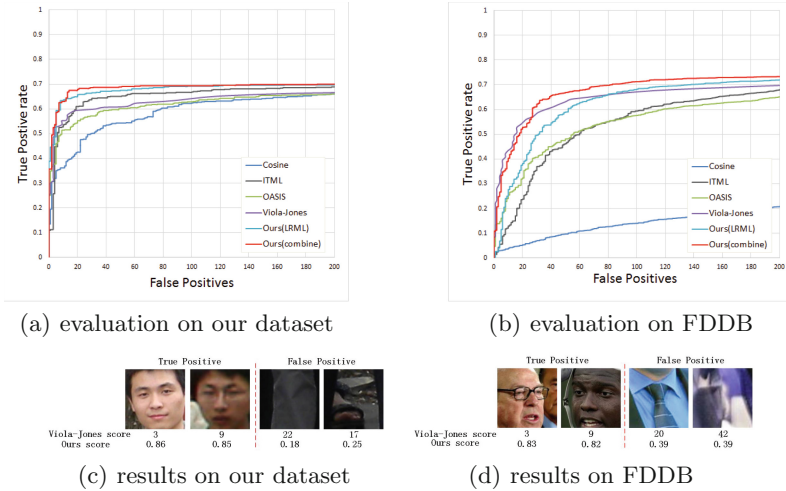


Fig. 4. (a) Evaluation on our dataset (Discontinuous score), (b) Evaluation on FDDB (Discontinuous score), (c)(d) some results of two detectors on two datasets, numbers are the score of two models, the greater score means the greater probability of face.

[4], the cosine similarity and two other classic metric learning based false positives filtering. From Fig. 4(a), we can see that our approach is slightly better than ITML [13] and OASIS [14] while outperforms the cosine similarity significantly. Besides, all of the three metric learning based methods show an efficient improvement to original Viola-Jones detector (implemented by OpenCv). The reason for four post-processing methods achieve good performance to improve Viola-Jones detector may be that most images in dataset are high resolution and have no complex background. Figure 4(b) shows some examples of easily misclassified images. From the result we can see that our approach outperforms the Viola-Jones detector in detecting these hard examples.

4.4 Experiments on FDDB

The Face Detection Data Set and Benchmark (FDDB) is a challenging dataset for evaluating the performance of face detector [5]. The FDDB contains 2845 images with a total of 5171 faces. For the evaluation protocols, we also use the discrete setting. From Fig. 4(b), we can see that our result outperforms cosine similarity significantly while achieves competitive result comparing with ITML [13] and OASIS [14]. Comparing with Viola-Jones detector, our single model of metric learning slightly outperforms it while the combining method is superior to both two single models and achieve the best result. Figure 4(d) shows some examples of easily misclassified images. From the result, we can see that two examples misclassified by Viola-Jones detector is well classified by our approach, which shows the better performance of our approach.

5 Conclusion

In this article, we propose to use affinity propagation clustering algorithm to learn some prototype faces and propose a robust framework of false positives filtering for face detection based on locally rejected metric learning. Our approach shows significant improvements for Viola Jones detector on challenging FDDB dataset and a self-collected dataset. As metric learning is a unified approach to learn discriminative feature, our approach should be able to improve other face detectors. So improvements of other face detectors with our approach will be our future work.

Acknowledgments. This work was partially sponsored by supported by the NSFC (National Natural Science Foundation of China) under Grant No. 61375031, No. 61573068, No. 61471048, and No. 61273217, the Fundamental Research Funds for the Central Universities under Grant No. 2014ZD03-01, This work was also supported by Beijing Nova Program, CCF-Tencent Open Research Fund, and the Program for New Century Excellent Talents in University.

References

1. Li, J., Wang, T., Zhang, Y.: Face detection using surf cascade. In: 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), pp. 2183–2190. IEEE (2011)
2. Li, H., Lin, Z., Shen, X., Brandt, J., Hua, G.: A convolutional neural network cascade for face detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5325–5334 (2015)
3. Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part-based models. *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(9), 1627–1645 (2010)
4. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001, vol. 1, p. I-511. IEEE (2001)
5. Jain, V., Learned-Miller, E.G.: FDDB: A benchmark for face detection in unconstrained settings. UMass Amherst Technical Report (2010)
6. Steinhaus, H.: Sur la division des corp materiels en parties. *Bull. Acad. Polon. Sci* **1**, 801–804 (1956)
7. Ng, A.Y., Jordan, M.I., Weiss, Y., et al.: On spectral clustering: analysis and an algorithm. *Adv. Neural Inf. Process. Syst.* **2**, 849–856 (2002)
8. Comaniciu, D., Meer, P.: Mean shift: a robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(5), 603–619 (2002)
9. Frey, B.J., Dueck, D.: Clustering by passing messages between data points. *Science* **315**(5814), 972–976 (2007)
10. Chen, D., Ren, S., Wei, Y., Cao, X., Sun, J.: Joint cascade face detection and alignment. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8694, pp. 109–122. Springer, Heidelberg (2014). doi:[10.1007/978-3-319-10599-4_8](https://doi.org/10.1007/978-3-319-10599-4_8)
11. Xing, E.P., Jordan, M.I., Russell, S., Ng, A.Y.: Distance metric learning with application to clustering with side-information. In: Advances in Neural Information Processing Systems, pp. 505–512 (2002)
12. Weinberger, K.Q., Blitzer, J., Saul, L.K.: Distance metric learning for large margin nearest neighbor classification. In: Advances in Neural Information Processing Systems, pp. 1473–1480 (2005)
13. Davis, J.V., Kulis, B., Jain, P., Sra, S., Dhillon, I.S.: Information-theoretic metric learning. In: Proceedings of the 24th International Conference on Machine Learning, pp. 209–216. ACM (2007)
14. Chechik, G., Sharma, V., Shalit, U., Bengio, S.: Large scale online learning of image similarity through ranking. *J. Mach. Learn. Res.* **11**, 1109–1135 (2010)
15. Hu, J., Lu, J., Tan, Y.P.: Discriminative deep metric learning for face verification in the wild. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1875–1882. IEEE (2014)
16. Mignon, A., Jurie, F.: PCCA: a new approach for distance learning from sparse pairwise constraints. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2666–2672. IEEE (2012)
17. Xiong, X., De la Torre, F.: Supervised descent method and its applications to face alignment. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 532–539. IEEE (2013)

Biometric Recognition

11th Chinese Conference, CCBR 2016, Chengdu, China,

October 14-16, 2016, Proceedings

You, Z.; Zhou, J.; Wang, Y.; Sun, Z.; Shan, S.; Zheng,

W.; Feng, J.; Zhao, Q. (Eds.)

2016, XVII, 778 p. 358 illus., Softcover

ISBN: 978-3-319-46653-8