Relationship Verification using Siamese Networks

Harsha Vardhan Pokkalla
Electrical & Computer Engineering
Carnegie Mellon University
Pittsburgh
hpokkall@andrew.cmu.edu

Divyaa Ravichandran
Electrical & Computer Engineering
Carnegie Mellon University
Pittsburgh
dravicha@andrew.cmu.edu

Abstract

ConvNets have shown the best performance in image classification and detection in recent years. In this project, we are exploring ConvNets to extract the hidden and more complex information in the images. We define our task: Given a pair of images, a trained model has to verify the relationship between two persons using their facial information. Here, as an initial step, we are considering faces - we try to predict the existence of a Kinship Relationship [Mother -Son, Mother- Daughter, Father - Son, Father- Daughter] between a pair of faces. Also, we extend this theory to a new dataset that we build to test the idea, containing faces of couples who are together in real life. We implemented Siamese networks to do this task. We achieved better performance scores with ConvNets compared to shallow learning techniques.

1. Introduction

Verifying relationship between two persons given only their face images is a very challenging task. Verifying Kinship relations has wide applications such as genetic information research, finding missing family member and social media analysis. Also, learning models on real life couples has even more applications i.e, social media (Tinder). For this task, we are targeting KinFaceW –II dataset. It has these Kinship Relationships [Mother -Son, Mother- Daughter, Father - Son, Father- Daughter] between a pair of faces. Also, we extend this same theory to a new dataset that we build to test the idea, containing faces of couples who are together in real life.

Most of the existing techniques to do the kinship verification are shallow learning techniques. Those techniques involves hand-pick features followed by metric learning algorithms. Such techniques are mentioned in section 3. In recent years, deep learning methods have outperformed on many computer vision tasks. As deep learning methods are end-to-end learning, they perform better by extracting appropriate features and classify them.

In this project, we apply the concept of siamese networks to solve this problem. A siamese network, instead of sampling data independently and identically from a fixed distribution, exploits the interaction between the samples. This approach makes use of the characteristic of real world data where they possess an inherent relational structure.

Experimental results demonstrate that siamese networks have performed better than general metric learning techniques and human level prediction.

2. Dataset

To perform this task, we used existing kin faces dataset 'KinFaceW - II' and our own dataset 'Celebrity Couples in the Wild'. Details about these datasets are as follows.

2.1 KinFaceW II (Kinship Faces in the Wild – Set II)

To have baseline implementation of our model, we used KinFaceW (kinship face in the wild) dataset, a dataset created to help study the kinship relations from unconstrained faces. The kin relations covered by this dataset are limited to father-son, father-daughter, mother-son and mother-daughter. This dataset contains 250 pairs of images (500 images) for each kinship relation. Each image is a cropped face, aligned and 64x64 in size.

Figure 1: Examples of KinShip Faces (KinFaceW - II)

2.2 Celebrity Couples in the Wild

This is a dataset we built ourselves for the purpose of this project. It consist of pictures of celebrity couples, and is sourced from the images on Bing. The dataset consists of ~600 pairs of images of faces of couple pairs, as they appear in the original images that were downloaded from Bing's image search API used while crawling the web. The subjects of these images were Hollywood, Bollywood (Indian) and Korean celebrity couples (so far).

Figure 2: Celebrity Couple Faces extracted from Bing Search

Images of celebrity couples are downloaded from Bing search. Viola Jones method is employed to detect the faces in each image. Image is considered for the dataset if there are only two faces in it. Frontal-faces are cropped and resized to 64x64x3 size and added to the face-pair dataset as was required for the experiment. The final dataset was manually checked to ensure correctness of the data that was collected. Currently, this dataset has 500 pairs of couple faces. We intend to continue adding more diversity to the dataset in terms of appearances of couple pairs across different geo-socio-cultural divides.

3. Related Work

In kinship verification, many papers have been published targeting KinFaceW –II Dataset. Most of the techniques uses low-level features such as Histogram of Oriented Gradients, Local Binary Patterns etc. and k-nearest neighbors, support vector machines as classifiers. There are other papers which employed different kinds of metric learning strategies [5, 6]. To the best of our knowledge, there is only one paper that makes use of deep convnets to do kinship verification [4]. In their paper, the authors trained a single CNN to which input is pair of images concatenated in channel dimension (input: 64x64x6) and network outputs binary class labels (output class: 0 or 1). They have achieved state-of-art performance on this dataset.

In this work, we are exploring Siamese Networks to do this task. In past, Siamese networks have performed very well in face verification as a similarity measure technique. We make use of Siamese networks with our own CNN. Network architecture details are explained in section 4.

4. Method

Siamese Networks:

Siamese Architecture consists of two deep networks with shared weights and connected with a loss function in the end as shown in **Figure 3**. This loss function is trained to make E_w small for matched pairs and large for mis-matched pairs. Siamese networks are explored in the field of learning image descriptors [2] and also in the face recognition [1] which achieved state of art performance.

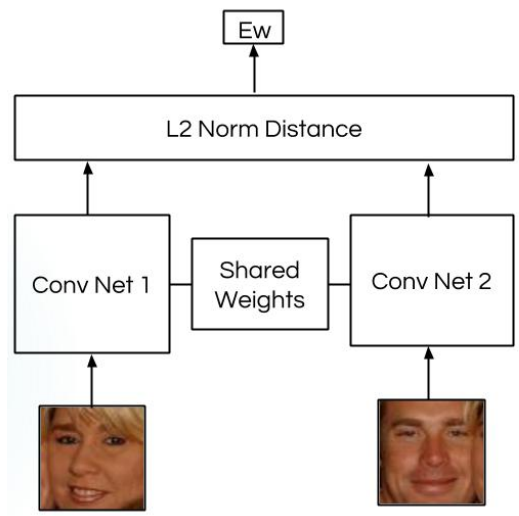

Figure 3: Siamese Network

We are building a network from scratch in the above illustrated model. We initially use the KinFaceW-II dataset to get a baseline implementation. 400 of the 500 images from each relationship set are used for training, while the remaining 100 are used to test. We perform a 5-fold cross-validation to counter overfitting, because the dataset is very limited in size.

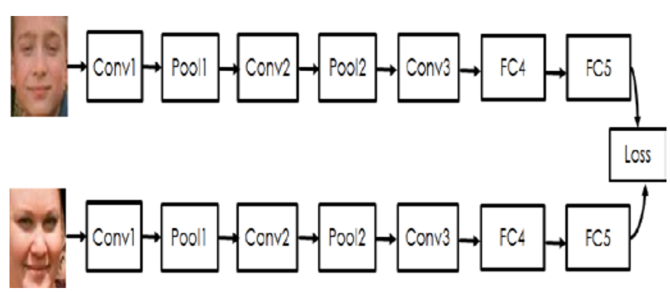

Figure 4: Detailed ConvNet Architecture

The loss function at the end of the Siamese network that connects the two networks and enables backprop is a contrastive loss function. This function ensures that the energy for the correct pair of inputs is low, and that of any wrong pair is large. In this implementation of the Siamese network, we use the hinge embedding criterion and cosine embedding criterion to determine if a pair is a correctly matched or mismatched pair.

5. Results

We trained the above described network separately for each relationship. Since our dataset is limited (only 500 pairs), to avoid overfitting, we implemented 5-fold cross validation on each of the relationship datasets. Results and comparison with existing techniques are shown in table 1.

Implementation Details:

We trained our model with margin embedding criterion (margin = 1) and cosine embedding criterion (margin = 0.5). Loss is back propagated with these criterions. We initialized weights of the network with Gaussian distribution (mean is 0, std is 0.01). Biases are initialized to be zero. At every iteration, batch-size of the input that if forwarded is 100. In all layers, momentum is set to 0.09 and weight decay to be 0.0005. Learning rate for all the iteration is 0.01. Loss is reduced over the iterations for all the models. This is shown in Figure 5.

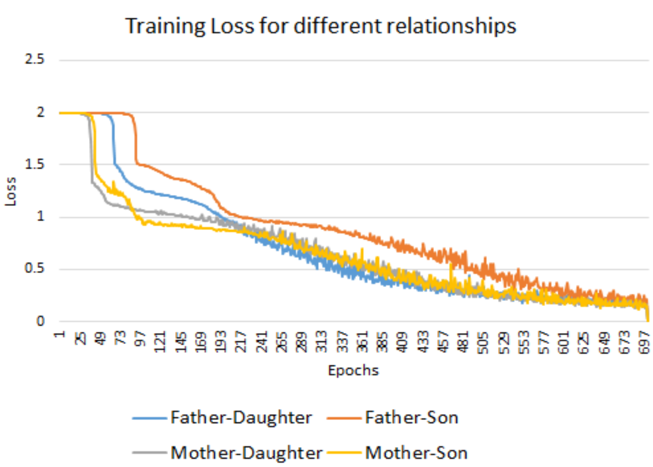

Figure 5: Error plot of four models

The below table provides the accuracy of identifying matching pairs using the KinFaceW-II dataset. Some of the approaches that we use to compare against are also indicated here. Our model performed better than existing non – deep learning techniques. But it cannot perform better than existing CNN model [4]. Reasons might include i) KinFaceW-II is small dataset for our model to generalize better and ii) Low resolution of face images in this dataset.

Method	Mother -Son	Mother-Daughter	Father -Son	Father-Daughter
CML [8]	76	76.5	73.5	73
Human A [9]	69	73	61	61
Human B [9]	78	80	70	68
Our CNN (Hinge)	**89**	**79**	**74**	**75.2**
Our CNN (Cosine)	**80.4**	**80.8**	**75.8**	**76.4**

Table 1: Accuracy of our methods and comparison with other methods on KinFaceW-II Dataset

Deep Network proposed by Zhang et. al [4] has achieved state of art performance using CNNs on this dataset . Their network has 3 convolutional layers, one fully connected layer and softmax layer. They have built an extra model which takes multiple regions of face as inputs instead of full face images. Mean accuracy of this model on KFW II Dataset is 85.3%.

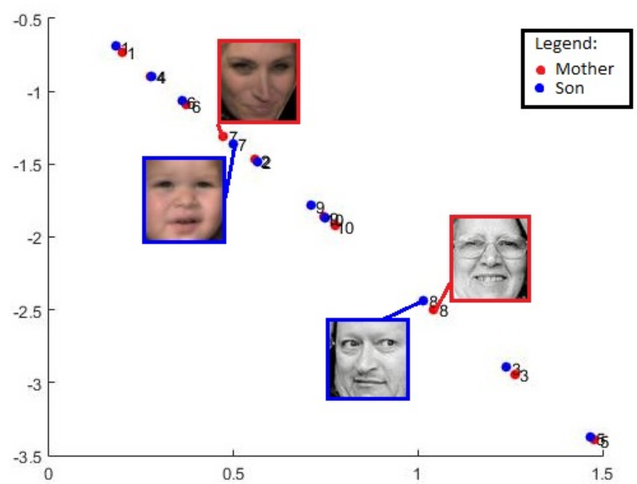

Figure 6: Visualization of last FC layer outputs of Mother-Son pair (2D space)

The graph depicted in figure 6 uses the final FC layer's 2-dimensional outputs from the Siamese network (just before the prediction of 'pair' or 'not pair' is output) to plot some sample points in the mother-son relation. The graph is visualized on the training set pairs. It can be seen that intra-pair distance is pretty small, while inter-pair distance is usually much more significant.

This ratifies our choice of using a Siamese network with hinge/cosine embedding criterion, which seem to be working well in minimizing the distance between the same pair, while increasing that across different pairs.

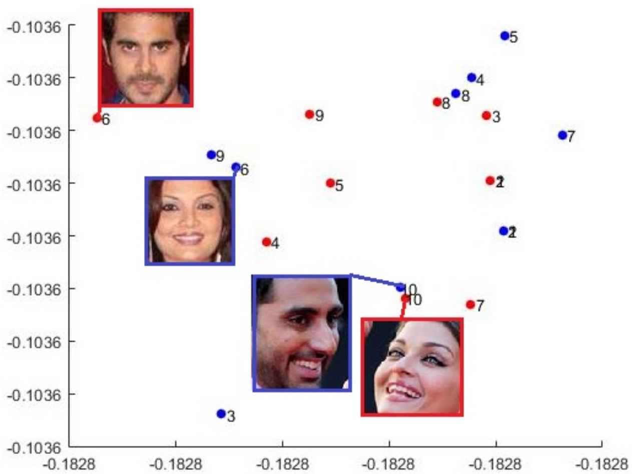

Figure 7: Visualization of last FC layer outputs of Indian Celebrity couples (2D space)

Figure 7 is plotted using the same parameters as Figure 6, but this time on the dataset trained on Indian celebrity couple faces. As is evident by looking at both the plots, the model doesn't perform comparably on the celebrity dataset as it did on the Kinface dataset.

6. Discussion

A possible reason as to why there is this stark disparity in the performance on the different data sets could be that our dataset doesn't have a lot of data. The Indian celebrity dataset has only 350 pairs of images. Some data augmentation in the form of random flips and contrast changes, as well as building the dataset to a more substantial number will likely help.

We carried out testing in the same way as we did for the Kinface dataset. What we think might however work better is to try a nearest-neighbor approach. For instance, feed one image into the network, and output a series of possible matches for that image, based on nearest neighbor probabilities. This approach could be more forgiving than the hard binary approach used earlier, and might also offer insight into how the network is actually deciding on the pairs.

We are currently in the process of testing out these ideas and building a probabilistic matching network using the celebrity datasets.

7. References

[1] Learning a Similarity Metric Discriminatively, with Application to Face Verification http://yann.lecun.com/exdb/publis/pdf/chopra05.pdf

[2]+ DeepFace: Closing the Gap to HumanLevel Performance in Face Verification, CVPR 2014. https://www.cs.toronto.edu/~ranzato/publications/taigman_cvpr14.pdf

[3]Fracking Deep Convolutional Image Descriptors, ICLR 2015 http://arxiv.org/pdf/1412.6537v2.pdf

[4] Kaihao Zhang, Yongzhen Huang, Chunfeng Song, Hong Wu and Liang Wang. Kinship Verification with Deep Convolutional Neural Networks. In Xianghua Xie, Mark W. Jones, and Gary K. L. Tam, editors, Proceedings of the British Machine Vision Conference (BMVC), pages 148.1-148.12. BMVA Press, September 2015.

[5] Jiwen Lu, Xiuzhuang Zhou, Yap-Peng Tan, Yuanyuan Shang, Jie Zhou. Neighborhood Repulsed Metric Learning for Kinship Verification. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), vol. 36, no. 2, pp. 331-345, 2014.

[6] Jiwen Lu, Junlin Hu, Xiuzhuang Zhou, Yuanyuan Shang, Yap-Peng Tan, Gang Wang. Neighborhood Repulsed Metric Learning for Kinship Verification. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2594-2601, 2012.

[7] Deep face recognition, O. M. Parkhi and A. Vedaldi and A. Zisserman, Proceedings of the British Machine Vision Conference (BMVC), 2015

[8] H. B. Yan, J. B. Lu, W. H. Deng, and X. Z. Zhou. Discriminative multimetric learning for kinship verification. Information Forensics and Security, IEEE Transactions on, 9 (7):1169–1178, 2014.

[9] J. W. Lu, J. L. Hu, X. Z. Zhou, Y. Y. Shang, Y. P. Tan, and G. Wang. Neighborhood repulsed metric learning for kinship verification. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pages 2594–2601. IEEE, 2012.