CS230

Family kinship Recognition Using Inception RestNet

Bruce Jianye Liu
Department of Computer Science
Stanford University
bruceliu@stanford.edu
github repo: https://github.com/brucelau-github/cs230

Abstract

Computer vision based face recognition had a significant progress over last decade and most of face identification tasks focus on comparing two face images to verify if they are the same person. However, Kinship recognition by face pictures is not quite popular in academia. As more open family face dataset publicly accessible and new well-engineered deep neural network architecture invented, state-of-art results on kinship identification becomes possible. In this paper, we trained a deep neural network architecture based on a fine-tuned Inception ResNet v2 to identify parent-child, siblings relationships by comparing two face pictures and achieved 82% accuracy on FIW test set, surpassing previous study about 7%.

1 Introduction

Parents and children share nearly half of genetic information and almost the same amount of DNA are also shared among siblings in a family. People biologically related often show some sort of delicate similarities among each other. This delicacy could be easily caught by human eyes, by observing faces of their family photos. As computer vision performance improving during the past decade, it becomes possible to use machine learning to capture the different. Computer vision based kinship recognition could lead variety of useful applications in life such as missing-children parents matching, family album organization, social networking apps, lost sibling/relatives searching, crime investigation. In this paper, we propose a fine-tuned KinNet model to classify the relationship between two faces – parent-children, sibling-sibling, none-kinship, and same person. We are able to make over 80 percent accuracy.

Previous research mainly focus on kinship verification and family classification. It is hard to achieve very high accuracy on these tasks. Facail recognition recently attained a new record of accuracy, and this motivated us to employ a similar approach to booster kinship recognition benchmarks.

2 Related Works

Even though many researchers have tried traditional approaches, Deep learning often showed state-of-art achievement out-performing other methods in image recognition tasks. Many neural network models are designed and invented, among them AlexNet[6] from University of Toronto in 2012, GoogLeNet (Inception)[7] from Google in 2014, VGG[8] from Oxford Vision Geometry Group in 2015, and ResNet[9] form Microsoft Research showing significant impact on academia. Their error rate and accuracy surpassed human performance on ImageNet[5] dataset.

Similarly to image classification problems, facial recognition employing deep neural networks error rates have dropped over the last two decades by at least several orders of magnitude. Many commercial

CS230: Deep Learning, Spring 2020, Stanford University, CA.

applications are deployed to verify users' identity, such as mobile phones unlocking features. Most current methods on face verification use hand-crafted deep neural nets. Yaniv et al.[12] designed DeepFace architecture using 3D face modeling to make affine transformation before feeding into a nine-layer dense neural network. It contains 120M parameters and reaches 97.32% accuracy on LFW[10] dataset. Yi et al.[13] developed a very deep neural architecture, DeepID3, building by stacking convolution and inception layers and achieved 96% accuracy in the same dataset. Schroff et al.[2] presented a system, called FaceNet, that directly learns a mapping from face images to embeddings and reached a new record accuracy of 99.63%.

Visual kinship recognition is one of tasks that attract many researchers to put an effect on it. Fang et al.[14] initially made an attempt on parent-child verification using K-Nearest-Neighbours over computed face features such as eye color, skin color, hair color, facial parts size and positions. They achieved a classification accuracy of 70.67% on the test set. Joseph et al.[3] contributed Family In the Wild (FIW) dataset, and introduced several benchmarks on popular image recognition architectures such as SIFT[15], LBP[16], pre-trained VGG-Face[17], ResNet CNN[18]. They achieved 72.15% accuracy, 15% higher that human average performance.

3 Dataset

Figure 1: 108x124 face sample

Families In The Wild (FIW) Database is one of the largest and most comprehensive databases available for kinship recognition, published by Robinson et al. in 2016. We use the latest version 0.1.2 at writing time, which includes 13,188 faces cropped from photos of 1018 families. The dataset contains 11 kinship types, divided into father-daughter (F-D), father-son (F-S), mother-daughter (M-D), mother-son (M-S), brother-brother (B-B), sister-sister (S-S), grandfather-granddaughter (GF-GD), grandfather-grandson (GF-GS), grandmother-granddaughter (GM-GD), grandmother-grandson (GM-GS). Sibling and parent-child types are the most relevant to our research. So by adding up 64,669 F-D, 46,143 F-S, 68,935 M-D, 48,940 M-S types, we get 22,687 pairs of parent-child photos and 55,937 pairs of sibling photos. All face images are cropped from public family photos with idential size 108*124*3 and manually labeled. For the same person faces, we generated them by selecting pictures with the same FaceID. Similarly non-related picture pairs are created by selecting one image in an unrealted folder under the family ID folder and one of the faces in the family.

face-pair image distribution		
pair types	face-pair number	percentage
parent-child	228,687	21.17%
siblings	55,937	5.18%
same	230,938	21.38%
unrelated	564,496	52.27%
	1,080,058	

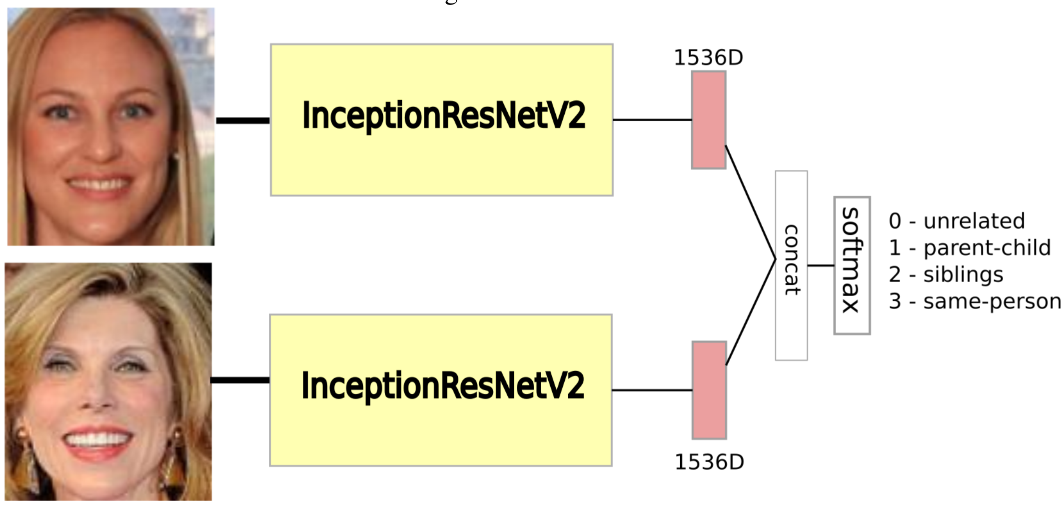
Table 1: Training data distribution

Table 1 shows all the data distribution that we used during the training session. We managed to assemble nearly 1 million image pairs from FIW dataset by scanning folders, permuting image pairs, and summing up existing face pairs. The unrelated image pairs is over 500,000, nearly 52% of all data. Parent-child and the same face pairs contain almost the equal number of images, about 230,000 respectively, about 21% of total images. On the other hand, only 5% of data are sibling face pairs, which may affect our model on predicting sibling kinship. In practice, we forsake some data in order to make our dataset even on each category.

4 Methods and Models

Convolutional architecture performances very well in face recognition tasks. In our method, We fine tuned the InceptionResNetV2 model to classify kinship face image pairs. Figure 2 shows our model architecture.

Figure 2: Our model



InceptionResNetv2[11], containing over 1 million parameters, demonstrated its performance on image classification challenge although it is very expensive to train from scratch. Instead of starting from randomized weights, we took a model with pre-trained weights on ImageNet. Using this transfer-learning approach could reduce our training time.

To handle the uneven training dataset, we take 50,000 images from each category, shuffle the whole 200K images, split 6000 samples as a test set, and leave the rest of the data as a training set. To fit our memory, at each epoch we randomly extract 10,000 training images from a training set and divide them into mini-batch sizes of 16. We trained two models with this training dataset arrangement.

In the first model, we remove the last full connection layers of the InceptionResNetV2 model, and add a global average last layer with 1536 dimension output. Then connect the output of convolutional layers with two dense layers, containing 1024, 128 units separately. The final layer is a 4 node softmax output. For each face pair, we feed them into InceptionResNet which would produce two 1536 dimensional vectors and concatenate them into 3072 dimensional vectors that are forwarded into the dense layers to produce 4 outputs. Table 2 shows the detailed architecture.

layers	**size-in**	**size-out**	**param**	**FLPS**
Input1	299x299x3		0	
Input2	299x299x3		0	
InceptionResNetV2	299x299x3	1x1536	75M	75M
InceptionResNetV2	299x299x3	1x1536	75M	75M
concat	2x1536	1x3072		0
fc1	1x3072	1x1024	3M	3M
fc2	1x1024	1x128	131K	131K
softmax	1x1024	1x4	4100	4100

Table 2: KinNet Layers Model I

After several epochs of training, model I converge gradually and the best result we get is 68% test accuracy, 1.58 loss. Due to computer memory limitation, it is impossible to fit all training images. We choose to shrink the training set further to 5000 images for each training session, in over 5 sessions. After the fourth session, it starts to overfit, the training accuracy jumping up to 98% while the test accuracy remains about 68%.

To handle overfitting problem with model I, we change InceptionResNetV2 output from global average to global max, delete these two middle fully-connected layers and keep only the last softmax layer of four units. Table 2 shows our model II structure. This modification works pretty well during training sessions. We don't choose to reuse the trained weights on model I. Instead we take the same starting point of Model I, and set learning rate of 0.01 with Adam optimization on the first 8 epochs, until its test accuracy reached 65%. The Model II starts over shooting, its training accuracy bumping

layers	size-in	size-out	param	FLPS
Input1	299x299x3		0	
Input2	299x299x3		0	
InceptionResNetV2	299x299x3	1x1536	75M	75M
InceptionResNetV2	299x299x3	1x1536	75M	75M
concat	2x1536	1x3072		0
softmax	1x3072	1x4	4100	4100

Table 3: KinNet Layers Model II

above 80% but not on the test set. We reduced the learning rate to 0.001 for the next 5 epochs, and its test accuracy slowly climbed over 75%. In the last training session, we further decreased the learning rate down to 0.0001 and achieved the best test accuracy to 82%. After 82% its performance won't improve even if we adjust learning rate again.

Figure 3: Results

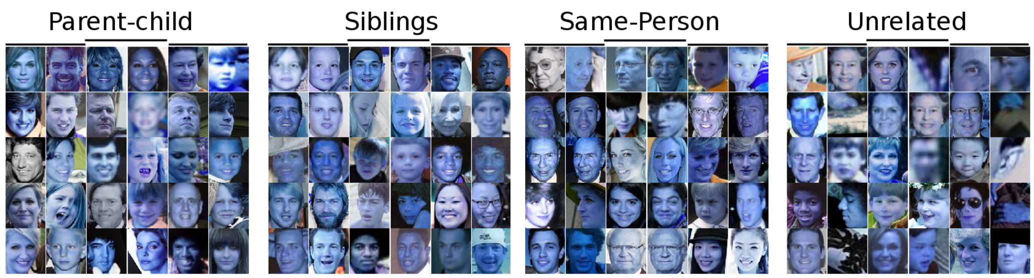
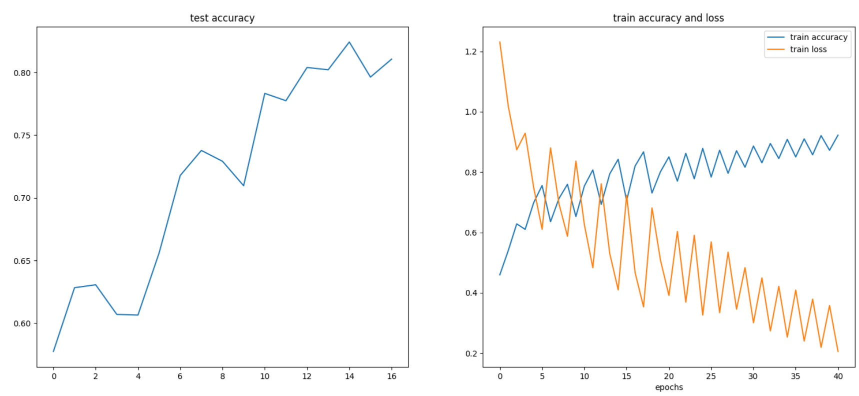


Figure 4: Test and Training Accuracy



5 Conclusion/Future Work

By iteratively adjusting hyperparameters of Model II, we attained 82% test accuracy on FIW dataset, though the first model didn't get an ideal result. It is proved that the InceptionResNet performed well not only on basic object recognition tasks but also on face recognition tasks. Further modification on our model could be also worthy to try in future experiments, such as computing cosine similarity of two face images encoding, using k-nearest neighbour algorithm to cluster the encoding, or changing some layers of InceptionResNet.

Even though approximate half of genes are shared among family members, solely relying on face pairs for kinship verification might be hard to achieve higher accuracy. If people without any family relation could look extremely like each other, even humans might misjudge their relationship. Beside facial likeness, gene similarity among siblings and generations could also exhibit in terms of height, skin color, nail shape, toes length, ear contours, hand size etc. This information couldn't be presented in face images. Adding these additional information into our model, it could boost our model significantly. Due to time limitation and team size, we aren't able to collect this data.

References

[1] Robinson, Joseph P., et al. Families in the Wild (FIW): Large-Scale Kinship Image Database and Benchmarks. *Proceedings of the 2016 ACM on Multimedia Conference* - MM '16, 2016, doi:10.1145/2964284.2967219.

[2] Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. doi: 10.1109/cvpr.2015.7298682

[3] Wang, S., Robinson, J. P., & Fu, Y. (2017, May). Kinship verification on families in the wild with marginalized denoising metric learning. *In 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017) (pp. 216-221). IEEE.*

[4] Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017, February). Inception-v4, inception-resnet and the impact of residual connections on learning. In *Thirty-first AAAI conference on artificial intelligence*.

[5] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K.,& Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248-255).

[6] Krizhevsky, Alex., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing system* (pp. 1097-1105).

[7] Szegedy, c., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9)

[8] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image reognition. *arXiv preprint arXiv*: 1409. 1556.

[9] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pateern recognition* (pp. 770-778).

[10] Huang, G. B., Ramesh, M., Berg T., & Learned-Miller, E. (2007) Labeled faces in the wild: a database for studying face recognition in unconstraned environments. In *Technical Report* 07-49, October 2017

[11] Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017, February). Inception-v4, inception-resnet and the impact of residual connections on learning. In *Thirty-first AAAI conference on artificial intelligence*.

[12] Taigman, Y., Yang, M., Ranzato, M. A., & Wolf, L. (2014). 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).

[13] Sun, Y., Liang, D., Wang, X., & Tang, X. (2015). Deepid3: Face recognition with very deep neural networks. *arXiv preprint arXiv*:1502.00873.

[14] Fang, R., Tang, K. D., Snavely, N., & Chen, T. (2010, September). Towards computational models of kinship verification. In 2010 *IEEE International conference on image processing* (pp. 1577-1580). IEEE.

[15] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2), 91-110.

[16] Ahonen, T., Hadid, A., & Pietikainen, M. (2006). Face description with local binary patterns: Application to face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 28(12), 2037-2041.

[17] Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep face recognition.

[18] Wen, Y., Zhang, K., Li, Z., & Qiao, Y. (2016, October). A discriminative feature learning approach for deep face recognition. In *European conference on computer vision* (pp. 499-515). Springer, Cham.