

From Same Photo: Cheating on Visual Kinship Challenges^{*}

Mitchell Dawson¹[0000–0002–6719–6584], Andrew Zisserman¹[0000–0002–8945–8573],
and Christoffer Nellaker²[0000–0002–2887–2068]

¹ Visual Geometry Group (VGG), Dept. of Engineering Science, University of Oxford
`{mdawson,az}@robots.ox.ac.uk`

² Nuffield Dept. of Women’s & Reproductive Health, Big Data Institute, IBME,
University of Oxford
`christoffer.nellaker@bdi.ox.ac.uk`

Abstract. With the propensity for deep learning models to learn unintended signals from data sets there is always the possibility that the network can “cheat” in order to solve a task. In the instance of data sets for visual kinship verification, one such unintended signal could be that the faces are cropped from the same photograph, since faces from the same photograph are more likely to be from the same family. In this paper we investigate the influence of this artefactual data inference in published data sets for kinship verification.

To this end, we obtain a large data set, and train a CNN classifier to determine if two faces are from the same photograph or not. Using this classifier alone as a naive classifier of kinship, we demonstrate near state of the art results on five public benchmark data sets for kinship verification – achieving over 90% accuracy on one of them. Thus, we conclude that faces derived from the same photograph are a strong inadvertent signal in all the data sets we examined, and it is likely that the fraction of kinship explained by existing kinship models is small.

Keywords: Kinship Verification · Data Set Bias · Deep Learning · Convolutional Neural Network

1 Introduction

Kinship verification is the task of determining whether two or more people share a close family relation, using only photographs of their respective faces. Potential uses of kinship verification include being able to organise digital photo albums, detecting lost children using photos of the child’s parents, and in distinguishing family traits from environmental or disease related changes in clinical research contexts.

^{*} This research was financially supported by the EPSRC programme grant Seebibyte EP/M013774/1, the EPSRC Systems Biology DTC EP/G03706X/1, and the MRC Grant MR/M014568/1.



Fig. 1. Representative examples of how cropping face images from the same original photo can make the kinship verification a trivial task. Confounding, non-kinship information includes camera specific noise, the background similarity (A,C), the historical era in which the photo was taken (B), image lighting and tone (A,B,C), and the relative age difference between parents and children (A,B,C).

Over the last few years, several kinship verification data sets have been developed for research groups to benchmark against. These data sets of images and videos have been produced by different groups, and each has its own unique characteristics including variations in the number of images, collection methods, and types of relation included.

In building these data sets for kinship verification, often different individuals’ images have been cropped from larger family photos. While this is a good way to find people in the same family it also builds in another clue – a bias that people from the same photo are related. This has been overlooked as a potential issue, as it is far easier to identify whether two face images are from the same original photo than it is to determine if they are kin. As will be demonstrated, in some data sets, this ‘from same photograph’ (FSP) signal can be used to achieve results comparable to the state of the art.

There are several cues that can be used to determine if two face images are cropped from the same photograph, and by inference are more likely to be kin. For example, common lighting and shadows, background, blur, resolution, tone, contrast, clothing, camera specific noise and exact regional overlap between crops. Another confounding signal, which is present when parent and child are cropped from the same photo, is that the age difference between the people shown in each crop will be roughly the same as the average age which parents give birth to children.

Deep neural networks are notorious for finding subtle data shortcuts to exploit in order to ‘cheat’ and thus not learn to solve the task in the desired manner; an example is the misuse of chromatic aberration in [6] to solve the relative-position task. Thus deep learning models are liable to pick up on these FSP cues to gain a boost in performance on kinship tasks.

This problem was raised by Lopez *et al.* [16], who recommended that the two KinFaceW data sets should no longer be used in kinship research. Their

work showed that comparable results could be achieved on the data sets using a method with no understanding of kinship: classifying pairs of images based upon the the distance between the chrominance averages of images in the Lab color space. Further to this work, Wu *et al.* [28] showed that results on kinship data sets can be improved by using colour images rather than grey-scale images. Although one might expect that skin colour could help to identify kinship relations, it also gives more information about the background and colour distribution of the image, a confounding variable which can be used to improve results on data sets containing images cropped from the same photo.

Other groups have attempted to avoid this issue by using images where each family member is cropped from different original photographs, such as the Family 101 [8] and UBKinFace [30] data sets. However, such images are expensive to find and so photographs of famous people and their children, who have many public images available, are frequently included, introducing a new bias into the data sets.

Many models have already reported on these data sets infected by the signal of pairs cropped from the same photo. Therefore, we would like to benchmark the extent to which these results could have used the ‘from same photo’ signal. This will provide a clear understanding of how much of the accuracy reported is based upon kinship signals, and how much could have been ‘from same photo’ signal.

In this paper we benchmark the accuracy which can be achieved on five popular kinship data sets, using only the signal of whether two face images are from the same photo. We achieve this by creating a new data set of 24,827 images, containing 145,618 faces, and 914,517 pairs of faces taken from non-familial group photos. We use this data set to train a CNN classifier to determine if two facial images are cropped from the same original photo or not. Crucially, this classifier has no understanding of kinship relations.

We present results on the KinFaceW-I [19], KinFaceW-II [19], Cornell KinFace [9], Families in the Wild (FIW) [24], and TSKinFace [23] data sets. We show that the ‘from same photo’ signal exists for all of these five data sets. For some, the signal is fairly weak, whereas for others it explains a large proportion of the state of the art results. In many cases we achieve comparable results with other submissions using only the ‘from same photo’ signal.

2 Training a Classifier to Detect Faces from the Same Photo

In this section, we describe our approach to train a CNN classifier to be able to determine whether two facial images are cropped from the same original photograph or not. For positive training data we use pairs of face images cropped from non-family group photographs. This was done to ensure the classifier could not learn kinship signals between faces. We evaluate the ability of our classifier to classify faces from the same photo on a held out test set.



Fig. 2. Examples of photos found using non-kin group search terms and the Bing Azure API [4]. Pairs of faces found in these images are used in creating the From Same Photo (FSP) data set.

2.1 Generating the FSP Data Set

We would like to build a data set of face images that were cropped from the same original photo, but crucially where the people shown in the face images do not share a kinship relation.

We began by creating a list of 125 search terms that describe groups of people, who are unlikely to share a kinship relation. These include terms such as ‘school students’, ‘business meeting’, ‘sport team photos’ (the complete list of terms is given in the supplementary material). We then use the Bing Azure API [4] to find URLs of images related to these search terms. We ensure that the images returned are large enough to potentially contain high resolution faces, and that each URL is returned only once in an attempt to avoid duplicate images. These searches result in a list of 81,784 URLs: an average of 654 per search term. We download each image and, with some loss due to moved content and unusable file formats, obtain 76,450 images.

The Dlib face detector [10] is used to find faces within these images. For our purposes, a usable group photo is defined to be an image which contains at least two faces, where the height and width of both face detection regions is not less than 50 pixels. This results in a maximum usable positive data set of 24,827 group photos, which contains 145,618 face images and 914,517 possible pairs of face images. We partition this data set into training (70%), validation (10%) and testing (20%), as shown in Table 2.2.

Although the vast majority of pairs of faces in the positive FSP data set are correctly labelled, it was found that there are some cases of negative pairs making it through the processing pipeline. For example, some of the group images collected were composite collages of separate photos, and so the faces in the collages were not originally taken from the same photo. Another example which could lead to falsely labelled positive pairs comes from photographs of people standing next to portrait paintings or photographs which contain a face. These faces are also detected by Dlib, and so lead to pairs of faces being falsely labelled as positives. Furthermore, we can not exclude that some images will have true kin related people present in them. Overall, these examples make up a very tiny proportion of the data set, and so it is not expected that they skew our results significantly.

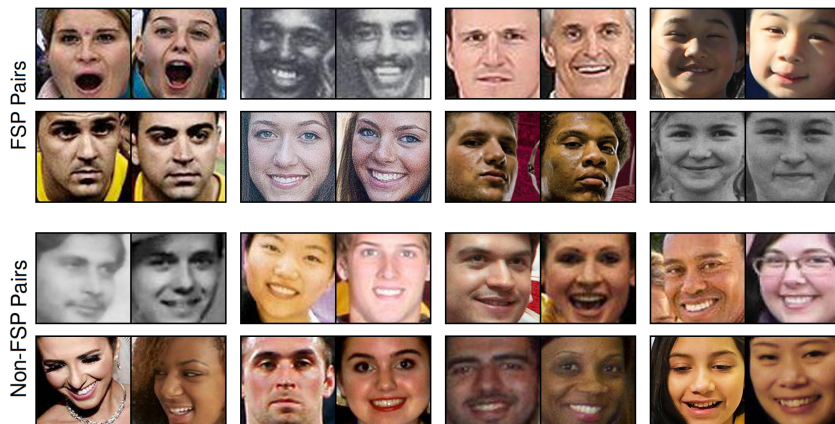


Fig. 3. Examples of pairs of faces used to train the From Same Photo (FSP) classifier. FSP pairs are cropped from the same non-kin group photo. Images in Non-FSP pairs are cropped from separate non-kin group photos.

Training the FSP classifier also requires creating a corresponding negative set of pairs of faces that are cropped from different original photographs. This is achieved by taking each image in the positive data set and randomly swapping one of the faces for another in the same training/validation/testing subset. This ensures that pairs of faces in the negative data set are not from the same photo, and that face photos do not leak across train, validation and testing splits. Furthermore, matching the total number of positive pairs and negative pairs in this way also helps to ensure that the FSP classifier does not learn a bias towards predicting a particular class.

2.2 The FSP Classifier

In this section we describe the CNN classifier used to detect whether two facial photographs are cropped from the same original image. For training data we use the balanced FSP data set described above in section 2.1. We evaluate the performance of the FSP classifier on a test set of images from the FSP data set, and analyse the results using the standard receiver operating characteristic (ROC) and precision-recall (PR) curves.

Architecture. The CNN architecture can be split into four distinct parts as shown in Figure 4. The first part consists of two parallel copies of a pre-trained VGG-M-Face, up to the output from the fully connected FC6 layer. This architecture was first used in [2] and has been trained here for facial recognition on the VGGFace data set. The VGG-M-Face networks generate a 4096-dimensional feature vector for each face image of the input pair. Both feature vectors are then fed to a trainable fully connected layer to reduce their dimensionality. These reduced

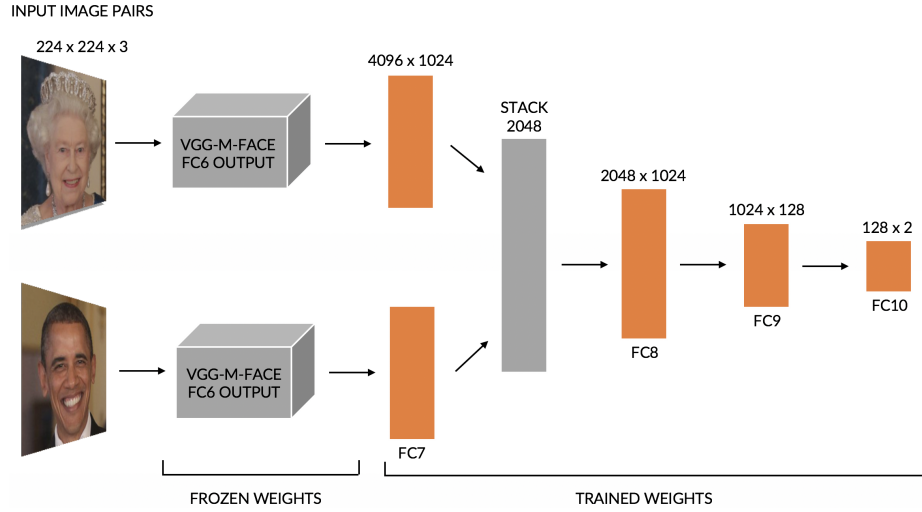


Fig. 4. The FSP network architecture.

feature vectors are then stacked into a single vector and fed through through three more fully connected layers and ReLU activation functions to produce a two-dimensional vector giving the probability that the pair belongs to the class ‘from same photo’ or ‘not from same photo’. The network is implemented in the PyTorch framework, and the code (and data set) will be made publicly available.

Training. The FSP network was trained using stochastic gradient descent to minimise a softmax cross entropy loss function. During training we freeze the weights of the VGG-M-Face part of the model, but train all other layers. The initial learning rate was set to 0.1, and decreased by a factor of 0.5 every five epochs, until the classifier’s performance on the validation set began to saturate. In training we use dropout with $p = 0.1$ between layers FC8 and FC9 to prevent overfitting. During training we augment our input images by resizing them up to 256×256 pixels and taking a random 224×224 pixel crop from the resulting image.

Table 1. Summary of the number of pairs of face images in the data sets used to train, validate and test the FSP classifier. The area under receiver operating characteristic curve (ROC AUC) is reported for the FSP classifier on the test set of images

Data set	Train	Validation	Test	ROC AUC (%)
FSP	1,321,440	184,358	323,236	98.1

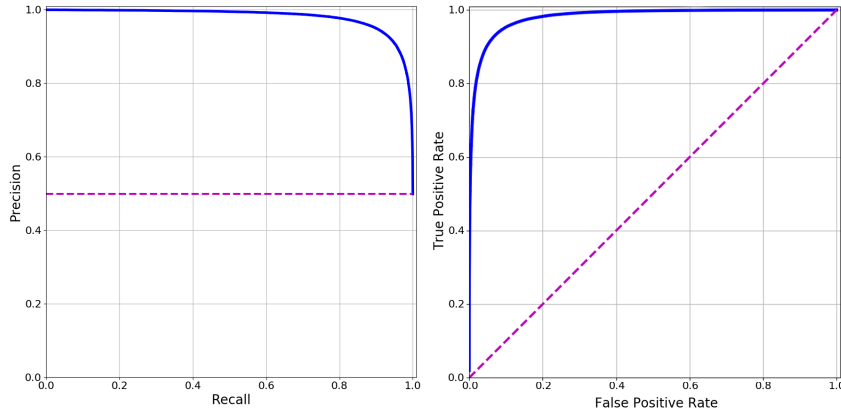


Fig. 5. Precision-recall and ROC curves for testing the FSP classifier on the FSP test set. The area under the ROC curve is 98.09%, and the classifier performs with an equal error rate of 6.97% and an average precision of 98.07%.

Testing. We evaluate the FSP network on a test set taken from the FSP data set that was not used in training or validation. The testing set consists of 323,236 pairs of face images, half of which are cropped from the same original photo. At test time we test a pair of images ten times using different permutations of region crops and horizontal flips. The result of these ten tests is then averaged to give the classifier score for a pair of images.

As can be seen in Figure 5, the FSP classifier is able to tell whether two images are from the same photograph with very high accuracy. Testing resulted in an area under ROC curve of 98.09%, with an equal error rate of 6.97%.

3 Testing FSP Classifier on Kinship Data Sets

Kinship verification tasks have branched into three major challenges, consequently multiple different types of kinship verification data sets have been built. The first type of task is one-to-one kinship verification, where one wishes to determine whether two individuals are related. Alternatively, tri-subject kinship verification is where given a picture of a mother-father and a third person, we wish to determine if the third person is the biological child of the two parents. A final challenge is one-to-many family classification, where one wishes to know which family a person belongs to among a data set. We examined five popular data sets and the degree to which FSP is a biasing factor for the respective one-to-one kinship verification task challenges.

3.1 Kinship Verification Data Sets

KinFaceW-I & *KinFaceW-II* (Lu et al. [19]). These are two widely used kinship verification data sets introduced in 2014. They contain images of public figures

paired with their parents or children. They are the most commonly tested on kinship data sets, but were also the first to receive criticism for using images cropped from the same photograph [16]. The KinFaceW-I data set contains pairs of images from the four main bi-subject kinship verification categories: mother-daughter (127 pairs), mother-son (116 pairs), father-daughter (134 pairs), father-son (156 pairs). The major differences between the two data sets are that KinFaceW-II is larger, and contains a greater number of photos cropped from the same original image compared to KinFaceW-I.

Typically these data sets are tested using five-fold cross validation, where both the positive and negative pairs are specified for each of the five folds. Here we collect all the positive and negative pairs across all five folds to create the balanced test set. The KinFaceW-I test set contains 1066 pairs of face images, and the KinFaceW-II test set contains 2000 pairs of face images, with 250 for each of the four major relations.

Cornell KinFace (Fang et al. [9]). This data set was published in 2010 and consists of parents and children of public figures and celebrities. It is the smallest data set among the five tested in this paper, consisting of only 144 positive pairs. Human level performance on this data set was benchmarked as 70.7%. For the negative set we randomly substitute a parent or child from the positive pair with a random parent or child from the remaining positive set. To avoid bias from choosing a particularly easy negative set we average over five different versions of the negative set, randomised uniquely each time.

Families in the Wild (FIW) (Robinson et al. [24]). Families in the wild is by far the largest of the data sets we test on. It contains 1000 families, 10,676 people and 30,725 face images. 418,060 pairs of images, heavily weighted towards over 37,000 pairs for each of the four major parent-child relationships, and over 75,000 for sibling-sibling relationships. The data set was made available as part of the 2018 Recognising Families in the Wild (RFIW) kinship verification challenge. Here we test on the 99,962 pairs of images in the challenge evaluation set.

TSKinFace (Qin et al. [23]). TSKinFace was introduced as an alternative to the growing number of bi-subject kinship verification data sets. For many of the use situations described for kinship verification, such as organising family photo albums or locating missing children, it is likely that pictures of both parents will be available. TSKinFace is the largest publicly available data set of triplet kinship images to date. The data set contains 787 images of 2589 individuals belonging to 1015 tri-subject kinship groups (father-mother-child). The images from TSKinFace are collected from family photographs.

TSKinFace specifies related triplets of father-mother-son and father-mother-daughter. For the negative set they randomly combine images of parents with an image of different child of the same gender as their own. To reduce the probability of reporting results on a particularly easy negative set, we repeated our experiments with a different set of permutations for the children and parents in the negative set each time. The data set is split by the gender of the child with

513 positive father-mother-son (FM-S) triplets and 502 positive father-mother-daughter (FM-D) triplets.

At test time we split each triplet into two pairs, father-child and mother-child. Each pair is then scored by the FSP classifier. We take the maximum of these two scores as the test score for a triplet. This corresponds to asking whether at least one of the parents is cropped from the same original photo as the child.

3.2 Implementation Details

We apply the FSP classifier to the kinship data sets naively, that is to say without further training for the intended kinship verification task. To achieve this we extract each image at three crop sizes. This is required as the Dlib face detector tends to propose tight square regions around the centre of a face, often cutting out chin, tops of heads and ears. However, the face images contained within two of the data sets we test on are more loosely cropped. We found that the greatest accuracy was achieved on KinFaceW-I, KinFaceW-II and FIW using the FSP classifier trained with images cropped to the standard Dlib size. Whereas on Cornell KinFace and TSKinFace, the best results were obtained by expanding the width and height of the images by 15%.

We report the accuracy our FSP classifier is able to obtain on each kinship test set. To determine the accuracy, we set the threshold of our FSP classifier using five-fold cross validation on the test sets. We determine the threshold which produces the maximum achievable accuracy for the FSP classifier on 80% of the test set, then calculate the accuracy on the remaining 20% of the test data for five splits. The mean accuracy across the splits is then reported.

4 Results

We show the results for the FSP classifier as a kinship classifier on the five data sets in Table 2. For each of the bi-subject kinship verification tasks we report high accuracies: KinFaceW-I 76.8% , KinFaceW-II 90.2%, FIW 58.6%, Cornell

Table 2. Results of using the FSP classifier to predict kinship on KinFaceW-I, KinFaceW-II, Cornell KinFace, FIW and TSKinFace data sets. Accuracy percentages are shown for the pair subsets of mother-daughter (MD), mother-son (MS), father-daughter (FD), and father-son (FS), and the triplet subsets of father-mother-daughter (FMD) and father-mother-son (FMS), as well as across the entire test set

Data set	MD	MS	FD	FS	FMD	FMS	All
KinFaceW-I	86.0	78.3	74.1	74.6			76.8
KinFaceW-II	94.8	90.3	84.5	92.3			90.2
FIW	60.3	59.3	59.0	57.5			58.6
TSKinFace					88.6	89.4	88.6
Cornell KinFace							76.7

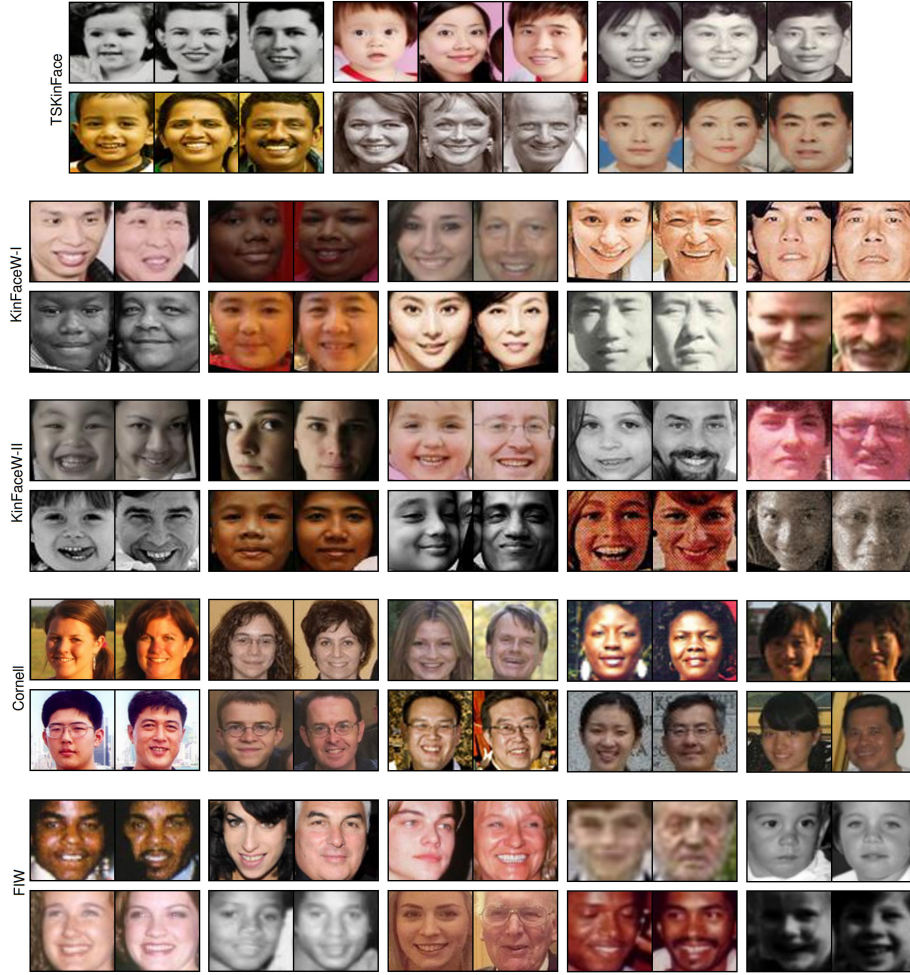


Fig. 6. Examples found in kinship data sets which we predict are taken from the same photo. Each row corresponds to examples from one of the five analysed kinship data sets. Note the similarities in background, facial pose, lighting, and overall image tone between images cropped from the same original photo.

KinFace 76.7%. There is variance between the tasks depending on the gender of the parent and child in the image. In kinship verification tasks this could be interpreted as perhaps biases in facial similarities between genders, however this is not something we should expect from the FSP classifier. It is more likely to be due to biases in likelihood for various gender pairs of family members to show up together in photographs, or even just stochastic variance from data set sampling.

Table 3. Results of using the FSP classifier to predict kinship, and previously published kinship classifiers on all five kinship data sets. KinFaceW is abbreviated as KFW. Note the high accuracies achieved by using the FSP classifier as a kinship classifier on each data set, in comparison with the average and state-of-the-art maximum accuracies over results reported in previous publications. Also note the available human annotation benchmarks which are all exceeded by the FSP classifier model

Paper	Year	KFW-I	KFW-II	Cornell KF	FIW	TSKinFace
FSP classifier (ours)	2018	76.8	90.2	76.7	58.6	88.6
Mean Accuracy ¹		78.3	82.0	79.2	67.2	85.7
Median Accuracy ¹		78.8	82.8	79.0	68.8	87.2
Max Accuracy		96.9	97.1	94.4	74.9	93.4
Human [35,9,23]		71.0	74.0	67.2		79.5
Aliradi et al. [1]	2018	80.6	88.6			
Moujahid et al. [21]	2018	88.2	88.2			
Lopez et al. [17]	2018	68.4	66.5			
Yan et al. [32]	2018	77.6	78.5			
Robinson et al. [25]	2018	82.4	86.6			
Kohli et al. [12]	2018	96.9	97.1	94.4		
Zhou et al. [36]	2018	82.8	85.7	81.4		
Mahpod et al. [20]	2018	79.8	87.0	76.6		
Zhao et al. [35]	2018	81.5	82.5	81.7		84.5
Wang et al. [26]	2018				69.5	
Xia et al. [29]	2018					90.7
Yang et al. [33]	2017	88.6	90.3			93.4
Chen et al. [3]	2017	83.3	84.3			
Lu et al. [18]	2017	83.5	84.3			
Patel et al. [22]	2017	78.7	80.6			
Kohli et al. [11]	2017	96.1	96.2	89.5		
Laiadi et al. [13]	2017			83.2	54.8	
Duan et al. [7]	2017				66.6	
Dahan et al. [5]	2017				65.0	
Li et al. [14]	2017				74.9	
Wang et al. [27]	2017				68.8	
Zhou et al. [37]	2016	78.8	75.7			
Xu et al. [31]	2016	77.9	77.1			
Liu et al. [15]	2016	77.9	81.4	75.9		
Zhang et al. [34]	2016					89.7
Robinson et al. [24]	2016				71.0	
Qin et al. [23]	2015					85.4
Lu et al [19]	2014	69.9	76.5	66.5		72.1
Fang et al. [9]	2010			70.7		

In Table 3, we show summary statistics across a large number of published kinship verification models on the various data sets, and show that the FSP

¹ Some results have been omitted due to page length constraints. These statistics are calculated using the full table of results, available in the Arxiv version of the paper.



Fig. 7. Examples of non-kinship pairs that the classifier incorrectly predicted as being from the same photo. Pairs shown, from left top right, are from the KinFaceW-I, Families in the Wild and Cornell KinFace data sets respectively.

classifier performs kinship classification as well or better than most of them. We also achieve high accuracies for the tri-subject verification data set beating many published models and near the state-of-the-art published results (Table 3, TSKinFace 88.6%). In this instance there is no significant gender bias in accuracy of verification of kinship for the gender of the child. Finally, we can see that the FSP classifier consistently outperforms human classification of kinship. We can expect that the deep learning models will have inadvertently learnt the FSP signal as a means to solve the kinship verification task, but can only speculate if human classifiers uses the same type of information to simplify the task. Examples of positive kin pairs and triplets found by the FSP classifier are shown in Figure 6. A selection of pairs the classifier falsely predicted as being from the same photo are displayed in Figure 7.

5 Conclusions

In this work, we have applied a ‘from same photograph’ (FSP) classifier as a naive classifier of kinship to five data sets regularly used for kinship verification tasks, and thereby have estimated the degree to which the FSP signal contaminates these data sets. The FSP classifier performs amongst the best published kinship classifier models, despite not being trained for this primary task. It is likely that deep neural network models built with the intention to verify kinship, are instead primarily using this much easier to detect FSP signal. We have also obtained a new data set for training classifiers to detect facial crops from the same photograph to begin to address this problem.

Furthermore, it is important to note that there are many other spurious signals that one should expect in kinship verification data sets. For instance, deep learning kinship classifiers would also be expected to learn biases in distributions of age, gender and particularly ancestral backgrounds. Due to the way the FSP classifier has been trained it is blind to many of these other confounding non-kinship signals contained within existing kinship data sets.

We recommend that an FSP classifier should be an important part of kinship data set production pipelines: either to ensure the FSP signal is removed entirely, or is balanced between positive and negative kin pairs. It should be considered that inappropriate construction of negative sets can introduce biases

by overly simplifying the task, such as generating father-father-child tri-subjects, mismatched ancestral backgrounds, and implausible age distributions (such as a child older than the parents). Kinship verification is a challenging task made more difficult by inherent biases in how data occurs in the wild. Sharing a photograph does not make us relatives, but we are likely to share a photograph if we are.

References

1. Aliradi, R., Belkhir, A., Ouamane, A., Elmaghraby, A.S.: Dieda: discriminative information based on exponential discriminant analysis combined with local features representation for face and kinship verification. *Multimedia Tools and Applications* pp. 1–18 (2018)
2. Chatfield, K., Simonyan, K., Vedaldi, A., Zisserman, A.: Return of the devil in the details: Delving deep into convolutional nets. *arXiv preprint arXiv:1405.3531* (2014)
3. Chen, X., An, L., Yang, S., Wu, W.: Kinship verification in multi-linear coherent spaces. *Multimedia Tools and Applications* **76**(3), 4105–4122 (2017)
4. Corporation, M.: Bing Image Search API (2018), <https://azure.microsoft.com/en-us/services/cognitive-services/bing-image-search-api/>
5. Dahan, E., Keller, Y., Mahpod, S.: Kin-verification model on fiw dataset using multi-set learning and local features. In: *Proceedings of the 2017 Workshop on Recognizing Families In the Wild*. pp. 31–35. ACM (2017)
6. Doersch, C., Gupta, A., Efros, A.A.: Unsupervised visual representation learning by context prediction. In: *International Conference on Computer Vision (ICCV)* (2015)
7. Duan, Q., Zhang, L.: Advnet: Adversarial contrastive residual net for 1 million kinship recognition. In: *Proceedings of the 2017 Workshop on Recognizing Families In the Wild*. pp. 21–29. ACM (2017)
8. Fang, R., Gallagher, A.C., Chen, T., Loui, A.: Kinship classification by modeling facial feature heredity. In: *Image Processing (ICIP), 2013 20th IEEE International Conference on*. pp. 2983–2987. IEEE (2013)
9. Fang, R., Tang, K.D., Snavely, N., Chen, T.: Towards computational models of kinship verification. In: *Image Processing (ICIP), 2010 17th IEEE International Conference on*. pp. 1577–1580. IEEE (2010)
10. King, D.E.: Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research* **10**, 1755–1758 (2009)
11. Kohli, N., Vatsa, M., Singh, R., Noore, A., Majumdar, A.: Hierarchical representation learning for kinship verification. *IEEE Transactions on Image Processing* **26**(1), 289–302 (2017)
12. Kohli, N., Yadav, D., Vatsa, M., Singh, R., Noore, A.: Supervised mixed norm autoencoder for kinship verification in unconstrained videos. *IEEE Transactions on Image Processing* (2018)
13. Laiadi, O., Ouamane, A., Benakcha, A., Taleb-Ahmed, A.: Rfiw 2017: Lpq-sieda for large scale kinship verification. In: *Proceedings of the 2017 Workshop on Recognizing Families In the Wild*. pp. 37–39. ACM (2017)
14. Li, Y., Zeng, J., Zhang, J., Dai, A., Kan, M., Shan, S., Chen, X.: Kinnet: Fine-to-coarse deep metric learning for kinship verification. In: *Proceedings of the 2017 Workshop on Recognizing Families In the Wild*. pp. 13–20. ACM (2017)
15. Liu, Q., Puthenputhussery, A., Liu, C.: A novel inheritable color space with application to kinship verification. In: *Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on*. pp. 1–9. IEEE (2016)

16. López, M.B., Boutellaa, E., Hadid, A.: Comments on the “kinship face in the wild” data sets. *IEEE transactions on pattern analysis and machine intelligence* **38**(11), 2342–2344 (2016)
17. Lopez, M.B., Hadid, A., Boutellaa, E., Goncalves, J., Kostakos, V., Hosio, S.: Kinship verification from facial images and videos: human versus machine. *Machine Vision and Applications* **29**(5), 873–890 (2018)
18. Lu, J., Hu, J., Tan, Y.P.: Discriminative deep metric learning for face and kinship verification. *IEEE Transactions on Image Processing* **26**(9), 4269–4282 (2017)
19. Lu, J., Zhou, X., Tan, Y.P., Shang, Y., Zhou, J.: Neighborhood repulsed metric learning for kinship verification. *IEEE transactions on pattern analysis and machine intelligence* **36**(2), 331–345 (2014)
20. Mahpod, S., Keller, Y.: Kinship verification using multiview hybrid distance learning. *Computer Vision and Image Understanding* (2017)
21. Moujahid, A., Dornaika, F.: A pyramid multi-level face descriptor: application to kinship verification. *Multimedia Tools and Applications* pp. 1–20 (2018)
22. Patel, B., Maheshwari, R., Raman, B.: Evaluation of periocular features for kinship verification in the wild. *Computer Vision and Image Understanding* **160**, 24–35 (2017)
23. Qin, X., Tan, X., Chen, S.: Tri-subject kinship verification: Understanding the core of a family. *IEEE Transactions on Multimedia* **17**(10), 1855–1867 (2015)
24. Robinson, J.P., Shao, M., Wu, Y., Fu, Y.: Families in the wild (fiw): Large-scale kinship image database and benchmarks. In: *Proceedings of the 2016 ACM on Multimedia Conference*. pp. 242–246. ACM (2016)
25. Robinson, J.P., Shao, M., Wu, Y., Liu, H., Gillis, T., Fu, Y.: Visual kinship recognition of families in the wild. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2018)
26. Wang, S., Ding, Z., Fu, Y.: Cross-generation kinship verification with sparse discriminative metric. *IEEE transactions on pattern analysis and machine intelligence* (2018)
27. Wang, S., Robinson, J.P., Fu, Y.: Kinship verification on families in the wild with marginalized denoising metric learning. In: *Automatic Face & Gesture Recognition (FG 2017)*, 2017 12th IEEE International Conference on. pp. 216–221. IEEE (2017)
28. Wu, X., Boutellaa, E., López, M.B., Feng, X., Hadid, A.: On the usefulness of color for kinship verification from face images. In: *Information Forensics and Security (WIFS)*, 2016 IEEE International Workshop on. pp. 1–6. IEEE (2016)
29. Xia, C., Xia, S., Zhou, Y., Zhang, L., Shao, M.: Graph based family relationship recognition from a single image. In: *Pacific Rim International Conference on Artificial Intelligence*. pp. 310–320. Springer (2018)
30. Xia, S., Shao, M., Fu, Y.: Kinship verification through transfer learning. In: *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume Three*. pp. 2539–2544. IJCAI’11, AAAI Press (2011)
31. Xu, M., Shang, Y.: Kinship verification using facial images by robust similarity learning. *Mathematical Problems in Engineering* **2016** (2016)
32. Yan, H.: Learning discriminative compact binary face descriptor for kinship verification. *Pattern Recognition Letters* (2018)
33. Yang, Y., Wu, Q.: A novel kinship verification method based on deep transfer learning and feature nonlinear mapping. *DEStech Transactions on Computer Science and Engineering* **aiea** (2017)
34. Zhang, J., Xia, S., Pan, H., Qin, A.: A genetics-motivated unsupervised model for tri-subject kinship verification. In: *Image Processing (ICIP)*, 2016 IEEE International Conference on. pp. 2916–2920. IEEE (2016)

35. Zhao, Y.G., Song, Z., Zheng, F., Shao, L.: Learning a multiple kernel similarity metric for kinship verification. *Information Sciences* **430**, 247–260 (2018)
36. Zhou, X., Jin, K., Xu, M., Guo, G.: Learning deep compact similarity metric for kinship verification from face images. *Information Fusion* (2018)
37. Zhou, X., Shang, Y., Yan, H., Guo, G.: Ensemble similarity learning for kinship verification from facial images in the wild. *Information Fusion* **32**, 40–48 (2016)