Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 198 (2022) 225–230

Procedia Computer Science

www.elsevier.com/locate/procedia

International Workshop on Artificial Intelligence Methods for Smart Cities (AISC 2021)
November 1-4, 2021, Leuven, Belgium

Kinship recognition: how far are we from viable solutions in smart environments?

Carmen Bisogni[a,*], Fabio Narducci[a]

[a] *University of Salerno, Via Giovanni Paolo II, 132 - 84084 Fisciano (SA), Italy*

Abstract

Automated kinship recognition from face is a relatively recent problem that is mainly studied by the application of Deep Learning techniques. Despite the impact that an accurate kinship recognition algorithm can reach in a controlled environment, its applicability in smart environments is limited due to the degradation of performances. In this study we investigate the limitations of recent approaches that lead to a difficult applicability in a real case use. We present several tests on Siamese Neural Networks (SNN) based on a VGGFace architecture to solve both the kinship-vs-not-kinship recognition and the kind-of-kinship recognition. To perform our tests we used two popular kinship recognition Datasets that are Faces in the Wild and KinFace-II, respectively. To examine the behavior of the SNNs in a real scenario, we applied them, properly trained on the above mentioned datasets, to a popular TV show in which the aim is to discover kinship in a set of people. The weaknesses demonstrated in those tests have confirmed that the recent literature and algorithm to solve the kinship recognition problem are still far to achieve the high performances required in a smart environment.

Keywords: kinship recognition; face recognition; smart environment; siamese networks; deep learning; face biometrics

1. Introduction

Kinship recognition is a relatively new branch of Biometrics developed only in the last few years [1]. The human ability to detect kinship between individuals is a well-known problem in psychology and cognitive sciences [2] [3]. This ability has been developed as an evolutive response to kin selection, that implies kin recognition, mediated in many species by facial traits [4]. As in several applications involving human understanding, also in this case the trend is the use of Artificial Intelligence (AI) techniques to solve the kinship recognition task, in particular by using deep learning. If, on the one hand the use of deep learning achieved high performances in recognizing the existence of kinship between individuals [5], on the other hand we can also observe that without intermediate steps as gender

* Corresponding author.
E-mail address: cbisogni@unisa.it

10.1016/j.procs.2021.12.232

recognition and age estimation, performances on this task can dramatically decrease [6], achieving results similar to the human accuracy between 60% and 75% [7]. Those results highlight some limitations in generalization ability of deep learning techniques in kinship recognition. If we consider a practical scenario, smart environments where kinship recognition is meant as a support to missing children searching or to detect relations between groups of people [8], how much the recent solutions are effective? The contributions of this work are as follows:

- Discuss about recent techniques in kinship recognition and their applicability in a real-case use;
- Present a set of experiments in which we perform kinship recognition by Siamese Neural Networks;
- Analysing the dependencies between recognition accuracy and kind of relation.

2. Related works

One of the most common analysis to find if two individuals are related involves blood group and DNA [9]. This kind of analysis is very effective in forensics, but its application in smart environment is strongly limited. When subjects are not cooperative this is not feasible and it is inapplicable in surveillance scenarios. Since also humans recognize their pairs by faces, the analysis of the facial traits to discover kinship became popular. In particular the techniques involved in kinship recognition can be split in two main categories: Features based and Machine Learning (ML) based. The work of Guo et al. [10] avoids the machine learning by creating graphs of families. Here, the features similarities are organized in graphs that increase or decrease the probability of kinship. The facial features are also used in [11], in which the facial parts are essential to find similarities between relatives. On the contrary, in [12] the face is analysed in its entirety, through a pyramid multi-level face descriptor. Much more popular are the methods using Machine Learning: from traditional ML techniques as K-Nearest Neighbor and Support Vector Machines [13] [14] to recent deep learning techniques [15] [16] [5]. However, the most significant question to be solved is if those solutions are applicable in real cases and in particular in smart environments. The first problem to be solved is about the source of image. For a realistic scenario, the images should be *in-the-wild*. Such an approach can be found in [17] where a pre-trained Convolutional Neural Network (CNN) has been fine-tuned to be applied to a new large dataset and its performance are evaluated in-the-wild and compared with human performances. The use of pre-trained models is not new to solve kinship recognition [18]. However, the limited availability of labeled images leads to several problems in the use of deep architectures, that require a large amount of data. Another difference between a classical Machine Learning problem is the use of Siamese Neural Networks (SNN), very popular in recent kinship recognition literature [6] [19] [20]. In SNN there are two identical network architectures that work in parallel to extract features from the image to be compared. Then, those features are concatenated and passed through fully connected layers to perform the final classification. Since the literature reveals that SNN represent a trend of the last years to solve the Kinship recognition problem, in this work we will build a typical SNN architecture, training it on popular kinship recognition datasets. We will test it on real cases to stress the configuration and to highlight weaknesses of those architectures that represent the gap between in-lab accuracy and practical uses in smart environments.

3. Experimental Environment

A very popular dataset used to study kinship is the Families In the Wild (FIW) Dataset, presented in [17]. This dataset collects over 11,932 family photos of 1,000 families and it is the biggest dataset of its kind. FIW was also used for a big competition to find if two given faces are related [1]. The leader board shown an accuracy of 0.923 on the task, however there are no information about the generalization ability reached by the leader performance because codes are not available. Regarding the recognition of the kind of relation, one of the most used dataset is KinFaceW-II [21]. KinFaceW-II is the new version of KinFaceW-I and it contains 250 pairs for each of the following relation: Father-Son (F-S), Father-Daughter (F-D), Mother-Son (M-S), and Mother-Daughter (M-D). In Figure 1 example images of the two datasets are shown. The two datasets were developed to solve two different tasks, the existence of kinship in the case of FIW, that we will call NK problem and the kind of kinship in the case of KinFaceW-II, that we will call KK problem. For this reason we will use both the dataset to evaluate the proposed SNN.

[1] https://www.kaggle.com/c/recognizing-faces-in-the-wild/data

Fig. 1: On the left, images from the Michael Jackson's family by FIW. On the right images of KinFaceW-II with various relations.

The SNN we use is composed of two VGGFace [22] networks work in parallel. In particular, the VGGFace we use have the structure of VGG16. This means we have 5 convolutional blocks followed by max-pooling for a total of 13 convolutional layers. In the classical architecture, those are followed by 3 fully connected layers. Since the two networks share the same fully connected layers, and preceded by the combination of the extracted features, their number and characteristic are different in NK and KK problems. In particular, we perform the combination using the following squared difference that is known to enhance the differences between the two sets of features:

$$C_1 = (x_1 - x_2)^2, \qquad C_2 = [C_1 \quad x_1 * x_2] \tag{1}$$

where x_1 and x_2 are the features array, the output of the two networks, and C_2 is the resulting final combined array. After the combination, the NK architecture is followed by a fully connected layer of 100 units and ReLu activation, a dropout layer of rate 0.01, and a final fully connected layer to obtain the binary response, with sigmoid activation. On the other hand, the KK architecture is more complex because it has to distinguish 5 different classes, including the classes of KinFaceW-II and not-in-relation (NR) class. For this reason, here we have 3 fully connected layers of 1024, 512 and 256 units, with ReLu as activation function, each one followed by a Dropout layer of rate 0.03 and finally a fully connected layer of 5 units with softmax as activation function. The two architectures are shown in Figure 2.

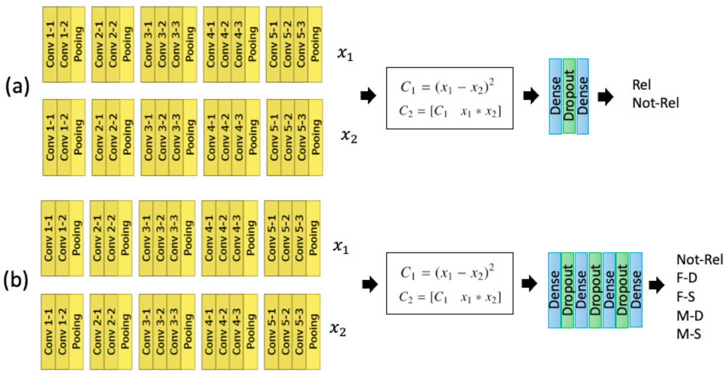

Fig. 2: The SNN architectures.(a) Kinship binary recognition problem (KN), (b) Kind of Kinship recognition problem (KK).

Since the amount of data is very limited, both for NK and KK problem, we adopt common solutions to avoid underfitting and overfitting on data. In both cases, we use the weights of FaceNet [23], pre-trained on 2,622 identities for a total of 982,803 face images. We use those weights in the VGG16 convolutional layers, thus reducing the computational time required to train just the fully connected layers. For the sake of completeness, we also tried to train the convolutional layers. As expected, the experiment reveals that the amount of data is insufficient to obtain satisfactory results. Then, both for NK than KK we also augment the amount of data in FIW and KinFAceW-II, respectively. On each image we perform horizontal and vertical shift, brightness variations, zoom-in/zoom-out, channel shift, horizontal and vertical flip, rotation, fill mode, color jitter, addition of noise, Gaussian, median and blur filter. This procedure is a standard in deep learning approaches as reported in [24]. In our case, 110 transformation are applied to each image. FIW went from 11,932 images to 1,312,520 images and KinFace from 1,000 to 110,000 images.Then, the obtained data are split in 80% for training and validation and 20% for testing, for both datasets. We ensure that the augmented images obtained from the same original image were all in the same subset. In the case of FIW, the authors of the competition provide both the positive than negative samples. In the case of KinFaceW-II the authors provide only the couple in relation. We use all the combination of not-related couples to extract the not-related class. In both cases, data in related and not related class are unbalanced. We did not try to avoid this unbalance, since this is the same environment that is present in a real case-use (the probability that two random people on a street are related is considerably lower than the probability that they are related).

The results we obtain on the two networks, together with the details about the training are summarised in Table 1 (a).

Table 1 – Results of NK and KK.

(a) The details of the two networks reported in terms of Learning Rate (LR), Training Time (TT) and number of step per epoch (S/E). In particular, results of NK are on FEW and results of KK on Kinface-II

SNN	Epoch	S/E	LR	TT	Acc
NK	112	222	10^{-5}	4.5 hours	**0.75**
KK	30	200	10^{-5}	3 hours	**0.72**

(b) The confusion matrix of KK problem (on KinFace-II).

Relation	F-D	F-S	M-D	M-S	NR
F-D	**0.52**	0.22	0.1	0.02	0.14
F-S	0.16	**0.72**	0.02	0.02	0.08
M-D	0	0.02	**0.76**	0.06	0.16
M-S	0.08	0.1	0.26	**0.38**	0.18
NR	0.07	0.03	0.07	0.04	**0.79**

As we can notice, there are not relevant differences between the performance of the two problems. However we have to highlight that more relations are included in NK. On the other hand, KK is not only able to find the presence of a relation but also its kind. The confusion matrix of KK is in Table 1(b). Here we can notice that the higher accuracy is reached in the Not-Related class, this means that the method is more able to find if there is a relation than the kind of relation. This is also due to the dataset unbalance. Among the kind of relations, the best accuracy is obtained for Father-Son and Mother-Daughter classes. This is not a surprisingly result. In fact, the gender of the subjects is the same in both cases. We can then assume that the network is implicitly learning to detect when the subjects have the same gender and what is the gender (there is no misunderstanding between F-S and M-D).

At this point, we want to test the method obtained to solve the 5 classes problem on a real scenario. To do so, we select 4 random episodes of a popular Italian television show (the name of the TV show is copyrights protected) that are publicly available online. In this TV show, a competitor has to guess the kin of person in a set of 8 candidates. The competitor can also ask for the kind of relation of the person (but not the gender of the kin, e.g., daughter or son, mother of father). If the competitor ask for the kind of relation the problem is NK, otherwise is KK. Since the 8 candidates have different age ranges, known the kind of relation is vital for the competitor, for this reason the kind of relation is often required by the competitor. Each episode is representative of a kind of relation (F-D, F-S, M-D, M-S). In total we have 32 couple to test, one for each relation and 28 not related. The class unbalance is part of the game. It is clear that for each episode one couple is in relation and the other should be classified as NR. The knowledge potentially inferred by this experiment lies in three main points:

- If the SNN is able to successfully find the existence of a relation;
- If the SNN can correctly classify the right couple, regardless of the other results;
- If the SNN is able to implicitly understand the gender, regardless of the existence of the relation.

We present the results of the first two points in Table 2. In particular, here we can notice that the only relation detected is M-D that also presented the most performing scenario in general. In opposition to the controlled scenario, here we do not see relevant improvement about the relation F-S. On the opposite, it seems that the not-in-relation condition is more easily detected when there is a Daughter. All of those results were not previously highlighted, when we apply the SNN to the test set. This makes us pay more attention on the difference between the environments when a SNN, trained on family photos of state-of-the-art dataset, want to be used in a real scenario.

Table 2 – Results of the study on a real scenario.

True relation	Correct NR	Relation Detected	Total test accuracy
F-D	42.85%	No	0.38
F-S	28,57%	No	0.38
M-D	71.42%	Yes	0.75
M-S	14.28%	No	0.25

Regard the last point, we examine the problem from the point of view of the gender in both the images. In particular, mothers were confused with fathers in 31,25% of cases and fathers with mothers in 25% of cases. This means that the gender of the parents is implicitly detected with a mean accuracy of 71.85%. Daughters are confused with sons in 25% of cases and sons with daughters in 43.75% of cases. This means that the gender of the sons is implicitly detected with a mean accuracy of 69.44%. This kind of results suggest us that the SNN we built is, implicitly but strongly, dependent on the gender of the images. It can be both a strength and a weakness of this kind of methods. What happens if we consider more relations when the gender of the subjects is all the same? (e.g. grandmother-granddaughter, uncle and niece, etc.).

4. Discussion

Taking into account the considerations resulting from the experiments made in the previous section, we want here discuss some similarities between the obtained results and the state-of-the-art methods for kinship recognition presented in Section 2. We split our considerations on the basis of the method used to recognize the kinship. Feature based techniques are in [11], [12] and [14]. Here, apart of [11] that not solve the multi-class problem, we can observe the same behaviour of our case. In other words, the best classified kinship are M-D and F-S. In the case of Deep Learning approach we can found very different behaviours, that are in most cases dataset-dependent. Also in [15] the best classes are F-S and M-D but with very few differences with the others. Then we have works that perform better with one or another parent. [16] performs better on the Fathers' relations, and [17] and [18] on the Mothers' relations. Then, there are the Siamese Network as the once we used, here the literature solves more the NK problem than the KK problem, on which we can only observe from [5] that the result variations are strongly dependent on the dataset. From those considerations, we can observe that similar techniques give similar unbalanced results and that maybe in the future the classes should be separated depending on the gender to ensure also to other classes a higher accuracy. We can also observe that, even if those methods do not perform a gender recognition, they are strongly and implicitly affected by the gender in the process of recognizing the kinship, both feature-based and deep-learning and SNN. Another observation we can make is that there is a strongly dependence between the performances and the data used in the training or to set the parameters of the method. In fact, in literature we noticed more than once that the distribution of the better accuracy over the classes is dependent on the dataset. The extreme case is observable in our experiment in the real scenario, where the performances drastically decrease and the expected behaviour on the classes is not observable. Those considerations lead to the conclusions that the state-of-the-art and the way of proceeding in solving kinship recognition problem lacks in several aspects that should be considered to better understand deep pattern learning when solving kinship recognition problems. As an example, the implicit learning of the gender may lead to the conviction to use balanced data (for the kinship) that results to be unbalanced for other inner pattern (the gender). Those aspect that must be further investigated can significantly affect accuracy of those system in a smart environment. In fact, in a smart environment, errors due to a hidden pattern can easily be propagated to further models in the environment without a real control over them. For this reason, to be applicable in this field, the kinship recognition should work in synergy with other image understanding techniques to produce a final accuracy and reliability score.

5. Conclusion

This work explores the potential and the limitation of kinship recognition in practical smart environments through computer vision approaches. Kinship recognition from faces is an inherently harder task than biometric face recognition. Even if significant improvements have been achieved by machine and deep learning techniques in the analysis of visual features for biometric identification/authentication, the same neural models do not exhibit a comparable level of performances in detecting the kin relation within a group of people. When supported by the knowledge on the gender and the age of the observed people, laboratory experiments have demonstrated desirable accuracy and precision of neural models based on convolutional neural networks. In this work a Siamese network built on VGGFace has been proposed. By replicating the CNN model twice, which are fed with face images of two people, during the training the network should learn how to relate two people, distinguishing among four classes of relations that are Father-Son, Father-Daughted, Mother-Son and Mother-Daughter. Experimental results reported in this work show strengths and weaknesses of the proposed approach. More precisely, the experimentation performed in this work reveals how intrinsically hard is recognizing the kin of a person and how significantly far are the state-of-the-art solutions from a feasible adoption in-the-wild scenarios and, consequently, in smart environment. The observations inferred by the experiments in this work can be considered as a preliminary step for further and more aimed considerations when dealing with kinship recognition in practical scenarios.

References

[1] J. Robinson, M. Shao, Y. Fu, Survey on the analysis and modeling of visual kinship: A decade in the making, in: IEEE Transactions on Pattern Analysis & Machine Intelligence, Vol. 01, 2020, pp. 1–1.
[2] G. Kaminski, S. Dridi, C. Graff, E. Gentaz, Human ability to detect kinship in strangers' faces: effects of the degree of relatedness, in: Proceedings of the Royal Society B: Biological Sciences, Vol. 276, 2009, pp. 3193–3200.
[3] A. Alvergne, R. Oda, C. Faurie, A. Matsumoto-Oda, V. Durand, M. Raymond, Cross-cultural perceptions of facial resemblance between kin, in: booktitle of vision, Vol. 9, 2009, pp. 1–10.
[4] M. J. E. Charpentier, M. Harté, C. Poirotte, J. M. de Bellefon, B. Laubi, P. M. Kappeler, J. P. Renoult, Same father, same face: Deep learning reveals selection for signaling kinship in a wild primate, in: Science Advances, Vol. 6, 2020, pp. 1–10.
[5] M. Wang, X. Shu, J. Feng, X. Wang, J. Tang, Deep multi-person kinship matching and recognition for family photos, in: Pattern Recognition, Vol. 105, 2020, pp. 107–342.
[6] A. Nandy, S. S. Mondal, Kinship verification using deep siamese convolutional neural network, in: 2019 14th IEEE International Conference on Automatic Face Gesture Recognition, 2019, pp. 1–5.
[7] M. Bordallo Lopez, A. Hadid, E. Boutellaa, J. Goncalves, V. Kostakos, S. Hosio, Kinship verification from facial images and videos: human versus machine, in: Machine Vision and Applications, Vol. 29, 2018, pp. 1–18.
[8] H. Yoo, T. Eom, J. Seo, S.-I. Choi, Detection of interacting groups based on geometric and social relations between individuals in an image, in: Pattern Recognition, Vol. 93, 2019, pp. 498–506.
[9] G. Gunther, I. Wirth, Genetic kinship investigation from blood groups to dna markers, in: Transfus Med Hemother, Vol. 39(3), 2012, pp. 163–175.
[10] Y. Guo, H. Dibeklioglu, L. Van Der Maaten, Graph-based kinship recognition, in: 2014 22nd International Conference on Pattern Recognition, 2014, pp. 4287–4292.
[11] G. Guo, X. Wang, Kinship measurement on salient facial features, in: IEEE Transactions on Instrumentation and Measurement, Vol. 61, 2012, pp. 2322–2325.
[12] A. Moujahid, F. Dornaika, A pyramid multi-level face descriptor: Application to kinship verification, in: Multimedia Tools Appl., Vol. 78, 2019.
[13] R. Fang, K. D. Tang, N. Snavely, T. Chen, Towards computational models of kinship verification, in: 2010 IEEE International Conference on Image Processing, 2010, pp. 1577–1580.
[14] H. Yan, Learning discriminative compact binary face descriptor for kinship verification, in: Pattern Recognition Letters, Vol. 117, 2019, pp. 146–152.
[15] J. Lu, J. Hu, Y.-P. Tan, Discriminative deep metric learning for face and kinship verification, in: IEEE Transactions on Image Processing, Vol. 26, 2017, pp. 4269–4282.
[16] X. Zhou, K. Jin, M. Xu, G. Guo, Learning deep compact similarity metric for kinship verification from face images, in: Information Fusion, Vol. 48, 2019, pp. 84–94.
[17] J. P. Robinson, M. Shao, Y. Wu, H. Liu, T. Gillis, Y. Fu, Visual kinship recognition of families in the wild, in: IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 40, 2018, pp. 2624–2637.
[18] F. Dornaika, I. Arganda-Carreras, O. Serradilla, Transfer learning and feature fusion for kinship verification, in: Neural Computing and Applications, Vol. 32, 2019, pp. 7139–7151.
[19] J. Yu, M. Li, X. Hao, G. Xie, Deep fusion siamese network for automatic kinship verification, in: 15th IEEE International Conference on Automatic Face and Gesture Recognition, 2020, pp. 892–899.
[20] J. Yu, G. Xie, M. Li, X. Hao, Retrieval of family members using siamese neural network, in: 15th IEEE International Conference on Automatic Face and Gesture Recognition, 2020, pp. 882–886.
[21] J. Lu, X. Zhou, Y.-P. Tan, Y. Shang, J. Zhou, Neighborhood repulsed metric learning for kinship verification, in: IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 36, 2014, pp. 331–345.
[22] O. M. Parkhi, A. Vedaldi, A. Zisserman, Deep face recognition, in: Proceedings of the British Machine Vision Conference, 2015, pp. 41.1–41.12.
[23] F. Schroff, D. Kalenichenko, J. Philbin, Facenet: A unified embedding for face recognition and clustering, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 815–823.
[24] C. Shorten, T. Khoshgoftaar, A survey on image data augmentation for deep learning, in: Big Data, Vol. 6, 2019, pp. 1–48.