Who Do I Look Like?
Determining Parent-Offspring Resemblance via Gated Autoencoders

Afshin Dehghan, Enrique G. Ortiz, Ruben Villegas, Mubarak Shah
Center for Research in Computer Vision, University of Central Florida
{adehghan, eortiz}@cs.ucf.edu, ruben.e.villegas@knights.ucf.edu, shah@crcv.ucf.edu

Abstract

Recent years have seen a major push for face recognition technology due to the large expansion of image sharing on social networks. In this paper, we consider the difficult task of determining parent-offspring resemblance using deep learning to answer the question "Who do I look like?" Although humans can perform this job at a rate higher than chance, it is not clear how they do it [2]. However, recent studies in anthropology [24] have determined which features tend to be the most discriminative. In this study, we aim to not only create an accurate system for resemblance detection, but bridge the gap between studies in anthropology with computer vision techniques. Further, we aim to answer two key questions: 1) Do offspring resemble their parents? and 2) Do offspring resemble one parent more than the other? We propose an algorithm that fuses the features and metrics discovered via gated autoencoders with a discriminative neural network layer that learns the optimal, or what we call genetic, features to delineate parent-offspring relationships. We further analyze the correlation between our automatically detected features and those found in anthropological studies. Meanwhile, our method outperforms the state-of-the-art in kinship verification by 3-10% depending on the relationship using specific (father-son, mother-daughter, etc.) and generic models.

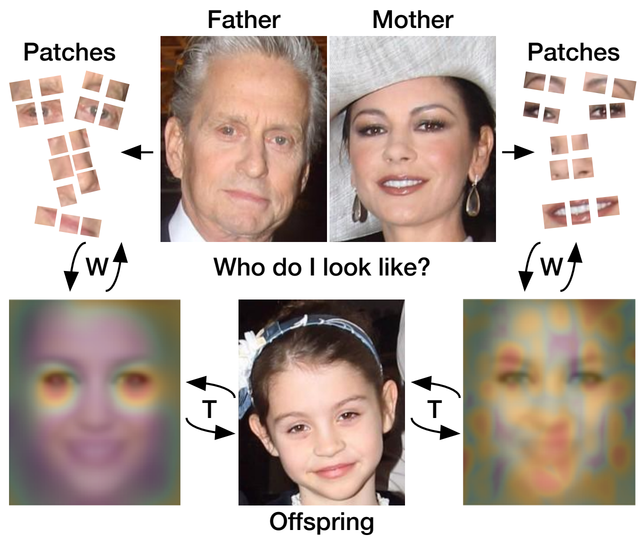

Figure 1. Our method, given a set of patches from parent-offspring pairs, learns the most discriminative features and metrics to describe parent-offspring relatedness.

1. Introduction

From the moment a baby is born, he is faced with the question,"Who does he look like more, the father or the mother?" With differing opinions, it is often difficult to ever answer this question completely. Anthropologists have tried to answer this question with quantifiable measurements for years. In [24], Naini and Moss claim to have cracked the code finding the most meaningful genetic features in determining relatedness. Further studies [2, 5, 6], claim to have correlated visual resemblance between parents and their offspring of varying ages. In this paper, we aim to bridge the gap between findings in the social sciences and computer vision to answer the age-old question, "Who do I look like?"

Sparking much research and debate, anthropology began to explore this question with Christenfeld and Hill's pioneering work [9], concluding that children do not look like parents, with the exception of one-year olds to their fathers. Brédart and French [5] later contradicted those findings stating there is a large resemblance between parents and children up to the age of 5. Subsequent studies have corroborated that offspring do in fact resemble parents more than random strangers and at different ages may resemble a particular parent more [2, 6].

When assessing the resemblance of parents and their offspring it is crucial to consider which features are the best (the most similar between parents and offspring), which we refer to as genetic features. Past anthropological studies have not only analyzed familial relationships, but more specifically the relationship between twins [7, 8, 20]. Naini and Moss [24] specifically compare facial features across 3D scans of dyzogotic and monozygotic twins. This twin method allows finding which features are genetic vs. environmentally influenced. In this work, we directly compare our automatically discovered features to the features from anthropology [24] to see if they correlate.

In the computer vision community, there has always been great interest in implementing the human visual system; however, finding a general principle that underlies most perception is still very challenging. Recently, deep learning architectures have shown great power in discovering biologically inspired features in an unsupervised fashion, and have been successful in tasks ranging from face verification to object recognition [14, 16, 17, 4, 15]. Therefore, we found these approaches well suited in discovering features that model parent-offspring resemblance. In learning the relationship between parents and their offspring, we further propose using a new generative and discriminative model based on the gated autoencoder [1, 22, 21]. Our method enables us to learn the transformation matrix and feature representation jointly through a hybrid system unlike existing approaches. Prevailing methods use simple, hand-crafted features [11, 12, 19, 28], discover the metrics between them separately [11, 19, 28], and/or only find a single metric [19, 28]. In general, none of the existing work address the questions we aim to answer with the quasi-genetic features discovered via deep learning.

While anthropological studies have found a correlation between parental effort and father-son resemblance [3, 25], and a further relationship between facial resemblance and trust [18], computer applications contrastingly aim to use facial resemblance to automate retrieval tasks. With the increasing pervasiveness of photos in social networks and in search engines, new methods for sifting through this data is essential. In these difficult scenarios, contextual information like the co-occurrence [23], position [13], and relationship [27] of individuals in an image has proven advantageous in increasing recognition rates. Recognition could be further aided by kinship relations. Besides providing contextual information, we believe familial resemblance can aid in reuniting parents with their missing children.

In this paper, we propose a method to discover the optimal features and metrics relating a parent and offspring via gated autoencoders. Moreover, we introduce a new layer that further enhances the relationship of a parent-offspring pair converging on a more discriminative function. Given our proposed method, we aim to answer two key questions from the perspective of a computer:

1. Do offspring resemble their parents?
2. Do offspring resemble one parent more than the other?

Given the answers to these questions, we can in turn conclude whether computer vision findings agree with anthropology and consequently discover which facial features lead to the best performance in parent-offspring recognition.

The rest of this paper is organized as follows: First, in Section 3, we introduce our method combining gated autoencoders with a discriminative layer to discover the best feature and metrics to describe the parent-offspring relationship. Next, in Section 5, we explore our core questions, compare our automatically discovered features to those in the anthropology literature, and evaluate the kinship verification task. Finally, in Section 6, we summarize our findings and propose some future work.

2. Related Work

In the computer vision community, most interest has been in kinship verification (family or not family). Fang *et al.* [12] first introduces the problem and postulates simple features like eye color, distances between facial parts, and skin color work well for verification. Subsequently, Xia *et al.* [28] claim that the appearance similarity between parents and their offspring is quite large, thus propose transfer learning between two photos of a parent, one young and one old, and an image of a child to close the gap. Unfortunately, in a real-world application we cannot always expect the availability of such data. Lu *et al.* [19] propose a metric learner specifically for kinship verification. Though effective, the features used are not necessarily the most discriminative. Finally, Fang *et al.* [11] present a method for kinship identification (family matching) using sparse representation-based classification for different facial components claiming them as 'genetic features'. However, they make no direct connections with existing anthropological works. In general, none of the existing works address the main questions we aim to answer.

3. Autoencoder

Recently autoencoders have shown promise in feature learning [26] and have been correlated to the way the human visual system processes imagery. Deep learning architectures provide the means to find a compact representation of the data while keeping the most important information. This property allows us to learn the most discriminative features that encode the resemblance between a parent and their offspring, which we refer to as 'genetic features'.

A simple way to train an autoencoder is minimizing the reconstruction error given a set of N randomly sampled local patches from the training set:

$$L = \sum_{n=1}^{N} \left\| \mathbf{y}^{(n)} - \mathbf{y}'^{(n)} \right\|^2, \tag{1}$$

where $\mathbf{y}^{(n)} \in \mathbb{R}^{N_y}$ represent the n^{th} image patch, $\mathbf{y}'^{(n)} = W^T W \mathbf{y}^{(n)}$, and $W \in \mathbb{R}^{N_y \times N_m}$ is the weight matrix that maps the input data to the hidden units, which is a discovered representation. N_y and N_m represent the dimension of the image patch and number of hidden units respectively. Once the learning method is finished the weights W will be used as filters for feature extraction.

4. Gated Autoencoders

Autoencoders have shown good performance in modeling the representation of a single image. However, we are interested in encoding the relationship between a pair of images. Thus, when we present a new pair of images, the hidden units will change even if the transformation between images remains the same. We move towards the use of relational autoencoders which help us learn the transformation between a pair of images, while still benefitting from the ability of autoencoders to represent the data.

One naïve way to achieve our goal in finding the relationship between two images is a standard feature learning on the concatenation of the two images. Although this feature learning may capture the transformation between images, it is still dependent on the content of each individual image. A better representation should be dependent on the input data as well as the transformation between them. Gated autoencoders [1, 21, 22] allow encoding the relation between images and frees the network to focus only on the transformation of the images rather than the representation of each individual image. In this case the activation of the hidden units will be dependent on both inputs, $\mathbf{x}$ and $\mathbf{y}$.

As shown in Figure 2, our proposed method takes parent-offspring pairs as input, from which we extract patches. Given these patches, we first learn the mapping units $\mathbf{z}$ via a gated autoencoder, which is the generative portion (Section 4.1) that learns the new feature representation that best describes the relationship between the pair. The next stage implements our discriminative model (Section 4.2) that finds the best features to differentiate between true pairs versus wrong pairs. The final output of our system is a relatedness or resemblance statistic, which we can use for classification. The details of each stage are discussed in the remainder of this section.

4.1. Generative Training

Our generative model closely follows [1, 22] to encode the transformation between a pair of images. If we recall from autoencoders, given an input image patch, $\mathbf{y}$, the hidden units are obtained via $f_k = \sum_j w_{jk} y_j$. In gated autoencoders, it is very similar, however the weights are a linear combination of one of the inputs. For example, given a pair of local image patches, $\mathbf{x}$ and $\mathbf{y}$, if we consider $w_{jk}(x) = \sum_i^{N_x} w_{ijk} x_i$, then the weights are obtained as follows:

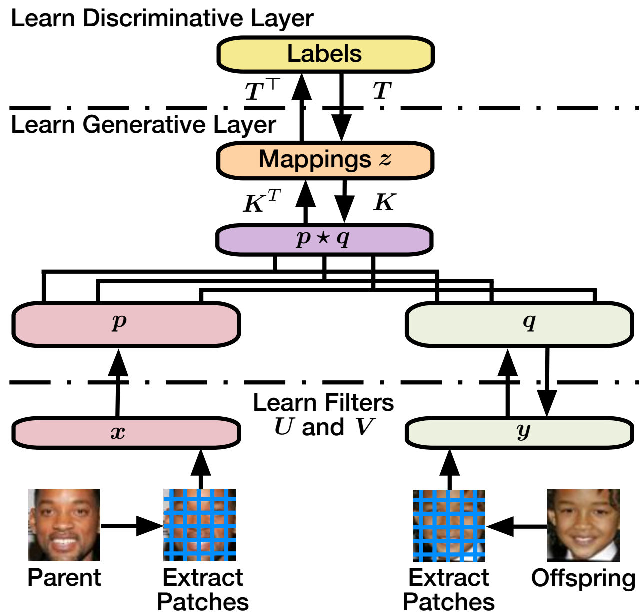

Figure 2. Gated Autoencoder Diagram. This diagram depicts the processing of a father-son pair and the process leading to the final discriminative filter that gives the relatedness prediction.

$$z_k = \sum_{j=1}^{N_y} \sum_{i=1}^{N_x} w_{ijk} x_i y_j. \tag{2}$$

We refer to the weights z_k as mapping units. Given a basis expansion of input x as well as the mapping units z one can get output vectors y' and x' similarly:

$$y'_j = \sum_{k=1}^{N_z} w_{jk}(x) z_k = \sum_{k=1}^{N_z} \sum_{i=1}^{N_x} w_{ijk} x_i z_k, \tag{3}$$

$$x'_i = \sum_{k=1}^{N_z} w_{jk}(y) z_k = \sum_{k=1}^{N_z} \sum_{j=1}^{N_y} w_{ijk} y_j z_k, \tag{4}$$

where N_x, N_y and N_z are the dimension of x, y and z respectively. Factorizing the parameter W into three matrices [22] results in the following equations:

$$w_{ijk} = \sum_{f=1}^{F} w^x_{if} w^y_{jf} w^z_{kf}$$

$$z_k = \sum_{f=1}^{F} w^z_{kf} \left(\sum_{i=1}^{N_x} w^x_{if} x_i\right)\left(\sum_{j=1}^{N_y} w^y_{jf} y_j\right), \tag{5}$$

where F is the number of hidden units. We can further simplify the equation and write it in the form of:

$$\mathbf{z} = K^T(\mathbf{p} \star \mathbf{q}), \tag{6}$$

where $\mathbf{p} = U^T\mathbf{x}$ and $\mathbf{q} = V^T\mathbf{y}$ are the hidden units, and $\star$ indicates elementwise multiplication. The columns of $U \in R^{N_x \times F}$ and $V \in R^{N_y \times F}$ contain our image filters that are learned along with $K \in R^{F \times N_z}$ from the data. It is worthwhile to mention that the mapping units, $\mathbf{z}$, encode only the transformation and not the content of each individual input.

Given $\mathbf{z}$ and $\mathbf{x}$, $\mathbf{y}'$ can be computed using Equation 3 and $\mathbf{x}'$ can be computed similarly given $\mathbf{y}$ and $\mathbf{z}$. Therefore, for learning, we simply minimize the reconstruction error using gradient-based optimization over the loss function:

$$L_{gen} = \sum_{n=1}^{N} \left\| \mathbf{y}'^{(n)} - \mathbf{y}^{(n)} \right\|^2 + \sum_{n=1}^{N} \left\| \mathbf{x}'^{(n)} - \mathbf{x}^{(n)} \right\|^2, \quad (7)$$

where N is the number of image patches used for training.

4.2. Discriminative Training

A good generative method can ensure that our model has preserved most of the information from the original training data. However, it does not necessarily give us optimal discriminative ability. Since label information is available for training, we propose to modify our objective function by adding a discriminative term, which takes into account the labels while learning the features. Therefore, we are able to learn features which are not only generative, but also discriminative, in other words able to differentiate between parent-offspring and not pairs.

A discriminative objective function computes an average loss between the predicted and ground-truth labels. Here the ground-truth labels take the values one (same family) or zero (not same family), therefore the discriminative objective function can be written as:

$$L_{disc} = \sum_{n=1}^{N} \left\| softmax(T(\mathbf{z}^{(n)})) - GT^{(n)} \right\|_1 \quad (8)$$

$$softmax(a_k) = \frac{exp(a_k)}{\sum_{k'} exp(a_{k'})} \quad k = 1, ..., K \quad \mathbf{a} \in \mathbb{R}^K,$$

where $GT \in [0, 1]$ and $T \in R^{2 \times N_z}$ is a classifier to be learned. Combining both the discriminative and generative models results in our final hybrid model:

$$L_{hybrid} = L_{gen} + \alpha L_{disc}, \quad (9)$$

where α is selected to avoid overfitting of our discriminative function. The best α is easily found by cross validation over the training data, which in our experiments came to be 0.4.

5. Experiments

In this section, we explore our two main questions:

1. Do offspring resemble their parents? (Section 5.2)
2. Do offspring resemble one parent more than the other? (Section 5.3)

Then, we compare our automatically discovered features to those found in anthropological studies (Section 5.4.1). Finally, we explore performance of our method on kinship verification (Section 5.4.2) and how well different methods produce generic kinship models (Section 5.4.3).

5.1. Experimental Setup

For all experiments involving the gated autoencoder method, we experimentally found the following optimal parameters. For each input pair, we extract 8×8 patches from an RGB image of size 64×64 (performance plateaued at this size). We set the number of filters to $F = 160$ and the number of mapping units to $N_z = 40$. During training, pairs are provided with their corresponding labels, *same* or *not same*. For training our generative model, we only input pairs with label *same*, but for training the discriminative model both positive and negative samples are required. For discriminative training, the number of negative patches are set to be equal to the number of positives. The parameter α is found through cross validation, which is 0.4. When using SVMs for classification, we use the RBF kernel with parameters selected via 4-fold cross validation.[1]

5.2. Relatedness

To explore the question, "Do offspring resemble their parents?", we look at the face identification task. In other words, can our algorithm match an offspring to the correct parent given a large gallery of parents. This experiment mimics anthropological studies [2, 6, 5], where a human subject is shown an offspring image and three images of adults and is asked to match the offspring to his/her most likely parent. In [2], all the images are taken under specific conditions which makes the decision easier for judges. All of the subjects are asked to express a neutral face and to look directly at the camera, the background is removed, and the illumination and contrast is normalized in all images. In our experiment, the problem is substantially more difficult, since we can easily show it more images and they are unconstrained from the web [11]. Moreover, this models a realistic usage scenario where an operator would want to match a missing person with a database of families.

We employ the Family 101 [11] dataset to investigate the relatedness between parents and their offspring. This

[1] For more information visit http://crcv.ucf.edu/projects/kinship/.

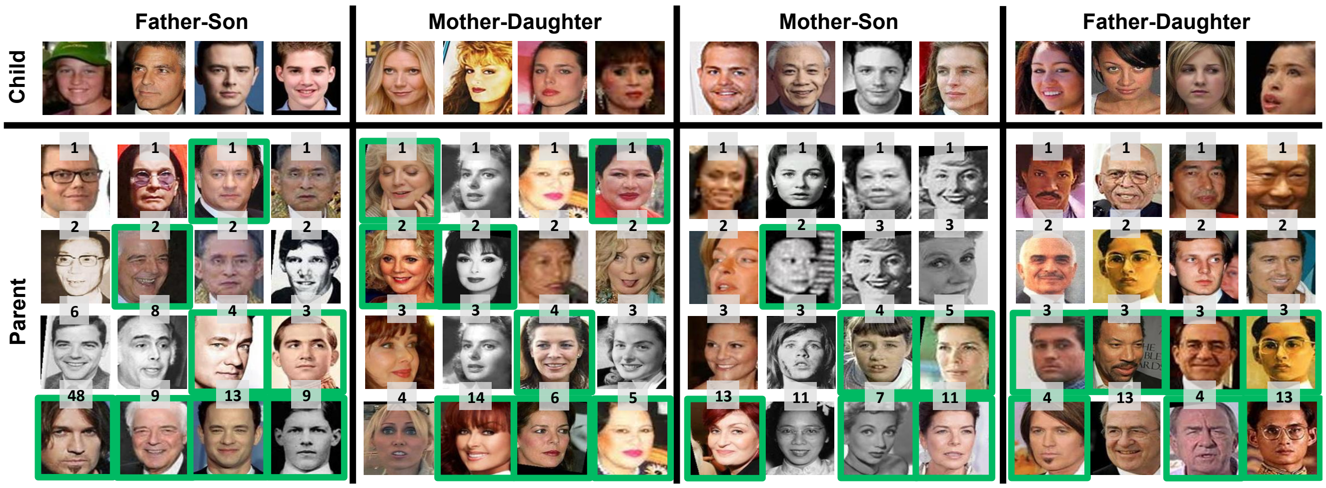

Figure 4. Rank Results. The first row of this figure shows the query child or offspring followed by the latter part of the figure showing the top ranked parent results. The first row of the parent matches shows the top rank 1 match followed by three other results, with a green box denoting the correct match.

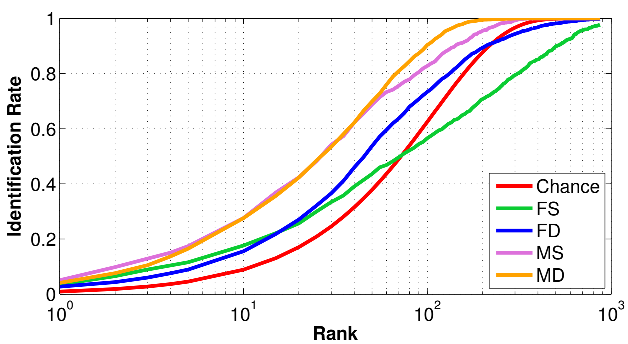

Figure 3. Rank vs. Identification. The graph shows the result of the identification task on Family 101 for each split (Father-Son, Father-Daughter, Mother-Son, and Mother-Daughter) compared to Chance.

dataset consists of 206 nuclear families, 101 unique family trees, and 14,816 images. We select 101 unique, nuclear families for our experiments and we split the set into 50 training families and 51 testing families for a total of 11,300 images. For training, we match every possible parent-offspring pair in a family for a maximum of ten images each. We apply our discriminative feature learning technique on the training relationships between all possible mother-daughter (MD), mother-son (MS), father-daughter (FD), and father-son (FS) pairs. Given a test image, we apply its respective model to compare it to every image of test parents; in other words, a test son is compared to all fathers and all mothers and the candidates are ranked based on their similarity. Figure 3 summarizes the results in terms of rank vs. the identification rate, showing that for the most part all of the splits perform better than chance (average shown - some performed better w.r.t. to own curve) at all ranks. This result corroborates anthropological findings stating that offspring resemble their parents more than random adults with a rate higher than chance [2, 6, 5].

Finally, Figure 4 shows qualitative results for our method on the Family 101 dataset. The first row depicts the child or offspring, followed by subsequent rows showing the 1st and a few other ranked matches with the correct matches highlighted in green. This figure highlights the true difficulty of this dataset with large variations in pose, illumination, image quality, and occlusions, which often causes the correct match to fall to a larger rank, *e.g*. first column at rank 48.

5.3. Resemblance

Next, we look at, “Do offspring resemble one parent more than the other?” To analyze this question, we setup the experiment in terms of training identically to the previous section, however for testing we only compare an offspring to its known parents. In this comparison, we record which parent the offspring resembles more as a frequency for each gender with a margin of 5% due to the tight distribution of predictions (e.g a FS resemblance prediction score of 60% and MS of 58% is not considered). Our results in Table 1 show that sons overall resemble their father more often than their mother and daughters resemble their mother more than their father. Our results parallel anthropological findings in which daughters resemble their mothers more often than sons do their fathers.

5.4. Genetic Features

Finally, we examine our method with respect to three factors: 1) how our discovered features compare to those

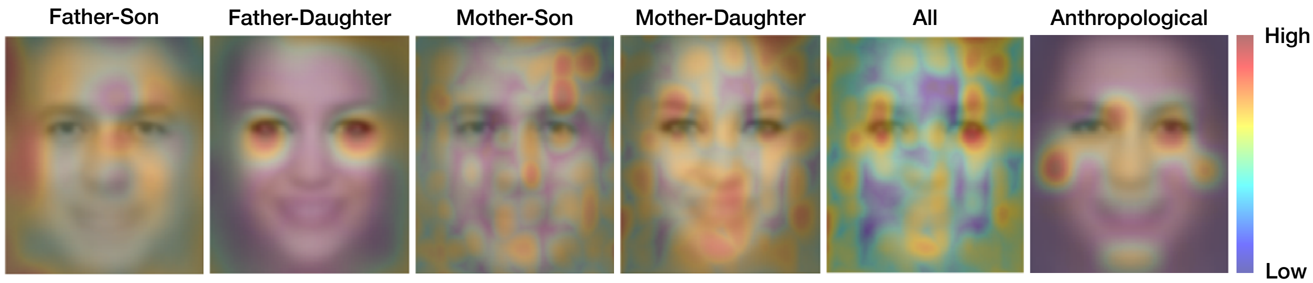

Figure 5. Genetic Features. This figure depicts the feature response of our method overlaid on the mean face from each data split separately of the Face 101 dataset. The last two show the overall scores for the trained models and an estimate of the anthropological weights from [24].

	Son	Daughter
Father	62.6%	18.5%
Mother	37.4%	81.5%

Table 1. Resemblance Results. This table shows results for the Family 101 dataset comparing a test offspring to their parents and counting how many times they match the father or the mother. Sons resemble their fathers and daughters resemble their mothers more often than not.

from anthropological studies, 2) how well our genetic features outscore the state-of-the-art in metric learning, and 3) how well the feature models generalize.

5.4.1 Computer vs. Anthropology

We first analyze the features and metrics discovered by our method compared to those determined by Naini and Moss [24]. As previously stated, Naini and Moss use the twin-method to discover genetic vs. environmental features, in other words, the features that best describe familial resemblance vs. all others. Using 3D scan measurements, they rate the importance of each facial component, for example the left-eye is rated higher than the base of the nose. Similarly, we aim to find which features our method finds most meaningful. We employ the Face 101 dataset for this task. Given this data, we train a father-son, father-daughter, mother-son, and mother-daughter model separately providing the discriminative output describing the importance of each patch for distinction during learning. Figure 6 shows automatically discovered feature filters with some strong edges and circular shapes. Next, we test each model on test data and the feature responses are overlaid on the mean test faces for each split as well as a fusion of them to show the overall response (All) in Figure 5. Overall, the eyes, chin, and parts of the forehead give a large feedback. The response to the eye locations are especially correlated to the anthropological results shown in the final image, which as-

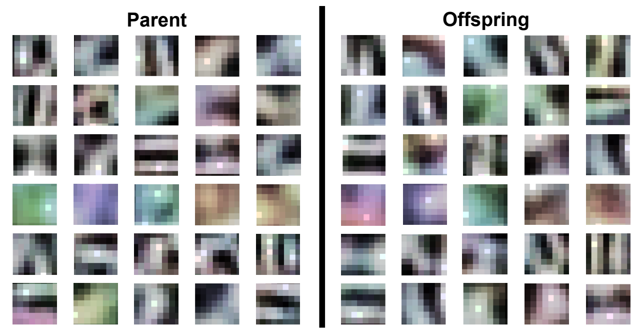

Figure 6. Discovered Filters. This figure shows sample filters automatically discovered using autoencoders trained on parent-offspring image patch pairs. The left column shows filters for parents and the right for offspring.

sign them a high weight as shown in the image on the far right. Interestingly, the mother-daughter relationship has large responses throughout the face, confirming the strong resemblance between mothers and daughters versus fathers and daughters found in [2].

5.4.2 Face Verification

Next, we explore the face verification task in order to compare how well our method performs against existing metric learning techniques and determine whether fusing the findings from anthropological studies with our method improves performance. For this task, we use the KinFaceW [19] dataset, which is comprised of two sets, KinFaceW-I with 533 parent-offspring pairs from different images and KinFaceW-II with 1,000 pairs from the same image. The data is split into 134 father-son, 156 father-daughter, 126 mother-son, and 116 mother-daughter relationships. Similar to [19], we follow a 5-fold cross-validation with balanced positive and negative pairs. To test our implementation, we were able to obtain comparable accuracy as reported in [19], often getting higher results. We find the best metric for analysis is mean average

	FS	FD	MS	MD	Mean
ITML [10]	75.3	64.3	69.3	76.0	71.2
NRML [19]	66.7	66.8	64.8	65.8	66.0
Generative	70.5	70.0	67.2	74.3	70.5
Anthropological	72.5	71.5	70.8	75.6	72.6
Discriminative	**76.4**	**72.5**	**71.9**	**77.3**	**74.5**

Table 2. KinFaceW-I Verification Results. This table summarizes results on the KinFaceW-I verification data where face pairs are from different images. Our discriminative model outperforms all other methods in terms of mean average precision from 7-15%.

	FS	FD	MS	MD	Mean
ITML [10]	69.1	67.0	65.6	68.3	67.5
NRML [19]	78.8	73.2	71.9	77.9	75.5
Generative	81.8	74.3	80.5	80.8	79.4
Discriminative	83.9	76.7	83.4	84.8	82.2

Table 3. KinFaceW-II Verification Results. This table summarizes results on the KinFaceW-II verification data where face pairs are from the same image. Our discriminative model outperforms all other methods in terms of mean average precision by up to 14%.

precision (MAP) because it summarizes verification performance over a wide range of operating thresholds for a set of queries, as opposed to the single one reported by accuracy.

Tables 2 and 3 summarize the results on the KinFaceW-I and II respectively. The first two entries are metric learners Information-Theoretic Metric Learning (ITML) [10] and Neighborhood Repulsed Metric Learning (NRML) [19]. ITML tends to pull together all pairs marked as similar making no distinction in classes, while NRML pulls together a kin-match, while pushing away all other uncorrelated pairs. The advantage of our method compared to these methods is that, we are learning the features and metrics jointly. Moreover, instead of learning one metric, we learn multiple metrics at the same time which more precisely guarantees to capture the transformation between a pair of images. The Generative entry in the table refers to our feature learning technique followed by SVM classification, whereas the Discriminative technique uses the proposed hybrid model. As can be seen, the generative+discriminative model is better than the pure generative model by ~4% and outperforms the metric learners by 5-14%. This means that learning the features and metrics jointly improves performance.

We further explore our method by introducing the weights in [24] to put more emphasis on parts we labeled followed by patch extraction. The weights show the relative contribution of heredity and environment for each facial

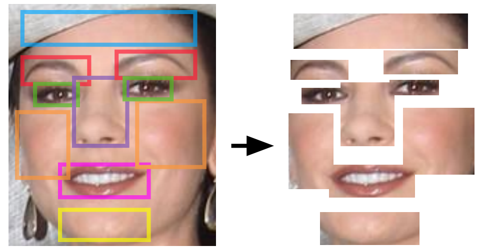

Figure 7. Facial Parts. Here we depict the parts we use in the experiments introducing weights for the most genetic features [24].

feature as shown in the last image of Figure 5. The intuition behind our experiment is that the more genetic is a facial feature, the more contribution it should have in finding parent-offspring pairs. Therefore, we increase the contribution of the parts with higher weights by repeating the corresponding patches during training. We manually selected the parts shown in Figure 7 in a way to be as close as possible to the ones in [24] for the KinFaceW-I split and results are shown under the Anthropological entry of Table 2. Interestingly, the weights help in MAP by about 2% over the generative results, however the discriminative model is able to find the best features without any part detection or annotation.

5.4.3 Generalizability

Finally, we explore the generalizability of each method by training a generic model for parent-offspring relationships combining all father-son, father-daughter, mother-son, and mother-daughter pairs. Intuitively, we want to see if we can learn the genetic relationship between parents and offspring overall and still perform well across domains. Moreover, having a generic model can be useful in finding family members and is more suitable when the gender of the test subject is unknown. Similar to previous verification experiments, we use the KinFaceW dataset, do a 5-fold cross-validation, and record the results in terms of MAP as shown in Table 4. Most interestingly, all of the models obtain competitive results using generic learned models versus the relationship specific ones and our discriminative method still outperforms other methods by approximately 6-7%.

6. Conclusions and Future Work

In this paper, we introduce a new method for learning discriminative, genetic features for describing the parent-offspring relationship. Using this method, we uncover three key insights that bridge the gap between anthropological studies and computer vision. First, we find that our results corroborate the finding that offspring resemble their parents with a probability higher than chance (Section 5.2). Second, we conclude that female offspring resemble their mothers

	KinFaceW-I	KinFaceW-II
ITML [10]	69	75.3
NRML [19]	67.7	74.1
Generative	68.3	78.9
Discriminative	**72.72**	**81.09**

Table 4. Generic Model Results. This table shows results for training a generic model combining all parent-offspring pairs. Our Discriminative model outperforms the next best metric learner by 6%.

more often than their fathers, while a male offspring only slightly favor the father(Section 5.3). Third, our algorithm discovers features similar to those found in anthropological studies, for example the eyes and parts of the nose, however our method generally uses additional information to make its decisions (Section 5.4.1). Moreover, we consider the face verification task and obtain a performance increase over existing methods by 5-14% depending on the relationships analyzed. Further, we consider the generalizability of each learning method and find that our method outperforms existing techniques by up to 7%.

In summary, we have made the first major strides towards bridging the gap between computer vision and anthropological studies by looking at the parent-offspring relationship, however there is still a large breadth of literature with further insights into vision problems. Moreover, kinship verification in the past has tended to focus on looking at individual relationships like mother-daughter, however as we showed, a joint, generic model is quite adequate at the task, thus requiring more research.

References

[1] D. Alain and S. Olivier. Gated autoencoders with tied input weights. *ICML*, 2013. 2, 3

[2] A. Alvergne, C. Faurie, and M. Raymond. Differential facial resemblance of young children to their parents: who do children look like more? *Evolution and Human Behavior*, 2007. 1, 4, 5, 6

[3] C. L. Apicella and F. W. Marlowe. Perceived mate fidelity and paternal resemblance predict men's investment in children. *Evolution and Human Behavior*, 2004. 2

[4] Y. Bengio, A. C. Courville, and P. Vincent. Unsupervised feature learning and deep learning:a review and new perspectives. *CoRR*, 2012. 2

[5] S. Brédart and R. M. French. Do babies resemble their fathers more than their mothers? A failure to replicate Christenfeld and Hill (1995). *Evolution and Human Behavior*, 1999. 1, 4, 5

[6] P. Bressan and M. Grassi. Parental resemblance in 1-year-olds and the Gaussian curve. *Evolution and Human Behavior*, 2004. 1, 4, 5

[7] P. Burke. Intrapair facial differences in twins. *Acta Geneticae Medicae et Gemellologiae*, 1989. 1

[8] V. C, K. K, S. R, R. LC, and T. GC. Genetic and environmental influences on facial profile. *Australian Dental Journal*, 1995. 1

[9] N. Christenfeld and E. Hill. Whose baby are you? *Nature*, 1995. 1

[10] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon. Information-theoretic metric learning. *ICML*, 2007. 7, 8

[11] R. Fang, A. C. Gallagher, T. Chen, and A. Loui. Kinship Classification by Modeling Facial Feature Heredity. *IEEE ICIP*, 2013. 2, 4

[12] R. Fang, K. D. Tang, N. Snavely, and T. Chen. Towards computational models of kinship verification. In *IEEE ICIP*, 2010. 2

[13] A. Gallagher and T. Chen. Understanding images of groups of people. *IEEE CVPR*, 2009. 2

[14] G. E. Hinton, S. Osindero, and Y. whye Teh. A fast learning algorithm for deep belief nets. *Neural Computation*, 2006. 2

[15] Q. V. Le, R. Monga, M. Devin, G. Corrado, K. Chen, M. Ranzato, J. Dean, and A. Y. Ng. Building high-level features using large scale unsupervised learning. *ICML*, 2012. 2

[16] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. *ICML*, 2009. 2

[17] H. Lee, R. Raina, A. Teichman, and A. Y. Ng.. Exponential family sparse coding with application to self-taught learning. *IJCAI*, 2009. 2

[18] D. L.M. Facial resemblance enhances trust. *Proceedings of the Royal Society of London B*, 2002. 2

[19] J. Lu, X. Zhou, Y.-P. Tan, Y. Shang, and J. Zhou. Neighborhood Repulsed Metric Learning for Kinship Verification. *IEEE TPAMI*, 2013. 2, 6, 7, 8

[20] N. G. Martin, L. J. Eaves, M. J. Kearsey, and P. Davies. The power of the classical twin study. *Heredity*, 1978. 1

[21] R. Memisevic. On multi-view feature learning. *ICML*, 2012. 2, 3

[22] R. Memisevic. Learning to relate images. *IEEE TPAMI*, 2013. 2, 3

[23] M. Naaman, R. Yeh, H. Garcia-Molina, and A. Paepcke. Leveraging context to resolve identity in photo albums. *ACM/IEEE-CS JCDL*, 2005. 2

[24] F. B. Naini and J. P. Moss. Three-dimensional assessment of the relative contribution of genetics and environment to various facial parameters with the twin method. *American Journal of Orthodontics and Dentofacial Orthopedics*, 2004. 1, 2, 6, 7

[25] S. M. Platek. An evolutionary model of the effects of human paternal resemblance on paternal investment. *Evolution and Cognition*, 2003. 2

[26] P. Vincent, H. Larochelle, Y. Bengio, and P. A. Manzagol. Extracting and composing robust features with denoising autoencoders. *ICML*, 2008. 2

[27] G. Wang, A. Gallagher, J. Luo, and D. Forsyth. Seeing people in social context: Recognizing people and social relationships. *ECCV*, 2010. 2

[28] S. Xia, M. Shao, J. Luo, and Y. Fu. Understanding Kin Relationships in a Photo. *IEEE TMM*, 2012. 2