On the Intrinsic Dimensionality of Face Representation

Sixue Gong, Vishnu Naresh Boddeti, and Anil K. Jain

Computer Science and Engineering
Michigan State University
{gongsixu,vishnu,jain}@cse.msu.edu

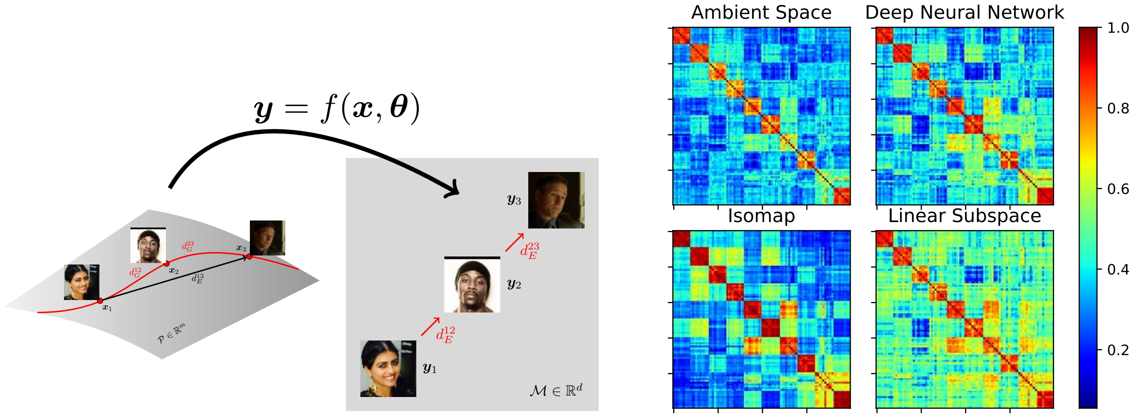
arXiv:1803.09672v1 [cs.CV] 26 Mar 2018

Abstract. The two underlying factors that determine the efficacy of face representations are, the embedding function to represent a face image and the dimensionality of the representation, e.g. the number of features. While the design of the embedding function has been well studied, relatively little is known about the compactness of such representations. For instance, *what is the minimal number of degrees of freedom or intrinsic dimensionality of a given face representation? Can we find a mapping from the ambient representation to this minimal intrinsic space that retains it's full utility?* This paper addresses both of these questions. Given a face representation, (1) we leverage intrinsic geodesic distances induced by a neighborhood graph to empirically estimate it's intrinsic dimensionality, (2) develop a neural network based non-linear mapping that transforms the ambient representation to the minimal intrinsic space of that dimensionality, and (3) validate the veracity of the mapping through face matching in the intrinsic space. Experiments on benchmark face datasets (LFW, IJB-A, IJB-B, PCSO and CASIA) indicate that, (1) the intrinsic dimensionality of deep neural network representation is significantly lower than the dimensionality of the ambient features. For instance, Facenet's [1] 128-d representation has an intrinsic dimensionality in the range of 9-12, and (2) the neural network based mapping is able to provide face representations of significantly lower dimensionality while being as discriminative (TAR @ 0.1% FAR of 84.67%, 90.40% at 10 and 20 dimensions, respectively vs 95.50% at 128 ambient dimension on the LFW dataset) as the corresponding ambient representation.

Keywords: face recognition, intrinsic dimensionality, face representation, dimensionality reduction, network based mapping

1 Introduction

A face representation is an embedding function that transforms the raw pixel representation of a face image to a point in a high-dimensional vector space. Learning or estimating such a mapping is motivated by two goals: (a) the compactness of the representation, and (2) the robustness of the mapping for the task at hand. While the latter aspect has received substantial attention, ranging from PCA based Eigenfaces [3] to state-of-the-art deep neural network based feature



(a) (b)

Fig. 1: Overview: (a) Illustration of the ambient space $\mathcal{P}$ and intrinsic manifold $\mathcal{M}$ of a face representation. In this example, while the ambient and linear dimension of the representation is three, it's intrinsic dimension is only two. Our goal in this paper is to estimate the intrinsic dimensionality of state-of-the-art face representations. (b) Heatmaps of similarity scores between image pairs from 10 classes with 10 images per class. The similarity is computed in four different spaces, the ambient space $\mathcal{P}$, space of linear dimensionality, intrinsic space estimated by Isomap [2] and intrinsic space $\mathcal{M}$ estimated by our proposed deep neural network based transformation.

representations, there has been relatively little focus on the dimensionality of the representation. The dimensionality of face representations has ranged from a few tens to thousands of dimensions. For instance, current state-of-the-art face representations have 128, 1024 and 4096 dimensions for FaceNet [1], Sphereface [4] and VGG-Face [5] respectively. The choice of dimensionality of face representation is often determined by other practical considerations, such as, parameter learning of the embedding functions [6], constraints on system memory, etc. This naturally raises the following fundamental question, *How compact can the representation be without any loss in recognition performance?* In other words, *what is the intrinsic dimensionality of the representation?* Addressing this question is the primary goal of this paper.

The intrinsic dimensionality (ID) of a representation refers to the minimum number of parameters (or degrees of freedom) necessary to capture the entire information present in the representation. Equivalently, it refers to the dimensionality of the d-dimensional representation manifold $\mathcal{M}$ embedded within the m-dimensional ambient space $\mathcal{P}$ where $d \leq m$. This notion of dimensionality is notably different from common linear dimensionality estimates obtained through principal component analysis (PCA). This dimension corresponds to the best linear subspace necessary to retain a desired fraction of the variations in the data, which in principle can be as large as the ambient dimension if the variation factors are highly entangled with each other. Figure 1 provides an illustration that highlights these differences.

Therefore, the ability to estimate the ID of a given face representation is useful in a number of ways.

- At a fundamental level the intrinsic dimensionality determines the true capacity and complexity of variations in the data captured by the representation, through the embedding function. In fact the ID can be used to guage the information content in the representation, due to its linear relation with Shannon entropy [7, 8]. Furthermore, it provides an estimate of the amount of redundancy built into the representation which can be related to it's generalization capability.
- On a practical level, knowledge of the intrinsic dimensionality is crucial for devising optimal unsupervised strategies to obtain face representations that are minimally redundant, while retaining it's full ability to recognize faces of different individuals. Recognition in the intrinsic space can provide significant savings, both in memory requirements as well as processing time, across a range of tasks like face matching, face retrieval etc. The gap between the ambient and intrinsic dimensionalities of representations can serve as a useful indicator and drive the development of algorithms that can directly learn highly compact representations. And, lastly, the ID can also serve as a dataset agnostic metric to compare the efficacy of different face representations, since it is a characteristic of the representation rather than the data itself.

Estimating the intrinsic dimensionality of data however is a challenging task. ID estimates are crucially dependent on the density variations in the representation, which in itself is difficult to estimate as faces often lie on a topologically complex curved manifold [9]. More importantly, given an estimate of ID, how do we verify that the estimate truly represents the dimensionality of complex high-dimensional embeddings for face representation? An indirect validation of the intrinsic dimensionality is possible through a mapping that transforms the ambient representation space to the intrinsic representation space while preserving it's discriminative ability. However, there is no certainty that such a mapping can be found efficiently. In practice, finding such mappings can be considerably harder than estimating the intrinsic dimensionality.

We overcome both of these challenges by (1) adopting a topological dimensionality estimation technique based on the geodesic distance between points in the manifold, and (2) relying on the ability of deep neural networks to approximate the complex mapping function from the ambient space to the intrinsic space. The latter enables a validation of the intrinsic dimensionality estimates through face verification experiments on the corresponding low-dimensional intrinsic representation of feature vectors.

The key contributions of this paper are: (1) an estimate of the intrinsic dimensionality of a 128-dimensional deep neural network based face representation, namely FaceNet [1], and (2) a deep neural network based dimensionality reduction method under the framework of multidimensional scaling. Our numerical experiments yield an intrinsic dimensionality estimate of 9-12 for FaceNet across different datasets, significantly lower than the ambient dimensionality of the representation. We show that an appreciatively designed deep neural network is able to learn a mapping from the ambient to the intrinsic space that is significantly better than other dimensionality reduction approaches with it's discriminative capability being close to that of the ambient space.

2 Related Work

Intrinsic Dimensionality: Existing approaches for estimating intrinsic dimensionality can be broadly classified into two groups: projection methods and geometric methods. The projection methods [10–12] determine the dimensionality by principal component analysis on small subregions of the data and estimating the number of dominant eigenvalues. These approaches have classically been used in the context of modeling facial appearance under different illumination conditions [13] and object recognition with varying pose [14]. While these methods serve as an efficient heuristic they cannot provide reliable estimates of intrinsic dimension. Geometric methods [15–20] on the other hand model the intrinsic topological geometry of the data and are based on the assumption that the volume of an m-dimensional set scales with it's size ϵ as ϵ^m and hence the number of neighbors less than ϵ also behave the same way. Our approach in this paper is based on the topological notion of correlation dimension [16, 17], the most popular type of fractal dimensions. The correlation dimension implicitly uses nearest-neighbor distance, typically based on the Euclidean distance. We instead follow the approach of Granata et.al. [21] utilizing the geodesic distance induced by a neighborhood graph of the data.

Face Recognition: The quest to develop face representations that are simultaneously robust and discriminative have led to extensive research in developing feature representations. For instance, among the earliest learning based approaches, Turk and Pentland proposed Eigenfaces [3] that relied on principal component analysis (PCA) of data. Later on, integrated and high-dimensional local features of faces became prevalent for face recognition, examples of which include local binary patterns (LBP) [22], scale-invariant feature transform (SIFT) [23] and histogram of oriented gradients (HoG) [24]. In contrast to these hand-designed representations the past decade has witnessed the development of end-to-end representation learning systems, leading to tremendous advances in the capabilities of face recognition systems. Convolutional neural network based features now typify the state-of-the-art face representations [25, 1, 5, 4]. All of these face representations are however characterized by features that are of hundreds to thousands of dimensions in number. While more compact representations are desirable, difficulties with optimizing deep neural networks with narrow bottlenecks [6] have proven to be the primary barrier towards realizing this goal. We believe that estimating the intrinsic dimensionality would serve as the first step towards understanding the bound on the minimal required dimensionality for representing faces and aid in the development of novel algorithms that can achieve this limit.

Dimensionality Reduction: There is a tremendous body of work on the topic of estimating low-dimensional approximations of the manifolds on which data lies. These include linear approaches such as such as Principal Component Analysis [26], Multidimensional Scaling [27] and Laplacian Eigenmaps [28] and their corresponding non-linear extensions, Locally Linear Embedding [29], Isomap [2] and Diffusion Maps [30]. Another class of dimensionality reduction algorithms leverage the ability of deep neural networks to learn complex non-linear mappings of data including deep autoencoders [31], denoising autoencoders [32, 33]. In this paper, we too leverage deep neural networks to find a mapping from the ambient to the intrinsic space while maintaining the utility of the data. However, we cast the learning problem within the framework of multidimensional scaling to preserve the geodesic distance between points in the ambient space after being mapped into the intrinsic space.

3 Approach

Our goal in this paper is two-fold: estimate the intrinsic dimensionality of a given ambient representation of a face embedding and develop a deep neural network based dimensionality reduction method to validate the intrinsic dimension estimate. Our intrinsic dimensionality estimation is based on the one presented by [21] which relies on two key ideas, (1) using graph induced geodesic distances to estimate the correlation dimension of the face representation topology, and (2) the similarity of the distribution of distances across different topological structures with the same intrinsic dimensionality. Given the estimate of the intrinsic dimensionality, we then learn a mapping that transforms the ambient representation $\mathcal{P} \in \mathbb{R}^m$ to the intrinsic representation $\mathcal{M} \in \mathbb{R}^d$ ($d < m$). We employ a deep neural network for this purpose and learn the parameters of the network to preserve distances in the intrinsic space.

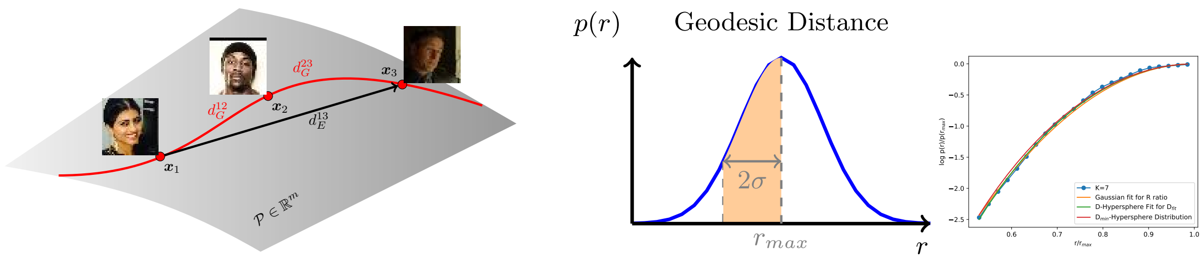
3.1 Estimating Intrinsic Dimension

We define the notion of intrinsic dimension through the classical concept of *topological dimension* of the support of a distribution $\boldsymbol{X}$. This is a generalization of the concept of dimension of a linear space [1] to a non-linear manifold. Methods for estimating the topological dimension are all based on the assumption that the behavior of the number of neighbors of a point in a d-dimensional set scales with it's size ϵ as ϵ^d. The cumulative distribution of the pairwise distances $C(r)$ can be estimated as,

$$C(r) = \frac{2}{n(n-1)} \sum_{i<j=1}^{n} H(r - \|\boldsymbol{x}_i - \boldsymbol{x}_j\|) = \int_0^r p(\boldsymbol{x}) d\boldsymbol{x} \tag{1}$$

where $p(x)$ is the probability distribution of the pairwise distances. In this paper we choose the correlation dimension [16], a particular type of topological

[1] Linear dimension is the minimum number of linearly independent vectors necessary to represent any given point in this space as a linear combination.



(a) Graph Induced Geodesic Distance **(b)** Topological Similarity

Fig. 2: Intrinsic Dimension: Our approach is based on two observations, (a) Graph induced geodesic distance between faces are able to capture the topology of the face representation manifold more reliably. (b) The distribution of the geodesic distances (for distance $r_{max} - 2\sigma \leq r \leq r_{max}$, where r_{max} is the distance at the mode) has been empirically observed [21] to be similar across different topological structures with the same intrinsic dimensionality.

dimension, to represent the intrinsic dimension of the face representation. It is is defined as,

$$d = \lim_{r \to 0} \frac{\ln C(r)}{\ln r} \tag{2}$$

Therefore the intrinsic dimension is crucially dependent on the accuracy with which the probability distribution can be estimated at very small length-scales, i.e., $r \to 0$. Significant efforts have been devoted to estimating the intrinsic dimension through line fitting in the $\log C(r)$ vs $\log r$ in the region where $r \to 0$ i.e.,

$$d = \lim_{(r_2 - r_1) \to 0} \frac{\ln C(r_2) - \ln C(r_1)}{\ln r_2 - \ln r_1} = \lim_{r \to 0} \frac{d \ln C(r)}{d \ln r} = \lim_{r \to 0} \frac{p(r)}{C(r)} r = \lim_{r \to 0} d(r) \tag{3}$$

The main drawback with this approach is the need for reliable estimates of $p(r)$ at very small length scales, which is precisely where the estimates are most unreliable with limited data, especially in very high-dimensional spaces. Granata et al. [21] present an elegant solution to this problem through three observations, (1) estimates of $d(r)$ can be stable even as $r \to 0$ if the distance between points is computed as the graph induced shortest path between points instead of the euclidean distance as is commonly the case, (2) the probability distribution $p(r)$ at intermediate length-scales around the mode of $p(r)$ i.e., $(r_{max} - 2\sigma) \leq r \leq r_{max}$ can be conveniently used to obtain reliable estimates of ID, and (3) the distributions $p(r)$ of different topological geometries are similar to each other as long as the intrinsic dimensionality is the same.

Even though this approach was used by Granata et al. [21] for problems with low complexity, swiss roll, MNIST images etc., we observed that this method can be leveraged to estimate the intrinsic dimensionality of face representations with an appropriate choice of geodesic distance. For instance, in the case of the FaceNet representation [1] that is usually normalized to lie on the surface of a hyper-sphere, the geodesic distance between nearest neighbors would correspond to the arc length in the hyper-sphere which is defined as $d_G = R\cos^{-1}\left(\frac{\boldsymbol{x}_1^T\boldsymbol{x}_2}{\|\boldsymbol{x}_1\|\|\boldsymbol{x}_2\|}\right)$, where R is the radius of the hyper-sphere on which the normalized features lie. Beyond the nearest neighbor, the distance between any pair of points is computed as the shortest path between the points as induced by the graph connecting all the points in the representation. Figure 2a. illustrates the difference between the euclidean distance and geodesic distance. Figure 2b. shows the distribution of $\frac{\log p(r)}{p(r_{max})}$ vs $\frac{r}{r_{max}}$ in the range $r_{max} - 2\sigma \leq r \leq r_{max}$ for the 128-dimensional FaceNet representation and different topological geometries, i.e., hyper-spherical and Gaussian distribution of the same intrinsic dimensionality. Therefore, by matching the distance distribution $p(r)$ to that of a known distribution, say the hyper-spherical and Gaussian distribution. For instance, the distribution of geodesic distance for a hyper-sphere is $p(r) = c\sin^{d-1}\left(\frac{\pi r}{r_{max}}\right)$ and $p(r) = cr^{d-1}\exp\left(-\frac{r^2}{4\sigma^2}\right)$, where d is the dimensionality of the hyper-sphere. Therefore, the intrinsic dimension can be estimated by comparing to the hyper-spherical distribution as,

$$\min_{c,d} \int_{r_{max}-2\sigma}^{r_{max}} \left\| \log p(r) - \log(c) - (d-1)\log\left(\sin\left[\frac{\pi r}{r_{max}}\right]\right)\right\|_2^2 \tag{4}$$

$$\min_{d} \int_{r_{max}-2\sigma}^{r_{max}} \left\| \log \frac{p(r)}{p(r_{max})} - (d-1)\log\left(\sin\left[\frac{\pi r}{r_{max}}\right]\right)\right\|_2^2 \tag{5}$$

where the need to estimate the constant c can be absorbed within the normalization of the probability distribution. Similarly, the intrinsic dimension can also be estimated by comparing to the Gaussian distribution as,

$$\min_{d} \int_{r_{max}-2\sigma}^{r_{max}} \left\| \log \frac{p(r)}{p(r_{max})} + (d-1)\frac{r^2}{4\sigma^2}\right\|_2^2 \tag{6}$$

The solutions of both the optimization problems could result in a fractional dimension. If one only requires integer solutions, the optimal value of d can be estimated by simple numerical evaluation of the objective and selecting the value that minimizes the objective. We refer to the former solution as d_{fit} and the later solution as d_{min} through the rest of the paper.

3.2 Estimating Intrinsic Space

The intrinsic dimensionality estimates obtained in the previous subsection alludes to the existence of a mapping, that can transform the ambient representation to the intrinsic space, but does not provide any solutions to find such a mapping. The mapping itself could potentially be complex and difficult to estimate.

We base our solution on Multidimensional scaling (MDS) [27], a classical mapping technique for analyzing data similarity or dissimilarity. MDS attempts to preserve the distances (dissimilarities) between points after embedding them

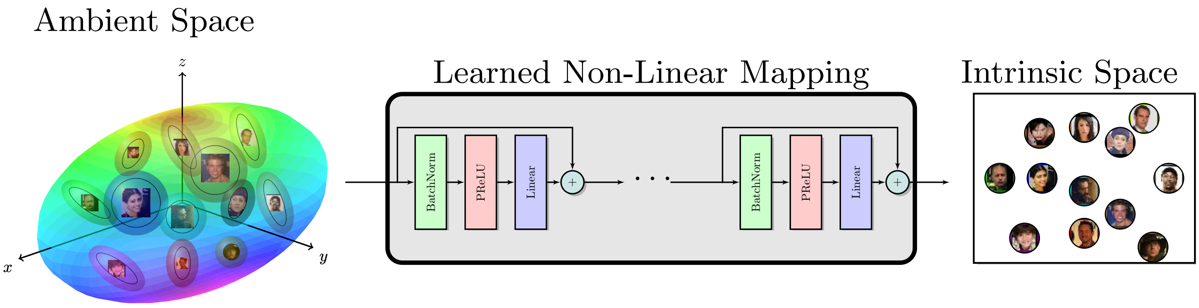


Fig. 3: Non-Linear Mapping: A neural network based non-linear mapping is learned to transform the ambient representation to the corresponding intrinsic space. The neural network is optimized to preserve corresponding distances between pairs of points in the ambient and the intrinsic representations.

in a low-dimensional space. Given data points $\boldsymbol{X} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ in the ambient space and $\boldsymbol{Y} = \{\boldsymbol{y}_1, \dots, \boldsymbol{y}_n\}$ the corresponding points in the intrinsic low-dimensional space, the MDS problem is formulated as,

$$\min \sum_{i<j} \left(d_H(\boldsymbol{x}_i, \boldsymbol{x}_j) - d_L(\boldsymbol{y}_i, \boldsymbol{y}_j)\right) \tag{7}$$

where $d_H(\cdot)$ and $d_L(\cdot)$ are distance metrics in the ambient and intrinsic space, respectively. Different choices of the distance metric, leads to different dimensionality reduction algorithms. For instance, classical metric MDS is based on Euclidean distance between the points while using geodesic distance induced by a neighborhood graph leads to Isomap [2]. Similarly, many different distance metrics have been proposed corresponding to non-linear mappings between the ambient space and the intrinsic space. However, these approaches suffer from a few drawbacks, (1) they do not scale well with the size of the number of data points, and (2) ambiguity in the choice of the right non-linear function.

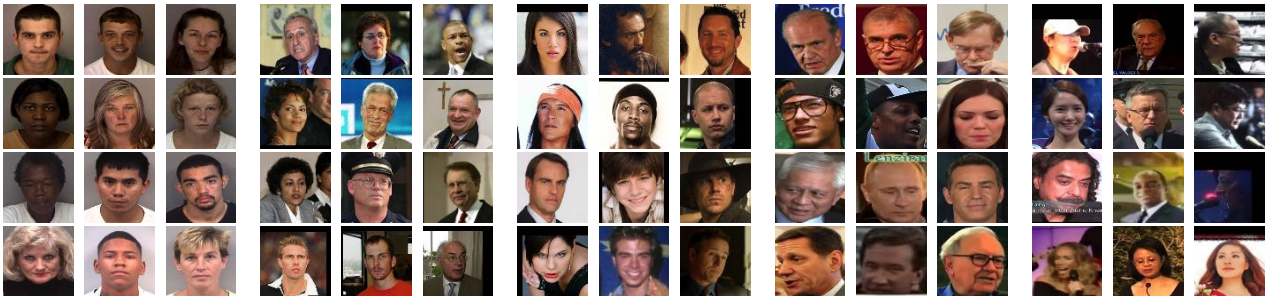
To overcome these limitations, we employ a deep neural network to approximate the non-linear mapping function that transforms the ambient representation, $\boldsymbol{x}$, to the intrinsic space, $\boldsymbol{y} = f(\boldsymbol{x}; \boldsymbol{\theta})$ with parameters $\boldsymbol{\theta}$. We pose the problem of learning the parameters within the MDS framework as,

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \sum_{j \in \mathcal{N}_k(i)} \left[d(\boldsymbol{x}_i, \boldsymbol{x}_j) - d(f(\boldsymbol{x}_i; \boldsymbol{\theta}), f(\boldsymbol{x}_j; \boldsymbol{\theta}))\right] + \lambda \|\boldsymbol{\theta}\|_2^2 \tag{8}$$

where for each data point $\boldsymbol{x}_i$ we consider it's k nearest neighbors $\mathcal{N}_k$. Considering only a subset of the neighbors eases the computational burden from considering $n \times k$ pairs instead of $n \times n$ pairs, for computing pairwise distances, while also allowing us to overcome the imbalance between the number of similar and dissimilar pairs in the dataset.

4 Experiments

In this section we will describe the estimates of the intrinsic dimensionality of FaceNet [1] using multiple datasets of varying complexity. We also evaluate the



(a) PCSO (b) LFW (c) CASIA (d) IJB-A (e) IJB-B

Fig. 4: Example images from each of the five datasets considered in our paper in increasing order of complexity. The IJB-A and IJB-B datasets exhibit large variations in pose, expressions and illumination. The LFW and CASIA datasets, however, only have limited variations in pose, illumination and expression, but LFW consists of only very few images per identity, on average. The PCSO dataset (not in the public-domain) is a mugshot dataset of frontal face images over a large and diverse set of identities.

efficacy of the deep neural network based dimensionality reduction in finding the mapping from the ambient to the intrinsic space while maintaining the discriminative ability of the ambient representation.

4.1 Datasets

We first provide a brief description of the face datasets used to estimate the intrinsic dimensionality and also to evaluate the ability of the deep neural network to find the mapping from the ambient to the intrinsic space through face matching. Figure 4 shows a few examples of the kind of faces in each dataset.

PCSO: A large collection of mugshot images (not in the public domain) acquired from the Pinellas County Sheriffs Office (PCSO), comprising of 1,447,607 images of 403,619 subjects. We use a subset of 84,164 images from 10,000 subjects for our experiments.

LFW [34]: 13,233 face images of 5,749 subjects, downloaded from the web. These images exhibit limited variations in pose, illumination, and expression, since only faces that could be detected by the Viola-Jones face detector [35] were included in the dataset.

CASIA [36]: A large collection of labeled images downloaded from the web (based on names of famous personalities) typically used for training deep neural networks. It consists of 494,414 images across 10,575 subjects, with an average of about 500 face images per subject. This was used to train the FaceNet representation.

IJB-A [37]: IARPA Janus Benchmark-A (IJB-A) contains 500 subjects with a total of 25,813 images (5,399 still images and 20,414 video frames), an average of 51 images per subject. Compared to the LFW and CASIA datasets, the IJB-A dataset is more challenging due to: i) full pose variation making it difficult to detect all the faces using a commodity face detector, ii) a mix of images and videos, and iii) wider geographical variation of subjects.

IJB-B [38]: IARPA Janus Benchmark-B (IJB-B) dataset consists of 1,845 subjects with a total of 76,824 images (21,798 still images and 55,026 video frames from 7,011 videos), an average of 41 images per subject. Images in this dataset are labeled with ground truth bounding boxes and other covariate meta-data such as occlusions, facial hair and skin tone.

4.2 Baseline Methods

Intrinsic Dimensionality: To compare the algorithms for estimators of the intrinsic dimensionality, we obtain ID estimates through two approaches, a K-nearest neighbor based estimator [15] and "Intrinsic Dimensionality Estimation Algorithm" (IDEA) [39]. Both of these estimators are known to underestimate the intrinsic dimensionality.

Dimensionality Reduction: We consider two dimensionality reduction algorithms, principle component analysis (PCA) for linear dimensionality reduction and Isomap [2] a non-linear multidimensional scaling using graph distances.

4.3 Implementation Details

The architecture of our deep neural network for dimensionality reduction is based on the idea of residual networks [40]. Our network consists of multiple linear layers to change the dimensionality with a residual layer between them. The parameters of the network are learned using the Adam [41] optimizer with a learning rate of 0.01 and the regularization parameter $\lambda = 3 \times 10^{-4}$. We observed that using the cosine-annealing scheduler and balance between the positive and negative in each minibatch was critical to learn an effective mapping.

Before we train the network, we extract 128-dimensional FaceNet feature vectors. These features are normalized to lie on the surface of a hyper-sphere i.e., L_2 norm of each feature vector is equal to 1. The first two linear layers reduce the original 128-dimension, D_0, to an intermediate space of dimensionality D_M, with a shortcut inserted to the connection. We repeat such dimension reduction units with residuals several times, before a final linear layer that decreases the dimension from the intermediate space to the target intrinsic dimension, D_I. To proceed the dimension reduction gradually in an orderly way, we evenly split the dimensions to be reduced into t stages. Thus, each dimension reduction unit will reduce the dimension by $D_R = \frac{D_0 - D_I}{t}$, so that the network will learn the mapping and disentangle the ambient representation in a steady stagewise fashion.

4.4 Intrinsic Dimensions

Table 1 reports the intrinsic dimensionality estimates for the FaceNet [1] model across the five different datasets. Our estimates are reported as we vary the number of neighbors k used to compute the parameters of the probability density for our baselines [15, 39], and the shortest path between two samples for the graph distance algorithm [21] that our estimates are based on. The ID estimates from the KNN [15] and IDEA [39] approaches have been rounded to integers. Note that for the graph distance method, we only show the results with four nearest neighbors on LFW dataset, because the number of neighbors k has to be large enough to include as many as samples as possible to construct a single fully connected graph. For the other larger and more challenging face datasets, a small k ($k = 4$) is not able to fully connect every point in the dataset. To select the $k-$nearest neighbors, we calculate the distance using the Euclidean metric in the embedding space for the two baseline approaches, [15] and [39]. Since the graph distance based algorithm approximates the geodesic distance by the shortest path connecting any two points, we consider both the Euclidean distance (chord length) and the great-circle distance (arc length), since all the FaceNet feature vectors all normalized features with unit L_2 norm, they reside on the surface of a unit hyper-sphere. The great-circle distance between two feature vectors $\mathbf{f}_1$ and $\mathbf{f}_2$ can be simply calculated as:

$$\sigma = \arccos\left(\frac{\mathbf{f}_1}{\|\mathbf{f}_1\|_2} \cdot \frac{\mathbf{f}_2}{\|\mathbf{f}_2\|_2}\right) \tag{9}$$

The intrinsic dimensionality estimates of the FaceNet representations, under both the Euclidean distance and great-circle distance, present similar results on all our datasets according to Table 4 and Table 5, showing the efficacy of approximation of intrinsic geodesic distances using the distances calculated in the embedding space of every small neighborhood. By comparing results of different estimators, we noticed that both the k-nearest neighbor based estimator [15] and the IDEA estimator [39] are less sensitive to the number of nearest neighbors than the graph distance based method [21]. However, the near neighbor estimator generally yields an underestimate of the intrinsic dimensionality for sets with high intrinsic dimensionality [12].

To find an appropriate number of nearest neighbors for graph distance based ID estimation, we generated different graphs by varying the number of neighbors k and connecting each point in the dataset to its k neighbors. On one hand, k should be small enough to avoid shortcuts between two points that are close to each other in the Euclidean space, but are actually far away in the corresponding intrinsic manifold due to highly complicated local curvatures. By performing a quadratic fit on the left side ($r_{max} - 2\sigma \leq r \leq r_{max}$) of the geodesic distance distribution, we can further analyze the curvature of the probability distribution, and select the appropriate number of neighbors k with a better quadratic fit. See the supplementary material for fitness of the curves on four different datasets, LFW. PCSO, IJB-B, and CASIA respectively, based on two distance metrics, i.e., Euclidean distance and great-circle distance. On the other hand, k has to be large enough to generate a graph that connects all points in the entire dataset i.e., there are no unreachable (unconnected) data samples. Our experiments suggests that across all of the five datasets, all the data samples in the graph can be fully connected by using seven or more than seven nearest neighbors for each given point.

Table 1: Intrinsic Dimensionality of Face Representation

Table 2: KNN [15]

k	dataset				
	LFW	IJB-A	IJB-B	PCSO	CASIA
4	10	9	10	8	10
7	10	8	9	7	11
9	11	8	9	6	11
15	11	7	9	6	11
30	12	7	8	6	10
90	13	7	8	9	10

Table 3: IDEA [39]

k	dataset				
	LFW	IJB-A	IJB-B	PCSO	CASIA
4	14	12	13	10	16
7	13	10	11	7	14
9	13	8	10	7	13
15	12	7	9	7	12
30	12	6	9	8	10
90	14	8	9	13	10

Table 4: Graph Distance (Euclidean)

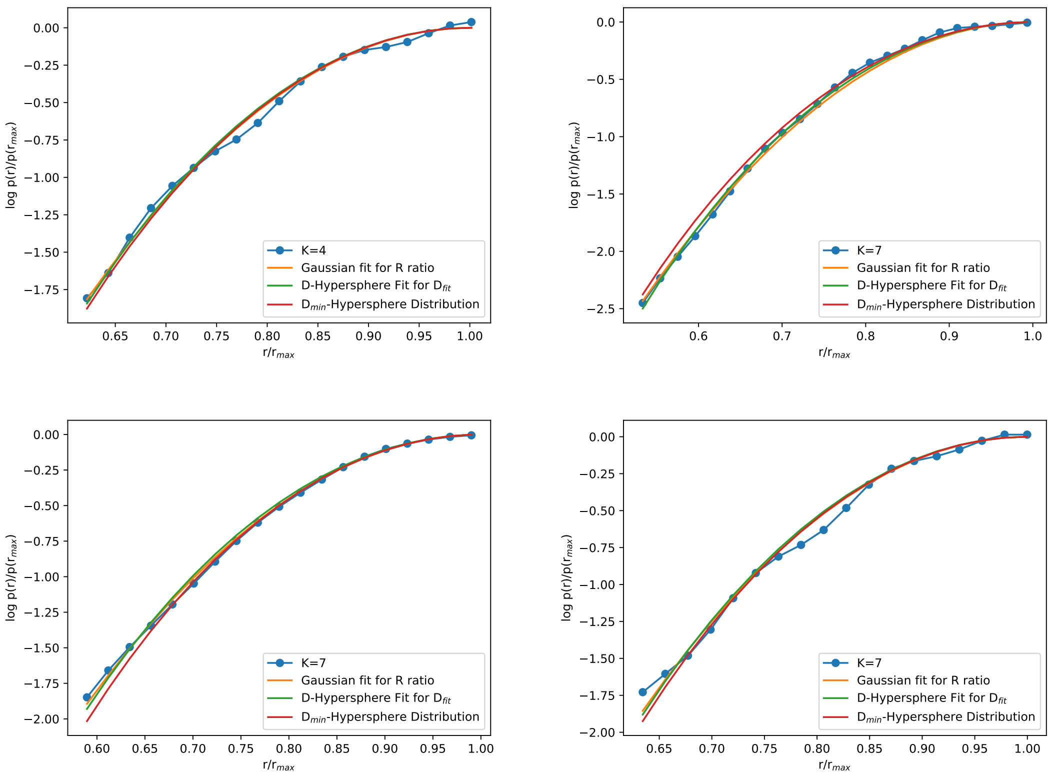
k	dataset				
	LFW	IJB-A	IJB-B	PCSO	CASIA
4	11	-	-	-	-
7	13	10	10	9	12
9	10	11	11	10	12
15	18	11	12	10	13
30	33	16	12	10	13
90	23	31	27	15	87

Table 5: Graph Distance (Great-Circle)

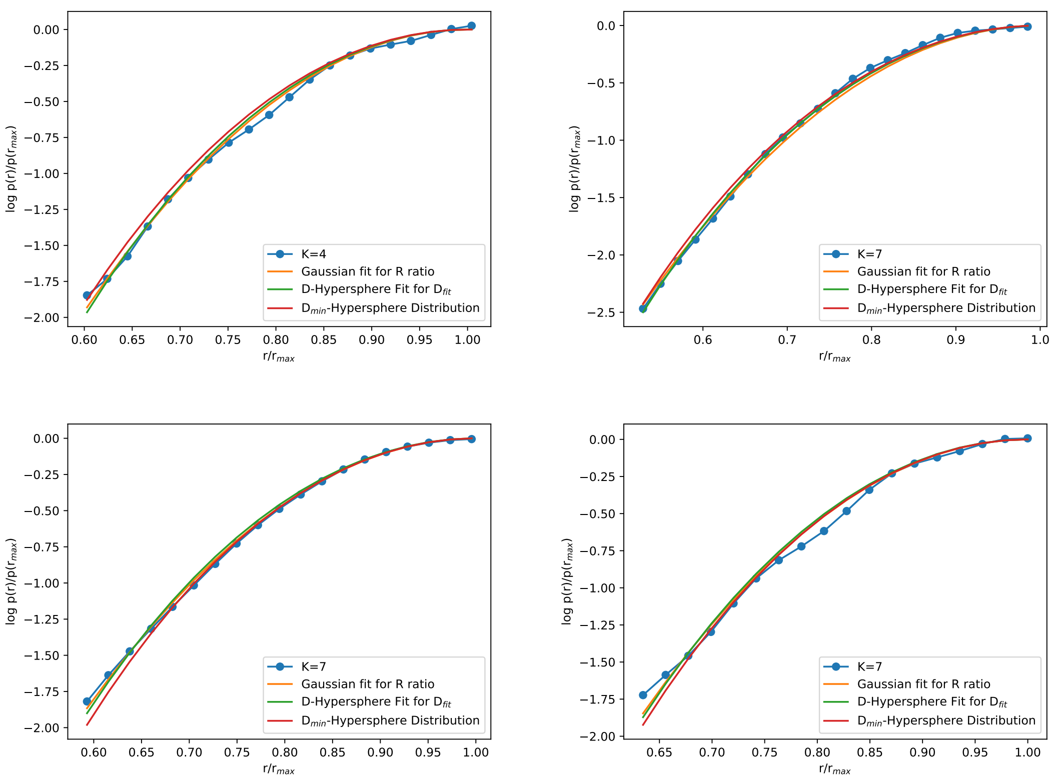
k	dataset				
	LFW	IJB-A	IJB-B	PCSO	CASIA
4	10	-	-	-	-
7	13	10	10	9	11
9	11	10	12	10	12
15	18	11	12	10	12
30	33	15	14	11	12
90	23	32	27	13	39

The intrinsic dimensionality estimator based on graph induced distance approximates the geodesic distance by the shortest path connecting any two points. In our experiments, we consider both the Euclidean distance and the great circle distance. The number of neighbors k is crucial for constructing the entire embedding graph. We thus analyze the curvature of the probability distribution to select the appropriate k. By performing quadratic fit on the left side geodesic distance distribution ($r_{max} - 2\sigma \leq r \leq r_{max}$), we select the appropriate number of neighbors by observing the fitness curves. Fig. 5 shows the fitness of the curves on four different datasets, LFW. PCSO, IJB-B, and CASIA respectively, based on two distance metrics.

By analyzing both the number of connected points and the quadratic curve fitness, we chose the parameter $k = 4$ for LFW dataset, and $k = 7$ for the rest of the datasets. Figure 7a presents the Euclidean distance distribution of the five datasets. In the geodesic distance space, all datasets share a similar distribution. We also reported the dimension estimates by analyzing the root mean square deviation (RMSD) between the observed distribution and the distribution of a D-dimensional hyper-sphere. The dimension that minimizes the RMSD is the corresponding intrinsic dimension. Our results indicate that across the five datasets the intrinsic dimensionality of the FaceNet representation is around ten. These estimates are significantly lower than the dimensionality of the ambient space of 128, suggesting that face representations could, in princple, be almost $10\times$ more compact.

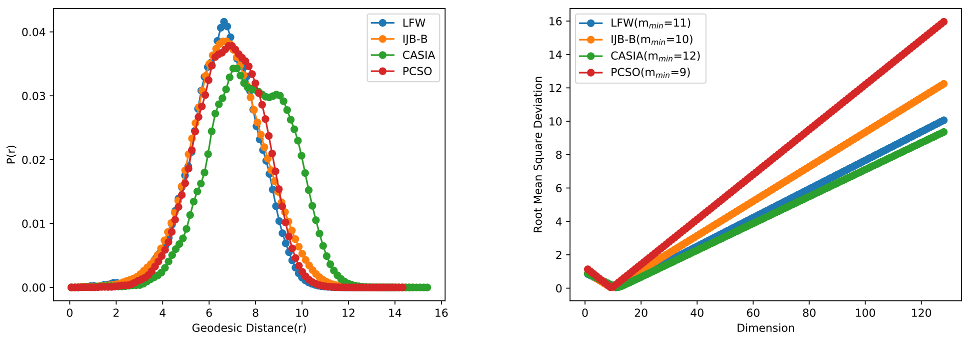


(a) Euclidean Distance



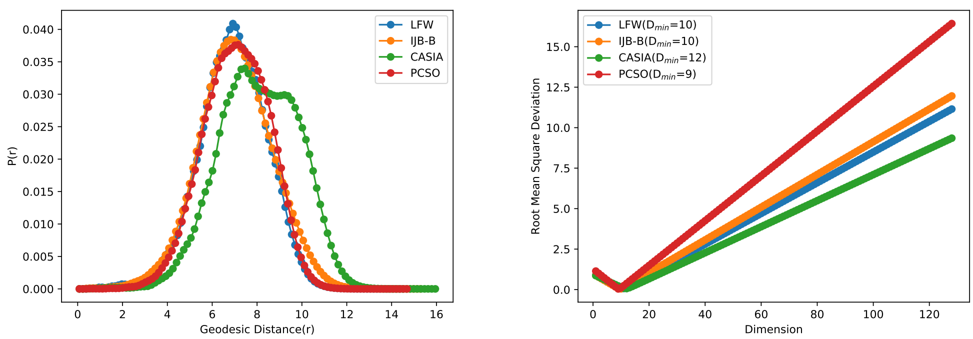
(b) Great-circle Distance

Fig. 5: Curve fit on the probability distribution of graph distances. Figure 5a shows the fit on the graph generated by using Euclidean distances between near neighbors. Figure 5b shows the fit on the graph generated by using great circle distances.



(a) Distance Distribution $p(r)$ **(b)** d_{min} Fitting

Fig. 6: Probability distribution of euclidean distances and the ID estimation based on the global minimum (d_{min}) of RMSD



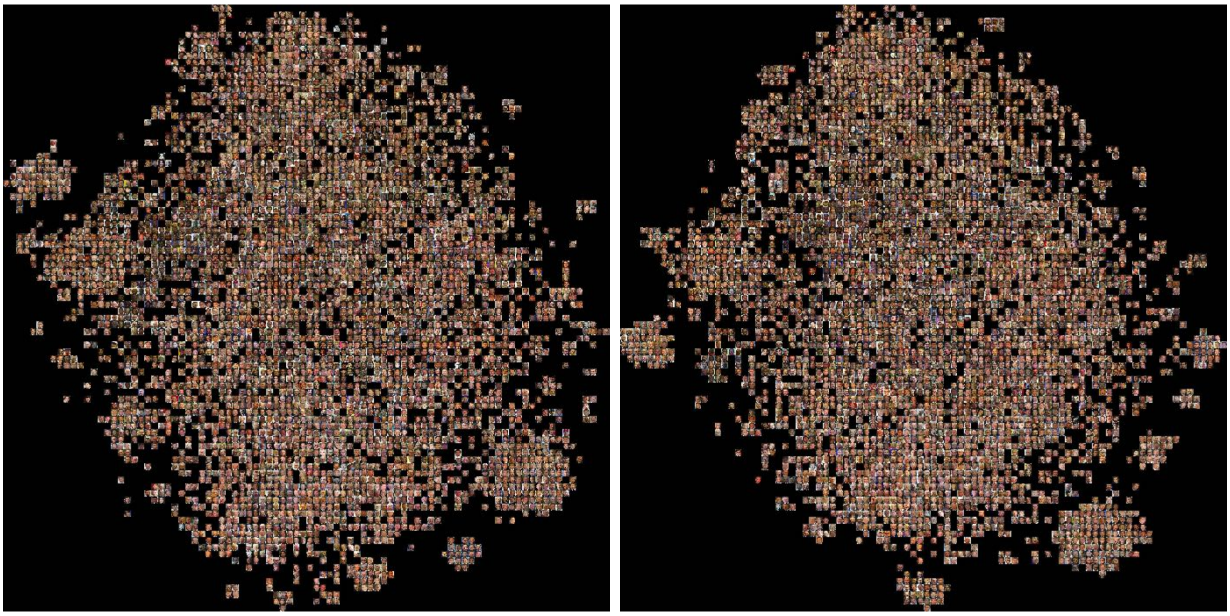
(a) Distance Distribution $p(r)$ **(b)** d_{min} Fitting

Fig. 7: Probability distribution of great circle distance distances and the ID estimation based on the global minimum (d_{min}) of RMSD

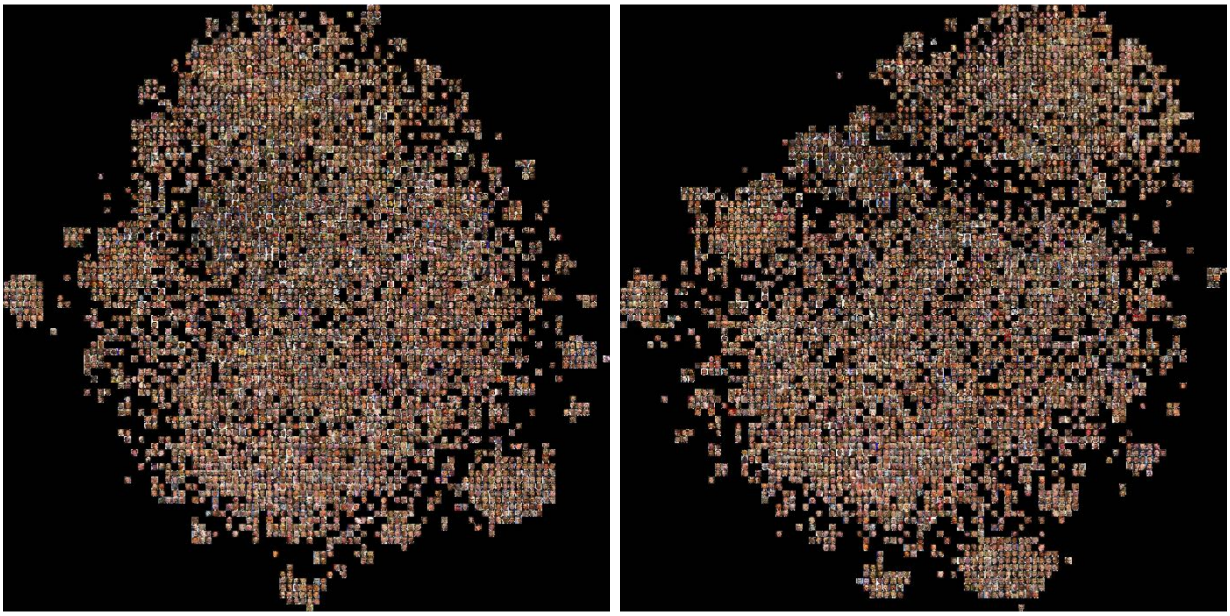
Visualization: Our main objective for training the mapping network is to preserve the distance between samples in the given dataset. It is critical that the projected features preserve the local relationships and the overall structure of the embedding space. We visualize the lower dimensional features projected by the proposed deep neural network, and the original 128-dimensional feature vectors. Fig. 8 and Fig. 9 provide visualizations of the embeddings after projecting the feature vectors onto the two-dimensional space by using t-SNE [42]. Fig. 8 illustrates the embeddings of the entire LFW dataset. In Fig. 9, we selected twenty subjects in LFW. We observe that the deep neural network is able to preserve local neighborhood relationships between images in the low-dimensional intrinsic space as well as the global structure of the embedding space.

4.5 Intrinsic Dimension Mapping

Validating the ID estimates directly is not feasible since the true dimensionality of face representation is unknown. Therefore we seek an indirectly by estimating a mapping from the ambient to the intrinsic space and verifying the discriminate power of the mapping. Therefore, we learn a feature mapping model to



(a) 10-dimension (b) 20-dimension



(c) 30-dimension (d) 128-dimension (ambient)

Fig. 8: Visualization of Embeddings on LFW Dataset

(a) 10-dimension (b) 20-dimension (c) 30-dimension (d) 128-dimension

Fig. 9: Visualization of Embeddings on LFW Dataset: subset of images from Fig. 8

project the original representations to the "intrinsic" distance space based on the intrinsic dimensionality. We estimate the mapping through the proposed deep neural network and traditional methods, including Isomap and PCA. We compare the discriminative ability of these mappings by evaluating them on five face datasets, namely, LFW, PCSO, IJB-A, IJB-B and CASIA. Face images in LFW and PCSO datasets correspond to constrained scenarios presenting low variations in comparison to those of IJB-A IJB-B and CASIA, a more challenging dataset consisting of faces captured in unconstrained environments. We evaluate the dimension reduction models on the LFW face dataset using the standard LFW protocol. For IJB-A and IJB-B, the protocol defines matching between templates, where each templates is composed of possibly multiple images of the class. We define the match score between templates as the average of the match scores between all pairs of images in the two templates. For PCSO and CASIA we follow the standard protocol of matching all pairs of images in the dataset.

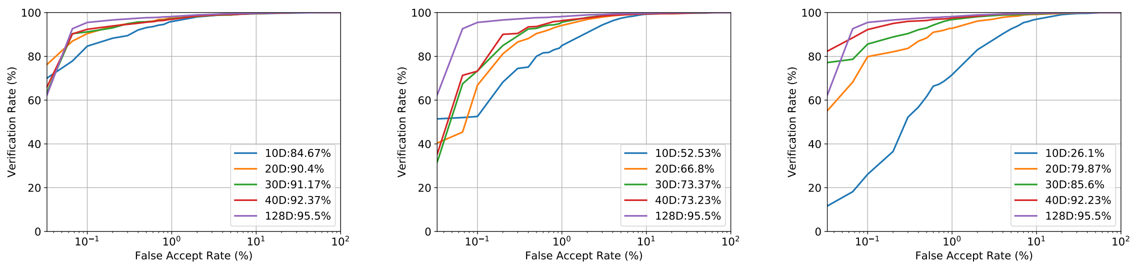
According to our ID estimators, the intrinsic dimensionality of both datasets is ten. However, designing an appropriate scheme for mapping the intrinsic manifold is much more complicated than the ID estimation itself. To show how the number of dimensions influence the performance of face presentations, we evaluate and compare the performance of the representation at different intrinsic dimensions. Specifically, the original 128-dimensional features are mapped into four different spaces with 10, 20, 30, and 40 dimensions, respectively. Figure 10, 11, 12, 13, and 14 report the performance of the proposed DNN, Isomap, and PCA, on LFW and IJB-A respectively. We make the following three observations: (1) on both datasets, the performance of 30-dimensional features and 40-dimensional features are extremely similar, and are comparable to the original 128-dimensional features. The 20-dimensional presentations of DNN based projection on LFW, consisting largely of frontal face images with minimal pose variations and facial occlusions, also achieves remarkable accuracy. (2) Our results indicate that the proposed deep network based dimensionality reduction model is able to learn a proper mapping to reduce the dimension of the given features with little loss in discriminative information as demonstrated by their recognition accuracy as well as the receiving operating curves. (3) The performance of DNN based projection out performs Isomap and PCA under the same dimension due to it's ability to better approximate the non-liner mapping from the ambient to intrinsic space. Furthermore, due to the iterative nature of Isomap, it does not provide an explicit mapping function for a new data sample, while the proposed DNN can be used to map a new data sample as well.

Table 6 shows the face verification results over different datasets for spaces with different intrinsic dimensions (10-D, 20-D, 30-D and 40-D). We compare the performance on the linear dimensionality reduction through PCA, Isomap and the proposed deep neural network based mapping. Our experimental results suggest that the mapping learned by DNN is more able to preserve discriminative information to about 20 dimensional intrinsic space. However, the mapping learned by our DNN architecture for 10-dimensional intrinsic space is not as

Table 6: Face verification performance (TAR@FAR) in the ambient space (128-D) across different lower-dimensional spaces.

Dataset	Method	10D 0.1%	10D 1%	20D 0.1%	20D 1%	30D 0.1%	30D 1%	40D 0.1%	40D 1%	128D 0.1%	128D 1%
LFW	DNN	**84.67**	**95.90**	90.40	97.43	91.17	97.13	92.37	96.90		
	Isomap	52.53	84.93	66.80	94.13	73.37	95.63	73.23	96.33	95.50	98.20
	PCA	26.10	71.60	79.87	92.80	85.60	96.80	92.23	97.53		
PCSO	DNN	**89.23**	**97.11**	91.94	98.14	94.03	98.68	94.70	98.78		
	Isomap	55.39	89.11	79.72	96.98	92.30	99.14	94.84	99.19	96.04	99.29
	PCA	63.08	89.06	80.82	96.19	88.31	98.26	92.77	98.89		
IJB-A	DNN	**35.32**	**59.23**	37.78	63.03	40.32	67.13	40.87	67.11		
	Isomap	21.08	45.03	33.06	61.69	36.63	65.89	37.42	67.93	45.98	70.26
	PCA	6.43	20.56	22.09	47.38	35.24	62.06	42.74	68.20		
IJB-B	DNN	**34.88**	58.62	41.43	63.45	46.43	71.29	48.60	72.58		
	Isomap	34.60	**63.32**	35.71	61.45	38.93	63.64	46.39	69.06	48.67	74.47
	PCA	6.57	26.37	20.93	38.91	31.58	58.95	44.46	67.60		
CASIA	DNN	**69.71**	**82.20**	71.34	82.55	73.33	85.26	76.78	86.68		
	Isomap	29.41	64.71	44.22	76.39	70.88	88.02	76.53	87.69	81.10	89.92
	PCA	26.21	55.60	47.51	71.50	59.72	79.19	67.75	83.49		

effective in learning a mapping that fully preserves the discriminative content of the representation.

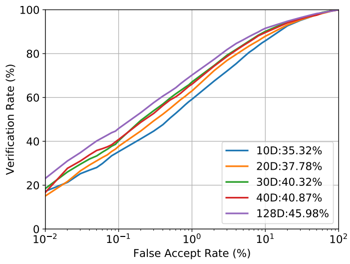


(a) Deep Neural Network **(b)** Isomap **(c)** PCA

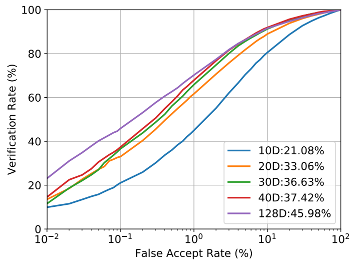
Fig. 10: Face Verification on LFW: Receiver Operating Curves

5 Concluding Remarks

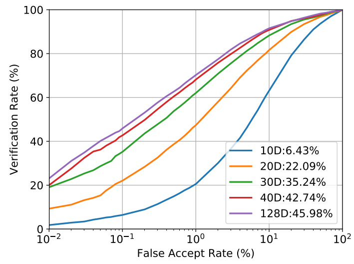
In this paper, we addressed two questions, given a face representation, what is the minimum degrees of freedom in the representation i.e., it's intrinsic dimension and can we find a mapping between the ambient and intrinsic space while maintaining the discriminative capability of the representation? We leverage a



(a) Deep Neural Network

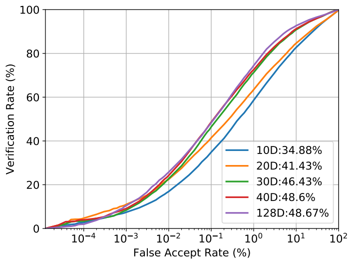


(b) Isomap

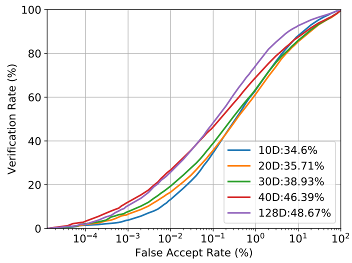


(c) PCA

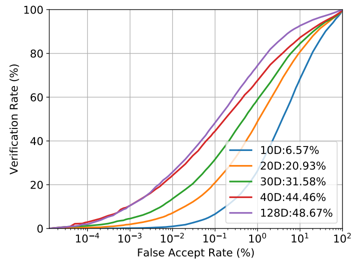
Fig. 11: Face Verification on IJB-A: Receiver Operating Curves



(a) Deep Neural Network

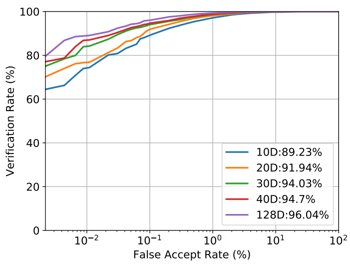


(b) Isomap

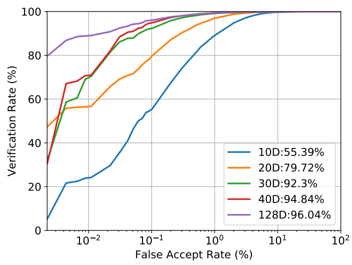


(c) PCA

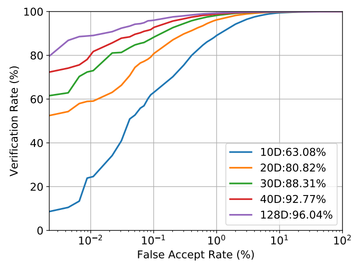
Fig. 12: Face Verification on IJB-B: Receiver Operating Curves



(a) Deep Neural Network

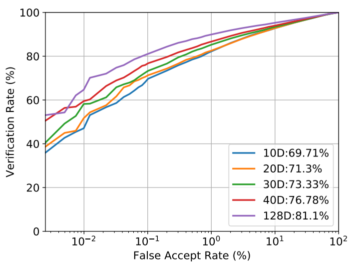


(b) Isomap

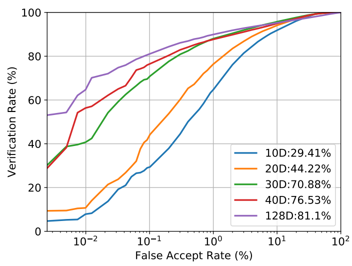


(c) PCA

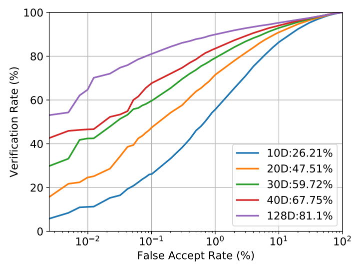
Fig. 13: Face Verification on PCSO: Receiver Operating Curves



(a) Deep Neural Network



(b) Isomap



(c) PCA

Fig. 14: Face Verification on CASIA: Receiver Operating Curves

geodesic graph distance based approach to estimate the intrinsic dimension and proposed a deep neural network to learn a mapping to transform the ambient space to the intrinsic space. Experiments on a deep neural network based face representation, 128-dimensional FaceNet, yielded an intrinsic dimension between 9-12 dimensions across multiple datasets, significantly less than the ambient dimension. The deep neural network was able to estimate a mapping that can transform the ambient to the intrinsic dimension while preserving it's discriminative ability, to an extent, on the LFW and IJB-A datasets. These results suggest that (1) the ambient dimension of face representations can be significantly lower, (2) finding the optimal mapping from ambient space to highly compact intrinsic space is challenging, especially so for linear dimensionality reduction, and (3) a deep neural network based model can be learned to significantly compress existing face representations while maintaining it's discriminative power to a large extent. Our findings in this paper call for the development of algorithms that can directly learn more compact face representations.

References

1. Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: A unified embedding for face recognition and clustering. In: IEEE Conference on Computer Vision and Pattern Recognition. (2015)
2. Tenenbaum, J.B., De Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. Science **290**(5500) (2000) 2319–2323
3. Turk, M.A., Pentland, A.P.: Face recognition using eigenfaces. In: IEEE Conference on Computer Vision and Pattern Recognition. (1991)
4. Liu, W., Wen, Y., Yu, Z., Li, M., Raj, B., Song, L.: Sphereface: Deep hypersphere embedding for face recognition. In: IEEE Conference on Computer Vision and Pattern Recognition. (2017)
5. Parkhi, O.M., Vedaldi, A., Zisserman, A.: Deep face recognition. In: British Machine Vision Conference. (2015)
6. Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: Web-scale training for face identification. In: IEEE Conference on Computer Vision and Pattern Recognition. (2015)
7. Theiler, J.: Estimating fractal dimension. JOSA A **7**(6) (1990) 1055–1073
8. Costa, J.A., Hero, A.O.: Geodesic entropic graphs for dimension and entropy estimation in manifold learning. IEEE Transactions on Signal Processing **52**(8) (2004) 2210–2221
9. Talwalkar, A., Kumar, S., Rowley, H.: Large-scale manifold learning. In: IEEE Conference on Computer Vision and Pattern Recognition, IEEE (2008)
10. Fukunaga, K., Olsen, D.R.: An algorithm for finding intrinsic dimensionality of data. IEEE Transactions on Computers **100**(2) (1971) 176–183
11. Bruske, J., Sommer, G.: Intrinsic dimensionality estimation with optimally topology preserving maps. IEEE Transactions on Pattern Analysis and Machine Intelligence **20**(5) (1998) 572–575
12. Verveer, P.J., Duin, R.P.W.: An evaluation of intrinsic dimensionality estimators. IEEE Transactions on Pattern Analysis and Machine Intelligence **17**(1) (1995) 81–86

13. Georghiades, A.S., Belhumeur, P.N., Kriegman, D.J.: From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE Transactions on Pattern Analysis and Machine Intelligence **23**(6) (2001) 643–660
14. Murase, H., Nayar, S.K.: Visual learning and recognition of 3-d objects from appearance. International journal of computer vision **14**(1) (1995) 5–24
15. Pettis, K.W., Bailey, T.A., Jain, A.K., Dubes, R.C.: An intrinsic dimensionality estimator from near-neighbor information. IEEE Transactions on Pattern Analysis and Machine Intelligence (1) (1979) 25–37
16. Grassberger, P., Procaccia, I.: Measuring the strangeness of strange attractors. In: The Theory of Chaotic Attractors. Springer (2004) 170–189
17. Camastra, F., Vinciarelli, A.: Estimating the intrinsic dimension of data with a fractal-based method. IEEE Transactions on Pattern Analysis and Machine Intelligence **24**(10) (2002) 1404–1407
18. Kégl, B.: Intrinsic dimension estimation using packing numbers. In: Advances in Neural Information Processing Systems. (2003)
19. Hein, M., Audibert, J.Y.: Intrinsic dimensionality estimation of submanifolds in r d. In: International Conference on Machine Learning. (2005)
20. Levina, E., Bickel, P.J.: Maximum likelihood estimation of intrinsic dimension. In: Advances in Neural Information Processing Systems. (2005)
21. Granata, D., Carnevale, V.: Accurate estimation of the intrinsic dimension using graph distances: Unraveling the geometric complexity of datasets. Scientific Reports **6** (2016) 31377
22. Ahonen, T., Hadid, A., Pietikäinen, M.: Face recognition with local binary patterns. European Conference on Computer Vision (2004)
23. Lowe, D.G.: Object recognition from local scale-invariant features. In: IEEE International Conference on Computer Vision. (1999)
24. Albiol, A., Monzo, D., Martin, A., Sastre, J., Albiol, A.: Face recognition using hog–ebgm. Pattern Recognition Letters **29**(10) (2008) 1537–1543
25. Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: Deepface: Closing the gap to human-level performance in face verification. In: IEEE Conference on Computer Vision and Pattern Recognition. (2014)
26. Jolliffe, I.T.: Principal component analysis and factor analysis. In: Principal Component Analysis. Springer (1986) 115–128
27. Kruskal, J.B.: Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika **29**(1) (1964) 1–27
28. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. Neural computation **15**(6) (2003) 1373–1396
29. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. science **290**(5500) (2000) 2323–2326
30. Coifman, R.R., Lafon, S.: Diffusion maps. Applied and computational harmonic analysis **21**(1) (2006) 5–30
31. Hinton, G.E., Salakhutdinov, R.R.: Reducing the dimensionality of data with neural networks. science **313**(5786) (2006) 504–507
32. Vincent, P., Larochelle, H., Bengio, Y., Manzagol, P.A.: Extracting and composing robust features with denoising autoencoders. In: Proceedings of the 25th international conference on Machine learning, ACM (2008) 1096–1103
33. Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.A.: Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. Journal of Machine Learning Research **11**(Dec) (2010) 3371–3408

34. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical report, Technical Report 07-49, University of Massachusetts, Amherst (2007)
35. Viola, P., Jones, M.J.: Robust real-time face detection. International Journal of Computer Vision **57**(2) (2004) 137–154
36. Yi, D., Lei, Z., Liao, S., Li, S.Z.: Learning face representation from scratch. arXiv:1411.7923 (2014)
37. Klare, B.F., Klein, B., Taborsky, E., Blanton, A., Cheney, J., Allen, K., Grother, P., Mah, A., Jain, A.K.: Pushing the frontiers of unconstrained face detection and recognition: Iarpa janus benchmark a. In: IEEE Conference on Computer Vision and Pattern Recognition. (2015)
38. Whitelam, C., Taborsky, E., Blanton, A., Maze, B., Adams, J., Miller, T., Kalka, N., Jain, A.K., Duncan, J.A., Allen, K., et al.: IARPA janus benchmark-b face dataset. In: CVPRW. (2017)
39. Rozza, A., Lombardi, G., Ceruti, C., Casiraghi, E., Campadelli, P.: Novel high intrinsic dimensionality estimators. Machine learning **89**(1-2) (2012) 37–65
40. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. In: European Conference on Computer Vision, Springer (2016) 630–645
41. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
42. Maaten, L.v.d., Hinton, G.: Visualizing data using t-sne. Journal of Machine Learning Research **9**(Nov) (2008) 2579–2605