On the Intrinsic Dimensionality of Image Representations (Supplementary Material)

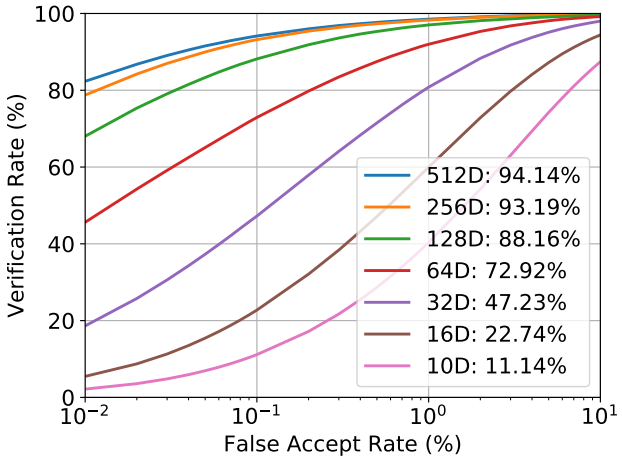
Sixue Gong Vishnu Naresh Boddeti Anil K. Jain
Michigan State University, East Lansing MI 48824
{gongsixu, vishnu, jain}@msu.edu

In this supplementary material we include; (1) Section 1: direct training of low-dimensional representations, (2) Section 2: intrinsic dimensionality estimates from the baseline approaches [4, 3], (3) Section 3: evaluation of the baseline dimensionality reduction techniques on the LFW and IJB-C datasets, (4) Section 4: derivations of the intrinsic dimensionality estimation process, (5) Section 5: RMSE and fitting plots for the graph distance based approach [1], (6) Section 6 intrinsic dimensionality estimation and learning and visualizing the learned projections on the Swiss Roll dataset, and (7) Section 7: finally visualizations of the embeddings using t-SNE [2].

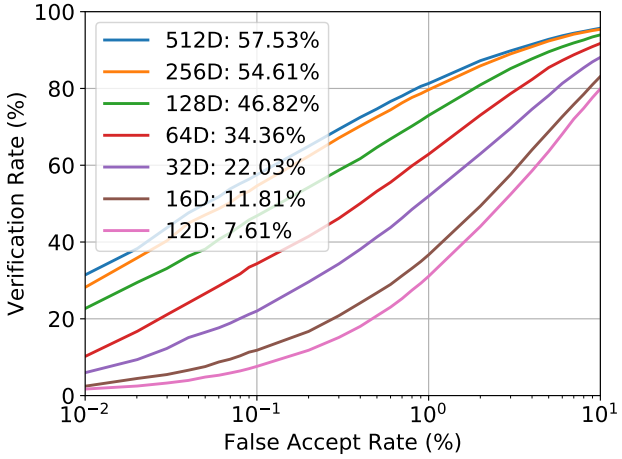
1. Direct Training

Our findings in this paper, that many current DNN representations can be significantly compressed, naturally begs the question: *can we directly learn embedding functions that yield compact and discriminative embeddings in the first place?* Taigman et al. [6] study this problem in the context of learning face embeddings, and noted that a compact feature space creates a bottleneck in the information flow to the classification layer and hence increases the difficulty of optimizing the network when training from scratch. Given the significant developments in network architectures and optimization tools since then, we attempt to learn highly compact embedding directly from raw-data, using current best-practices, while circumventing the chicken-and-egg problem of not knowing the target intrinsic dimensionality before learning the embedding function.

We train[1] the Inception ResNet V1 [5] on the CASIA-WebFace [8] for embeddings of different sizes. Figure 1 shows the ROC curves on the LFW and IJB-C datasets. The models suffer significant loss in performance as we decrease the dimensionality of the embeddings. In comparison the proposed DeepMDS based dimensionality reduction retains its discriminative ability even at high levels of compression. These results call for the development of algorithms that can directly learn compact and effective image representations.



(a) LFW



(b) IJB-C

Figure 1: ROC curve on LFW and IJB-C datasets for the Inception ResNet V1 [5] model trained with different embedding dimensionality on the CASIA-WebFace [8] dataset.

[1]We build off of the publicly available implementation at https://github.com/davidsandberg/facenet

2. Intrinsic Dimensionality Estimation

Table 1 and Table 2 reports the ID estimates from the k-nearest neighbor approach [3] and IDEA [4], respectively, for different representation models across different datasets that we consider. These approaches are known to underestimate the intrinsic dimensionality [7]. We observe the same as our ID estimates for the baselines are lower than the estimates of the graph distance based approach that we use.

Table 1: Intrinsic Dimensionality: KNN [3]

Representation	dataset	k			
		4	7	9	15
FaceNet-128	LFW	10	10	11	11
	IJB-C	10	10	9	9
FaceNet-512	LFW	8	8	8	9
	IJB-C	10	10	9	9
Sphereface	LFW	6	7	7	8
	IJB-C	6	6	5	5
ResNet-101	ImageNet-100	25	20	19	16

Table 2: Intrinsic Dimensionality: IDEA [4]

Representation	dataset	k			
		4	7	9	15
FaceNet-128	LFW	14	13	13	12
	IJB-C	14	11	10	9
FaceNet-512	LFW	12	10	10	10
	IJB-C	14	11	10	9
Sphereface	LFW	10	9	9	9
	IJB-C	8	7	6	5
ResNet-101	ImageNet-100	21	21	20	20

3. Intrinsic Dimension Mapping

In this section we present results of DeepMDS on LFW (BLUFR) dataset and the baseline dimensionality reduction methods for mapping from the ambient to the intrinsic space. Figure 2 show the face verification ROC curves of DeepMDS on LFW dataset for FaceNet-128, FaceNet-512 and SphereFace representation models. Figure 3 show the face verification ROC curves of Principal Component Analysis on the IJB-C and LFW (BLUFR) datasets for all the three representation models. Similarly, Fig. 4 and Fig. 5 show the face verification ROC curves of the Isomap and Denoising Autoencoder baselines, respectively.

4. Intrinsic Dimensionality Estimation (Derivations)

We first show the derivation for estimating the intrinsic dimensionality m that minimizes the RMSE with respect to a m-hypersphere,

$$\min_m \int_{r_{max}-2\sigma}^{r_{max}} \left\| \log \frac{\hat{p}_{\mathcal{M}}(r)}{\hat{p}_{\mathcal{M}}(r_{max})} - (m-1)\log\left(\sin\left[\frac{\pi r}{2r_{max}}\right]\right)\right\|^2$$

First we estimate σ for the m-hypersphere by approximating the distribution $\hat{p}_{\mathcal{M}}(r)$ by a univariate Gaussian distribution around the mode of $p_{\mathcal{M}}(r)$. So, given samples $S = \{r_1, \ldots, r_T\}$ from the distribution $p(r)$, the variance around the mode can be estimated as, $\sigma^2 = \frac{1}{T}\sum_{t=1}^{T}(r_t - r_{max})^2$, where r_{max} is the radius at the mode of $\hat{p}_{\mathcal{M}}(r)$. Then, we estimate the distribution $\log \frac{\hat{p}_{\mathcal{M}}(r)}{\hat{p}_{\mathcal{M}}(r_{max})}$ vs $\log\left(sin\left[\frac{\pi r}{2r_{max}}\right]\right)$ and solve the following least-squares fit problem:

$$\min_m \sum_{S \cap r_{max}-2\sigma \le r_i \le r_{max}} (y_i - (m-1)x_i)^2$$

where $y_i = \log \frac{\hat{p}\_\mathcal{M}(r_i)}{\hat{p}\_\mathcal{M}(r_{max})}$ and $x_i = \log\left(sin\left[\frac{\pi r}{2r_{max}}\right]\right)$.

In the case of comparison to a Gaussian distribution, the intrinsic dimensionality can also be estimated by comparing to the geodesic distance distribution for points sampled from a Gaussian distribution as,

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

The solution of this optimization problem can be found following the same procedure described above for a m-hypersphere.

5. Intrinsic Dimensionality Estimation Fitting

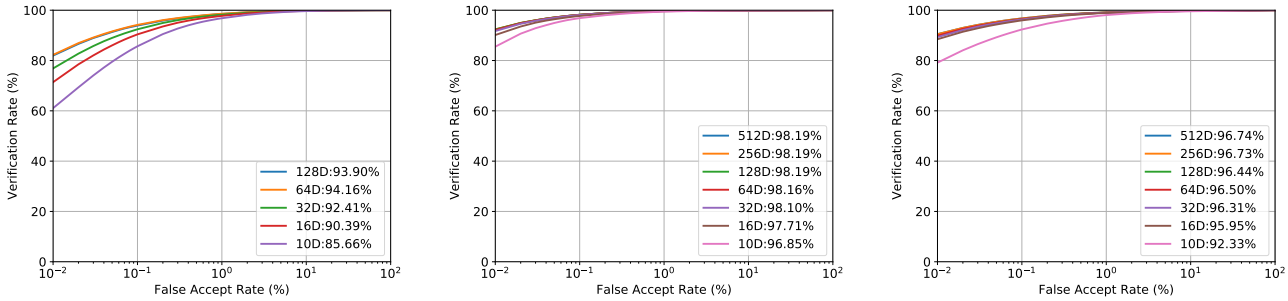
Figure 6 shows the distribution of geodesic distances $p(r)$ for each of the datasets and representation models. Figure 7 shows the plot of $\log \frac{\hat{p}_{\mathcal{M}}(r)}{\hat{p}_{\mathcal{M}}(r_{max})}$ vs $\log \frac{r}{r_{max}}$, as we vary the number of neighbors k, for the SphereFace representation model on the LFW and IJB-C datasets and ResNet-34 on the ImageNet dataset.

6. Swiss Roll

In this section we consider the swiss roll dataset, as a means of providing visual validation of the estimated intrinsic space on a known dataset. First we estimate the intrinsic dimensionality of the swiss roll dataset and then we learn a low-dimensional mapping from the ambient 3-dim space to the intrinsic space. We sample 2000 points from the swiss roll dataset and use these points for the experiments. For this dataset, the intrinsic dimensionality estimate is 2 dimensions (see Figure 8, which is indeed the ground truth intrinsic dimensionality of swiss-roll.

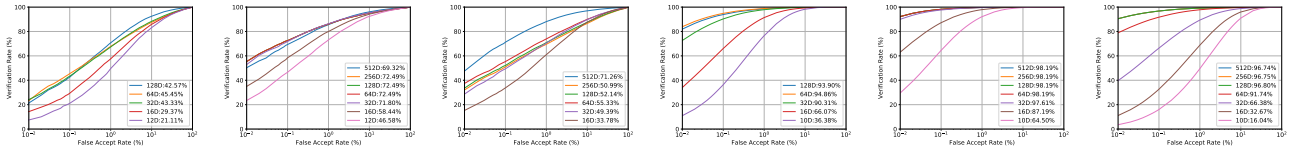
7. Visualizing Embeddings

Our main objective for training the mapping network is to preserve the distance between samples in the given



(a) LFW: FaceNet-128 (b) LFW: FaceNet-512 (c) LFW: SphereFace

Figure 2: **DeepMDS:** Face Verification on LFW (BLUFR) dataset for the (a) FaceNet-128, (b) FaceNet-512 and (c) SphereFace embeddings.



(a) IJB-C: FaceNet-128 (b) IJB-C: FaceNet-512 (c) IJB-C: SphereFace (d) LFW: FaceNet-128 (e) LFW: FaceNet-512 (f) LFW: SphereFace

Figure 3: **PCA:** Face Verification on IJB-C and LFW (BLUFR) dataset for the (a) FaceNet-128, (b) FaceNet-512 and (c) SphereFace embeddings.

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 DeepMDS, and the original 128-D FaceNet-128 feature vectors. Fig. 10 illustrates the embeddings of the entire LFW dataset. In Fig. 11, we selected twenty subjects in LFW. Fig. 10a and Fig. 11a provide visualizations of a 2-dim DeepMDS embeddings learned from the original 128-dim FaceNet-128 features. Fig. 10b and Fig. 11b provide visualizations of the 10-dim intrinsic space embedding learned from the original 128-dim FaceNet-128 features. We use t-SNE [2] to visualize the 10-dim space in two dimensions. We observe that DeepMDS 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.

References

[1] D. Granata and V. Carnevale. Accurate estimation of the intrinsic dimension using graph distances: Unraveling the geometric complexity of datasets. *Scientific Reports*, 6:31377, 2016. 1

[2] L. v. d. Maaten and G. Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(Nov):2579–2605, 2008. 1, 3

[3] K. W. Pettis, T. A. Bailey, A. K. Jain, and R. C. Dubes. An intrinsic dimensionality estimator from near-neighbor information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (1):25–37, 1979. 1, 2

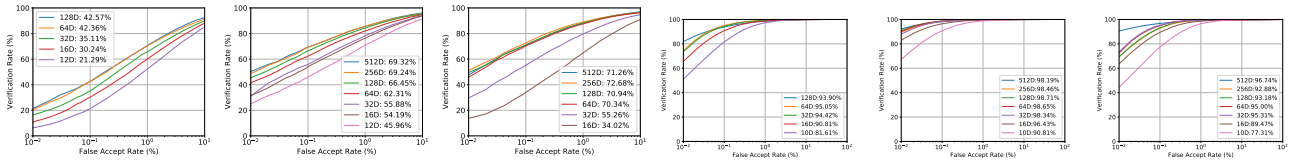
[4] A. Rozza, G. Lombardi, C. Ceruti, E. Casiraghi, and P. Campadelli. Novel high intrinsic dimensionality estimators. *Machine Learning*, 89(1-2):37–65, 2012. 1, 2

[5] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *AAAI Conference on Artificial Intelligence*, volume 4, page 12, 2017. 1

[6] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. Web-scale training for face identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2015. 1

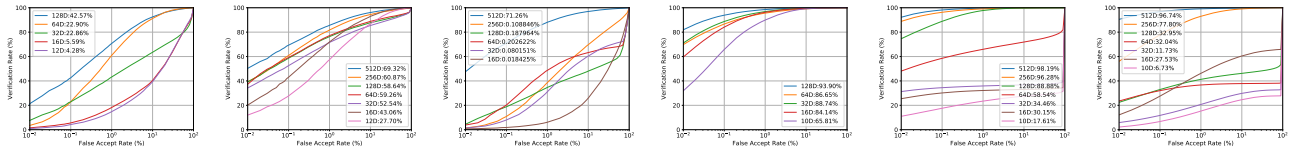
[7] P. J. Verveer and R. P. W. Duin. An evaluation of intrinsic dimensionality estimators. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(1):81–86, 1995. 2

[8] D. Yi, Z. Lei, S. Liao, and S. Z. Li. Learning face representation from scratch. *arXiv:1411.7923*, 2014. 1



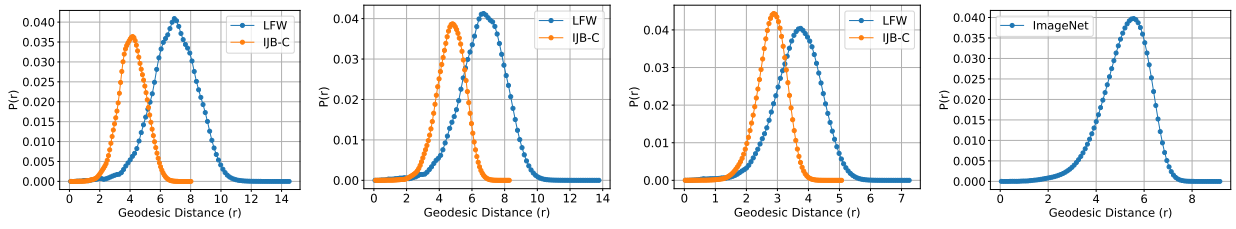
(a) IJB-C: FaceNet-128 (b) IJB-C: FaceNet-512 (c) IJB-C: SphereFace (d) LFW: FaceNet-128 (e) LFW: FaceNet-512 (f) LFW: SphereFace

Figure 4: **Isomap:** Face Verification on IJB-C and LFW (BLUFR) dataset for the (a) FaceNet-128, (b) FaceNet-512 and (c) SphereFace embeddings.



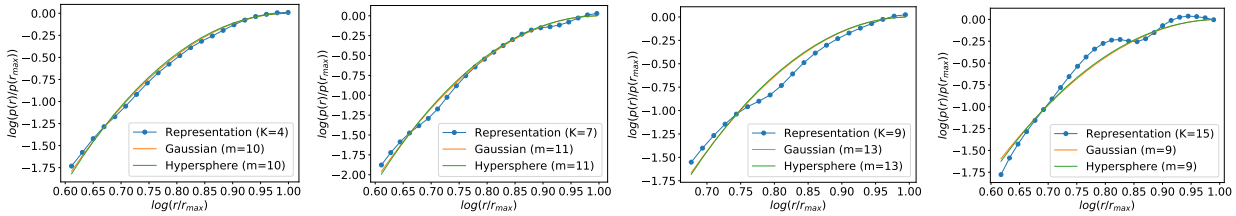
(a) FaceNet-128 (b) FaceNet-512 (c) SphereFace (d) FaceNet-128 (e) FaceNet-512 (f) SphereFace

Figure 5: **Denoising Autoencoder:** Face Verification on IJB-C and LFW (BLUFR) dataset for the (a) FaceNet-128, (b) FaceNet-512 and (c) SphereFace embeddings.

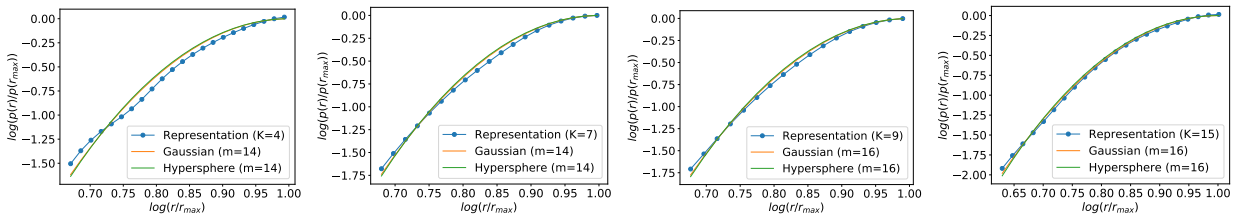


(a) FaceNet-128 (b) FaceNet-512 (c) SphereFace (d) ResNet-34

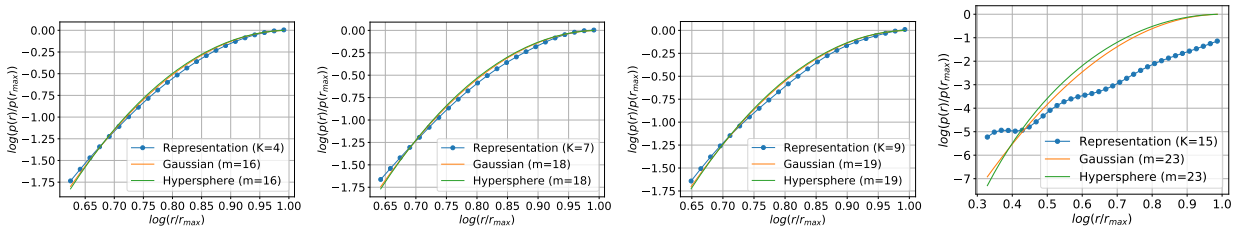
Figure 6: Distribution of geodesic distances for different representation models and datasets.



(a) SphereFace on LFW Dataset

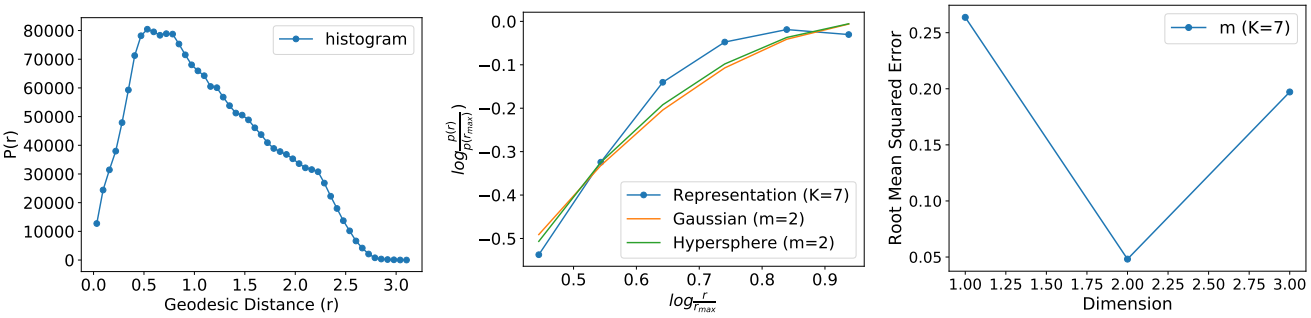


(b) SphereFace on IJB-C Dataset



(c) ResNet-34 on ImageNet Dataset

Figure 7: $\log \frac{\hat{p}_{\mathcal{M}}(r)}{\hat{p}_{\mathcal{M}}(r_{max})}$ vs $\log \frac{r}{r_{max}}$ plots as we vary number of neighbors k for different representation models and datasets.

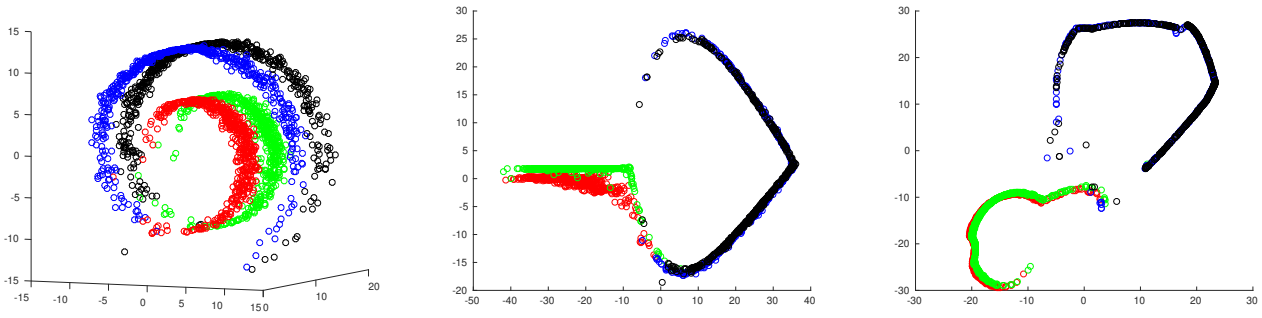


(a) Histogram of Swiss Roll

(b) $\log \frac{\hat{p}_{\mathcal{M}}(r)}{\hat{p}_{\mathcal{M}}(r_{max})}$ vs $\log \frac{r}{r_{max}}$ of Swiss Roll

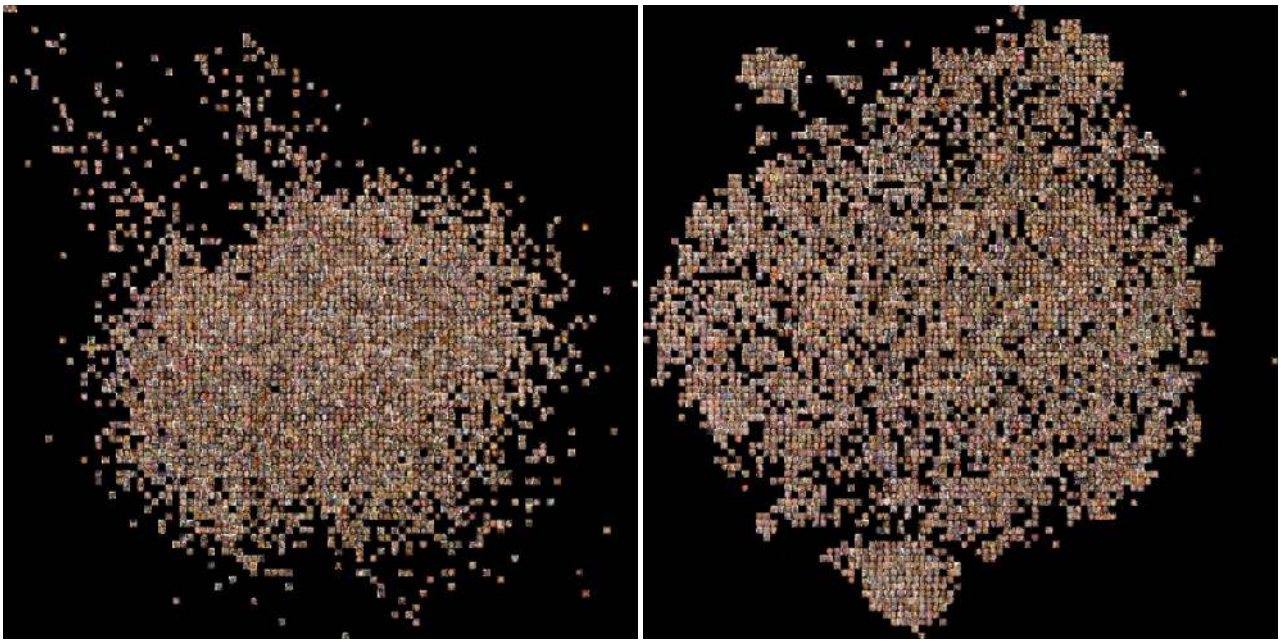
(c) Dimensionality of Swiss Roll

Figure 8: Intrinsic Dimensionality of Swiss Roll



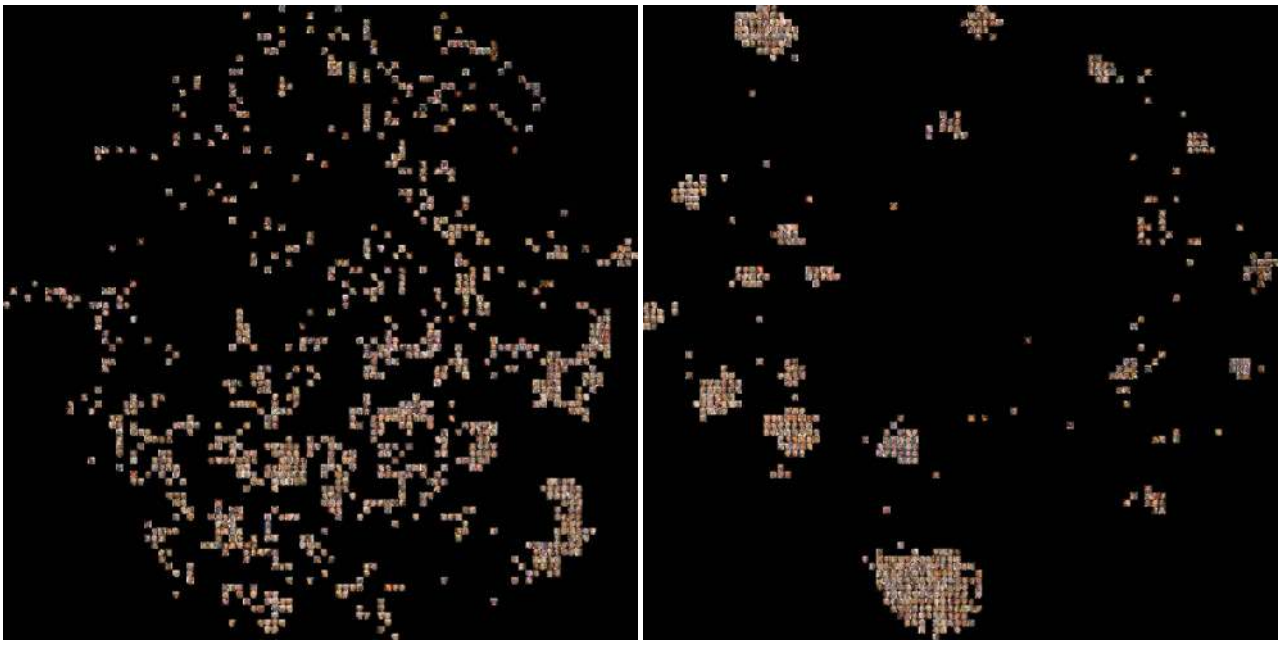
(a) Swiss Roll (b) Projection of Isomap (c) Projection of DeepMDS

Figure 9: **Swiss Roll:** (a) the original 2000 points from the swiss roll manifold, (b) the 2-dim intrinsic space estimated by Isomap, and (3) the 2-dim intrinsic space estimated by our proposed method DeepMDS. In both cases, the blue and black points, and correspondingly green and red points, are close together in both the intrinsic and ambient space.



(a) 2-D (b) 10-D (Intrinsic)

Figure 10: Visualization of Embeddings on LFW Dataset



(a) 2-D (b) 10-D (Intrinsic)

Figure 11: Visualization of Embeddings on LFW Dataset: subset of images from Fig. 10. The subset consists of 20 subjects from LFW dataset.