

Manifold Topology Divergence: a Framework for Comparing Data Manifolds

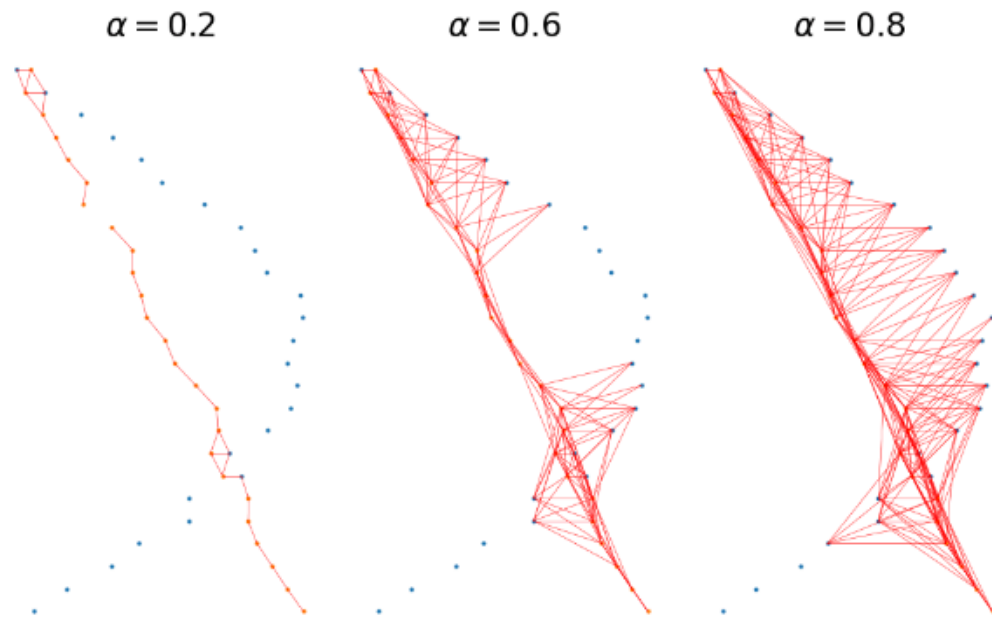
Barannikov, S., Trofimov, I., Sotnikov, G., Trimbach, E., Korotin, A., Filippov, A., & Burnaev, E.

Skoltech

Skolkovo Institute of Science and Technology



Framework for Comparing Data Manifolds



We develop a framework for comparing data manifolds, aimed, in particular, towards the evaluation of deep generative models. We describe a novel tool, Cross-Barcode(P, Q), that, given a pair of distributions in a high-dimensional space, tracks multi-scale topology spacial discrepancies between manifolds on which the distributions are concentrated. It is one of the first TDA-based practical methodologies that can be applied universally to datasets of different sizes and dimensions.

$$P \sim \mathcal{P}_{\text{data}} \quad Q \sim \mathcal{Q}_{\text{model}}$$

Comparing Data Manifolds

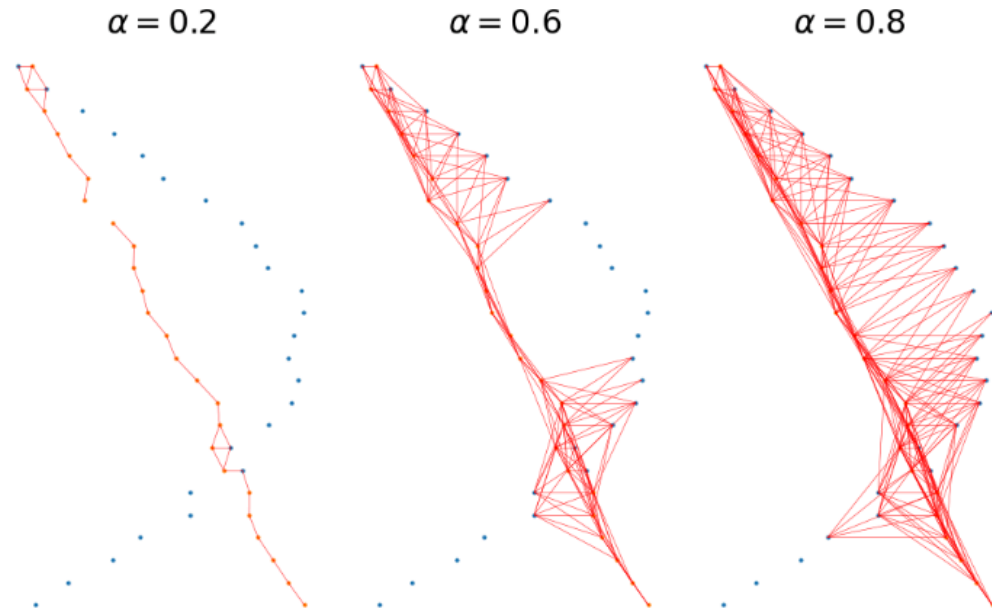


Figure 1: Edges(red) connecting P -points(red) with Q -points(blue), and also P -points between them, are added for three thresholds: $\alpha = 0.2, 0.4, 0.6$

$$P \sim \mathcal{P}_{\text{data}} \quad Q \sim \mathcal{Q}_{\text{model}}$$

IDEA: compare manifolds

$$M_{\text{data}}, M_{\text{model}}$$

via calculating topological features of

$$(M_{\text{data}} \cup M_{\text{model}}) / M_{\text{model}}$$

$$\text{and } (M_{\text{data}} \cup M_{\text{model}}) / M_{\text{data}}$$

in the setting of manifolds represented by point clouds

Comparing Data Manifolds

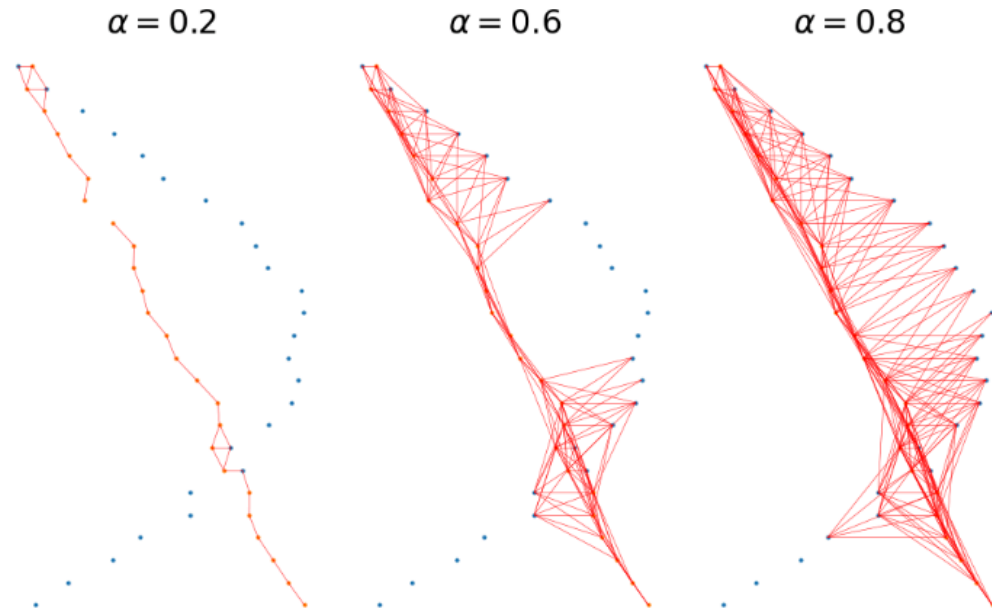


Figure 1: Edges (red) connecting P -points (red) with Q -points (blue), and also P -points between them, are added for three thresholds: $\alpha = 0.2, 0.4, 0.6$

$$P \sim \mathcal{P}_{\text{data}} \quad Q \sim \mathcal{Q}_{\text{model}}$$

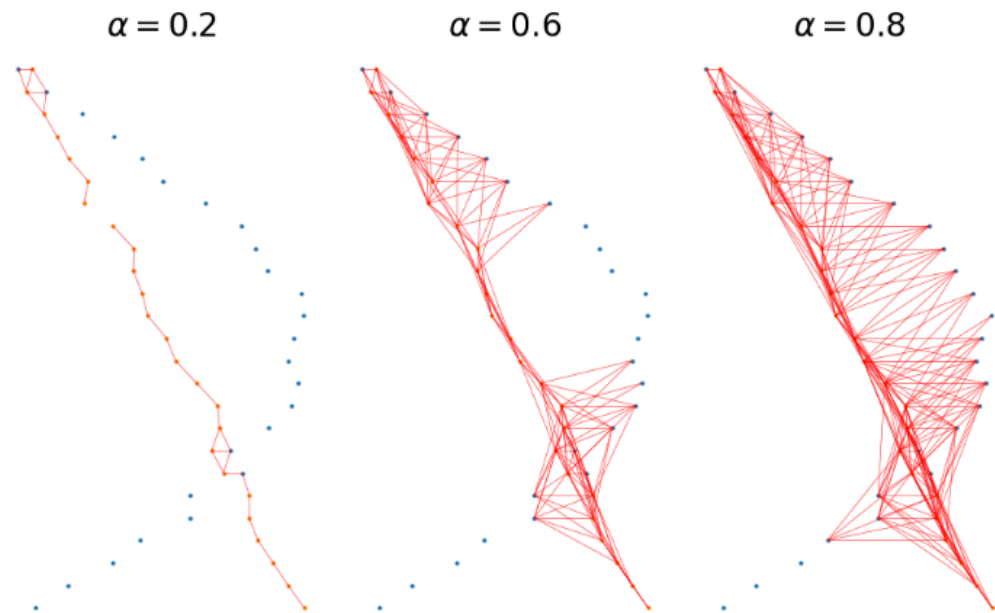
To calculate topological features of $(M_{\text{data}} \cup M_{\text{model}}) / M_{\text{model}}$ in the setting of manifolds represented by point clouds

we replace $(M_{\text{data}} \cup M_{\text{model}})$ by simplicial approximations at varying scale $\alpha > 0$

by picking simplexes with vertices in samples $P \cup Q$ with edges not exceeding α , for all $\alpha > 0$

and then for the calculation of topological features, as the counterpart of taking quotient by M_{model} , we set the distances within Q -cloud **to zero**

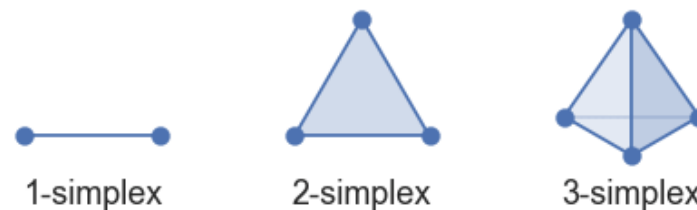
Comparing Data Manifolds



Our method measures the differences in the simplicial approximation of the two manifolds, represented by samples P and Q , by constructing sets of simplices, describing discrepancies between the two manifolds. To construct these sets of simplices we take the edges connecting P -points with Q -points, and also P -points between them, ordered by their length, and start adding these edges one by one, beginning from the smallest edge and gradually increasing the threshold, see Figure 1.

Figure 1: Edges (red) connecting P -points (red) with Q -points (blue), and also P -points between them, are added for three thresholds: $\alpha = 0.2, 0.4, 0.6$

$$P \sim \mathcal{P}_{\text{data}} \quad Q \sim \mathcal{Q}_{\text{model}}$$



Comparing Data Manifolds

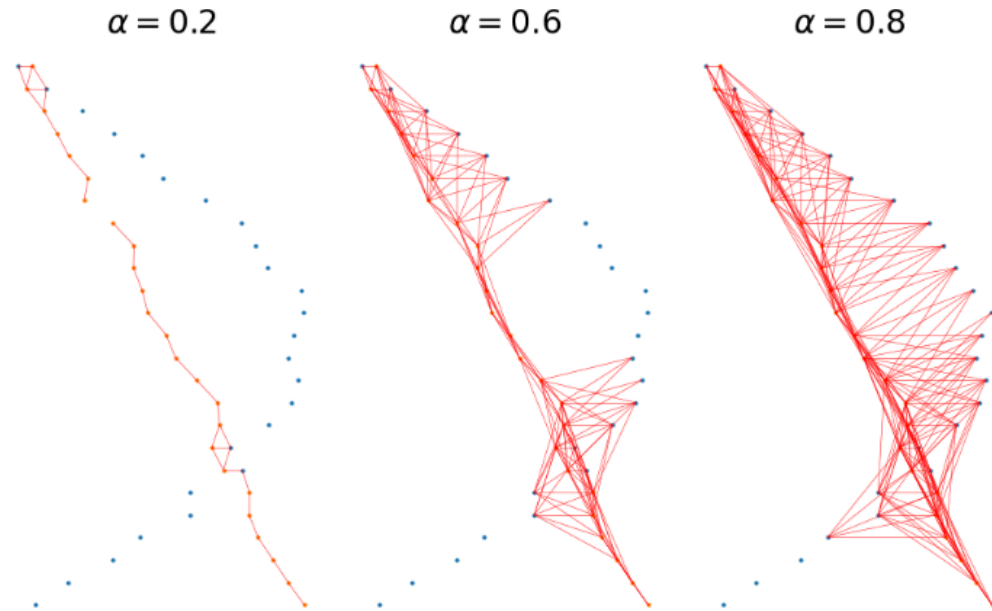


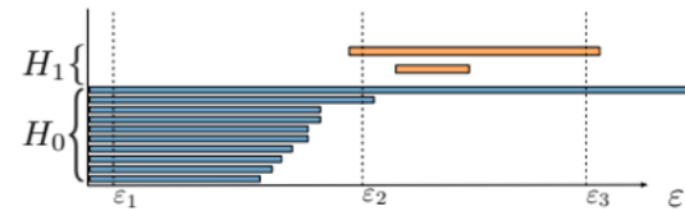
Figure 1: Edges (red) connecting P -points (red) with Q -points (blue), and also P -points between them, are added for three thresholds: $\alpha = 0.2, 0.4, 0.6$

$$P \sim \mathcal{P}_{\text{data}} \quad Q \sim \mathcal{Q}_{\text{model}}$$

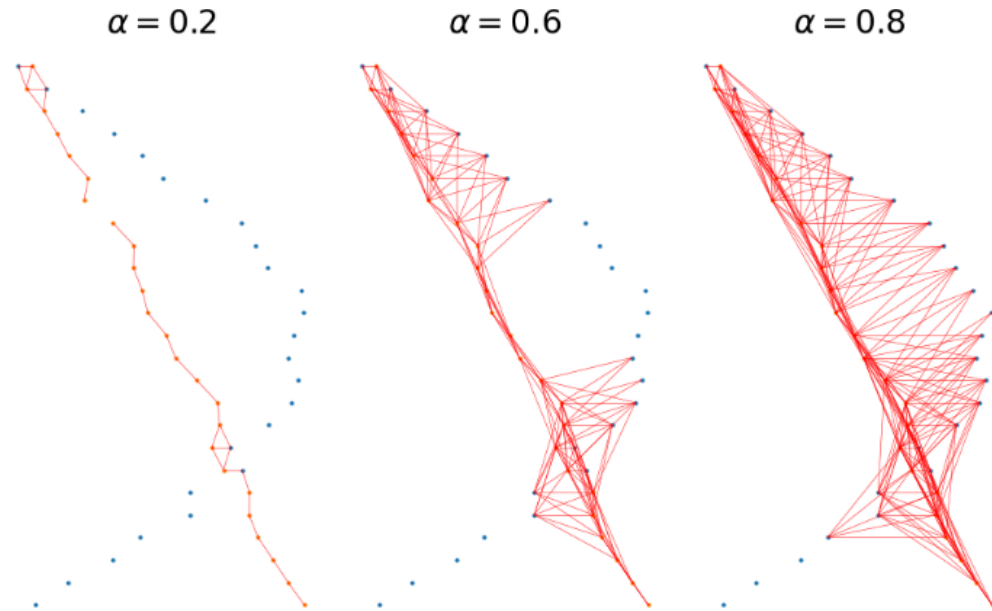
We track in this process the appearance of nontrivial k -cycles, i.e. collections of k -simplexes formed by some P - and Q -points, such that the simplexes' boundaries cancel each other. Any collection with boundary in Q can be completed to a cycle since all simplexes formed only by Q -points have been added at zero scale.

We track the scales at which such nontrivial cycles appear and disappear.

The longer the lifespan of such topological feature across the change of threshold the bigger the discrepancy between the two manifolds



Comparing Data Manifolds



The process of adding longer edges can be visually assimilated to the building of a "spider's web" that tries to bring the cloud of red points closer to the cloud of blue points.

Figure 1: Edges(red) connecting P -points(red) with Q -points(blue), and also P -points between them, are added for three thresholds: $\alpha = 0.2, 0.4, 0.6$

$$P \sim \mathcal{P}_{\text{data}} \quad Q \sim \mathcal{Q}_{\text{model}}$$

Cross-Barcode

Cross-barcode(P, Q) is calculated for two point clouds: P and Q .

All the pairwise distances within Q are set to **zero**.

To define Cross-Barcode(P, Q) we construct first the following filtered simplicial complex. Let $(\Gamma_{P \cup Q}, m_{(P \cup Q)/Q})$ be the weighted graph with the distance-like weights on edges defined as the complete graph on the union of point clouds $P \cup Q$ with the distance matrix given by the pairwise distance in \mathbb{R}^D for the pairs of points (p_i, p_j) or (p_i, q_j) and with all pairwise distances within the cloud Q that we set to zero. Our filtered simplicial complex is the Vietoris-Rips complex of $(\Gamma_{P \cup Q}, m_{(P \cup Q)/Q})$.

The Vietoris-Rips complex $R_\alpha(\Gamma, m)$ is the abstract simplicial complex with simplices that correspond to the non-empty subsets of vertices of Γ whose pairwise distances are less than α as measured by m . Increasing parameter α adds more simplices and this gives a nested family of collections of simplices known as filtered simplicial complex.

Definition The Cross-Barcode $_i(P, Q)$ is the set of intervals recording the “appearance” and “disappearance” scales of i -dimensional topological features in the filtered simplicial complex $R_\alpha(\Gamma_{P \cup Q}, m_{(P \cup Q)/Q})$.

Cross-Barcode

Cross-barcode(P, Q) is calculated for two point clouds: P and Q .
All the pairwise distances within Q are set to **zero**.

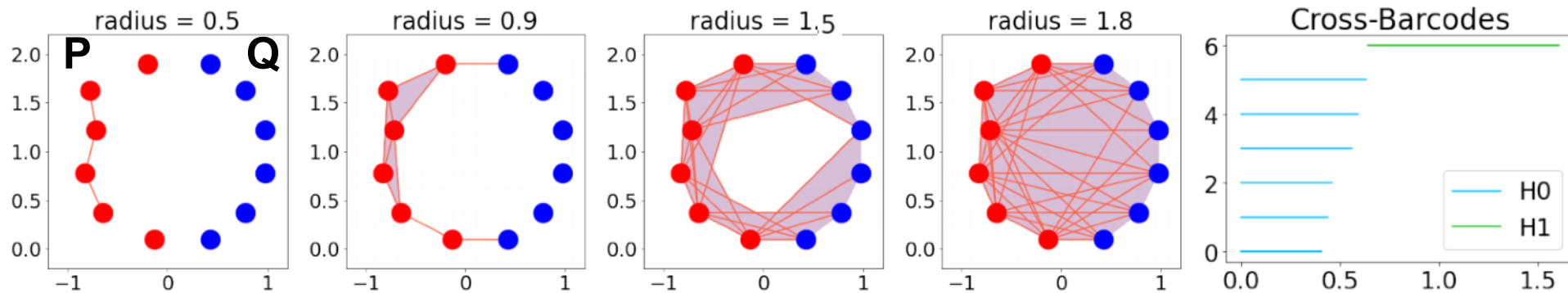


Figure 11: The process of adding the simplices between the P -cloud (red) and Q -cloud (blue) and within the P -cloud. Here we show the consecutive adding of edges together with simultaneous adding of triangles. All the edges and simplices within Q -cloud are assumed added at $\alpha = 0$ and are not shown here for perception's ease. Notice the 1-cycle born between $\alpha = 0.5$ and 0.9 , it corresponds to the green segment in the shown Cross-Barcode

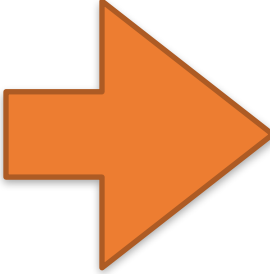
Cross-Barcode: the Algorithm

Algorithm 1 Cross-Barcode_{*i*}(*P*, *Q*)

Input: $m[P, P]$, $m[P, Q]$: matrices of pairwise distances within point cloud *P*, and between point clouds *P* and *Q*

Require: $\text{VR}(M)$: function computing filtered complex from pairwise distances matrix *M*

Require: $\text{B}(C, i)$: function computing persistence intervals of filtered complex *C* in dimension *i*



$b_Q \leftarrow$ number of columns in matrix $m[P, Q]$
 $m[Q, Q] \leftarrow$ zeroes(b_Q, b_Q)
 $M \leftarrow \begin{pmatrix} m[P, P] & m[P, Q] \\ m[P, Q] & m[Q, Q] \end{pmatrix}$

Cross-Barcode_{*i*} $\leftarrow \text{B}(\text{VR}(M), i)$

Return: list of intervals **Cross-Barcode**_{*i*}(*P*, *Q*) representing "births" and "deaths" of topological discrepancies

Basic Properties of Cross-Barcode

1. $\text{Cross-Barcode}(P, P) = \emptyset$
2. $\text{Cross-Barcode}(P, \emptyset) = \text{Barcode}(P)$
3. $\text{Cross-Barcode}(P, Q)$ is not symmetric
4. $\|\text{Cross-Barcode}(P, Q)\|_B$ is bounded from above by the Hausdorff distance between P and Q , where $\|\cdot\|_B$ is the bottleneck distance.

Manifold Topology Divergence (MTop-Div)

MTop-Div(P, Q) by definition equals the **sum of lengths of segments** in Cross-Barcode₁(P, Q)

Proposition. MTop-Div(P, Q) equals the Earth-Mover's Distance between Relative Living Time histogram for the Cross-Barcode₁(P, Q) and the histogram of the empty barcode, multiplied by the parameter α_{\max} from the definition of RLT.

Relative Living Times is a discrete distribution $RLT(k)$ over non-negative integers $k \in \{0, 1, \dots, +\infty\}$. For a given $\alpha_{\max} > 0$, $RLT(k)$ is a fraction of “time”, that is, parts of horizontal axis $\tau \in [0, \alpha_{\max}]$, such that exactly k segments $[b_i, d_i]$ include τ .

MTop-Divergence: the Algorithm

Algorithm 2 MTop-Divergence(\mathcal{P}, \mathcal{Q}), see section 2.6 for details, default suggested values: $b_{\mathcal{P}} = 1000, b_{\mathcal{Q}} = 10000, n = 100$

Input: $X_{\mathcal{P}}, X_{\mathcal{Q}}: N_{\mathcal{P}} \times D, N_{\mathcal{Q}} \times D$ arrays representing datasets

for $j = 1$ **to** n **do**

$P_j \leftarrow$ random choice($X_{\mathcal{P}}, b_{\mathcal{P}}$)

$Q_j \leftarrow$ random choice($X_{\mathcal{Q}}, b_{\mathcal{Q}}$)

$\mathcal{B}_j \leftarrow$ list of intervals Cross-Barcode₁(P_j, Q_j) calculated by Algorithm 1

$mtd_j \leftarrow$ sum of lengths of all intervals in \mathcal{B}_j

end for

MTop-Divergence(\mathcal{P}, \mathcal{Q}) \leftarrow mean(mtd)

Return: number MTop-Divergence(\mathcal{P}, \mathcal{Q}) representing discrepancy between the distributions \mathcal{P}, \mathcal{Q}

Evaluation of Generative models. Methods:

We have compared the MTop-Div against 7 established evaluation methods: FID, discriminative score, MMD, JSD, 1-coverage, IMD and Geometry score and found that MTop-Div outperforms many of them and captures well subtle differences in data manifolds.

For images, the Fréchet Inception Distance (**FID**) is a distance between two multivariate Gaussians. These Gaussians approximate the features of generated and true data extracted from the last hidden layer of the pretrained Inception network. FID is the most popular GAN evaluation measure. However, FID is limited only to 2D images since it relies on pre-trained on ImageNet "Inception" network. FID unrealistically approximates point clouds by Gaussians in embedding space. Surprisingly, FID can't be applied to compare adversarial and non-adversarial generative models since it is overly pessimistic to the latter ones.

The Geometry Score (**GScore**) is the L2-distance between mean Relative Living Times (RLT) of topological features calculated for the model distribution and the true data distribution. The GScore is domain agnostic, does not involve auxiliary pretrained networks and is not limited to 2D images. However, GScore is not sensitive even to some simple transformations - like constant shift, dilation, or reflection. The barcodes in GScore are calculated approximately, based only on the approximate witness complexes on 64 landmark points sampled from each distribution. That's why the procedure is stochastic and should be repeated several thousand times for averaging. Thus, the calculation of GScore can be prohibitively long for large datasets.

Experiments. Pairs of rings.

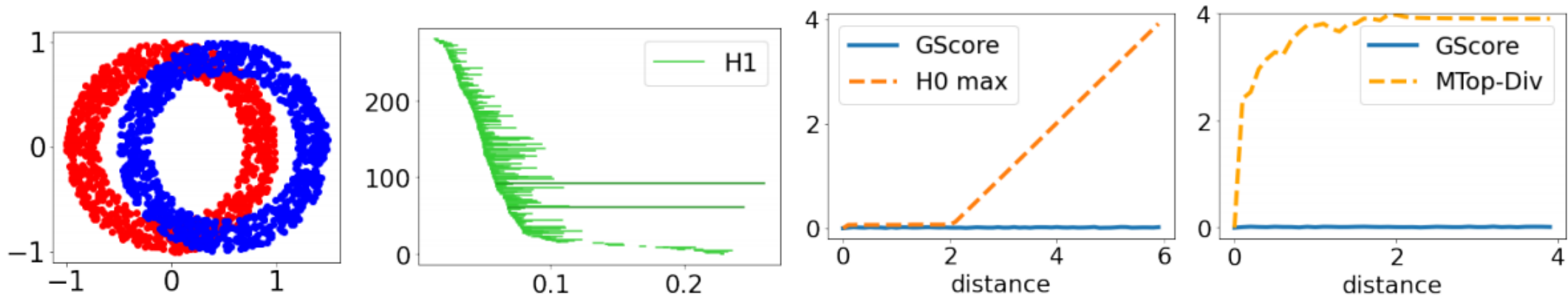
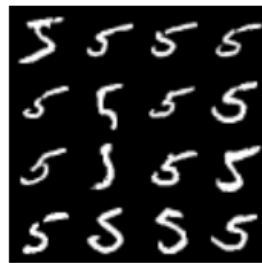
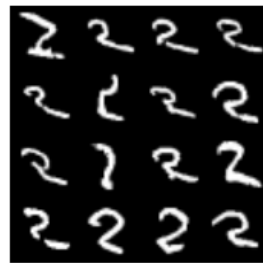


Figure 2: MTop-Div and H0 max compared with GScore, for two ring clouds of 1000 points, as function of $d = \text{distance}$ between ring centers, the $\text{Cross-Barcode}_1(P, Q)$ is shown at $d = 0.5$

Experiments. '5's vs. flipped '5's.



vs



Geometry Score = 0.0
MTop-Div = 6154.0

Figure 4: Two point clouds: “5”s from MNIST vs. vertically flipped “5”s from MNIST (resembling rather “2”s). The two clouds are indistinguishable for Geometry Score, while the MTop-Divergence is sensitive to such flip as it depends on the positions of clouds with respect to each other.

Experiments. Modifications of CIFAR10.

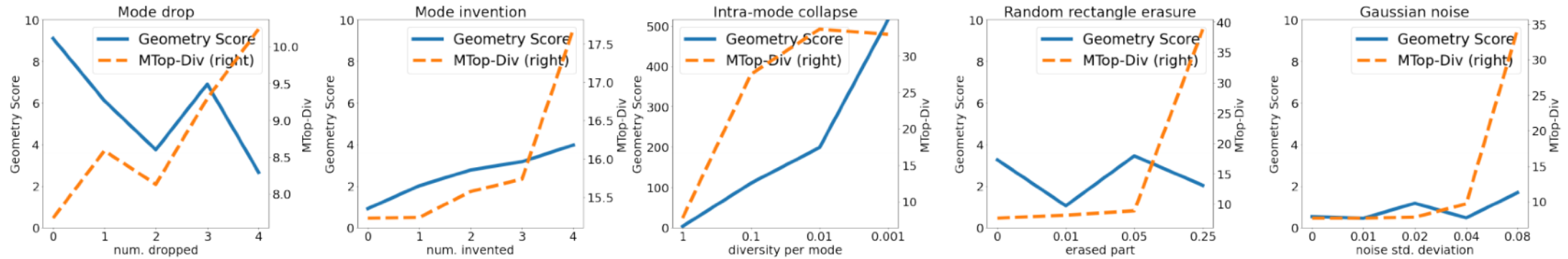


Figure 5: Experiment with modifications of CIFAR10. The disturbance level rises from zero to a maximum. Ideally, the quality score should monotonically increase with the disturbance level.

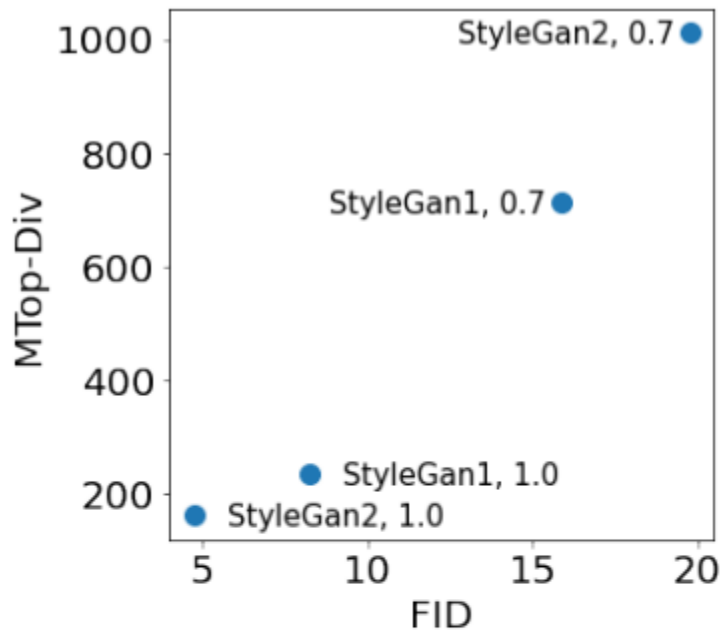
The average Kendall-tau rank correlation between MTop-Div(D,M) and disturbance level is **0.89**, while for Geometry Score the rank correlation is only 0.36. FID performs well on this benchmark, not shown for ease of perception

Experiments. GAN model selection.

| Dataset | FID | | MTop-Div(D,M) | |
|--------------|--------------|-------------|---------------|---------------|
| | WGAN | WGAN-GP | WGAN | WGAN-GP |
| CIFAR10 | 154.6 | 399.2 | 353.1 | 1637.4 |
| SVHN | 101.6 | 154.7 | 332.0 | 963.2 |
| MNIST | 31.8 | 22.0 | 2042.8 | 1526.1 |
| FashionMNIST | 52.9 | 35.1 | 919.6 | 660.4 |

Table 1: MTop-Div is consistent with FID for model selection of GAN's trained on various datasets.

Experiments. StyleGAN, StyleGAN2.



Comparison of two quality measures: FID vs. MTop-Div on StyleGAN, StyleGAN2 trained on FFHQ with different truncation levels.

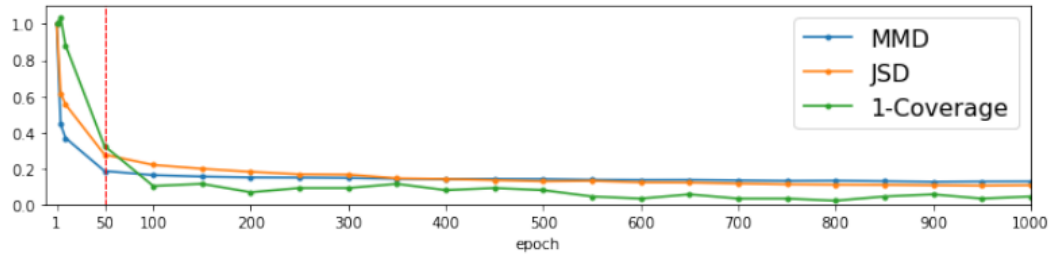
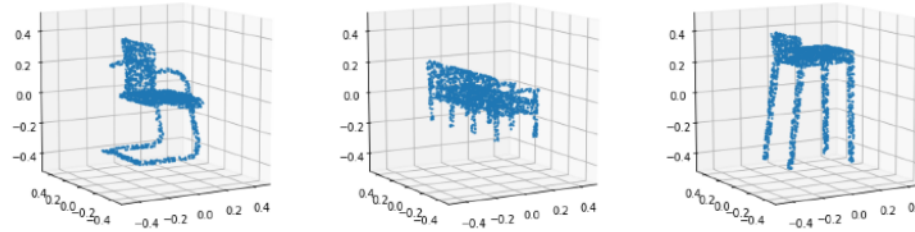
MTop-Div is monotonically increasing in good correlation with FID.

FFHQ is the largest dataset in our experiments, number of samples = $2 \cdot 10^4$, $D=10^7$

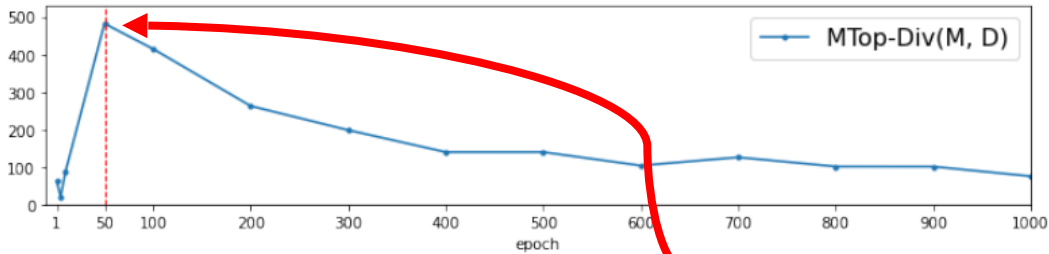
Calculation of MTop-Div took 30 sec.

Calculation of Geometry Score didn't finish in a reasonable time.

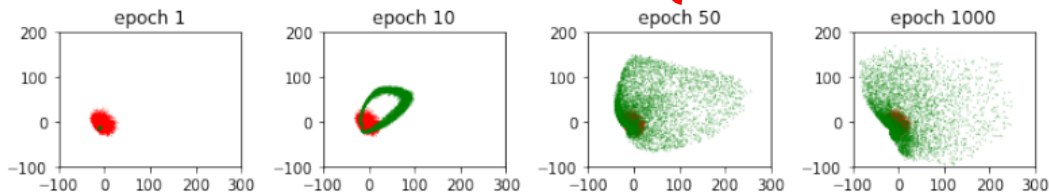
Experiments. 3D GAN.



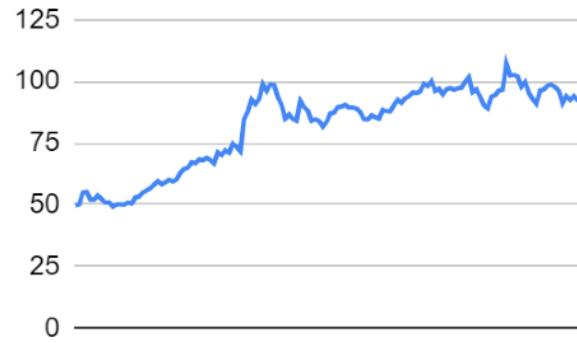
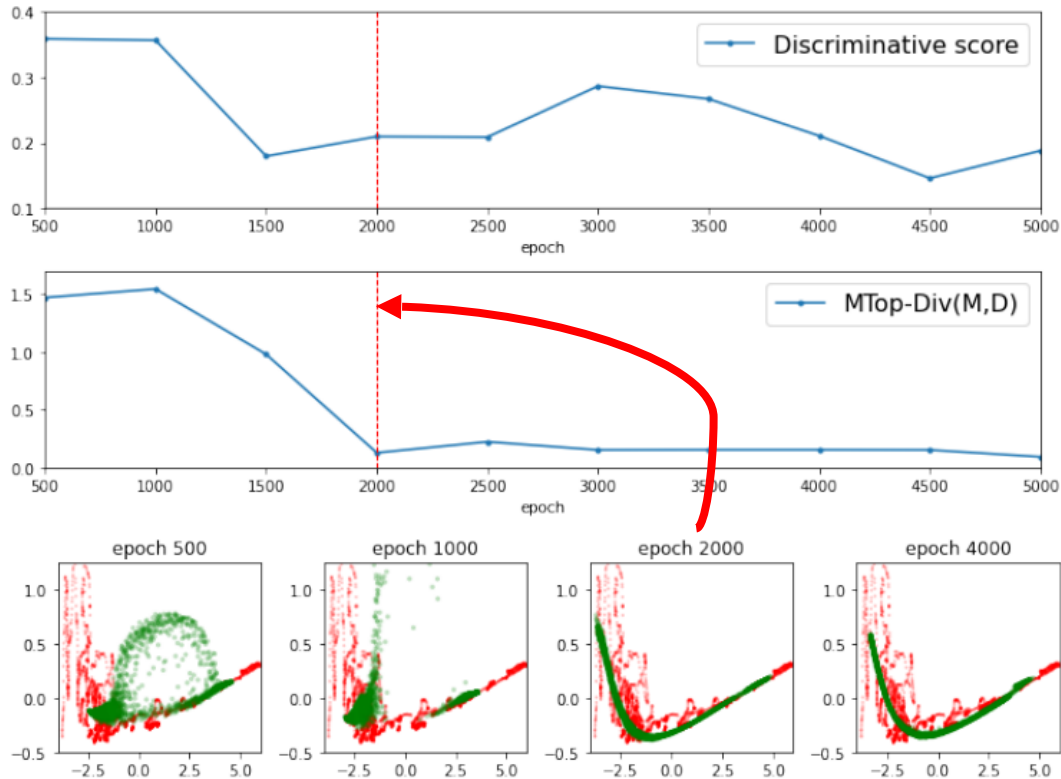
Training process of GAN applied to 3D shapes. Normalized quality measures MMD, JSD, 1-Coverage, MTop-Div vs. epoch. Lower is better. MTop-Div is more sensitive than standard quality measures.



PCA projection of real objects (red) and generated objects (green). Vertical red line (epoch 50) depicts the moment, when the manifold of generated objects “explodes” and becomes much more diverse.



Experiments. TimeGAN.



Training dynamics of TimeGAN applied to market stock data. Discriminative score vs. epoch, MTop-Div vs. epoch. Lower is better. MTop-Div agrees with discriminative score.

PCA projection of real time-series (red) and generated time-series (green). Vertical red line (epoch 2000) depicts the moment when manifolds of real and generated objects become close.

Experiments. Chest X-ray data.

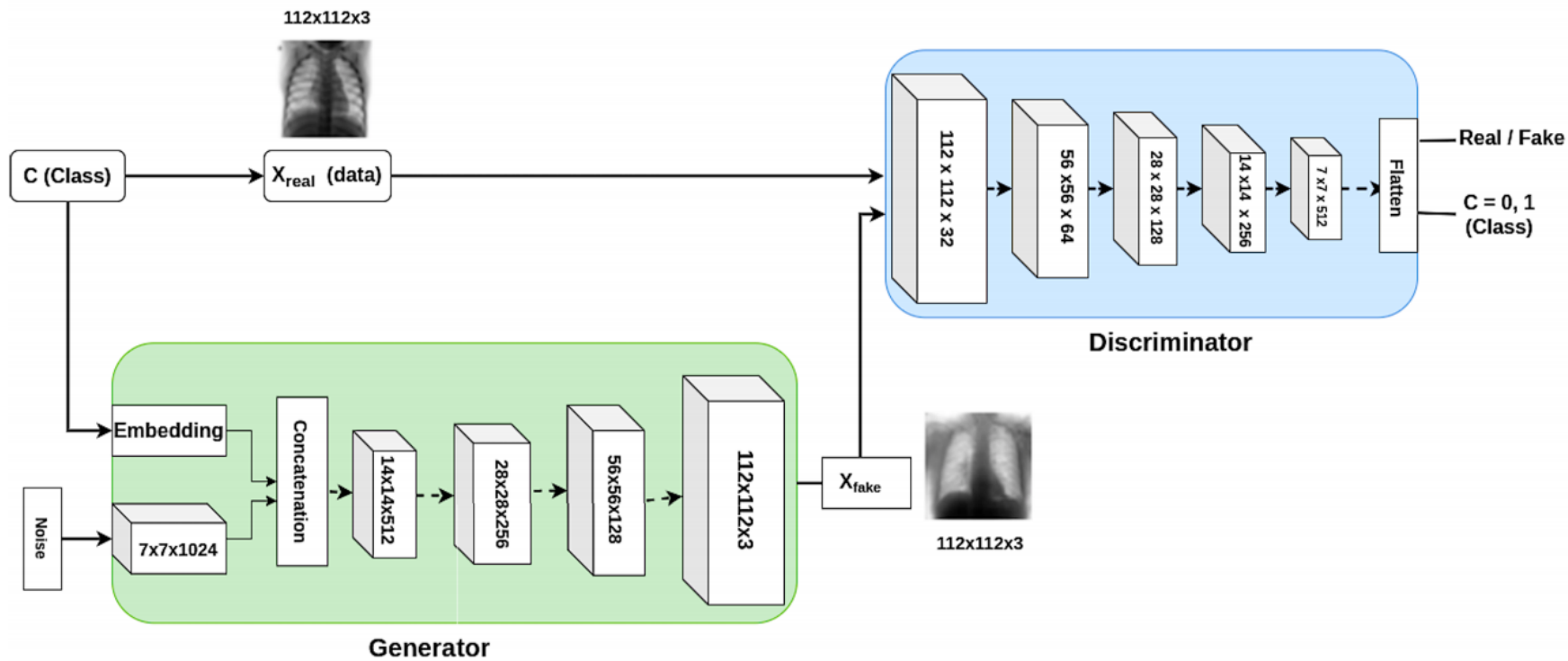
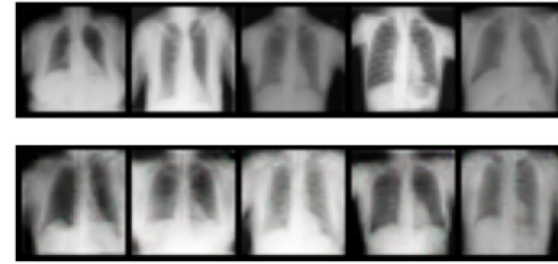


FIGURE 2. CovidGAN complete Architecture with generator and discriminator.

CovidGAN generates images to augment chest X-ray dataset for COVID diagnosis.

Waheed, A., Goyal, M., Gupta, D., Khanna, A., Al-Turjman, F., & Pinheiro, P. R. (2020). CovidGAN: data augmentation using auxiliary classifier gan for improved covid-19 detection. *IEEE Access*, 8, 91916-91923.

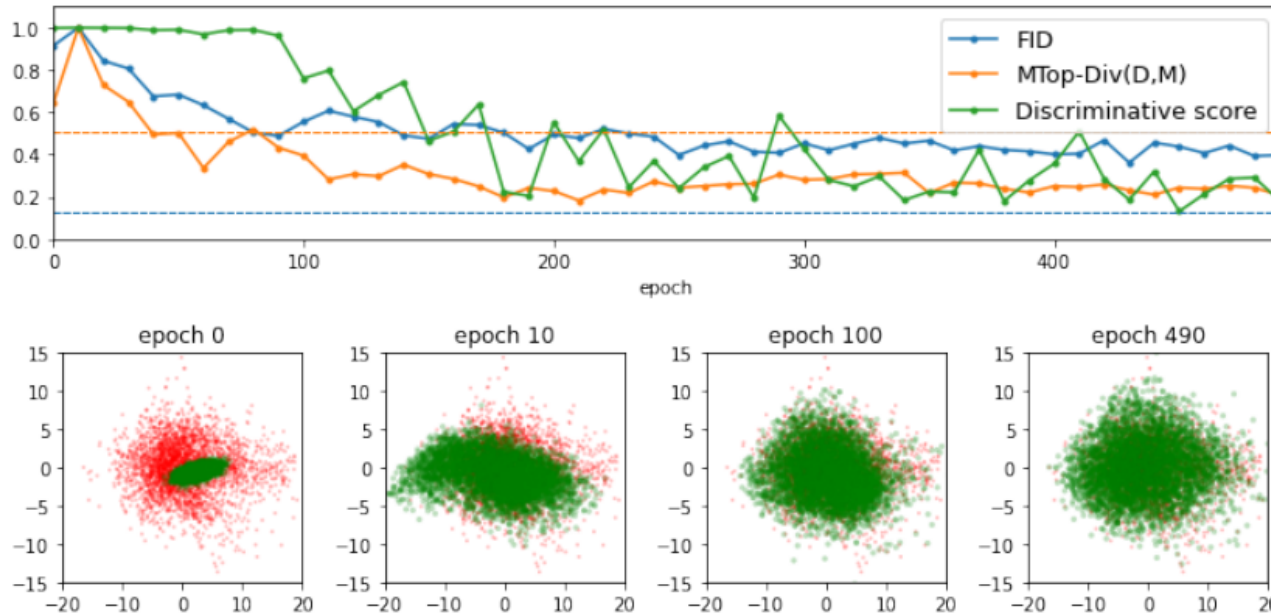
Experiments. Chest X-ray data.



Training process of CovidGAN applied to chest X-ray data. Normalized quality measures FID, MTop-Div, Disc. score vs. epoch. Lower is better.

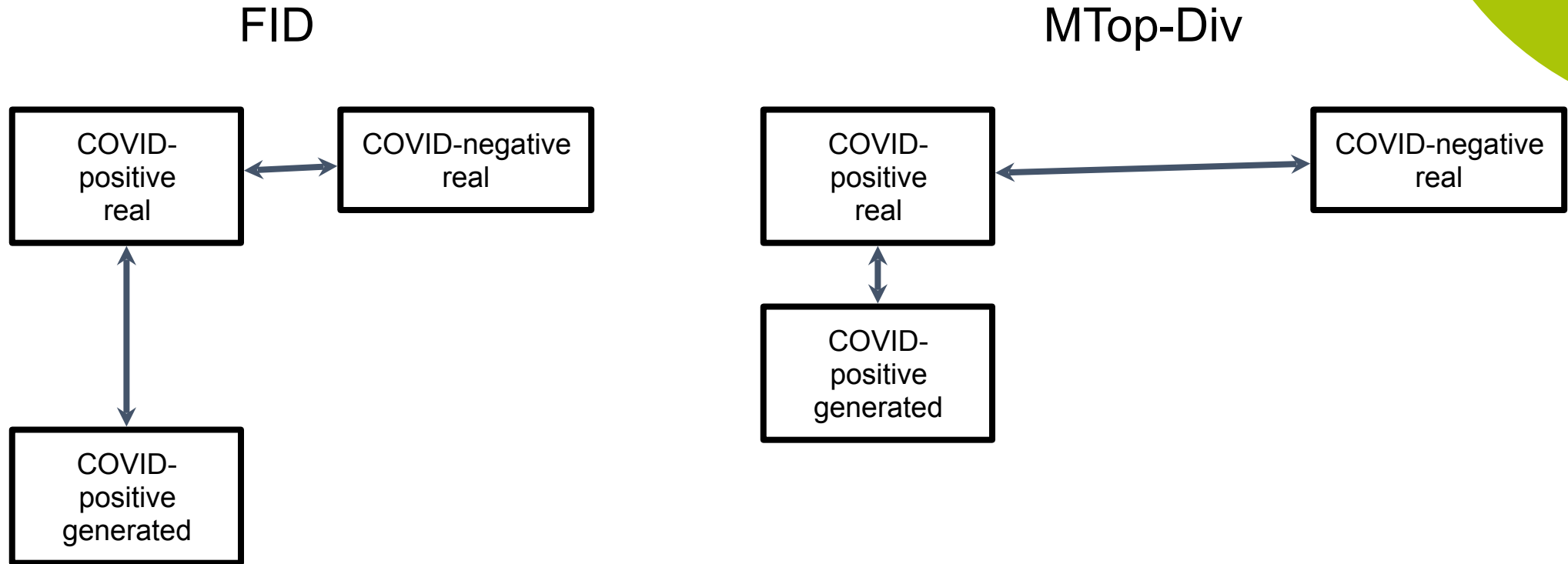
MTop-Div agrees with standard measures.

PCA projections of real objects (red) and generated objects (green).



Dashed horizontal lines depict comparison of real COVID-positive and COVID-neg. chest X-rays.

Experiments. Chest X-ray data.



Counterintuitively, for FID real COVID-positive images are closer to real COVID-negative ones than to generated COVID-positive images.

Probably because FID is overly sensitive to textures.

Evaluation by MTop-Div is consistent

Conclusions

1. We introduced a new tool: Cross-Barcode(P, Q). For a pair of point clouds P and Q , the Cross-Barcode(P, Q) records the differences in multiscale topology between two manifolds approximated by the point clouds;
2. We proposed a new measure for comparing two data manifolds approximated by point clouds: Manifold Topology Divergence (MTop-Div);
3. We applied the MTop-Div to evaluate performance of GANs in various domains: 2D images, 3D shapes, time-series. We show that the MTop-Div correlates well with domain-specific measures and can be used for model selection. Also it provides insights about evolution of generated data manifold during training;

Conclusions

4. We have compared the MTop-Div against 6 established evaluation methods: FID, discriminative score, MMD, JSD, 1-coverage, and Geometry score and found that MTop-Div is able to capture subtle differences in data geometry;
5. We have essentially overcome the known TDA scalability issues and in particular have carried out the MTop-Div calculations on most recent datasets such as FFHQ, with dimensions D up to 10^7

Thank you for your attention!