TopoFR: A Closer Look at Topology Alignment on Face Recognition

Jun Dan[1,2] **Yang Liu**[2,3] **Jiankang Deng**[4] **Haoyu Xie**[2,5]

Siyuan Li[2] **Baigui Sun**[2,5] **Shan Luo**[3]

[1]Zhejiang University [2]FaceChain Community [3]King's College London

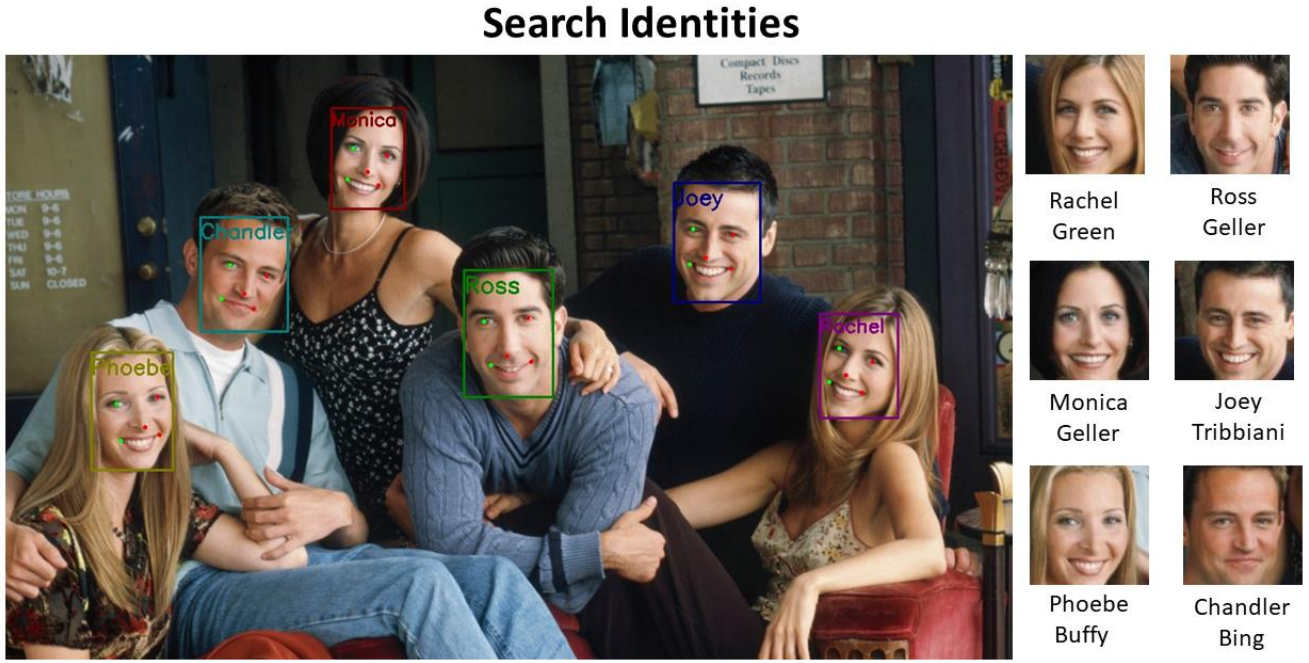
[4]Imperial College London [5]Alibaba Group

Motivation

Existing studies on FR primarily focuses on constructing more discriminative face features by developing:

1) margin-based loss functions

2) powerful network architectures



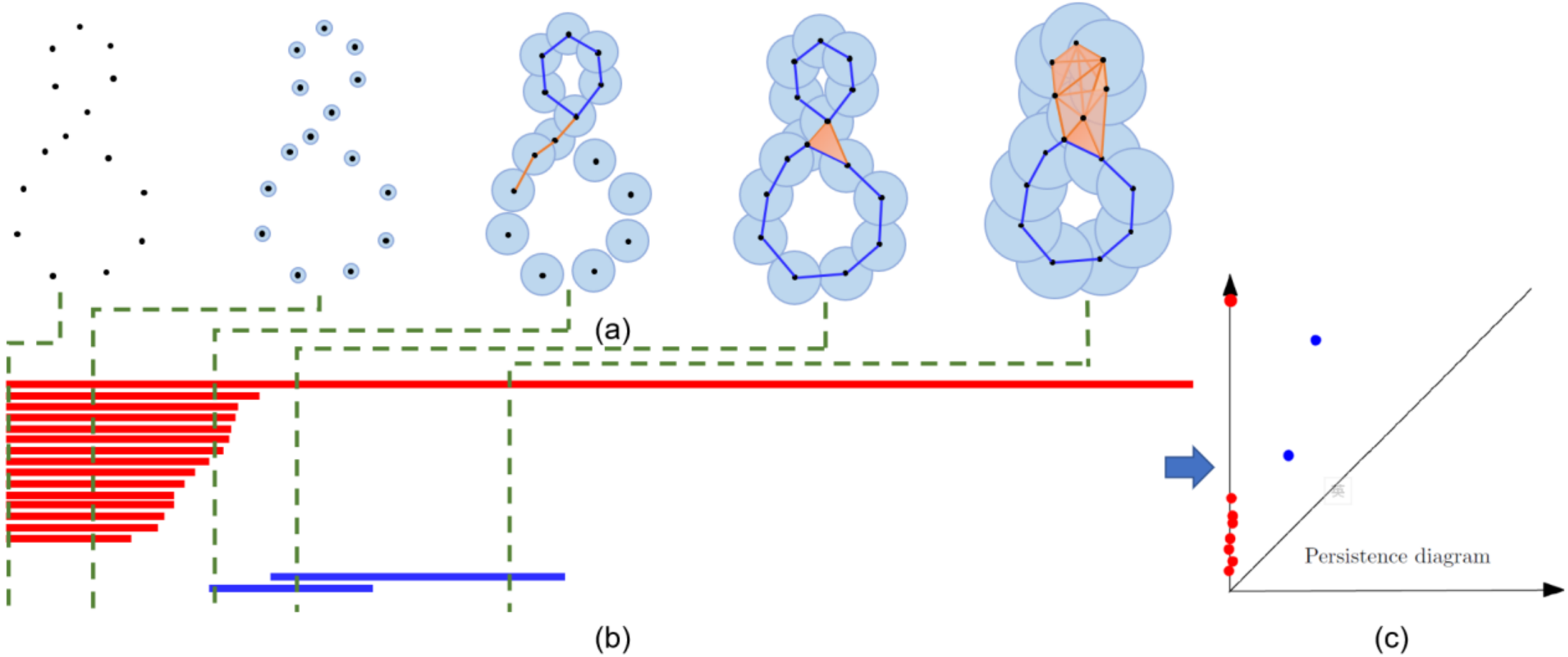
Recently, the success of unsupervised learning and graph neural networks has demonstrated the importance of data structure information in improving model generalization.

Considering that the FR task can leverage large-scale training data, which intrinsically contains significant structure information. Thus, in this paper, we extend our interests on building a cutting-edge FR framework through exploiting such powerful and substantial **structure information**.

Topological Data Analysis

Persistent Homology (PH) is a method in computational topology used to analyze and capture the underlying topological structure information of complex point clouds.

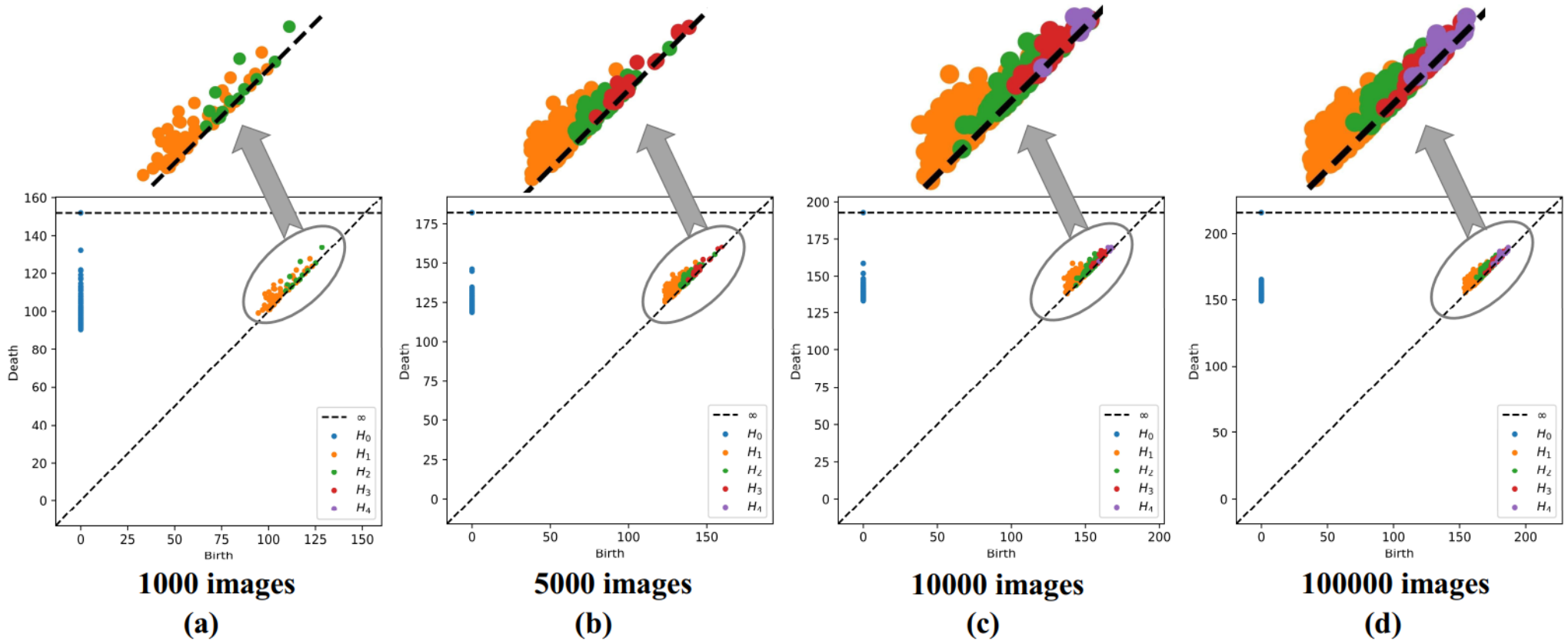
Vietoris–Rips Complex



Novel Findings

We use Persistent Homology (PH) to investigate the evolution trend of structure information in existing FR framework and illustrate **3 interesting findings**:

(i) As the amount of data increases, the topological structure of the input space becomes more and more complex.



Novel Findings

(ii) **As the amount of data increases, the topological structure discrepancy between the input space and the latent space becomes increasingly larger.**

(iii) **As the depth of the network increases, the topological structure discrepancy becomes progressively smaller.**

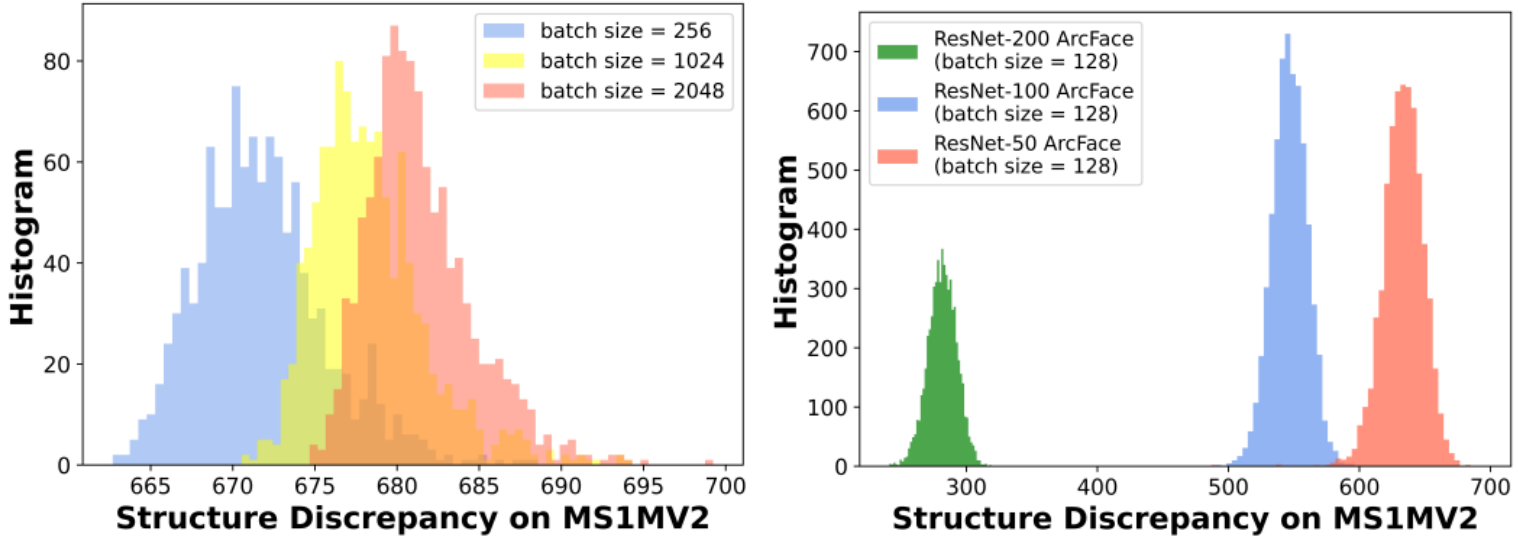
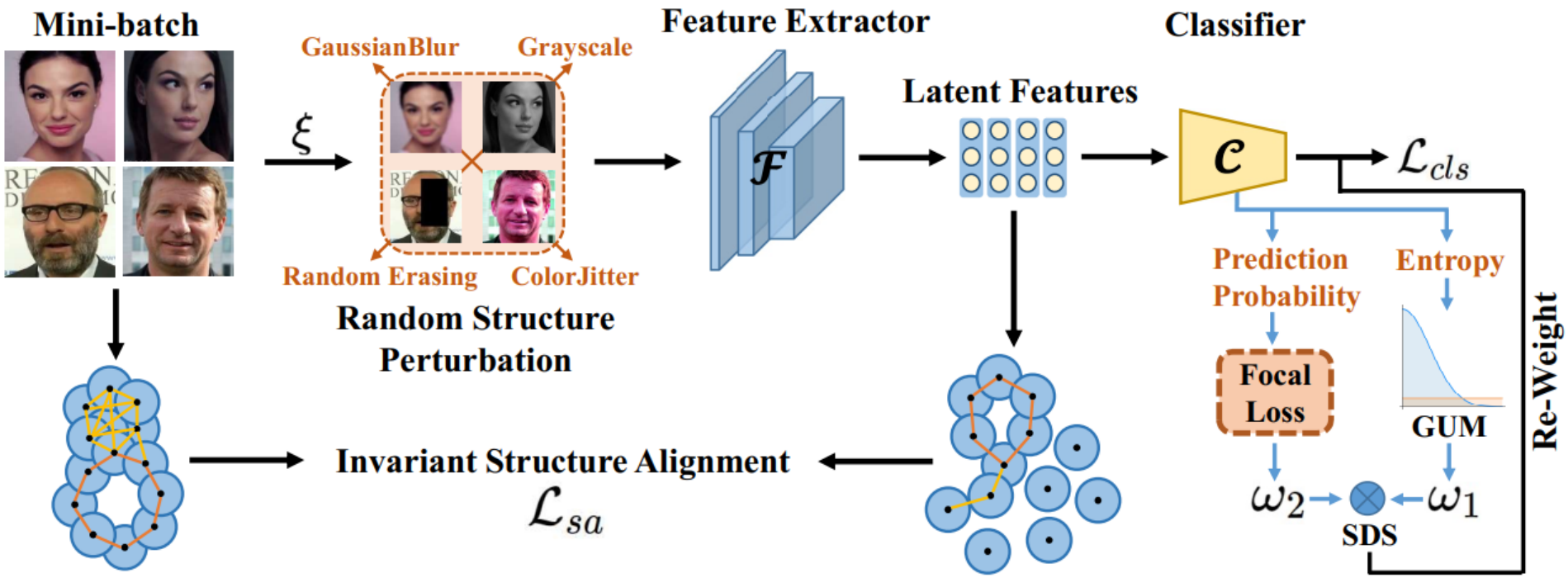


Figure 2: **(a):** We investigate the relationship between the amount of data and the topological structure discrepancy by employing ResNet-50 ArcFace model [1] to perform inferences on MS1MV2 training set. Inferences are conducted for 1000 iterations with batch sizes of 256, 1024, and 2048, respectively. Histograms are used to approximate these discrepancy distributions. **(b):** We investigate the relationship between the network depth and the topological structure discrepancy by performing inference on MS1MV2 training set (batch size=128) using ArcFace models with different backbones.

Methodology

In FR tasks with large-scale datasets, the structure of face data will be severely destroyed during training, which limits the generalization ability of FR models in practical application scenarios.

A fundamental idea is to align the structures of the input and latent spaces in order to maximize the preservation of the topological structure information of face data.



Perturbation-guided Topological Structure Alignment (PTSA)

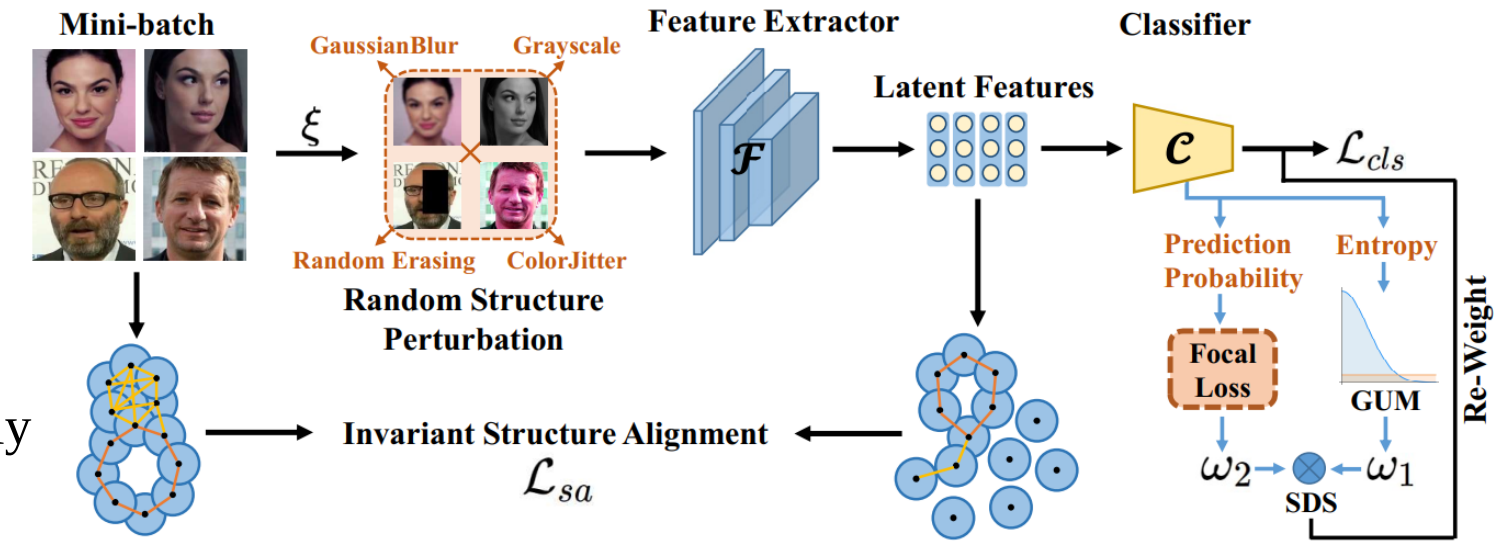
To remedy this problem, we propose a Perturbation-guided Topological Structure Alignment (**PTSA**) strategy that includes two mechanisms: Random Structure Perturbation **(RSP)** and Invariant Structure Alignment **(ISA)**.

RSP introduces a data augmentation list

$$\mathcal{A} = \{\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \mathcal{A}_4\}$$

For each training sample, RSP will randomly select an data augmentation operation to perturb the sample in order to increases the structure diversity of the latent space.

$$\widetilde{x}_i = \mathcal{A}_r(x_i)$$



We adopt ArcFace loss as the basic classification loss

$$\mathcal{L}_{arc}(\widetilde{x}_i, y_i) = -\log \frac{e^{s(\cos(\theta_i^y + m))}}{e^{s(\cos(\theta_i^y + m))} + \sum_{k=1, k\neq y}^{K} e^{s\cos\theta_i^k}}$$

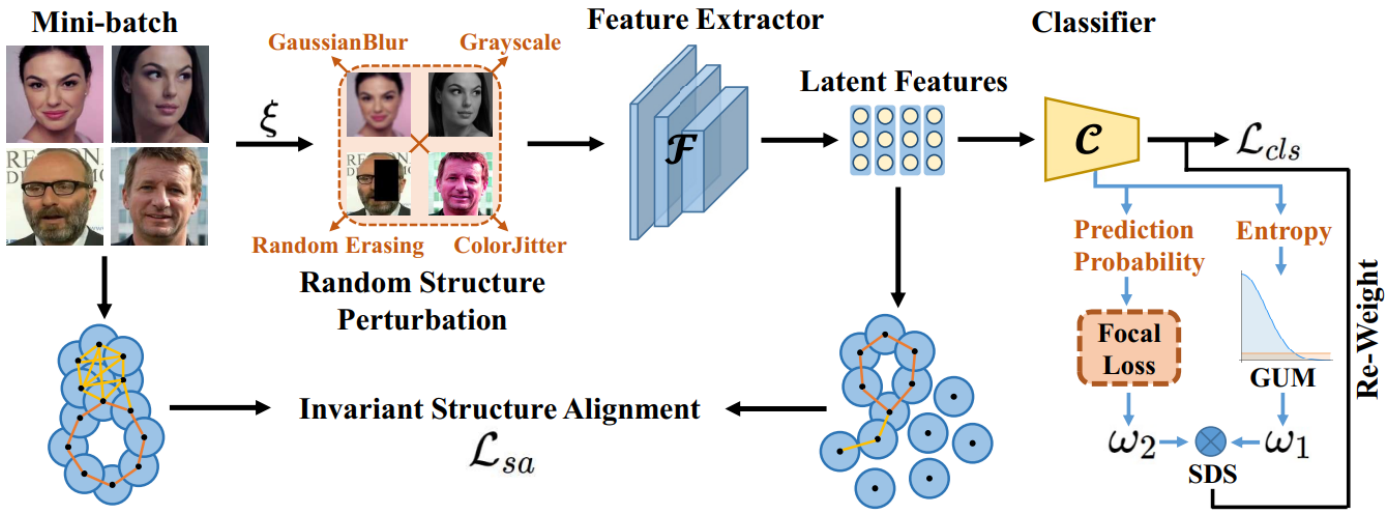
Perturbation-guided Topological Structure Alignment (PTSA)

During forward propagation, we can construct the Vietoris-Rips complexes for the original input space and the perturbed latent space. Then we can utilize persistent homology to analyze the topological structures of two complexes, and obtain their corresponding persistence diagrams and persistence pairing, respectively.

We choose to align the original input space with the perturbed latent space.

$$\mathcal{L}_{sa}(\mathcal{D}^{\mathcal{X}}, \mathcal{D}^{\widetilde{\mathcal{Z}}}) = \frac{1}{2}\left(\left\|\mathcal{M}^{\mathcal{X}}[\gamma^{\mathcal{X}}] - \mathcal{M}^{\widetilde{\mathcal{Z}}}[\gamma^{\mathcal{X}}]\right\|^2 + \left\|\mathcal{M}^{\widetilde{\mathcal{Z}}}[\gamma^{\widetilde{\mathcal{Z}}}] - \mathcal{M}^{\mathcal{X}}[\gamma^{\widetilde{\mathcal{Z}}}]\right\|^2\right)$$

Ideally, no matter how the face image is perturbed, the position of the encoded face feature in the latent space should remain unchanged, and the topological structure of the perturbed latent space should also be consistent with the original input space.



Structure Damage Estimation (SDE)

In practical FR scenarios, low-quality face samples, also known as "hard samples", are commonly included in the training set, which will disrupt the latent space's topological structure and further hinder the alignment of structures.

To address this issue, we propose a novel hard sample mining strategy called Structure Damage Estimation (SDE) to identify hard samples with serious structure damage and guide them back to the reasonable positions during optimization.

Prediction Uncertainty:

To accurately select hard samples, we propose using a Gaussian-uniform mixture (GUM) model to model sample difficulty, which utilizes prediction entropy as the distribution variable.

$$p\left(E(\widetilde{g}_i)|\widetilde{x}_i\right) = \pi\mathcal{N}^{+}(E(\widetilde{g}_i)|0,\Sigma) + (1-\pi)\mathcal{U}(0,\Omega),$$

$$\mathcal{N}^{+}(a|0,\Sigma) = \begin{cases} 2\mathcal{N}(a|0,\Sigma), & a \geq 0. \\ 0, & a < 0. \end{cases}$$

Structure Damage Estimation (SDE)

Then the posterior probability that the sample to be hard (*i.e.*, high-uncertainty) can be computed as follows:

$$h_\varphi(\widetilde{x}_i) = P_\varphi\left(u_i = 1|\widetilde{x}_i\right) = \frac{(1-\pi)\mathcal{U}(0,\Omega)}{\pi\mathcal{N}^+(E(\widetilde{g}_i)|0,\Sigma) + (1-\pi)\mathcal{U}(0,\Omega)}$$

Structure Damage Score (SDS):

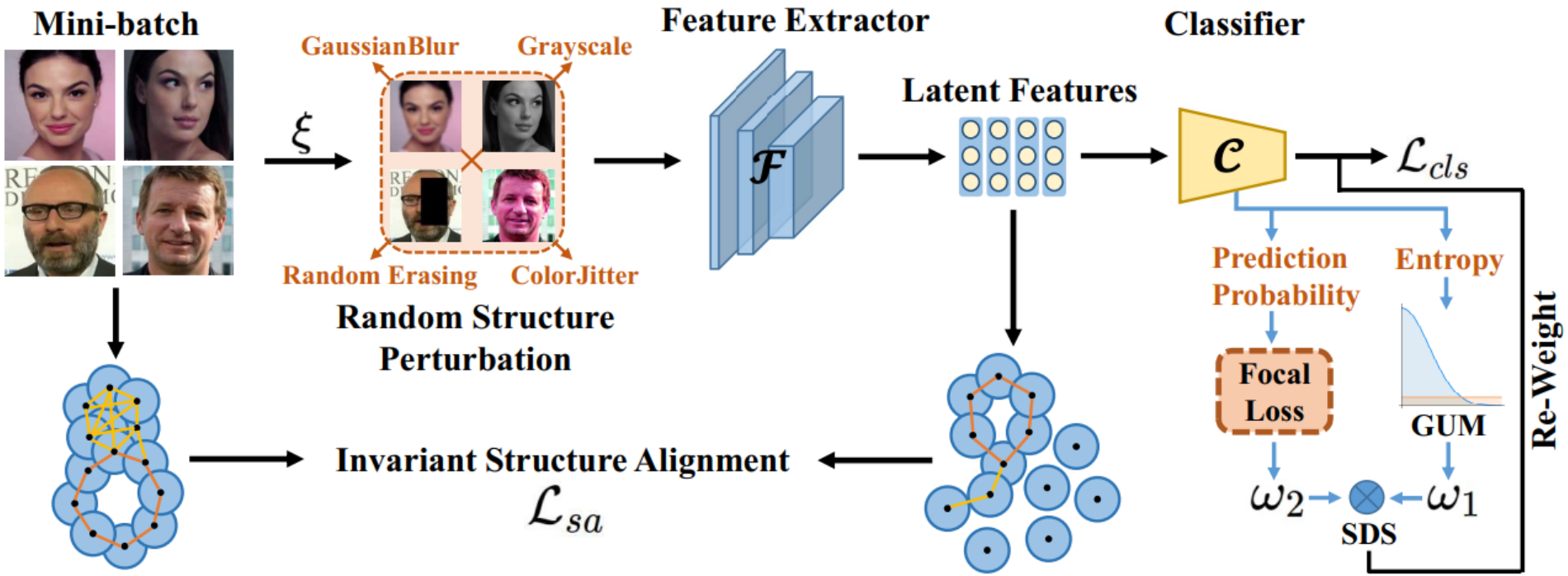
Inspired by the Focal loss, we design a probability-aware scoring mechanism that combines prediction uncertainty and prediction accuracy to adaptively compute SDS for each sample.

$$\omega(\widetilde{x}_i) = \omega_1(\widetilde{x}_i) \times \omega_2(\widetilde{x}_i) = (1 + h_\varphi(\widetilde{x}_i))^\lambda \times (1 - \widetilde{g}_i^{gt})$$

By assigning higher scores to hard samples, the model is encouraged to focus more on learning these challenging samples, boosting the FR system's generalization.

$$\mathcal{L}_{cls} = \omega(\widetilde{x}_i) \times \mathcal{L}_{arc}(\widetilde{x}_i, y_i)$$

Model Optimization



The overall objective of TopoFR:

$$\min_{\mathcal{F},\mathcal{C}} \mathcal{L}_{cls} + \alpha \mathcal{L}_{sa}$$

Experiments

Datasets：

For training, we employ three distinct datasets, namely **MS1MV2** (5.8 Mimages, 85K identities), **Glint360K** (17M images, 360K identities), and **WebFace42M** (42.5M facial images, 2M identities).

For evaluation, we adopt LFW, AgeDB-30, CFP-FP, CPLFW, CALFW, IJB-C, and IJB-B as the benchmarks to test the performance of our models.

Backbones：

ResNet-50, ResNet-100, and ResNet-200.

Experiments

Results on LFW, CFP-FP, AgeDB-30, IJB-B and IJB-C:

(1) TopoFR's performance on easy benchmarks has nearly reached saturation and is significantly higher than that of compared methods.

(2) On IJB-B/C, TopoFR has achieved SOTA performance across different ResNet backbones. Notably, our R50-based TopoFR model even surpasses most R100-based competitors.

Table 2: Verification accuracy (%) on LFW, CFP-FP, AgeDB-30, IJB-C and IJB-B benchmarks.

Training Data	Method	Venue	LFW	CFP-FP	AgeDB-30	IJB-C		IJB-B
						1e-5	1e-4	1e-4
MS1MV2	R50, ArcFace [1]	CVPR19	99.68	97.11	97.53	88.36	92.52	91.66
	R50, MagFace [5]	CVPR21	99.74	97.47	97.70	88.95	93.34	91.47
	R50, AdaFace [3]	CVPR22	99.82	97.86	97.85	-	96.27	94.42
	R50, **TopoFR**†	-	**99.83**	**98.24**	98.23	**94.79**	96.42	95.13
	R50, **TopoFR**	-	**99.83**	**98.24**	**98.25**	94.71	**96.49**	**95.14**
	R100, CosFace [2]	CVPR18	99.78	98.26	98.17	92.68	95.56	94.01
	R100, ArcFace [1]	CVPR19	99.77	98.27	98.15	92.69	95.74	94.09
	R100, MV-Softmax [71]	AAAI20	99.80	98.28	97.95	-	95.20	93.60
	R100, URL [53]	CVPR20	99.78	98.64	-	95.00	96.60	-
	R100, BroadFace [72]	ECCV20	**99.85**	98.63	98.38	94.59	96.38	94.97
	R100, CurricularFace [4]	CVPR20	99.80	98.37	98.32	-	96.10	94.80
	R100, MagFace+ [5]	CVPR21	99.83	98.46	98.17	94.08	95.97	94.51
	R100, SCF-ArcFace [22]	CVPR21	99.82	98.40	98.30	94.04	96.09	94.74
	R100, DAM-CurricularFace [73]	ICCV21	-	-	-	-	96.20	95.12
	R100, ElasticFace-Cos+ [74]	CVPR22	99.80	98.73	98.28	-	96.65	95.43
	R100, AdaFace [3]	CVPR22	99.82	98.49	98.05	-	96.89	95.67
	TransFace-B [9]	ICCV23	99.82	98.39	98.27	94.15	96.55	-
	R100, **TopoFR**†	-	**99.85**	**98.83**	**98.42**	**95.28**	**96.96**	**95.70**
	R100, **TopoFR**	-	**99.85**	98.71	**98.42**	95.23	96.95	**95.70**
	R200, ArcFace [1]	CVPR19	99.79	98.44	98.19	94.67	96.53	95.18
	R200, AdaFace [3]	CVPR22	99.83	98.76	98.28	94.88	96.93	95.71
	TransFace-L [9]	ICCV23	99.83	98.65	98.23	94.55	96.59	-
	R200, **TopoFR**†	-	**99.85**	**99.09**	**98.54**	**95.19**	**97.12**	95.77
	R200, **TopoFR**	-	**99.85**	99.05	98.52	95.15	97.08	**95.82**
Glint360K	R50, ArcFace [1]	CVPR19	99.78	98.77	98.28	95.29	96.81	95.30
	R50, AdaFace [3]	CVPR22	99.82	99.07	98.34	95.58	96.90	95.66
	R50, **TopoFR**	-	**99.85**	**99.28**	**98.47**	**95.99**	**97.27**	**95.96**
	R100, ArcFace [1]	CVPR19	99.81	99.04	98.31	95.38	96.89	95.69
	R100, AdaFace [3]	CVPR22	99.82	99.20	98.58	96.24	97.19	95.87
	TransFace-B [9]	ICCV23	99.85	99.17	98.53	96.18	97.45	-
	R100, **TopoFR**	-	**99.85**	**99.43**	**98.72**	**96.57**	**97.60**	**96.34**
	R200, ArcFace [1]	CVPR19	99.82	99.14	98.49	95.71	97.20	95.89
	R200, AdaFace [3]	CVPR22	99.83	99.24	98.61	95.96	97.33	96.12
	TransFace-L [9]	ICCV23	99.85	99.32	98.62	96.29	97.61	-
	R200, **TopoFR**	-	**99.87**	**99.45**	**98.82**	**96.71**	**97.84**	**96.56**

Code

Code and pre-trained models are available at:

https://github.com/modelscope/facechain/tree/main/face_module/TopoFR