Unbiased Manifold Augmentation for Coarse Class Subdivision

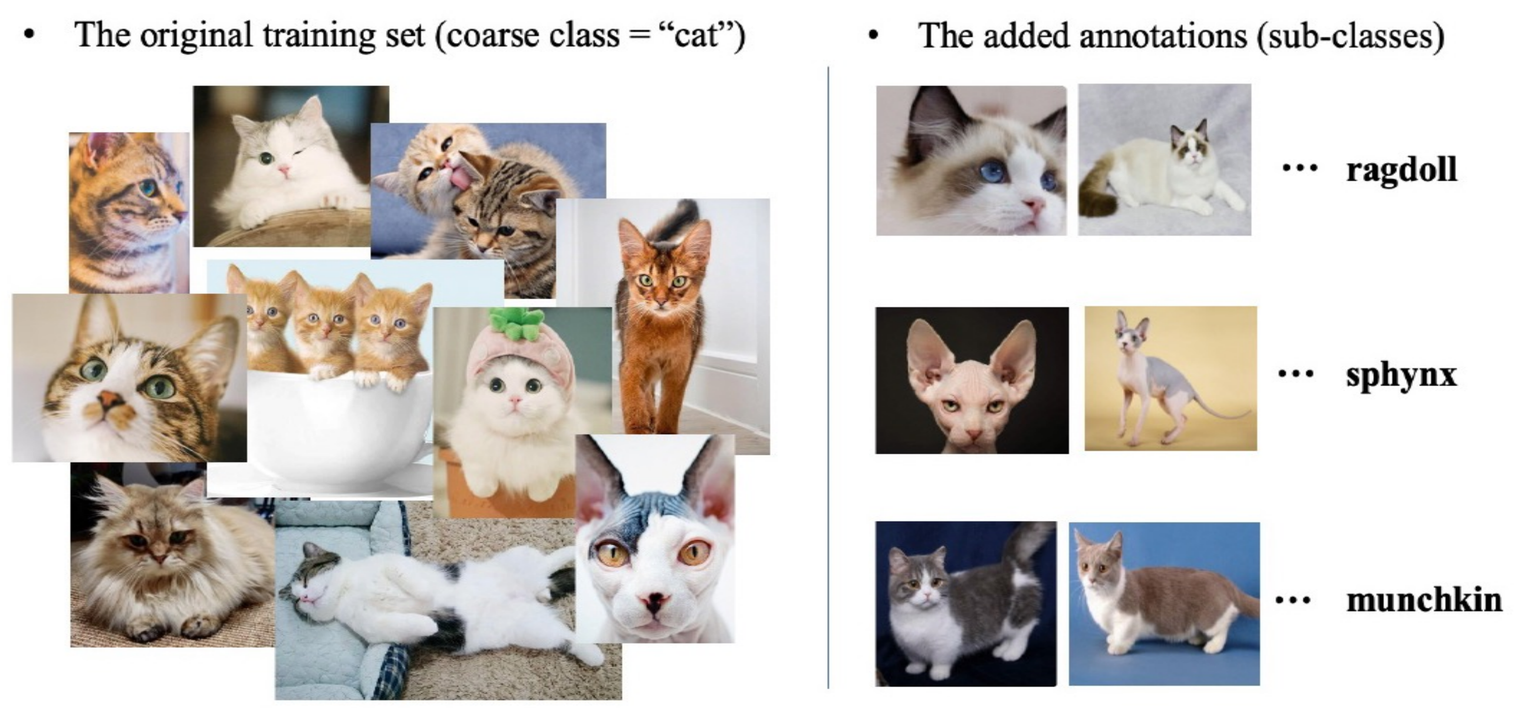
Baoming Yan, Ke Gao*, Bo Gao, Lin Wang, Jiang Yang, and Xiaobo Li

Alibaba Group： {andy.ybm, gaoke.gao, leo.gb, youlin.wl, yangjiang.yj, xiaobo.lixb}@alibaba-inc.com

ECCV TEL AVIV 2022

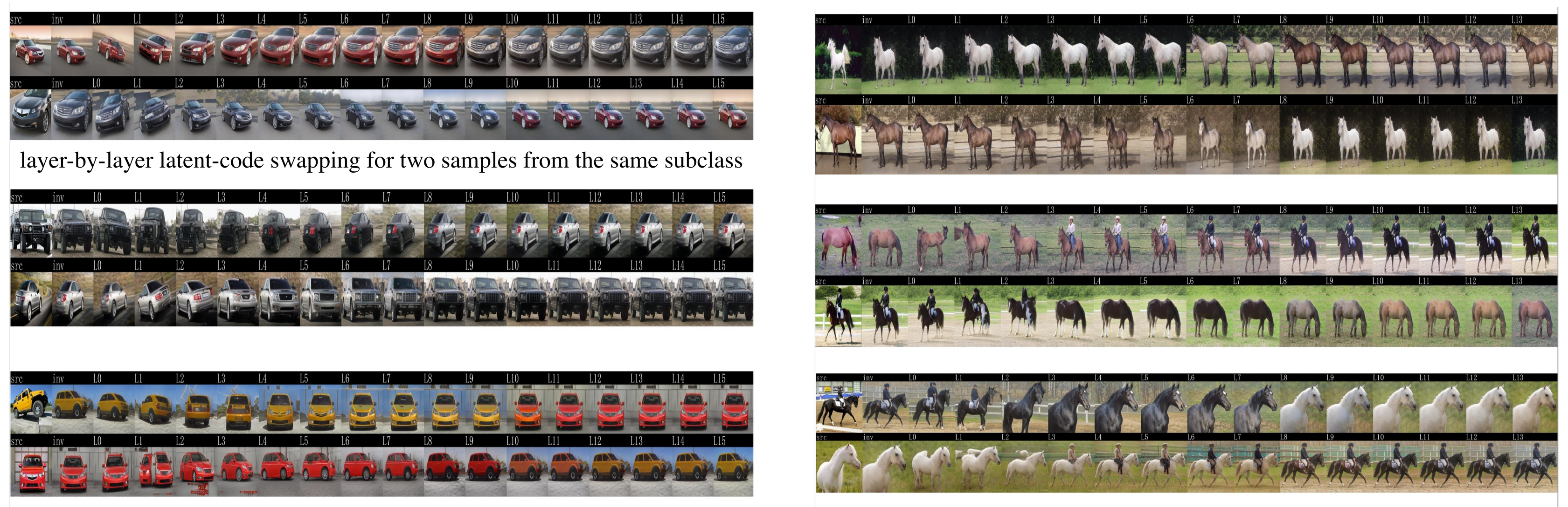
【MOTIVATIONS】

- Different from the conventional classification tasks where the original class and the new target are of similar level of semantic granularity, the **target of CCS (Coarse Class Subdivision) is to further recognize its sub-classes with minimum fine-grained labeling cost**.
- From the perspective of causal representation learning, the sub-classes in CCS task inherit the same determinative factors of the coarse class, and their difference lies only in values. This task requires **identifying the determinative factors of these sub-classes and distinguishing subtle differences between them**. Traditional data augmentation methods show limited effect on this challenging task.
- Although the rapid development of factors-disentangled and controllable generative models illuminates an entirely new avenue for overcoming this problem, the correspondence of existing generators are far from ideal, especially **the correspondence between editable latent-codes and the generating factors are implicit**.



The illustration of Coarse Class Subdivision task

【OBSERVATIONS】



layer-by-layer latent-code swapping for two samples from the same subclass

layer-by-layer latent-code swapping for two samples from different subclasses

layer-by-layer latent-code swapping for non-rigid objects

【CONTRIBUTIONS】

- A novel and systematic data augmentation mechanism called Unbiased Manifold Augmentation (UMA) is proposed for coarse class subdivision problem. Given an existing training set for a coarse class, our UMA can support subclasses recognition with minimum fine-grained labeling cost.
- **The UMA is conducted on latent-code manifolds of a controllable generator pretrained for a coarser category**, instead of the traditional highly-coupled image or feature space. **Using a simple and effective progressive synthesis strategy, an approximate unbiased augmentation at the granularity of generating factors is achieved**, even with limited labeled samples and agnostic bias. By revealing the unbiased mutual information between the target class and all of its impact factors, the classifier can be guided to focus on the right determining factors of the target sub-classes. **In conjunction with it, a phase of progressive robust learning is further integrated, to keep a good balance of the diversity and reliability of these synthetic samples**.
- Extensive experiments have shown that in the case of small data learning (less than 1% fine-grained samples of commonly used), our UMA can achieve 10.37% average improvement compared with existing data augmentation methods. On challenging tasks with severe bias, the accuracy is improved by up to 16.79%.

【SOLUTIONS】

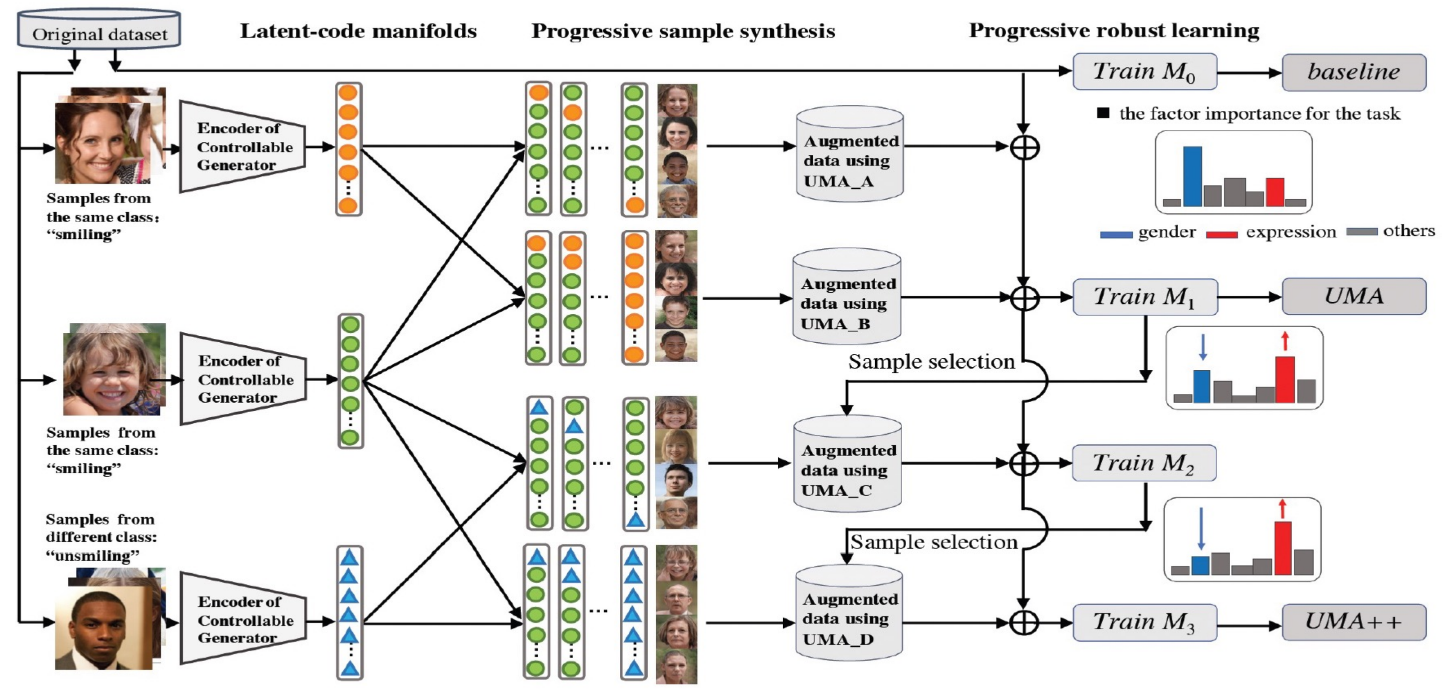


Fig. 2. The architecture of the proposed Unbiased Manifold Augmentation (UMA) for coarse class subdivision task. Take the coarse class "face" for example, here the target sub-classes are "smiling" and "unsmiling". Randomly selected fine-grained samples often have inevitable bias on the confounding factor "gender", instead of the determining factor "expression" for the target task. Our UMA can generate unbiased samples for each sub-class and lead the classifier to focus on determining factors with better generalization.

Fig. 6. Visualization of the proposed UMA and other wildly used data augmentation methods. The reliability and diversity of the synthetic samples using the UMA are better than other augmentation methods conducted in image space or feature space.

【EXPERIMENTAL RESULTS】

Method	Publication	Black Hair	Eye-glasses	Makeup	Smilling	Bald	CA#1
MixUp [28]	ICLR'18	74.28	77.39	67.58	55.76	76.56	64.45
CutMix [27]	ICCV'19	70.68	74.45	62.89	51.76	76.17	65.82
Auto Augment [5]	CVPR'19	71.55	77.76	59.67	54.98	74.22	63.77
Manifold Mixup [24]	ICML'19	<u>75.76</u>	75.74	67.29	58.40	76.95	66.70
Random Erasing [30]	AAAI'20	74.24	75.55	60.94	50.78	73.05	62.79
Random Augment [6]	NIPS'20	71.52	72.06	59.28	54.69	73.83	59.18
MoEx [13]	CVPR'21	74.09	78.68	<u>69.73</u>	53.22	<u>81.64</u>	66.80
StyleMix [9]	CVPR'21	68.71	<u>78.86</u>	63.67	<u>61.23</u>	71.88	51.76
StyleCutMix [9]	CVPR'21	72.34	69.30	64.84	60.94	75.39	66.11
StyleCutMix(auto-γ) [9]	CVPR'21	68.67	71.32	63.09	58.98	74.21	<u>67.09</u>
UMA	-	**79.20** (3.44↑)	**90.99** (12.13↑)	**74.41** (4.68↑)	**67.87** (6.64↑)	**84.77** (3.13↑)	**76.95** (9.86↑)
UMA++	-	**79.89** (4.13↑)	**93.01** (14.15↑)	**77.05** (7.32↑)	**69.53** (8.30↑)	**87.50** (5.86↑)	**78.03** (10.94↑)

Method	Publication	CA#2	Sedan	SUV	Brown Horse	White Horse	Average
MixUp [28]	ICLR'18	<u>74.51</u>	62.89	62.30	79.69	84.38	70.89
CutMix [27]	ICCV'19	71.00	65.04	62.50	79.69	79.69	69.06
Auto Augment [5]	CVPR'19	72.75	<u>68.75</u>	66.99	<u>85.94</u>	78.91	70.48
Manifold Mixup [24]	ICML'19	74.32	63.96	61.91	76.56	80.47	70.73
Random Erasing [30]	AAAI'20	72.07	61.52	59.08	78.91	78.12	67.91
Random Augment [6]	NIPS'20	71.78	66.99	<u>68.55</u>	81.25	81.25	69.13
MoEx [13]	CVPR'21	71.97	60.16	57.62	85.16	<u>86.72</u>	<u>71.44</u>
StyleMix [9]	CVPR'21	73.34	63.96	62.50	78.12	81.25	68.66
StyleCutMix [9]	CVPR'21	72.75	64.36	59.57	75.00	78.91	69.05
StyleCutMix(auto-γ) [9]	CVPR'21	72.27	64.84	62.65	76.56	85.16	69.53
UMA	-	**76.86** (2.35↑)	**72.56** (3.81↑)	**71.29** (2.74↑)	**89.84** (3.90↑)	**88.28** (1.56↑)	**79.37** (7.93↑)
UMA++	-	**81.74** (7.23↑)	**74.61** (5.86↑)	**74.90** (6.35↑)	**91.41** (5.47↑)	**92.19** (5.47↑)	**81.81** (10.37↑)

Table 1. Accuracy comparison on 11 coarse class subdivision tasks with 10 widely used data augmentation methods. All methods use an identical backbone architecture and the same default parameters. The number of training samples is only 32. More details of the task settings please refer to section 4.

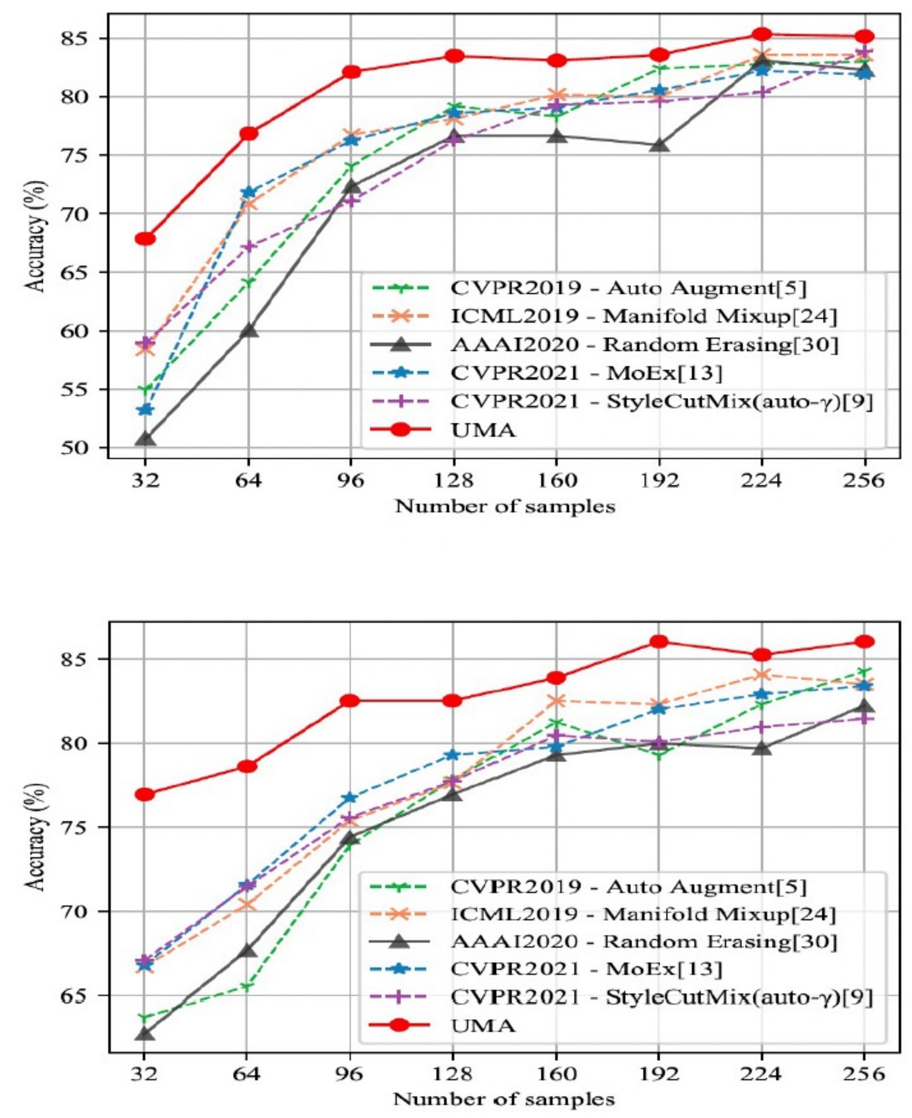
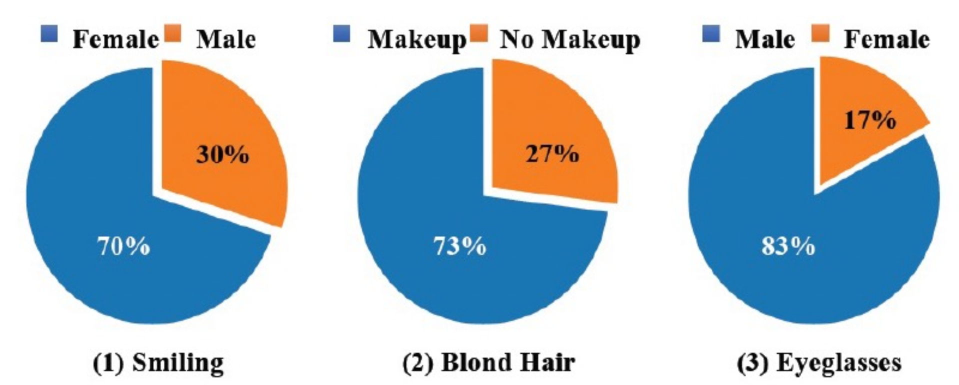
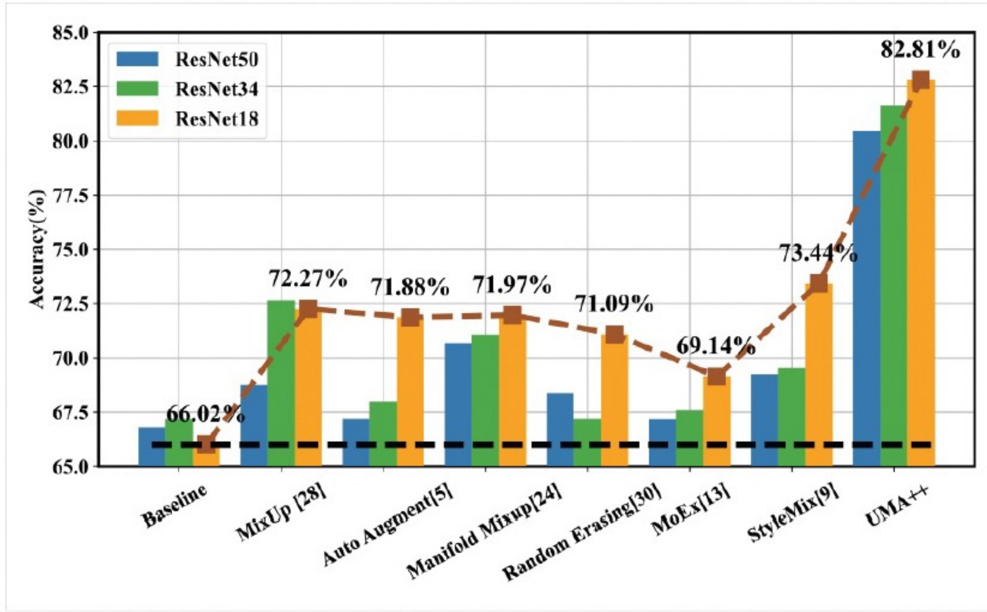


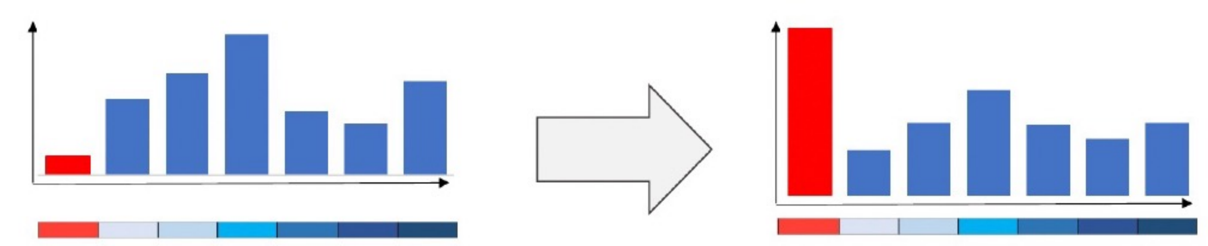
Fig. 4. Accuracy comparison on various sub-class recognition tasks, with different number of training samples. To simulate the challenging small-data-learning scenario of coarse class subdivision task, we only provide less than 1% fine-grained samples of commonly used.



(a) Severe bias observed in well-known facial dataset [14]



(b) Accuracy comparison on the recognition task of "gender" with severe bias on "wearing eyeglasses", using three different backbones.



(c) The change of factor importance for a target sub-class before/after using UMA

Table 2. Performance comparison with several state-of-the-art FGVC methods, on a dataset with severe bias. Baseline stands for resnet-18 backbone with one fully connected layers. UMA++ and various FGVC methods are added to evaluate the performance.

Method	Publication	Baseline	Baseline + UMA++	Baseline + FGVC	Baseline +FGVC +UMA++
SPS [16]	ICCV'21	66.02%	82.81%	72.27%	83.59%(**11.31**↑)
ProtoTree [10]	CVPR'21			66.80%	84.38%(**17.58**↑)
CAL [17]	CVPR'21			73.05%	84.77%(**11.72**↑)

Fig. 5. The observation of severe bias in wildly-used datasets, and the accuracy comparison between our UMA and the state-of-the-art data augmentation methods.