Feature Generating Networks for Zero-Shot Learning

Reporter：陈思玉

2023.02.25

Xian Y, Lorenz T, Schiele B, et al. Feature generating networks for zero-shot learning[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 5542-5551.

Author

Yongqin Xian

Google
在 google.com 的电子邮件经过验证 - 首页
Computer Vision Machine Learning

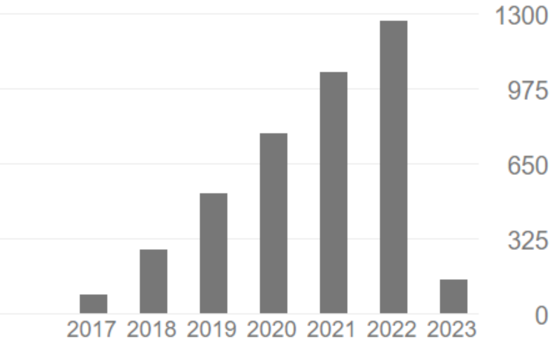
关注

创建我的个人资料

标题	引用次数	年份
Zero-shot learning—a comprehensive evaluation of the good, the bad and the ugly EI检索 SCI基础版 工程技术1区 SCI Q1 SCIIF 24.31 SWUFE A+ Y Xian, CH Lampert, B Schiele, Z Akata IEEE transactions on pattern analysis and machine intelligence 41 (9), 2251-2265	1208	2018
Feature generating networks for zero-shot learning Y Xian, T Lorenz, B Schiele, Z Akata Proceedings of the IEEE conference on computer vision and pattern …	790	2018
Latent embeddings for zero-shot classification Y Xian, Z Akata, G Sharma, Q Nguyen, M Hein, B Schiele Proceedings of the IEEE conference on computer vision and pattern …	712	2016
Zero-shot learning-the good, the bad and the ugly Y Xian, B Schiele, Z Akata Proceedings of the IEEE conference on computer vision and pattern …	654	2017
f-vaegan-d2: A feature generating framework for any-shot learning Y Xian, S Sharma, B Schiele, Z Akata Proceedings of the IEEE/CVF conference on computer vision and pattern …	362	2019
Attribute prototype network for zero-shot learning EI检索 W Xu, Y Xian, J Wang, B Schiele, Z Akata Advances in Neural Information Processing Systems 33, 21969-21980	112	2020

引用次数

	总计	2018 年至今
引用	4184	4073
h 指数	13	13
i10 指数	15	15



开放获取的出版物数量 查看全部

2 篇文章 6 篇文章

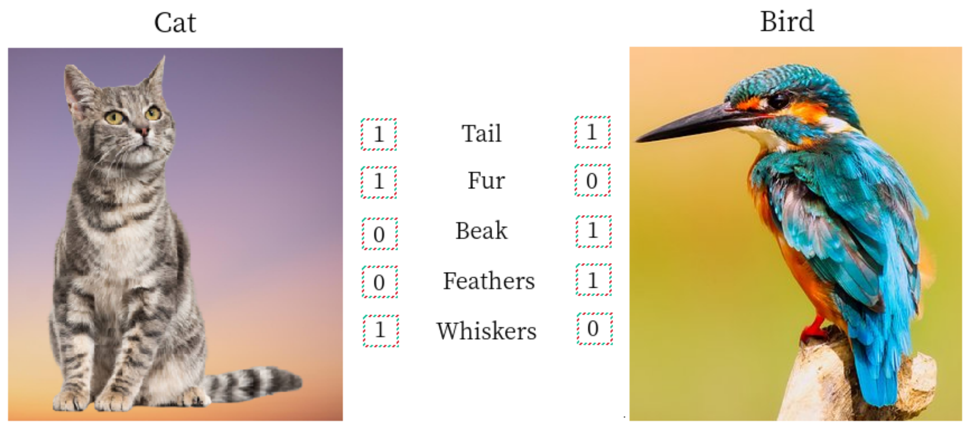
无法查看的文章 可查看的文章

根据资助方的强制性开放获取政策

Zero-Shot Learning

ZSL 目标

ZSL 旨在训练个模型，该模型能够通过语义信息的辅助，利用从 seen classes 中学到的知识来对 unseen classes 进行分类。



ZSL 所用数据

- seen classes：X^s（图像特征），Y^s（类别标签），A^s（语义信息）
- unseen classes：A^u（语义信息）

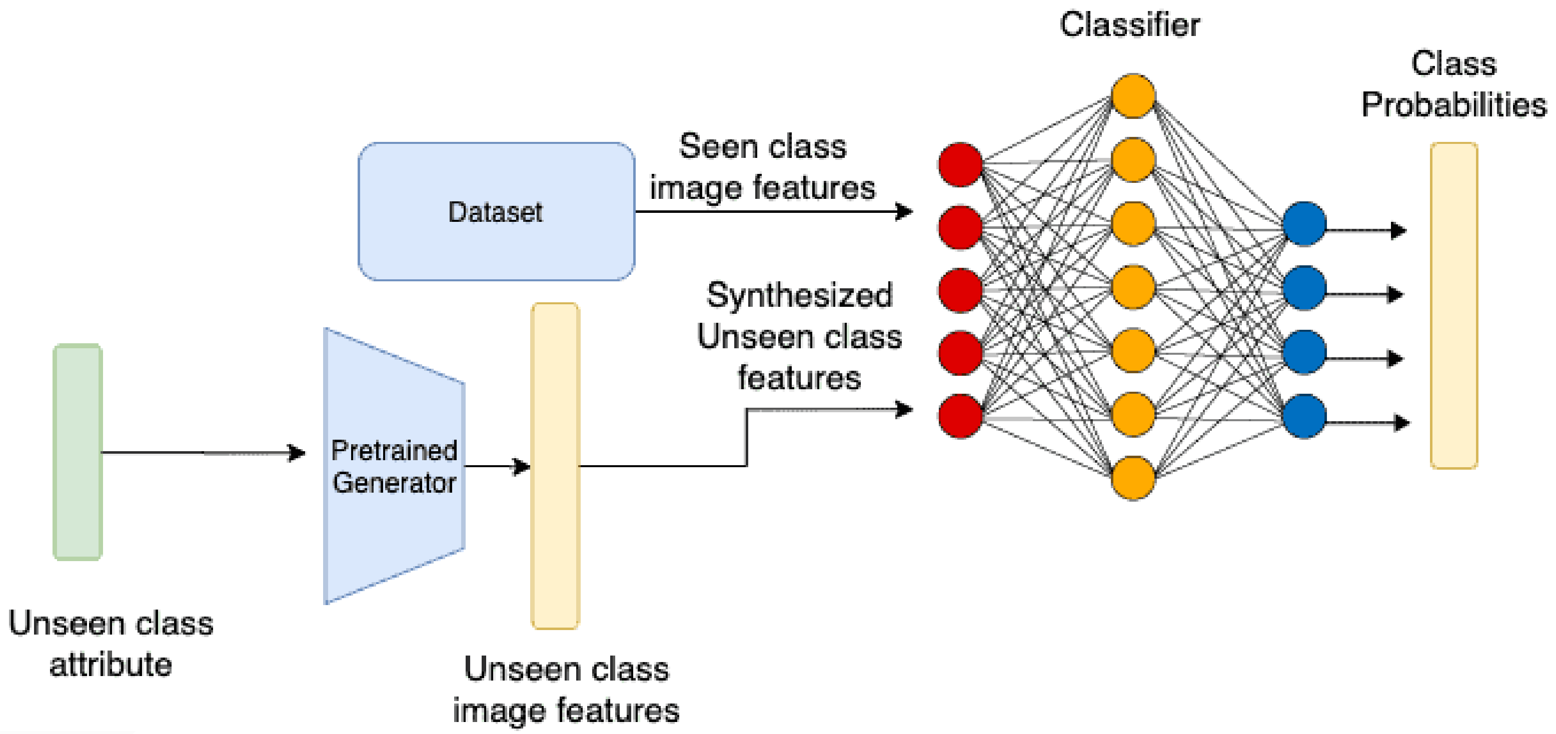
举例说明

1. 训练集有马、老虎、熊猫的图片
2. 语义信息有形状、条纹、颜色等属性
3. 给出斑马的定义：马的形状、老虎的条纹、熊猫的颜色
4. 输入斑马的图像，分类器能输出斑马的类别

Generative-based Methods

主要思想

1. 训练一个生成模型，该模型能够使用语义信息进行条件生成
2. 向训练好的模型输入 unseen classes 的语义信息，从而生成 unseen 的样本
3. 将训练集中的 seen 样本和生成的 unseen 的样本组合成数据集
4. 将数据集输入分类器进行学习，从而使得分类器能对 seen 和 unseen classes 进行分类



论文贡献

- 提出了新的 GAN——f-xGAN，以语义信息为指导生成特定的特征
- 设置分类 loss，生成 discriminative 特征
- 做了大量实验，得出了一些重要的结论

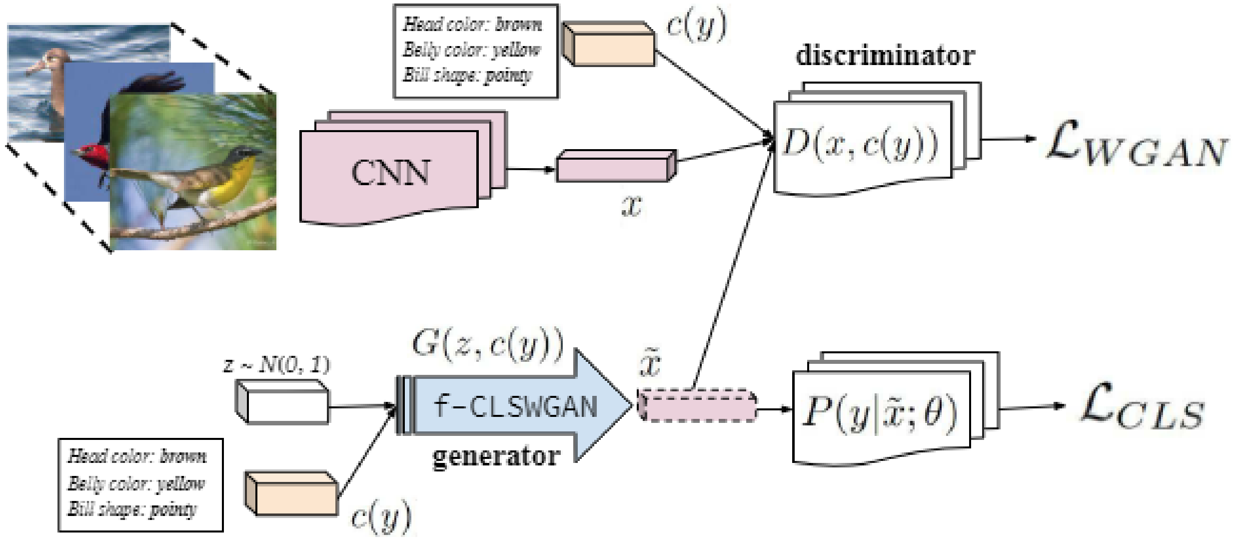


Figure 2: Our f-CLSWGAN: we propose to minimize the classification loss over the generated features and the Wasserstein distance with gradient penalty.

f-GAN

$$\min_G \max_D \mathcal{L}_{GAN} = E[\log D(x, c(y))] + E[\log(1 - D(\tilde{x}, c(y)))]$$

- $D : \mathcal{X} \times \mathcal{C} \to [0, 1]$
  - 最后一层是 sigmoid
- x：seen classes 的图像特征
- $\tilde{x} = G(z, c(y))$：生成的 seen classes 图像特征
  - z：高斯噪声
  - $c(y)$：seen classes 的语义信息

f-WGAN

$$\min_G \max_D \mathcal{L}_{WGAN} = E[D(x, c(y))] - E[D(\tilde{x}, c(y))] - \lambda E[(\|\nabla_{\hat{x}} D(\hat{x}, c(y))\| - 1)^2]$$

- $D : \mathcal{X} \times \mathcal{C} \to \mathbb{R}$
  - 最后一层去除了 sigmoid，输出实数
- $\hat{x} = \alpha x + (1-\alpha)\tilde{x}$
  - $\alpha \sim U(0,1)$
- 前两项代表 Wasserstein distance
- 第三项是 gradient penalty
  - 驱使 D 的梯度沿着真实值和生成值之间的直线具有单位范数
- 不能保证生成 discriminative 的特征

f-CLSWGAN

$$\mathcal{L}_{CLS} = -E_{\tilde{x}\sim p_{\tilde{x}}}[\log P(y|\tilde{x};\theta)]$$

- 对于生成的特征，加入一个分类器 loss，驱使生成的特征易于分类
- θ：提前用 seen classes 特征训练好的分类器
    - 类别概率计算使用 linear softmax

$$\min_G \max_D \mathcal{L}_{WGAN} + \beta\mathcal{L}_{CLS}$$

Classification

Multimodal Embedding

$$f(x) = \arg\max_{x} F(x, c(y); W)$$

- F：基于嵌入的模型
- 相比传统的基于嵌入的方法，这里作者将生成的特征也加入训练

Softmax

$$\text{train} \begin{cases} \min_\theta -\frac{1}{T}\sum_{(x,y)\in\mathcal{T}} \log P(y|x;\theta) \\ \log P(y|x;\theta) = \frac{\exp(\theta_y^T x)}{\sum_i^N \exp(\theta_i^T x)} \end{cases}$$

$$f(x) = \arg\max_{y} P(y|x;\theta)$$

- $\theta \in \mathbb{R}^{d_x \times N}$

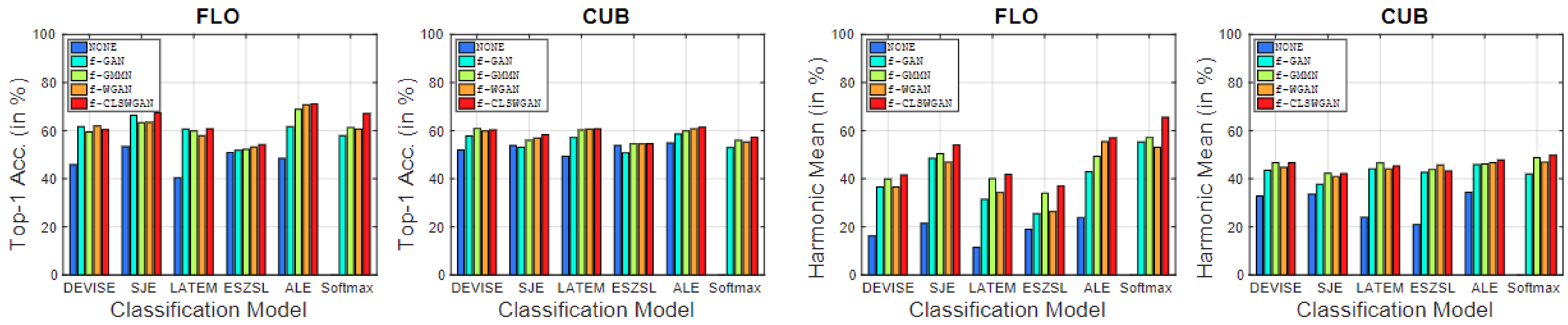
Experiments

		Zero-Shot Learning				Generalized Zero-Shot Learning											
		CUB	FLO	SUN	AWA	CUB			FLO			SUN			AWA		
Classifier	FG	T1	T1	T1	T1	u	s	H	u	s	H	u	s	H	u	s	H
DEVISE [14]	none	52.0	45.9	56.5	54.2	23.8	53.0	32.8	9.9	44.2	16.2	16.9	27.4	20.9	13.4	68.7	22.4
	f-CLSWGAN	60.3	60.4	60.9	66.9	52.2	42.4	46.7	45.0	38.6	41.6	38.4	25.4	30.6	35.0	62.8	45.0
SJE [3]	none	53.9	53.4	53.7	65.6	23.5	59.2	33.6	13.9	47.6	21.5	14.7	30.5	19.8	11.3	74.6	19.6
	f-CLSWGAN	58.4	67.4	56.5	66.9	48.1	37.4	42.1	52.1	56.2	54.1	36.7	25.0	29.7	37.9	70.1	49.2
LATEM [45]	none	49.3	40.4	55.3	55.1	15.2	57.3	24.0	6.6	47.6	11.5	14.7	28.8	19.5	7.3	71.7	13.3
	f-CLSWGAN	60.8	60.8	61.3	**69.9**	53.6	39.2	45.3	47.2	37.7	41.9	42.4	23.1	29.9	33.0	61.5	43.0
ESZSL [40]	none	53.9	51.0	54.5	58.2	12.6	63.8	21.0	11.4	56.8	19.0	11.0	27.9	15.8	6.6	75.6	12.1
	f-CLSWGAN	54.7	54.3	54.0	63.9	36.8	50.9	43.2	25.3	69.2	37.1	27.8	20.4	23.5	31.1	72.8	43.6
ALE [2]	none	54.9	48.5	58.1	59.9	23.7	62.8	34.4	13.3	61.6	21.9	21.8	33.1	26.3	16.8	76.1	27.5
	f-CLSWGAN	**61.5**	**71.2**	**62.1**	68.2	40.2	59.3	47.9	54.3	60.3	57.1	41.3	31.1	35.5	47.6	57.2	52.0
Softmax	none	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	f-CLSWGAN	57.3	67.2	60.8	68.2	43.7	57.7	**49.7**	59.0	73.8	**65.6**	42.6	36.6	**39.4**	57.9	61.4	**59.6**

Table 2: ZSL measuring per-class average Top-1 accuracy (T1) on $\mathcal{Y}^u$ and GZSL measuring **u** = T1 on $\mathcal{Y}^u$, **s** = T1 on $\mathcal{Y}^s$, **H** = harmonic mean (FG=feature generator, none: no access to generated CNN features, hence softmax is not applicable). f-CLSWGAN significantly boosts both the ZSL and GZSL accuracy of all classification models on all four datasets.

- 传统的基于嵌入的方法，加上 FG 后，准确率提高，且平衡了 seen 和 unseen 的分类准确率
- 在 GZSL 下，简单的 Softmax 分类器优于传统的方法

Experiments



(a) Zero-Shot Learning (b) Generalized Zero-Shot Learning

Figure 3: Comparing f-xGAN versions with f-GMMN as well as comparing multimodal embedding methods with softmax.

- 加上 FG 的结果优于 none，即不加 FG
- 在不同的生成模型下，f-CLSWGAN 的效果普遍最好

Experiments

对于生成的 seen classes 样本进行分类器的训练，并在测试集上测试 seen 准确率

- 稳定性：随着 Epoch 的增加，分类器对于 seen 的准确率呈稳定的上升趋势

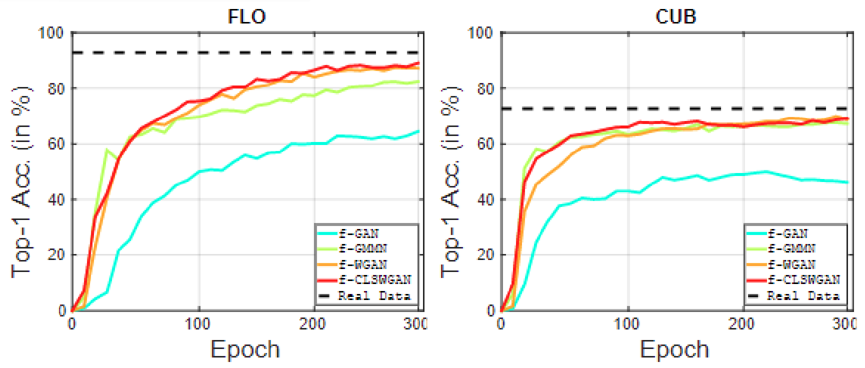


Figure 4: Measuring the seen class accuracy of the classifier trained on generated features of seen classes w.r.t. the training epochs (with softmax).

改变生成的 unseen 样本的数量训练分类器，并在测试集上测试 unseen 准确率

- 泛化性：随着生成的 unseen 样本的增加，分类器对于 unseen 的准确率呈稳定的上升趋势

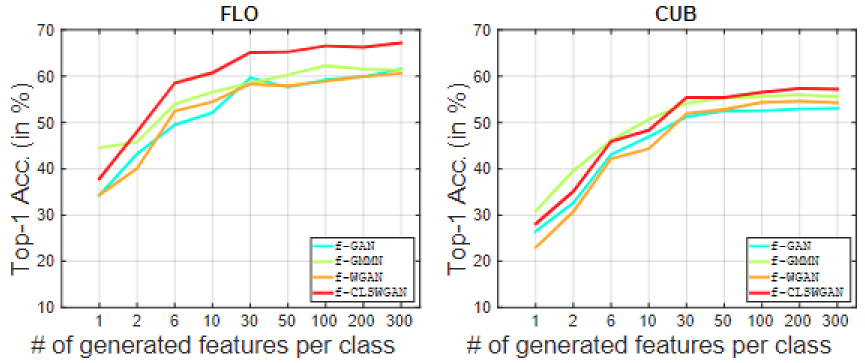


Figure 5: Increasing the number of generated f-xGAN features wrt unseen class accuracy (with softmax) in ZSL.

Experiments

CNN	FG	**u**	**s**	**H**
GoogLeNet	none	20.2	35.7	25.8
	`f-CLSWGAN`	35.3	38.7	36.9
ResNet-101	none	23.7	62.8	34.4
	`f-CLSWGAN`	43.7	57.7	49.7

Table 3: GZSL results with GoogLeNet vs ResNet-101 features on CUB (CNN: Deep Feature Encoder Network, FG: Feature Generator, **u** = T1 on $\mathcal{Y}^u$, **s** = T1 on $\mathcal{Y}^s$, **H** = harmonic mean, "none"= no generated features).

Experiments

C	FG	u	s	H
Attribute (att)	none	23.7	62.8	34.4
	f-CLSWGAN	43.7	57.7	49.7
Sentence (stc)	none	38.8	53.8	45.1
	f-CLSWGAN	50.3	58.3	54.0

Table 4: GZSL results with conditioning f-xGAN with stc and att on CUB (C: Class embedding, FG: Feature Generator, u = T1 on $\mathcal{Y}^u$, s = T1 on $\mathcal{Y}^s$, H = harmonic mean, "none"= no generated features).

- 使用 stc 进行生成的效果优于使用 att 进行生成的效果，说明 stc 能反映出更多的特征

Experiments

Image 代表从 StackGAN 依据 stc 生成的 256 × 256 的图像中提取特征

	CUB			FLO		
Generated Data	u	s	H	u	s	H
none	38.8	53.8	45.1	13.3	61.6	21.9
Image (with [48])	23.8	48.5	31.9	39.4	64.9	49.0
CNN feature (Ours)	50.3	58.3	**54.0**	59.0	73.8	**65.6**

Table 5: Summary Table (u = T1 on $\mathcal{Y}^u$, s = T1 accuracy on $\mathcal{Y}^s$, H = harmonic mean, class embedding = stc). "none": ALE with no generated features.

- Image 在 FLO 上有所提高，但在 CUB 上反而降低了
  - 论文中认为生成鸟类比生成花类更难
  - 论文通过目视观察发现，尽管生成了许多鸟类或花类的外观，但缺乏分类所需的判别细节
- 生成图像特征比生成图像的效果要好，论文认为原因在于
  - 生成的图像特征的数量是无限的
  - 图像特征是从ImageNet上训练的模型学习到的低维特征，因此可用规模小的生成模型进行生成
  - 生成的图像特征具有更高的辨识度，因为生成图像的维度要高的多，导致更难区分