

# 一种融合生成对抗网络的零样本图像分类方法

刘 帅<sup>1</sup>, 黄 刚<sup>1</sup>, 戴晓峰<sup>1</sup>, 颜金花<sup>2</sup>

(1. 南京邮电大学 计算机学院, 江苏 南京 210023;

2. 南京工业大学 计算机科学与技术学院, 江苏 南京 211816)

**摘要:**目前很多零样本图像分类方法是采用学习语义信息空间和图像特征空间之间的映射关系来实现图像分类,但是这些方法会产生枢纽化和域漂移问题。通过利用生成对抗网络实现图像特征生成可以缓解以上两种问题,但该方法容易产生模式崩溃从而导致生成的图像特征不真实。因此提出一种改进的生成对抗网络方法,通过在生成器网络上增加一个重构网络,将生成器生成的图像特征重构回语义信息,以此实现生成器网络生成的图像特征更加符合语义信息的图像特征。实验结果表明,该方法相较于原本的生成对抗网络模型而言,在AWA、CUB、FLO、SUN四个数据集上的分类准确率分别提升了1.0、0.1、1.2和0.9个百分点,证明了通过融合改进的生成对抗网络实现零样本图像分类方法的有效性和可行性。

**关键词:**零样本学习;生成对抗网络;语义信息;图像特征生成;信息重构

中图分类号:TP301

文献标识码:A

文章编号:1673-629X(2022)07-0087-06

doi:10.3969/j.issn.1673-629X.2022.07.015

## A Zero-shot Classification Based on Generative Adversarial Network

LIU Shuai<sup>1</sup>, HUANG Gang<sup>1</sup>, DAI Xiao-feng<sup>1</sup>, YAN Jin-hua<sup>2</sup>

(1. School of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210023, China;

2. School of Computer Science and Technology, Nanjing Technology University, Nanjing 211816, China)

**Abstract:** Since most zero-shot image classification adopts the mapping relationship between the semantic information space and the image feature space to achieve the image classification, which will cause hubness and domain drift. The above two problems can be alleviated by using the generated adversarial network to generate the image feature, but this method is prone to mode collapse, which will result in unrealistic image features. Therefore, an improved generative adversarial network method is proposed. By adding a reconstruction network to the generator network, the image features generated by the generator are reconstructed back to semantic information, so that it is more in line with the semantic information. The experiment shows that compared with the original generative adversarial network model, the proposed method has increased by 1.0, 0.1, 1.2, and 0.9 percent in classification accuracy on the four datasets of AWA, CUB, FLO, and SUN, respectively. It proves the effectiveness and feasibility of the zero-shot image classification method by fusing the improved generative adversarial network.

**Key words:** zero-shot learning; generative adversarial network; semantic information; image feature generation; information reconstruction

## 0 引言

近年来,随着深度学习的发展,深度学习在自然图像识别领域(如图像识别与分类、图像的文本描述以及图像分割等)都取得了突破性的进展,在对人脸、车牌号等特定的对象识别与分类方面表现的尤为突出。但算法的发展过程中也逐渐暴露出泛化能力差、训练所需数据量大等缺点。尤其是近年来的图像分类数据量都非常大,训练数据过大则导致人工标注难度大,而传统方法分类图像需要大量被标注的标签,因此传统

的方法无法满足现有需求。

为了解决上述问题,Larochelle 等于 2008 年提出了零样本学习这一概念<sup>[1]</sup>。零样本可以将可见类拓展到不可见类而不需要额外的标签<sup>[2-8]</sup>。零样本图像分类的过程与人类识别未见过的物体很相似,但是人类可以只通过一句话的描述就能识别出物体。例如,如果一个孩子以前见过马,并且知道“斑马”就像一匹黑白条纹的马,那么他/她识别“斑马”就没有问题。零样本分类依赖于一个有标记的可见类,以及一个拥有

收稿日期:2021-07-26

修回日期:2021-11-26

基金项目:江苏省教育基金资助项目(17JS010);中国电信公司江苏分公司基金资助项目(DGJ02)

作者简介:刘 帅(1996-),男,硕士,研究方向为计算机视觉;黄 刚,教授,研究方向为数据挖掘。

语义信息描述的不可见类。传统的零样本方法主要是建立视觉空间和语义特征空间之间的映射函数,包括视觉空间到语义空间的映射<sup>[2]</sup>,语义空间到视觉空间的映射<sup>[9]</sup>,视觉空间和语义空间映射到一个共享的隐空间<sup>[10]</sup>。但是由于视觉和语义之间的差异,相互映射会导致语义信息丢失,最终对图像分类的效果不理想。

同时,由于深度网络的训练缺少注释样本,训练数据生成成为了研究热点<sup>[11]</sup>。生成对抗网络<sup>[12]</sup>特别有吸引力,因为它们允许生成真实而清晰的图像,例如,生成以对象类别为条件的图像<sup>[13-14]</sup>。尽管生成对抗网络已经可以解决多数问题,但是其难以训练和梯度消失等问题仍是一大阻碍。强大的 WGAN<sup>[15]</sup>解决了这些问题。

因此,可以通过融合生成对抗网络来解决零样本存在的问题,利用高斯噪声和不可见类的语义信息生成图像特征,然后利用可见类和生成的图像特征训练分类器。虽然上述方法可以解决零样本存在的问题,但是也引入了新的问题:如何训练出能够生成符合语义信息的图像特征。因此,该文改进了生成器,将生成的图像特征重构回语义信息,从而约束生成器生成高质量图像特征。

主要贡献如下:改进生成对抗网络的生成器网络,在原本生成器网络基础上引入重构网络,将生成的图像特征重构回语义信息,以此来限制生成器,使得生成的图像特征能够更加贴合语义信息,从而确保图像和语义直接的结构相关性。

## 1 相关工作

本节介绍零样本学习和生成对抗网络,为后面改进融合生成对抗网络的零样本图像分类模型做铺垫。

### 1.1 零样本学习

在早期的零样本方法中,主要可以分为直接语义预测和模型嵌入两类。其中第一类,直接预测方法,最具代表性的为 2013 年 Lampert 提出的两种拓扑结构零样本学习模型,即直接属性预测(DAP)模型以及间接属性预测(IAP)模型。其中 DAP,就是对输入的每一个属性训练一个分类器,然后将训练得出的模型用于属性的预测,测试时,对测试样本的属性进行预测,再从属性向量空间里面找到和测试样本最接近的类别。IAP 则是利用可见类建立一种关联关系来将测试样本和不可见类相关联,该方法是将测试样本分类到可见类别中,随后根据可见类和不可见类的语义信息建立的语义关系对测试样本进行预测。文献[16]提出利用随机森林算法来训练模型以提高对语义信息的预测准确度,通过对每个类别的属性预测损失来提升预测准确率。基于嵌入的方法则是将可见类和不可见

类关联在一个高维向量空间中,即语义嵌入空间,该空间可以是语义属性空间或语义词向量空间。由于视觉和语义的差异,三种语义和视觉空间的映射会导致语义丢失问题。Chen 等于 2017 年提出了 SP-AEN 模型<sup>[17]</sup>,通过引入一个独立的视觉到语义空间嵌入器来防止语义丢失,该嵌入器将语义空间分解为两个子空间,以实现两个可能冲突的目标:分类和重构。通过两个子空间的对抗学习,SP-AEN 可以将语义从重构子空间转移到判别子空间,实现对不可见类的零样本识别。

虽然传统的零样本方法能够在一定程度上实现零样本图像分类,但取得的效果并不理想。于是 Xian 等<sup>[18]</sup>提出了利用生成对抗网络体系结构来直接生成图像特征。通过利用可见类训练生成器和鉴别器,以及不可见类的语义信息生成类似不可见类的图像特征来训练分类器,最终使得分类器的效果好于之前的传统方法。

### 1.2 生成对抗网络

生成对抗网络,最初是由 Goodfellow 作为一种学习生成模型的方法提出,该模型从一个特定的领域(图像等)捕获任意的数据分布。生成对抗网络由生成器和鉴别器两个部分组成,生成器输入随机噪声试图生成“真实图像”来欺骗鉴别器,鉴别器则区分真假图像,两个网络通过相互对抗的方式进行。在随后的发展中,通过将类别标签,句子描述输入到生成器和鉴别器中,GAN 也被扩展为条件 GAN。GAN 的理论在文献[19]中进行了研究,表明 Jensen-Shannon 发散会导致不稳定性问题。为了解决 GAN 的不稳定训练问题,文献[20]提出了 Wasserstein-GAN(WGAN),优化了 Wasserstein 距离的有效逼近。虽然 WGAN 获得了比原始 GAN 更好的理论性能,但它仍然受到梯度消失和爆炸问题的困扰。因此,文献[15]提出了一个改进的 WGAN 版本,通过梯度惩罚对鉴别器进行 1-Lipschitz 约束来缓解梯度消失和爆炸问题。

## 2 基于生成对抗网络模型

### 2.1 问题描述

首先设  $S = \{(x, y, c(y)) \mid x \in X, y \in Y, c(y) \in C\}$ ,其中  $S$  代表可见类的训练数据,  $x \in R^d$  为经过 CNN 处理的图像特征,  $y$  表示类别标签  $Y^s = \{y_1, \dots, y_k\}$ ,由  $K$  个离散的可见类组成,  $c(y) \in R^d$  是类别嵌入,例如类  $y$  的属性,它模拟类之间的语义关系。另外,还有一个不相交的类标签集  $Y^u = \{u_1, \dots, u_L\}$ ,包含  $L$  个不可见类,其类别嵌入集合  $U = \{u, c(u) \mid u \in Y^u, c(u) \in C\}$  可用,但缺少图像特征,以此来训练分类器,最终使得分类器的效果好于之前的传统方法。

### 2.2 f-CLSWGAN 模型

f-CLSWGAN 模型就是利用生成对抗网络生成图像特征,提供更多的图像特征训练分类器以提高分类器的精准度。如图 1 所示,先将真实的图像特征输入鉴别器,让鉴别器给真实的图像特征高分,随后将随机高斯噪声  $z \in Z \in R^d$  和语义信息  $c(y) \in C$  作为生成器的输入,让生成器  $G:Z \times C \rightarrow X$  输出一个类  $y$  的 CNN 图像特征  $\tilde{x} \in X$ ,然后将生成器生成的图像特征输入鉴别器,让鉴别器给生成的特征低分。随着  $D$  和  $G$  的相互对抗,生成器  $G$  会生成接近真实的图像特征。最后将不可见类的语义信息  $c(y)$  和高斯噪声  $z$  输入生成器以生成类似不可见类的图像特征并和可见类一起来训练分类器。f-CLSWGAN 模型主要通过优化生成对抗网络损失函数和分类器损失函数,同样让  $G$  和  $D$  相互竞争对抗学习:

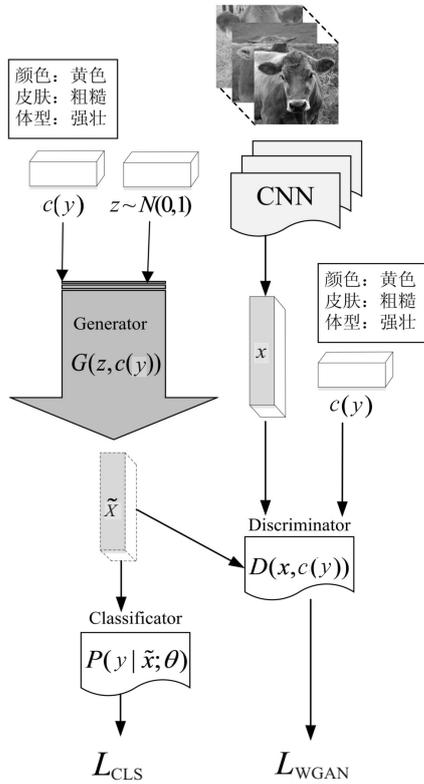


图 1 CLSWGAN 网络结构

生成对抗网络损失函数  $L_{WGAN}$  :

$$L_{WGAN} = E[D(x, c(y))] - E[D(\tilde{x}, c(y))] - \lambda E[(\|\nabla_{\tilde{x}} D(\hat{x}, c(y))\|_2 - 1)^2]$$

$$\tilde{x} = G(z, c(y)), \hat{x} = \alpha x + (1 - \alpha)\tilde{x} \quad (1)$$

式中,前两项近似于 Wasserstein 距离,第三项是梯度惩罚,  $\lambda$  为惩罚系数。

分类器的损失函数  $L_{CLS}$  :

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

$$\tilde{x} = G(z, c(y)) \quad (2)$$

式中,  $y$  是  $\tilde{x}$  的类标签,  $P$  是用其真实类别标签  $y$  预测

$\tilde{x}$  的概率。

整体的损失函数为:

$$\min_G \max_D L_{WGAN} + \beta L_{CLS} \quad (3)$$

### 2.3 改进生成器的模型 f-WGRAN

由于生成对抗网络容易产生模式崩溃的问题,仅通过生成器和鉴别器之间相互对抗会使得最终生成的图像特征趋向单一化。计算机视觉的另一种方法—变分自动编码器(VAE)也可以生成图像<sup>[21]</sup>。VAE 由编码器 Encoder 和解码器 Decoder 构成。如图 2 所示,VAE 本质上是提取图像的隐性特征,构建从特征生成目标图像的模式。编码器从输入图像中提取均值向量和标准差向量,然后对编码得到的结果向量加上高斯噪声来约束,生成潜在变量,潜在变量服从高斯分布的隐含特征。解码器将潜在变量映射到重新生成的概率分布中。编码之后的分布要与原始的分布越接近越好。

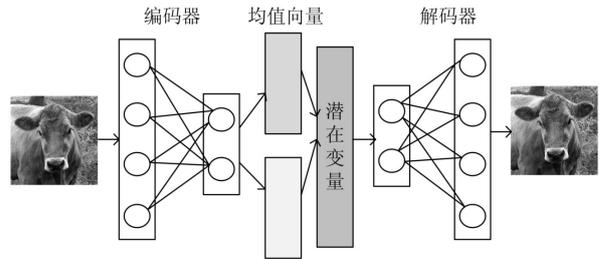


图 2 VAE 网络结构

VAE 的损失函数如下:

$$L(\varphi, \theta, x) = E_{q_{\varphi}(z|x)}[\log p_{\theta}(x|z)] - D_{KL}(q_{\varphi}(z|x) \| p_{\theta}(z)) \quad (4)$$

式中,  $E$  为编码器网络,  $D$  为解码器网络,  $q_{\varphi}(z|x)$  表示从图像编码成向量的分布,  $p_{\theta}(x|z)$  表示从向量重构为图像的分布。第一项为解码器  $D$  的误差,驱使分布  $p_{\theta}(x|z)$  接近于输入分布  $p_{\theta}(x)$ , 第二项旨在减少 KL 散度,令  $q_{\varphi}(z|x)$  更接近于先验分布  $p_{\theta}(z)$ 。

VAE 本质上就是在对 Encoder 的均值向量结果加上了高斯噪声,使得 Decoder 能够对噪声有鲁棒性。而额外的 KL 损失目的是让均值为 0, 方差为 1, 事实上就是相当于对 Encoder 的一个正则项,希望 Encoder 出的结果有均值。另外,Encoder 的标准差向量用于动态调节噪声的强度。当 Decoder 还没有训练好即重构误差远大于 KL 损失时,就会适当降低噪声使得拟合起来容易一些;反之,如果 Decoder 训练的效果不错,重构误差小于 KL 损失,这时候噪声就会增加,使得拟合更加困难,这时候 Decoder 就要提高 Decoder 生成能力。

基于 VAE 网络的原理,推理出可以在 f-CLSWGAN 的基础上引入重构网络来比较生成器生成的图像特征和原始的语义信息之间的差异来约束生

成器,以改进 f-CLSWGAN 模型,提升分类准确度。

因此,为了使生成的图像特征更加多样化,在生成对抗网络上增加一个重构网络,如图 3 所示。通过将生成器生成的视觉特征重构回语义信息,以此来限制生成器生成更符合属性的图像特征,将得到的新模型命名为 f-WGRAN。在 f-CLSWGAN 的基础上,将  $G$  生成的图像特征  $\tilde{x}$  通过重构网络  $R$  重构回语义信息  $\tilde{c}(y)$ ,然后与将用于生成图像特征的语义信息  $c(y)$  进行对比,计算出重构的损失  $L_R$  来改进  $G$ ,让  $G$  生成的图像特征与语义信息的相关性更加紧密。

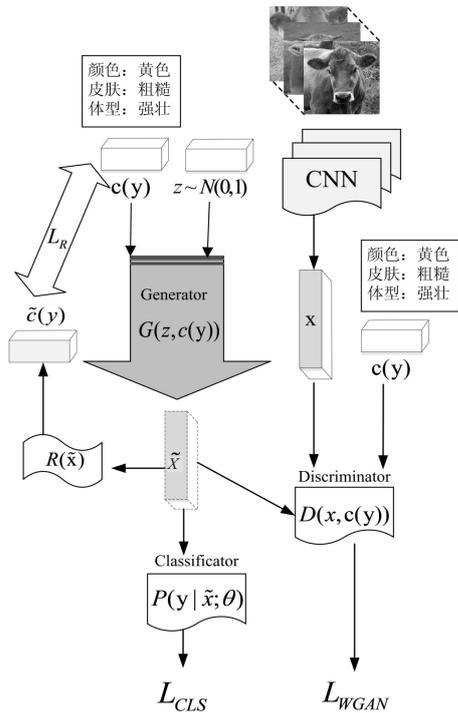


图 3 f-WGRN 网络结构

重构网络的损失函数如下:

$$L_R = E(\|c(y) - R(\tilde{x})\|_2^2)$$

$$\tilde{x} = G(z, c(y)) \quad (5)$$

生成对抗网络损失函数  $L_{WGRAN}$  :

$$L_{WGRAN} = E[D(x, c(y))] - E[D(\tilde{x}, c(y))] - \lambda E[(\|\nabla_{\tilde{x}} D(\tilde{x}, c(y))\|_2 - 1)^2] + E(\|c(y) - R(\tilde{x})\|_2^2)$$

$$\tilde{x} = G(z, c(y)), \hat{x} = \alpha x + (1 - \alpha)\tilde{x} \quad (6)$$

整体的损失函数为:

$$\min_G \max_D L_{WGRAN} + \beta L_{CLS} \quad (7)$$

### 3 实验

#### 3.1 实验数据集

实验采用了 4 个不同的经典数据集,分别是 AWA<sup>[4]</sup>、CUB<sup>[22]</sup>、FLO<sup>[23]</sup>、SUN<sup>[24]</sup>,其中 CUB、FLO、SUN 都是细粒度数据集,AWA 是粗粒度数据集。CUB 包含来自 200 种不同的类型鸟类的 11 788 张图

片,这些图片带有 312 种属性。FLO 有 102 种不同类型的鲜花图片共 8 189 张,但是没有属性注释,所以通过视觉描述对 FLO 进行注释。SUN 是一个场景数据集,其包含 717 个场景的 14 340 张,且场景标注有 102 种属性。AWA 是一个动物数据集,包含 30 475 张图片,共用 50 个类别,包含 85 种属性。

数据集具体参数如表 1 所示,同时将数据集分割为两部分,可见类作为训练集,不可见类作为测试集,可见类与不可见类之间没有交集。

表 1 AWA、CUB、FLO、SUN 数据集

数据集	类别	属性	语句	可见类	不可见类
AWA	50	85	否	27+13	10
CUB	200	312	是	100+50	50
FLO	102	--	是	62+20	20
SUN	717	102	否	580+65	72

#### 3.2 数据集处理

参考 Xian 等的对于数据集的操作,采用 Res101<sup>[25]</sup> 对整幅图像提取 2 048 维的图像特征,并在 ImageNet 上进行了预训练。没有对图像进行预处理,例如裁剪或者其他的数据增强技术。此外,把每个类别的属性 AWA(85 维),CUB(312 维)和 SUN(102 维)作为语义信息。对于 SUN 和 FLO,从细粒度的视觉描述(每幅图像 10 句)中提取了 1 024 维的基于字符的 CNN-RNN 特征。

#### 3.3 实验结果与分析

该模型在 ubuntu16.04、Python2.6、pytorch 环境下搭建,在一张 Tesla P100 显卡上训练并测试。实验在四个不同数据集上分别进行了 100 轮训练,图 4 为四个不同数据集的分类准确度。

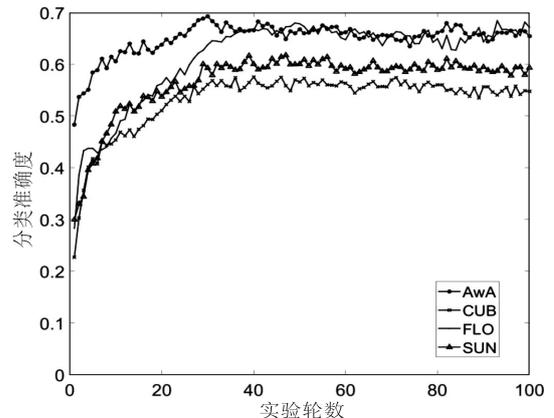


图 4 实验结果

可以看出,该模型在训练 AWA 上只经过了 30 轮训练便达到了最大值,随之便逐渐下降;在 CUB 数据集上数据震荡较大,在 45 轮左右达到最大值;在 FLO 数据集上经过 40 轮训练后开始在最大值附件震荡,直到 95 轮左右达到了最大值;在 SUN 数据集上,经过 30

轮训练后,便逐渐趋于稳定。分类准确度的训练轮次能够很快达到稳定,主要原因在于 f-WGRAN 模型采用了重构网络,可以约束生成器使得生成的图像特征与语义信息的相关性更加紧密。

### 3.4 与不同方法的比较

与不同方法对比结果如表 2 所示。观察实验结果可知,f-WGRAN 关于四种不同数据集的分类准确度

表 2 不同模型方法在 4 个数据集上的分类准确度比较

模型方法	分类准确度/%			
	AWA	CUB	FLO	SUN
DEVISE	54.2	52.0	45.9	56.5
SJE	65.5	53.9	53.4	53.7
EZSL	58.2	53.9	51.0	54.5
ALE	59.9	54.9	48.5	58.1
f-CLSWGAN	68.2	57.3	67.2	60.8
f-WGRAN	69.2	57.4	68.4	61.7

该模型在 AWA 数据集取得了 69.2% 的分类准确度,比 f-CLSWGAN 模型提升了 1.0 个百分点,在 FLO 和 SUN 模型分别取得了 68.4% 和 61.7% 的分类准确度,相比 f-CLSWGAN 分别提升了 1.2 和 0.9 个百分点,在 CUB 数据集上几乎没有提升。实验证明,f-WGRAN 能够取得更高的分类准确度,证明该模型方法是可行的。

## 4 结束语

零样本图像分类经过十多年的发展仍是深度学习领域的具有挑战性的研究方向<sup>[26-27]</sup>。该文通过分析以往的零样本分类方法,分析存在的问题,结合生成对抗网络提出了一种利用生成对抗网络生成图像特征来进行判别的模型 f-WGRAN,以此来解决零样本图像分类问题。f-WGRAN 在生成对抗网络上增加一个重构网络,通过将生成器生成的视觉特征重构回语义信息,以此来限制生成器生成更符合属性的图像特征,加强了生成器的特征生成能力,并且在很少的轮次就可以达到最大分类准确度。经过实验论证,该模型分别在 AWA、CUB、FLO、SUN 四个数据集提升了 1.0、1.2 和 0.9 个百分点。该文主要针对生成对抗网络来进行改进,并且此改进可以应用于不同的生成对抗网络上。在后续的工作中将考虑如何提升鉴别器的鉴别能力。

### 参考文献:

[1] LAROCHELLE H, ERHAN D, BENGIO Y. Zero-data learning of new tasks[C]//Proceedings of the association for the advancement of artificial intelligence. Palo Alto: AAAI Press, 2008: 646-651.

均优于之前的方法。同时,可以看出引入生成对抗网络的模型分类准确度明显高于传统的零样本图像分类方法。原始的 f-CLSWGAN 网络模型只是利用生成器和鉴别器直接相互对抗,使得生成的图像特征越来越单一。而提出的 f-WGRAN 模型对生成器进行改进,引入重构网络来限制生成器生成更符合属性的图像特征,从而使得分类准确度有所提升。

[2] AKATA Z, PERRONNIN F, HARCHAOUI Z, et al. Label-embedding for image classification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(7): 1425-1438.

[3] AKATA Z, REED S, WALTER D, et al. Evaluation of output embeddings for fine-grained image classification[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Boston: IEEE, 2015: 2927-2936.

[4] LAMPERT C H, NICKISCH H, HARMELING S. Attribute-based classification for zero-shot visual object categorization [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 36(3): 453-465.

[5] FROME A, CORRADO G S, SHLENS J, et al. DeViSE: a deep visual semantic embedding model[C]//Proceedings of the neural information processing systems. Cambridge: MIT Press, 2013: 2121-2129.

[6] BUCHER M, HERBIN S, JURIE F. Improving semantic embedding consistency by metric learning for zero-shot classification[C]//European conference on computer vision. Amsterdam: Springer, 2016: 730-746.

[7] FU Y, SIGAL L. Semi-supervised vocabulary-informed learning[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas: IEEE, 2016: 5337-5346.

[8] FU Z, XIANG T, KODIROV E, et al. Zero-shot object recognition by semantic manifold distance[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Boston: IEEE, 2015: 2635-2644.

[9] CHANGPINYO S, CHAO Weilun, SHA Fei. Predicting visual exemplars of unseen classes for zero-shot learning[C]//2017 IEEE international conference on computer vision (ICCV). Venice: IEEE, 2017: 3496-3505.

- [10] ZHANG Ziming, SALIGRAMA V. Zero-shot learning via semantic similarity embedding[C]//2015 IEEE international conference on computer vision (ICCV). Santiago; IEEE, 2015;4166-4174.
- [11] CHEN Q, KOLTUN V. Photographic image synthesis with cascaded refinement networks[C]//Proceedings of the IEEE international conference on computer vision. Venice; IEEE, 2017;1520-1529.
- [12] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//Proceedings of the neural information processing systems. Cambridge; MIT Press, 2014;2672-2680.
- [13] REED S, AKATA Z, YAN X, et al. Generative adversarial text to image synthesis[C]//Proceedings of the 33rd international conference on machine learning. New York; JML-Rorg, 2016;1060-1069.
- [14] ZHANG H, XU T, LI H, et al. Stackgan: text to photo-realistic image synthesis with stacked generative adversarial networks[C]//Proceedings of the IEEE international conference on computer vision (ICCV). Venice; IEEE, 2017;5908-5916.
- [15] GULRAJANI I, AHMED F, ARJOVSKY M, et al. Improved training of wasserstein gans[J]. arXiv:1704.00028, 2017.
- [16] JAYARAMAN D, GRAUMAN K. Zero shot recognition with unreliable attributes[J]. arXiv:1409.4327, 2014.
- [17] CHEN L, ZHANG H, XIAO J, et al. Zero-shot visual recognition using semantics-preserving adversarial embedding networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City; IEEE, 2018;1043-1052.
- [18] 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. Salt Lake City; IEEE, 2018;5542-5551.
- [19] ARJOVSKY M, BOTTOU L. Towards principled methods for training generative adversarial networks[J]. arXiv:1701.04862, 2017.
- [20] ARJOVSKY M, CHINTALA S, BOTTOU L. Wasserstein generative adversarial networks[C]//Proceedings of the international conference on machine learning. Madison; Omnipress, 2017;214-223.
- [21] KINGMA D P, WELING M. Auto-encoding variational bayes[J]. arXiv:1312.6114, 2013.
- [22] WAH C, BRANSON S, WELINDER P, et al. The caltech-ucsd birds-200-2011 dataset[R]. Pasadena, California Institute of Technology, 2011.
- [23] NILSBACK M E, ZISSERMAN A. Automated flower classification over a large number of classes[C]//2008 sixth Indian conference on computer vision, graphics & image processing. Bhubaneswar; IEEE, 2008;722-729.
- [24] XIAO J, HAYS J, EHINGER K A, et al. Sun database: large-scale scene recognition from abbey to zoo[C]//2010 IEEE computer society conference on computer vision and pattern recognition. San Francisco; IEEE, 2010;3485-3492.
- [25] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas; IEEE, 2016;770-778.
- [26] 张鲁宁, 左信, 刘建伟. 零样本学习研究进展[J]. 自动化学报, 2020, 46(1): 1-23.
- [27] 冀中, 汪浩然, 于云龙, 等. 零样本图像分类综述: 十年进展[J]. 中国科学: 信息科学, 2019, 49(10): 1299-1320.
- 
- (上接第 86 页)
- AIS data and accident reports[J]. Reliability Engineering & System Safety, 2018, 176: 174-186.
- [15] LI H, LIU J, WU K, et al. Spatio-temporal vessel trajectory clustering based on data mapping and density[J]. IEEE Access, 2018, 6: 58939-58954.
- [16] AURADKAR P, PRASHANTH T, ARALIHALLI S, et al. Performance tuning analysis of spatial operations on spatial Hadoop cluster with SSD[J]. Procedia Computer Science, 2020, 167: 2253-2266.
- [17] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of the 2016 IEEE conference on computer vision and pattern recognition. Las Vegas, NV, USA; IEEE, 2016; 779-788.
- [18] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//Proceedings of the 2017 IEEE conference on computer vision and pattern recognition. Honolulu, HI, USA; IEEE, 2017; 6517-6525.
- [19] REDMON J, FARHADI A. Yolov3: an incremental improvement[EB/OL]. (2018-04-08) [2019-12-15]. <https://arxiv.org/pdf/1804.02767.pdf>.
- [20] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for object detection and semantic segmentation[C]//IEEE conference on computer vision and pattern recognition (CVPR). Columbus; IEEE, 2014; 580-587.
- [21] GIRSHICK R. Fast R-CNN[C]//Proceedings of IEEE international conference on computer vision. Los Alamitos; IEEE, 2015; 1440-1448.
- [22] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards realtime object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [23] REDMON J, FARHADI A. YOLO v3: an incremental improvement[EB/OL] [2018-11-02]. <http://arxiv.org/pdf/1804.02767v1.pdf>.
- [24] 杨剑, 刘方涛, 张涛, 等. 基于改进型残差网络烟雾图像识别[J]. 科学技术与工程, 2019, 19(32): 236-243.