

中图法分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(2025)05-1389-15

论文引用格式: Zhang G M, Yan W S and Huang J Y. 2025. Generative zero-shot image recognition combined with double contrastive embedding learning. Journal of Image and Graphics, 30(5):1389-1403(张桂梅, 闫文尚, 黄军阳. 2025. 结合双重对比嵌入学习的生成式零样本图像识别. 中国图象图形学报, 30(5):1389-1403)[DOI:10.11834/jig.240526]

结合双重对比嵌入学习的生成式零样本图像识别

张桂梅*, 闫文尚, 黄军阳

南昌航空大学图像处理与模式识别江西省重点实验室, 南昌 330063

摘要: 目的 零样本学习(zero-shot learning, ZSL)是解决样本数据缺失情况下目标识别的有效方法。传统的零样本识别是通过带对标签的可见类数据训练,实现对无标签的未见类新数据的识别。根据任务设置的不同,分为传统零样本学习(conventional zero-shot learning, CZSL)和广义零样本学习(generalized zero-shot learning, GZSL)。生成式零样本识别方法由于可以生成未见类的视觉特征,从而将零样本学习问题转换为常规监督学习问题。但是生成式零样本识别存在特征判别性信息不足、伪视觉特征与语义信息不一致以及域偏移等问题。针对上述问题,提出结合双重对比嵌入学习的生成式零样本图像识别方法。方法 首先,针对生成的特征判别性不足问题,基于VAE-GAN(variational autoencoder-generative adversarial network)生成框架,集成对比嵌入模块,多个网络协同训练,提高零样本图像识别精度;其次,以条件VAE-GAN为生成网络,提出双重对比学习策略。一方面,在现有可见类对比学习的基础上,引入未见类伪样本实例—原型域内对比学习,使生成的伪视觉特征与语义信息对齐,缓解可见类和未见类的语义混淆;另一方面,提出跨域中心—原型对比学习,缓解模型过于偏向于可见类,一定程度上减轻域偏移。结果 在AWA1(animals with attributes1)、AWA2、CUB(Caltech-UCSD birds-200-2011)和SUN(SUNAttribute)数据集上进行零样本和广义零样本识别实验,并与最新相关方法进行比较。在零样本识别任务中,提出的方法在AWA1和CUB数据集取得最优值,相比性能次优的模型, T1值分别提高2.2%和2.7%;在AWA2和SUN数据集均取得次优值。在广义零样本识别中,在AWA1、AWA2和CUB数据集H值均取得最优,相比次优值分别提升0.6%、0.8%和2.8%;在SUN数据集取得次优值。消融实验验证了提出算法的有效性。结论 实验结果表明,提出的方法可提高零样本和广义零样本图像识别的精度,并具有较好的泛化性能。

关键词: 零样本学习(ZSL);广义零样本学习(GZSL);生成对抗网络(GAN);嵌入空间;对比学习

Generative zero-shot image recognition combined with double contrastive embedding learning

Zhang Guimei*, Yan Wenshang, Huang Junyang

Jiangxi Province Key Laboratory of Image Processing and Pattern Recognition, Nanchang Hangkong University, Nanchang 330063, China

Abstract: Objective With the rapid advancement of deep learning technologies, image recognition accuracy has substantially improved. However, traditional deep learning models are heavily dependent on large-scale, well-labeled datasets. This reliance on manually annotated data not only incurs high costs but also poses challenges when models need to adapt to new, unseen data or operate in data-scarce situations. Zero-shot learning (ZSL) has emerged as a promising solution to

收稿日期: 2024-09-06;修回日期: 2024-10-20;预印本日期: 2024-10-27

* 通信作者: 张桂梅 guimei.zh@163.com

基金项目: 国家自然科学基金项目(62361043)

Supported by: National Natural Science Foundation of China(62361043)

these challenges. This approach enables models to recognize unseen, unlabeled data by training on labeled data from seen classes. Depending on the task setting, ZSL is typically categorized into conventional zero-shot learning (CZSL) and generalized zero-shot learning (GZSL). CZSL occurs when the test samples only contain unseen class samples, while GZSL is applicable when the test samples include seen and unseen class ones. The GZSL is more aligned with real-world situations, thereby improving the accuracy of generalized zero-shot image recognition and offering notable practical importance. Generative methods are the most effective approach for addressing GZSL because they can generate visual features for unseen classes, transforming the ZDL problem into a conventional supervised learning task. However, generative zero-shot recognition encounters challenges, such as insufficient discriminative information in generated features, inconsistencies between pseudo-visual features and semantic information, and domain shift, ultimately reducing recognition accuracy. Generative zero-shot image recognition combined with double contrastive embedding learning is proposed to address these issues.

Method A generative framework based on the variational autoencoder-generative adversarial network (VAE-GAN) is constructed to address the main challenge of insufficient discriminability in generated features, and a contrastive embedding module is integrated. Through a collaborative training strategy involving multiple network components, the quality of pseudo-features is notably enhanced, leading to substantial improvements in the accuracy of zero-shot image recognition. Additionally, leveraging conditional VAE-GAN as the core generative network, an innovative dual contrastive learning strategy, which effectively integrates intra-domain and cross-domain information, is proposed, maximizing information utilization. Specifically, intra-domain contrastive learning among pseudo-sample instances of unseen classes and their prototypes is introduced, ensuring that generated pseudo-visual features align closely with semantic information, which mitigates confusion between visible and unseen classes. Cross-domain center-prototype contrastive learning, which strengthens the alignment between visual centers and semantic prototypes, is then implemented. This approach effectively reduces inter-class variance, facilitating more efficient cross-domain knowledge transfer. Consequently, the model becomes less reliant on visible classes, partially alleviating domain shift, which results in exceptional performance even in unknown domains.

Result The experimental framework was rigorously evaluated on zero-shot and generalized zero-shot recognition tasks across four distinct datasets, and the results were compared to the latest field advancements. In zero-shot recognition, the proposed method achieved state-of-the-art performance on the AWA1 and CUB datasets, showing a notable improvement of 2.2% and 2.7% in T1 values, respectively, over the second-best models. On the AWA2 and SUN datasets, the proposed approach demonstrated competitive performance, further displaying its robustness across diverse data environments. For generalized zero-shot recognition, the algorithm outperformed competitors by achieving the highest H values on the AWA1, AWA2, and CUB datasets, revealing improvements of 0.6%, 0.8%, and 2.8%, respectively, over the second-best methods. Notably, this approach is the only one to achieve an accuracy over 70% across all three datasets, emphasizing its exceptional generalization capability. The method performed competitively on the SUN dataset, further reinforcing its overall effectiveness. The observed performance gains can be attributed to the generated pseudo-features, which exhibit a high degree of similarity to the genuine features of the visible classes. This approach helps mitigate semantic confusion and effectively addresses domain shift. Ablation studies were conducted on the AWA1 and CUB datasets to further validate the effectiveness of the proposed approach, providing empirical evidence of its contributions. Additionally, the impact of the number of generated samples on model performance was evaluated by conducting extensive experiments with varying sample sizes in zero-shot and generalized zero-shot scenarios across all four datasets, offering valuable insights for optimizing the generation process to enhance recognition accuracy. Finally, t-SNE dimensionality reduction and visualization experiments were performed on randomly selected unseen classes from the coarse-grained AWA2 and fine-grained CUB datasets to assess the effectiveness of using a low-dimensional embedding space in the algorithm. These experiments visually represent the embedded features and demonstrate the advantages of the proposed approach in capturing discriminative and meaningful representations in a low-dimensional space. **Conclusion** Experimental results indicate that the proposed method enhances the accuracy of zero-shot and generalized zero-shot image recognition, effectively addressing the challenges associated with current contrastive learning-based generative zero-shot image recognition, while also demonstrating good generalization performance. Future research will consider a trade-off between recognition accuracy and efficiency by designing lightweight networks to reduce model complexity and improve recognition efficiency.

Key words: zero-shot learning (ZSL); generalized zero-shot learning (GZSL); generative adversarial network (GAN); embedding space; contrastive learning

0 引言

随着深度学习技术的快速发展,图像识别的精度得到显著提升。但是,传统的深度学习模型严重依赖于大规模、强标注的数据,这种对人工标注数据的强烈依赖不仅导致高昂成本,还使得模型难以适应新数据或数据稀缺的情境,如变异的病毒样本、濒临灭绝的物种等。针对这一挑战,Larochelle等人(2008)提出零样本学习(zero-shot learning, ZSL),旨在通过对带标签的可见类数据训练,实现对无标签的未见类新数据的识别。根据任务设置的不同,可将其分为传统零样本学习(conventional zero-shot learning, CZSL)和广义零样本学习(generalized zero-shot learning, GZSL)(Chao等,2016),在CZSL中,测试集仅有未见类样本。而GZSL中测试集同时包括可见类和未见类样本,这种任务设置更加贴近现实世界的场景,即测试样本可以来源于任意的目标类别。

零样本图像识别(Yang等,2022)是零样本学习任务之一。在早期研究中,通常利用可见类与未见类共享的语义知识,如人工标注的属性(Lampert等,2009)、词向量(Li等,2019)以及文本描述(张桂梅等,2020)等,建立从可见类到未见类的知识迁移。其中,属性作为连接底层视觉特征和高层抽象标签的中间桥梁,因其语义性和判别性强的特点而备受关注。然而,传统基于属性的零样本学习方法在处理属性互斥性、增量学习等问题时存在局限性,导致模型对未见类的识别精度不佳。针对上述问题,研究者(Wang等,2021;Chen等,2023a;Xu等,2022)探索了基于映射的零样本学习方法。这类方法试图通过构建“视觉特征—语义特征”双层架构的映射模型,寻求不同模态之间的对应关系。然而,在识别未见类数据时往往存在域偏移,导致模型在广义零样本任务的识别性能较低。此外,由于语义空间与视觉空间的流形构成不同,直接建立映射关系可能导致语义间隔问题,无法准确反映特征的真实语义信息。

随着生成网络的成功应用,研究者(Chen等,

2021a;Cetin等,2022;Cavazza等,2023)利用生成模型实现零样本图像识别任务,从而将零样本学习问题转换为常规监督学习问题。通过将未见类伪特征加入到模型中进行训练,以改善类不平衡,在一定程度上缓解域偏移。早期的生成式ZSL/GZSL通常采用单一的变分自编码器(variational auto-encoder, VAE)(Schönfeld等,2019)或生成对抗网络(generative adversarial network, GAN)(Goodfellow等,2014)模型。然而这些模型生成的伪视觉特征效果不理想,导致模型在零样本和广义零样本图像识别中精度较低。Xian等人(2019a)提出基于VAE-GAN的零样本识别方法,将变分自编码器(VAE)和生成对抗网络(GAN)集成到统一的框架内。实验结果表明,VAE-GAN具有更好的特征表示能力,生成的伪特征更接近真实特征。Zhang等人(2024)提出双判别器VAE-GAN框架,分别学习可见类和未见类样本,并计算真实样本与伪样本之间的马氏距离以优化生成模型。但是该模型存在较大的域偏移,导致未见类识别效果不佳。Wang等人(2023)基于VAE-GAN模型提出一种微调方法,仅通过提供类名生成未见类特征,操作简单易于实现,但是该方法生成的特征判别性不足,对细粒度识别易出现类别间边界混淆,因此,在细粒度数据集上的识别精度不理想。

近年来,对比学习在自监督学习领域取得惊人的效果,甚至在某些数据集上超越了监督学习模型的性能,如Google团队发布的SimCLR(simple framework for contrastive learning of visual representations)模型(Chen等,2020)仅利用无标签数据进行自监督学习,在ImageNet上的分类精度超越了监督学习AlexNet(Krizhevsky等,2017)模型的分类精度。鉴于对比学习在特征表示方面的优良特性,一些学者将对比学习引入到零样本学习任务中。如Han等人(2022)设计了一个混合生成框架,该框架将特征生成模型WGAN(Gulrajani等,2017)和对比嵌入模型集成到统一框架上,通过为每个锚样本创建多对正/负样本对,在嵌入空间中进行正/负样本对的对比学习,较好地提升了零样本图像识别的精度。Gao等人(2022)提出将实例集群的对比学习方法用于广义

零样本识别,以缓解域偏移。Du 等人(2024)提出基于属性原型的对比学习方法,以凸显不同属性之间的差异性和判别性,在粗粒度数据集上较好地提高了零样本图像识别性能。Paul 等人(2023)提出基于边缘的原型(Chen 等,2023b)对比学习网络,以缓解细粒度聚类时的重叠现象。

上述研究的对比学习核心理念如图 1(a)所示,均采用可见类实例和原型进行对比学习训练模型,尽管取得了一定进展,但现有的基于对比学习的生成式零样本图像识别仍面临以下困难和挑战:1)仅考虑在嵌入空间中建立可见类的视觉特征和语义特征的对比学习,忽略了未见类语义信息的指导,使得模型学习到的特征偏向可见类,导致广义零样本图像识别时产生域偏移。2)伪视觉特征与语义信息不一致,特别是未见类伪视觉特征,脱离类的聚簇结构,难以获得紧致的视觉特征,导致在细粒度数据集

上识别精度下降。

针对生成式零样本识别方法存在特征判别性信息不足、伪视觉特征与语义信息不一致以及域偏移等导致的识别精度低下问题,本文以条件 VAE-GAN 为基本网络框架,提出一种结合双重对比嵌入学习的生成式零样本图像识别方法。本文在零样本学习中引入嵌入模型和对比学习,以提高零样本图像识别的精度。提出双重对比学习策略,同时考虑域内和跨域的双重信息。双重对比学习策略分别是:未见类域内的实例—原型对比学习如图 1(b)所示;跨域中心—原型对比学习如图 1(c)和图 1(d)所示。提出未见类伪样本进行实例—原型对比学习策略,目的是得到准确且与语义一致的视觉特征;提出跨域中心—原型对比学习策略,以促进跨域知识更好地迁移,在一定程度上缓解域偏移,从而提高零样本识别和广义零样本识别的精度。

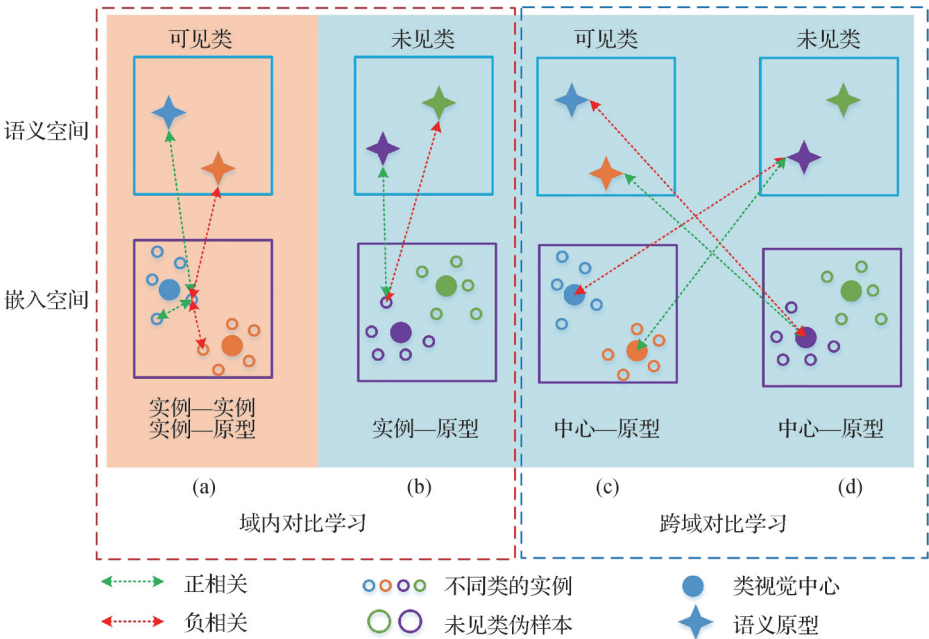


图 1 双重对比学习(图中的实例为视觉特征,原型为语义特征,中心为实例的类中心)

Fig. 1 The dual contrastive learning (the instances in the figure are visual features, the prototypes are semantic features, and the centers are the class centers of instances)

本文的工作总结如下:1)针对生成式零样本识别模型的特征判别性不足的问题,基于 VAE-GAN 基本框架,集成嵌入模块和对比学习,多个网络协同训练,从而增强伪特征的生成质量,提高模型的识别精度;2)基于 VAE-GAN 生成框架,在可见类域内对比学习的基础上,创新性地提出在未见类域内进行实例—原型对比学习,通过兼顾可见类和未见类的语义信息,生成准确且与语义一致的伪视觉特征;3)提

出跨域中心—原型对比学习,通过视觉特征中心和语义原型的强对齐来修正类间方差,促进跨域知识更好地迁移,缓解模型过于偏向于可见类的问题,在一定程度上减轻域偏移;4)在 AWA1 (animals with attributes1)、AWA2、CUB (Caltech-UCSD birds-200-2011)和 SUN (SUNattribute)这 4 个标准数据集上对零样本图像识别和广义零样本图像识别任务进行验证,实验结果表明提出的方法相比于现有较先进的

零样本识别方法,识别精度均有一定提高。

1 本文方法

1.1 问题描述

在零样本图像识别任务中,已知可见类数据集 D_s 和未见类数据集 D_u ,且 D_s 和 D_u 是互斥的,即 $D_s \cap D_u = \emptyset$ 。定义可见类的数据: $D_s = \{(x_s, a_s, y_s) | x_s \in X_s, a_s \in A_s, y_s \in Y_s\}$, 未见类数据: $D_u = \{(x_u, a_u, y_u) | x_u \in X_u, a_u \in A_u, y_u \in Y_u\}$ 。零样本图像识别的思想是利用可见类数据集中的样本以及与样本对应的类标签对模型进行训练,在测试过程中利用在训练中得到的信息及测试类的辅助信息对模型进行补充,以预测未见类属性的类标签 y_u 。文中出现的符号意义如表1所示。

表1 本文的符号表示

Table 1 Symbol representations in this paper

名称	意义
D_s	可见类数据
D_u	未见类数据
x_s	可见类的真实视觉特征
x_u	未见类的真实视觉特征
y_s	可见类的标签
y_u	未见类的标签
a_s	可见类属性
a_u	未见类属性
$\tilde{x}_s$	可见类的伪视觉特征
$\tilde{x}_u$	未见类的伪视觉特征
z_p	高斯噪声

1.2 本文的模型

本文提出的结合双重对比嵌入学习的生成式零样本图像识别模型(generative zero-shot image recognition combined with double contrastive embedding learning, ZSIR-DCEL)由VAE、GAN和对比嵌入3个模块构成,网络结构如图2所示。图2中,对比嵌入分支中的浅红色背景为前人研究,浅蓝色背景为本文提出的双重对比学习损失。VAE由1个编码器E和1个解码器DEC构成;GAN由生成器G和判别器D构成;对比嵌入模块由嵌入层 E_{emb} 、非线性投影层H

和比较网络R构成。解码器DEC/生成器G和判别器D均为一个具有4 096维度的隐藏单元的多层感知器(multilayer perceptron, MLP),隐藏单元输出经过Leaky ReLU(leaky rectified linear unit)激活。对比嵌入层 E_{emb} 和比较网络R均为MLP,且都具有一个2 048维度的隐藏单元,并均使用Leaky ReLU函数激活。

1.3 损失函数

1.3.1 视觉特征生成损失

F-VAEGAN(Chen等,2021a)将条件变分自编码器VAE和WGAN集成到一个统一的生成模型中,使得模型训练更稳定,提高了特征生成的质量。因此,本文也选用该生成框架,VAE和WGAN均使用可见类样本进行模型的监督训练,再以未见类的属性作为模型的输入生成伪视觉特征。条件VAE分支的损失函数为

$$L_{VAE} = L_{KL} + L_{rec} = KL(q(z|x, a_s) || p(z|a_s)) - \mathbb{E}_{q(z|x, a_s)}[\log p(x|z, a_s)] = KL(E(x, a_s) || p(z|a_s)) - \mathbb{E}_{E(x, a_s)}[\log G(z, a_s)] \quad (1)$$

式中, L_{VAE} 由KL(Kullback-Leibler)损失(L_{KL})和视觉特征重建损失(L_{rec})组成。 KL 表示KL散度,条件分布 $q(z|x, a_s)$ 由 $E(x, a_s)$ 建模得到, $p(z|a_s)$ 表示假设服从标准正态分布 $N(0, 1)$ 的先验分布, z 表示编码器E生成的潜在特征。

在VAE-GAN框架中,VAE的解码器和GAN的生成器共享网络参数。GAN分支训练过程为:首先,将高斯噪声 z_p 和属性 a_s 拼接并输入生成器G,输出伪视觉特征 $\tilde{x}_s$ 。将属性 a_s 与真实视觉特征 x_s 以及伪视觉特征 $\tilde{x}_s$ 分别拼接,输入到判别器D,输出视觉特征的真伪程度。判别器通过迭代训练增强区分真实特征和伪特征的能力,而生成器G则通过迭代训练生成更逼真的伪特征以欺骗判别器D,两者对抗博弈迭代训练,最终达到纳什平衡。为了使训练更稳定,本文选择WGAN(Gulrajani等,2017),其损失函数为

$$L_{WGAN} = \mathbb{E}[\log(D(x, a_s))] - \mathbb{E}[\log(D(\tilde{x}, a_s))] - \lambda \mathbb{E}[(\|\nabla D(\hat{x})\|_2 - 1)^2] \quad (2)$$

$$\hat{x} = \alpha x + (1 - \alpha)\tilde{x}, \alpha \sim N(0, 1) \quad (3)$$

$$L_{VAEGAN} = L_{VAE} + L_{WGAN} \quad (4)$$

式(2)中, $\lambda \mathbb{E}[(\|\nabla D(\hat{x})\|_2 - 1)^2]$ 是判别器的梯度惩罚, λ 是一个正值惩罚权重, $\hat{x}$ 是惩罚鉴别器网络的

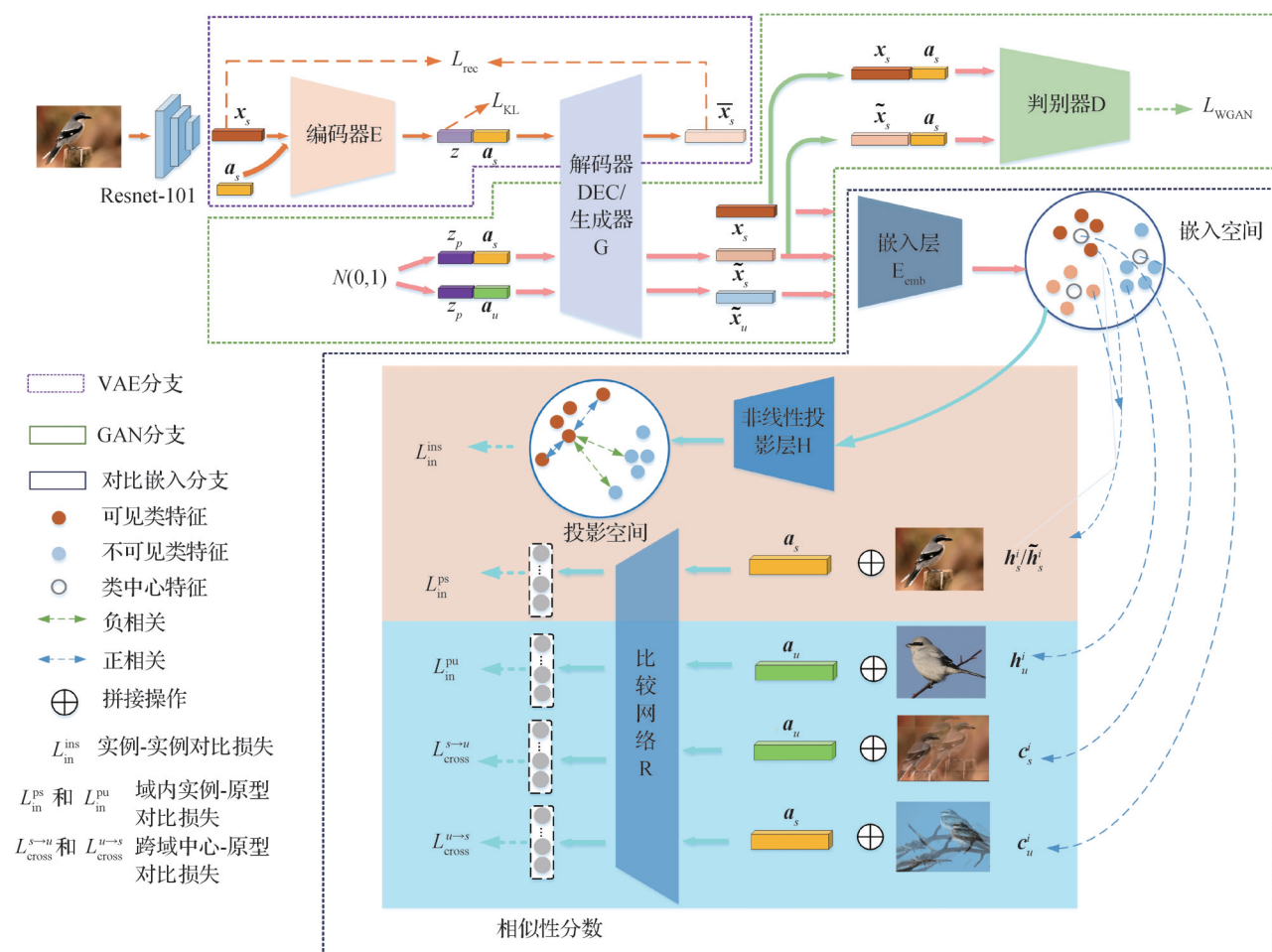


图2 网络结构

Fig. 2 The network architecture

梯度,作用是防止因梯度过大导致生成器网络无法有效学习。 α 是插值权重,控制真实样本 x 和伪样本 $\tilde{x}$ 在插值过程中的权重比例,其服从均值为0、方差为1的标准正态分布。

1.3.2 域内实例—实例对比损失

在细粒度识别(魏秀参等,2022)任务中,由于原始视觉特征的判别性信息不足,且生成的伪视觉特征的质量无法得到保证,导致图像识别精度受到显著影响,而对比学习可以更好地学习类别之间的差异性。本文参照Han等人(2022)提出的实例对比学习方法,构建了域内实例—实例对比损失函数,具体为

$$L_{in}^{ins}(e_i, e^+) = \mathbb{E} \left[-\log \frac{\exp \frac{e_i^T e^+}{\tau_e}}{\exp \frac{e_i^T e^+}{\tau_e} + \sum_{k=1}^{N_{neg}} \exp \frac{e_i^T e_k^-}{\tau_e}} \right] \quad (5)$$

式中, e_i 为随机选取的样本点(称为锚样本), e^+ 表示与 e_i 具有相同标签的任意一个正样本, e^- 表示与 e_i 具

有不同标签的任意一个负样本, e_k^- 表示负样本集合 N_{neg} 中第 k 个负样本, $\tau_e > 0$ 为视觉级温度系数,用于调节正样本和负样本之间距离。

1.3.3 域内实例—原型对比损失

由于语义空间与视觉空间维度不同,属于异构的流形空间,利用生成模型从语义特征生成伪视觉特征的过程中存在语义间隔,导致嵌入特征不能较好地反映真实的语义特征。尤其是细粒度数据集,类间差异较小,容易造成语义混淆。基于此,本文考虑分别在可见类和未见类执行域内实例—原型对比学习,以获得语义一致的特征表示。并使用交叉熵损失优化嵌入函数 E_{emb} 和比较器网络 R ,域内实例—原型对比损失函数为

$$L_{in}^p(h^i, a^+) = L_{in}^{ps}(h_s^i, a_s^+) + L_{in}^{pu}(\tilde{h}_u^i, a_u^+) = \mathbb{E} \left[-\log \frac{\exp \frac{R(h_s^i, a_s^+)}{\tau_s}}{\sum_{k=1}^S \exp \frac{R(h_s^i, a_s^+)}{\tau_s}} - \log \frac{\exp \frac{R(\tilde{h}_u^i, a_u^+)}{\tau_s}}{\sum_{l=1}^U \exp \frac{R(\tilde{h}_u^i, a_u^+)}{\tau_s}} \right] \quad (6)$$

式中, τ_i 为原型级温度系数, $\mathbf{h}_s^i, \mathbf{h}_u^i$ 分别表示可见类和未见类第 i 个实例, $\mathbf{a}_s^+, \mathbf{a}_u^-$ 分别表示与实例 $\mathbf{h}^i$ 相同标签、不同标签的语义原型, L_{in}^s, L_{in}^u 分别是可见类和未见类域内实例与原型的对比损失。

1.3.4 跨域中心—原型对比嵌入损失

由于训练过程缺乏未见类的视觉特征, 本文通过训练 VAE-GAN 模型来生成伪视觉特征。然而可见类和未见类特征之间存在差异, 仅使用可见类数据训练生成模型, 往往会导致域偏移, 使得模型的识别精度下降。在 GZSL 任务中, 这种现象尤为严重。针对该问题, 提出跨域中心—原型对比损失, 以类中心来捕获细粒度的聚类结构。通过比较网络学习嵌入空间中可见类的视觉中心与未见类语义原型之间的潜在结构关系, 有效地将知识从可见类迁移到未见类。尽管在训练过程中未见类的标签不可知, 但其语义原型是已知的, 因此, 本文首先计算可见类的语义原型与未见类语义原型 $s \rightarrow u$ 方向的相似性矩阵 $\mathbf{M}^{s \rightarrow u}$, 其中矩阵元素 y_{ij} 表示第 i 个可见类原型和第 j 个未见类原型之间的内积。同理, 计算未见类的伪视觉中心与可见类语义原型 $u \rightarrow s$ 方向的对比损失。跨域对比损失函数为

$$L_{cross}^{s \rightarrow u}(\mathbf{c}_s^i, \mathbf{a}_u^j) = \mathbb{E} \left[- \sum_{i=1}^K \sum_{j=1}^U \mathbf{M}_{y_{ij}}^{s \rightarrow u} \ln R(\mathbf{c}_s^i, \mathbf{a}_u^j) + (1 - \mathbf{M}_{y_{ij}}^{s \rightarrow u}) \ln (1 - R(\mathbf{c}_s^i, \mathbf{a}_u^j)) \right] \quad (7)$$

$$L_{cross}(\mathbf{c}_s^i, \mathbf{a}_u^j) = L_{cross}^{s \rightarrow u}(\mathbf{c}_s^i, \mathbf{a}_u^j) + L_{cross}^{u \rightarrow s}(\mathbf{c}_u^i, \mathbf{a}_s^j) \quad (8)$$

式中, $\mathbf{c}_s^i$ 定义为可见类的第 i 个类的类中心, $\mathbf{a}_u^j$ 为未见类的第 j 个类的语义原型。通过最小化损失函数提高模型对目标类的迁移能力, 从而达到缓解域偏移的目的。

1.3.5 总损失

综上, 本文模型的总损失函数为

$$L_{all} = L_{VAEGAN} + \lambda L_{in}^s(\mathbf{e}_i, \mathbf{e}^+) + \beta L_{in}^u(\mathbf{h}^i, \mathbf{a}^+) + \rho [L_{cross}^{s \rightarrow u}(\mathbf{c}_s^i, \mathbf{a}_u^j) + L_{cross}^{u \rightarrow s}(\mathbf{c}_u^i, \mathbf{a}_s^j)] \quad (9)$$

式中, λ 为域内实例—实例对比损失权重因子, β 为域内实例—原型对比损失权重因子, ρ 为跨域中心—原型对比损失权重因子。

1.4 模型训练的算法伪代码

算法 1 ZSIR-DCEL 训练过程

输入: 可见类视觉特征 $\mathbf{x}_s$ 和类别标签 $\mathbf{y}_s$, 可见类语义特征 $\mathbf{a}_s$ 和未见类的语义特征 $\mathbf{a}_u$, 训练轮数 T_1 和 T_2 。

输出: 各网络最佳参数;

1) 初始化: 初始化编码器 E、解码器 DEC/生成器 G、判别器 D、嵌入网络 E_{emb} 、非线性投影网络 H、比较网络 R 的参数。

2) for epoch = 1 to T_1 do;

根据式(4)计算生成对抗损失 L_{VAEGAN} ;

反向传播更新 E、DEC/G、D 网络参数;

输出: $\tilde{\mathbf{x}}_s$ 和 $\tilde{\mathbf{x}}_u$ 。

3) for epoch = 1 to T_2 do;

将 $\tilde{\mathbf{x}}_s, \tilde{\mathbf{x}}_u$ 和 $\mathbf{x}_s$ 输入网络 E_{emb} , 低维映射至嵌入空间;

根据式(5)计算域内实例—实例对比损失;

根据式(6)计算域内实例—原型对比损失;

根据式(8)计算跨域中心—原型对比损失;

反向传播更新 E_{emb} 、H、R 的网络参数。

end for

根据式(9)计算总损失函数;

通过 Adam 优化器优化所有网络参数。

end for;

4) 输出最佳网络参数。

1.5 识别阶段

1.5.1 零样本图像识别

模型训练完成后, 固定网络参数, 将高斯噪声 z_p 和未见类的属性 $\mathbf{a}_u$ 输入生成器 G, 即可获得未见类伪视觉特征, 然后再将其输入到对比嵌入函数获取未见类的嵌入伪特征 $\tilde{\mathbf{h}}_u$, 与对应类标签构成集合 $\tilde{\mathcal{C}}_u$ 。由于获得了对应类的未见类伪视觉特征, 零样本学习问题将转化为监督训练一个未见类分类器, softmax 分类器的目标函数为

$$L_{cls} = - \frac{1}{|\tilde{\mathcal{C}}_u|} \sum_{(\tilde{\mathbf{h}}_u, \mathbf{y}_u) \in \tilde{\mathcal{C}}_u} \log(p(\mathbf{y}_u | \tilde{\mathbf{h}}_u; \theta_{cls})) \quad (10)$$

$$\tilde{\mathbf{h}}_u = H(G(z_p, \mathbf{a}_u)) \quad (11)$$

式中, θ_{cls} 表示 softmax 分类器的网络参数, $\mathbf{y}_u$ 表示未见类标签, 通过最小化损失函数式(11), 得到训练好的未见类分类器。将未见类伪视觉特征输入到对比嵌入层获取未见类的嵌入伪特征 $\mathbf{h}_u$, 使用分类器对 $\mathbf{h}_u$ 进行识别, 识别函数为

$$f(\mathbf{h}_u; \theta_{cls}) = \arg \max_{\mathbf{y}_u \in \mathcal{C}_u} p(\mathbf{y}_u | \mathbf{h}_u; \theta_{cls}) \quad (12)$$

1.5.2 广义零样本图像识别

对于GZSL任务,可利用嵌入函数获取未见类嵌入伪特征 $\tilde{h}_u$ 和可见类的嵌入真实特征 h_s ,构成集合 D 训练一个softmax分类器。训练完成后,再使用该分类器对未见类的嵌入伪特征 $\tilde{h}_u$ 和可见类的嵌入真实特征 h_s 进行识别,识别函数为

$$f(h_i; \theta_{cls}) = \arg \max_{(y, \cup y_u)} p(y | h_i; \theta_{cls}), h_i \in D(h_u, h_s) \quad (13)$$

2 实验

本文实验的硬件环境为Intel i7-9700K CPU、Nvidia RTX2080Ti GPU;软件环境为Ubuntu18.04 LTS操作系统、CUDA10.1.0、cuDNN7.5.1。模型的训练和测试均在PyTorch1.8.0环境下完成。

2.1 数据集

在ZSL的4个基准数据集上评估本文提出的方法,数据集分别是AWA1(Lampert等,2009)、AWA2(Xian等,2019a)、CUB(Wah等,2011)和SUN(Patterson和Hays,2012)。

本文将每个数据集上的属性注释作为语义描述,使用在ImageNet 1K上预训练好的ResNet-101提取图像的2048维视觉特征,参照Xian等人(2019a)提出的数据拆分方式将每个数据集划分为可见类和未见类。数据集详细信息如表2所示。

表2 数据集的统计信息
Table 2 Statistical information of the datasets

数据集	属性维度	可见类/未见类	训练集图像数	测试集图像数	
				可见类	未见类
AWA1	85	40/10	19 832	4 958	5 685
AWA2	85	40/10	23 527	5 882	7 913
CUB	312	150/50	7 057	1 764	2 967
SUN	102	645/72	10 320	2 580	1 440

2.2 网络参数设置和评估指标

在本文的网络框架中,参照Han等人(2022)的方法,设置嵌入维数 $h = 2\,048$,并将非线性投影的输出尺寸设置为512维,设置实例—实例对比嵌入损失的权重因子 $\lambda = 0.001$ 。在AWA1、CUB和SUN上,温度系数 τ_e 、 τ_s 均设置为0.1。在AWA2上,温度系数 τ_e 、 τ_s 分别设置为10.0和1.0, β 和 ρ 的选取将

在2.6.1节进一步讨论。本文网络使用Adam优化器,判别器D的学习率设置为0.0005,解码器/生成器G的学习率设置为0.0001。从每个批量中随机选取数据,其中来自同类的实例为正实例,来自不同类的实例为负实例。

参照Xian等人(2019a)提出的评估策略,在ZSL任务中,采用T1作为精度评估指标,T1为每个Top1分类准确率的均值。在GZSL任务中,分别评估可见类和未见类样本的Top-1准确率,分别表示为S和U。GZSL的性能用其调和平均值H来衡量,以H作为综合评价指标评估GZSL任务下模型的性能。

2.3 零样本图像识别

为了验证提出方法的有效性,在AWA1、AWA2、CUB和SUN这4个基准数据集上进行识别实验,并与近年来较为先进的12种相关模型比较,实验结果如表3所示。为了展示生成式方法以及本文提出方法的优势,选择了包括非生成式和生成式在内的多种方法进行比较,其中前5种方法为非生成式零样本图像识别方法,后7种方法是近年来提出的生成式零样本图像识别方法。从表3可以看出,本文提出的方法在AWA1和CUB数据集上均取

表3 在4个数据集上零样本识别的准确率
Table 3 Accuracy of zero-shot learning on four datasets

方法	Top1 分类准确率/%			
	AWA1	AWA2	CUB	SUN
DAP(Larochelle等,2008)	—	—	40.0	39.9
DPPN(Wang等,2021)	<u>73.3</u>	73.3	77.8	61.5
TransZero++(Chen等,2023a)	72.6	72.6	78.3	64.6
DUET(Chen等,2023c)	69.9	69.9	72.3	64.4
IAAC-net(Chen和Zhou,2024)	—	70.7	<u>80.3</u>	63.8
F-CLSWGAN(Gulrajani等,2017)	68.2	—	57.3	60.8
F-VAEGAN(Xian等,2019b)	71.1	—	61.0	64.7
HSVA(Chen等,2021b)	—	—	62.8	63.8
SCE-GZSL(Han等,2022)	71.5	69.4	78.6	62.5
DFTN(Zhai等,2023)	71.1	71.2	74.2	—
PCE-GZSL(Paul等,2023)	71.5	73.9	79.2	66.8
Co-GZSL(Li等,2024)	69.4	69.4	78.6	—
本文	75.5	<u>73.4</u>	81.9	<u>64.8</u>

注:加粗、下划线字体表示各列最优、次优结果。“—”表示引用文献未对此数据集进行实验。

得最优值,分别比相应数据集上次优的方法精度提升2.2%和1.6%。在AWA2和SUN数据集上均取得次优值,比最优方法的精度略有降低。本文方法与SCE-GZSL方法相同之处是都采用对比嵌入学习,但是本文方法识别精度在4个数据集上都高于SCE-GZSL方法,表明本文提出的域内和跨域对比嵌入学习方法更具优越性。

2.4 广义零样本图像识别

零样本学习(ZSL)的设置条件过于严格,而广义零样本学习(GZSL)测试样本可以来源于任意的目标类别。本实验的目的是验证提出方法在广义零样本识别任务上的有效性。

在4个基准数据集AWA1、AWA2、CUB和SUN上进行实验,结果如表4所示。参与比较的方法均为近年来提出的生成式广义零样本图像识别方法。

从表4可以看出,本文提出的方法在AWA1、AWA2和CUB数据集上H值均取得最优,分别为70.9%、71.5%和70.4%,比次优方法分别提高0.4%、0.8%和2.6%,而且是唯一在3个数据集上精度都达到70%以上的方法。在SUN数据集上H值取得次优,原因是SUN数据集是一个包含717个类别的中等规模细粒度场景数据集,每个类别平均仅有20个图像样本,正样本数量不足,从而导致提出的模型难以在每个类别上充分执行对比学习。此外,在CUB数据集上,本文方法的表现提升最为显著,原因分析如下:CUB作为中等规模的细粒度鸟类数据集,相比于粗粒度数据集AWA1和AWA2,更容易出现类间混淆。通过引入未见类的中心—原型对比损失有效增强了模型区分细微差异的能力,从而实现性能的显著提升。

表4 在4个数据集上广义零样本识别的精度
Table 4 Accuracy of generalized zero-shot learning on four datasets

方法	/%											
	AWA1			AWA2			CUB			SUN		
	S	U	H	S	U	H	S	U	H	S	U	H
F-CLSWGAN(Gulrajani等,2017)	61.4	57.9	59.6	—	—	—	57.7	43.7	49.7	36.6	42.6	39.4
f-VAEGAN(Xian等,2019b)	69.8	58.1	64.3	70.6	57.6	63.5	60.1	48.4	53.6	38.0	45.1	41.3
FREE(Chen等,2021a)	69.4	62.9	66.0	75.4	60.4	67.1	59.9	55.7	57.7	37.2	47.4	41.7
HSVA(Chen等,2021b)	<u>76.6</u>	59.3	66.8	79.8	56.7	66.3	58.3	52.7	55.3	39.0	48.6	43.3
ESZSL(Cetin等,2022)	74.7	55.2	63.5	78.0	51.3	61.8	60.4	51.6	55.7	36.5	48.2	41.5
TDCSS(Feng等,2022)	69.8	54.4	60.9	74.9	59.2	66.1	62.8	44.2	51.9	—	—	—
SCE-GZSL(Han等,2022)	75.1	65.1	69.7	77.5	64.3	70.3	68.6	66.5	67.6	41.7	45.9	43.7
CLSWGAN+DSP(Chen等,2023b)	—	—	—	86.0	60.0	<u>70.7</u>	63.8	51.4	56.9	<u>43.0</u>	48.3	45.5
CDL+OSCO(Cavazza等,2023)	63.0	34.0	44.2	71.0	48.0	57.1	65.0	32.0	42.9	69.0	29.0	40.6
DFTN(Zhai等,2023)	83.6	56.3	67.3	78.5	61.1	68.7	67.2	61.8	64.4	—	—	—
PCE-GZSL(Paul等,2023)	73.8	<u>67.2</u>	70.3	74.8	67.0	70.6	65.8	<u>66.9</u>	66.4	37.9	<u>51.3</u>	43.6
CEZSL+DGZSL(Wang等,2024)	76.2	65.7	<u>70.5</u>	76.7	64.6	70.1	<u>69.1</u>	66.7	<u>67.8</u>	—	—	—
本文	74.7	67.5	70.9	<u>79.9</u>	<u>64.7</u>	71.5	71.9	68.9	70.4	36.0	51.4	<u>43.9</u>

注:加粗、下划线字体表示各列最优、次优结果,“—”表示引用文献未对此数据集进行实验。

本文的模型相比于同样采用对比嵌入的SCE-GZSL模型识别精度有较好的提升,这是因为本文的模型能生成更多具有判别性的特征,且增加的域内实例—原型对比嵌入损失在一定程度上改善了类内识别错误。此外,中心—原型跨域对比损失,使得模型可以学习到更多的可见类与未见类间的联系,在

一定程度上增强了可见类和未见类之间的决策边界,有效缓解了域偏移问题。

2.5 消融实验

为了进一步探明提出方法中不同模块的有效性,在AWA1和CUB数据集上,分别在ZSL和GZSL任务上进行消融实验,实验结果如表5所示。

表 5 在两个数据集上的消融实验
Table 5 Ablation study on two datasets

方法	AWA1				CUB			
	ZSL		GZSL		ZSL		GZSL	
	T1	S	U	H	T1	S	U	H
SCE-GZSL(WGAN)	71.0	73.4	65.3	69.1	77.5	66.8	63.9	65.3
+VAEGAN(基线)	72.7	76.9	63.2	69.4	79.2	69.2	62.1	65.5
基线 + L_{in}^{pu}	72.9	75.7	65.6	70.3	80.8	69.0	64.8	66.8
基线 + L_{cross}	75.2	74.6	66.8	70.5	80.7	72.2	66.8	69.4
基线 + $L_{in}^{pu} + L_{cross}$	75.5	74.7	67.5	70.9	81.9	71.9	68.9	70.4

注:加粗字体表示各列最优结果。

首先评估SCE-GZSL模型在ZSL/GZSL任务上的性能,该模型使用的网络框架为WGAN。本文在SCE-GZSL基础上,用VAEGAN替换WGAN,并以此作为基线模型。相比于SCE-GZSL方法,本文提出的方法在ZSL和GZSL任务上,在两个数据集上,T1和H值均有提升。在AWA1和CUB数据集上对于GZSL任务,可见类的识别精度分别提高3.5%和2.4%,表明VAE-GAN与WGAN相比,拟合复杂特征分布的能力更强,生成的伪特征更接近可见类的真实特征。

从表5可以看出,基线+ L_{in}^{pu} (域内未见类实例—原型对比损失)相比于基线,在AWA1和CUB数据集上H值分别提升0.9%和1.3%,T1值提升0.2%和1.6%,验证了 L_{in}^{pu} 对两个任务都有贡献,其中GZSL的改进比ZSL的改进更显著,表明域内实例—原型对比损失增强了特征的语义一致性,能较好地缓解同类别的特征混淆。

基线网络 + L_{cross} (跨域中心—原型对比损失)相比于基线模型,在ZSL任务上,在两个数据集上T1分别提升2.5%和1.5%,这种改进归功于:AWA1是粗粒度的数据集,类间语义区分性较强,在进行跨域对比学习过程中可使得可见类的知识更好地迁移到未见类,从而使模型在测试前尽可能地适应未见类的分布。CUB是一个细粒度的数据集,类间语义差异较小,这可能会导致类间特征混淆,表5中,在CUB数据集上,GZSL任务的H值提升3.9%,表明本文提出的中心—原型对比学习有助于改善类间距离并减少错误识别,从而在CUB数据集上获得良好的

性能。细粒度数据集更依赖于可靠的知识迁移。

基线网络 + $L_{in}^{pu} + L_{cross}$ 是本文提出的方法,从表5可以看出,在AWA1和CUB数据集上,在ZSL/GZSL任务上,本文的模型在可见类和未见类均获得了比基线更优的表现,域偏移问题有较大改善,这是因为提出的模型受益于双重对比学习。

2.6 模型中参数的分析

2.6.1 参数敏感性分析

对于总损失函数式(9)中的参数 λ ,参照Han等人(2022)的选取方法。本文主要对参数 β 和 ρ 进行实验分析,在4个数据集AWA1、AWA2、CUB和SUN上对识别任务的权重超参数敏感性进行实验评估,实验结果如图3和图4所示。图3和图4分别为ZSL/GZSL任务下的超参数实验,横轴表示超参数值,纵轴表示T1值。通过多次实验,选择如图3和图4所示的5个不同超参数值进行分析,每种颜色代表不同的超参数值。

对于ZSL任务,从图3(a)可以看出,在AWA1和AWA2数据集上,随着 β 值的变化,模型精度变化较小,当 $\beta = 0.001$ 时,模型性能达到最优,说明参数 β 对于这两个数据集不敏感,本文认为可能是粗粒度数据集语义之间差异较大,特征混淆程度较小,域内实例—原型对比嵌入损失改善ZSL决策边界的效果不明显。而对于细粒度数据集CUB,随着 β 值变化模型精度变化较为明显,T1值先升高而后缓慢下降,当 $\beta = 0.001$ 时,模型性能最优。该实验表明 β 值对CUB数据集更敏感,这是由于细粒度数据集不同实例语义差异小,容易造成特征混淆,增加域内

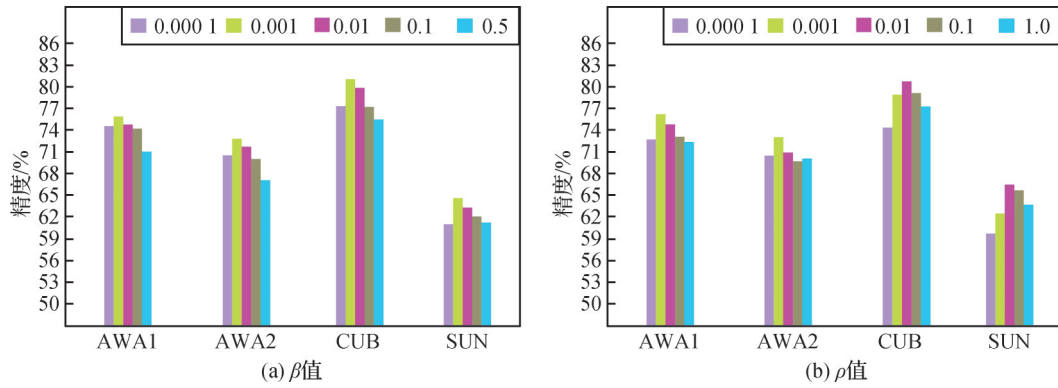


图3 在4个数据集上取不同超参数值时的精度(ZSL任务)

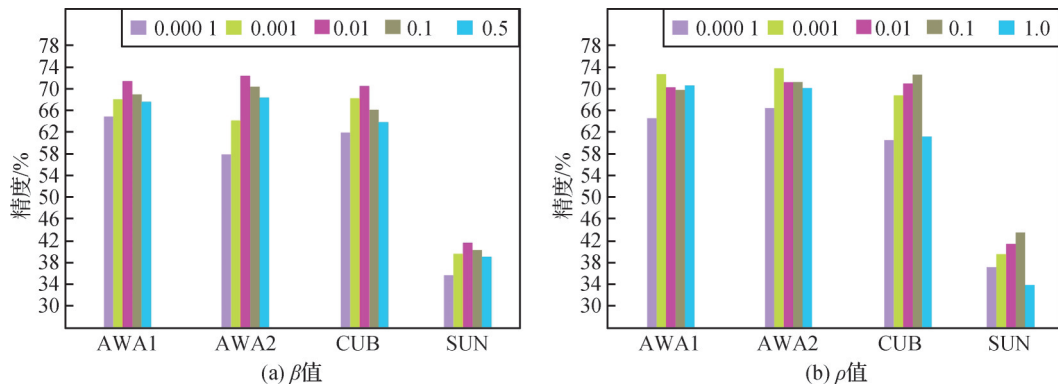
Fig. 3 Accuracy (for ZSL task) with different hyperparameter values on four datasets ((a) β value; (b) ρ value)

图4 在4个数据集上取不同超参数值时的精度(GZSL任务)

Fig. 4 Accuracy (for GZSL task) with different hyperparameter values on four datasets ((a) β value; (b) ρ value)

实例—原型对比损失后缓解了特征混淆。此外,可以看到在SUN数据集上, β 值变化也不大,当 $\beta = 0.001$ 时达到最优。

从图3(b)可以看出,在AWA1和AWA2数据集上,随着 ρ 值变化模型精度整体变化幅度不大,当 $\rho = 0.001$ 达到最优值,而CUB和SUN数据集上,随着 ρ 值变化精度先迅速上升后缓慢下降,当 $\rho = 0.01$ 时,在CUB、SUN上获得最优。这是因为本文提出的跨域中心—原型对比嵌入损失对参数较敏感,在较大参数范围内都能实现性能提升,表明本文方法具有较好的鲁棒性。

对于GZSL任务,从图4(a)可以看出,随着 β 值的增大模型在4个数据集上的精度整体得到提升,趋势为先升高而后缓慢下降,当 $\beta = 0.01$ 时,获得最优值。但 β 值在GZSL上比ZSL任务上更敏感,原因是广义零样本测试集中包含了可见类和未见类,特征混淆程度相对较大。

从图4(b)可以看出,在AWA1和AWA2数据集

上,随着 ρ 值变化识别精度先迅速上升而后基本变化不大,当 $\rho = 0.001$ 时获得最优。而在CUB、SUN上,随着 ρ 值的增大模型性能逐步提升,当 $\rho = 0.1$ 时获得最优,超过该值后精度迅速下降,可以看出跨域中心—原型对比嵌入损失在GZSL上对 ρ 值比较敏感,表明了细粒度数据集更依赖于生成高质量的未知类样本。此外,本文认为导致精度迅速下降的原因可能是:过分加强跨域中心—原型对比损失,也有可能产生负迁移,从而降低识别精度。

2.6.2 生成样本数量

本实验研究生成样本数量对模型性能的影响,结果如图5和图6所示。其中,横坐标表示生成样本数量,纵坐标表示模型的识别精度。

如图5所示,当在ZSL任务上为每个未见类生成1 000, 1 000, 700, 200个样本时,本文方法在AWA1, AWA2, CUB和SUN获得最佳精度。从图6可以看出,当在GZSL任务上为每个未知类生成1 300, 900, 400, 200个样本时,本文方法在AWA1,

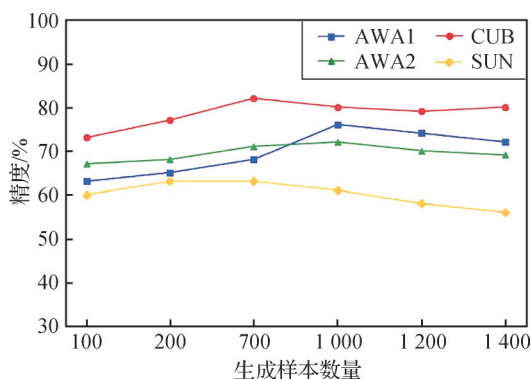


图5 生成样本数量对ZSL精度的影响

Fig. 5 The impact of the number of generated samples on ZSL accuracy

AWA2, CUB和SUN获得最高识别精度。

2.6.3 可视化实验

由于本文实验的基准数据集类别众多,用生成式零样本识别方法不便于做图像识别的可视化实验。参照 Han 等人(2022)和 Wang 等人(2024)的方法,本文进行t-SNE可视化实验,目的是验证本文方法中使用低维嵌入空间的有效性。本次实验在粗粒度数据集 AWA2 和细粒度数据集 CUB 的测试集上随机挑选了5个未见类进行t-SNE降维可视化实验,

结果如图7所示。用不同颜色表示不同的类别。具体来说,利用生成器G分别为AWA2数据集和CUB数据集生成伪视觉特征,并将其映射到嵌入空间得到嵌入特征,从这些特征中随机抽取50个样本进行可视化。因为本文方法是在SCE-GZSL进行改进,所以选择与SCE-GZSL进行对比实验。从图7(a)(c)可以看出,SCE-GZSL方法对某些类的类簇结构不明显,存在一定程度的特征混淆,分析原因是存在域偏移、边界判别性信息约束不强等;而从图7(b)(d)可以看出,本文方法生成的特征类簇结构更为明显,边界更为清晰。这是因为本文模型的未见类域内实例—原型对比损失可以有效地改善特征混淆,加强特征的判别性;其次,提出的跨域中心—原型对比损失在一定程度上促进了知识迁移,使得捕获聚类结构的效果更佳。

3 结论

本文以条件VAE-GAN为基本网络框架,提出一种结合双重对比嵌入学习的生成式零样本图像识别方法。首先,针对特征的判别性不足问题,通过在生

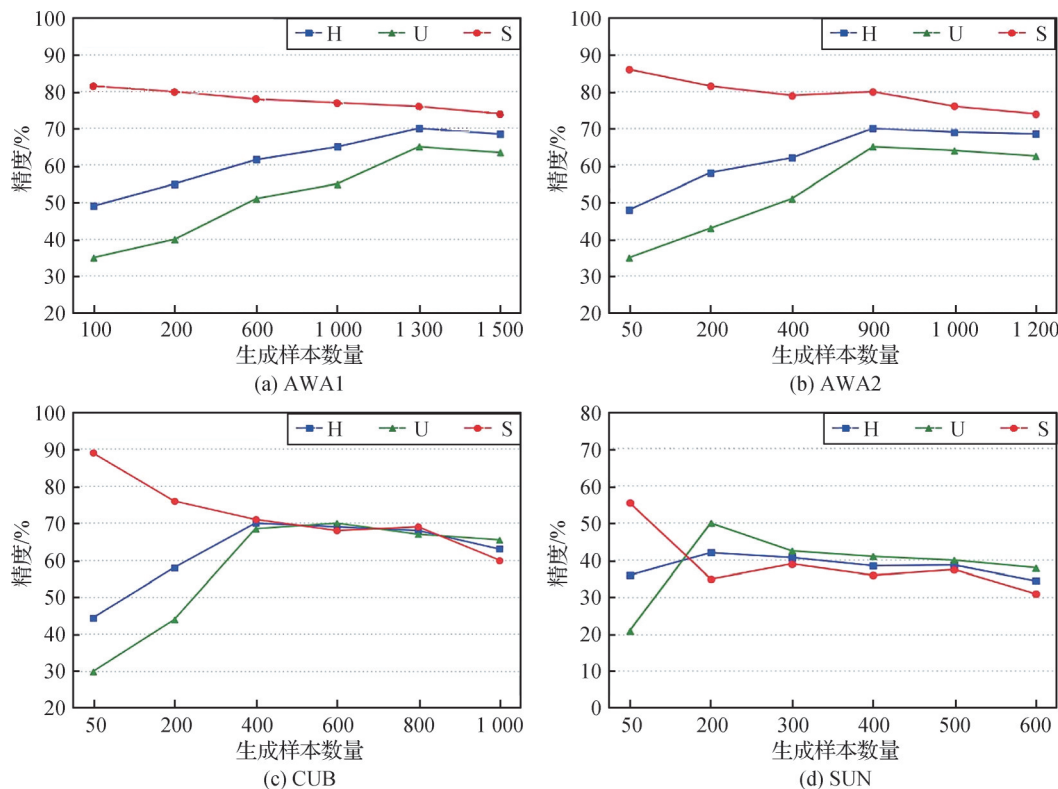


图6 生成样本数量对GZSL的影响

Fig. 6 The impact of the number of generated samples on GZSL ((a) AWA1; (b) AWA2; (c) CUB; (d) SUN)

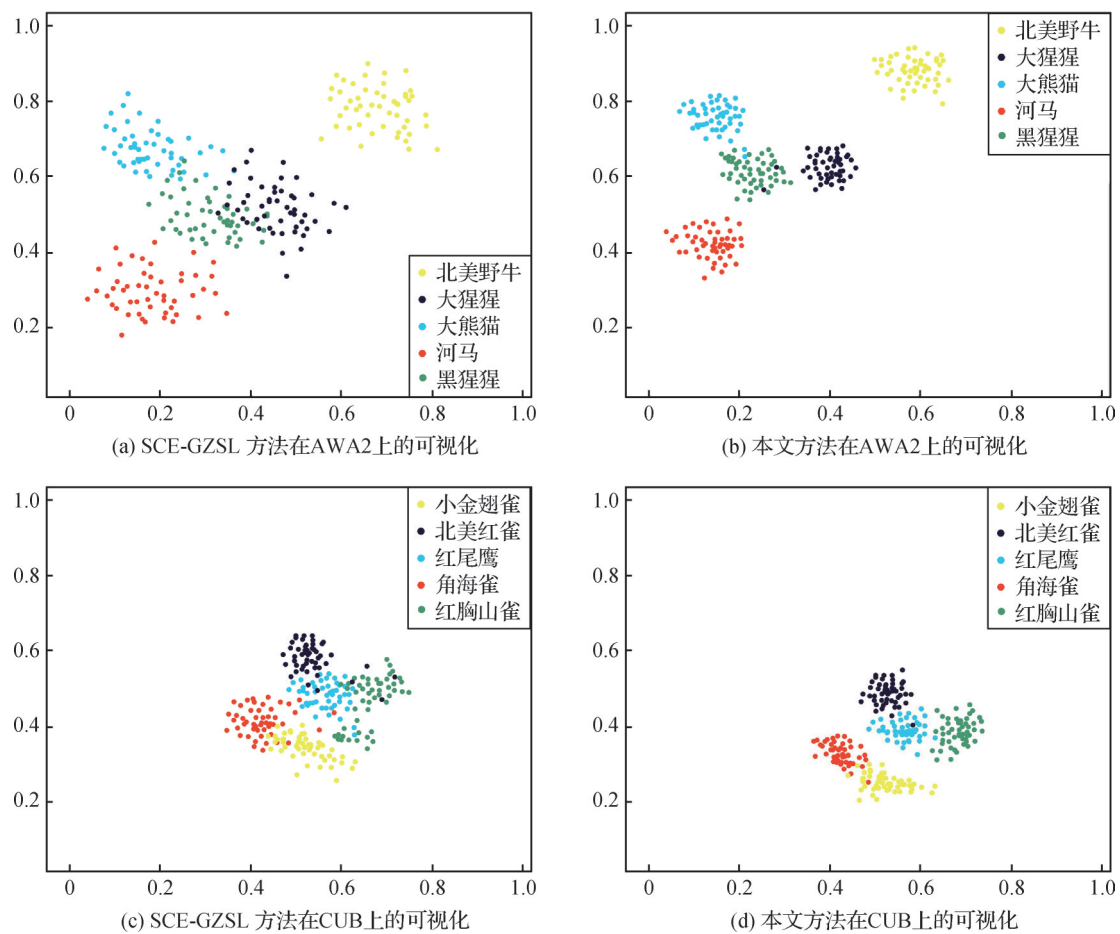


图7 AWA2和CUB数据集上的t-SNE可视化图

Fig. 7 t-SNE visualization on AWA2 and CUB datasets ((a) visualization of the SCE-GZSL method on AWA2; (b) visualization of the method in this paper on AWA2; (c) visualization of the SCE-GZSL method on CUB; (d) visualization of the method in this paper on CUB)

成式零样本学习中引入嵌入模型和对比学习,多个网络协同训练,增强伪特征的生成质量,从而提高零样本图像识别的精度;其次,本文提出双重对比学习策略,同时考虑了域内和跨域的双重信息。一方面,在现有可见类对比学习的基础上,引入未见类伪样本实例—原型域内对比学习,使得生成的伪视觉特征与语义信息保持一致性,缓解可见类和未见类的语义混淆;另一方面,提出跨域中心—原型对比学习,通过视觉中心和语义原型的强对齐来修正类间方差,促进跨域知识更好地迁移,缓解模型过于偏向于可见类,在一定程度上减轻域偏移。最后,在4个标准数据集上进行验证,实验结果表明提出的方法可提高零样本识别和广义零样本识别的精度。在后续研究中,将考虑在识别精度和识别效率间进行折中,设计轻量化网络,以降低模型的复杂度,提高识别效率。

参考文献 (References)

- Cavazza J, Murino V and Del Bue A. 2023. No adversaries to zero-shot learning: distilling an ensemble of Gaussian feature generators. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(10): 12167-12178 [DOI: 10.1109/TPAMI.2023.3282971]
- Cetin S, Baran O B and Cinbiş R G. 2022. Closed-form sample probing for learning generative models in zero-shot learning//*Proceedings of 2022 International Conference on Learning Representations (ICLR)*. openreview.net
<https://openreview.net/forum?id=ljxWpdBl4V>
- Chao W L, Changpinyo S, Gong B Q and Sha F. 2016. An empirical study and analysis of generalized zero-shot learning for object recognition in the wild//*Proceedings of the 14th European Conference on Computer Vision—ECCV 2016*. Amsterdam, the Netherlands: Springer International Publishing: 52-68 [DOI: 10.1007/978-3-319-46475-6_4]
- Chen S M, Hong Z M, Hou W J, Xie G S, Song Y B, Zhao J, You X

- G, Yan S C and Shao L. 2023a. TransZero++ : cross attribute-guided Transformer for zero-shot learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(11): 12844-12861 [DOI: 10.1109/TPAMI.2022.3229526]
- Chen S M, Hou W J, Hong Z M, Ding X H, Song Y B, You X G, Liu T L and Zhang K. 2023b. Evolving semantic prototype improves generative zero-shot learning//*Proceedings of the 40th International Conference on Machine Learning*. Honolulu, USA: JMLR.org: 4611-4622 [DOI: 10.5555/3618408.3618590]
- Chen S M, Wang W J, Xia B H, Peng Q M, You X G, Zheng F and Shao L. 2021a. FREE: feature refinement for generalized zero-shot learning//*Proceedings of 2021 IEEE/CVF International Conference on Computer Vision*. Montreal, Canada: IEEE: 122-131 [DOI: 10.1109/ICCV48922.2021.00019]
- Chen S M, Xie G S, Liu Y, Peng Q M, Sun B G, Li H, You X G and Shao L. 2021b. HSVA: hierarchical semantic-visual adaptation for zero-shot learning//*Proceedings of the 35th Conference on Neural Information Processing Systems*. 34: 16622-16634 [DOI: 10.48550/arXiv.2109.15163]
- Chen T, Kornblith S, Norouzi M and Hinton G. 2020. A simple framework for contrastive learning of visual representations//*Proceedings of the 37th International Conference on Machine Learning*. [s.l.]: JMLR.org: 1597-1607 [DOI: 10.5555/3524938.3525087]
- Chen Y and Zhou Y. 2024. Incorporating attribute-level aligned comparative network for generalized zero-shot learning. *Neurocomputing*, 573: #127188 [DOI: 10.1016/j.neucom.2023.127188]
- Chen Z, Huang Y F, Chen J Y, Geng Y X, Zhang W, Fang Y Z, Pan J and Chen H J. 2023c. DUET: cross-modal semantic grounding for contrastive zero-shot learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(1): 405-413 [DOI: 10.1609/aaai.v37i1.25114]
- Du Y, Shi M J, Wei F Y and Li G Q. 2024. Boosting zero-shot learning via contrastive optimization of attribute representations. *IEEE Transactions on Neural Networks and Learning Systems*, 35(11): 16706-16719 [DOI: 10.1109/TNNLS.2023.3297134]
- Feng Y G, Huang X W, Yang P B, Yu J and Sang J T. 2022. Non-generative generalized zero-shot learning via task-correlated disentanglement and controllable samples synthesis//*Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans, USA: IEEE: 9346-9355 [DOI: 10.1109/CVPR52688.2022.00913]
- Gao Y, Tang C W and Lyu J C. 2022. Cluster-based contrastive disentangling for generalized zero-shot learning [EB/OL]. [2024-09-06]. <https://arxiv.org/pdf/2203.02648.pdf>
- Goodfellow I J, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y. 2014. Generative adversarial nets//*Proceedings of the 28th International Conference on Neural Information Processing Systems*. Montreal Canada: MIT Press: 2672-2680 [DOI: 10.5555/2969033.2969125]
- Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V and Courville A. 2017. Improved training of Wasserstein GANs//*Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach, USA: Curran Associates Inc.: 5769-5579 [DOI: 10.5555/3295222.3295327]
- Han Z Y, Fu Z Y, Chen S and Yang J. 2022. Semantic contrastive embedding for generalized zero-shot learning. *International Journal of Computer Vision*, 130(11): 2606-2622 [DOI: 10.1007/s11263-022-01656-y]
- Krizhevsky A, Sutskever I and Hinton G E. 2017. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6): 84-90 [DOI: 10.1145/3065386]
- Lampert C H, Nickisch H and Harmeling S. 2009. Learning to detect unseen object classes by between-class attribute transfer//*Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition*. Miami, USA: IEEE: 951-958 [DOI: 10.1109/CVPR.2009.5206594]
- Larochelle H, Erhan D and Bengio Y. 2008. Zero-data learning of new tasks//*Proceedings of the 23rd National Conference on Artificial Intelligence*. Chicago, Illinois: AAAI: 646-651 [DOI: 10.5555/1620163.1620172]
- Li B F, Drozd A, Guo Y H, Liu T, Matsuoka S and Du X Y. 2019. Scaling Word2vec on big corpus. *Data Science and Engineering*, 4(2): 157-175 [DOI: 10.1007/s41019-019-0096-6]
- Li Q, Zhan Z X, Shen Y Y and Bhanu B. 2024. Co-GZSL: feature contrastive optimization for generalized zero-shot learning. *Neural Processing Letters*, 56(2): #99 [DOI: 10.1007/s11063-024-11557-5]
- Patterson G and Hays J. 2012. SUN attribute database: discovering, annotating, and recognizing scene attributes//*Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition*. Providence, USA: IEEE: 2751-2758 [DOI: 10.1109/CVPR.2012.6247998]
- Paul R, Vora S and Li B X. 2023. Instance adaptive prototypical contrastive embedding for generalized zero shot learning [EB/OL]. [2024-09-06]. <https://arxiv.org/pdf/2309.06987.pdf>
- Schönfeld E, Ebrahimi S, Sinha S, Darrell T and Akata Z. 2019. Generalized zero- and few-shot learning via aligned variational autoencoders//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, USA: IEEE: 8247-8255 [DOI: 10.1109/CVPR.2019.00844]
- Wah C, Branson S, Welinder P, Perona P and Belongie S. 2011. The Caltech-UCSD Birds-200-2011 Dataset. California Institute of Technology
- Wang C Q, Min S B, Chen X J, Sun X Y and Li H Q. 2021. Dual progressive prototype network for generalized zero-shot learning//*Proceedings of the 35th International Conference on Neural Information Processing Systems*. [s.l.]: Curran Associates Inc.: 2936-2948 [DOI: 10.5555/3540261.3540486]
- Wang Y J, Hong M J, Huangfu L W and Huang S. 2024. Data distribu-

- tion distilled generative model for generalized zero-shot recognition. Proceedings of the AAAI Conference on Artificial Intelligence, 38(6): 5695-5703 [DOI: 10.1609/aaai.v38i6.28381]
- Wang Z B, Liang J, He R, Xu N, Wang Z L and Tan T N. 2023. Improving zero-shot generalization for CLIP with synthesized prompts//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. 3032-3042 [DOI: 10.48550/arXiv.2307.07397]
- Wei X S, Xu Y Y and Yang J. 2022. Review of webly-supervised fine-grained image recognition. Journal of Image and Graphics, 27(7): 2057-2077 (魏秀参, 许玉燕, 杨健. 2022. 网络监督数据下的细粒度图像识别综述. 中国图象图形学报, 27(7): 2057-2077) [DOI: 10.11834/jig.210188]
- Xian Y Q, Lampert C H, Schiele B and Akata Z. 2019a. Zero-shot learning — A comprehensive evaluation of the good, the bad and the ugly. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(9): 2251-2265 [DOI: 10.1109/TPAMI.2018.2857768]
- Xian Y Q, Sharma S, Schiele B and Akata Z. 2019b. F-VAEGAN-D2: a feature generating framework for any-shot learning//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 10275-10284 [DOI: 10.1109/CVPR.2019.01052]
- Xu W J, Xian Y Q, Wang J N, Bernt S and Zeynep A. 2022. VGSE: visually-grounded semantic embeddings for zero-shot learning//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 9316-9325 [DOI: 10.1109/CVPR52688.2022.00910]
- Yang G Y, Ye Z H, Zhang R and Huang K Z. 2022. A comprehensive survey of zero-shot image classification: methods, implementation, and fair evaluation. Applied Computing and Intelligence, 2(1): 1-31 [DOI: 10.3934/aci.2022001]
- Zhai Z B, Li X and Chang Z H. 2023. Center-VAE with discriminative and semantic-relevant fine-tuning features for generalized zero-shot learning. Signal Processing: Image Communication, 111: #116897 [DOI: 10.1016/j.image.2022.116897]
- Zhang C, Jin M Y, Yu Q K, Xue H C, Gowda S N and Jin X B. 2024. Bridging the projection gap: overcoming projection bias through parameterized distance learning//Proceedings of the 17th Asian Conference on Computer Vision. Hanoi, Vietnam: Springer-Verlag [DOI: 10.1007/978-981-96-0966-6_15]
- Zhang G M, Long B Y and Lu F F. 2020. Zero-shot text recognition combining transfer guide and bidirectional cycle structure GANs. Pattern Recognition and Artificial Intelligence, 33(12): 1083-1096 (张桂梅, 龙邦耀, 鲁飞飞. 2020. 结合迁移引导和双向循环结构GAN的零样本文本识别. 模式识别与人工智能, 33(12): 1083-1096) [DOI: 10.16451/j.cnki.issn1003-6059.202012003]

作者简介

张桂梅, 女, 教授, 主要研究方向为计算机视觉和零样本图像识别。E-mail: guimei.zh@163.com

闫文尚, 男, 硕士研究生, 主要研究方向为计算机视觉和零样本学习。E-mail: y_wenshang@163.com

黄军阳, 男, 硕士研究生, 主要研究方向为计算机视觉和零样本学习。E-mail: 1160554026@qq.com