

控制与决策

Control and Decision

基于未知类语义约束自编码的零样本图像分类

王雪松, 张淳, 程玉虎

引用本文:

王雪松, 张淳, 程玉虎. 基于未知类语义约束自编码的零样本图像分类[J]. *控制与决策*, 2023, 38(12): 3499–3506.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2022.0274>

您可能感兴趣的其他文章

Articles you may be interested in

基于分类特征约束变分伪样本生成器的类增量学习

Class incremental learning based on variational pseudo-sample generator with classification feature constraints

控制与决策. 2021, 36(10): 2475–2482 <https://doi.org/10.13195/j.kzyjc.2020.0228>

一种基于多层语义特征的图像理解方法

An image understanding method based on multi-level semantic features

控制与决策. 2021, 36(12): 2881–2890 <https://doi.org/10.13195/j.kzyjc.2020.0927>

改进集成深层自编码器在轴承故障诊断中的应用

Application of improved ensemble deep auto-encoder in bearing fault diagnosis

控制与决策. 2021, 36(1): 135–142 <https://doi.org/10.13195/j.kzyjc.2019.0270>

融合稀疏编码与深度学习的草图特征表示

A feature representation of sketch based on fusion of sparse coding and deep learning

控制与决策. 2021, 36(3): 699–704 <https://doi.org/10.13195/j.kzyjc.2019.0941>

基于协同聚类和权重注意力稀疏自编码网络的变化检测方法

Change detection approach based on cooperative clustering and weighted-attention sparse autoencoder

控制与决策. 2021, 36(10): 2442–2450 <https://doi.org/10.13195/j.kzyjc.2019.1633>

基于未知类语义约束自编码的零样本图像分类

王雪松, 张 淳, 程玉虎[†]

(中国矿业大学 信息与控制工程学院, 江苏 徐州 221116)

摘要: 为缓解传统零样本图像分类模型中存在的领域偏移问题, 提出一种基于未知类语义约束自编码的零样本图像分类模型. 首先, 利用预训练的 ResNet101 网络提取所有已知类和未知类图像的视觉特征; 其次, 通过编码器将提取的图像深度视觉特征从视觉空间映射到语义空间; 然后, 通过解码器将映射后得到的语义向量重构为视觉特征向量, 在语义自编码器的训练过程中, 利用未知类图像的聚类视觉中心和未知类语义类原型的分布对齐施加约束, 以缓解领域偏移问题; 最后, 基于经编码器预测得到的测试图像语义向量和各测试类语义类原型之间的相似性, 采用最近邻算法实现零样本图像分类. 在 AwA2 和 CUB 数据集上的实验结果表明, 所提出模型具有较高的分类准确度.

关键词: 零样本图像分类; 语义约束; 领域偏移; 自编码器

中图分类号: TP273

文献标志码: A

DOI: 10.13195/j.kzyjc.2022.0274

引用格式: 王雪松, 张淳, 程玉虎. 基于未知类语义约束自编码的零样本图像分类[J]. 控制与决策, 2023, 38(12): 3499-3506.

Zero-shot image classification based on unknown-class semantic constraint autoencoder

WANG Xue-song, ZHANG Chun, CHENG Yu-hu[†]

(School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China)

Abstract: In order to alleviate the domain shift problem in traditional zero-shot image classification models, a zero-shot image classification model based on the unknown-class semantic constraint autoencoder is proposed. Firstly, a pre-trained ResNet101 network is used to extract visual features of all known-class and unknown-class images. Then, the extracted deep visual features are mapped from the visual space to semantic space through an encoder. Next, the obtained semantic vectors after mapping are reconstructed into the visual feature vectors through a decoder. In the training process of the semantic autoencoder, a constraint is imposed based on the distribution alignment between the clustering visual centers of unknown-class images and the unknown-class semantic prototypes to alleviate the domain shift problem. Finally, based on the similarity between the semantic vector of testing image predicted by the encoder and the semantic prototype of each testing class, the nearest neighbor algorithm is used to achieve zero-shot image classification. Experimental results on AwA2 and CUB datasets show that the proposed model has high classification accuracy.

Keywords: zero-shot image classification; semantic constraint; domain shift; autoencoder

0 引言

近年来,深度学习在图像分类、图像超分辨率重建和目标检测等领域均取得了较为成功的应用^[1-2]. 现有基于深度学习的图像分类模型大多属于监督型^[3], 在训练前需要准备大量带标签的样本数据, 而这些带标签的样本数据往往是通过人工标注得到的, 需要耗费巨大的人工成本. 此外, 现实世界中物体的种类还在不断增加, 存在大量没有标签的样本数据, 利用传统的监督学习模型难以对缺失标签的样本进

行识别. 因此, 如何识别未知类样本已成为目前研究热点.

2008 年, Larochelle 等^[4] 首次提出了零样本学习 (zero-shot learning, ZSL) 的概念, 即通过学习已知类样本数据的分类模型, 借助属性^[5]、词向量^[6] 等辅助信息实现对未知类样本数据的分类. 零样本学习实质上是迁移学习^[7] 的一个分支, 它与传统的监督分类问题^[8] 的区别在于, 零样本学习中训练集样本所属的类别与测试集样本所属的类别标签是不相交的. 零

收稿日期: 2022-02-20; 录用日期: 2022-06-10.

基金项目: 国家自然科学基金项目 (61976215, 62176259).

[†] 通讯作者. E-mail: chengyuhu@163.com.

样本学习的对象包含文字和图像,本文仅针对零样本图像分类问题进行研究。

现有的零样本图像分类方法大多根据源域的已知类数据,学习语义空间和视觉空间之间的映射函数,然后直接应用到目标域的未知类图像上,实现对未知类图像的识别^[9]。由于源域和目标域类别不同,两者的数据分布存在差异,导致零样本图像分类存在严重的领域偏移问题^[10]。比如斑马和猪都具有“尾巴”这一属性,但两者“尾巴”的视觉特征完全不同,如果将斑马的分类模型直接应用于猪的分类,则会产生较大的偏差。

为了缓解零样本图像分类存在的领域偏移问题,Kodirov等^[11]利用无监督领域自适应方法学习语义到视觉特征的映射,并进一步提出了语义自编码(semantic autoencoder, SAE)模型,将自动编码器的隐藏层赋予语义,在学习视觉特征到语义空间映射的同时对视觉特征进行重构,使学习到的映射模型对于新的测试类具有更好的泛化性。Romera-Paredes等^[12]采用两层线性网络对特征、属性和类之间的关系进行建模,提出一种极为简单的零样本学习(embarrassingly simple zero-shot learning, ESZSL)模型。Akata等^[13]将基于属性的图像分类问题视为标签嵌入问题,并利用一个线性模型将视觉特征映射到语义空间,提出了基于属性标签嵌入(attribute label embedding, ALE)的零样本学习模型。通过分析发现上述3种模型存在以下问题:1)均采用简单的线性模型,对于视觉空间和语义空间之间的映射关系描述能力有限;2)训练时缺乏未知类的信息,在对未知类样本进行分类时会产生较大的误差。特别是在广义零样本设置下,分类结果严重向已知类偏移,大量未知类样本被误判为已知类,导致未知类样本的分类准确度极低。为此,一些学者提出采用直推式的方式,在训练中充分利用未知类的信息以缓解领域偏移问题。Song等^[14]提出了一种准完全监督学习(quasi-fully supervised learning, QFSL)模型,通过在损失函数中增加未知类的偏置项损失使得未知类数据映射到未知类标签空间的概率最大化;Li等^[15]提出了基于双路视觉-语义映射路径(dual visual-semantic mapping paths, DMaP)模型,从流形对齐的角度研究语义嵌入空间的流形结构与映射模型迁移能力间的关系,能够根据先验知识预测图片特征空间的潜在流形结构,且可以生成最优的语义嵌入空间。近年来,许多学者采用生成模型的方法来缓解领域偏移问题。Schönfeld等^[16]提出了基于交叉-分布对齐的变分

自动编码器(variational autoencoders based on cross alignment and distribution alignment, CADA-VAE)模型,将第三空间作为嵌入空间,采用两个变分自编码器将已知类的视觉特征和语义描述同时嵌入到隐含空间中,利用分布对齐损失和交叉重建损失对自编码器进行训练。Verma等^[17]提出了基于生成样本(synthesized examples, SE)的ZSL模型,以变分自编码器为基础构建编码器和解码器(生成器),并利用判别器将生成的特征映射到相应的类别属性向量上,结合反馈、驱动机制训练生成器,再利用训练好的生成器生成未知类样本数据。Xian等^[18]提出了基于变分自编码和对抗生成网络的特征生成(feature generator-variational autoencoder and GANs-discriminator, F-VAEGAN-D2)模型,模型的主体结构结合了变分自编码和对抗生成网络且加入了翻译模块,使生成的特征具有可解释性。生成模型通过生成未知类数据,将零样本图像分类问题转化为监督图像分类问题,由于大多数生成模型生成的未知类视觉特征质量不佳,导致最后的分类模型表现并不尽人意。

综上所述,本文提出一种基于未知类语义约束自编码的零样本图像分类(zero-shot image classification based on unknown-class semantic constraint autoencoder, USCAE)模型,主要工作如下:

1) 利用语义自编码器,在学习视觉空间到语义空间映射的同时重构语义空间到视觉空间的映射。编码器和解码器均采用多层全连接网络,使学习到的非线性映射模型能够更好地描述视觉空间与语义空间之间的关系。

2) 利用未知类图像的聚类视觉中心和未知类语义类原型在语义空间中的分布对齐,对语义自编码器的训练进行约束,从而缓解领域偏移,提高模型的泛化性能。

3) 计算待测样本的预测语义向量与各测试类语义类原型之间的欧氏距离,采用最近邻方法实现零样本图像分类。

1 基于未知类语义约束自编码的零样本图像分类

1.1 模型框架

图1给出了基于未知类语义约束自编码的零样本图像分类模型的框架,主要分为3个阶段。

阶段1是提取各个类的视觉中心。利用预训练的ResNet101网络对所有已知类和未知类样本图像进行特征提取,得到各个样本的视觉特征。由于已知类

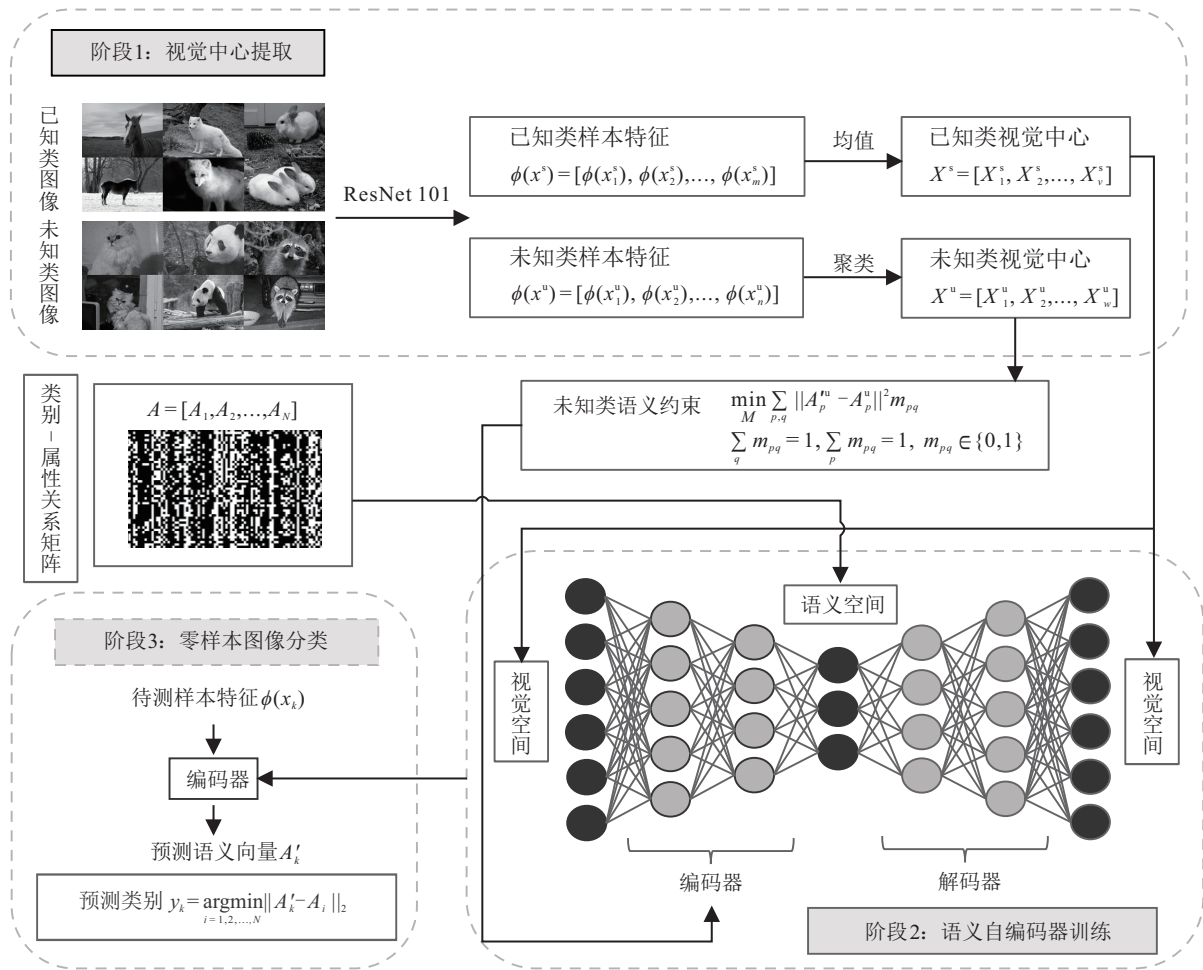


图1 基于未知类语义约束自编码的零样本图像分类

样本的标签在训练阶段已知,通过直接计算每一类所有样本的视觉特征的均值即可得到各个已知类的视觉中心;由于未知类的标签在训练阶段未知,采用无监督聚类方法对未知类所有样本的视觉特征进行聚类,得到未知类的近似视觉中心.

阶段2是语义自编码器训练.将已知类的视觉中心送入编码器中,得到已知类的预测语义向量,再通过解码器进行视觉特征的重建,通过迭代训练学习视觉空间到语义空间的映射,利用未知类的聚类视觉中心与未知类的语义类原型的分布对齐作为约束训练语义自编码器.

阶段3是零样本图像分类.将测试样本的视觉特征送入训练好的语义自编码器中,得到预测的语义向量,再利用欧氏距离计算预测的语义向量与各个测试类语义原型的相似性,采用最近邻方法即可得到测试样本的预测类别.

1.2 视觉中心提取

由于同一类的样本图像在视觉空间中的分布具有聚集性,本文将同一类所有样本的视觉特征点的中心点作为该类的视觉中心.相较于利用大量单个样

本的视觉特征进行训练,采用视觉中心进行训练来学习从视觉空间到语义空间的映射能够极大提高训练速度,同时避免错误样本或噪声对模型训练产生的不利影响.

首先,使用预训练 ResNet101 网络提取各个类别的每一个样本 x 的视觉特征,得到视觉特征向量 $\phi(x) \in R^{2048 \times 1}$.由于已知类的标签在训练时已知,可以直接计算每个已知类所有样本视觉特征的均值,得到已知类的视觉中心 $X^s = \{X_1^s, X_2^s, \dots, X_v^s\}$.其中: s 表示已知类, v 表示已知类的总类别数.假设第 i 类已知类有 N_i^s 个样本, $\phi(x_m^s)$ 是第 m 个样本经过 ResNet101 网络提取得到的视觉特征,则第 i 类已知类的视觉中心为

$$X_i^s = \frac{1}{N_i^s} \sum_{m=1}^{N_i^s} [\phi(x_m^s)], X_i^s \in R^{2048 \times 1}. \quad (1)$$

由于未知类的标签在训练时是不可知的,不能用提取已知类视觉中心的方法提取未知类的视觉中心.为此,本文采用 K 均值聚类算法对所有未知类样本的视觉特征进行聚类,得到各个未知类的近似视觉中心,第 j 类未知类的视觉中心为 $X_j^u \in R^{2048 \times 1}$.

利用 K 均值聚类算法提取未知类视觉中心的主要步骤如下。

输入: 未知类样本特征集 $\phi(x^u) = \{\phi(x_1^u), \phi(x_2^u), \dots, \phi(x_n^u)\}$, 聚类视觉中心个数 w , 最大迭代次数 R ;

输出: 未知类聚类视觉中心 $X^u = \{X_1^u, X_2^u, \dots, X_w^u\}$ 。

step 1: 随机初始化 w 个聚类视觉中心 $\{X_1^u, X_2^u, \dots, X_w^u\}$;

step 2: 计算样本特征 $\phi(x_i^u)$ 与各个聚类视觉中心 $\{X_1^u, X_2^u, \dots, X_w^u\}$ 之间的距离, 将其划分为距离最近的视觉中心的集合中;

step 3: 所有样本特征都划分完之后, 计算各个集合中所有样本特征的均值, 将其作为更新后的聚类视觉中心 $\{X_1^u, X_2^u, \dots, X_w^u\}$;

step 4: 重复 R 次 step 2 和 step 3, 得到最终的未知类聚类视觉中心 $\{X_1^u, X_2^u, \dots, X_w^u\}$ 。

1.3 语义自编码器训练

USCAE 模型中的编码器和解码器均采用两层全连接网络。相较于 Kodirov 等^[11]提出的 SAE 模型, USCAE 模型是非线性模型, 能够更好地描述视觉空间与语义空间之间的映射关系, 提高模型的泛化性能。

首先, 将各个已知类的视觉中心 X_i^s 送入编码器中, 得到各个已知类的映射语义向量 A_i^s , 有

$$A_i^s = \theta_2(W_2^T \theta_1(W_1^T X_i^s)). \quad (2)$$

其中: θ_1 和 θ_2 表示非线性网络层, W_1 和 W_2 是待学习编码器中两个全连接层的权值。

编码器将已知类的视觉中心映射到语义空间, 目的是使得已知类的映射语义向量和已知类的语义类原型在语义空间中进行对齐。由于已知类的样本标签已知, 可以做到一一对齐。编码器的目标函数为

$$L_{\text{encoder}} = \frac{1}{v} \sum_{i=1}^v \|A_i^s - A_i^s\|_2^2. \quad (3)$$

其中: v 为已知类的总类别数, A_i^s 为第 i 类已知类的映射语义向量, A_i^s 为第 i 类已知类的语义类原型。

然后, 将已知类的映射语义向量 A_i^s 送入解码器中, 得到重构的视觉向量 X_i^s , 有

$$X_i^s = \theta_4(W_4^T \theta_3(W_3^T A_i^s)). \quad (4)$$

其中: θ_3 和 θ_4 表示非线性网络层, W_3 和 W_4 是待学习解码器中两个全连接层的权值。因此, 解码器的目标函数定义为

$$L_{\text{decoder}} = \frac{1}{v} \sum_{i=1}^v \|X_i^s - X_i^s\|_2^2. \quad (5)$$

将解码器的重构损失作为正则化项, 采用最小均方差损失, 于是自编码器的目标函数为

$$L_{\text{seen}} = \frac{1}{v} \sum_{i=1}^v \|A_i^s - A_i^s\|_2^2 + \lambda_1 \times \frac{1}{v} \sum_{i=1}^v \|X_i^s - X_i^s\|_2^2, \quad (6)$$

其中 λ_1 是正则化系数。

1.4 未知类语义约束

现有的零样本图像分类模型大多利用已知类的学习视觉空间到语义空间的映射, 然后将未知类样本直接映射到语义空间中。由于源域和目标域的数据分布不同, 从源域学到的模型直接应用于目标域, 会导致未知类待测样本映射到语义空间中偏离该样本所属类别的真实语义类原型, 从而造成误分类。

USCAE 在训练时利用未知类的视觉特征, 将未知类样本聚类得到的近似视觉中心 X^u 也送入编码器, 从而得到未知类的映射语义向量 A^u 。进一步, 将未知类的映射语义向量 A^u 与未知类的语义类原型 A^u 进行对齐, 作为语义自编码器的约束项, 由此可以得到视觉空间与语义空间之间更好的映射关系, 进一步缓解领域偏移问题。

未知类视觉中心是通过无监督聚类算法得到的近似视觉中心, 在训练时类标签是未知的, 通过编码器得到的未知类的映射语义向量也是无标签的, 因此, 与未知类的语义类原型无法直接做到一一对应。本文采用全局最小分布距离作为未知类语义约束项的损失函数。假设矩阵 M 为未知类的映射语义向量集合 A^u 与未知类真实语义类原型集合 A^u 之间的匹配矩阵^[9], 点 p 为未知类的映射语义向量集合 A^u 中的任意一个点, 点 q 为未知类语义类原型集合 A^u 中的任意一个点, 矩阵 M 中的元素 m_{pq} 表示点 p 与点 q 之间的匹配关系, 则未知类的语义约束项可以表示为

$$L_{\text{unseen}} = \min_{m_{pq} \in M} \sum_{p,q} \|A_p^u - A_q^u\|_2^2 m_{pq}, \quad (7)$$

$$\sum_q m_{pq} = 1, \quad \sum_p m_{pq} = 1, \quad m_{pq} \in \{0, 1\}.$$

矩阵 M 遵循一对一的匹配原则, 采用匈牙利算法对矩阵 M 进行求解, 目的是找到两个集合之间的全局最小距离。

将未知类语义约束损失作为原语义自编码器目标函数的正则化项, 因此最终的目标函数为

$$L_{\text{USCAE}} = L_{\text{seen}} + \lambda_2 L_{\text{unseen}}, \quad (8)$$

其中 λ_2 是未知类语义约束项的正则化系数。

1.5 零样本图像分类

在进行零样本图像分类时, 首先将第 k 个待测样本经过 ResNet101 网络提取的视觉特征 $\phi(x_k)$ 送入编码器中, 得到该样本映射到语义空间中的预测语义向量 A'_k , 有

$$A'_k = \theta_2(W_2^T \theta_1(W_1^T \phi(x_k))). \quad (9)$$

然后, 计算该样本的预测语义向量与每一个测试类语义类原型之间的距离, 距离最近的类即为该样本的预测类别。

当进行传统零样本图像分类时, 测试类仅包含未知类, 分类公式为

$$y_k = \operatorname{argmin}_{i=1,2,\dots,w} \|A'_k - A_i^u\|_2. \quad (10)$$

其中: w 表示未知类的总类别数, A_i^u 表示第 i 类未知类的语义类原型。

当进行广义零样本图像分类时, 测试类同时包含已知类和未知类, 分类公式为

$$y_k = \operatorname{argmin}_{i=1,2,\dots,N} \|A'_k - A_i^{s+u}\|_2, \quad (11)$$

其中 N 表示已知类和未知类的总类别数。

2 实验

2.1 实验数据集

为验证 USCAE 的有效性, 在动物数据集 AwA2 和鸟类数据集 CUB 上分别进行实验。AwA2 是粗粒度数据集, 共包含 30 475 张动物图片, 有 50 个类, 每一类用 85 种属性来描述。CUB 是细粒度数据集, 共包含 11 788 张鸟类图片, 有 200 个类, 每一类用 312 种属性来描述。两个数据集包含的图像所属类别和属性数量差异较大, 能够充分验证 USCAE 的有效性。

为方便与现有方法进行对比, 按照文献 [20] 中的划分方式对数据集进行划分: AwA2 数据集的 40 类作为训练集 (共 24 295 个样本), 10 类作为测试集 (共 6 180 个样本); CUB 数据集的 150 类作为训练集 (共 8 855 个样本), 50 类作为测试集 (共 2 933 个样本)。实验过程中采用已知的属性空间作为语义空间, 数据集的属性均采用连续属性。

2.2 评价指标

实验过程中, 采用的评价指标包括: 类平均准确度、谐调平均准确度、ROC 曲线和 AUC 值。类平均准确度是零样本学习领域最常用的评价指标, 先对每个类统计类内的分类准确度, 再通过求均值计算得到类平均准确度。在传统零样本学习 (traditional ZSL, TZSL) 实验中, 测试类仅包含未知类, 采用类平均准确度进行评估。类平均准确度的计算公式^[20]为

$$\operatorname{Acc} = \frac{1}{w} \sum_{i=1}^w \frac{N_i}{N_i^u}. \quad (12)$$

其中: N_i 表示第 i 个未知类中被正确分类的样本数, N_i^u 表示第 i 个未知类的总样本数。

在广义零样本学习 (generalized ZSL, GZSL) 实验中, 测试类既包含未知类, 也包含已知类, 因此, 无法用类平均准确度作为评价指标。为了能够更合理地衡量模型的性能, 在 GZSL 实验设置下, 采用 Xian 等^[21] 提出的谐调平均准确度进行评估。谐调平均准确度的计算公式为

$$H = 2 \times \frac{\operatorname{Acc}^s \times \operatorname{Acc}^u}{\operatorname{Acc}^s + \operatorname{Acc}^u}. \quad (13)$$

其中: Acc^s 表示已知类数据的类平均准确度, Acc^u 表示未知类数据的类平均准确度。

ROC 曲线展示了假阳率与真阳率之间的关系。假阳率是指被误判为阳性的样本占有所有真实阴性样本的比例, 真阳率是指正确判别为阳性的样本占有所有真实阳性样本的比例。AUC 值定义为 ROC 曲线下方与坐标轴围成的面积。ROC 曲线越靠近左上方, 该 ROC 曲线下方的面积即 AUC 值越大, 说明分类效果越好。

2.3 实验设置与参数分析

所有实验基于 Pytorch 1.7.1 深度学习框架完成, 实验平台为: 英特尔 Core i7-6850 六核, 64 GB 内存, GTX1080Ti×4 显卡, Ubuntu16.04 系统。为提取到更具判别性的视觉特征, 采用预训练的 ResNet101 网络进行特征提取。在特征提取前, 将所有图像的尺寸调整为 224×224 , 经过 ResNet101 网络提取到 2 048 维深度视觉特征。模型参数设置为: 自编码器的编码器和解码器均采用两层全连接网络, 使用 LeakyReLU 函数作为激活函数; 编码器的隐含层为 {1 024, 512}, 解码器的隐含层为 {512, 1 024}; 模型训练过程中使用 Adam 优化器, 以 0.000 1 的固定学习率迭代训练 5 000 次。

在传统零样本学习设置下, AwA2 数据集上参数 λ_1 的取值范围设置为 {0.05, 0.1, 0.15, 0.2, 0.3, 0.7, 1}, 参数 λ_2 的取值范围设置为 {0, 0.1, 0.12, 0.125, 0.13, 0.15, 0.2}; CUB 数据集上参数 λ_1 的取值范围设置为 {0.1, 0.3, 0.5, 0.7, 0.75, 0.8, 1}, 参数 λ_2 的取值范围设置为 {0, 0.01, 0.03, 0.04, 0.05, 0.08, 0.1}。在广义零样本学习设置下, AwA2 数据集上参数 λ_1 的取值范围设置为 {0.1, 0.15, 0.2, 0.25, 0.3, 0.5, 1}, 参数 λ_2 的取值范围设置为 {0, 0.1, 0.115, 0.12, 0.125, 0.15, 0.2}; CUB 数据集上参数 λ_1 的取值范围设置为 {0.1, 0.3, 0.5, 0.7, 0.75, 0.8, 1}, 参数 λ_2 的取值范围设置为 {0, 0.1, 0.15,

0.19, 0.2, 0.21, 0.25, 0.3}.

图2给出了在传统零样本学习设置下, AwA2和CUB数据集上参数λ₁和λ₂对USCAE模型的类平均准确度的影响. 可以看出: 固定λ₂的取值, 两个数据集上的类平均准确度均随着λ₁的增加先增加后减小; 固定λ₁的取值, 两个数据集上的类平均准确度同样随着λ₂的增加先增加后减小. 因此, 在AwA2数据集上, 参数λ₁和λ₂的取值分别为0.15和0.125; 在CUB数据集上, 参数λ₁和λ₂的取值分别为0.75和0.04.

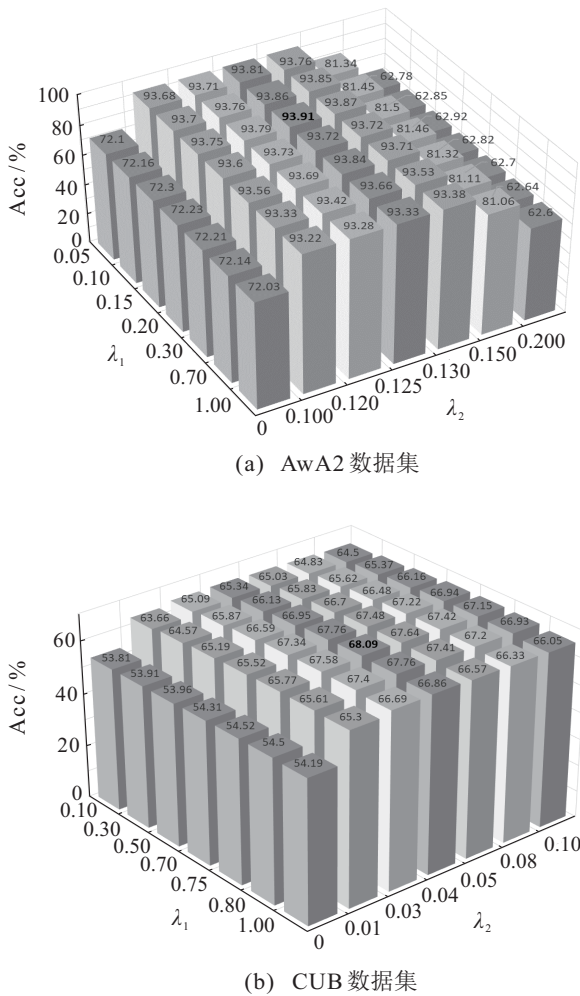


图2 USCAE的类平均准确度与参数λ₁和λ₂的关系(TZSL)

图3给出了在广义零样本学习设置下, AwA2和CUB数据集上参数λ₁和λ₂对USCAE模型的谐调平均准确度的影响. 可以看出: 固定λ₂的取值, 两个数据集上的谐调平均准确度均随着λ₁的增加先增加后减小; 固定λ₁的取值, 两个数据集的谐调平均准确度同样随着λ₂的增加先增加后减小. 当λ₂设置为0时, 即模型去除未知类语义约束时, 由于测试类别包含已知类, 受到已知类的干扰, USCAE的谐调平均准确度极低, AwA2和CUB数据集上均仅有18%左右的准

确度; 加上未知类语义约束后, USCAE的谐调平均准确度大大提升, 可以看出未知类语义约束极大地缓解了领域偏移问题. 因此, 在AwA2数据集上, 参数λ₁和λ₂的取值分别为0.25和0.12; 在CUB数据集上, 参数λ₁和λ₂的取值分别为0.75和0.2.

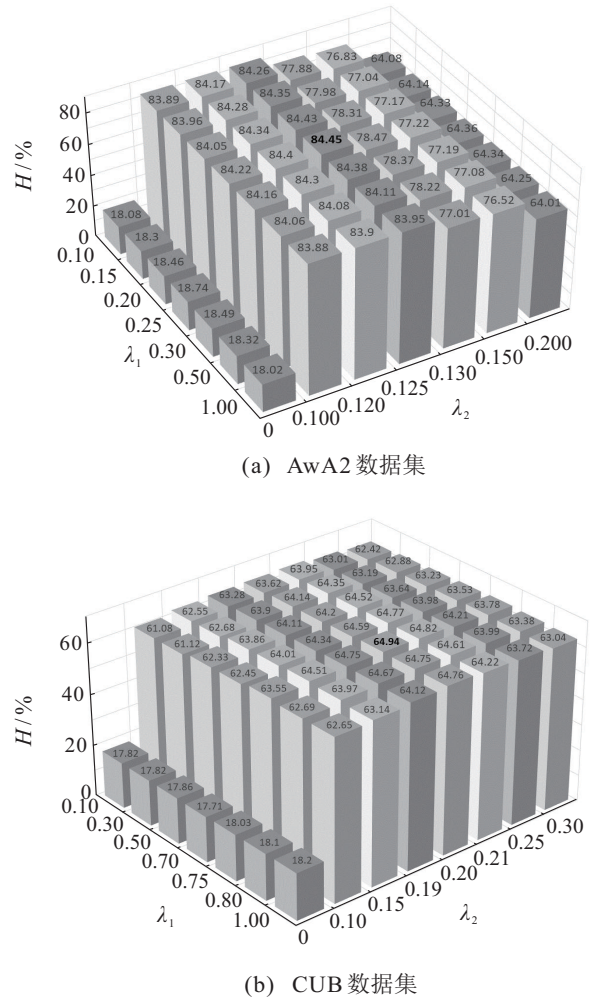


图3 USCAE的谐调平均准确度与参数λ₁和λ₂的关系(GZSL)

2.4 对比实验

为更准确、公平地评估USCAE在零样本图像分类上的分类性能, 在传统零样本学习和广义零样本学习两种不同设置下分别进行实验, 并与SAE、ESZSL、ALE、CADA-VAE、SE、F-VAEGAN-D2、QFSL以及DMaP进行对比(前5种模型为归纳式学习, 后3种模型为直推式学习).

表1给出了在传统零样本学习设置下的对比结果, 其中ESZSL和ALE的类平均准确度从文献[21]中引用, 其他模型的类平均准确度从对应的原文中取得. 由表1可以看出: USCAE在AwA2数据集上达到了93.91%, 比SAE高了9.21%, 比对比方法中最高的DMaP高了3.42%; 在CUB数据集上达到了68.09%, 比SAE高了6.69%, 略低于F-VAEGAN-D2和QFSL.

表1 TZSL设置下类平均准确度对比 %

模型	AwA2	CUB
SAE ^[11]	84.7	61.4
ESZSL ^[21]	75.6	55.1
ALE ^[21]	80.3	53.2
CADA-VAE ^[16]	—	—
SE ^[17]	80.8	60.3
F-VAEGAN-D2 ^[18]	—	71.1
QFSL ^[14]	84.8	69.7
DMaP ^[15]	90.49	67.69
USCAE	93.91	68.09

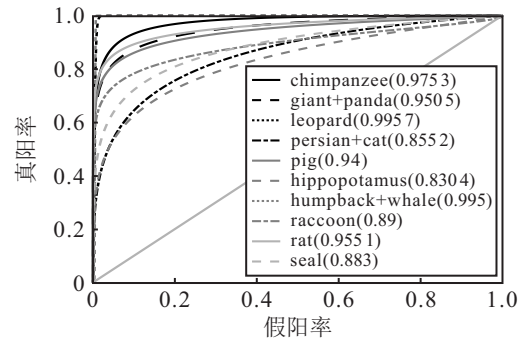
表2给出了在广义零样本学习设置下的对比结果,其中SAE、ESZSL和ALE的谐调平均准确度从文献[21]中引用,其他模型的谐调平均准确度从对应的原文中取得.由表2可以看出,与SAE、ESZSL、ALE、CADA-VAE和SE等归纳式学习模型相比,USCAE在两个数据集上均取得了最高的谐调平均准确度,这是由于USCAE在训练模型时利用目标域样本信息对模型进行了约束,从而可以有效缓解领域偏移问题.

表2 GZSL设置下谐调平均准确度对比 %

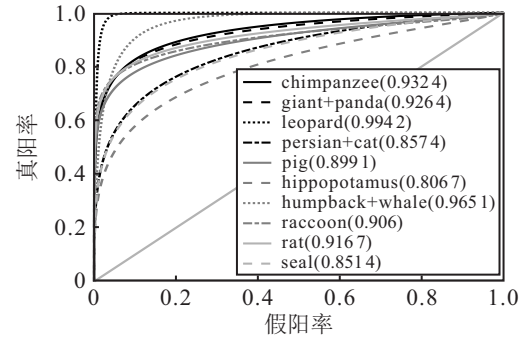
模型	AwA2			CUB		
	Acc ^s	Acc ^u	H	Acc ^s	Acc ^u	H
SAE ^[21]	82.2	1.1	2.2	54.0	7.8	13.6
ESZSL ^[21]	77.8	5.9	11.0	63.8	12.6	21.0
ALE ^[21]	81.8	14.0	23.9	62.8	23.7	34.4
CADA-VAE ^[16]	75.0	55.8	63.9	53.5	51.6	52.4
SE ^[17]	68.1	58.3	62.8	53.3	41.5	46.7
F-VAEGAN-D2 ^[18]	—	—	—	65.1	61.4	63.2
QFSL ^[14]	93.1	66.2	77.4	74.9	71.5	73.2
DMaP ^[15]	—	—	—	—	—	—
USCAE	93.44	77.04	84.45	77.56	55.86	64.94

由表1和表2的实验结果可以看出,所有模型在AwA2数据集上的准确度均高于在CUB数据集上的准确度,主要原因是:AwA2数据集由多种动物的样本组成,不同物种原型之间的差异较大,使得训练出的模型在目标域中具有较强的泛化能力;而CUB数据集是关于鸟类的细粒度数据集,不同鸟类的类原型之间虽有差别,但差别很小,导致训练的模型识别率偏低.

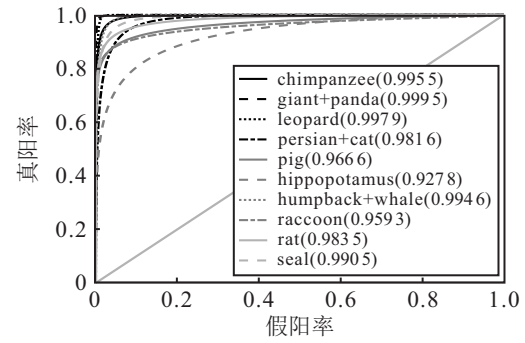
为进一步分析误判率与灵敏性的关系,图4给出了在传统零样本学习设置下,SAE、ESZSL、DMaP和USCAE在AwA2数据集上的ROC曲线与AUC值.由图4可以看出,USCAE在10个测试类上的ROC曲线均靠近左上方且其AUC值均高于SAE、ESZSL和DMaP,表明USCAE在各个测试类上均具有良好的分类性能.



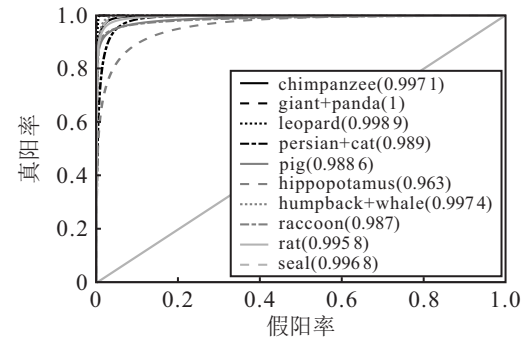
(a) SAE模型



(b) ESZSL模型



(c) DMaP模型



(d) USCAE模型

图4 AwA2数据集的ROC曲线与AUC值对比(TZSL)

3 结论

在零样本图像分类中,由于测试集包含训练集没有的类别,将训练集上学习到的知识直接用于测试集容易出现严重的领域偏移问题.为此,本文提出了一种基于语义约束语义自编码的零样本图像分类模型USCAE,主要思路为:利用语义自编码器学习从视觉空间到语义空间的映射,在训练过程中利用未知类样本的聚类视觉中心和未知类语义类原型在语义空间中的分布对齐,对语义自编码器的训练进行约束.在

传统零样本学习和广义零样本学习设置下的实验结果验证了USCAE模型具有较好的分类性能。

参考文献(References)

- [1] 柳长源, 王琪, 毕晓君. 多目标小尺度车辆目标检测方法[J]. 控制与决策, 2021, 36(11): 2707-2712.
(Liu C Y, Wang Q, Bi X J. Multi-target and small-scale vehicle target detection method[J]. Control and Decision, 2021, 36(11): 2707-2712.)
- [2] Jiang D H, Zhang S, Dai L, et al. Multi-scale generative adversarial network for image super-resolution[J]. Soft Computing, 2022, 26(8): 3631-3641.
- [3] Zhang L B, Ma J. Salient object detection based on progressively supervised learning for remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(11): 9682-9696.
- [4] Larochelle H, Erhan D, Bengio Y. Zero-data learning of new tasks[C]. Proceedings of International Conference on Artificial Intelligence. Chicago: AAAI Press, 2008: 646-651.
- [5] Ji Z, Sun Y X, Yu Y L, et al. Attribute-guided network for cross-modal zero-shot hashing[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31(1): 321-330.
- [6] Biswas S, Annadani Y. Preserving semantic relations for zero-shot learning[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 7603-7612.
- [7] 姜志彬, 潘兴广, 周洁, 等. 基于DLSR的归纳式迁移学习[J]. 控制与决策, 2021, 36(12): 2982-2990.
(Jiang Z B, Pan X G, Zhou J, et al. DLSR based inductive transfer learning method[J]. Control and Decision, 2021, 36(12): 2982-2990.)
- [8] 赵静, 王弦, 王奔, 等. 基于神经网络的多类别目标识别[J]. 控制与决策, 2020, 35(8): 2037-2041.
(Zhao J, Wang X, Wang B, et al. Multi-category target recognition based on neural network[J]. Control and Decision, 2020, 35(8): 2037-2041.)
- [9] Wang X S, Li Q Y, Gong P, et al. Zero-shot learning based on multitask extended attribute groups[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2021, 51(3): 2003-2011.
- [10] Wang X S, Huang W W, Cheng Y H, et al. Multisource domain attribute adaptation based on adaptive multikernel alignment learning[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2020, 50(5): 1897-1908.
- [11] Kodirov E, Xiang T, Gong S G. Semantic autoencoder for zero-shot learning[C]. IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, 2017: 4447-4456.
- [12] Romera-Paredes B, Torr P H S. An embarrassingly simple approach to zero-shot learning[C]. Proceedings of International Conference on Machine Learning. Switzerland: Springer Cham, 2015: 11-30.
- [13] Akata Z, Perronnin F, Harchaoui Z, et al. Label-embedding for attribute-based classification[C]. IEEE Conference on Computer Vision and Pattern Recognition. Portland, 2013: 819-826.
- [14] Song J, Shen C C, Yang Y Z, et al. Transductive unbiased embedding for zero-shot learning[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 1024-1033.
- [15] Li Y N, Wang D H, Hu H H, et al. Zero-shot recognition using dual visual-semantic mapping paths[C]. IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, 2017: 5207-5215.
- [16] Schönfeld E, Ebrahimi S, Sinha S, et al. Generalized zero-shot and few-shot learning via aligned variational autoencoders[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, 2019: 8239-8247.
- [17] Verma V K, Arora G, Mishra A, et al. Generalized zero-shot learning via synthesized examples[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 4281-4289.
- [18] Xian Y Q, Sharma S, Schiele B, et al. F-VAEGAN-D2: A feature generating framework for any-shot learning[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, 2019: 10267-10276.
- [19] Hao G S, Lim M H, Ong Y S, et al. Domination landscape in evolutionary algorithms and its applications[J]. Soft Computing, 2019, 23(11): 3563-3570.
- [20] 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.
(Ji Z, Wang H R, Yu Y L, et al. A decadal survey of zero-shot image classification[J]. Scientia Sinica: Informationis, 2019, 49(10): 1299-1320.)
- [21] Xian Y Q, Lampert C H, Schiele B, et al. Zero-shot learning—A comprehensive evaluation of the good, the bad and the ugly[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(9): 2251-2265.

作者简介

王雪松(1974—), 女, 教授, 博士生导师, 从事人工智能、机器学习及模式识别等研究, E-mail: wangxuesongcumt@163.com;

张淳(1997—), 女, 硕士生, 从事零样本学习的研究, E-mail: zctracy1997@163.com;

程玉虎(1973—), 男, 教授, 博士生导师, 从事机器学习、模式识别与智能系统等研究, E-mail: chengyuhu@163.com.