

基于特征增强模型的广义零样本学习

梁嘉豪

广东工业大学自动化学院, 广东 广州

收稿日期: 2024年3月5日; 录用日期: 2024年4月5日; 发布日期: 2024年4月16日

摘要

传统零样本学习旨在通过训练模型实现对未知类别的样本的精确分类, 使模型具备对新任务和新环境的适应能力。但广义零样本学习的目标更为艰巨, 它不仅要求模型能够辨识并分类未知类别的样本, 还需确保对已知类别的样本也能准确无误地归入其对应的类别。由于在实际的训练过程中, 我们仅能获得已知类别的样本, 这使得零样本学习在分类任务中面临巨大的挑战。为了克服这一难题, 我们创新性地提出了特征增强模型(Feature Enhancement Model, 简称FE)。该模型不仅具备生成高质量未知类别样本的能力, 以弥补训练样本的不足, 而且能够构建每个样本的虚拟语义信息。此外, FE模型还配备了特征过滤模块, 用于筛选出每个样本的核心特征。最终, 模型将这些核心特征、样本本身以及虚拟语义信息相结合, 作为最终的特征进行分类。这种方法通过凸显每类样本的独特性, 有效地提升了分类的准确性和性能。

关键词

传统零样本学习, 广义零样本学习, 特征增强模型

Feature Enhancement Model for Generalized Zero-Shot Learning

Jiahao Liang

School of Automation, Guangdong University of Technology, Guangzhou Guangdong

Received: Mar. 5th, 2024; accepted: Apr. 5th, 2024; published: Apr. 16th, 2024

Abstract

Traditional zero-shot learning aims to achieve accurate classification of unseen class samples by training the model, so that the model can adapt to new tasks and new environments. However, the goal of generalized zero-shot learning is more difficult. It not only requires the model to identify and classify samples of unseen classes, but also needs to ensure that the samples of seen classes

can be accurately classified into their corresponding classes. Because in the actual training process, we can only obtain samples of seen classes, which makes zero-shot learning in the classification task faces a huge challenge. To overcome this problem, we proposed an innovative feature enhancement model (FE). This model not only has the ability to generate high-quality unseen class samples to make up for the deficiency of training samples, but also can construct virtual semantic descriptions of each sample. In addition, the FE model is equipped with a feature filtering module to screen out the core features of each sample. Finally, the model combines these core features, the sample itself and the virtual semantic descriptions as the final features for classification. By highlighting the uniqueness of each type of sample, this method effectively improves the accuracy and performance of classification.

Keywords

Traditional Zero-Shot Learning, Generalized Zero-Shot Learning, Feature Enhancement Model

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

由于深度学习的蓬勃发展，图像分类取得了令人兴奋的高准确率。然而，它需要大量的标记数据作为支持，而且标记不仅费时费力，而且成本高昂。最重要的是，在某些情况下，根本没有标记数据，所以这对我们来说是不现实的。人们可以根据他们以前见过的东西，结合对新事物的描述，想象出新事物的样子。一个人不知道长颈鹿长什么样，但如果他以前见过豹子和大象，一旦有人告诉他长颈鹿有豹子和大象的一些特征，他就能想象出长颈鹿的样子。机器为什么不能呢？受此启发，零样本学习在2009年应运而生[1]。它先对已知类别的样本进行训练，然后用训练好的模型对未知类别的样本进行分类。“豹”和“象”对应于已知类，而“长颈鹿”对应于未知类。其实现的关键原因是利用语义描述将知识从已知类转移到未知类。2016年提出广义零样本学习[2]。与传统的零样本学习类似，已知类用于训练模型。不同之处在于在测试阶段使用已知类和未知类。

在过去，使用嵌入式方法进行样本分类的零样本学习很流行[3]。Xu等[4]使用语义空间作为嵌入空间，利用流形正则化来保留视觉特征和语义特征的几何流形结构。Akata等[5]提出了一种基于属性的分类与标签嵌入(ALE)方法来比较标签嵌入与图像的兼容性。他们把分类问题看作是标签嵌入问题。Shen等[6]认为从视觉空间到语义空间的转换会有偏差，并且存在噪声类嵌入，因此语义空间不适合分类。他们提出在公共汉明空间中嵌入样本和类，以缓解视觉和语义之间的偏差，从而实现更准确的分类。

近年来，生成对抗网络在零样本学习中得到了广泛的应用[7]。由于它能在一定程度上解决域转移问题，所以在分类中取得了很好的效果[8]。Sun等[9]提出了一种合作耦合生成网络(cooperative coupled generative networks, CCGN)模型，该模型以深度神经网络作为GAN的基本结构，并利用残差结构减少合成过程中的信息损失。Zhang等[10]提出了一种由随机属性选择(RAS)和条件生成对抗网络组成的混合模型。Han等[11]认为GAN生成的样本是欠约束的，根本不属于任何有意义的对象类。为了解决这一问题，他们提出了一个并行推理网络的特征生成框架。Li等人[12]认为使用手工语义生成的特征的多样性和可分离性较差。他们提出了一种语义增强方法，该方法将随机噪声结合到GAN中以产生更多的判别样本。

2. 本文方法

2.1. 零样本学习定义

我们假设总共有 K 类，包括 K_1 已知类和 K_2 未知类。它们分别表示为集合 \mathbb{L}_s 和 \mathbb{L}_u 。然后，类集可以表示为 $\mathbb{L} = \mathbb{L}_s \cup \mathbb{L}_u = \{l_i\}_{i=1}^{K_1+K_2}$ ， l_i 是第 i 类的标签。 $\mathbb{L}_s \cap \mathbb{L}_u = \emptyset$ 。据此，我们可以得到语义特征集 $\mathbb{S} = \mathbb{S}_s \cup \mathbb{S}_u$ 。具体来说已知类的语义特征集是 $\mathbb{S}_s = \{a_s^{(i)}\}_{i=1}^{K_1}$ ，其中 $a_s^{(i)}$ 是第 i 个已知类的 p 维语义特征。类似地，未知类的语义特征集是 $\mathbb{S}_u = \{a_u^{(i)}\}_{i=1}^{K_2}$ ，其中 $a_u^{(i)}$ 是第 i 个未知类的 p 维语义特征。所看到的类样本记为 $X^s = [x_i^s]_{i=1}^N \in \mathbb{R}^{N \times d}$ ，其中 N 为样本数量， x_i^s 为第 i 个样本的 d 维特征。另外，将每个样本 x_i^s 对应的标签记为 $y_i^s \in \mathbb{L}_s$ ，对应的语义描述记为 $\hat{a}_i^s \in \mathbb{S}_s$ 。此外，假设未知类有 M 个样本 $X^u = [x_i^u]_{i=1}^M \in \mathbb{R}^{M \times d}$ ，而 x_i^u 是第 i 个样本的 d 维特征。对应的语义表示和标签分别是 $\hat{a}_i^u \in \mathbb{S}_u$ 和 $y_i^u \in \mathbb{L}_u$ ，但是 x_i^u 在训练阶段不可用。在测试时，传统的零样本学习学习一个分类器 $f_{zsl}: x_i^u \rightarrow \mathbb{L}_u$ ，希望将未知类的每张图像分类到正确的类中。而广义零样本学习需要将已知类和未知类的图像都分类到相应的类中。假设用于测试的两个类中的一个样本为 x^s ，则 GZSL 的任务表示为 $f_{gzsl}: \{x^s, x_i^u\} \rightarrow \mathbb{L}_s \cup \mathbb{L}_u$ 。

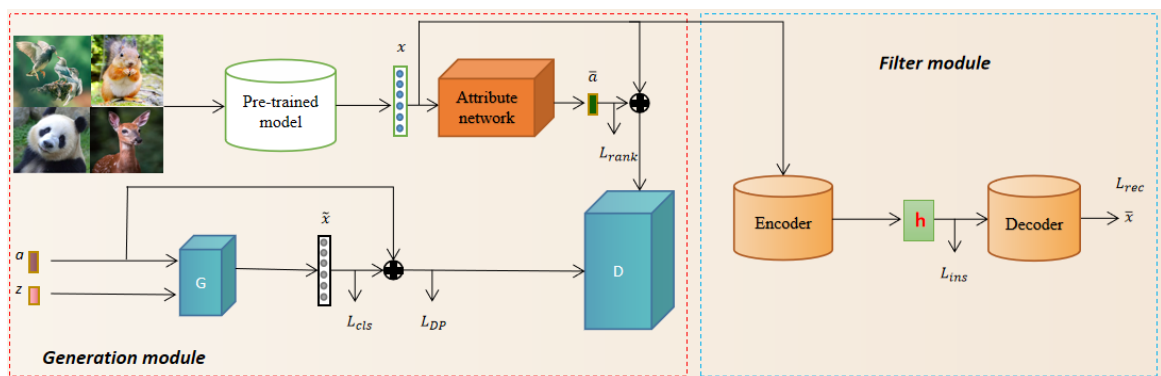


Figure 1. The overall framework of feature enhancement model

图 1. 特征增强模型的总体框架

2.2. 模型结构

如图 1 所示，它由生成模块和过滤模块组成，两个模块分别用红色和蓝色虚线隔开。与传统生成对抗网络不同的是，我们将生成器生成的样本和该类的真实语义配对，同时将语义网络生成的虚拟语义和真实图片样本进行配对，再一起放进鉴别器进行判断真伪。这种做法相比于传统做法不同的是不仅可以生成有质量的图片样本，而且可以生成以假乱真的语义描述。在这个过程中，我们使用排列损失对虚拟语义描述进行约束，用预训练的分类器对生成的图片样本进行约束，同时用依赖损失使生成的图片样本依赖于真实语义描述。过滤模块的作用是提取图片样本的核心信息，具体做法是将编码器对图片样本的编码部分进行实例约束，使其从一定程度上反映该类的特点。在训练最终分类器时，我们使用图片样本、虚拟语义描述和图片样本核心信息的联合作为最终的特征。

2.3. 生成模块

一般在零样本学习生成对抗网络中，无论是与真实图片样本配对，还是与生成样本配对的都是真实的语义描述，我们提出的特征增强模型中的生成模块则与之不同，我们将生成器生成的样本和该类的真实语义配对，同时将语义网络生成的虚拟语义和真实图片样本进行配对，再一起放进鉴别器进行判断真伪。但是，这种做法中虚拟语义和生成样本是极度缺乏约束的。

为此，我们通过使用已知类样本预训练的分类器对生成样本进行分类，得到其分类损失，其损失函数为：

$$L_{cls} = -\mathbb{E}[\log P(y|\tilde{x};\theta_c)] \quad (1)$$

其中， y 是 \tilde{x} 的类标签真值， $P(y|\tilde{x};\theta_c)$ 表示 \tilde{x} 被预测为 y 的概率，约束分类损失分类器由 θ_c 参数化的 softmax 分类器，该分类器在已知类训练集上进行预训练。

另外，为了让鉴别器分辨不出生成样本是伪造的，我们利用依赖损失使生成样本高度依赖该类的语义描述，从而进一步改善生成样本的质量。依赖损失测量两个随机变量之间的相关性，并定义为在给定一个随机变量的情况下，另一个随机变量的不确定性减少。具体来说，依赖损失的目标定义为：

$$\mathcal{L}_{DP} = \mathbb{E}_{a \sim P_{S_s}} \mathbb{E}_{\tilde{x} \sim P_G} [-(H(\tilde{x}) - H(\tilde{x}|a))] \quad (2)$$

其中， P_{S_s} 是已知类的语义描述符的经验边际分布， P_G 是生成图片样本的条件分布， $H(\tilde{x})$ 是生成图片样本的熵， $H(\tilde{x}|a)$ 在给定语义描述符 a 的情况下，生成图片样本 \tilde{x} 的条件熵。

以上是对生成样本设计的约束，力求生成样本足够逼真。接着我们需要对真实图片样本经过语义网络得到的虚拟语义描述进行约束，以使得虚拟语义描述达到以假乱真的目的。我们采用排名损失，该损失目标要求根据语义描述符空间中的点积相似性，将真值语义描述符排名高于任何其他语义描述符。具体而言，排名损失的目标定义为：

$$\mathcal{L}_{rank} = \mathbb{E}_{(x,a) \sim P_{X_{tr} \times S_s}} \mathbb{E}_{a' \neq a} \left[\max\left(0, \Delta - a^T N(x) + (a')^T N(x)\right) \right] \quad (3)$$

其中， $P_{X_{tr} \times S_s}$ 是已知类训练集的视觉和语义描述符的经验联合分布， $X_{tr} \in X^s$ ， $\Delta > 0$ 是裕度的超参数， a 是真值语义描述符， a' 是从其他类中随机选择的语义描述符，虚拟语义描述 $\bar{a} = N(x)$ 。

总体来说，生成器、鉴别器和语义网络的联合损失函数为：

$$V = \mathbb{E}_{x \sim P_{X_{tr}}} \left[\underbrace{\mathbb{E}_{\bar{a} \sim P_N(\cdot|x)} [\log D(x, \bar{a})]}_{\log D(x, N(x))} \right] + \mathbb{E}_{z \sim P_Z} \mathbb{E}_{a \sim P_{S_s}} \left[\underbrace{\mathbb{E}_{\tilde{x} \sim P_G(\cdot|a,z)} [\log(1 - D(\tilde{x}, a))]}_{\log(1 - D(G(a,z), a))} \right] \quad (4)$$

其中， P_N 是虚拟语义描述符的分布， P_Z 为高斯噪声分布。

目前为止，生成模块不仅能够生成出高质量图片样本，而且能够输出逼真的语义描述。下一步需要对图片样本进行过滤，提取出每一类的核心信息。

2.4. 过滤模块

过滤模块的主体为编码器和译码器，我们将真实图片样本输入编码器，得到编码部分，详细公式如下：

$$h = E_{nc}(X) \quad (5)$$

接着，我们会对编码部分 h 进行实例监督的约束，此做法与第三章提及的想法一样，具体公式如下：

$$L_{ms} = -\log \frac{\exp(z_i^T \cdot z_i^+ / \varepsilon_0)}{\exp(z_i^T \cdot z_i^+ / \varepsilon_0) + \sum_{k=3}^{k=1} \exp(z_i^T \cdot z_i^k / \varepsilon_0)} \quad (6)$$

它的目的是试图最大化 z_i 与其同类别的正样本 z_i^+ 之间的相似性，最小化其与其不同类别的负样本 z_i^k 之间的相似性。

最后，我们也会将编码部分输入到解码器，力求解码出原本的图片样本。我们通过重构损失来约束解码部分，具体公式如下：

$$L_{rec} = \|X^s - D_{ec}(h_F^2)\| \quad (7)$$

最终，我们不仅可以生成未知类样本来补全训练集，而且可以生成虚拟的语义描述，外加上每个图片样本的核心信息。

3. 实验评估与分析

3.1. 实验细节

提出的 FE 通过 PyTorch 实现，并在具有两个 NVIDIA GeForce RTX 2080 Ti gpu, AMD Ryzen 9 5950x 3.52 GHz CPU 和 126 G RAM 的服务器上运行。

对于语义网络的具体结构，我们先通过一个线性层，输入维度和输出维度分别为该类的图片样本维度和语义描述维度，然后通过一个 Tanh 激活函数，最终构成完整的语义网络。

我们将特征增强模型(FE)与多种方法进行了比较，其中大多数方法都是基于 GAN 的，其中许多是近年来的优秀模型。

我们在三个数据集上验证 FE 的有效性，分别是 FLO、CUB 和 SUN。FLO 数据集由 8189 张花卉图片组成，是一个细粒度数据集，包含 102 个花卉类别，其中 82 个已知类包含 7034 个样本，20 个未知类包含 1155 个样本。对于已知类，1640 个样本用于训练，5394 个样本用于测试。未知类的样本都在测试阶段使用；CUB 是一个细粒度数据集。样本总数为 11,788 个。具体来说，150 个已知类包含 8821 个样本，50 个未知类包含 2967 个样本。对于已知类，7057 个样本用于训练，1764 个样本用于测试。未知类的样本都在测试阶段使用；SUN 也是一个细粒度的数据集。样本总数为 14,340 个。具体来说，645 个已知类包含 12,900 个样本，以及 72 个未知类包含 1440 个样本。对于已知类，10,320 个样本用于训练，2580 个样本用于测试。未知类的样本都在测试阶段使用。

另外，我们使用每个类的 TOP-1 精度作为我们的评估指标。对于传统零样本学习，我们只评估未知类，用 T1 表示。对于广义零样本学习，我们通过谐波平均 $H = \frac{2 \cdot S \cdot U}{S + U}$ 来验证模型的精度，其中 S 和 U 分别表示已知类和未知类的精度。

3.2. 实验结果

3.2.1. 传统零样本学习

传统零样本学习方法的实验结果如表 1 所示，AVG 表示每种方法在三个数据集上的平均性能，最好的结果已加粗表示。我们可以得出以下结论：

1) 经过比较，FE 在三个数据集上都取得了最好的结果，平均值也理所当然的实现最好的准确率。在 FLO、CUB、SUN 上分别领先第二名 0.3、0.3 和 0.1 个百分点。究其原因，无疑是我们形成最终分类特征的合理性，不仅有原始图片样本，而且加入该图片样本经过语义网络和编码器分别得出的虚拟语义描述和核心信息。得益于此，才能在各方面都取得优胜的结果。

2) 对于 FLO 数据集，它的语义描述维度为 1024，是一个相当大的维度。对于我们的特征增强模型 (FE) 来说也许是一个难点，因为我们的模型需要生成虚拟语义描述，越高的维度就意味着越难以生成出高质量的语义描述。但是，由结果可以看出，特征增强模型克服了此困难，取得了良好的分类性能。

3) 当我们将映射方法与基于 GAN 的方法(例如 CCGN 和 FE)进行比较时，可以明显看出基于 GAN 的方法表现更优。这进一步证明了将 GAN 生成的未知类别的虚拟样本融入训练过程的合理性。

4) 相较于基于 GAN 的方法，如 LisGAN、CCGN 和 MGA-GAN，FE 表现出更出色的性能，这得益于其精心设计。

3.2.2. 广义零样本学习

对于更具挑战性的广义零样本学习，我们在表 2 中展示了实验结果，最好的结果已加粗表示。我们可以得出以下结论：

1) 我们提出的特征增强模型(FE)在 FLO 和 CUB 数据集上获得第一名，分别比第二名高 0.6 和 0.3 个百分点。

2) FE 在 SUN 数据集上的表现明显落后了。比第一名差了 4 个百分点。但是在 FLO 和 CUB 训练集上都比它好，所以总体来说 FE 比 MM-WAE 的表现略优。这可能是由于 SUN 上的类太多，每个类的样本比较少，在语义网络的训练时就缺乏足够的训练样本，造成了对 FE 模型的巨大影响。

3) 通过比较大多数基于生成对抗网络的方法和特征增强模型(FE)的结果，发现 FE 总体表现更好，原因就是 we 补充了原始图片样本的不足，增加了可以反映该类不同于其他类别的特征，即虚拟语义描述和核心信息。

Table 1. Experimental results of traditional zero-shot learning

表 1. 传统零样本学习实验结果

	FLO	CUB	SUN	AVG
ALE [5]	48.5	54.9	58.1	53.8
f-CLSWGAN [7]	67.2	57.3	60.8	61.8
LisGAN [13]	69.6	58.8	61.7	63.4
CCGN [9]	-	53.5	62.8	58.2
LsrGAN [14]	-	60.3	62.5	61.4
Inf-FG [11]	67.3	58	61.1	62.1
MGA-GAN [8]	70.6	58.9	61.8	63.8
MM-WAE [16]	68.5	55	58.2	60.6
FE	70.9	60.6	62.9	64.8

Table 2. Experimental results of generalized zero-shot learning

表 2. 广义零样本学习实验结果

	FLO	CUB	SUN
ALE [5]	21.9	34.4	26.3
f-CLSWGAN [7]	65.6	49.7	39.4
LisGAN [13]	68.3	51.6	40.2
Inf-FG [11]	66.7	50.8	40.5
NereNet [15]	-	53.6	41.6
MGA-GAN [8]	68.2	50.8	40.1
MM-WAE [16]	66.5	53.2	46.7
FE	68.9	53.9	42.7

3.2.3. 消融

为了验证每个部分在特征增强模型(FE)中的作用，我们在传统零样本学习下设计了几个变体。具体来说，FE-latent 意味着最终的分类特征不带有核心信息；FE-att 意味着最终的分类特征不带有虚拟语义

描述。根据特征增强模型(FE)的结果,我们可以从表3中得出以下结论:

1) 与 FE 相比, FE-latent 的实验结果 T1。结果表明,提取图片样本的核心信息融入最后的分类特征可以有效提高分类性能。

2) 与 FE 相比, FE-att 的实验结果 T1 略有降低。可以得出采用语义网络生成虚拟语义描述从而融入分类特征对分类性能起积极作用。

3) 与 FE-att 相比, FE-latent 的实验结果 T1 降低的幅度较大。这说明虽然它们对分类准确率都有正面的影响,但是提取图片样本的核心信息比生成虚拟语义描述的作用大。

Table 3. Comparison results of DI-GAN with its variants

表 3. FE 与其变体的比较结果

	FLO	CUB	SUN
FE-latent	67.1	58.9	61.9
FE-att	69.3	60.1	62.3
FE	70.9	60.6	62.9

4. 结束语

本文提出了一种用于广义零样本学习的特征增强模型(FE),一方面,我们在生成模块中不仅生成能够鱼目混珠的图片样本,而且虚拟语义描述的出现也为模型分类准确率的优越性奠定了基础;另一方面,我们采用实例监督对编码器的输出进行约束,使其过滤掉该图片样本的无关信息,而留下核心信息。比如一张熊猫的图片,熊猫的黑眼圈就是必须关注的特点。在三个数据集上的实验证明了 FE 在优越性。

参考文献

- [1] Lampert, C.H., Nickisch, H., Harmeling, S. (2009) Learning to Detect Unseen Object Classes by between-Class Attribute Transfer. 2009 *IEEE Conference on Computer Vision and Pattern Recognition*, Miami, 20-25 June 2009, 951-958. <https://doi.org/10.1109/CVPRW.2009.5206594>
- [2] Chao, W.L., Changpinyo, S., Gong, B., et al. (2016) An Empirical Study and Analysis of Generalized Zero-Shot Learning for Object Recognition In: Leibe, B., Matas, J., Sebe, N. and Welling, M., Eds., *Computer Vision—ECCV 2016. Lecture Notes in Computer Science*, Vol. 9906, Springer, Cham, 52-68. https://doi.org/10.1007/978-3-319-46475-6_4
- [3] Akata, Z., Perronnin, F., Harchaoui, Z., et al. (2015) Label-Embedding For Image Classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **38**, 1425-1438. <https://doi.org/10.1109/TPAMI.2015.2487986>
- [4] Xu, X., Shen, F., Yang, Y., et al. (2017) Matrix Tri-Factorization With Manifold Regularizations For Zero-Shot Learning. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 3798-3807. <https://doi.org/10.1109/CVPR.2017.217>
- [5] Akata, Z., Perronnin, F., Harchaoui, Z., et al. (2013) Attribute-Based Classification with Label-Embedding. NIPS 2013 Workshop on Output Representation Learning.
- [6] Shen, F., Zhou, X., Yu, J., et al. (2019) Scalable Zero-Shot Learning via Binary Visual-Semantic Embeddings. *IEEE Transactions on Image Processing*, **28**, 3662-3674. <https://doi.org/10.1109/TIP.2019.2899987>
- [7] Xian, Y., Lorenz, T., Schiele, B., et al. (2018) Feature Generating Networks for Zero-Shot Learning. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, San Juan, 18-23 June 2018, 5542-5551.
- [8] Xie, G.S., Zhang, Z., Liu, G., et al. (2021) Generalized Zero-Shot Learning with Multiple Graph Adaptive Generative Networks. *IEEE Transactions on Neural Networks and Learning Systems*, **33**, 2903-2915. <https://doi.org/10.1109/TNNLS.2020.3046924>
- [9] Sun, L., Song, J., Wang, Y., et al. (2020) Cooperative Coupled Generative Networks for Generalized Zero-Shot Learning. *IEEE Access*, **8**, 119287-119299. <https://doi.org/10.1109/ACCESS.2020.3000347>
- [10] Zhang, H., Long, Y., Liu, L., et al. (2019) Adversarial Unseen Visual Feature Synthesis for Zero-Shot Learning. *Neurocomputing*, **329**, 12-20. <https://doi.org/10.1016/j.neucom.2018.10.043>

-
- [11] Han, Z., Fu, Z., Li, G., *et al.* (2021) Inference Guided Feature Generation for Generalized Zero-Shot Learning. *Neuro-computing*, **430**, 150-158. <https://doi.org/10.1016/j.neucom.2020.10.080>
- [12] Li, Z., Chen, Q., Liu, Q. (2021) Augmented Semantic Feature Based Generative Network for Generalized Zero-Shot Learning. *Neural Networks*, **143**, 1-11. <https://doi.org/10.1016/j.neunet.2021.04.014>
- [13] Li, J., Jing, M., Lu, K., *et al.* (2019) Leveraging the Invariant Side of Generative Zero-Shot Learning. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 15-20 June 2019, 7402-7411. <https://doi.org/10.1109/CVPR.2019.00758>
- [14] Vyas, M.R., Venkateswara, H., Panchanathan, S. (2020) Leveraging Seen and Unseen Semantic Relationships For Generative Zero-Shot Learning. In: Vedaldi, A., Bischof, H., Brox, T. and Frahm, J.M., Eds., *Computer Vision—ECCV 2020. Lecture Notes in Computer Science*, Vol. 12375, Springer, Cham, 70-86. https://doi.org/10.1007/978-3-030-58577-8_5
- [15] Liu, J., Bai, H., Zhang, H., *et al.* (2021) Near-Real Feature Generative Network for Generalized Zero-Shot Learning. 2021 *IEEE International Conference on Multimedia and Expo (ICME)*, Shenzhen, 5-9 July 2021, 1-6. <https://doi.org/10.1109/ICME51207.2021.9428229>
- [16] Chen, X., Li, J., Lan, X., *et al.* (2020) Generalized Zero-Shot Learning Via Multi-Modal Aggregated Posterior Aligning Neural Network. *IEEE Transactions on Multimedia*, **24**, 177-187. <https://doi.org/10.1109/TMM.2020.3047546>