

[文章编号] 1003-4684(2022)01-0034-06

多尺度通道注意力机制的小样本图像分类算法

王 奇, 靳华中, 李文莹, 李晴晴

(湖北工业大学计算机学院, 湖北 武汉 430068)

[摘 要] 为了提升关系网络图像分类的准确度,在网络中引入多尺度通道注意力机制,提出了一种新的小样本图像分类算法。由于多尺度通道注意力机制能够关注样本特征空间的重要信息,该方法能够提取图像更丰富的多尺度特征,并通过关系度量,改善了分类结果。实验结果表明,在 MiniImageNet 和 Omniglot 数据集上,该算法对图像分类精度有明显的提高。

[关键词] 小样本学习; 元学习; 关系网络; 注意力机制

[中图分类号] TP311 [文献标识码] A

随着大数据时代的到来,基于数据驱动的深度学习模型在图像分类任务中取得了巨大成功^[1]。一般来说,深度学习的成功可归因于三个关键因素:充足的计算资源、复杂的神经网络和大规模数据集。然而,许多现实的应用场景中,例如在医学、军事和金融领域,由于涉及隐私安全问题和较高的人力成本等因素,无法得到足够多的训练样本。训练模型时没有足够多的有监督样本,容易产生过拟合现象,即模型在训练样本上表现良好,在测试集上的泛化效果不佳^[2]。小样本学习^[3]使用远小于深度学习所需要的数据样本量,达到接近甚至超越大数据深度学习的效果。为了让小样本学习进一步提高模型泛化性能,国内外学者对其进行大量的研究。阿里巴巴智能服务事业部团队^[4]根据内部实现机制将小样本学习方法分为基于模型、基于度量和基于优化三类。李新叶等人^[5]和赵凯琳等人^[6]从跨任务学习知识的角度将基于元学习和基于迁移学习的方法也归为小样本学习。祝钧桃等人^[7]根据小样本分类中所采用的技术将小样本学习的解决方案分为数据增强、度量学习、外部记忆、参数优化四个策略。由于真实世界中样本采集困难或成本昂贵,小样本学习已成为当前机器学习的研究热点。

从宏观角度来看,研究小样本学习的理论和实践意义主要来自三个方面^[8]。1)小样本学习方法不依赖大规模的训练样本,从而避免了某些特定场景中数据准备的高昂成本;2)小样本学习可以缩小人

工智能和人类智能之间的差距,是开发通用人工智能的必要研究;3)小样本学习可以为一个新出现的任务实现低成本和快速的模型部署,为新任务阐明早期的规律。

在图像分类中,依据不同的建模方式,小样本图像分类算法分为卷积神经网络模型和图神经网络模型两大类^[9]。其中基于卷积神经网络模型的算法包括四种学习范式:迁移学习^[10]、元学习^[11]、对偶学习^[12]和贝叶斯学习^[13]。元学习包括基于度量的元学习^[14]、基于模型的元学习、基于优化的元学习。基于度量的元学习方法依据不同的度量方式对成对的样本进行相似度度量,根据不同的度量方式,可被分为暹罗网络^[15]、匹配网络^[16]、原型网络^[17]以及关系网络^[18]。在基于度量的元学习方法中,暹罗网络、匹配网络和原型网络的分类器采用手动设计的方法,使用欧氏距离或者余弦距离,关系网络改进了距离的度量方式,使用神经网络作为分类器来学习特征之间度量方式。

基于度量的元学习首先专注学习一个可嵌入的模块,并预先定义固定的度量方式,在此基础上关系网络用来学习一个可转移的深度度量模块以提高图像的分类精度。2018 年,sung 等人^[18]首次提出关系网络。从深度学习出发,关系网络延续基于度量的元学习的基本原理,通过设计模型学习“如何度量”样本间的距离。相比于固定距离的度量方式,关系网络通过学习一种非线性度量方式取得了更好

[收稿日期] 2021-08-02

[基金项目] 大学生创新创业训练计划项目(S201910500074)

[第一作者] 王 奇(1997-),女,山西侯马人,湖北工业大学硕士研究生,研究方向为计算机视觉

[通信作者] 靳华中(1973-),男,湖北洪湖人,湖北工业大学副教授,研究方向为计算机视觉

的分类效果。但是关系网络中的特征提取模块忽略了很多重要的样本信息,对样本的关键信息提取不充分,导致仍然存在识别准确率低的问题。通过将图像裁剪与关系网络相结合,庞振全等人^[19]构造的网络取得了较好的图像分类精度,但裁剪操作会造成样本有效信息的损失。金璐等人^[20]在关系网络中引入 inception 模块,融合多尺度特征后对红外空中目标分类取得了良好效果。王年等人^[21]通过引入自注意力机制与感受野模块提升了网络的特征表达能力和度量能力。

注意力机制^[22]能够使网络模型更加关注重要信息而减少与目标任务无关的信息。2018 年,胡杰等人提出 SE-Net^[23],将它引入 CNN 中取得良好的图像分类效果。此后,研究者提出 ECA-Net^[24]、SK-Net^[25]、ResNeSt^[26]、CBAM^[27] 等注意力机制,分别从通道域与空间域两个方面增强了注意力机制的性能。注意力机制大幅度提升模型训练的速度与任务效果,同时其即插即用的特性极大方便了模型的设计。小样本学习的训练样本较少,能提取到的信息相对有限,而使用注意力机制可以在有限的训练样本下提取到更多的有效信息。

基于上述分析,在关系网络的基础中本文引入多尺度空间与通道的注意力机制,使关系网络的嵌入模块能够学习更丰富的多尺度特征,同时自适应地对多维度的通道注意力权重进行特征重标定,提升了模型对小样本的分类能力。在 MiniImageNet^[15] 与 Omniglot^[28] 数据集上进行实验,实验表明本文方法的分类精度比原方法明显提高。

1 相关技术原理

1.1 小样本问题描述

受人类智能的启发,研究小样本学习的目的是希望模型(算法)能够像人脑一样,在学习大量的基类后,仅需要少量样本就能快速得到新类,获取新的知识,做到举一反三。

小样本图像分类流程(图 1)包括数据集处理,特征提取网络和分类器三个部分。

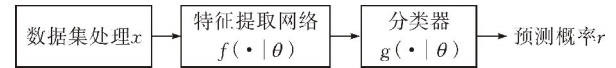


图 1 小样本图像分类流程

小样本学习的基本模型定义为 $r = g(f(x | \theta) | \omega)$,它由特征提取网络 $f(\cdot | \theta)$ 和分类器 $g(\cdot | \omega)$ 组成, θ 和 ω 分别表示 f 和 g 的参数, x 表示待识别的图像, $f(x | \theta)$ 表示对图像 x 提取的特征, r 表示对图像 x 识别的结果。

为了提取图像的有效特征,分类时需要建立特

征提取模型。对于模型,提取的图像特征应该尽可能有效地描述图像,使分类器能够更好地利用它们。可见,提高图像特征描述和提取能力,将能够获得更好的分类结果。注意力机制、记忆力机制等是常见的图像特征有效提取的技术手段。

分类器通过对特征进行相似度度量,来获取对象的不同类别。小样本图像分类中所使用的分类器大多数是在卷积神经网络的最后一层构建带有 Softmax 的全连接层,或者对提取的图像特征应用 K 近邻算法等。本文分类器采用的是前者。

1.2 关系网络

关系网络是一个灵活、通用、端到端的小样本学习框架,通过计算支持集和查询集样本图像的关系得分,对新类别的样本图像进行分类。关系网络结构见图 2,它主要由嵌入模块 f_φ 和关系模块 g_φ 组成。

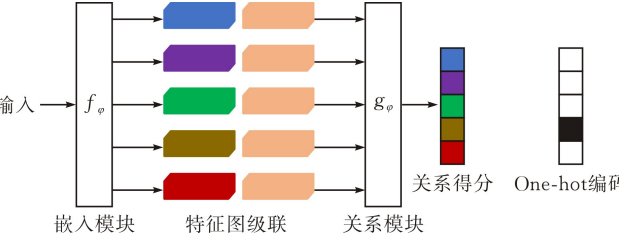


图 2 关系网络结构

嵌入模块用于提取图像样本的抽象特征,生成支持集和查询集图像的特征表示,如图 3 中 f_φ 所示,由 4 个卷积块和 2 个最大池化层组成,每个卷积块包含一个卷积核大小为 3×3 ,通道数为 64 的卷积层,一个 BatchNorm 层,一个 ReLU 层。嵌入模块能提取出样本的抽象特征,为模型提供可以用于对比的特征信息。支持集与查询集图片分别输入嵌入模块,首先提取图像特征,接着将特征图级联后送入关系模块。

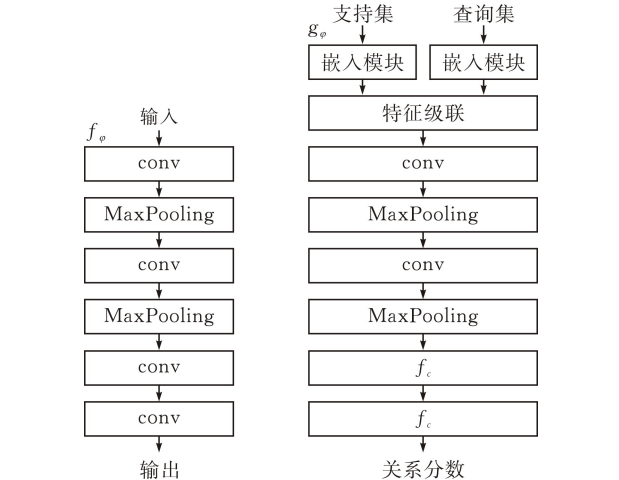


图 3 嵌入模块与关系模块

如图 3 中 g_φ 所示,关系模块由 2 个卷积块,2

个最大池化层,2个全连接层组成。该模块最终产生0-1范围内的标量,该标量表示支持集样本和测试集样本之间的相似性,为关系得分。计算该得分的公式如下:

$$r_{i,j} = g_{\varphi}(C(f_{\varphi}(x_i), f_{\varphi}(x_j))), i = 1, 2, \dots, S \quad (1)$$

其中, $r_{i,j}$ 为相似度得分即关系得分, $C(\cdot)$ 为特征级联。

训练中的损失函数使用均方误差 MSE, 如:

$$L(\varphi, \varphi) = \sum_{i=1}^m \sum_{j=1}^n (r_{i,j} - 1(y_i == y_j))^2 \quad (2)$$

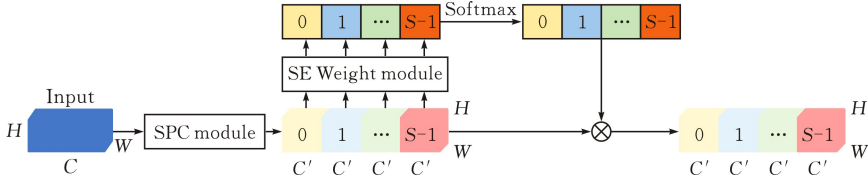


图4 金字塔切分注意力机制框架

PSA 模块包括以下四个部分:

1) 使用分割和连接模块 (SPC module) 将输入转换为通道方向上的多尺度特征图。其中, 分割和连接模块的实现步骤如下, 将输入的特征图 X 分割成 S 个部分, 每个部分用 $[X_0, X_1, \dots, X_{S-1}]$ 和通道维度 C' 表示, $C' = C/S$ 。引入群卷积的方法对这些部分分别用多尺度并行处理。

SPC 设计了新的准则来处理多尺度内核与组大小的关系:

$$G = 2^{\frac{K-1}{2}} \quad (3)$$

其中, G 是组大小, K 是内核大小。此外, 当内核大小 $K=3$ 时, $G=1$ 。经过 SPC 模块的多尺度特征图的生成函数如下:

$$F_i = \text{conv}(k_i \times k_i, G_i)(X_i), i = 1, 2, \dots, S-1 \quad (4)$$

其中, 第 i 个核的大小 $k_i = 2 \times (i+1) + 1$, 第 i 个组的大小是 $G_i = 2^{\frac{K_i-1}{2}}$ 。将得到的多尺度特征图拼接在一起:

$$F = \text{cat}([F_0, F_1, \dots, F_{S-1}]) \quad (5)$$

2) 使用 SE 权重模块提取不同比例的特征图的注意力权重, 获得通道方向的注意力向量。

$$Z_i = \text{SEWeight}(F_i), i = 1, 2, \dots, S-1 \quad (6)$$

$$Z = Z_0 \oplus Z_1 \oplus \dots \oplus Z_{S-1} \quad (7)$$

其中, \oplus 是级联运算符, Z_i 是 F_i 的注意力权重, Z 是多尺度注意力权重向量。

SE 模块是通道注意力机制中的经典模型, 包括 Squeeze 操作与 Excitation 操作。SE 模块结构见图 5。

对于任何给定的变换, 将输入 X 映射到特征映射 U , 其中 $U \in R^{H \times W \times C}$ 。特征 U 首先通过 Squeeze 操作在空间维度上被压缩为 $1 \times 1 \times C$ 向量, 即全局平均池化。Squeeze 操作的公式如下:

2 基于多尺度通道注意力机制的小样本图像分类模型

2.1 金字塔切分注意力模块

金字塔切分注意力模块 (PSA) 是一种多尺度空间与通道注意力机制^[29]。金字塔切分注意力机制框架的结构见图 4。

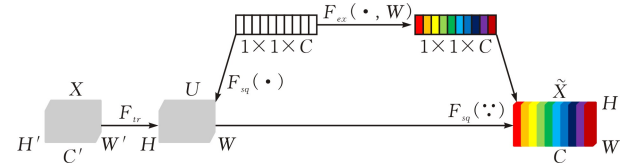


图5 SE 模块结构

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (8)$$

Squeeze 操作之后是 Excitation 操作, 该操作通过两个全连接层组成一个 Bottleneck 结构去建模通道间的相关性, 为每个通道生成一个权重值, 得到各个通道权重的集合。Excitation 操作的公式如下:

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z)) \quad (9)$$

式中, δ 为 ReLU 激活函数, W_1 、 W_2 分别为两个全连接层的参数。

最后通过 Scale 操作将得到的归一化权重加权到每个通道的特征上, 以生成 SE 模块的输出。Scale 操作的公式如下:

$$\tilde{x}_c = F_{scale}(u_c, s_c) = s_c u_c \quad (10)$$

3) 使用 Softmax 重新校准通道方向的注意力向量, 得到多尺度通道重新校准的注意力权重。

$$\text{att}_i = \text{Softmax}(Z_i) = \frac{\exp(Z_i)}{\sum_{i=0}^{S-1} \exp(Z_i)} \quad (11)$$

其中, Softmax 用于获得多尺度通道的重新校准的权重, 其包含空间上的所有位置信息和通道中的注意力权重。

$$\text{att} = \text{att}_0 \oplus \text{att}_1 \oplus \dots \oplus \text{att}_{S-1} \quad (12)$$

其中, att 表示注意力级联后的多尺度通道权重。

4) 将得到的多尺度通道的权重加权到对应的特征图 F_i 中:

$$Y_i = F_i \odot \text{att}_i, i = 1, 2, \dots, S-1 \quad (13)$$

其中, \odot 表示通道乘法。

最后得到一个多尺度特征信息更丰富的细化特征图作为输出。

out = cat([Y₀,Y₁,⋯,Y_{i-1}])

(14)

2.2 基于多尺度通道注意力机制的小样本图像分类算法

本文提出的基于多尺度通道注意力机制的小样本图像分类网络模型见图 6。

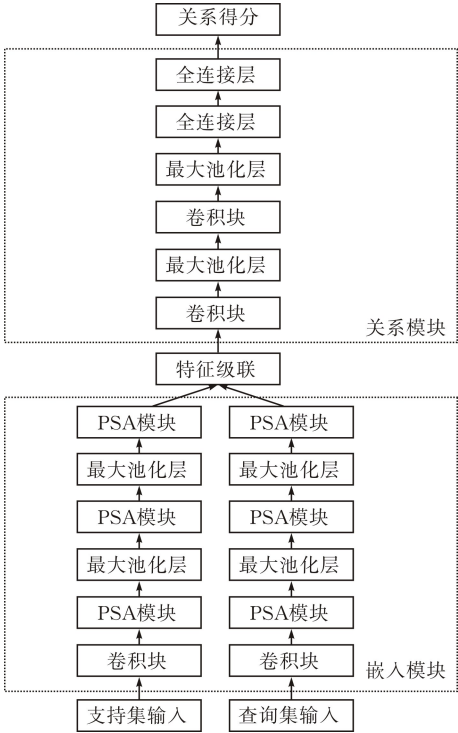


图 6 小样本图像分类网络模型

嵌入模块由一层卷积块、三层 PSA 模块以及两层最大池化层组成;关系模块由两层卷积块、两层最大池化层，两层全连接层组成。将得到的支持集图

像特征与查询集图像特征级联并输入关系模型。

PSA 模块网络结构是一个金字塔切分注意力模块^[29],其网络结构如图 7 所示。它将输入的特征图分为四个尺度,用大小不同的卷积核并行处理,其中, $K_0=3,K_1=5,K_2=7,K_3=9$ 。图中,SE 权重模块得到通道注意力先计算不同通道特征图的权重,再通过 Softmax 层对权重校准,最后得到通道权值不同的特征图。

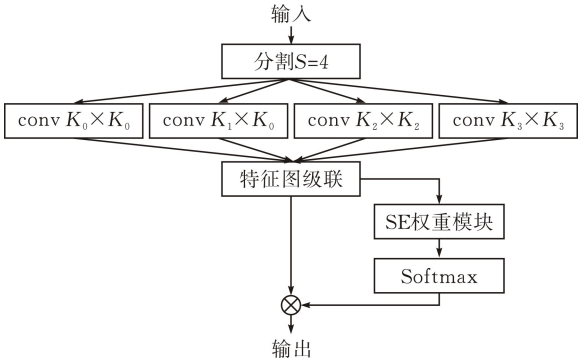


图 7 PSA 模块网络结构

本文使用 PSA 模块替换嵌入模块中后三层卷积块,从而得到改进后的嵌入模块(图 7)。

3 实验

3.1 数据集

MiniImageNet 数据集是元学习和小样本领域的基准数据集,由 google DeepMind 团队 Oriol Vinyals 等人在 ImageNet 的基础上提取并提出,数据集复杂,包含 100 类共 60 000 张彩色图片,其中每类有 600 个样本,每张图片的规格为 84×84 ,适合进行原型设计和实验研究(图 8)。



图 8 MiniImageNet 数据集图像示例

Omniglot 数据集包含来自 50 个不同字母的 1623 个不同手写字符。每个字符由 20 个不同的人通过亚马逊的 Mechanical Turk 在线绘制的。对现有数据旋转 90°,180°和 270°来扩充数据集,1200 个原始类加上来进行训练,剩余的 423 个类来进行测试。输入图像都调整到 28×28(图 9)。



图 9 Omniglot 数据集图像示例

3.2 实验设置

遵循大多数现有的小样本学习工作采用的标准设置,对于 MiniImagenet 数据集,使用 5-way 1-shot 和 5-way 5-shot 两种模式进行训练,在训练阶段随机采样并构建 500 000 个 episode,在每个训练的 episode 中,5-way 1-shot 每类包含 15 个查询图像,5-way 5-shot 每类包含 10 个查询图像。即,在一个 episode 中 5-way 1-shot 有 $15 \times 5 + 1 \times 5 = 80$ 个图像,5-way 5-shot 有 $10 \times 5 + 5 \times 5 = 75$ 个图像,实验设置见表 1。对于 Omniglot 数据集,20-way 1-shot 和 20-way 5-shot 两种模式进行训练,在训练阶段随机采样并构建 1 000 000 个 episode,在每个训练的 episode 中,20-way 1-shot 每类包含 10 个查询图像,20-way 5-shot 每类包含 5 个查询图像。即,在一个 episode 中 20-way 1-shot 有 $10 \times 20 + 1 \times 20 = 220$ 个图像,20-way 5-shot 有 $5 \times 20 + 5 \times 20 = 200$ 个图像,实验设置见表 2。

3.3 实验结果

在 MiniImagenet 和 Omniglot 数据集上的实验结果见表 3 和表 4。

表 1 MiniImagenet 实验设置		
MiniImagenet		
C-way K-shot 设置	5-way 1-shot	5-way 5-shot
支持集	5×1 个	5×5 个
查询集	5×15 个	5×10 个
训练的 episode 数量	500000	
测试的 episode 数量	6000	

表 2 Omniglot 实验设置		
Omniglot		
C-way K-shot 设置	20-way 1-shot	20-way 5-shot
支持集	20×1 个	20×5 个
查询集	20×10 个	20×5 个
训练的 episode 数量	1000000	
测试的 episode 数量	6000	

表 3 MiniImagenet 数据集上识别率		
模型(算法)	5-way	
	1-shot/%	5-shot/%
匹配网络 ^[16]	43.56	53.11
原型网络 ^[17]	49.42	68.20
MAML ^[30]	48.70	63.11
关系网络 ^[18]	50.44	65.32
本文方法	51.22	65.86

表 4 Omniglot 数据集上识别率		
模型(算法)	20-way	
	1-shot/%	5-shot/%
匹配网络 ^[16]	93.5	98.7
原型网络 ^[17]	96.0	98.9
MAML ^[30]	95.8	98.9
关系网络 ^[18]	97.6	99.1
本文方法	97.8	99.2

从实验数据可以看出,本文改进后的网络在 MiniImageNet 数据集上 5-way 1-shot 与 5-way 5-shot 的分类精度分别提升到了 51.22%与 65.86%,比原网络分别提高了 0.78%和 0.54%;在 Omniglot 数据集上 20-way 1-shot 与 20-way 5-shot 的分类精度,比原网络提高了 0.2%和 0.1%。

4 结论

本文在关系网络的嵌入模块中引入多尺度空间与通道注意力机制 PSA 模块,提出了一种基于多尺度通道注意力机制的小样本图像分类方法,用不同尺度的卷积核对图像进行特征提取,丰富了特征空间。实验表明,在标准的 MiniImageNet 与 Omniglot 数据集中,本文提出的方法提升了关系网络中嵌入模块的特征提取效率,提高了小样本图像的分类精度。

[参 考 文 献]

[1] KRIZHEVSKY A,SUTSKEVER I,HINTON G.Imagenet classification with deep convolutional neural networks[J]. Advances in neural information processing systems, 2012, 25(2):1097-1105.

[2] RUSSAKOVSKY O,DENG J,SU H, et al. Imagenet

- large scale visual recognition challenge[J]. International Journal of Computer Vision, 2014;1-42.
- [3] WANG Y, YAO Q, KWOK J, et al. Generalizing from a few examples: a survey on few-shot learning [EB/OL]. [2019-04-10]. arXiv:1904.05046, <https://arxiv.org/abs/1904.05046>.
 - [4] 耿瑞莹, 李永彬, 黎槟华. 小样本学习综述[EB/OL]. [2020-12-15]. <https://zhuanlan.zhihu.com/p/61215293>.
 - [5] 李新叶, 龙慎鹏, 朱婧. 基于深度神经网络的少样本学习综述[J]. 计算机应用研究, 2020, 37(8): 2241-2247.
 - [6] 赵凯琳, 靳小龙, 王元卓. 小样本学习研究综述[J]. 软件学报, 2020, 32(2): 1-16.
 - [7] 祝钧桃, 姚光乐, 张葛祥, 等. 深度神经网络的小样本学习综述[J]. 计算机工程与应用, 2021, 57(07): 22-33.
 - [8] LU J, GONG P, YE J, et al. Learning from very few samples: a survey[J/OL]. [2020-09-06]. arXiv:2009.02653, <https://arxiv.org/abs/2009.02653>.
 - [9] 刘颖, 雷研博, 范九伦, 等. 基于小样本学习的图像分类技术综述[J]. 自动化学报, 2021, 47(2): 297-315.
 - [10] PAN S J, QIANG Y. A Survey on transfer learning [J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345-1359.
 - [11] HUISMAN M, RIJN J, PLAAT A. A survey of deep meta-learning [J/OL]. [2020-10-07]. arXiv: 2010.03522. <https://arxiv.org/abs/2010.03522>.
 - [12] PAN W, LI B, CHONG Y. Image block compression based on dual-learning dictionaries[C]// 2020 12th International Conference on Advanced Computational Intelligence (ICACI). Dali: IEEE, 2020, 79-85.
 - [13] ZHU J, CHEN J, HU W, et al. Big learning with bayesian methods[J]. National Science Review, 2017 (4): 627-651.
 - [14] YANG, LIU, JIN R. Distance metric learning: a comprehensive survey. michigan state universiy[J], 2006: 1-51.
 - [15] ZAGORUYKO S, KOMODAKIS N. Learning to compare image patches via convolutional neural networks [C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston: IEEE, 2015, 4353-4361.
 - [16] VINYALS O, BLUNDELL C, LILICRAP T, et al. Matching networks for one shot learning [J/OL]. [2016-06-13]. arXiv:1606.04080, <https://arxiv.org/abs/1606.04080>.
 - [17] JAKE S, KEVIN S, RICHARD Z. Prototypical networks for few-shot learning[C]// Conference on Neural Information Processing Systems. Long Beach: NeurIPS, 2017, 4077-4087.
 - [18] SUNG F, YANG Y, ZHANG L, et al. Learning to Compare: Relation Network for Few-Shot Learning [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake: IEEE, 2018, 1199-1208.
 - [19] 张碧陶, 庞振全. 融合强化学习和关系网络的样本分类[J]. 计算机工程与应用, 2019, 55(21): 189-196+253.
 - [20] 金璐, 刘士建, 王霄, 等. 基于改进关系网络的小样本红外空中目标分类方法[J]. 光学学报, 2020, 40(8): 87-96.
 - [21] 王年, 孟树林, 吴洛天, 等. 基于改进关系网络的小样本学习[J]. 安徽大学学报(自然科学版), 2020, 44(4): 38-44.
 - [22] 张宸嘉, 朱磊, 俞璐. 卷积神经网络中的注意力机制综述[J/OL]. 计算机工程与应用: 1-10 [2021-09-24]. <http://kns.cnki.net/kcms/detail/11.2127.tp.20210726.0839.002.html>.
 - [23] JIE H, LI S, GANG S, et al. Squeeze-and-Excitation Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017: 99.
 - [24] WANG Q, WU B, Zhu P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020, 11 531-11 539.
 - [25] LI X, WANG W, HU X, et al. Selective kernel networks[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, 510-519.
 - [26] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition[C]// IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, 770-778.
 - [27] 付国栋, 黄进, 杨涛, 等. 改进 CBAM 的轻量级注意力模型[J/OL]. 计算机工程与应用: 1-8 [2021-05-10]. <http://kns.cnki.net/kcms/detail/11.2127.TP.20210420.0924.006.html>.
 - [28] LAKE B, SALAKHUTDINOV R, GROSS J, et al. Humanlevel concept learning through probabilistic program induction [J]. Science, 2015, 350 (6266): 1332-1338.
 - [29] ZHANG HU, ZU KEKE, LU JIAN, et al. EPSANet: An efficient pyramid split attention block on convolutional neural network [J/OL]. [2021-05-30]. arXiv: 2105.14447. <https://arxiv.org/abs/2105.14447>.
 - [30] FINN C, ABBEEL P, LEVINE S, et al. Model-agnostic meta-learning for fast adaptation of deep network [C]// International Conference on Machine Learning. Sydney: ICML, 2017, 1126-1135.