

基于元学习和位置信息的小样本学习方法

张水利*, 张 军, 白宗文

延安大学物理与电子信息学院, 陕西 延安

收稿日期: 2023年4月4日; 录用日期: 2023年4月24日; 发布日期: 2023年4月29日

摘 要

随着小样本学习的发展, 元学习已经成为一种流行的小样本学习框架, 其作用是开发能够快速适应有限数据和低计算成本的小样本分类任务模型。最近有关注意的研究已经证明了通道注意力对于特征提取效果有一定的提升, 但是它忽略了位置信息的作用, 位置信息对于在小样本任务中更好地从有限的数据中学习来说很重要。基于这一事实, 本文提出一种新的方法, 通过在所有基类上预先训练一个加入位置信息注意力的分类器, 然后在基于最近质心的小样本分类算法上进行元学习, 实现了将位置信息和提取特征有效的结合。通过在两个标准的数据集上实验, 和当下主流的小样本图像分类方法相比, 该方法在 Mini-ImageNet 数据集的 1-shot 与 5-shot 任务上分别提升 1.23% 和 1.02%, 在 Tiered-ImageNet 数据集上, 也分别提升 0.85% 和 0.78%。实验表明该方法有效的发挥了位置信息的作用, 可以提升小样本图像分类的准确率。

关键词

小样本学习, 元学习, 位置信息注意力, 长期依赖关系, 最近邻分类器

Few-Shot Learning Method Based on Meta-Learning and Location Information

Shuili Zhang*, Jun Zhang, Zongwen Bai

School of Physics and Electronic Information, Yan'an University, Yan'an Shaanxi

Received: Apr. 4th, 2023; accepted: Apr. 24th, 2023; published: Apr. 29th, 2023

Abstract

With the development of few-shot learning, meta-learning has become a popular few-shot learning framework, and its role is to develop models for few-shot classification tasks that can quickly

*通讯作者。

adapt to limited data and low computational cost. Recent studies on attention have demonstrated that channel attention can improve feature extraction to a certain extent, but it ignores the role of location information, which is important for better learning with limited data in few-shot tasks. Based on this fact, this paper proposes a new method to effectively combine location information and extracted features by pre-training a classifier with location information attention on all base classes, and then performing meta-learning on few-shot classification algorithm based on the nearest centroid. Through experiments on two standard datasets, compared with the current mainstream few-shot image classification methods, this method was improved by 1.23% and 1.02% on the 1-shot and 5-shot tasks of the Mini-ImageNet dataset, and also by 0.85% and 0.78% on the Tiered-ImageNet dataset. Experiments show that the method effectively plays the role of location information and can improve the accuracy of few-shot image classification.

Keywords

Few-Shot Learning, Meta Learning, Location Information Attention, Long Term Dependencies, Nearest Neighbor Classifier

Copyright © 2023 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

人类可以通过观察很少的几个例子,就能够适应一项新的任务,这是因为人类的大脑具有良好的学习能力。目前人工智能算法往往需要大量训练数据,但是在很多时候,去收集这些大量的带标签数据的成本是非常高,有时甚至是不可能做到的。比如,冷启动建议[1]和药物发现[2]。为了使人工智能去模仿人类的学习能力,小样本学习[3][4]被提出用于训练网络。小样本学习可以被看作是更广泛的元学习[5]的一个特殊案例,与传统机器学习的不同点在于,小样本学习的目标是从一组具有丰富标记样本的基类中训练分类器,然后转移到具有少量标记样本的新类[6]。

在当前的环境下,小样本学习的研究方法大致分为四种,即基于度量的方法[7][8]、基于优化的方法[9]、基于图形的方法[10]和基于语义的方法[11]。虽然他们的方法不同,但是都遵循元学习中“学会学习”的关键思想,通过元训练和元测试两个阶段来解决小样本问题。具体来说,就是先在基类的训练样本中抽取小样本任务,并训练模型使得其在这些任务上表现良好。任务通常采用 N-way 和 K-shot 的形式,其中每个类都包含 K 个支持样本和 Q 个查询样本。小样本学习的目标是在学习完毕之后,将这 Q 个查询样本准确地分类。基于这一思想,领域内最近的工作集中于改进元学习结构,越来越多的元学习方法[12][13]被提议用于小样本学习。近期在文献[14]中发现通过额外的预训练阶段可以明显的提升效果。文章首先通过在整个基类上预训练一个分类器基线,用以学习视觉表示,之后去除最后一个与类相关的全连接层,把它当作一个特征提取器。然后给定一个样本数较少的新类,计算给定样本的平均特征,并利用特征空间中的余弦距离,即余弦最近质心,按最近质心对查询样本进行分类。在元训练阶段,用基于度量的元训练方式对预先训练好的分类器基线进行微调。在元测试阶段,通过具有余弦距离的最近邻分类器和基于均值的原型对新类进行分类。从结论上看,元训练可以进一步改善分类器基线的性能。

虽然基于预训练的元学习方法取得了不错的改进,但是这种方法忽略了位置信息的重要性,位置信息对于在视觉任务中捕获对象结构至关重要[15]。迄今为止,最流行的挤压和激发注意力机制[16]只考虑了编码通道间的信息,通过二维全局池化计算通道注意,也没有考虑到位置信息。后来的工作,如 BAM

[17]和 CBAM [18], 试图通过减少输入张量的通道维数, 然后使用卷积计算空间注意力, 来利用位置信息。然而, 卷积只能捕获局部关系, 无法建模对视觉任务至关重要的长期依赖关系[19]。

基于这一事实, 本文提出一种新的基于原型和元训练的小样本学习方法。在预训练特征提取器的时候, 受坐标注意力[20]启发, 本文引入了一种新颖且有效的注意力机制。为了缓解二维全局池化造成的位置信息丢失, 将通道注意分解为两个并行的一维特征编码过程, 有效地将空间坐标信息整合到生成的注意力映射中。具体来说, 该机制利用两个一维全局池化操作分别将垂直方向和水平方向上的输入特征聚合成两个独立的方向感知特征映射。这两个带有特定方向信息的特征映射分别被编码成两个注意力映射, 每个注意力映射捕获了输入特征映射在一个空间方向上的长期依赖关系。因此, 位置信息可以保存在生成的注意力映射中。通过乘法将这两种注意力映射应用于输入特征映射, 来表示感兴趣的特征。特征提取器预训练好之后, 再用基于余弦距离度量的元训练进行微调。最后在元测试时, 用最近邻分类器进行小样本的分类。

2. 相关工作

2.1. 小样本学习

小样本学习的目标是使分类模型适应只有少量标记样本的新类, 元学习是解决小样本问题的有效途径。最近, 一些研究转向小样本问题的预训练技术, 并取得了良好的效果。在文献[21]中首先提出并研究了小样本问题中的预训练技术, 分别考虑了基于线性和基于余弦距离的分类器。这些方法虽然提供了很好的性能, 但是没有充分发挥位置信息的重要性, 位置信息对于在视觉任务中捕获对象结构至关重要。因此, 在本文中, 我们提出一个方法来解决这个问题。

2.2. 注意力机制

注意力可以被看作是一种工具, 它将可用的处理资源分配给输入信号中信息量最大的部分, 它在文献[22]中已被证明在各种计算机视觉任务中 useful, 例如图像分类[23]和图像分割[24]。其中一个成功的例子是 SENet [16], 它简单地压缩每个二维特征图, 以有效地建立通道之间的相互依赖关系, 旨在增强整个网络中基本模块的表示能力。CBAM [18]进一步推进了这一想法, 通过大尺寸核卷积引入空间信息编码。后来的作品, 如 GENet [23], GALA [25]和 TA [26], 通过采用不同的空间注意机制或设计高级注意块, 扩展了这一想法。

非局部/自注意网络最近非常流行, 因为它们能够建立空间或通道性注意。典型的例子包括 NLNet [27]和 GCNet [28], 它们都利用非局部机制来捕获不同类型的空间信息。

与这些自注意网络的方法不同, 我们的方法考虑了一种更有效的方式来捕获位置信息和通道关系, 以增强网络的特征表示。通过将二维全局池操作分解为两个一维编码过程, 本文提出的方法比其他的注意力方法(如 SENet [16], CBAM [18]和 TA [26])性能更优。

3. 方法

3.1. 问题定义

在小样本分类的一般工作中, 给定一个基类 C_{base} 带有标签的数据集, 每个类中有大量的图像, 目标是学习新类 C_{novel} 中类别的概念, 每个新类中有几个样本, 其中基类和新类之间的标签空间不相交。在 N-wayK-shot 的小样本任务中, 支持集包含 N 个类别, 每个类别中有 K 个样本, 查询集包含相同的 N 个类别, 每个类别中有 Q 个样本, 目标是将查询集中的 $N \times Q$ 个未标记的样本分类为 N 个类别。在评估表现时, 从 C_{novel} 数据中采集的很多任务来计算平均精度。

3.2. 坐标注意力

3.2.1. SE 注意力

在文献[16]中,标准卷积本身很难模拟通道关系。明确地建立通道相互依赖关系可以提高模型对信息通道的敏感性,这些信息通道对最终分类决策的贡献更大。此外,使用全局平均池化还可以帮助模型捕获全局信息,这是卷积操作不能完成的。

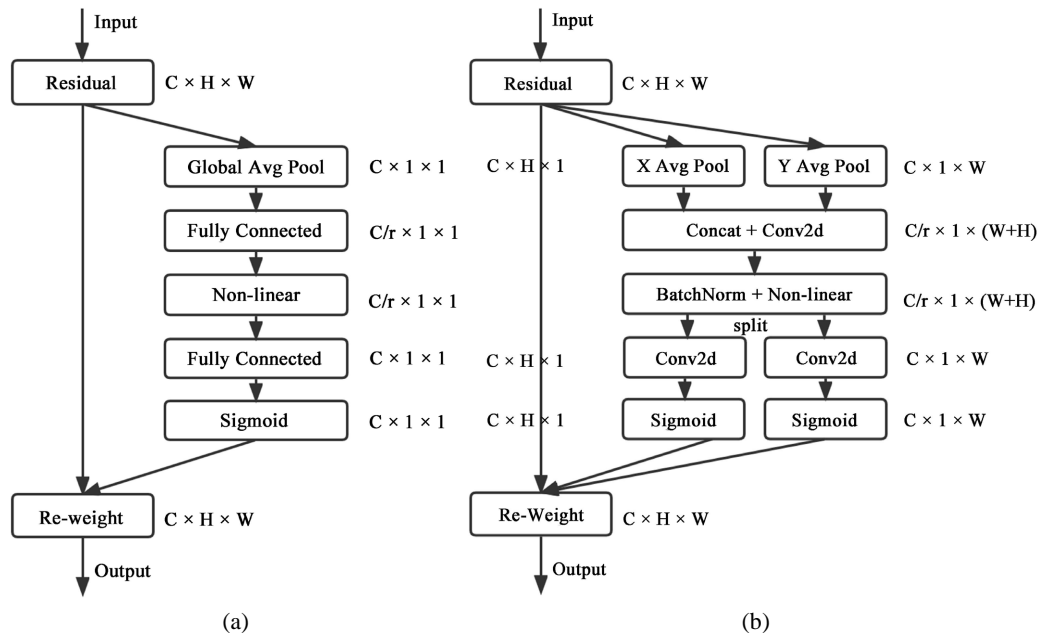


Figure 1. Schematic diagram of SE channel attention (a) and coordinate attention (b)

图 1. SE 通道注意力(a)与坐标注意力(b)示意图

在结构上,如图 1(a)所示,SE 注意块可以分解为两个步骤:挤压和激发,分别用于全局信息嵌入和通道关系的自适应重新校准。给定输入 X ,第 c 个通道的挤压步骤如下:

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W x_c(i, j), \quad (1)$$

其中 z_c 是与第 c 通道相关的输出。输入 X 直接来自具有固定核大小的卷积层,因此可以被视为局部描述符的集合。挤压操作使收集全局信息成为可能。

第二步,激发,旨在捕获通道依赖性,表述为:

$$\hat{X} = X \cdot \sigma(\hat{z}), \quad (2)$$

其中 \cdot 表示乘法, σ 是 sigmoid 函数, \hat{z} 表示变换函数生成的结果,公式如下:

$$\hat{z} = T_2(\text{ReLU}(T_1(z))). \quad (3)$$

这里, T_1 和 T_2 是两个线性变换,可以学习它们来捕获每个通道的重要性。

SE 注意力在计算机视觉任务中得到了广泛应用,并被证明是实现优越性能的关键角色。但是它只考虑了通过建模通道关系来重新评估每个通道的重要性,而忽略了位置信息,位置信息对于生成空间选择性注意图非常重要。接下来本文将介绍坐标注意力,它同时考虑了通道间关系和位置信息。

3.2.2. 坐标注意力

坐标注意力[20]编码通道关系与精确的位置信息分为两个步骤：坐标信息嵌入和坐标注意力生成。坐标注意力的示意图如图 1(b)所示，下面将详细描述坐标注意力。

首先是坐标信息的嵌入，为了鼓励注意块通过精确的位置信息在空间上捕获远程交互，将全局平均池化公式(1)分解为一对一维特征编码操作。具体来说，给定输入 X ，使用两个空间范围的池化核($H, 1$)或($1, W$)分别沿水平坐标和垂直坐标对每个通道进行编码。因此，第 c 个通道在高度 h 处的输出可以表示为：

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i < W} x_c(h, i), \quad (4)$$

类似的，第 c 个通道在宽度 w 处的输出可以表示为：

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j < H} x_c(j, w), \quad (5)$$

上述两种变换分别沿两个空间方向聚合特征，生成一对方向感知特征映射。这与产生单一特征向量的通道注意力方法中的挤压操作(公式(1))有很大不同。这两种变换还允许注意块沿着一个空间方向捕获长期依赖，并沿着另一个空间方向保留精确的位置信息，这有助于网络更准确地定位感兴趣的对象。

第二步，坐标注意力生成。上述公式(4)和公式(5)启用全局感受野对精确的位置信息进行编码。为了利用这种表达性表征，提出了第二种变换，坐标注意力生成。它可以充分利用捕获到的位置信息，从而准确地突出显示感兴趣的区域，它还能够有效地捕获通道间的关系。具体地说，给定由公式(4)和公式(5)生成的聚合特征映射，将它们连接起来，然后发送到共享的 1×1 卷积变换函数 F_1 ，得到：

$$f = \delta \left(F_1 \left(\left[z^h, z^w \right] \right) \right), \quad (6)$$

其中 $[\cdot, \cdot]$ 表示沿空间维度的连接合并操作， δ 是非线性激活函数， $f^w \in \mathbb{R}^{C/r \times (H+W)}$ 是在水平方向和垂直方向上编码空间信息的中间特征图。这里， r 是用于控制块大小的缩小比，如在 SE 块中一样。然后沿着空间维度将 f 拆分为两个独立的张量 $f^h \in \mathbb{R}^{C/r \times H}$ 和 $f^w \in \mathbb{R}^{C/r \times W}$ 。利用另外两个 1×1 卷积变换 F_h 和 F_w 分别将 f^h 和 f^w 变换成与输入 X 具有相同通道数的张量，得到：

$$g^h = \sigma \left(F_h \left(f^h \right) \right), \quad (7)$$

$$g^w = \sigma \left(F_w \left(f^w \right) \right), \quad (8)$$

这里， σ 是 sigmoid 函数。将输出的 g^h 和 g^w 分别展开并用作坐标注意力权重。最后，通过坐标注意力的输出 Y 可以写成：

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j). \quad (9)$$

3.3. 分类器基线

本文使用的是加入了坐标注意力的分类器基线，如图 2。具体来说，输入一张图片，通过卷积神经网络(CNN)得到它的特征图 X ，然后通过坐标注意力嵌入位置信息以后，得到了定位感兴趣对象后的特征图 Y 。把得到的带有位置信息的特征图 Y ，展开为一维数组，得到全连接层 FC，最后进行分类。

本文在所有基类上训练一个加入了坐标注意力的分类器，并使用余弦最近质心方法执行小样本分类任务。具体来说，在所有基类上用标准交叉熵损失训练这个分类器，然后移除最后一个 softmax 层，得到编码器 f_θ ，它将输入映射到特征空间。给定一个带有支持集 S 的小样本任务，让 S_c 表示 c 类别中的样例数量，计算平均特征 w_c 作为 c 类的质心。

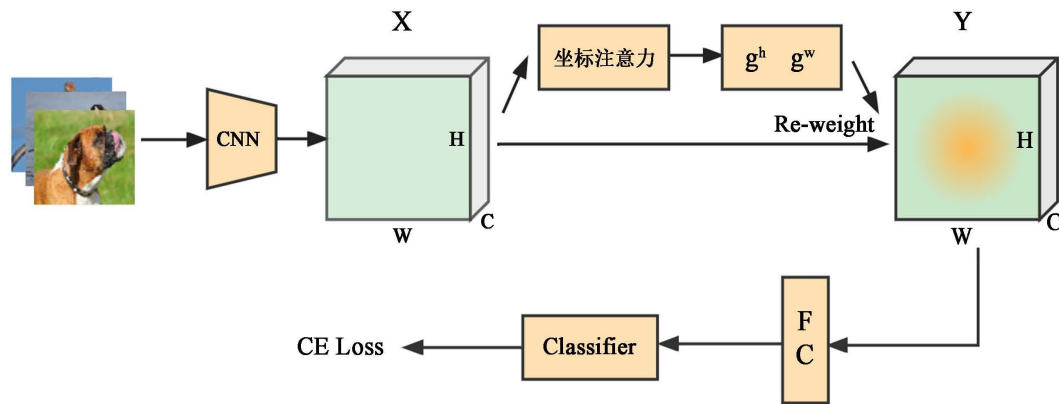


Figure 2. Classifier baseline with coordinate attention

图 2. 加入坐标注意力的分类器基线

$$w_c = \frac{1}{|S_c|} \sum_{x \in S_c} f_\theta(x), \quad (10)$$

然后对于小样本任务中的查询样本 x ，通过样本 x 的特征向量与类别 c 质心之间的余弦相似度来预测样本 x 属于类别 c 的概率：

$$p(y=c|x) = \frac{\exp(\langle f_\theta(x), w_c \rangle \cdot \gamma)}{\sum_c \exp(\langle f_\theta(x), w_c \rangle \cdot \gamma)}. \quad (11)$$

其中 $\langle \cdot, \cdot \rangle$ 表示两个向量的余弦相似性， γ 是可学习的尺度参数。注意 w_c 也可以被看作类别 c 的新 FC 层的预测权重。

3.4. 元基线

元基线包含两个训练阶段，第一阶段是预训练阶段，即训练加入了坐标注意力的分类器基线方法(即在所有基类上训练这个分类器，并移除其最后一个 FC 层以获得编码器 f_θ)。第二阶段是元学习阶段，即对分类器基线评估算法的模型进行优化。具体地说，给定预先训练的特征编码器 f_θ ，然后从基类中的训练数据中采样 N-wayK-shot 任务(使用 $N \times Q$ 查询样本)。为了计算每个任务的损失，在支持集中，计算等式(10)中定义的 N 个类的质心，然后计算等式(11)用于定义查询集中每个样本的预测概率分布。该损失是由 p 和查询集中样本的标签计算出的交叉熵损失。注意本方法将每个任务视为训练中的一个数据点，每个 batch 可能包含多个任务，并计算平均损失。

4. 结果与分析

4.1. 数据集

Mini-ImageNet 数据集[4]是小样本学习的常见基准。它包含从 ILSVRC-2012 [29]中抽样的 100 个类，然后将这些类随机分为 64, 16, 20 个类，分别作为训练集、验证集和测试集。每个类包含 600 张尺寸大小为 84×84 的图像。

Tiered-ImageNet 数据集[30]是最近提出的另一个通用基准，其规模要大得多。它是 ILSVRC-2012 的一个子集，包含来自 34 个高级类别的 608 个类，然后将这些类拆分为 20, 6, 8 个高级类别，分别产生 351, 97, 160 个类作为训练集，验证集和测试集。图像尺寸大小为 84×84 。注意这个数据集更具挑战性，因为基类和新类来自不同的高级类别。

4.2. 实验细节

本方法在 Mini-ImageNet 和 Tiered-ImageNet 上使用 ResNet12 作为特征提取器进行了实验。在训练加入坐标注意力的分类器基线时，本方法使用动量为 0.9 的 SGD 优化器，学习率从 0.1 开始，衰减因子为 0.1。在 Mini-ImageNet 上，训练了 100 个 batch size 为 128 的 epoch，学习率在 epoch 90 时衰减。在 Tiered-ImageNet 上，训练了 120 个 batch size 为 512 的 epoch，学习率在 epoch 40 和 80 时衰减。ResNet-12 的权重衰减为 0.0005。应用标准数据增强，包括随机裁剪和水平翻转。在元训练时，本方法使用动量为 0.9 的 SGD 优化器，固定学习率为 0.001，batch size 为 4，即每个 batch 包含 4 个小样本任务，以计算平均损失。余弦尺度参数 γ 初始化为 10，缩小比 r 为 32。

4.3. 实验结果与分析

按照标准设置，本方法在 Mini-ImageNet 和 Tiered-ImageNet 上进行了实验，结果分别如表 1 和表 2 所示。虽然之前的工作有不同的数据增强策略，但本方法选择不在元学习阶段应用数据增强。在这两个数据集上，本方法在 Mini-ImageNet 上的 1-shot 和 5-shot 精度分别提升了 1.23% 和 1.02%，在 Tiered-ImageNet 上的 1-shot 和 5-shot 精度分别提升了 0.85% 和 0.78%。本方法利用坐标注意力去嵌入位置信息，从而定位了有感兴趣对象的特征图，然后结合元学习的优点，通过结果验证了本方法的有效性。

Table 1. Classification accuracy of 5-way tasks on Mini-ImageNet

表 1. Mini-ImageNet 上的 5-way 任务的分类精度

| Method | Backbone | 1-Shot | 5-Shot |
|---|-----------|------------------|------------------|
| LEO (Rusu <i>et al.</i> , 2019) | WRN-28-10 | 61.76 \pm 0.08 | 77.59 \pm 0.12 |
| SNAIL (Mishra <i>et al.</i> , 2018) | ResNet-12 | 55.71 \pm 0.99 | 68.88 \pm 0.92 |
| AdaResNet (Munkhdalai <i>et al.</i> , 2017) | ResNet-12 | 56.88 \pm 0.62 | 71.94 \pm 0.57 |
| TADAM (Oreshkin <i>et al.</i> , 2018) | ResNet-12 | 58.50 \pm 0.30 | 76.70 \pm 0.30 |
| MTL (Sun <i>et al.</i> , 2019) | ResNet-12 | 61.20 \pm 1.80 | 75.50 \pm 0.80 |
| MetaOptNet (Lee <i>et al.</i> , 2019) | ResNet-12 | 60.33 \pm 0.61 | 76.61 \pm 0.46 |
| Meta-Baseline(Chen <i>et al.</i> , 2021) | ResNet-12 | 63.17 \pm 0.23 | 79.26 \pm 0.17 |
| Our Method | ResNet-12 | 64.40 \pm 0.18 | 80.28 \pm 0.19 |

Table 2. Classification accuracy of 5-way tasks on Tiered-ImageNet

表 2. Tiered-ImageNet 上的 5-way 任务的分类精度

| Method | Backbone | 1-Shot | 5-Shot |
|---|-----------|------------------|------------------|
| MAML (Finn <i>et al.</i> , 2017) | ConvNet-4 | 51.67 \pm 1.81 | 70.30 \pm 1.75 |
| Prototypical Networks* (Snell <i>et al.</i> , 2017) | ConvNet-4 | 53.31 \pm 0.89 | 72.69 \pm 0.74 |
| Relation Networks* (Sung <i>et al.</i> , 2018) | ConvNet-4 | 54.48 \pm 0.93 | 71.32 \pm 0.78 |
| LEO (Rusu <i>et al.</i> , 2019) | WRN-28-10 | 66.33 \pm 0.05 | 81.44 \pm 0.09 |
| MetaOptNet (Lee <i>et al.</i> , 2019) | ResNet-12 | 65.99 \pm 0.72 | 81.56 \pm 0.53 |
| Meta-Baseline(Chen <i>et al.</i> , 2021) | ResNet-12 | 68.62 \pm 0.27 | 83.29 \pm 0.18 |
| Our Method | ResNet-12 | 69.47 \pm 0.30 | 84.07 \pm 0.25 |

4.4. 消融实验

为了证明坐标注意力的重要性,本文在 Mini-ImageNet 和 Tiered-ImageNet 上进行了一系列消融实验,相应的结果都在表 3 和表 4 中。本文以 Meta-Baseline 为基线,分别对比了加入 SE 注意力, CBAM 注意力以及坐标注意力的表现。实验结果发现,使用 SE 注意力和 CBAM 注意力有相近的结果,可以使性能得到提升,证明了通道信息的重要性。但是,坐标注意力的表现最好,因为它同时编码了通道信息和位置信息。

Table 3. Ablation experiment of 5-way task on Mini-ImageNet

表 3. Mini-ImageNet 上的 5-way 任务的消融实验

| Settings | Backbone | 1-Shot | 5-Shot |
|---------------|-----------|------------------|------------------|
| Meta-Baseline | ResNet-12 | 63.17 \pm 0.23 | 79.26 \pm 0.17 |
| +SE | ResNet-12 | 63.66 \pm 0.25 | 79.74 \pm 0.28 |
| +CBAM | ResNet-12 | 63.73 \pm 0.42 | 79.87 \pm 0.35 |
| +坐标注意力 | ResNet-12 | 64.40 \pm 0.18 | 80.28 \pm 0.19 |

Table 4. Ablation experiment of 5-way task on Tiered-ImageNet

表 4. Tiered-ImageNet 上的 5-way 任务的消融实验

| Settings | Backbone | 1-Shot | 5-Shot |
|---------------|-----------|------------------|------------------|
| Meta-Baseline | ResNet-12 | 68.62 \pm 0.27 | 83.29 \pm 0.18 |
| +SE | ResNet-12 | 68.97 \pm 0.35 | 83.74 \pm 0.19 |
| +CBAM | ResNet-12 | 69.10 \pm 0.22 | 83.89 \pm 0.32 |
| +坐标注意力 | ResNet-12 | 69.47 \pm 0.30 | 84.07 \pm 0.25 |

5. 结论

本文提出了一种有效的小样本学习方法,即在训练特征提取时加入了坐标注意力,它不仅继承了通道注意(例如 SE 注意)的优点,可以模拟通道间的关系,同时还利用了精确的位置信息捕获长期依赖关系,可以准确地定位感兴趣的对象,充分发挥了位置信息的作用。然后在基于最近质心的小样本分类算法上进行元学习。实验表明,本文的方法在两个数据集上都取得了较好的表现。未来的工作是探索更加有效的捕获位置信息的方法,以便于提升小样本图像分类的准确率。

基金项目

工业人工智能中跨媒体协同深度安全态势感知理论和应用研究(62266045)。

参考文献

- [1] Vartak, M., Thiagarajan, A., Miranda, C., et al. (2017) A Meta-Learning Perspective on Cold-Start Recommendations for Items. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 6907-6917.
- [2] Altae-Tran, H., Ramsundar, B., Pappu, A.S., et al. (2017) Low Data Drug Discovery with One-Shot Learning. *ACS Central Science*, **3**, 283-293. <https://doi.org/10.1021/acscentsci.6b00367>
- [3] Li, F.F., Fergus, R. and Perona, P. (2006) One-Shot Learning of Object Categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **28**, 594-611. <https://doi.org/10.1109/TPAMI.2006.79>

- [4] Vinyals, O., Blundell, C., Lillicrap, T., *et al.* (2016) Matching Networks for One Shot Learning. *Proceedings of the 30th International Conference on Neural Information Processing Systems*, Barcelona, 5-10 December 2016, 3637-3645.
- [5] Hospedales, T., Antoniou, A., Micaelli, P., *et al.* (2020) Meta-Learning in Neural Networks: A Survey.
- [6] Wang, Y., Yao, Q., Kwok, J.T., *et al.* (2020) Generalizing from a Few Examples: A Survey on Few-Shot Learning. *ACM Computing Surveys*, **53**, 1-34. <https://doi.org/10.1145/3386252>
- [7] Chen, J., Zhan, L.M., Wu, X.M., *et al.* (2020) Variational Metric Scaling for Metric-Based Meta-Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, **34**, 3478-3485. <https://doi.org/10.1609/aaai.v34i04.5752>
- [8] 叶萌, 杨娟, 汪荣贵, 薛丽霞, 李懂. 基于特征聚合网络的小样本学习方法[J]. 计算机工程, 2021, 47(3): 77-82.
- [9] Finn, C., Abbeel, P. and Levine, S. (2017) Model-Agnostic Meta-Learning for Fast Adaptation of Deep Network. *International Conference on Machine Learning*, Sydney, 6-11 August 2017, 1126-1135.
- [10] Garcia, V. and Bruna, J. (2017) Few-Shot Learning with Graph Neural Networks.
- [11] Xing, C., Rostamzadeh, N., Oreshkin, B., *et al.* (2019) Adaptive Cross-Modal Few-Shot Learning. *33rd Annual Conference on Neural Information Processing Systems (NeurIPS 2019)*, Vancouver, 8-14 December 2019, 4847-4857.
- [12] Snell, J., Swersky, K. and Zemel, R. (2017) Prototypical Networks for Few-Shot Learning. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 4080-4090.
- [13] Sung, F., Yang, Y., Zhang, L., *et al.* (2018) Learning to Compare: Relation Network for Few-Shot Learning. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 1199-1208. <https://doi.org/10.1109/CVPR.2018.00131>
- [14] Chen, Y., Wang, X., Liu, Z., *et al.* (2020) A New Meta-Baseline for Few-Shot Learning.
- [15] Wang, H., Zhu, Y., Green, B., *et al.* (2020) Axial-deeplab: Stand-Alone Axial-Attention for Panoptic Segmentation. *European Conference on Computer Vision*, Glasgow, 23-28 August 2020, 108-126. https://doi.org/10.1007/978-3-030-58548-8_7
- [16] Hu, J., Shen, L. and Sun, G. (2018) Squeeze-and-Excitation Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 7132-7141. <https://doi.org/10.1109/CVPR.2018.00745>
- [17] Park, J., Woo, S., Lee, J.Y., *et al.* (2018) Bam: Bottleneck Attention Module.
- [18] Woo, S., Park, J., Lee, J.Y., *et al.* (2018) CBAM: Convolutional Block Attention Module. *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, 8-14 September 2018, 3-19. https://doi.org/10.1007/978-3-030-01234-2_1
- [19] Zhao, H., Shi, J., Qi, X., *et al.* (2017) Pyramid Scene Parsing Network. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 2881-2890. <https://doi.org/10.1109/CVPR.2017.660>
- [20] Hou, Q., Zhou, D. and Feng, J. (2021) Coordinate Attention for Efficient Mobile Network Design. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, 20-25 June 2021, 13713-13722. <https://doi.org/10.1109/CVPR46437.2021.01350>
- [21] Chen, W.Y., Liu, Y.C., Kira, Z., *et al.* (2019) A Closer Look at Few-Shot Classification.
- [22] Tsotsos, J.K. (1990) Analyzing Vision at the Complexity Level. *Behavioral and Brain Sciences*, **13**, 423-445. <https://doi.org/10.1017/S0140525X00079577>
- [23] Hu, J., Shen, L., Albanie, S., *et al.* (2018) Gather-Excite: Exploiting Feature Context in Convolutional Neural Networks. *32nd Conference on Neural Information Processing Systems (NeurIPS 2018)*, Montréal, 2-8 December 2018, 9401-9411.
- [24] Hou, Q., Zhang, L., Cheng, M.M., *et al.* (2020) Strip Pooling: Rethinking Spatial Pooling for Scene Parsing. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 13-19 June 2020, 4003-4012. <https://doi.org/10.1109/CVPR42600.2020.00406>
- [25] Linsley, D., Shiebler, D., Eberhardt, S., *et al.* (2018) Learning What and Where to Attend.
- [26] Misra, D., Nalamada, T., Arasanipalai, A.U., *et al.* (2021) Rotate to Attend: Convolutional Triplet Attention Module. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, Waikoloa, 3-8 January 2021, 3139-3148. <https://doi.org/10.1109/WACV48630.2021.00318>
- [27] Wang, X., Girshick, R., Gupta, A., *et al.* (2018) Non-Local Neural Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 7794-7803. <https://doi.org/10.1109/CVPR.2018.00813>
- [28] Cao, Y., Xu, J., Lin, S., *et al.* (2019) Gcnet: Non-Local Networks Meet Squeeze-Excitation Networks and Beyond. *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, Seoul, 27-28 October 2019,

-
- 1-10. <https://doi.org/10.1109/ICCVW.2019.00246>
- [29] Russakovsky, O., Deng, J., Su, H., *et al.* (2015) Imagenet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, **115**, 211-252. <https://doi.org/10.1007/s11263-015-0816-y>
- [30] Ren, M., Triantafillou, E., Ravi, S., *et al.* (2018) Meta-Learning for Semi-Supervised Few-Shot Classification.