

# 基于小样本学习融合随机深度和多尺度卷积的 SDM-RNET 网络<sup>①</sup>



刘馨瑶, 梁 军, 余嘉琳

(华南师范大学 软件学院, 佛山 528225)  
通信作者: 余嘉琳, E-mail: 2022024208@m.scnu.edu.cn

**摘 要:** 针对神经网络难以利用少量标注数据获取足够的信息来正确分类图像的问题, 提出了一种融合随机深度网络和多尺度卷积的关系网络——SDM-RNET. 首先在模型嵌入模块引入随机深度网络用于加深模型深度, 然后在特征提取阶段采用多尺度深度可分离卷积替代普通卷积进行特征融合, 经过骨干网络后再采用深浅层特征融合获取更丰富的图像特征, 最终学习预测出图像的分类. 在 mini-ImageNet、RP2K、Omniglot 这 3 个数据集上对比该方法与其他小样本图像分类方法, 结果表明在 5-way 1-shot 和 5-way 5-shot 分类任务上该方法准确率最高.

**关键词:** 深度学习; 小样本学习; 图像分类

引用格式: 刘馨瑶, 梁军, 余嘉琳. 基于小样本学习融合随机深度和多尺度卷积的 SDM-RNET 网络. 计算机系统应用, 2024, 33(4): 93-102. <http://www.c-s-a.org.cn/1003-3254/9460.html>

## SDM-RNET Network Based on Small-sample Learning Fusing Stochastic Depth and Multi-scale Convolution

LIU Xin-Yao, LIANG Jun, YU Jia-Lin

(School of Software, South China Normal University, Foshan 528225, China)

**Abstract:** To solve the problem that it is difficult for neural networks to obtain enough information to correctly classify images by using a small amount of labeled data, this study proposes a new relational network, SDM-RNET, which combines random deep network and multi-scale convolution. First, a stochastic deep network is introduced into the model embedding module to deepen the model depth. Then, in the feature extraction stage, multi-scale depth-separable convolution is adopted to replace ordinary convolution for feature fusion. After the backbone network, deep and shallow layer feature fusion is applied to obtain richer image features and finally learn to predict the categories of images. Compared with other small sample image classification methods on mini-ImageNet, RP2K, and Omniglot datasets, the results show that the proposed method has the highest accuracy on 5-way 1-shot and 5-way 5-shot classification tasks.

**Key words:** deep learning (DL); small-sample learning; image classification

随着数据量爆发式增长, 深度学习模型在计算机视觉领域取得了前所未有的成功. 深度学习模型由多层神经网络构成, 在图像分类<sup>[1]</sup>、机器翻译<sup>[2]</sup>、文本分类<sup>[3]</sup>和语音建模<sup>[4]</sup>等任务中表现出卓越性能. 然而, 这些成就离不开大规模数据的支持, 而数据的采集和标

注成本十分昂贵. 例如, ImageNet 是当前全球最大的图像识别数据集, 包含了超过 2 万个类别, 每个类别有数百张图像. 与之相反, 人类可以通过少量的数据进行快速学习, 这启发了小样本图像分类学习<sup>[5]</sup>的研究. 小样本图像分类可以帮助模型提升泛化能力. 当模型只能

① 基金项目: 广东省基础与应用基础研究基金 (2022A1515140110, 2020B1515120089, 2021A1515110673); 佛山市高等教育高层次人才项目

收稿时间: 2023-10-15; 修改时间: 2023-11-15; 采用时间: 2023-11-24; csa 在线出版时间: 2024-01-30

CNKI 网络首发时间: 2024-02-01

依赖少量标记样本进行学习时,它被迫从有限的信息中学习一种更通用的特征表示,从而使模型在处理未见过的类别时具备更好的泛化能力。

小样本图像分类作为小样本学习的重要子领域,近年来备受关注。国内外许多学者都致力于解决小样本学习问题<sup>[6]</sup>。在小样本图像分类任务中,我们需要训练一个分类器,该分类器具备像人类一样的快速学习能力,可以正确对新类别的样本进行分类,即依赖于少量带标签的样本。然而,由于神经网络在处理少量标记数据时难以获取足够的信息进行准确分类,因此准确率较低。为了解决这个问题,本文改进了关系网络,并提出了基于度量学习的小样本图像分类方法。

针对现有关系网络中特征提取模块深度不足的问题,无法捕捉数据中的深层次特征,以及单一尺度卷积提取信息不全面的问题。本文提出了一种融合随机深度网络和多尺度卷积的关系网络(SDM-RNET),同时提升了嵌入模块的特征提取能力,以提高整体分类效果。提高网络性能的最常见的方法是加深网络的深度、宽度,但是过深过于复杂的网络不适合应用在小样本学习当中,因为过深的网络可能造成模型的过拟合<sup>[7-9]</sup>。因此,本文在嵌入模块中使用随机深度网络来加深网络。由于单一尺度的特征无法为网络结构提供更丰富的语义表示并且缺乏上下文信息,因此多尺度特征信息聚合非常重要。然而,使用多尺度卷积会增加计算量,不适用于小样本。为了降低网络复杂度,本文提出使用MSCA多尺度深度可分离卷积来提取和聚合多尺度特征信息,同时减少计算量。

底层特征提取通常指从原始图像中提取较低级别的特征,例如边缘和角点。高层特征提取则指从底层特征中提取更抽象的特征,例如纹理、形状和物体部件等。将这些不同层次的特征结合在一起,可以提高特征表达的能力,从而更好地表达数据。因此,本文将多层次特征进行信息融合,以提高表达能力。

综上所述,本文提出了基于度量学习的SDM-RNET模型,旨在通过以下3个方面的改进来优化关系网络:(1)增加网络深度;(2)采用多尺度深度可分离卷积替代普通卷积进行特征融合;(3)将高层和底层特征进行融合。

## 1 SDM-RNET 网络结构

### 1.1 SDM-RNET 整体框架

本文提出的基于度量学习的SDM-RNET模型整

体框架如图1所示,模型由4个阶段构成:第1阶段是基于Stochastic Depth的特征增强;第2阶段是基于MSCA<sup>[10]</sup>的多尺度特征融合;第3阶段是骨干网络多级特征融合;第4阶段是特征度量。其中基于Stochastic Depth特征增强、基于MSCA的多尺度特征融合以及骨干网络多级特征融合这3个部分构成特征提取阶段,用于图像特征提取,特征度量部分用于图像特征相似度度量。

基于Stochastic Depth的特征增强由多个残差块构成,在训练阶段,每个残差块以一定的概率被丢弃来防止过拟合;基于MSCA多尺度卷积融合是将两个MSCA多尺度深度可分离卷积应用到Stochastic Depth的后两个BasicBlock中;骨干网络多级特征融合是将Stochastic Depth的前两个BasicBlock的输出特征与经过MSCA处理的两个BasicBlock的输出特征进行拼接融合操作;特征度量骨干网络包括4个卷积块,其中前两个卷积块之后各有一个的最大池化层,用于进行图像特征提取。本文将详细介绍基于Stochastic Depth的特征增强,基于MSCA的多尺度特征融合,骨干网络多级特征融合以及特征度量的实现过程。

### 1.2 基于Stochastic Depth的特征增强

基于Stochastic Depth的特征增强采用多个残差块构成的设计是为了解决过拟合问题,并提高模型的泛化能力。过拟合是指在模型训练过程中,模型过度地适应了训练数据集的特点,导致在面对新数据时无法良好泛化的现象。这种情况往往发生在训练数据集较小的情况下,特别是在存在高度复杂的模型中。在图像分类任务中,由于类别特征的差异性以及数据的限制,小样本数据集可能无法充分涵盖全局图像分布,也容易导致过拟合问题。基于Stochastic Depth的特征增强对每个残差块都引入了丢弃机制。在训练阶段,随机地以一定的概率丢弃每个残差块,从而有效避免过拟合。通过丢弃残差块,模型被迫依赖于剩余的残差块进行学习,从而鼓励模型提取更鲁棒和泛化的特征。这种设计可以有效扩大模型的表示能力,并防止模型对训练数据的过度拟合,提高模型在未见样本上的分类性能。

基于Stochastic Depth的特征增强模块由两个阶段实现,第1阶段是基础网络的搭建,第2阶段是随机丢弃基础网络残差层的过程,具体实现如下。

(1)搭建基础网络的过程(残差块未丢弃)

因为小样本学习算法中的特征提取模块不宜过复杂,所以本文使用ResNet-18<sup>[11]</sup>来搭建。

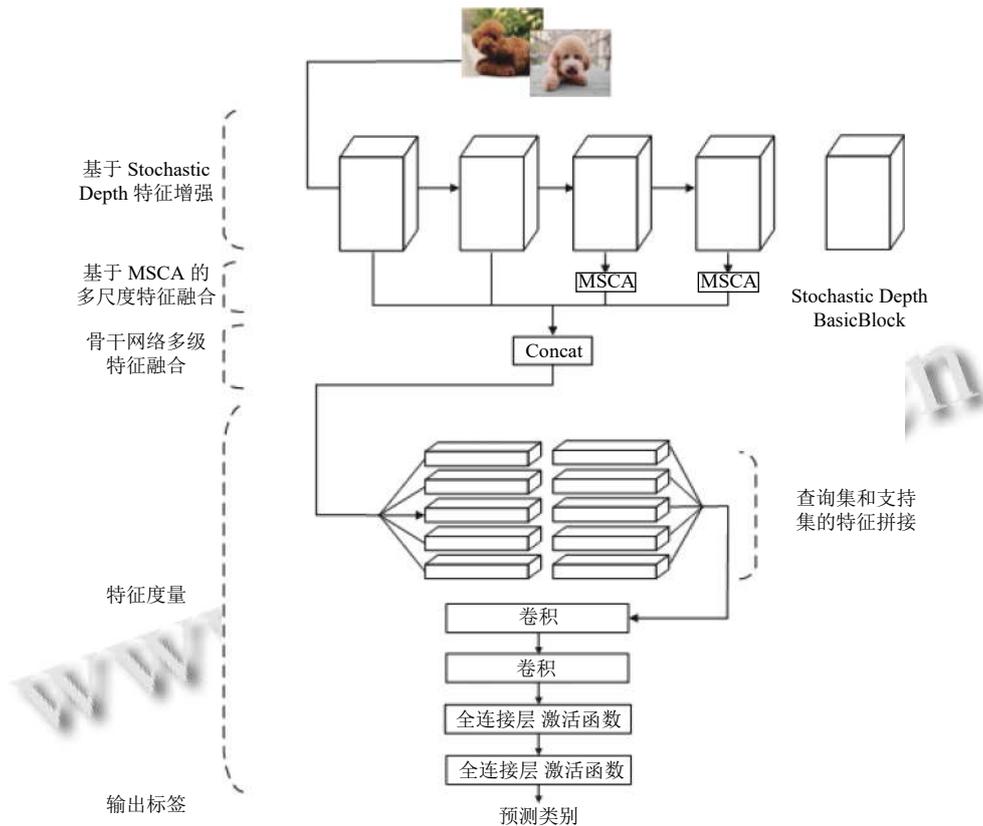


图1 SDM-RNET 模型整体框架图

ResNet-18 网络一共包含 5 个卷积层, 分别是 Conv1 层、Layer1 层、Layer2 层、Layer3 层、Layer4 层。Conv1 层包含 1 个通道数为 64, 卷积核大小为  $7 \times 7$  的卷积层, 图像经过 Conv1 卷积层后输出  $64 \times 112 \times 112$  大小的特征图。接着把 Conv1 层的输出作为 Layer1 的输入, Layer1 层包含 2 个 BasicBlock, 每 1 个 BasicBlock 包含 2 个通道数为 64, 卷积核大小为  $3 \times 3$  的卷积层和 1 个恒等映射, Layer2 层包含 2 个 BasicBlock, 每 1 个 BasicBlock 包含 2 个通道数为 128, 卷积核大小为  $3 \times 3$  的卷积层和 1 个恒等映射, Layer3 层包含 2 个 BasicBlock, 每 1 个 BasicBlock 包含 2 个通道数为 256, 卷积核大小为  $3 \times 3$  的卷积层和 1 个恒等映射, Layer4 层包含 2 个 BasicBlock, 每 1 个 BasicBlock 包含 2 个通道数为 512, 卷积核大小为  $3 \times 3$  的卷积层和 1 个恒等映射。

## (2) 随机丢弃残差块的过程

残差模块加深网络的同时不可避免造成网络的冗余, 参数量会大量增加。所以在完成 ResNet-18 的搭建后, 本文需要在模型训练过程对 ResNet-18 中的冗余残差模块进行随机丢弃来防止过拟合。整个过程可以看

作使用 dropout 的方法对残差块进行随机丢弃<sup>[12]</sup>, 这里只在训练阶段丢弃的残差层, 不包括测试阶段, 测试阶段本文使用的是完整的残差网络, 即残差模块未被丢弃。具体丢弃实现过程如下。

图 2 是原始的 ResNet 残差模块的结构, 其中  $f$  代表的是残差部分, 即权重层,  $id$  代表的是恒等映射, 把这两部分求和经过  $ReLU$  函数激活, 然后再输出, 这个过程可以用下面的式 (1) 来表示:

$$H_t = ReLU(f_t(H_{t-1}) + id(H_{t-1})) \quad (1)$$

在训练中, 如果一个特定的残差块被启用了, 那么它的输入就会同时流经恒等映射和权重层; 否则输入就只会流经恒等映射。在训练的过程中, 每一个残差模块层都有一个“生存概率”, 并且都会被任意丢弃。式 (2) 描述了 Stochastic Depth 在 ResNet 上随机丢弃一些残差层的过程。具体为, 随机网络其在训练时加入了一个随机变量  $b$ ,  $b$  的概率分布满足一个伯努利分布, 然后将  $f$  乘以  $b$ , 对残差部分做了随机丢弃。如果  $b = 1$ , 这个结构就是原始的 ResNet 结构, 而当  $b = 0$  时, 残差支路  $f$  便没有被激活, 整个结构退化为一个恒

等函数.

$$H_l = ReLU(b_l f_l(H_{l-1}) + id(H_{l-1})) \quad (2)$$

上面提到的  $b$  满足伯努利分布 (0-1) 分布, 它的取值只有 0 和 1 两种, 其中取 0 的概率为  $1-p$ , 取 1 的概率是  $p$ .  $p$  又被称为生存概率, 这个  $p$  即代表了  $b=1$  的可能性, 因此  $p$  的设置相当重要.  $p$  的设置主要有以下两种: 一种设置方式是将其设置为同一个值, 即  $p_l=p$ , 其中  $p_l$  代表每个残差块  $l$  的  $p$  参数取值. 另外一种设置

方式是将其设置为残差层数  $l$  的平滑函数. 从  $p_0=1$  线性递减到  $p_l=0.5$ , 一共有  $l$  个残差块. 具体式 (3) 如下:

$$p_l = 1 - \frac{l}{L}(1 - p_l) \quad (3)$$

本文选择了第 2 种设置方式, 即将线性衰减规律应用于每一层的生存概率的设置, 这是因为较早的层会提取低级特征, 而这些低级特征会被后面的特征应用, 因此前面的不应该频繁地被丢弃, 最终  $p$  产生的规则如图 3 所示.

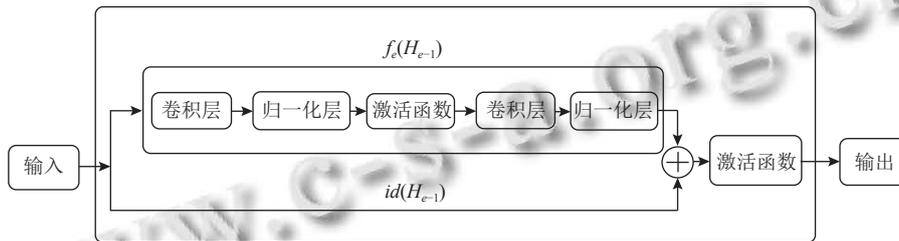


图 2 原始 ResNet 残差模块结构<sup>[13]</sup>

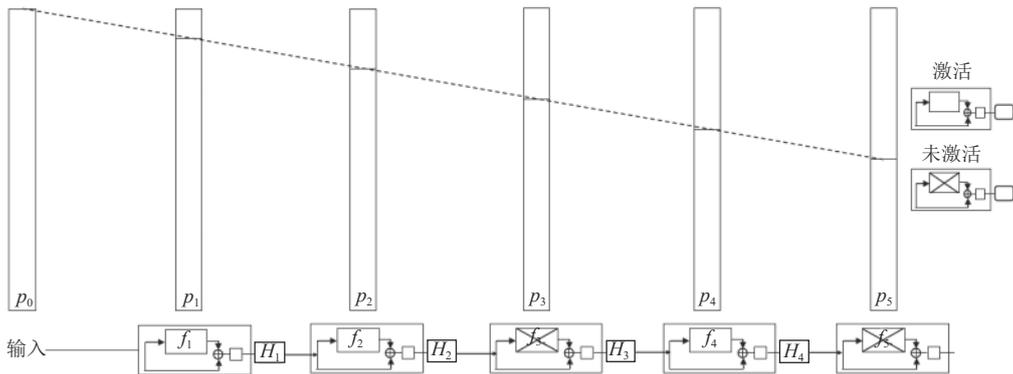


图 3 线性衰减规律

在测试过程中, 所有的残差块都将保持被激活状态, 并且残差块的权重将根据其在训练中的生存概率进行调整.

通过随机丢弃残差层, 形成 Stochastic Depth-18 网络, 这里 18 层不代表训练阶段丢弃的层, 指的是测试过程中的 18 层, 因为测试过程中的网络并不丢弃残差层. 为了方便后续的特征度量阶段进行卷积操作, 本文除去 Stochastic Depth-18 网络最后的平均池化层和全连接层, 形成以 Stochastic Depth-17 为特征提取模块的骨干网络.

以上便是基于 Stochastic Depth 的特征增强具体实现过程. 下文将具体介绍在 Stochastic Depth 基础上的

MSCA 特征融合的实现细节.

### 1.3 基于 MSCA 的特征融合

#### (1) 多尺度特征融合<sup>[14]</sup>

当前的图像分类任务中, 一个常见的问题是如何提高模型的感受野, 即通过一定的上下文信息来增强对图像特征的理解和判断.

尽管基于 Stochastic Depth 的特征增强简化并优化了残差网络的训练过程, 但是它使用的是单一尺度的卷积核进行卷积计算. 由于单一卷积只能在固定尺度下提取特征, 感受野单一, 无法捕捉到不同尺度下的特征. 这也导致它无法获取更广泛的上下文信息. 当处理尺度变化较大的图像时, 使用单一卷积操作可能效果

不佳<sup>[15]</sup>。相比之下,使用多尺度特征融合的方法,可以使用不同尺度的卷积核进行卷积操作,从而提取不同尺度的特征并进行融合。这种方法可以更充分地利用图像的多尺度信息,并提高模型的分类准确率。

因此,本文考虑使用多尺度卷积代替单一尺度卷积来对 Stochastic Depth 的输出特征进行多尺度特征融合。多尺度特征融合网络常见的有两种:第1种是并行多分支结构<sup>[16]</sup>,第2种是串行的跳层连接结构,都是在不同的感受野下进行特征提取。本文选用并行多分支特征结构进行特征融合。

并行多分支多尺度特征融合结构是一种常用的深度网络架构,用于图像分类任务中的特征提取和融合。其主要特点是将不同尺度的图像特征,在多个分支中进行提取和融合,从而提高模型的特征学习和分类能力。

具体而言,这种架构通常包括一个主干网络,以及多个并行的支路。其中,主干网络负责从原始图像像素级别开始提取较低层的特征信息;而每个支路则负责在其他特定的尺度下提取更加抽象和高级别的特征信息。在支路提取完各自的特征后,这些特征将会通过特定的方法进行融合,从而生成用于分类的最终特征表示。

与传统的网络模型相比,并行多分支多尺度特征融合结构具有多个优势。首先,它可以提取多个尺度下的图像特征,更全面地描述了图像的细节和空间结构,提高了模型的识别率和准确性。其次,该结构的并行设计可以大幅缩短训练时间,提高了模型训练的效率。此外,该结构的分支设计也增加了模型的可解释性,使得对图像的理解更加直观和易于理解。

图4为并行多分支多尺度特征融合结构示例,输入特征图进入到3个并行分支中,通过3个分支中的不同尺度的卷积进行特征提取,然后将3个分支提取到的进行融合,作为下一层的特征输入。

### (2) 深度可分离卷积

通过多尺度卷积的特征融合可以获取更多的特征信息,但是多尺度卷积需要对输入数据进行多次卷积计算,这会导致计算复杂度高、网络的参数量过大,不利于模型的训练和优化,同时也容易出现过拟合的问题,不适用于基于度量学习的小样本图像分类研究学习。所以本文考虑利用深度可分离卷积代替普通卷积来对 Stochastic Depth 输出的特征进行多尺度特征融

合。相比于传统卷积,深度可分离卷积可以更加高效地进行卷积运算,并且能够大幅度减少计算和参数量。

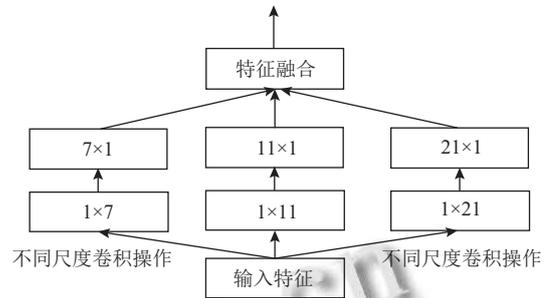


图4 并行多分支多尺度特征融合结构

深度可分离卷积是将卷积分为深度卷积和分离卷积两部分,能够在不改变特征图大小的情况下减少计算量。深度卷积是逐通道进行的,只对每个通道进行卷积操作。分离卷积则是使用 $1 \times 1$ 卷积进行通道上信息的整合。最终,经过深度卷积和分离卷积两部分操作,得到的特征图大小与标准卷积相同,但计算量却大大减少。

### (3) MSCA 多尺度深度可分离卷积的特征融合

综上,为了获取更大范围内的上下文信息的同时控制网络的复杂度,本文将(1)多尺度卷积融合和(2)深度可分离卷积进行结合,即使用 MSCA 多尺度深度可分离卷积来对 Stochastic Depth 输出的特征进行多尺度特征融合操作。MSCA 具体结构如图5所示。

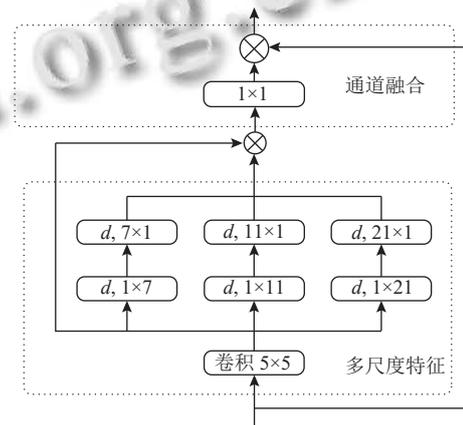


图5 MSCA 结构图

MSCA (multiscale spatial and channel attention) 多尺度卷积是一种用于特征提取和融合的方法,它由3个主要部分组成:用于聚集局部信息的深度卷积、用于捕获多尺度上下文的多分支深度卷积和用于聚合信息的 $1 \times 1$ 卷积。首先,在 MSCA 多尺度卷积中,深度卷

积用于聚集局部信息. 它通过卷积操作, 从输入特征中提取出局部感知能力. 这有助于捕捉特征的细节和局部结构. 多分支深度卷积用于捕获多尺度上下文信息. 为了获得不同尺度的特征表示, MSCA 采用了 3 种不同尺度的深度可分离卷积核, 分别为  $7 \times 7$ 、 $11 \times 11$  和  $21 \times 21$ . 每个尺度的卷积核都负责捕捉不同尺度的特征信息. 为了减少计算量, 每个卷积核还被分解成两个更小的卷积核, 例如将  $7 \times 7$  的卷积核分解为  $1 \times 7$  和  $7 \times 1$  的卷积核. 分解后的效果与原始尺寸的卷积核相当, 但计算量更小. 通过这样的多尺度卷积操作, MSCA 能够有效地捕捉图像中的多尺度信息. 随后,  $1 \times 1$  卷积用于聚合信息和重新加权. 它的输出直接作为注意力

权重, 重新加权 MSCA 的输入特征. 这个重加权的過程可以使 MSCA 更加关注重要的特征和通道, 从而增强特征表达能力. 在 MSCA 中, 多个尺度的特征图通过相加叠加的方式进行融合. 叠加后的特征图再通过  $1 \times 1$  卷积进行信息聚合, 然后作用回原始的特征图上. 这样的特征融合过程可以增强特征的表达能力和多样性.

考虑到增加多块 MSCA 可能会导致参数过多、计算量过大的问题, 该方法在 Stochastic Depth 的最后两个 BasicBlock (BasicBlock3 和 BasicBlock4) 后加入 MSCA 来处理输出的特征. 这样的设计既能充分利用 MSCA 的优势, 又能有效控制模型的复杂性, 具体如图 6 所示.

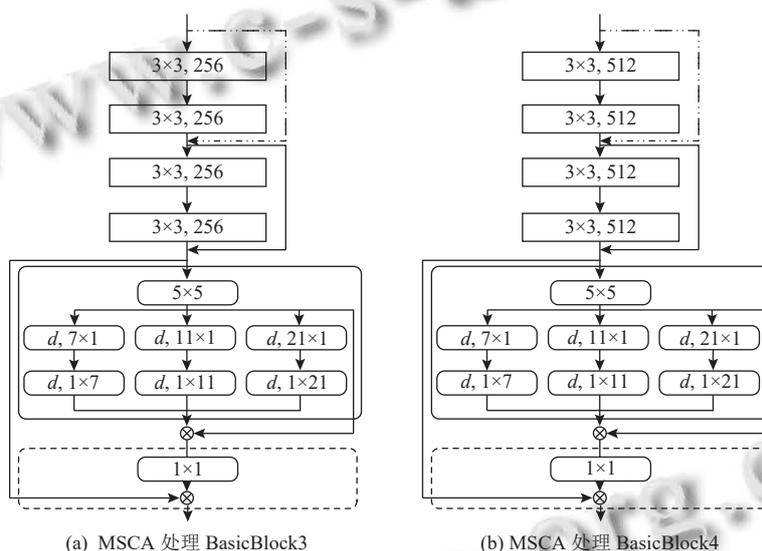


图 6 使用 MSCA 处理 BasicBlock

经过 BasicBlock 输出的特征进入到 MSCA 多尺度深度可分离卷积阶段, 首先通过  $5 \times 5$  的深度卷积进行特征提取, 深度卷积用于聚合局部信息, 再经过 3 个并行分支中, 进行多尺度深度卷积特征提取来捕获多尺度上下文信息, 最后使用卷积  $1 \times 1$  卷积进行通道维度进行相关性建模. 由此完成本文提出的 MSCA 多尺度深度可分离卷积的特征再增强过程.

#### 1.4 骨干网络多级特征融合

骨干网络多级特征融合将经过 Stochastic Depth 处理的前两个 BasicBlock 的输出特征与经过 MSCA 处理的两个 BasicBlock 的输出特征进行拼接融合操作. 这种融合操作可以将不同层级、不同尺度的特征信息有机地结合在一起, 进一步增强模型的特征表示能力.

通过多级特征融合, 模型可以更好地理解图像的多尺度特征, 有利于提高图像分类任务的性能. 这种融合方法能够更好地捕捉图像的多尺度信息, 有效提升了模型的性能.

特征在经过基于 MSCA 多尺度深度可分离卷积的特征融合后, 本文拟继续通过聚合高层特征和底层特征来提高模型对图像的理解.

高层特征通常是通过底层特征进行组合和抽象而得到的. 高层特征<sup>[17]</sup>具有更高的语义信息, 能够描述物体的整体形状、结构、物体部位、动作等高层次的语义概念. 高层特征通常是全局特征, 能够对整个图像进行全局描述, 例如通过将底层特征组合成物体的轮廓、部位等来描述物体. 将底层特征和高层特征聚合在

一起可以提高计算机视觉任务的准确性和性能,从而提高模型在各种计算机视觉任务上的准确性和鲁棒性<sup>[18]</sup>。

本文旨在通过聚合底层和高层特征,提高模型对图像的理解能力。具体地说,本文将 Stochastic Depth 的 BasicBlock1 和 BasicBlock2 的输出特征与经过 MSCA 处理后的 BasicBlock3 和 BasicBlock4 的输出特征拼接起来,实现底层和高层特征的融合,从而提高特征表达的能力,以更好地表达数据。最后,本文将这些特征输入到特征度量阶段。具体的结构图如图 7 所示。这些步骤完成了骨干网络多级特征融合的阶段。

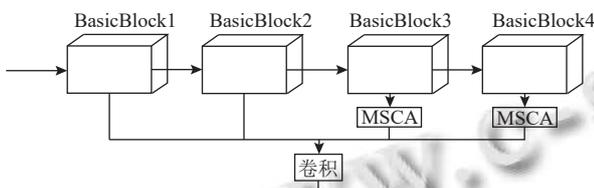


图 7 底层特征和高层特征信息融合过程

总的来说,本文提出的基于 Stochastic Depth、基于 MSCA 的多尺度特征融合以及骨干网络多级特征融合构成的特征提取模块有以下几方面的优势。

(1) 更强的表达能力减少过拟合。可以从输入数据中学习更抽象和高级的特征,这些特征可以更好地区分不同的类别,同时将这些不同层次的特征组合在一起可以提高特征表达的能力,从而更好地表达数据。

(2) 提高感受野。多尺度特征融合可以通过合并不同尺度的特征信息,增加模型对输入图像的感受野,从而更好地捕捉图像中的上下文信息。

(3) 提高鲁棒性。多尺度特征融合可以通过利用多个尺度的特征信息,降低对输入图像中物体尺寸和形状的敏感性,提高模型的鲁棒性。

## 2 SDMSCA-RNE 网络算法流程

本文提出的 SDM-RNET 模型分为以下几个阶段:数据预处理、图像特征提取、特征相似度测量并输出最终结果。具体阶段如图 8 所示,第 1 阶段为数据预处理,其目的是对原始图像进行清洗、转换、增强和归一化等处理,以提高图像质量和可用性;第 2 阶段是特征提取阶段,这个阶段包括基于 Stochastic Depth 的特征增强和基于 MSCA 的特征融合两个模块,用于对查询集和支持集图像进行提取;第 3 个阶段为特征度量,用于将提取后的查询集和支持集特征进行拼接,再测

量查询集特征图和支持集特征图之间的关系,最后得到一个范围为 0-1 的标量,表示查询集和支持集之间的相似度,称为关系得分。并将得分最高的支持集样本类别作为查询集的分类结果。

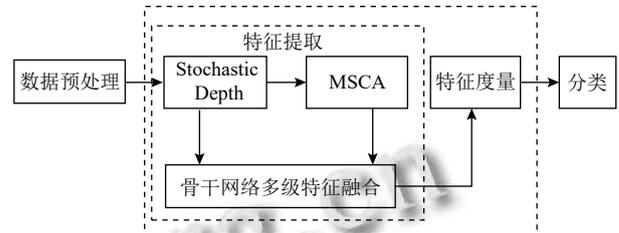


图 8 SDM-RNET 算法流程

## 3 实验结果与分析

### 3.1 实验环境

本文在 mini-ImageNet<sup>[19]</sup>, Omniglot<sup>[20]</sup>和 RP2K<sup>[21]</sup>这 3 个数据集上评估了 SDM-RNET 网络。所有实验基于 PyTorch 框架,Python 3.8 语言实现。在 64 位 Ubuntu 16.04 LTS 操作系统上进行, GPU 采用 NVIDIA P100。采用 PyTorch 框架进行评估。实验的软件和硬件平台的实现信息列在表 1 中。

表 1 软硬件环境

环境	配置
CPU	Intel(R) Xeon(R) CPU E5-2630 v4
GPU	NVIDIA P100
显存	16 GB
内存	64 GB
操作系统	Ubuntu 16.04
开发语言	Python 3.8

### 3.2 参数设置

本文使用 Adam 优化器从头到尾地训练模型。通过 episode 训练机制在 SDM-RNET 模型上进行学习。使初始学习率设置为 0.001,并且每过 2000 个 episode 学习率减少一半。在 mini-ImageNet 数据集上的训练轮次设置为 500000,即 episode=500000; RP2K 数据集上, episode=100000; Omniglot 数据集上, episode=300000。测试过程, episode=600,取平均值作为测试结果。

### 3.3 对比试验

本文将 SDM-RNET 模型与近些年的主流小样本学习模型进行比较,比较模型包括 Matching Net<sup>[19]</sup>、Meta Nets<sup>[22]</sup>、Meta-learn LSTM<sup>[23]</sup>、MAMAL<sup>[24]</sup>、Siamese

Nets<sup>[25]</sup>、Prototypical Net<sup>[26]</sup>、Relation Network<sup>[27]</sup>和 Baseline+<sup>[28]</sup>、CMD-ML<sup>[29]</sup>, 实验结果如表 2 所示。

表 2 SDM-RNET 的对比实验结果 (%)

数据集	模型	5-way accuracy	
		1-shot	5-shot
mini-ImageNet数据集	Matching Nets	43.56	55.31
	Meta Nets	49.21	—
	Meta-learn LSTM	43.33	60.60
	MAMAL	48.70	63.11
	Prototypical Nets	49.42	68.20
	Relation Network	50.44	65.32
	Baseline++	48.24	66.43
	CMD-MLA	51.01	68.72
	<b>SDM-RNET (Ours)</b>	<b>53.07</b>	<b>68.78</b>
RP2K 数据集	MAMAL	93.04	95.18
	Prototypical Nets	95.03	97.54
	Relation Network	95.64	98.82
	Baseline++	93.89	97.25
	CMD-MLA	94.57	96.98
		<b>SDM-RNET (Ours)</b>	<b>97.03</b>
Omniglot数据集	Siamese Nets	97.3	98.4
	Matching Network	98.1	98.9
	MAMAL	98.7	99.9
	Prototypical Nets	98.8	99.7
	Relation Network	99.6	99.8
	Baseline++	99.6	99.7
	CMD-MLA	99.1	99.6
	<b>SDM-RNET (Ours)</b>	<b>99.8</b>	<b>99.9</b>

通过表 2 可以看出, 在 mini-ImageNet 数据集上, 与 Relation Network 关系网络相比, SDM-RNET 在 5-way 1-shot 和 5-way 5-shot 分类任务的准确率分别提高了 2.63% 和 3.46%; 在 RP2K 数据集上, 与 Relation Network 关系网络相比, SDM-RNET 在 5-way 1-shot 和 5-way 5-shot 分类任务的准确率分别提高了 1.39% 和 0.52%; 在 Omniglot 数据集上, 与 Relation Network 相比, SDM-RNET 模型在 5-way 1-shot 和 5-way 5-shot 分类任务的准确率分别提高了 0.2% 和 0.1%。这表明 SDM-RNET 模型在这些任务上具有更好的分类效果, 同时也表明该模型在 3 个数据集上都得到了一定程度的提升。

由此说明, 在关系网络的特征提取阶段中使用随机深度网络代替关系网络中原来的 4 个卷积块来加深网络, 同时在随机深度网络的基础上融合多尺度可分离卷积 MSCA, 最后将骨干网络的多级特征进行融合, 通过这 3 个改进方法, 极大地提高了模型的泛化能力, 这里的泛化能力指的是模型可以对未见过的图像类别进行较好的分类。因为预测的是未见过的类别图像, 所以模型分类准确率的提高意味着模型的泛化能力就

更强。

### 3.4 消融实验

为了证明 SDM-RNET 模型的每一个模块的有效性, 本文对 SDM-RNET 模型进行了消融实验。具体如表 3 所示, RN 代表关系网络; H1 代表将 Stochastic Depth 替换关系网络的 4 个卷积块; H2 代表在 H1 的基础上, 在 Stochastic Depth 后两个 BasicBlock 中分别加入 MSCA 多尺度深度可分离卷积进行特征融合; H3 代表在 H1 的基础上将 Stochastic Depth 的多级特征进行融合 (4 个 BasicBlock 的输出特征进行融合); H4 代表的是本文所提出的 SDM-RNET 模型, 即先将 Stochastic Depth 替换关系网络的 4 个卷积块, 然后在 Stochastic Depth 后两个 BasicBlock 加入 MSCA 多尺度深度可分离卷积进行特征融合, 最后将 Stochastic Depth 前两个 BasicBlock 的输出特征与后两个融合了 MSCA 多尺度深度可分离卷积的 BasicBlock 的输出特征进行不同层级的特征融合。

表 3 SDM-RNET 模型进行消融实验 (%)

数据集	实验	Stochastic Depth	MSCA	骨干网络多级特征融合	5-way accuracy	
					1-shot	5-shot
mini-ImageNet 数据集	RN	—	—	—	50.44	65.32
	H1	√	—	—	51.92	66.83
	H2	√	√	—	52.60	67.72
	H3	√	—	√	52.35	67.26
	H4	√	√	√	<b>53.07</b>	<b>68.78</b>
RP2K 数据集	RN	—	—	—	95.64	98.82
	H1	√	—	—	95.89	99.08
	H2	√	√	—	96.55	98.94
	H3	√	—	√	96.78	99.12
	H4	√	√	√	<b>97.03</b>	<b>99.34</b>
Omniglot 数据集	RN	—	—	—	99.6	99.8
	H1	√	—	—	99.6	99.7
	H2	√	√	—	99.5	99.8
	H3	√	—	√	99.7	99.9
	H4	√	√	√	<b>99.8</b>	<b>99.9</b>

### 3.5 综合分析

通过 SDM-RNET 模型在 mini-ImageNet、RP2K、Omniglot 这 3 个数据集上的实验, 可以发现将随机深度网络和多尺度卷积结合代替关系网络特征提取模块的 4 个卷积块后, 网络在小样本图像分类任务上的分类准确率有了较大的提高, 验证了 SDM-RNET 的有效性。Stochastic Depth 可以通过随机地去掉一些冗余的层次来优化残差网络的训练过程, 使得网络变得更加简单。提取到更准确的支持集图像特征和查询集图像特征从而进一步提高查询集的分类判断。实验证明, 随

机深度在 SDM-RNET 模型中发挥了重要的作用, 主要优势如下.

(1) 计算效率: 通过在训练期间随机丢弃层, 网络变得更小, 计算效率更高. 这意味着网络可以更快地训练, 从而更容易实验不同的模型.

(2) 改进的泛化: 通过随机删除层, 网络将接触到更广泛的特征, 这有助于提高其泛化到看不见的数据的能力.

(3) 它提供了图像特征的多尺度和组合表示, 这比仅只关注局部特征的传统特征提取方法更加可靠和准确. MSCA 可以捕获图像中的局部和全局特征.

(4) MSCA 同时又具有注意力的优势, 它可以使网络更多地关注在有效信息上, 而非提炼出无效的信息, 减少了网络的复杂度和参数量.

总之, 融合多尺度卷积注意力的随机深度网络作为特征提取模块可以显著提高小样本图像分类的准确性. 融合多尺度卷积注意力的随机深度网络提供了更强大、更全面的图像特征表示, 从而改善了结果.

#### 4 结论与展望

本文提出了融合随机深度网络和多尺度卷积注意力机制的小样本图像分类模型 SDM-RNET. 首先分析了关系网络可改进的方向, 从改进特征提取模块角度出发, 通过改进关系网络的特征提取模块来提高网络的特征提取能力. 对此, 本文提出了两个创新点: 其一, 提出使用随机深度网络 Stochastic Depth-17 作为特征提取模块, 以加深网络深度, 提高特征提取能力. 其二, Stochastic Depth-17 融入多尺度卷积 MSCA 模块, 获取多尺度信息的同时尽可能减少参数量, 从而进一步提升网络的特征提取能力, 获取更加准确的图像特征. 最后为了验证 SDM-RNET 模型的有效性、鲁棒性, 本文还分别在 mini-ImageNet、Omniglot、RP2K 数据集上进行了大量的实验, 实验证明了 SDM-RNET 小样本图像分类任务上的有效性, 对比关系网络和其他经典的小样本学习算法, 其分类效果提高了不少.

随着主动学习的兴起, 后续本研究将把主动学习应用到模型中来, 主动学习通常基于不确定性采样 (uncertainty sampling) 的思想, 即从未标注的样本中选择出对模型预测结果不确定性最大的样本, 让人类标注, 然后将这些标注数据用于重新训练模型. 这样做可以有效地减少人类标注数据量, 并且提高模型的性能

和泛化能力.

#### 参考文献

- 1 Mai Z, Li RW, Jeong J, *et al.* Online continual learning in image classification: An empirical survey. *Neurocomputing*, 2022, 469: 28–51. [doi: [10.1016/j.neucom.2021.10.021](https://doi.org/10.1016/j.neucom.2021.10.021)]
- 2 Liu YH, Gu JT, Goyal N, *et al.* Multilingual denoising pre-training for neural machine translation. *Transactions of the Association for Computational Linguistics*, 2020, 8: 726–742. [doi: [10.1162/tacl\\_a\\_00343](https://doi.org/10.1162/tacl_a_00343)]
- 3 Dhar A, Mukherjee H, Dash NS, *et al.* Text categorization: Past and present. *Artificial Intelligence Review*, 2021, 54(4): 3007–3054. [doi: [10.1007/s10462-020-09919-1](https://doi.org/10.1007/s10462-020-09919-1)]
- 4 Wang CH, Tang Y, Ma XT, *et al.* Fairseq S2T: Fast speech-to-text modeling with Fairseq. *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing: System Demonstrations*. Suzhou: Association for Computational Linguistics, 2020. 33–39.
- 5 Lake BM, Salakhutdinov R, Gross J, *et al.* One shot learning of simple visual concepts. *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society*. Boston: The Cognitive Science Society, 2011. 2568–2573.
- 6 Li ZJ, Hu ZP, Luo WW, *et al.* SaberNet: Self-attention based effective relation network for few-shot learning. *Pattern Recognition*, 2023, 133: 109024. [doi: [10.1016/j.patcog.2022.109024](https://doi.org/10.1016/j.patcog.2022.109024)]
- 7 葛轶洲, 刘恒, 王言, 等. 小样本困境下的深度学习图像识别综述. *软件学报*, 2022, 33(1): 193–210. [doi: [10.13328/j.cnki.jos.006342](https://doi.org/10.13328/j.cnki.jos.006342)]
- 8 赵凯琳, 靳小龙, 王元卓. 小样本学习研究综述. *软件学报*, 2021, 32(2): 349–369. [doi: [10.13328/j.cnki.jos.006138](https://doi.org/10.13328/j.cnki.jos.006138)]
- 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. *自动化学报*, 2021, 47(2): 297–315. [doi: [10.16383/j.aas.c190720](https://doi.org/10.16383/j.aas.c190720)]
- 10 Guo MH, Lu CZ, Hou QB, *et al.* SegNeXt: Rethinking convolutional attention design for semantic segmentation. *arXiv: 2209.08575*, 2022.
- 11 Huang G, Sun Y, Liu Z, *et al.* Deep networks with stochastic depth. *Proceedings of the 14th European Conference on Computer Vision*. Amsterdam: Springer, 2016. 646–661.
- 12 谭涛. 基于卷积神经网络的随机梯度下降优化算法研究 [硕士学位论文]. 重庆: 西南大学, 2021. [doi: [10.27684/d.cnki.gxndx.2020.002742](https://doi.org/10.27684/d.cnki.gxndx.2020.002742)]
- 13 He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for

- image recognition. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 770–778.
- 14 史静, 朱虹. 多特征融合的场景图像分类算法. 计算机系统应用, 2018, 27(5): 171–175. [doi: 10.15888/j.cnki.csa.006343]
- 15 回文君. 基于浅层特征的弱监督目标定位研究 [硕士学位论文]. 秦皇岛: 燕山大学, 2022. [doi: 10.27440/d.cnki.gysdu.2022.001486]
- 16 孙有晓, 冯锐成, 关伟鹏, 等. 基于图像超分辨率的多尺度特征交互传播研究. 集成技术, 2022, 11(2): 41–54.
- 17 蔡国永, 储阳阳. 基于双注意力多层特征融合的视觉情感分析. 计算机工程, 2021, 47(9): 227–234. [doi: 10.19678/j.issn.1000-3428.0058303]
- 18 赵志成, 罗泽, 王鹏彦, 等. 基于深度残差网络图像分类算法研究综述. 计算机系统应用, 2020, 29(1): 14–21. [doi: 10.15888/j.cnki.csa.007243]
- 19 Vinyals O, Blundell C, Lillicrap T, *et al.* Matching networks for one shot learning. Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona: Curran Associates Inc., 2016. 3637–3645.
- 20 Lake BM, Salakhutdinov R, Tenenbaum JB. Human-level concept learning through probabilistic program induction. Science, 2015, 350(6266): 1332–1338. [doi: 10.1126/science.aab3050]
- 21 Peng JT, Xiao C, Li YF. RP2K: A large-scale retail product dataset for fine-grained image classification. arXiv:2006.12634, 2020.
- 22 Munkhdalai T, Yu H. Meta networks. Proceedings of the 34th International Conference on Machine Learning. Sydney: JMLR.org, 2017. 2554–2563.
- 23 Ravi S, Larochelle H. Optimization as a model for few-shot learning. Proceedings of the 5th International Conference on Learning Representations. Toulon: OpenReview.net, 2017.
- 24 Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks. Proceedings of the 34th International Conference on Machine Learning. Sydney: JMLR.org, 2017. 1126–1135.
- 25 Koch G, Zemel R, Salakhutdinov R. Siamese neural networks for one-shot image recognition. ICML Deep Learning Workshop, 2015, 2(1): 1–8.
- 26 Cai Q, Pan YW, Yao T, *et al.* Memory matching networks for one-shot image recognition. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 4080–4088.
- 27 Sung F, Yang YX, Zhang L, *et al.* Learning to compare: Relation network for few-shot learning. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 1199–1208.
- 28 Chen WY, Liu YC, Kira Z, *et al.* A closer look at few-shot classification. Proceedings of the 7th International Conference on Learning Representations. New Orleans: OpenReview.net, 2019.
- 29 孙牧野. 基于度量学习的小样本图像分类技术研究 [硕士学位论文]. 西安: 西安电子科技大学, 2020.

(校对责编: 牛欣悦)