

# 基于多尺度特征和元学习的图像识别算法研究

黄勇杰 高乐\* 杨田

(五邑大学智能制造学部 广东 江门 529000)

**摘要** 在计算机视觉与图像识别技术中,随着输入分辨率的变化,卷积神经网络方法的识别效果也不同。多尺度特征学习可以将图像的精度和细节结合在一起,结合图像的多个尺度信息进行分析。元学习让计算机模拟人的大脑,学习如何去学习,可以更高效灵活地实现图像分类。因此,结合多尺度特征和元学习进行图像识别算法研究具有较高的研究价值。通过膨胀卷积产生了不同分辨率的输入图像;使用元学习产生不同输入分辨率下神经网络的卷积权重;对于不同输入分辨率的模型使用知识蒸馏。

**关键词** 图像识别 多尺度特征学习 元学习 知识蒸馏 人工智能

中图分类号 TP391

文献标志码 A

DOI:10.3969/j.issn.1000-386x.2024.08.029

## IMAGE RECOGNITION ALGORITHMS WITH MULTISCALE FEATURES AND META-LEARNING

Huang Yongjie Gao Le\* Yang Tian

(Department of Intelligent Manufacturing, Wuyi University, Jiangmen 529000, Guangdong, China)

**Abstract** In computer vision and image recognition technology, the recognition effect of convolutional neural network methods varies with the change of input resolution. Multiscale feature learning can combine the accuracy and details of an image, combining the information of multiple scales of the image for analysis. Meta-learning allows computers to simulate the human brain and learn how to learn, which allows for more efficient and flexible image classification. Therefore, combining multi-scale features and meta-learning for image recognition algorithm research has high research value. In this study, we generated input images with different resolutions by inflated convolution, used meta-learning to generate convolutional weights of neural networks with different input resolutions, and used knowledge distillation for models with different input resolutions.

**Keywords** Image recognition Multi-scale feature learning Meta-learning Knowledge distillation Artificial intelligence

## 0 引言

随着计算机视觉与图像识别技术的迅猛发展,神经卷积网络已经充分展示了它的优势,但是它的识别效果会随着输入分辨率的不同而变化。传统的神经网络一般由多个卷积模块组成,卷积模块包含了卷积、归一化、激活等操作,然后通过全局池化层大大地减少了参数量,最后通过全连接层得到结果。

不同分辨率的图片被神经卷积网络模型处理后会得到不同大小的特征图像,由于进行了全局池化操作,最终得到的输出数据的维度是一致的,但是最后所得到的识别效果是大不相同的。目前已有许多学者在多尺度特征对图像识别的影响中展开了大量的研究,并获得了较好的成果<sup>[1-5]</sup>。多尺度特征学习的有效性在各种不同算法中已经被证实。FPN<sup>[6]</sup>使用了特征金字塔,小尺度特征图像通过上采样的方法获得与大尺度特征图像一样的尺寸,多种尺度结合在一起进行预测,在目

标检测和语义分割中取得了很好的效果。DeepLab<sup>[7]</sup>使用了不同的膨胀系数进行并行化计算,得到不同尺度的特征图像,然后再结合在一起。PSPNet<sup>[8]</sup>使用了不同大小的卷积核进行并行计算。大尺度能把握整体的信息,小尺度更能把握细节信息。多种尺度结合起来,对算法的性能有很大的提升。本文使用了膨胀卷积方法得到多个不同尺度大小的图像,并对其进行研究。

元学习是近年来计算机视觉研究的热点之一,在众多的方法研究中,有基于记忆的方法,利用以往训练所得到的经验进行学习,通过添加 Memory 模块来实现<sup>[9-10]</sup>;有基于预测梯度的方法,利用以往的任务学习如何预测梯度<sup>[11]</sup>;有利用注意力机制的方法,利用以往的任务来训练一个 Attention 模型,在执行新的任务时,可以将注意力放在最重要的地方<sup>[12]</sup>;也有借鉴 LSTM 的方法,参考 LSTM 制定一个更新机制,根据当前的网络参数,得到新的更新参数<sup>[13]</sup>;还有预测损失函数的方法,构造一个模型通过以往学到的经验来学习如何预测损失函数等<sup>[14]</sup>。不同领域的科研工作者在元学习方法的应用上也进行着不断的探索和深入研究<sup>[15-17]</sup>。

在本文中,为不同分辨率的输入提供了单独的网络进行训练,利用元学习从不同的尺度信息中学习得到不同网络的权重。每个网络的归一化层都是独立的,归一化层的参数由各自网络进行维护,不参与共享。各个网络进行全局池化操作后再通过公用的全连接层得到输出结果。各个网络共享全连接层参数,这样可以更好整合从不同分辨率网络中学习到的信息。本研究为每个网络的输出赋予不同的权重从而得到最后的输出结果。另外,本文采用了知识蒸馏的技术,利用大分辨率所获得的知识提升小分辨率的识别效果。研究在 Corel5k 上进行了图像识别实验,实验结果证明了该方法有巨大的潜力。

## 1 相关方法简介

### 1.1 多尺度特征学习

多尺度特征学习先通过使用不同的预处理方法或者尺寸不同的卷积核产生不同分辨率的图像,再结合不同分辨率的图像信息进行分析,最后通过分类算法完成图像识别任务。在特征提取的过程中,随着卷积神经网络的加深,特征图像越来越小,特征信息会丢失,结合不同分辨率的图像信息进行分析,可以减少特征信息的丢失。

### 1.2 元学习

元学习,也称为学习如何去学习。在人工智能技术迅速发展的今天,元学习的重要性逐渐被凸显出来。计算机不具备跟人类一样高效的学习能力,而深度学习所展现出来的优异效果往往是通过数以百万的数据集不停训练得到的。同时由于人类的生产生活需要,人工智能所面临的任务也日益复杂。为了解决这些问题,人们对人工智能模型的要求也越来越高。从机器学习到深度学习,再到深度强化学习。

人工智能是模拟人的大脑去学习,那人工智能就应该像人的大脑一样具备高效的学习能力,通过以往学到的知识主动学习未知的知识,学习如何去学习,学会思考怎样学习效果才是最好的<sup>[18]</sup>。元学习就是在这环境下被提出了。元学习的目标是在两个耦合层次下推进学习过程,利用一个基本学习算法在一组辅助任务中提取可转移的先验知识,让元学习算法结合先验知识更好地应对未知任务。通过元学习学习网络的权重也是一个热门的研究方向。在本文的研究方法中,我们为输入数据设定一个尺度知识,通过该尺度知识生成神经卷积网络的权重。

### 1.3 知识蒸馏

知识蒸馏的主要目的就是压缩神经网络模型,通过一个已经训练好的大型神经网络模型一步步指导另外一个相对比较小的神经网络模型如何学习,最终得到一个与大型神经网络性能相近的小型神经网络。具体实现方法就是将知识从教师网络提取到学生网络。教师网络对于学生网络有着指导意义,可以帮助学生网络更好地提高性能,引导学生网络的优化。通常是先训练一个教师网络,通过教师网络的输出结果和真实标签来辅助学生网络训练。知识蒸馏可以缩小神经网络的规模,并且缩小后神经网络的性能与缩小前相近;也可以将多个神经网络转移至一个神经网络中,得到的神经网络性能也十分优异。近年来人们对知识蒸馏也有很多的研究<sup>[19-21]</sup>。本文将知识蒸馏技术引入到研究方法中,在训练的过程中大分辨率的输入图像指导小分辨率图像进行学习,知识蒸馏过程在各个不同尺度的网络之间产生,更有利于各个网络的性能优化。

## 2 网络结构

图 1 是研究方法的流程,其中: $l_1, l_2, \dots, l_k$  是使用不同膨胀系数卷积后得到的输入图,  $\eta_1, \eta_2, \dots, \eta_k$  是膨胀卷积后图像对应的尺度标量,然后作为元学习器  $\Phi_1, \Phi_2, \dots, \Phi_l$  的输入,生成卷积核。归一化层进行私

有化处理,卷积层和全连接层各个网络共享,知识蒸馏以自下而上的方式进行。

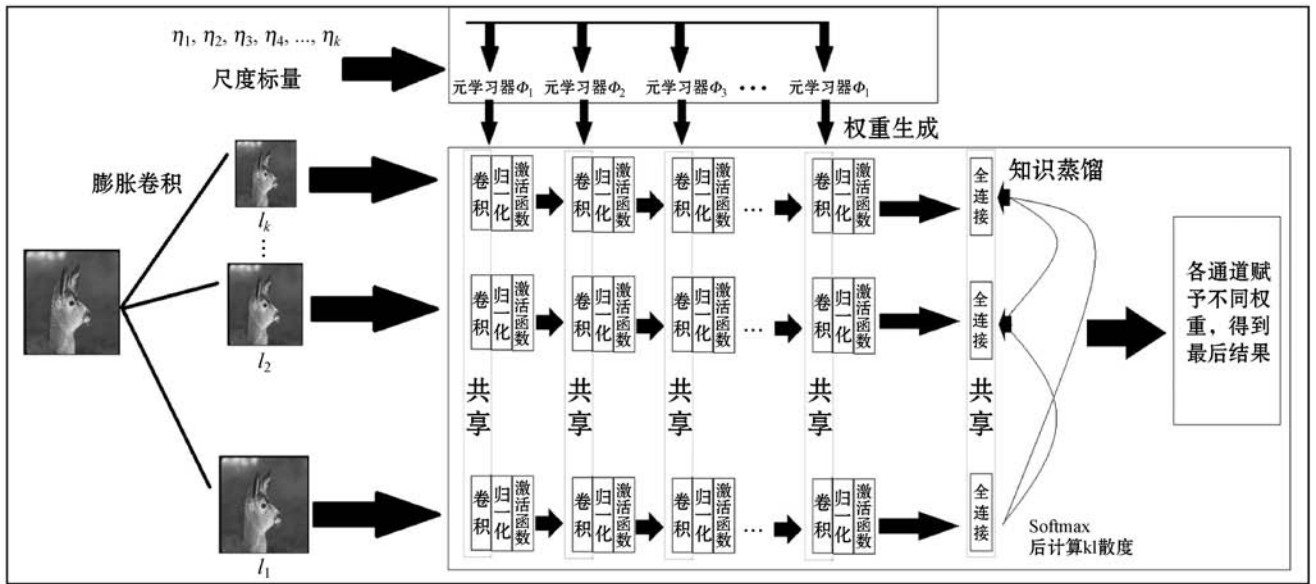


图 1 研究方法流程

本文对于不同尺度的网络设定一个尺度标量,将该尺度标量的信息作为各个网络的先验知识,以此生成卷积层的参数。不同尺度网络同层次的元学习器是共享的,这样可以更好地进行跨尺度进行学习。元学习器根据该先验知识生成卷积核,使其能更好地完成不同分辨率的机器视觉任务。卷积完成后再通过共享的全连接层,赋予各个尺度网络不同权重,得到最终结果。

### 2.1 多尺度图像生成

为了实现神经网络的多个尺度输入,研究设置了一组膨胀系数  $r = \{r_1, r_2, \dots, r_k\}$ ,本文对图像使用不同的膨胀系数进行卷积,得到不同分辨率的输入图像,再将输入图像作为特定的尺度网络的输入。本文在生成多尺度图像的方法上与其他多尺度研究方法不同,本文是通过使用不同的膨胀系数得到的。

### 2.2 网络权重生成

我们为每一个不同尺度的网络设定一个尺度标量集合  $\eta_1, \eta_2, \dots, \eta_k$ 。我们的元学习任务是将尺度标量作为各个子网络包含尺度信息的先验知识,然后输入到元学习器中,再通过元学习器得到卷积层所需参数。此外,本文还构建了一组元学习器  $\Phi = \{\Phi_1, \Phi_2, \dots, \Phi_l\}$ ,为每个神经网络的卷积层生成卷积核。例如,输入维度为 1,输出维度为  $C_{out} C_{in} K^2$ ,其中  $C_{out}$  是输出通道数,每个输出通道都具有  $C_{in} \times K \times K$  个卷积核。利用这些已经定义好的元学习器,可以生成各个网络中卷积层的参数,并与特定的尺度联系在一起,如式(1)所示。由于本研究的元学习器都是经过线性操作生

成,因此是可微的,反向传播时可以不断调整元学习器参数权重。

$$\phi_l = (\eta_k W_l + b_l) \quad (1)$$

式中: $\eta_k$ 是各个子网络的尺度标量; $W_l$ 是第  $l$  个元学习器的权重; $b_l$ 是第  $l$  个元学习器的偏量。

为了避免不同尺度间数据的影响,归一化层应该进行精细处理,在各个网络中私有化。通常情况下,归一化层的参数只占整个网络参数中很小一部分,我们为各个网络设定独立的归一化层是可行的。与此相比,全连接层参数所占的参数量太大,各个网络应共享全连接层。

### 2.3 基于尺度知识的知识蒸馏

高分辨率比低分辨率包含更多语义描述和信息<sup>[22]</sup>,因此大尺度网络可以辅导小尺度网络进行学习,我们称之为基于尺度的知识蒸馏技术。具体来说,本研究计算了所有尺度网络每对输出概率分布之间的 Kullback-Leibler 散度(KL 散度),从而得到了一个额外的损失函数,如式(2)所示。

$$L_{KL} = \sum_i^k \sum_j^k KL(p^i \| p^j) \quad i > j \quad (2)$$

式中:KL 是 KL 散度函数; $p$  是各个尺度网络 Softmax 后输出的概率分布。

本文方法的知识转移过程如下:(1) 我们将所有尺度网络进行联合训练,通过额外的损失函数,不同网络尺度间知识可以相互学习。(2) 我们的元学习器是共享的,不同网络之间的知识也在隐式学习。

### 2.4 预测

一般来说分辨率高的通道得到的结果比较精确,

权重比较高;分辨率低的通道可以分配较小权重。将大分辨率的精确信息结合小分辨率的细节信息结合在一起进行预测,可以发挥多尺度的优势,并且能大大增加识别的准确率。我们将各个通道得到的最后结果根据不同权重相加,相加后所得结果进行预测,如式(3)所示。

$$p_{\text{redit}} = \sum_{i=1}^k \alpha_i p_{\text{redit}_i} \quad (3)$$

式中: $\alpha_i$ 是各个通道的权重, $p_{\text{redit}_i}$ 是各个通道计算得到的结果。

## 2.5 损失函数

各个网络的损失函数如式(4)所示。

$$L_w = \text{celoss} \left( \sum_{i=1}^k \text{OUT}_w(\theta^i; x^i, y) \right) \quad (4)$$

式中: $\theta^i$ 是网络的权重参数; $x^i$ 是输入特征图像; $y$ 是真实标签; $\text{celoss}$ 是 Softmax 后再计算交叉损失熵。

通过 Softmax 和计算损失熵。再计算各个尺度网络间的 KL 散度。总的损失函数如式(5)所示。

$$L_{\text{oss}} = L_w + \beta L_{\text{KL}} \quad (5)$$

式中: $\beta$ 是辅助损失函数系数。

## 2.6 结构推理

本研究选择用不同的膨胀系数生成不同分辨率的输入图像。然后为每一个子网络设定一个不同尺度的尺度标量,再将尺度标量作为元学习器的输入,得到网络层的权重。元学习器在所有尺度网络中共享,归一化层和激活函数在各个网络中是私有的。各个网络的输入经过共享的全连接层后按不同权重进行计算得到结果,再将结果进行 Softmax 运算,计算损失熵。最后结合 KL 散度的辅助损失函数得到最后的损失值,再进行反向计算。我们选择在 ResNet-50<sup>[23]</sup>的基础上结合我们算法,在第一个卷积层中我们使用膨胀卷积方法得到不同分辨率图像,其他卷积层使用元学习的方法根据不同的尺度信息来学习卷积参数。然后再根据图 1 的操作,结合各个通道得到最后预测结果。

## 3 实验

本文实验平台为 Tesla v100 32 GB GPU,使用深度学习框架为 PyTorch,在 Corel5k 数据集上尝试了不同的结构进行实验,采用 top1、top5 准确率来验证方法的性能。

### 3.1 主实验

我们的方法在 Corel5k 图像集上进行评估,这是一

个公用的数据集,共 5 000 幅图片,其中包括 4 500 幅训练图像,500 幅测试图像,共 50 个场景。可以用于图像分类,图像检索。因为方法中各个子网络的输入是使用不同的膨胀系数进行卷积产生的,因此数据集中图像的尺寸不能太小。Corel5k 图像集中,图像大小为 192 × 128 像素,符合我们实验要求,故选择 Corel5k 作为实验数据集。

该数据集图片大小如图 2 所示,里面显示了其中 3 个场景,分别是日落、赛车场和北极。我们使用了数据增强的方法,利用随机大小调整和裁剪操作以及水平翻转来生成大小为 224 × 224 像素的原始图像。再通过不同的膨胀系数进行卷积得到各个通道的输入。



图 2 Corel5k 数据集

我们的方法选择轻量级网络 ResNet-50 进行优化,以证明网络的可扩展性和有效性,并且将多种当前流行算法与我们的方法进行对比。方法使用了 SGD 优化器进行训练,其动量为 0.1,权重衰减为 1e-4,循环迭代 180 次,初始学习率为 0.1,每 30 次衰减 50%,批量大小为 64。我们的方法参考了 DeepLab 神经网络,将膨胀系数设置为  $r = \{r_1, r_2, r_3, r_4\} = \{1, 6, 12, 18\}$ 。将辅助损失函数系数  $\beta$  设置为 0.3,将 4 个子网络的预测权重  $\alpha = \{0.4, 0.3, 0.2, 0.1\}$ 。在消融实验中,对  $\alpha$ 、 $\beta$ 、膨胀系数进行深入探究。

表 1 是实验结果,可以看出,我们的方法 top1 准确率比 ResNet-50 高了 4.21 百分点,top5 准确率比 ResNet-50 高 0.81 百分点,由于我们的方法是基于 ResNet-50 进行优化的,实验结果说明我们的方法是有效的并且准确率得到很大提升。同时,我们的方法比 ResNet-34 高了 5.41 百分点,比 ShuffleNet 高了 3.01 百分点,比 MobileNet 高了 4.21 百分点,比 MnasNet 高了 6.01 百分点,比效果最好的 GoogLeNet 高了 2.40 百分点,而 top5 准确率每个方法相差不大,说明我们方法的分类能力比当前一些流行算法更优异。

表 1 实验结果(%)

方法	top1 准确率	top5 准确率
ResNet-50	68.33	89.97
ResNet-34	67.13	88.37
ShuffleNetV2 <sup>[24]</sup>	69.53	90.98
GoogLeNet <sup>[25]</sup>	70.14	91.78
MobileNetV2 <sup>[26]</sup>	68.33	91.38
MnasNet <sup>[27]</sup>	66.53	88.17
本文	72.54	90.78

表 2 为我们的方法各个通道预测的准确率和结合各个通道进行预测的准确率,可以看出,在进行图像识别的任务时,随着膨胀系数的增加,该通道的预测结果明显降低,但是将各个网络所得到的结果按权重结合比单独的网络所得到的识别效果高得多。

表 2 使用膨胀系数为{1,6,12,18}效果(%)

类别	rate = 1	rate = 6	rate = 12	rate = 18	各个网络结合预测
top1	67.53	63.12	55.31	49.49	72.54
top5	89.97	84.96	82.96	78.15	90.78

### 3.2 消融实验

我们的方法的核心在于使用元学习产生卷积核权重和使用膨胀卷积生成不同尺度的输入图像,因此有必要针对元学习和膨胀卷积进行实验。因为每个尺度之间是通过知识蒸馏和预测时赋予不同的权重关联在一起,所以有必要针对由知识蒸馏技术所产生的辅助函数和各个子网络的权重进行实验。

#### 3.2.1 基于元学习和膨胀卷积实验

基于 ResNet-50 结构设置了 4 个子网络,进行了 4 组实验。第一组实验,不使用元学习器生成卷积核权重并且将图片处理为{224,192,160,128}大小作为 4 个子网络的输入。在这样的设置下,将不使用元学习和膨胀卷积的方法和我们提出的方法进行对比。第二组实验,将图片处理为{224,192,160,128}大小作为 4 个子网络的输入。在这样的设置下,将不使用膨胀卷积的技术和我们提出的方法进行对比。第三组实验,不使用元学习器生成卷积核权重。在这样的设置下,将不使用元学习技术和所提出的方法进行对比。第四组实验,也就是我们的方法,使用元学习器生成卷积核权重并且使用膨胀卷积方法生成 4 个子网络的输入。另外,因为当前流行的卷积的下采样率一般为 $1/32$ ,因此我们将图片处理为{224,192,160,128}大小。实验参数与主实验的实验参数一致。

表 3 是基于元学习和膨胀卷积进行实验的结果,可以看出,使用元学习器生成卷积核的权重并且使用膨胀卷积生成多个子网络的输入,得到的效果是最好的,为 72.54%。而不使用元学习和膨胀卷积方法比单独使用其两种之一方法效果更好。仅使用元学习器生成卷积核的权重或者膨胀卷积产生多个子网络的输入得到的效果与 ResNet-50 相比有一定提升。

表 3 基于元学习和膨胀卷积实验结果(%)

方法	top1 准确率	top5 准确率
不使用元学习和膨胀卷积	70.14	91.18
不使用膨胀卷积	68.73	91.18
不使用元学习	69.13	89.37
使用元学习和膨胀卷积*	72.54	90.78

由于元学习可以根据不同的尺度标量信息生成卷积核,因此每个网络所生成的图像信息都把握不同尺度信息,而膨胀卷积生成的图像信息在各个网络中都有不同效果,大尺度把握整体信息,小尺度把握细节信息。结合这两种方法,再通过知识蒸馏和结合各个网络信息进行预测的方法可以得到最好的效果。但是单独使用膨胀卷积和元学习的方法,每个网络所掌握图像信息的差异性不足,通过知识蒸馏和结合各个网络信息进行预测的方法得到的提升有限,甚至比两种方法都不使用的差。

实验结果证明同时使用元学习和膨胀卷积方法,增加每个网络所掌握信息的差异性,再对多个尺度的网络使用知识蒸馏并且赋予不同权重进行预测的方法对模型分类能力的提升效果最明显。

#### 3.2.2 基于辅助损失函数系数实验

基于辅助损失函数系数进行实验,主实验设置的辅助损失函数系数为 0.3。为了探究辅助损失函数系数对我们方法的影响,另外设置了五组实验,辅助损失函数系数分别为 1、0.9、0.7、0.5、0.1。其他实验参数与主实验的实验参数一致。

表 4 是基于辅助损失函数系数进行实验的结果,可以看出,系数为 1、0.9、0.7 时,随着系数的减少,准确率逐渐增高,系数为 0.7 时,准确率达 71.74%,比 ResNet-50 高了 3.41 百分点。但是系数为 0.5 时,准确率却降到了 69.73%,仅比 ResNet-50 高了 1.40 百分点。0.3 是主实验所使用的系数,准确率达到峰值,为 72.54%。系数为 0.1 时,效果甚至与 ResNet-50 相差不大。由此可见,损失函数系数对我们的方法具有一定的影响,0.3 是实验中效果最好的系数值。

表 4 基于辅助损失函数系数实验

辅助损失函数系数 $\beta$	top1 准确率/%	top5 准确率/%
1	69.73	90.38
0.9	70.34	91.38
0.7	71.74	89.97
0.5	69.73	89.77
0.3*	72.54	90.78
0.1	67.93	90.18

### 3.2.3 基于各子网络预测权重实验

基于各子网络预测权重进行实验,主实验的权重设置为 $\{0.4, 0.3, 0.2, 0.1\}$ 。为了探究辅助损失函数系数对我们方法的影响,另外设置了权重均为 1,权重为 $\{0.4, 0.2, 0.3, 0.1\}$ 、 $\{0.2, 0.4, 0.3, 0.1\}$ 、 $\{0.2, 0.3, 0.4, 0.1\}$ 四组实验。

除了权重均为 1(既每个网络的预测结果都是同样重要的)的实验外,其他实验的权重设置为 4 个子网络权重和为 1。可以认为是一个网络拆成四份,最后再合并进行预测。尺度大的网络往往准确率比较高,所以设置的权重比较大,而中间两个子网络往往通过知识蒸馏提高了准确率,往往权重是不能确定的,因此设置了 $\{0.4, 0.3, 0.2, 0.1\}$ 和 $\{0.4, 0.2, 0.3, 0.1\}$ 两组实验。中间两个通过知识蒸馏提升性能的网络对我们方法的贡献可能比尺度最大的网络高,因此设置了 $\{0.2, 0.4, 0.3, 0.1\}$ 、 $\{0.2, 0.3, 0.4, 0.1\}$ 两组实验。

表 5 是基于各子网络预测权重进行实验的结果,可以看出,不同的权重对实验结果的影响很大,尺度大的网络对预测结果的贡献更高,中间两个子网络对最后预测结果的贡献相差不大, $\{0.4, 0.3, 0.2, 0.1\}$ 是实验效果最好的一组权重。

表 5 基于各子网络预测权重实验(%)

各子网络预测权重	top1 准确率	top5 准确率
均为 1	70.94	89.37
$\{0.4, 0.3, 0.2, 0.1\}$ *	72.54	90.78
$\{0.4, 0.2, 0.3, 0.1\}$	72.34	90.58
$\{0.2, 0.4, 0.3, 0.1\}$	70.54	89.57
$\{0.2, 0.3, 0.4, 0.1\}$	69.93	89.77

### 3.2.4 基于膨胀系数实验

基于膨胀系数进行实验,主实验的膨胀系数参考了 DeepLab 神经网络,设置为 $\{1, 6, 12, 18\}$ 。为了探究膨胀系数对我们方法的影响,以 6 为基准每间隔 2 设置一组实验,共设置了 4 组实验,膨胀系数分为: $rate1 = \{1, 4, 8, 12\}$ ,  $rate2 = \{1, 6, 12, 18\}$ ,  $rate3 = \{1, 8,$

$16, 24\}$ ,  $rate4 = \{1, 10, 20, 30\}$ ,  $rate5 = \{1, 12, 24, 36\}$ 。

表 6 是基于膨胀系数进行实验的结果,可以看出,  $rate1$  的准确率达到 71.34%,比 resnet-50 高 3.01 百分点。而  $rate3$  准确率为 69.53%,比 resnet-50 高 1.20 百分点。 $rate4$  和  $rate5$  准确率分别为 68.73% 和 68.13%,与 resnet-50 相差不大。而我们的主实验  $rate2$  的效果最好,达到 72.54%。由此可知,膨胀系数过大或者过小对于图像识别任务都不是最适合的选择。 $rate2 = \{1, 6, 12, 18\}$  是实验组中效果最好的一组。

表 6 基于膨胀系数实验(%)

膨胀系数	top1 准确率	top5 准确率
$\{1, 4, 8, 12\}$	71.34	90.98
$\{1, 6, 12, 18\}$ *	72.54	90.78
$\{1, 8, 16, 24\}$	69.53	89.57
$\{1, 10, 20, 30\}$	68.73	90.98
$\{1, 12, 24, 36\}$	68.13	89.97

## 4 结 语

多尺度特征学习可以更好地结合图像的精度和细节特征,有利于图像识别准确率的提高。元学习让计算机模拟人的大脑学会如何去学习,更高效灵活地完成图像识别任务。多尺度特征学习和元学习都是现在机器视觉领域研究的热点。本文提出了一种基于多尺度和元学习的神经网络模型,首先使用膨胀卷积产生不同尺度的输入图像,再用元学习的方法动态生成卷积核,最后各个通道间通过知识蒸馏的方法,高分辨率通道指导低分辨率通道,提高图像识别的准确率。

我们使用了四组不同的膨胀系数进行实验得到以下结论:与单独训练单个网络相比,结合各个网络输出进行图像识别可以获取更好的性能。准确率最低的通道准确率为 49.49%,准确率最高的通道准确率为 67.53%,将各个通道按不同权重结合后准确率达 72.54%,提升效果十分显著。与基线算法 ResNet50 相比提高了 4.21 百分点,有很好研究前景。它兼容大部分神经网络架构,可以基于现在流行的神经网络模型进行使用。该方法的有效性和可扩展性具有巨大潜力,有望在其他领域中发挥巨大作用。

## 参 考 文 献

- [1] Li D, Yao A, Chen Q. Learning to learn parameterized classification networks for scalable input images[EB]. arXiv: 2007.06181, 2020.
- [2] Yang T, Zhu S, Chen C, et al. MutualNet: Adaptive Conv-

- Net via mutual learning from network width and resolution [EB]. arXiv:1909.12978,2019.
- [ 3 ] Zhang D, Zhang H, Tang J, et al. Feature pyramid transformer[EB]. arXiv:2007.09451,2020.
- [ 4 ] Tang C, Liu X, Zheng X, et al. DeFusionNET: Defocus blur detection via recurrently fusing and refining multi-scale deep features [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2020,44(2):955–968.
- [ 5 ] Wang H, Wang Q, Gao M, et al. Multi-scale location-aware kernel representation for object detection [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE,2018:1248–1257.
- [ 6 ] Lin T, Dollár P, Girshick R, et al. Feature pyramid networks for object detection [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. IEEE,2017:936–944.
- [ 7 ] Chen L, Papandreou G, Kokkinos I, et al. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018,40(4):834–848.
- [ 8 ] Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network [EB]. arXiv:1612.01105,2016.
- [ 9 ] Santoro A, Bartunov S, Botvinick M, et al. Meta-learning with memory-augmented neural networks [C]//33rd International Conference on Machine Learning. ACM,2016:1842–1850.
- [10] Munkhdalai T, Yu H. Meta networks [EB]. arXiv:1703.00837,2017.
- [11] Andrychowicz M, Denil M, Gomez S, et al. Learning to learn by gradient descent by gradient descent [EB]. arXiv:1606.04474,2016.
- [12] Vinyals O, Blundell C, Lillicrap T, et al. Matching networks for one shot learning [EB]. arXiv:1606.04080,2016.
- [13] Ravi S, Larochelle H. Optimization as a model for few-shot learning [C]//International Conference on Learning Representations,2017:1–11.
- [14] Sung F, Zhang L, Xiang T, et al. Learning to learn: Meta-critic networks for sample efficient learning [EB]. arXiv:1706.09529,2017.
- [15] Lee K, Maji S, Ravichandran A, et al. Meta-Learning with differentiable convex optimization [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE,2019:10649–10657.
- [16] Yao H, Liu Y, Wei Y, et al. Learning from multiple cities: A meta-learning approach for spatial-temporal prediction [EB]. arXiv:1901.08518,2019.
- [17] Behl H, Najafi M, Arnab A, et al. Meta learning deep visual words for fast video object segmentation [EB]. arXiv:1812.01397,2018.
- [18] Vanschoren J. Meta-Learning: A survey [EB]. arXiv:1810.03548,2018.
- [19] Huang Z, Wang N. Like what you like: Knowledge distill via neuron selectivity transfer [EB]. arXiv:1707.01219,2017.
- [20] Xu Z, Hsu Y, Huang J. Training shallow and thin networks for acceleration via knowledge distillation with conditional adversarial networks [EB]. arXiv:1709.00513,2018.
- [21] Furlanello T, Lipton Z, Tschannen M, et al. Born again neural networks [EB]. arXiv:1805.04770,2018.
- [22] Sun K, Xiao B, Liu D, et al. Deep high-resolution representation learning for human pose estimation [EB]. arXiv:1902.09212,2019.
- [23] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. IEEE,2016:770–778.
- [24] Ma N, Zhang X, Zheng H, et al. ShuffleNet V2: practical guidelines for efficient CNN architecture design [EB]. arXiv:1807.11164,2018.
- [25] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions [EB]. arXiv:1409.4842,2014.
- [26] Sandler M, Howard A, Zhu M, et al. MobileNetV2: Inverted residuals and linear bottlenecks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE,2018:4510–4520.
- [27] Tan M, Chen B, Pang R, et al. MnasNet: Platform-aware neural architecture search for mobile [EB]. arXiv:1807.11626,2019.
- ~~~~~
- (上接第 139 页)
- [20] Cui X, Liu J, Huang T, et al. A novel metric for cache size allocation scheme in Content Centric Networking [C]//National Doctoral Academic Forum on Information and Communications Technology. IEEE,2013:1–6.
- [21] 杨博,刘大有,金弟,等. 复杂网络聚类方法 [J]. 软件学报,2009,20(1):54–66.
- [22] Saino L, Psaras I, Pavlou G. Icarus: A caching Simulator for Information Centric Networking (ICN) [C]//7th International ICST Conference on Simulation Tools and Techniques, 2014:66–75.
- [23] Li Y, Wang J, Han R. An on-path caching scheme based on the expected number of copies in information-centric networks [J]. Electronics,2020,9(10):1705.
- [24] Zeng L, Ni H, Han R. An incrementally deployable IP-compatible-information-centric networking hierarchical cache system [J]. Applied Sciences,2020,10(18):6228.
- [25] Escrivá R, Wong B, Sireer E. HyperDex: A distributed, searchable key-value store [C]//ACM SIGCOMM Computer Communication Review. ACM,2012:25–36.
- [26] Knight S, Nguyen H, Falkner N, et al. The Internet topology zoo [J]. IEEE Journal on Selected Areas in Communications,2011,29(9):1765–1775.