

基于多分支结构的不确定性局部通道注意力机制

伍邦谷¹,张苏林²,石红¹,朱鹏飞¹,王旗龙¹,胡清华¹

(1. 天津大学智能与计算学部,天津 300350; 2. 中汽数据(天津)有限公司,天津 300380)

摘要: 近几年的研究表明视觉注意力机制是提升深层卷积神经网络性能的有效途径. 然而,现有的视觉注意力方法更多地致力于建模所有卷积通道之间的相关性,在一定程度上限制了模型的计算效率. 此外,这些方法尚未明确考虑相关性建模过程中不确定性带来的影响,缺少对注意力机制在泛化能力和稳定性方面的探索. 为解决上述问题,提出了一种多分支局部通道注意力模块(Multi-Branch Local Channel Attention, MBLCA). 通过建模通道之间的局部相关性学习各个通道的权重,提升了模型的计算效率. 并采用蒙特卡洛(Monte Carlo, MC)Dropout近似的深度贝叶斯学习方法对局部通道注意力模块进行不确定性建模,从而得到一个多分支的局部通道注意力模块. 提出的MBLCA模块可以灵活地应用于各种深层卷积神经网络架构中,与同类型的工作相比,嵌入MBLCA模块的ResNet-50网络结构在ImageNet-1K和MS COCO数据集上分别取得了2.58%的分类精度提升和1.9%的AP提升.

关键词: 通道注意力机制; 不确定性; 多分支结构; 深层卷积神经网络

中图分类号: TP183;TP391.4

文献标识码: A

文章编号: 0372-2112(2022)02-0374-09

电子学报 URL: <http://www.ejournal.org.cn>

DOI:10.12263/DZXB.20201204

Multi-Branch Structure Based Local Channel Attention with Uncertainty

WU Bang-gu¹, ZHANG Su-lin², SHI Hong¹, ZHU Peng-fei¹, WANG Qi-long¹, HU Qing-hua¹

(1. College of Intelligence and Computing, Tianjin University, Tianjin 300350, China;

2. Automotive Data of China Co., Ltd., Tianjin 300380, China)

Abstract: Recent researches demonstrate that attention mechanism is an effective way to improve performance of deep convolution neural networks(CNNs). However, most of existing attention methods more dedicate to modeling the correlation between all channels, which limits the computational efficiency of the model. In addition, these methods have not considered the impact of uncertainty in the correlation modeling process, and lack the exploration of the generalization ability and stability of the attention mechanism. A multi-branch local channel attention(MBLCA) module is proposed to handle above issues. MBLCA learns channel attention by capturing correlation across channels in a local range instead of global ones, improving the computational efficiency, and models the uncertainty of local channel attention by deep Bayesian learning, which is approximated by Monte Carlo(MC) Dropout, leading a multi-branch structure. The proposed MBLCA can be flexibly adopted to various deep CNN architectures. For example, ResNet-50 with the MBLCA module has achieved 2.58% improvement in classification accuracy and 1.9% improvement in average precise on the ImageNet-1K and MS COCO datasets against state-of-the-art counterparts.

Key words: channel attention mechanism; uncertainty; multi-branch structure; deep convolutional neural networks

1 引言

深层卷积神经网络已经被广泛应用到多种人工智能相关的任务中,例如图像分类、目标检测、实例分割等,并在相应任务中都取得了卓越的性能和效果. 但是如何在复杂场景和任务上进一步提升卷积神经网络的准确率和稳定性仍是一个非常具体挑战性的问题. 解

决该问题的一个重要思路就是提升卷积神经网络的特征学习能力,从而提升神经网络的表达和泛化能力^[1]. 沿着这个思路,近年来研究者在如何进一步提升深层卷积神经网络性能方面做了大量的工作^[2-8],包括不断扩增深层卷积神经网络的深度和宽度^[3-5],利用跳跃连接^[6]和密集连接^[7]提升深层卷积神经网络的训练稳

定性和特征利用率. 特别地,视觉注意力机制^[8]通过自适应学习特征的重要性,对不同的特征进行加权,突出对特定任务重要的特性,压制对任务无用或者有害的特征,实现样本级别的自适应特征选择,从而增强特征学习能力,提升卷积神经网络的表达和泛化能力.

SENet^[8]是利用视觉注意力机制提升深层卷积神经网络性能最具有代表性的工作之一,它通过聚合和激发操作来学习通道之间的相关性,并生成通道级的注意力图对卷积特征进行加权. 其中,聚合操作通过全局平均池化对输入的卷积特征进行全局编码,激发操作通过两层全连接层学习通道之间的全局相关性,最后使用 Sigmoid 函数对学习的相关性进行激活从而获得各个通道的权重. SENet 在图像分类任务上取得了非常优异的性能,并且启发了后续一系列关于视觉注意力机制的探索^[9-14]. 例如, CBAM^[9]在聚合操作中同时引入了全局均值池化和全局极大值池化,并且额外加入空间维度的视觉注意力模块,进一步提升了视觉注意力机制的性能. 此外,还有一些工作^[10-14]通过设计一些更为精巧的视觉注意力模块学习所有通道之间的非局部非线性相关性. 虽然这些方法在性能上得到了一定的提升,但是不可避免地引入了更多的参数和计算量,从而提升了模型复杂度,并限制了计算效率.

对于通道注意力机制而言,通道相关性建模是一个至关重要的环节. ECA-Net^[15]提出一种局部通道相关性建模的思想,打破了之前全局通道相关性建模的必要性假设,在效率和性能上得到了非常好的平衡. 然而,无论是 ECA-Net 还是现有其他的视觉注意力方法都忽略了卷积通道相关性建模中不确定性带来的影响. 例如,由输入数据或标签噪声和缺失导致的卷积特征噪声和异常值会给通道相关性估计带来明显的偏差和不确定性,在一定程度上限制了注意力机制的泛化能

力和稳定性. 类似的不确定性问题已经在深层神经网络模型中得到了研究者的关注. 贝叶斯神经网络^[16]通过对深度神经网络的输出进行概率化建模,是实现神经网络不确定性建模的重要手段. 然而这些贝叶斯神经网络通常具有较高的计算复杂度,特别是其推理过程. 为了缓解贝叶斯神经网络计算复杂度高的问题,研究者针对后验分布的建模提出多种近似算法^[17,18]. 其中,文献^[18]提出了一种 MC Dropout 方法,并取得了较好的性能.

受到 ECA-Net^[15]和 MC Dropout^[18]的启发,本文提出一种多分支局部通道注意力模块(Multi-Branch Local Channel Attention, MBLCA). 如图 1 所示,在模型训练阶段,MBLCA 模块对聚合后特征向量进行多重采样,使用 Dropout^[19]得到多个随机生成重采样的特征,利用权重共享的 1D 卷积学习所有重采样特征各自的局部通道之间的相关性,最终通过平均融合和 Sigmoid 函数得到通道权重. 因此,MBLCA 模块可以通过快速 1D 卷积和 MC Dropout 实现局部通道注意力机制的不确定性建模. 在模型推理阶段,为了避免随机性,直接使用训练好的 1D 卷积进行运算,这也使得 MBLCA 方法的推理效率与 ECA-Net^[15]完全一致. 通过训练不确定性局部通道注意力模块,可以进一步提升深层卷积神经网络中视觉注意力机制的泛化能力和稳定性. 为了验证所提方法的有效性,本文在 ImageNet^[3]和 MS COCO^[20]数据集上进行了大量的实验. 主要贡献总结如下.

(1) 尝试通过对局部通道注意力机制进行不确定性建模,以提高视觉注意力模块的泛化能力和稳定性. 为此,提出了一种多分支局部通道注意力模块(MBLCA),该模块采用权重共享的快速 1D 卷积和 MC Dropout 来学习通道之间的局部相关性和建模模型的不确定性.

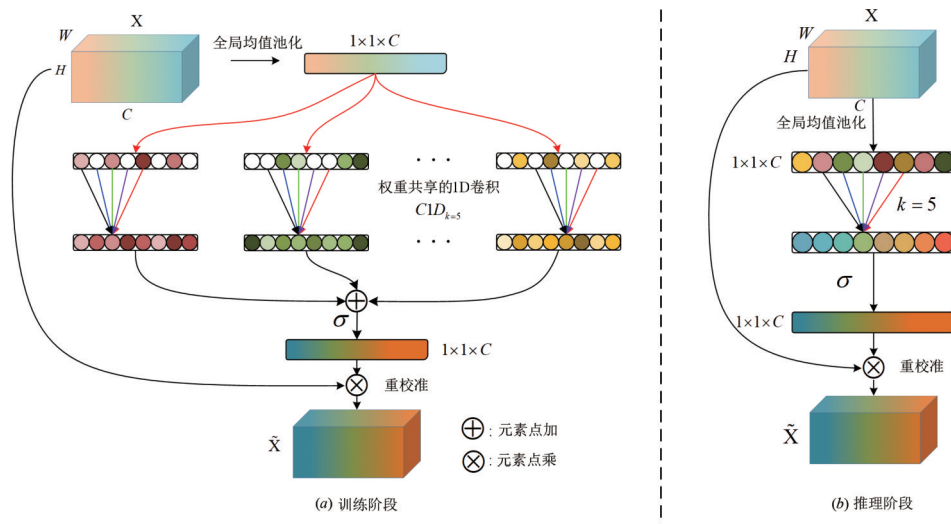


图1 多分支局部通道注意力(MBLCA)模块整体框图

(2)提出的 MBLCA 方法在 ImageNet-1K 和 MS COCO2017 数据集上针对图像分类、目标检测以及实例分割等多种任务进行了大量的实验验证. 实验结果表明, MBLCA 方法相比于其他同类方法具有更低的模型复杂度和更好的性能, 在模型复杂度和性能之间实现了较好的平衡.

2 相关工作

2.1 视觉注意力机制

视觉注意力机制在提升深层卷积神经网络性能方面展现了强大的潜力. SENet^[8]在学习视觉注意力时建模所有通道之间的相互依赖关系, 利用了通道的全局上下文信息. 在 SENet^[8]的基础上, CBAM^[9]同时引入了通道维度和空间维度的注意力模块, 并且探索了 2 个模块之间不同融合方式的影响. 文献[21]将注意力学习与相关滤波跟踪算法结合, 引入对目标的空时域特性的考虑, 提升了算法的精确度和成功率. DSCA-Net^[22]提出了一种基于双流信息与通道注意力的室内检测方法以解决计数问题中目标尺度多样性的问题. 此外, 一些研究者探索了更加复杂的通道相关性学习方法以获得更高的性能. 其中, 非局部神经网络^[14]提出了一种非局部模块用于捕获跨通道和空间位置的全局依赖关系, 在视频识别任务中取得了显著的性能提升. 文献[23]将基于深度学习的非局部注意力模块应用于增强网络的图像去雨效果, 有效地解决了不同雨密度和不同大小雨条纹的问题. GCNet^[24]简化了非局部模块^[14], 并且将之与 SE 模块相结合得到一种更高效的全局通道依赖关系学习方式. DANet^[12]、A²-Net^[11]和 AANet^[10]在非局部模块^[14]的基础上探索了更加有效的全局的依赖关系学习模块. GSoP 模块^[13]从二阶统计信息的角度发掘通道和空间的全局二阶相关性. 上述的方法主要基于 2 种思路对注意力模块进行设计: 第一种思路是计算空间维度或者通道维度的注意力图, 通过利用注意力图重加权策略实现特征增强^[8,9,13]; 另一种思路是学习并利用空间与通道之间的长距离非局部的相关性以提升模型的表达能力的方法^[10-12,14]. 但是, 这些方法在通过利用注意力机制提升深层神经网络模型性能的同时, 也引入了大量参数数量和计算复杂度, 影响了网络模型的计算效率. 与这些方法不同, MBLCA 模块采用局部通道视觉注意机制, 用一种更加高效的方式达到同样甚至更好的性能. 本文方法与 ECANet^[15]在建模通道相关性上具有相似思想, 它们均采用高效的 1D 卷积学习通道局部相关性. 然而, 与 ECANet^[15]不同的是, MBLCA 通过对通道注意力模块进行不确定性建模, 进一步提升模型的泛化能力和稳定性, 同时不引入额外的推理代价.

2.2 模型不确定性

近年来, 深度学习中的不确定性建模引起了研究者的广泛关注. 贝叶斯神经网络^[16]随之被提出来用于解决这个问题. 然而其存在计算复杂度高的问题, 很难应用到实际任务中. 为此, 研究者提出了多种贝叶斯神经网络的近似算法, 其中包括基于抽样的变分推理和随机变分推理方法^[18,25-27]. 特别地, 文献[18]提出了使用 MC Dropout 方法对贝叶斯推断进行近似以完成深层神经网络的不确定性建模. 与此同时, 不确定性建模在多个领域都得到了广泛的应用. 文献[28]将不确定性建模应用于高密度椒盐噪声降噪方法中, 并显著提升了算法的综合性能. 文献[29]提出了一种新的空间数据不确定性建模方法, 并在图像数据上验证了该方法的有效性. 上述现有的不确定性建模方法主要关注数据的不确定性和整体模型预测的不确定性, 极少关注整体模型中局部模块的不确定性. 本文提出的 MBLCA 受 MC Dropout^[18]的启发, 引入 MC Dropout 策略对局部注意力模块进行不确定性建模, 从而进一步提升整体神经网络模型的泛化能力和稳定性.

2.3 多分支结构

在深度学习的模型设计中, 多分支结构经常被用于学习多尺度、多视角的信息等, 以提升模型的表征能力. 文献[30]提出一种全局多分支 RNN 对时间序列数据中时间延迟进行建模, 以增强网络的表征能力. 不同于文献[30], 本文的方法 (MBLCA) 所设计的多分支结构是一种局部模块, 其目的是建模通道注意力机制的不确定性, 以提高神经网络的泛化能力和稳定性. Inception 网络^[5]和双通路网络 (Dual Path Network, DPN)^[31]设计了局部多分支结构用于提升神经网络的性能, 其中 Inception 网络^[5]采用不同尺寸的卷积运算, 形成了一种多分支结构用于提取图像多尺度特征; DPN^[31]结合了 ResNet^[6]和 DenseNet^[7]的优势, 一边利用残差旁支通路复用特征, 另一边利用密集连接旁支通路探索新特征. 不同于上述这些方法, MBLCA 基于 MC Dropout 方法设计了多分支结构对通道注意力机制进行不确定性建模.

3 多分支局部通道注意力机制

3.1 SENet 中的通道注意力机制

SE 模块通过建模所有通道之间的关系学习每个通道的权重. 对于任意卷积层的输出 $X \in \mathbb{R}^{W \times H \times C}$, 其中 W, H 和 C 分别表示特征图的宽度, 高度以及通道的维度, SE 模块通过引入聚合, 激发以及重校准三个连续的操作得到加权后的特征图 \tilde{X} , 具体计算过程如下:

$$\tilde{X} = \mathcal{F}_{\text{scale}}(\mathcal{F}_{\text{ex}}(\mathcal{F}_{\text{sq}}(X), W), X) \quad (1)$$

其中, W 是一组可学习参数, 用于建模聚合后特征所有

通道之间的相关性.

SE 模块的聚合操作 $\mathcal{F}_{sq}(\mathbf{X})$ 通过对卷积特征进行全局均值池化得到其维度上的全局分布. 具体而言, 对于给定的特征图 \mathbf{X} , $\mathcal{F}_{sq}: \mathbf{X} \rightarrow \mathbf{s} \in \mathbb{R}^C$ 对 \mathbf{X} 进行聚合操作, 其中 \mathbf{s} 的第 c 个元素计算如下:

$$s_c = \mathcal{F}_{sq}(X_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X_c(i, j) \quad (2)$$

其中, X_c 表示卷积特征的第 c 维通道的特征图.

随后, 激发操作 $\mathcal{F}_{ex}(\mathbf{s}, \mathbf{W})$ 采用 2 个全连接层和 Sigmoid 激活函数生成每一个通道的权重, 其中两个全连接层用于建模所有通道之间的相关性. 最终生成的权重 ω 计算如下:

$$\omega = \sigma(\mathbf{W}_2 \text{ReLU}(\mathbf{W}_1 \mathbf{s})) \quad (3)$$

其中, $\mathbf{W} = \{\mathbf{W}_1 \in \mathbb{R}^{C \times \frac{C}{r}}, \mathbf{W}_2 \in \mathbb{R}^{\frac{C}{r} \times C}\}$, ReLU 表示线性整流函数^[32], σ 为 Sigmoid 激活函数. 为了降低模型的复杂度, \mathbf{W}_1 和 \mathbf{W}_2 为 2 个低秩矩阵, 对特征向量进行了降维, 而 r 表示降维的比例. 最后, 权重 ω 通过通道级的点积运算对卷积特征进行加权重校准.

3.2 多分支局部通道注意力机制 (MBLCA)

MBLCA 模块同样采用全局均值池化 (Global Average Pooling, GAP) 生成特征向量 \mathbf{s} . 与以往关注于建模全局通道相关性的注意力方法不同, MBLCA 方法旨在更高效建模通道之间的相关性, 从而降低模型复杂度. 受 ECA-Net^[15] 的启发, 本文采用快速 1D 卷积建模通道之间的局部相关性. 具体而言, 在聚合后的特征向量上利用快速 1D 卷积计算通道的局部相关性, 采取 Sigmoid 函数获得各通道的权重, 整体运算过程如下:

$$\omega = \sigma(\text{C1D}_k(\mathbf{s})) \quad (4)$$

其中, $\text{C1D}_k(\cdot)$ 表示卷积核大小为 k 的 1D 卷积, σ 为 Sigmoid 激活函数.

上述的局部通道注意力机制 (式 (4)) 可以非常高效地建模通道之间的相关性, 并生成较为有效的通道权重. 然而, 通道局部交互范围的限制 (例如, 不同的通道存在不同的交互范围) 以及输入信息自身复杂性 (例如, 通道特征存在潜在的噪声或者异常点) 都会对这种通道局部相关性的建模, 带来明显的不确定性. 为了能够建模局部通道注意力机制存在的不确定性, 在式 (4) 的基础上引入了 MC Dropout 机制, 通过近似贝叶斯推断的方式, 对视觉注意力模型中的不确定性进行建模.

在具体实现过程中, MBLCA 模块中引入了一个可学习参数 $\tilde{\mathbf{W}}$. 对于给定的输入 \mathbf{s} 和期望权重参数 $\tilde{\mathbf{s}}$, MBLCA 模块期望可以学习到以下分布:

$$p(\tilde{\mathbf{W}}|\mathbf{s}, \tilde{\mathbf{s}}) \quad (5)$$

然而, 式 (5) 是很难计算得到的. 因此, 本文采用蒙特卡洛方法在较小计算量的前提下, 对 \mathbf{s} 进行有限次数

的重采样. 同时, Dropout 作为一个天然的随机生成器, 对重采样的特征进行随机扰动, 以近似后验分布 $p(\tilde{\mathbf{W}}|\mathbf{s})$. 于是, 构造了一种多分支结构, 利用蒙特卡洛采样和 Dropout 生成多分支通道特征向量 $\tilde{\mathbf{s}}$:

$$\tilde{\mathbf{s}}_i = \mathcal{F}_{\text{dropout}}(\mathbf{s}) = \text{Bernoulli}(p) \cdot \mathbf{s} \quad \text{for } i = 1, 2, \dots, n \quad (6)$$

其中 $\tilde{\mathbf{s}} = \{\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \dots, \tilde{\mathbf{s}}_n\}$, n 表示分支的数量, p 表示 dropout 的概率, $\text{Bernoulli}(\cdot)$ 表示 Bernoulli 分布.

如图 1(a) 所示, 在训练过程中对于聚合的通道特征先进行重采样和 Dropout 操作, 利用带有共享权重 $\tilde{\mathbf{W}}$ 的 1D 卷积计算多分支通道特征向量的局部相关性, 最后对多分支的输出结果进行平均聚合. 根据式 (4), MBLCA 模块在训练过程中的权重 ω 可以有如下公式计算:

$$\omega = \sigma\left(\frac{1}{n} \sum_{i=1}^n \text{C1D}_k(\tilde{\mathbf{s}}_i)\right) \quad (7)$$

由式 (7) 可以看出, 通过对输入特征进行多重随机采样, 可以建模共享权重 $\tilde{\mathbf{W}}$ 的不确定性, 从而获得更好的泛化能力和稳定性.

在训练阶段, 所有的分支都共享同样的 1D 卷积核, 因此在推理的过程中, 如图 1(b) 所示, 未使用 MC Dropout 策略以消除随机性, 并采用单分支用于最终预测. 与 SENet 和其他的注意力方法相比, MBLCA 模块具有更低的模型复杂度, 详见 3.3 节.

3.3 模型复杂度分析

本节从引入的参数数量和每秒浮点运算量 (FLOPs) 的角度分析了提出的 MBLCA 模块的模型复杂度. 对比了 3 个同类方法 (即 ECA 模块^[15], SE 模块^[8] 和 CBAM 模块^[9]). 4 种注意力模块的参数数量和每秒浮点运算量如表 1 所示, 其中 B 表示训练批次的大小, H 和 W 分别表示输入特征图的高度和宽度, k 和 n 分别表示 ECA 模块和 MBLCA 模块中 1D 卷积核的大小以及分支数量. 如表 1 所示, SE 模块包含 2 个全连接层, 它的参数数量和计算量分别为 $2C^2/r$ 和 $2C^2 \cdot B/r$. CBAM 将通道和空间维度的注意力机制相结合, 其包含的参数数量和计算量分别为 $2C^2/r + 49 \cdot 2$ 和 $2C^2 \cdot B/r + H \cdot W \cdot 49 \cdot 2$. ECA 模块采用 1D 卷积的方式学习局部通道之间的相关性, 其包含的参数数量和计算量分别为 k 和 $kC \cdot B$. MBLCA 模块和 ECA 模块具有相同的参数数量. 在模型训练阶段, MBLCA 模块的计算量是 ECA 模块的 n 倍 (n 为分支的数量, 通常 $n=6$). 但值得一提的是, MBLCA 模块在模型推理阶段的计算量与 ECA-Net 完全相同. 因此一旦模型训练完成, MBLCA 模块相比 ECA 模块未带来额外的计算代价, 但提高了泛化能力和稳定性. 显然, 相比于 SE 模块和 CBAM 模块, MBLCA 模块具有更低的模型复杂度. 值得注意的是, 这里 k 和 n 的取值通常不超过 15 和 6.

表1 不同注意力机制方法的模型复杂度分析

模型	MBLCA	ECA	SE	CBAM
参数量	k	k	$\frac{2C^2}{r}$	$\frac{2C^2}{r} + 49 \cdot 2$
计算量	$knC \cdot B$	$kC \cdot B$	$\frac{2C^2 \cdot B}{r}$	$\frac{2C^2 \cdot B}{r} + H \cdot W \cdot 49 \cdot 2$

4 实验分析

为了验证 MBLCA 方法在提升深层卷积神经网络的泛化能力和稳定性方面的有效性. 本节利用 ImageNet-1K^[3]和 MS COCO^[20]数据集在大规模图像分类、目标检测和实例分割任务,以及噪声图像分类上对 MBLCA 模块进行验证. 首先,为了探索 MBLCA 模块中所包含的超参数对模型的影响,同时找到最优的参数组合与相关工作进行公平比较,本节在大规模图像分类任务上对 MBLCA 模块中分支数量 n , Dropout 概率 p 以及快速 1D 卷积核大小 k 进行了消融实验. 之后,基于选取的最优超参数组合在 ImageNet-1K 数据集上对比了当前最新的同类相关方法,以验证 MBLCA 模块的有效性. 此外,为了进一步验证本文方法在不同任务上的泛化能力,以 ImageNet 预训练的网络模型作为骨干网络,采用 Mask R-CNN^[33]作为基础检测器,在 MS COCO 数据集上进行了目标检测和实例分割实验. 最后,为了验证 MBLCA 方法在提升噪声鲁棒性和稳定性方面的有效性,采用 ImageNet 预训练的网络模型在 ImageNet-C 数据集^[34]进行相关的对比实验. 相关程序均使用 PyTorch 实现,并运行在装备 4 块 Titan Xp 显卡和 Intel Xeon (R) Silver 4112 @ 2.60 GHz CPU 的服务器上.

4.1 实验设置

在 ImageNet-1K 数据集上训练网络模型时,采用了与 ResNet^[6]相同的数据增广方式,包括随机裁剪、随机水平翻转和缩放. 采用权重衰减为 $1e-4$,动量为 0.9,批次大小为 256 的随机梯度下降算法 (Stochastic Gradient Descent, SGD) 优化网络模型参数. 初始学习率为 0.1,每 30 轮迭代减小为原来的 $1/10$. 所有的模型都训练 100 轮迭代. 对于 MS COCO 数据集,本文以 Resnet-50 和 ResNet-101 作为基础网络,并采用带有 FPN^[35]的 Mask R-CNN^[33]作为基础检测器. 输入图像的短边调整为 800,所有模型都使用权重衰减为 $1e-4$,动量为 0.9 的 SGD 优化算法进行训练. 训练批次大小为 8 (即每张显卡上运算 2 张图片). 初始学习率为 0.01,分别在训练 8 和 10 轮迭代后缩减为原来的 $1/10$,总计训练 12 轮迭代.

4.2 在 ImageNet-1K 上的大规模分类实验

4.2.1 消融实验

本文首先在 ImageNet-1K 数据集上分析 MBLCA 方法中一些关键的超参数带来的影响,主要包括分支的数量 (n), Dropout 的概率 (p) 以及快速 1D 卷积的卷积核

大小 (k). 在该部分实验中采用 ResNet-50 为基础模型.

(1) **分支数量和 Dropout 概率的影响.** MBLCA 模块中 1D 卷积主要是为了学习局部通道之间的相关性,其在 ECA-Net^[15]中通常设置为 7 或者 9. 分支数量和 Dropout 概率是不确定性建模的 2 个重要影响因素. 为了更好地验证 2 个因素的影响,此处将局部通道相关性建模范围进行扩大,将 k 设置为 11. 因为通常较大的交互范围中的不确定性越显著. 这里的实验将 p 和 n 分别设置为 0.25、0.50、0.75 和 4、6、8,共计 9 种组合. 实验结果如图 2 所示,可以看到 n 和 p 对 MBLCA 的性能有着明显的影响. 当 $n=6, p=0.5$ 时, MBLCA 方法得到了最优的性能 (77.56%). 相比于基线模型 ($n=1, p=0, 77.04%$), MBLCA 方法引入不确定性建模后可以得到 0.52% 的性能提升,这证明了适当的不确定性引入以及多样性扰动对于 MBLCA 方法的不确定性建模十分重要. 特别是当分支数量固定时,Dropout 概率取 0.5 时一直取得最优的性能. 当 Dropout 概率固定时,分支数量取 6 时一直取得最优的性能. 从实验结果可以看出 $n=6$ 和 $p=0.5$ 是最优的参数组合.

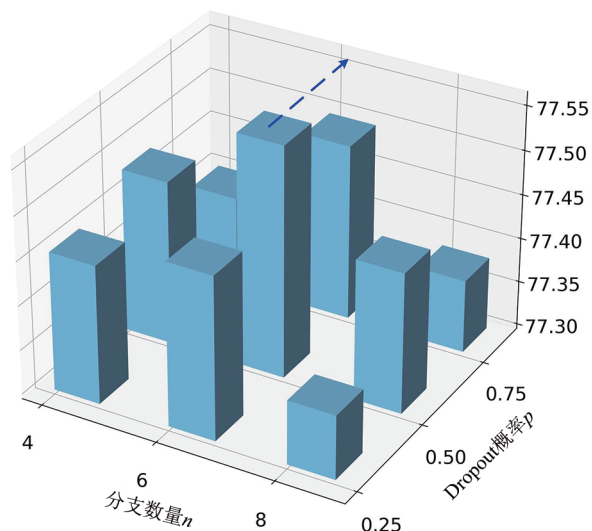
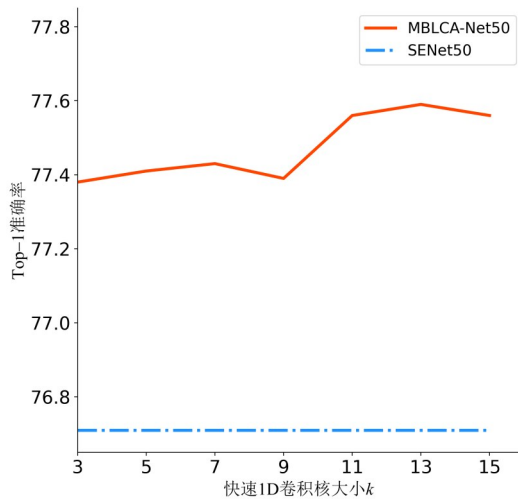


图2 MBLCA 模块对于不同分支数量 (n) 和 Dropout 概率 (p) 组合的消融实验

(2) **快速 1D 卷积核大小 k 的影响.** 接下来的实验采用上述最优的参数 ($n=6, p=0.5$) 进一步验证卷积核大小 k 的影响. 1D 卷积核的大小 k 控制着局部通道注意力机制的交互范围. 如图 3 所示,同样使用 ResNet-50 作为骨干网络,分别将 k 设置为 3, 5, 7, 9, 11, 13, 15 进行了 7 组实验. 相比于 SENet, MBLCA 模块在不同的设置下均保持着明显的优势. 同时, k 的变化也对模型的性能有一定的影响. 当 $k=13$ 时, MBLCA 方法取得了最优性能 (77.59%). 当 $k>13$ 时, MBLCA 的性能有所损

图3 MBLCA模块对于卷积核大小 k 的消融实验

失,同时会引入更多的计算代价.当 $k \leq 9$ 时,MBLCA模块的性能比较稳定,且均低于 $k=13$ 时的结果.同时可以看出此时MBLCA模块对于 k 的敏感度也比较低,说明相对较大局部交互范围中的不确定性会更突出,不确定性建模的作用也相应越明显.这种现象表明适当范围的局部通道相关性建模是十分重要的.实验结果表明,快速1D卷积核大小的最优设置是 $k=13$.综上所述,在后续的实验中均以 $n=6, p=0.5, k=13$ 的超参数组合作为默认设置.

4.2.2 比较当前最优的同类方法

首先,以ResNet-50和ResNet101作为骨干网络,本文比较了多种当前最优的视觉注意力方法,包括SENet^[8],CBAM^[9],A²-Net^[11],AA-Net^[10],GSoPNet1^[13]和ECA-Net^[15],实验结果如表2所示.在该部分实验中,与文献[15]一样使用自适应尺寸的1D卷积,可以得到0.1%~0.2%的性能提升.当ResNet-50作为骨干网络时,MBLCA方法相比于原始的网络模型,在几乎相同的模型复杂度的情况下可以带来2.58%的Top-1精度提升.与其他注意力模型相比,MBLCA方法利用更低的模型复杂度获得了几乎相同甚至更好的分类精度.当ResNet-101作为骨干网络时,MBLCA方法的Top-1精度比原始的ResNet-101高出2.02%,与其他同类方法相比,在效率和效果上实现了更佳平衡.与ECA-Net相比,嵌入MBLCA模块的模型具有一样的模型复杂度,同时带来了0.3%左右的精度提升.此外,10次随机试验中MBLCA方法具有更小的方差,即更好的稳定性.上述实验证明了MBLCA方法可以提升多种网络架构的分类准确率,同时令分类结果更加稳定.

为了验证MBLCA方法在更轻量化的网络上的有效性,在上述实验的基础上还以ResNet-18作为骨干网络与上述系列方法进行了对比,实验结果表明,在保持

与ResNet和ECA-Net几乎相同的模型复杂度的情况下,MBLCA方法分别带来了0.48%和0.1%的性能提升.同时,在10次重复实验中,MBLCA方法具有更稳定的性能.

最后,还与更深更复杂的卷积神经网络模型进行了对比,包括ResNet-200^[36],DenseNet^[7],Inception-v3^[37].如表2所示,与ResNet-200相比,MBLCA方法使用ResNet-101作为基础网络只有其近一半的模型复杂度,但是在Top-1的准确率上提升了0.65%.另外,基于ResNet-50的MBLCA模型比Inception-v3拥有更低的模型复杂度和更优的性能.与DenseNet-161相比,基于ResNet-50的MBLCA模型拥有相似的分类精度,但是仅利用接近一半的计算量(FLOPs).基于ResNet-101的MBLCA模型与DenseNet-161具有几乎一样的计算量,但是在Top-1精度上提升了1.2%.综上所述,MBLCA方法在模型分类准确率与模型复杂度之间实现了更好的平衡,在较低模型复杂度的条件下取得了与目前主流方法相近甚至更优的分类准确率.

表2 各种当前最优的视觉注意力方法的分类结果

方法	骨干网络	参数量	计算量	Top-1	Top-5	
ResNet	ResNet-18	11.15 M	1.70 G	70.40	89.45	
SENet		11.23 M	1.70 G	70.59	89.78	
CBAM		11.23 M	1.70 G	70.73	89.91	
ECA-Net		11.15 M	1.70 G	70.78 $\pm_{0.12}$	89.92 $\pm_{0.07}$	
MBLCA		11.15 M	1.70 G	70.88 $\pm_{0.06}$	89.89 $\pm_{0.03}$	
ResNet	ResNet-50	24.37 M	3.86 G	75.20	92.52	
SENet		26.77 M	3.87 G	76.71	93.38	
CBAM		26.77 M	3.87 G	77.34	93.69	
A ² -Net		33.00 M	6.50 G	77.00	93.50	
AA-Net		25.80 M	4.15 G	77.70	93.80	
GSoP-Net1		28.05 M	6.18 G	77.68	93.98	
ECA-Net		24.37 M	3.86 G	77.48 $\pm_{0.25}$	93.68 $\pm_{0.15}$	
MBLCA		24.37 M	3.86 G	77.78 $\pm_{0.12}$	93.89 $\pm_{0.07}$	
ResNet		ResNet-101	42.49 M	7.34 G	76.83	93.48
SENet			47.01 M	7.35 G	77.62	93.93
CBAM	47.01 M		7.35 G	78.49	94.31	
AA-Net	45.40 M		8.05 G	78.70	94.40	
ECA-Net	42.49 M		7.35 G	78.65 $\pm_{0.23}$	94.34 $\pm_{0.15}$	
MBLCA	42.49 M		7.35 G	78.85 $\pm_{0.11}$	94.48 $\pm_{0.08}$	
ResNet-200		74.45 M	14.10 G	78.20	94.00	
Inception-v3		25.90 M	5.36 G	77.45	93.56	
DenseNet-161		27.35 M	7.34 G	77.65	93.80	

4.3 MS COCO上的目标检测与实例分割实验

本节使用Mask R-CNN^[30]作为基础检测器在目标检测和实例分割任务上验证MBLCA模块的泛化能力.实验中所有检测器的训练都是在MS COCO2017训练集上完成的,并在MS COCO2017验证集上进行测试.实验结果如表3所示,本节在参数量、每秒浮点运算量以

表3 各种方法在MS COCO2017验证集上的目标检测和实例分割结果

方法	参数量	计算量	目标检测						实例分割					
			AP	AP ₅₀	AP ₇₅	AP _S	AP _M	AP _L	AP	AP ₅₀	AP ₇₅	AP _S	AP _M	AP _L
ResNet-50	44.18 M	275.58 G	37.2	58.9	40.3	22.2	40.7	48.0	34.1	55.5	36.2	16.1	36.7	50.0
+SE block	46.67 M	275.69 G	38.7	60.9	42.1	23.4	42.7	50.0	35.4	57.4	37.8	17.1	38.6	51.8
+ECA	44.18 M	275.69 G	39.0	61.3	42.1	24.2	42.8	49.9	35.6	58.1	37.7	17.6	39.0	51.8
+MBLCA	44.18 M	275.69 G	39.1	61.2	42.4	23.3	42.8	50.1	35.5	57.7	37.5	17.0	38.6	51.9
ResNet-101	63.17 M	351.65 G	39.4	60.9	43.3	23.0	43.7	51.4	35.9	57.7	38.4	16.8	39.1	53.6
+SE block	67.89 M	351.84 G	40.7	62.5	44.3	23.9	45.2	52.8	36.8	59.3	39.2	17.2	40.3	53.6
+ECA	63.17 M	351.83 G	41.3	63.1	44.8	25.1	45.8	52.9	37.4	59.9	39.8	18.1	41.1	54.1
+MBLCA	63.17 M	351.83 G	41.4	63.4	45.2	24.8	45.9	53.3	37.4	59.9	39.9	18.0	41.0	54.3

及平均准确率(mAP)等多个指标上对所有方法进行评估. 实验结果表明, MBLCA模块在目标检测和实例分割任务上都明显优于原始的ResNet. 用ResNet-50和ResNet-101作为骨干网络, 提出的MBLCA方法在目标检测(实例分割)任务上分别比SENet提高了0.4%(0.1%)和0.7%(0.6%)的平均准确率(mAP), 同时降低了模型的复杂度. 与ECA-Net50(ECA-Net101)相比, MBLCA方法在保证相同的推理复杂度的前提下, 也分别得到了0.1%的平均准确率提升. 上述结果表明, 提出的MBLCA模块在多种视觉任务上均具有良好的表现能力, 能够有效提升神经网络在下游任务上的泛化能力.

4.4 鲁棒性和稳定性实验

为了进一步验证MBLCA方法对噪声数据的鲁棒性和稳定性, 本文利用ImageNet-C^[34]噪声数据集进行了相关实验. ImageNet-C^[34]数据集在ImageNet-1K^[3]验证集数据的基础上引入了15种不同的扰动类型, 主要涵盖噪声(Noise)、模糊(Blur)、天气情况(Weather)、数字压缩(Digital)四大类扰动, 每种扰动的强度分为5个级别. 评估阶段采用文献[34]中的评价指标mCE来评价模型的泛化能力和稳定性, 该指标的数值越低表示模型的泛化能力和稳定性更强. 实验结果如表4所示, 从表4中可以看到: 当ResNet50(ResNet101)作为骨干网络时, MBLCA方法相比于基础模型ResNet50(ResNet101)的mCE值分别降低了5.5(3.77); 与SENet和ECA-Net相比, MBLCA方法在mCE指标上分别降低了1.6(2.6)和0.87(0.54). 上述结果进一步证明了MBLCA模块可以较为明显地提升模型对于噪声的鲁棒性和稳定性.

表4 各种方法在ImageNet-C数据集上测试结果

方法	错误率↓	mCE↓	Noise↓	Blur↓	Weather↓	Digital↓
ResNet50	24.80	76.70	81.67	80.50	69.00	77.50
+SE block	23.29	72.80	73.33	78.00	66.75	73.25
+ECA	22.52	72.07	73.67	77.75	67.50	69.75
+MBLCA	22.22	71.20	71.33	76.25	66.00	71.25
ResNet101	23.17	70.30	75.33	73.75	64.50	69.00
+SE block	22.38	69.13	70.33	75.00	62.50	69.00
+ECA	21.35	67.07	64.67	73.00	63.25	66.75
+MBLCA	21.15	66.53	64.33	73.25	62.25	65.75

5 结论

本文提出了一种多分支局部通道注意力模块(MBLCA), 核心思想是在局部通道注意力机制中引入不确定性建模, 从而提升注意力模型的泛化能力和稳定性. 实验结果表明, 该方法在图像分类、目标检测以及实例分割任务上都具有良好的性能. 本文主要探索了高效通道注意力机制中局部通道交互情况下的不确定性建模方法, 如何引入非局部通道交互和空间维度的注意力机制, 并对其进行不确定性建模仍是一个值得研究的方向. 未来将在该方向上进行深入探索, 并将该思想应用到更多的深度卷积神经网络架构中, 如MobileNets^[38]和Inception^[5,36,37,39].

参考文献

- [1] YOSHUA B, AARON C, PASCAL V. Representation learning: A review and new perspectives[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(8): 1798-1828.
- [2] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C]//Advances in Neural Information Processing Systems. Lake Tahoe, Ne: MIT Press, 2012: 1097-1105.
- [3] DENG J, DONG W, SOCHER R, LI L, LI K, LIFF. ImageNet: A large-scale hierarchical image database[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Miami, FL, USA: IEEE, 2009: 248-255.
- [4] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv: 1409.1556, 2015.
- [5] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA: IEEE, 2015: 1-9. DOI: 10.1109/CVPR.2015.7298594.
- [6] HE K, ZHANG X, REN S, SUN J. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Ve-

- gas, NV, USA: IEEE, 2016: 770-778.
- [7] HUANG G, LIU Z, MAATEN L, WEINBERGER K Q. Densely connected convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017: 2261-2269.
- [8] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 18-23.
- [9] WOO S, PARK J, LEE J, KWEON I S. CBAM: Convolutional block attention module[C]//European Conference on Computer Vision. Munich, Germany: Springer, 2018: 3-19.
- [10] BELLO I, ZOPH B, VASWANI A, SHLENS J, LE Q. Attention augmented convolutional networks[C]//Proceedings of the IEEE International Conference on Computer Vision. Seoul, Korea (South): IEEE, 2019: 3285-3294.
- [11] CHEN Y, KALANTIDIS Y, LI J, YAN S, FENG J. A²-Nets: Double attention networks[C]//Advances In Neural Information Processing Systems. Montreal, Canada: MIT Press, 2018: 352-361.
- [12] FU J, LIU J, TIAN H, LI Y, BAO Y, FANG Z, LU H. Dual attention network for scene segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE, 2019: 3146-3154.
- [13] GAO Z, XIE J, WANG Q, LI P. Global second-order pooling convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE, 2019: 3024-3033.
- [14] WANG X, GIRSHICK R, GUPTA A, HE K. Non-local neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 7794-7803.
- [15] WANG Q, WU B, ZHU P, LI P, ZUO W, HU Q. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE, 2020: 11531-11539.
- [16] DAVID J C MACKAY. A practical Bayesian framework for backpropagation networks[J]. *Neural Computation*, 1992, 4(3): 448-472.
- [17] GRAVES A. Practical variational inference for neural networks[C]//Advances in Neural Information Processing Systems. Red Hook, USA: MIT Press, 2011: 2348-2356.
- [18] GAL Y, GHAHRAMANI Z. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning[C]//International Conference on Machine Learning. New York, NY, USA: ACM, 2016: 1050-1059.
- [19] SRIVASTAVA N, HINTON G E, KRIZHEVSKY A, SUTSKEVER I, SALAKHUTDINOV R. Dropout: A simple way to prevent neural networks from overfitting[J]. *The Journal of Machine Learning Research*, 2014, 15(1): 1929-1958.
- [20] LIN T, MAIRE M, BELONGIE S J, HAYS J, PERONA P, RAMANAN D, DOLL'AR P, ZITNICK C L. Microsoft COCO: Common objects in context[C]//European Conference on Computer Vision. Munich, Germany: Springer, 2014: 8693: 740-755.
- [21] 仇祝令, 查宇飞, 吴敏, 等. 基于注意力学习的正则化相关滤波跟踪算法[J]. *电子学报*, 2020, 48(9): 1762-1768.
- QIU Z L, ZHA Y F, WU M, et al. Learning attentional regularized correlation filter for visual tracking[J]. *Acta Electronica Sinica*, 2020, 48(9): 1762-1768. (in Chinese)
- [22] QIN P, SHEN W, ZENG J. DSCA-Net: Indoor head detection network using dual-stream information and channel attention[J]. *Chinese Journal of Electronics*, 2020, 29(6): 1102-1109.
- [23] 盖杉, 王俊生. 基于深度学习的非局部注意力增强网络图像去雨算法研究[J]. *电子学报*, 2020, 48(10): 1899-1908.
- GAI S, WANG J S. Image raindrop algorithm research using nonlocal attention enhanced network based on deep learning[J]. *Acta Electronica Sinica*, 2020, 48(10): 1899-1908. (in Chinese)
- [24] CAO Y, XU J, LIN S, WEI F, HU H. GCnet: Non-local networks meet squeeze-excitation networks and beyond [C]//International Conference on Computer Vision Workshops. Seoul, Korea(South): IEEE, 2019: 1971-1980.
- [25] PAISLEY J W, BLEI D M, JORDAN M I. Variational Bayesian inference with stochastic search[C]//International Conference on Machine Learning. Edinburgh: ACM, 2012.
- [26] KINGMA D P, WELING M. Auto-encoding variational Bayes[C]//International Conference on Learning Representations. Banff, AB, Canada: IEEE, 2014: 1-14.
- [27] HOFFMAN M D, BLEI D M, WANG C, PAISLEY J W. Stochastic variational inference[J]. *The Journal of Machine Learning Research*. 2013, 14(1): 1303-1347.
- [28] 齐现英, 刘伯强, 徐建伟. 基于不确定性信息融合的高密度椒盐噪声降噪方法[J]. *电子学报*, 2016, 44(4): 878-885.
- QI X Y, LIU B Q, XU J W. A novel algorithm for removing high-density salt-and-pepper noise based on fusion of indeterminacy information[J]. *Acta Electronica Sinica*,

2016, 44(4): 878-885. (in Chinese)

- [29] 张挺, 刘金华. 一种新的空间数据不确定性重建方法[J]. 电子学报, 2018, 46(3): 641-645.
ZHANG T, LIU J H. A New indefinite reconstruction method for spatial data[J]. Acta Electronica Sinica, 2018, 46(3): 641-645. (in Chinese)
- [30] TAKASHI Y, SHINGO M, KOTARO H, et al. Recurrent neural networks with multi-branch structure[J]. IEEE Transactions on Electronics Information & Systems, 2007, 127(9): 1430-1435.
- [31] CHEN Y, LI J, XIAO H, JIN X, YAN S, FENG J. Dual path networks[C]//Advances in Neural Information Processing Systems. Long Beach: MIT Press, 2017: 4467-4475.
- [32] NAIR V, HINTON G E. Rectified linear units improve restricted Boltzmann machines[C]//International Conference on Machine Learning. Haifa: ACM, 2010: 807-814.
- [33] HE K, GKIOXARI G, DOLL'AR P, GIRSHICK R B. Mask R-CNN[C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE. 2017: 2980-2988.
- [34] DAN H, THOMAS G D. Benchmarking neural network robustness to common corruptions and perturbations[C]//International Conference on Learning Representation. New Orleans, LA, USA: IEEE, 2019.
- [35] LIN T, DOLL'AR P, GIRSHICK R B, HE K, HARIHARAN B, BELONGIE S J. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017: 936-944.
- [36] HE K, ZHANG X, REN S, SUN J. Identity mappings in deep residual networks[C]//European Conference on Computer Vision. Amsterdam, Netherlands: Springer, 2016: 630-645.
- [37] SZEGEDY C, VANHOUCKE V, IOFFE S, SHLENS J, WOJNA Z. Rethinking the inception architecture for computer vision[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016: 2818-2826.
- [38] SANDLER M, HOWARD A G, ZHU M, ZHMOGINOV A, CHEN L. Mobilenetv2: Inverted residuals and linear bottlenecks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 4510-4520.
- [39] SZEGEDY C, IOFFE S, VANHOUCKE V, ALEMI A A. Inception-v4, inception-resnet and the impact of residual connections on learning[C]//Proceedings of the Thirty-

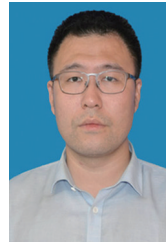
First AAAI Conference on Artificial Intelligence. San Francisco, USA: AAAI Press, 2017: 4278-4284.

作者简介



伍邦谷 男, 1996年生于湖北省. 硕士研究生, 计算机科学与技术专业, 研究方向为计算机视觉.

E-mail: wubangu@tju.edu.cn



张苏林 男, 1987年生于辽宁省. 中汽数据(天津)有限公司高级技术经理, 主要研究方向为智能网联汽车、自动驾驶控制及感知系统.

E-mail: zhangsulin@catarc.ac.cn



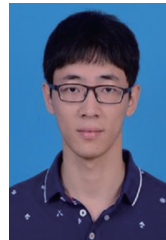
石红 女, 1975年生于内蒙古自治区. 副教授, 硕士生导师. 主要研究方向为粗糙集、机器学习.

E-mail: serena@tju.edu.cn



朱鹏飞 男, 1986年生于河南省. 副教授, 博士生导师. 主要研究方向为无人机视觉、人机协同学习、度量学习.

E-mail: zhupengfei@tju.edu.cn



王旗龙(通讯作者) 男, 1989年生于黑龙江省. 副教授, 硕士生导师. 主要研究方向为视频图像分析、深度学习等.

E-mail: qlwang@tju.edu.cn



胡清华 男, 1976年生于湖南省. 教授, 博士生导师. 主要研究方向为数据不确定性建模、多模态数据学习.

E-mail: huqinghua@tju.edu.cn