

一种并行不对称空洞卷积模块^①

张智杰, 尉 飞, 葛青青, 赵宝奇, 孙军梅, 李秀梅

(杭州师范大学 信息科学与技术学院, 杭州 311121)

通讯作者: 李秀梅, E-mail: xiumei_li@hotmail.com



摘 要: 构建卷积神经网络要耗费大量的人力资源, 且训练过程中需要消耗大量的算力资源. 利用空洞卷积代替卷积神经网络中的池化操作, 能有效增加感受野, 降低运算复杂度, 但是空洞卷积会带来空间层次和信息连续性的丢失. 本文提出了一种并行不对称空洞卷积模块, 该模块能够补全空洞卷积所丢失的信息, 可以嵌入到现有的卷积神经网络中, 代替 3×3 卷积进行网络训练, 从而加速网络的收敛, 提高网络的性能. 实验结果表明, 利用本文所提出的并行不对称空洞卷积模块, 可以显著提高不同网络在 CIFAR-10 等数据集上的分类效果.

关键词: 深度学习; 空洞卷积; 非对称卷积; 感受野

引用格式: 张智杰, 尉飞, 葛青青, 赵宝奇, 孙军梅, 李秀梅. 一种并行不对称空洞卷积模块. 计算机系统应用, 2021, 30(9): 206-211. <http://www.c-s-a.org.cn/1003-3254/8071.html>

Parallel Asymmetric Dilated Convolution Module

ZHANG Zhi-Jie, YU Fei, GE Qing-Qing, ZHAO Bao-Qi, SUN Jun-Mei, LI Xiu-Mei

(School of Information Science and Technology, Hangzhou Normal University, Hangzhou 311121, China)

Abstract: Creating a convolutional neural network consumes substantial human resources, and much computing power is needed during training. The application of dilated convolution instead of the pooling operation in the convolutional neural network can considerably increase the receptive field and reduce the computational complexity, but the dilated convolution will bring about the loss of spatial hierarchy and information continuity. This study proposes a parallel asymmetric dilated convolution module, which can fill in the information lost by dilated convolution and be embedded in the current convolutional neural networks to replace the 3×3 convolution for network training. As a result, network convergence is accelerated and network performance is improved. The experimental results show that the proposed module can significantly improve the classification of various classical networks on CIFAR-10 and other data sets.

Key words: deep learning; dilated convolution; asymmetric convolution; receptive field

卷积神经网络广泛应用于图像分类、语义分割和图像生成等计算机视觉领域, 其通过卷积核对图像中的局部区域做卷积, 以提取图像中的特征, 在每一层的卷积中通过参数共享以减少模型复杂度, 之后结合池化操作实现位不变性的识别. 现有的卷积神经网络普遍将 3×3 卷积作为其基本构建模块. 对于卷积神经网络而言, 卷积核的感受野、深度以及通道数决定了

网络的性能. 感受野越大, 表示特征图上像素点映射的区域越大; 深度决定了网络的抽象能力或学习能力; 通道数决定了卷积核所包含的信息量大小. 感受野和通道数共同决定了卷积层学习有效信息与空间的能力.

2012 年, 卷积神经网络在复杂模型下的有效性得到证明^[1]; 2015 年 Inception Net v1^[2] 以及 v2、v3、v4^[3] 证明了更多的卷积和更深的网络可以得到更好的分类

① 基金项目: 国家自然科学基金 (61801159, 61571174); 杭州师范大学星光计划

Foundation item: National Natural Science Foundation of China (61801159, 61571174); Starlight Program of Hangzhou Normal University

收稿时间: 2020-12-01; 修改时间: 2021-01-04; 采用时间: 2021-01-20; csa 在线出版时间: 2021-09-02

效果; 2016年, ResNet 解决了深度网络结构的训练问题^[4]. 传统的卷积操作主要有3个缺点: (1) 使用的是局部操作, 不能得到比较大的范围甚至图像的全局特征, 且卷积核是固定的尺寸; (2) 对物体的形状、姿态变化缺少适应性; (3) 当特征的通道数变大后, 卷积核的参数也变得庞大, 增加了运算量. 在传统的卷积神经网络中, 图像输入到卷积神经网络后, 先对图像做传统的卷积, 然后进行池化操作, 通过降低特征点的数量来扩大感受野^[5]. 由于图像分割是像素级的预测, 对池化后较小尺寸的图像上采样到的原始图像尺寸进行预测, 这种先池化后上采样的操作使得特征图的尺寸依次进行了缩小和增大, 该过程带来了信息的损失.

2015年, 空洞卷积被提出, 它是在标准的卷积核上注入空洞, 以达到扩大感受野的目的. 相较于传统卷积, 空洞卷积不仅可以保留数据的内部结构, 也可以获得上下文信息, 而且不需要下采样、池化等操作来进行特征的提取. 在 WaveNet^[6], ByteNet^[7] 等网络中, 也用到了空洞卷积, 以提升网络性能. 空洞卷积起源于图像分割领域, 也广泛应用于在分类任务中^[8]. 空洞卷积避免了池化所带来的信息丢失, 弥补了上述传统卷积的3个缺点^[9], 但同时带来了空间层次和信息连续性丢失等问题.

不对称卷积通常用于模型的加速和压缩. 在现有研究中, 研究者将 $n \times n$ 的对称卷积分解为 $1 \times n$ 和 $n \times 1$ 卷积, 减少了模型的参数. 如果二维卷积核的秩为1, 其运算可以等价转换成一系列的一维卷积. 在实际情况中, 由于卷积核具有分布特征值, 因此卷积核的秩要高得多, 直接应用会导致信息大量损失. Denton 等人用奇异值分解^[10], 找到了一个低秩逼近, 对上层进行精细化来恢复卷积性能. Jaberberg 等人通过最小化重构误差, 成功学习了水平核和垂直核^[11]. Jin 等人应用结构约束使二维卷积可分离, 在精度不变同时使计算时间加速2倍^[12]. 非对称卷积也广泛应用于网络结构的构成, 如 Inception Net 中, 1×7 和 7×1 卷积代替了 7×7 卷积. ENet 网络中也采用了不对称卷积设计网络, 虽然分类效果有所下降, 但是模型参数降低了33%, 降低了网络的复杂度^[13].

针对空洞卷积所带来的问题, 本文提出了一种新的并行不对称空洞卷积模块, 用 2×2 的空洞卷积、 1×3 和 3×1 的不对称卷积进行并列相加, 代替原有的 3×3 卷积, 以增强原有的卷积核. 本文主要贡献如下:

(1) 本文提出的并行不对称空洞卷积模块引入了空洞卷积, 在不增加参数的情况下, 可以扩大感受野, 捕获多尺度上下文信息.

(2) 本文提出的并行不对称空洞卷积模块, 着重在特定维度寻找特征, 补全空洞卷积在空间层次上缺失的信息, 使整个网络具有更好的连续性.

(3) 本文提出的并行不对称空洞卷积模块在 CIFAR-10^[3], 孟加拉文 (<https://www.kaggle.com/c/bengali-ai-cv19/data>) 等数据集上得到了较好的分类效果, 并可嵌入到任意的卷积神经网络中, 提高模型的特征表达能力.

1 并行不对称空洞卷积

本文提出了一种并行不对称空洞卷积模块, 如图1所示. 通过不对称卷积结构的特殊性, 使得模块在不改变感受野的情况下充分利用特征图所含信息, 提高了网络模型的特征表达能力. 新模块可以用来代替传统的连续卷积, 在不增加模型复杂度的情况下, 提升整个模型的准确率.

本文所提出模块的网络结构如表1所示.

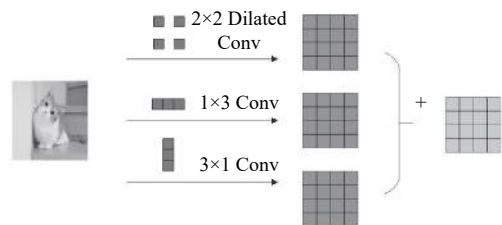


图1 并行不对称空洞卷积模块

表1 并行不对称空洞卷积模型结构

层	含义	结构
Conv	Conv	3×3 , 64, stride 2
	Dilated Conv	2×2 , 64, stride 2, dilated rate 2
DAA	Asymmetric Conv	1×3 , 64, stride 2
	Asymmetric Conv	3×1 , 64, stride 2
Conv	Conv	1×1 , 32, stride 1

1.1 空洞卷积的感受野

本文将从卷积操作出发, 在普通卷积上注入空洞, 连续卷积核进行卷积操作如式(1)^[9]:

$$(x * F)(p) = \sum_{s+t=p} x(t)F(t) \quad (1)$$

其中, $*$ 是二维卷积操作, 对于输入 $x \in \mathcal{R}^{C \times H \times W}$, $n \times n$ 卷积核 $f \in \mathcal{R}^{C \times K \times H}$, 输出 n 维特征图. 对于该层的第 j 个

卷积核, 相应的输出特征映射通道如式 (2):

$$Y_{i,j} = \sum_k^c x_{i,j,k} F_{i,j,k}^{(j)} \quad (2)$$

其中, $x_{i,j,k}$ 是 x 的第 k 个通道的特征图, 而 $F_{i,j,k}^{(j)}$ 代表的第 k 个尺寸为 $H * W$ 卷积核. 设 F 是一个离散过滤器即空洞卷积, 定义 l 为空洞率, $*$ 为空洞卷积操作:

$$(x * l F)(p) = \sum_{s+lt=p} x(t) F(t) \quad (3)$$

空洞卷积支持指数扩展的感受野, 不会丢失分辨率或覆盖范围. 将偶数尺寸的卷积核^[14]通过注入奇数个空洞, 其感受野范围等价于一个奇数尺寸连续卷积, 其对应的输出映射通道等价于式 (2), 因此空洞卷积能够代替连续卷积操作.

1.2 卷积并行性

对于经过普通卷积后再池化的操作而言, 空洞卷积不仅能够达到与其一致的效果, 而且减少了因池化所丢失的信息. 与此同时, 空洞卷积也带来了一些弊端. 空洞卷积操作会使一些信息因特征空间上的连续而丢失, 如图像角落和边界的信息. 注入过多的空洞数量可能会导致预测结果出现棋盘效应. 如果可以补充所丢失的连续信息, 则空洞卷积可以比普通卷积提取到更有效的特征.

将几个大小兼容的二维卷积核在相同的输入数据上进行相同步幅的卷积操作, 可以得到相同分辨率的输出^[15]. 通过将这些输出求和并在相应的位置上进行相加, 可以得到等效核的输出特征图. 即使卷积核大小不同, 二维卷积的可加性依旧成立. 因此可以引入两个不同类型的不对称卷积来补充所丢失的信息^[15].

$$I * K^{(1)} + I * K^{(2)} = I * (K^{(1)} \oplus K^{(2)}) = I * K^{(3)} \quad (4)$$

其中, I 是一个输入图片, $K^{(1)}$ 与 $K^{(2)}$ 是两个相兼容的卷积核, 将它们的输出相加, 操作上等价于卷积核 $K^{(3)}$ 的输出. 卷积核 $K^{(3)}$ 能够同时兼容 $K^{(1)}$ 与 $K^{(2)}$. 兼容是指尺寸较小的卷积核可以填充到较大的卷积核里.

如图 2 所示, $1 * 3$ 卷积核、 $3 * 1$ 卷积核和 $2 * 2$ 空洞卷积核因为形状的兼容性, 可以被 $3 * 3$ 的卷积核填充. 将 $1 * 3$ 卷积核、 $3 * 1$ 卷积核和 $2 * 2$ 空洞卷积核所形成的特征图依次相加, 其得到的特征图和 $3 * 3$ 卷积核所形成的特征图等价.

1.3 不对称卷积的补充作用

对于对称卷积而言, 不对称卷积更加注重特定维

度的特征, 其感受野与对称卷积的感受野不同. 例如 $1 * 3$ 卷积核在卷积的过程中, 更加倾向于列间特征的提取.

$$f_i = \sum_{k=1}^K \tau_k^i (h_k^i)^T \quad (5)$$

其中, f_i 表示 $n * n$ 大小的第 i 个卷积核, τ 和 h 是两个列向量, k 是 f 的秩. 因为卷积核的尺寸不同, 不对称卷积的特征提取操作与普通卷积的特征提取相比, 不仅减少了对称卷积带来的信息冗余, 而且在不对称卷积之间又引入了非线性激活函数. 因此, 不对称卷积的非线性提高了网络模型的拟合能力. 将互相兼容的不对称卷积得到的特征图相加, 也可以增强特征图的特征表达能力.

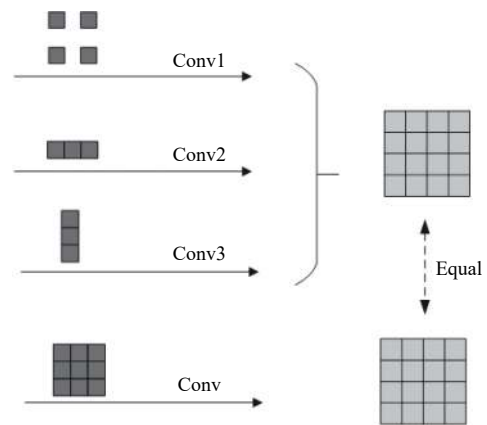


图 2 卷积并行性

1.4 计算量

$3 * 3$ 卷积核的感受野为 9, 计算量为 $9 * N * H * W * C$, 其中 N 是参数量, H, W 分别为输入特征图的长宽, C 是输出通道大小; 卷积核尺寸为 $2 * 2$, 空洞率为 2 的空洞卷积与卷积核尺寸为 $3 * 3$ 的卷积相比, 计算量为 $4 * N * H * W * C$, 在感受野相同的情况下, 计算量减少了一半. 与 $3 * 3$ 卷积相比, $1 * 3$ 卷积、 $3 * 1$ 卷积和 $2 * 2$ 空洞卷积三者的计算量总和仅有少量的参数增长.

2 实验及分析

2.1 实验设置

本文采用 CIFAR-10 与孟加拉文数据集对所提出模块的性能进行验证. CIFAR-10 数据集共有 60000 张彩色图像, 这些图像大小为 $32 * 32$, 分为 10 个类, 每类 6000 张图. 孟加拉文数据集为 Kaggle2020 比赛提供的数据集, 包括 10000 张孟加拉文手写体图片.

将本文所提出的并行不对称空洞卷积模块嵌入到不同的基础网络,如 VGG19、ResNet-34、ResNet-50、SENet-34 和 SENet-50 网络,并与原基础网络进行对比.本文使用十折交叉验证方法,损失函数为交叉熵损失函数,学习率初始化为 0.001, batchsize 为 128, epoch 为 100,均在 TITAN V 上进行试验.

2.2 实验结果分析

(1) 与基础网络进行比较

我们采用 CIFAR-10 与孟加拉文数据集,比较了 VGG19、ResNet-34、ResNet-50、SENet-34、SENet-50 五种经典的基础网络及其嵌入了本文所提出的并行不对称空洞卷积模块后网络的分类效果.在嵌入模块的网络中,均将模块嵌入到网络输出层位置,将连续卷积核和不对称空洞卷积模块的感受野面积都固定为 9.卷积核大小为 2*2,空洞率为 1 的空洞卷积在不对称卷积的信息补充下,其准确率与原基础网络相比有所提高.实验结果如表 2、表 3 及图 3、图 4 所示.通过对比可见,嵌入模块后网络的准确率均有比较明显的提升.实验结果表明,通过空洞卷积和不对称卷积对特征图的信息补偿,使得网络中所嵌入的并行不对称空洞卷积模块可以更好地提取特征,提高了模型的特征表达能力,提升了网络的分类效果.

表 2 在 CIFAR-10 数据集上的分类准确率比较 (%)

基础网络	基础网络的准确率	嵌入并行不对称空洞卷积模块后网络的准确率
VGG19	75.33	77.45
ResNet-34	85.53	87.28
ResNet-50	86.45	87.28
SENet-34	86.32	89.07
SENet-50	87.90	88.57

表 3 在孟加拉文数据集的准确率比较 (%)

基础网络	基础网络的准确率	嵌入并行不对称空洞卷积模块后网络的准确率
VGG19	58.33	59.36
ResNet-34	87.76	88.14
ResNet-50	91.64	92.35
SENet-34	93.28	93.32
SENet-50	93.89	94.06

(2) 与 ACNet 卷积比较

在该部分中,我们将并行不对称空洞卷积模块中的 2*2 空洞卷积替换成 3*3 的普通连续卷积后,则模块变成 ACNet 的不对称卷积模块 ACB^[15].我们将并行不对称空洞卷积模块与替换成普通连续卷积的 ACB

模块分别嵌入到 VGG19、ResNet-34、SENet-34 网络的输出层位置,保持相同的网络结构,对 CIFAR-10 数据集上的分类效果进行比较,实验结果如表 4 及图 5 所示.实验结果表明,嵌入并行不对称空洞卷积模块的网络比嵌入 ACB 模块的网络准确率更高,模型复杂度也更低.这是因为不对称卷积结构的特殊性,使其与空洞卷积的结构相匹配,与连续卷积搭配相比,准确率高;由于空洞卷积比连续卷积的参数要少,使得模型的复杂度也降低了很多.只有与空洞卷积相结合,不对称卷积的特征补偿能力才能更好地体现.

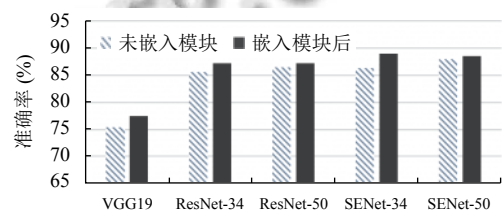


图 3 在 CIFAR-10 数据集上的准确率比较

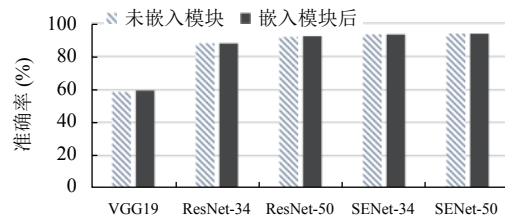


图 4 在孟加拉文数据集上的准确率比较

表 4 与 ACB 模块的准确率及网络复杂度比较

网络	嵌入 ACB 模块的网络准确率 (%)	嵌入并行不对称空洞卷积模块的网络准确率 (%)	嵌入 ACB 模块的网络复杂度 (M)	嵌入并行不对称空洞卷积模块的网络复杂度 (M)
VGG19	76.22	77.45	33.34	33.33
ResNet-34	81.53	89.07	33.30	33.21
SENet-34	84.33	86.32	34.50	34.46

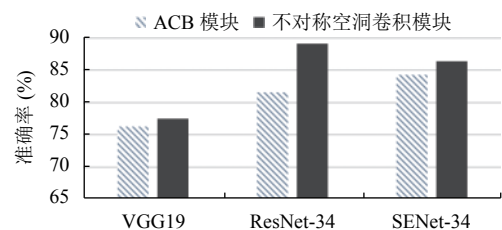


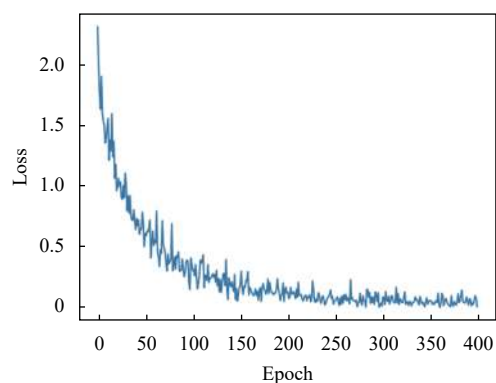
图 5 与普通卷积模块的准确率比较

2.3 损失特性分析

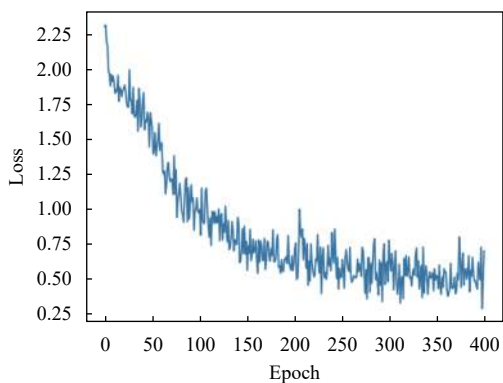
该部分将并行不对称空洞卷积模块分别嵌入到

ResNet-34 网络中的不同层并进行比较,其中损失函数的振荡变化及收敛情况如图 6 所示。

当模块嵌入到 ResNet-34 模型输出层位置时,损失函数的收敛趋势如图 6(a) 所示。当模块嵌入到 ResNet-34 模型网络输入层位置时,损失函数的收敛趋势如图 6(b) 所示。两者的迭代次数和运行环境均一致。可见,模块嵌入到模型的网络输出层位置时,其损失函数的振荡比模块嵌入到网络输入层时的振荡要小。这是因为该模块中含有不对称卷积,当放在与输入层比较近的位置时,会导致网络的信息损失变大,收敛速度变慢。而将模块嵌入到网络输出层,则会加速收敛的过程,提升网络的性能。



(a) 嵌入到模型输出层位置



(b) 嵌入到模型输入层位置

图 6 模块嵌入到模型不同位置时模型损失函数的收敛趋势

3 结论

本文提出了一种并行不对称空洞卷积模块,用不对称卷积补充空洞卷积,通过不对称卷积结构的特殊性,在不同维度上收集特征,使得模块在不改变感受野的情况下充分利用特征图的信息,提高了网络模型的特征表达能力。同时,该模块通过引入空洞卷积,使得

模型在与普通卷积复杂度相近的情况下,达到使用同样感受野尺寸的空洞卷积的分类效果。不对称卷积补充了空洞卷积所丢失的信息,补充了不同维度的信息,而且加速了整个网络的收敛过程,提高了网络的效率。本文将该模块嵌入到五种基本网络,并对 CIFAR-10 与孟加拉文数据集进行了分类比较,结果准确率均有较大的提高。该模块可嵌入其它任何网络模型中,帮助提高网络模型的分类效果,也可迁移到其它图像的分类任务当中。

参考文献

- 1 Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 2017, 60(6): 84–90. [doi: 10.1145/3065386]
- 2 Szegedy C, Liu W, Jia YQ, *et al.* Going deeper with convolutions. *Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition*. Boston, MA, USA. 2015. 1–9. [doi: 10.1109/CVPR.2015.7298594]
- 3 Szegedy C, Ioffe S, Vanhoucke V, *et al.* Inception-v4, inception-ResNet and the impact of residual connections on learning. *Proceedings of the 31st AAAI Conference on Artificial Intelligence*. San Francisco, CA, USA. 2016. 4278–4284.
- 4 He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV, USA. 2016. 770–778. [doi: 10.1109/CVPR.2016.90]
- 5 Lee CY, Gallagher PW, Tu ZW. Generalizing pooling functions in convolutional neural networks: Mixed, gated, and tree. *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*. Cadiz, Spain. 2016. 464–472.
- 6 van den Oord A, Dieleman S, Zen HG, *et al.* WaveNet: A generative model for raw audio. *Proceedings of the 9th ISCA Speech Synthesis Workshop*. Sunnyvale, CA, USA. 2016. 125.
- 7 Kalchbrenner N, Espeholt L, Simonyan K, *et al.* Neural machine translation in linear time. *arXiv*: 1610.10099, 2016.
- 8 Strubell E, Verga P, Belanger D, *et al.* Fast and accurate entity recognition with iterated dilated convolutions. *Proceedings of 2017 Conference on Empirical Methods in Natural Language Processing*. Copenhagen, Denmark. 2017. 2670–2680.
- 9 Yu F, Koltun V. Multi-scale context aggregation by dilated

- convolutions. Proceedings of the 4th International Conference on Learning Representations. San Juan, Puerto Rico. 2015.
- 10 Denton E L, Zaremba W, Bruna J, *et al.* Exploiting linear structure within convolutional networks for efficient evaluation. Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal, QC, Canada. 2014. 1269–1277.
- 11 Jaderberg M, Vedaldi A, Zisserman A. Speeding up convolutional neural networks with low rank expansions. Proceedings of 2014 British Machine Vision Conference. Nottingham, UK. 2014.
- 12 Jin J, Dondar A, Culurciello E. Flattened convolutional neural networks for feedforward acceleration. Proceedings of the 3rd International Conference on Learning Representations. San Diego, CA, USA. 2014.
- 13 Paszke A, Chaurasia A, Kim S, *et al.* ENet: A deep neural network architecture for real-time semantic segmentation. arXiv: 1606.02147, 2016.
- 14 Wu S, Wang GR, Tang P, *et al.* Convolution with even-sized kernels and symmetric padding. Advances in Neural Information Processing Systems 32. Vancouver, BC, Canada. 2019. 1194–1205.
- 15 Ding XH, Guo YC, Ding GG, *et al.* ACNet: Strengthening the kernel skeletons for powerful CNN via asymmetric convolution blocks. Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Republic of Korea. 2019. 1911–1920. [doi: [10.1109/ICCV.2019.00200](https://doi.org/10.1109/ICCV.2019.00200)]