引文格式：何民华,张润达,赵胜利. 基于 Att-DConv 的遥感舰船检测方法研究[J]. 地理空间信息, 2024, 22(3): 24-28. doi:10. 3969/j.issn. 1672-4623.2024.03.006

基于 Att-DConv 的遥感舰船检测方法研究

何民华[1]，张润达[1]，赵胜利[2]
（1. 交通运输部南海航海保障中心 广州海事测绘中心，广东 广州 510000；
2. 广州市地铁设计研究院股份有限公司，广东 广州 510000）

摘　要：针对遥感影像的舰船目标检测问题，提出了一种基于深度学习的舰船检测模型。首先利用空洞卷积组与通道注意力模块构成骨干网络，然后对所有特征提取层输出的不同尺度特征图进行拼接，再以融合后特征层分别进行上、下采样的方式构建了 4 个检测尺度的特征增强网络，最后采用改进的 NMS 算法优化最终的检测框输出。利用开源数据集 UCMerced_LandUse 与 FAIR1M 混合数据集对模型进行训练和测试，利用多种图像增强算法优化训练集质量，采用马赛克处理获取正样本更多的训练影像，并在未经处理的原始影像上进行测试。结果表明，该模型的精度均值可达 0.91，检测速度可达 34 f/s，对于不同复杂程度背景和尺度的舰船样本具有稳定的检测能力。

关键词：遥感影像；舰船检测；空洞卷积；通道注意力；融合特征增强
中图分类号：P237　　**文献标志码：**B　　**文章编号：**1672-4623（2024）03-0024-05

Research on Remote Sensing Ship Detection Method Based on Att-DConv

HE Minhua[1], ZHANG Runda[1], ZHAO Shengli[2]
(1. Guangzhou Maritime Surveying and Mapping Center, South China Sea Navigation Support Center, Ministry of Transport, Guangzhou 510000, China;
2. Guangzhou Metro Design and Research Institute Co., Ltd., Guangzhou 510000, China)

Abstract: Aiming at the problem of ship target detection in remote sensing images, we proposed a ship detection model based on deep learning. We used parallel dilated convolution groups and channel attention modules to form the backbone network at first. Then, we spliced the feature maps of different scales output by all feature extraction layers, and constructed four detection scales by upsampling and downsampling of the fused feature layers respectively. Finally, we adopted the improved NMS algorithm to optimize the final detection frame output. We used the the open source data set UCMerced_LandUse and FAIR1M combinative data set to train and test model, and used a variety of image enhancement algorithms to optimize the quality of training set. We used mosaic processing to obtain more training images with positive samples, performed tests on raw, unprocessed images. The results show that the average accuracy of model proposed in this paper can reach 0.91, the detection speed can reach 34 f/s, which has stable detection capability for ship samples with different complex backgrounds and scales.

Key words: remote sensing image, ship detection, dilated convolution, channel attention, fusion feature enhancement

航天遥感卫星技术的快速发展使得遥感卫星影像在时间、空间分辨率上取得了很大进步，以卫星遥感影像为基础的相关技术也在林业资源管理、军事侦察、应急救援等领域发挥了越来越重要的作用[1]。基于遥感影像的船舶目标检测技术对于提高海上航行效率、军事情报侦察以及海上交通安全等方面具有重要的应用价值，是当前遥感影像智能处理的热点研究方向之一。许多结合人工算子与机器学习的方法曾被用于检测遥感影像中的舰船目标，Tran T H[2]等曾提出一种结合视觉背景提取器与频谱残差的海上船舶检测算法；da Silva B R F[3]等也曾提出一种在候选图像中使用梯度直方图，再利用支持向量机和 K 近邻算法对舰船目标进行分类的算法，上述方法大多先通过传统算子来描述舰船目标特征，再通过浅层机器学习算法实现特征目标的检测。然而，受传统算子对目标非线性特征表达能力不足[4]、浅层机器学习模型泛化能力较差[5]等因素影响，上述经典方法在实施以小尺度目标为主的遥感影像船舶检测任务时，普遍存在误检、漏检率高，对不同背景下的船只目标不敏感等问题。近两年不依赖人工设计算子、非线性特征提取能力更强的深度学习卷积神经网络算法正成为目标检测、实例分割等领域的主流方法，已在多个领域取得了很多的应用

收稿日期：2022-03-30。
第一作者简介：何民华（1989—），本科，工程师，从事海洋测绘方面的工作，E-mail：hmhhua_198905@sina.cn。

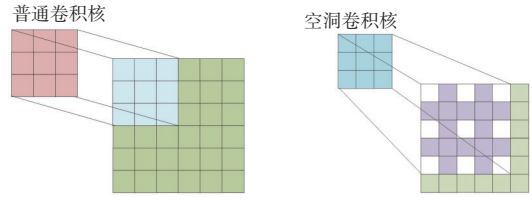
成果。刘冰洁[6]等运用Faster R-CNN与U-Net的组合设计了基于遥感影像的尾矿库检测方法，能在稀少样本条件下，通过特征提取与最优特征组合实现不同区域尾矿库高精度识别与提取，可为尾矿库科学管理工作提供一定技术支持；黄文涵[7]等针对复杂驾驶场景下的目标检测问题，以RetinaNet为基本框架提出了一种基于扩张卷积特征自适应融合的检测方法，在复杂驾驶场景数据集BDD100 K上的测试结果表明，该方法能有效提升目标检测算法在复杂驾驶场景下的检测性能；黄西尧[8]等针对地球同步轨道目标的检测问题，结合卷积神经网络与拓扑扫描的检测方法，在测试数据集上取得了80.7%召回率的检测结果；周佳玮[9]等提出了一种结合高分辨率和多时序遥感影像的深度学习水稻提取方法，通过全卷积网络对BJ-2号高分遥感影像进行分类，使用长短期记忆网络与随机森林Sentinel-2多时序遥感影像进行分类，再通过面向对象的分割和投票对3种方法的分类结果进行融合，最终取得了较好的分类效果，上述结果都充分说明了深度学习在目标检测与实例分割任务领域的优势；Kathiravan M[10]等基于R-CNN模型提出了一种基于合成孔径雷达影像的船舶检测算法，能对海上的船舶目标进行有效分类。

然而，相对于一般目标的检测任务，基于遥感影像的船舶检测任务具有更多的特殊性。首先，在目标尺度方面，不同类型船舶的跨度较大，这就需要检测模型拥有较好的小目标检测能力；其次，由于船舶目标的特点，除码头以外的大部分单幅遥感影像内仅包含极少数船舶目标，这就需要大量且丰富的样本类型才能充分训练模型。鉴于此，本文提出了一种基于卷积神经网络的舰船目标检测方法（Att-DConv），针对小目标特征，同时设计了4层全局融合结构作为特征增强网络，并利用改进的非极大值抑制算法优化选取最终的检测框。为了让检测结果与实际目标更贴合，使用带有旋转角度的锚点框实施检测；针对图像内样本个数较少且类型不丰富的问题，采用数字图像处理手段优化图像质量，以马赛克增强对训练数据进行处理，使模型得到更充分的训练。

1　Att-DConv模型结构设计

1.1　骨干网络

卷积核感受野是指卷积处理后获取的特征图在原始图像中映射的区域大小。骨干网络作为整个模型的前端结构，通过多个卷积特征提取层与下采样层提取到不同感受野的特征图来实现对不同尺度目标的检测输出。随着感受野的扩张，图像中的目标信息会不断丢失，一些小尺寸样本的特征信息甚至会被完全抹去。为了充分保留小尺寸样本特征对模型进行充分训练，在特征提取层中采用空洞卷积核进行特征提取。常规卷积核与空洞卷积核的计算过程见图1。



a 常规卷积　　b 空洞卷积（d=1）

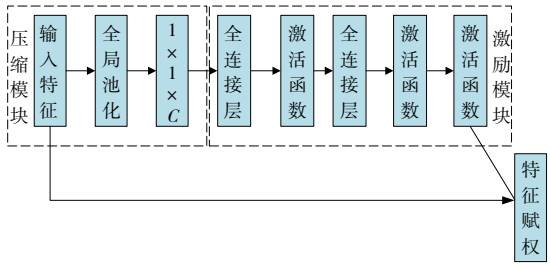
图1　常规卷积与空洞卷积

$$C_i^l \times k_i^l(f) = \sum_{a+b=f} C_i^l k_i^l(b) \tag{1}$$

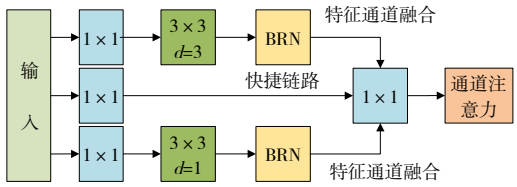
$$C_i^l \times d \times k_i^l(f) = \sum_{a+bd=f} C_i^l k_i^l(b) \tag{2}$$

式中，C_i^l为第l层特征提取层中第i个卷积核的输出；$k_i^l(f)$为进行操作的卷积核；f和a、b分别为对应的特征及其元素；d为空洞卷积的分离率。

在特征提取层中使用空洞卷积核能够保留较多的小目标特征，但持续使用相同分离率的空洞卷积核则会造成采样不连续的问题[11]，进而影响小目标特征信息的完整度。因此，在特征提取层中采用并联的方式组合不同分离率的空洞卷积核，同时添加快捷链路拼接原始特征图与提取后特征图，以提高特征图的复杂程度。为了避免模型因层数过深而产生训练数据的内部协变量偏移问题，以及训练阶段每次送入模型的训练样本数量batch设置较小时产生的非独立同分布和minibatch问题[12]，在每个卷积核的输出端后采用批量规范化层对输出特征进行规范化处理，避免模型在训练过程中出现梯度消失问题。遥感影像通常尺寸较大且背景复杂，而舰船目标在影像内的数量较少，这就产生了较严重的正负样本不平衡现象。为缓解该问题，在特征提取层末端设置了压缩—激励组合的通道注意力模块，让模型更聚焦于学习正样本特征信息。通道注意力模块首先在压缩操作层中将输入的多通道二维平面特征图拉长为$1\times1\times C$维度的向量，C为输入特征的通道维度；然后使用带有激活函数与全连接层的激励操作层通过两个全连接层生成每个通道的权重参数；最后将每个通道的权重与原始特征图相乘，实现对正负样本特征的聚焦与抑制。通道注意力机制和特征提取层结构见图2，以输入图像尺寸为608×608为例，骨干网络详细结构参数见表1。



a 通道注意力机制模块



b 特征提取层结构

图2　特征提取模块结构图

表1　骨干网络结构表

层数	输入尺寸	操作层	卷积核大小	是否包含注意力层
	608×608×3	Focus	–	–
1	304×304×32	DConv	3×3	√
1	304×304×64	DownSample	3×3(setp=2)	–
2	152×152×64	DConv	3×3	√
2	152×152×128	DownSample	3×3(setp=2)	–
3	76×76×128	DConv	3×3	√
3	76×76×256	DownSample	3×3(setp=2)	–
4	38×38×256	DConv	3×3	√
4	38×38×512	DownSample	3×3(setp=2)	–
5	19×19×512	DConv	3×3	√
5	19×19×1 024	DownSample	3×3(setp=2)	–

为了让模型更好地学习非线性特征，选取参数化的ReLU函数作为特征提取后的激活函数，具体公式为：

$$\mathrm{PReLU}(x)=\begin{cases}x_i & x_i>0\\ a_i x_i & x_i\leqslant 0\end{cases} \tag{3}$$

式中，x_i为第i个通道的输入特征；a_i为可学习动态更新的参数。

为尽量保留原始的目标特征信息，避免池化层在下采样时造成大量的特征损失，利用1×1的可学习卷积核作为算子，采用两倍步长下采样的方式实现对特征图的尺寸下采样。

1.2　特征增强网络

特征增强网络能在骨干网络的基础上提高特征图的信息复杂度，并输出多尺度的特征图参与最终的检测框输出。为更好地实现不同感受野下特征图间的信息交互，首先采用全局融合模式的特征增强网络，以全部的输出特征图为基础，通过下采样与上采样实现尺寸归一化；再将其融合为一张多通道的特征图，该特征图内包含了全部感受野下特征图的信息；然后根据船舶以小目标为主的尺寸特征，以融合特征图为基础进行两次上采样与两次下采样；最后输出4个层次的特征图进行多尺度目标的检测，具体结构见图3。

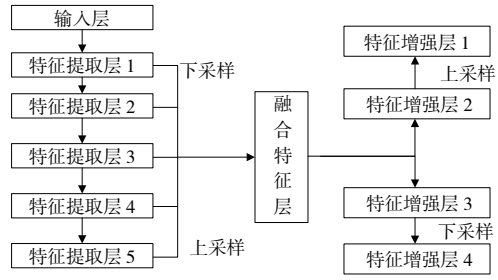


图3　特征增强层结构图

为了让检测框与真实目标更贴近，本文采用带有旋转角度的目标框来实施检测。Att-DConv在训练过程中使用的损失函数由边框回归损失函数L_{loc}与目标置信度回归损失函数L_{conf}的加权和组成，具体公式为：

$$\begin{cases}L_{\mathrm{loc}}=\dfrac{1}{N}\mathrm{obj}_n\sum\limits_{i=1}^{N}\dfrac{L_{\mathrm{reg}}(\grave{v}_n-v_n)}{|L_{\mathrm{reg}}(\grave{v}_n-v_n)|}+\mathrm{SkewIoU}\\ L_{\mathrm{reg}}=\mathrm{smooth}_{L1}\left(\grave{v}_\theta-v_\theta\right)-\mathrm{IoU}(\grave{v}_n-v_n)\end{cases} \tag{4}$$

$$L_{\mathrm{conf}}(x,c)=-\sum\nolimits_{i\epsilon\mathrm{Pos}}^{N}x_{ij}^{p}\log\left(c_j^p\right)-\sum\nolimits_{i}^{N}x_{ij}^{p}\log\left(c_j^o\right) \tag{5}$$

式中，N为锚点框的数量；obj_n为区分正、负样本的二元参数；L_{reg}为改进的Smooth L1函数；$\grave{v}_n$、v_n分别为预测框与真实框的偏移量；$\grave{v}_\theta$、v_θ分别为预测角度与真实角度的偏移量；SkewIoU为旋转预测框与旋转真实框间的重叠度；c为样本的目标置信度；x_{ij}^{p}为第j个正样本。

1.3　改进NMS算法

NMS算法通过对众多近似的检测框结果进行打分而剔除冗余边框。经典NMS算法是以目标框与真实框的交并比（IoU）为标准进行打分，在保留得分最高目标框的基础上直接剔除剩余所有候选框。该方法剔除标准严格，且仅考虑了两框之间的重叠面积，对各种类型检测框的描述不够全面。因此，本文采用加权平均GIoU算法作为最终检测框的筛选算法，同时以多个检测框的加权结果生成最终的输出框，具体公式为：

$$M=\frac{\sum\limits_i w_i B_i}{\sum\limits_i w_i}\quad B_i\in\{B|\mathrm{GIoU}(M,B)\geqslant\mathrm{therehold}\}\cup M \tag{6}$$

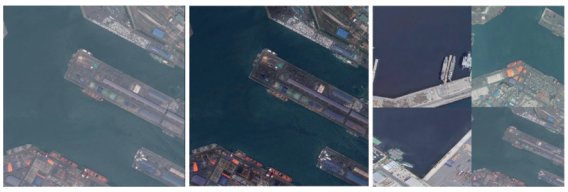
$$\mathrm{GIoU}=\mathrm{IoU}-\frac{|C(A\cup B)|}{|C|} \tag{7}$$

式中，w_i为权重值；B_i为第i个检测框；M为输出的检测框；C为检测框与实际框的最大包含面积。

2　实验与结果分析

2.1　数据集预处理

以开源数据集UCMerced_LandUse与FAIR1M为数据源获取混合遥感影像船舶数据集，对Att-DConv进行训练和测试。将所有包含舰船目标的遥感影像按照8∶2的比例划分为训练集和测试集，首先对训练集中的影像尺寸进行归一化处理；然后为提高训练后模型的泛化性，对训练数据进行直方图均衡化、暗通道先验滤波、高斯滤波等增强处理，并选取部分样本进行马赛克拼接，即通过将4个训练样本拼接为一张的方式来提高样本的丰富程度，部分处理结果见图4；最后将增强处理和马赛克处理后的影像放入训练集中，参与训练的影像为874张，包括目标1 458个，测试数据集内的影像未经过任何处理。



a 原始图像　b 增强影像　c 马赛克处理

图4　数据集预处理结果图

2.2　评价指标

为了客观全面地衡量模型性能，本文采用多种指标来评价提出模型的检测精度、检测速度和模型体量。在检测精度方面，采用精确率、召回率以及IoU阈值为0.5的精度均值（$AP_{0.5}$）作为评价指标；在检测速度方面，采用每秒检测图像帧数（FPS）作为评价指标；在模型体量方面，采用训练后模型权重文件大小和检测运行参数量作为评价指标。AP为精度—召回率曲线下的面积，相关计算公式为：

$$\begin{cases} \text{precision}=\dfrac{TP}{TP+FP} \\ \text{recall}=\dfrac{TP}{TP+FN} \\ AP=\displaystyle\int_0^1 P(r)\mathrm{d}r \end{cases} \tag{8}$$

式中，TP为真样本；FP为错误检测的真样本；FN为错误检测的假样本。

2.3　模型训练

模型的训练和测试均在相同的环境下进行，具体配置为：CPU Intel I7 i9800k、GPU Nvidia 1080Ti、内存大小16 GB、硬盘1 TB、操作系统Ubuntu 18.04、CUDA版本11.1，机器学习框架为以TensorFlow（1.13.1）为backend的Keras（2.2.4）。为使模型在训练过程中快速充分地拟合，在训练过程中配置了多种优化策略，经过多次调参后完成训练，具体参数配置：输入模型训练的图像尺寸为608×608，Batch Size设置为10，采用退火算法实现学习率的动态调节，初始学习率设置为10e-4，使用自适应矩估计算法更新训练权重，动量参数设置为0.9，迭代次数设置为1 000。为了验证本文算法的有效性，以单阶段的SSD、YoloV4、YoloV5s、RetinaNet模型，二阶段的Faster RCNN、Cascade RCNN模型，带有旋转框的Yolov5m-obb模型为基准，采用相同的运行环境与数据集进行训练，并使用同一组指标对所有模型进行评价。完整技术流程见图5。

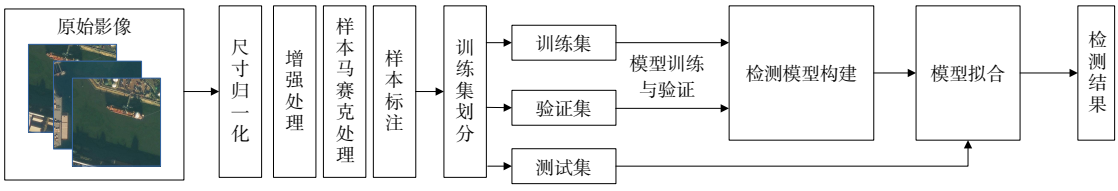


图5　完整技术流程图

2.4　测试与结果分析

全部模型的测试结果见表2，本文模型的部分检测结果见图6，可以看出，本文模型在检测精度方面优于所有对比模型，$AP_{0.5}$达到了0.85，比单阶段的SSD、YoloV4、YoloV5s、RetinaNet以及YoloV5m-obb分别提高了24.65%、8.33%、15.18%、62%和5.8%，相较于经典Faster RCNN和Cascade RCNN分别提高了49.2%、37.9%，说明Att-DConv在船舶检测任务中具有更明显的精度优势；在检测速度方面，测试环境下Att-DConv能达到34 f/s，相较于YoloV4、SSD模型均有显著优势，但稍低于YoloV5s模型；在模型结构方面，本文模型在骨干网络中设置了空洞卷积核组注意力模块，同时输出了4个尺度的特征图实现最终检测，而YoloV5s作为YoloV5系列的轻量级模型，拥有较小的参数量，但在检测精度方面表现不理想。

表2 模型评价结果对比表

模型名称	精确率/%	召回率/%	$AP_{0.5}$	FPS/(f/s)
Att-DConv	0.83	0.87	0.91	34
SSD	0.68	0.74	0.73	19
YoloV4	0.81	0.86	0.84	38
YoloV5s	0.78	0.81	0.79	62
YoloV5m-obb	0.82	0.89	0.86	54
RetinaNet	0.52	0.58	0.56	17
Faster RCNN	0.57	0.62	0.61	8
Cascade RCNN	0.63	0.68	0.66	9

进一步通过不同复杂程度背景的影像来直观说明本文模型的检测能力，在图6 a、6 b、6 e、6 f中包含较多的港口和邻近建筑物，背景复杂且负样本居多，剩余图像则没有太多负样本且样本个数较少，从检测结果可以看出，Att-DConv模型能准确检出不同背景复杂程度、不同密集程度的舰船目标，说明本文模型在具有较高检测精度的同时，还具有良好的泛化能力。

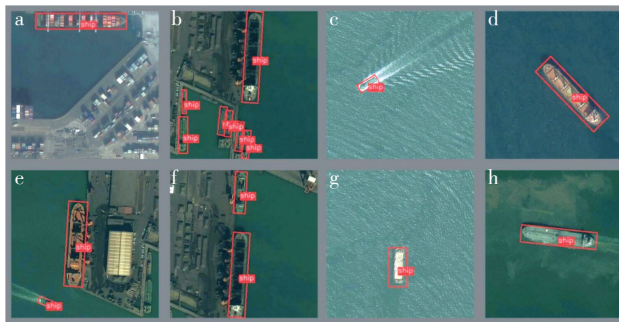


图6 部分检测结果

3 结 语

针对基于卫星遥感影像船舶场景下的背景复杂、小尺寸样本居多、正负样本不均衡等问题，本文提出了一种带有空洞卷积与通道注意力组合的遥感影像舰船检测模型Att-DConv。在骨干网络中使用并联空洞卷积核组进行特征提取，同时在提取层内引入通道注意力机制来聚焦正样本特征；再通过上下采样组合融合骨干网络输出的全部尺度特征图；最后利用增强数据集完成模型训练。实验结果表明，在多种复杂场景下Att-DConv对不同密集程度的舰船目标均表现出良好稳定的性能，$AP_{0.5}$达到0.91，显著优于现有主流深度学习模型，在测试环境中的检测速度可达到34 f/s，低于YoloV5s，但可开展对舰船目标的实时检测。在下一阶段的研究中，将尝试采用结构剪枝、模型量化等方法对现有模型进行轻量化压缩，然后通过部署在Nvidia Jetson等边缘硬件上实现低功耗环境下的离线推理；还将尝试通过生成对抗网络来模拟极端天气等复杂场景下的舰船遥感影像，并使用带有角度的目标框对全部样本进行重新标注，通过追加训练，进一步提高模型的泛化能力和目标定位的准确性。

参考文献

[1] 薛白,王懿哲,刘书含,等. 基于孪生注意力网络的高分辨率遥感影像变化检测[J]. 自然资源遥感,2022,34(1):61-66

[2] Tran T H, Le T L. Vision Based Boat Detection for Maritime Surveillance[C]. IEEE International Conference on Electronics,Information,and Communications,2016:1-4

[3] da Silva B R F, Nogueira M B, Alsina P J, et al. Study on Detection of Boats Using Satellite Imagery for Use on Unmanned Aerial Vehicles[C]//IEEE Latin American Robotics Symposium and Brazilian Symposium on Robotics, 2017:1-5

[4] 谢富,朱定局. 深度学习目标检测方法综述[J]. 计算机系统应用,2022,31(2):1-12

[5] 刘伟权,王程,臧彧,等. 基于遥感大数据的信息提取技术综述[J]. 大数据,2022,8(2):28-57

[6] 刘冰洁,邢旭东,吴浩,等. 深度学习支持下的尾矿库遥感识别方法[J]. 测绘科学,2021,46(12):129-139

[7] 黄文涵,殷国栋,耿可可,等. 基于扩张卷积特征自适应融合的复杂驾驶场景目标检测[J]. 东南大学学报(自然科学版),2021,51(6):1 076-1 083

[8] 黄西尧,何羿霆,杜华军,等. 地球同步轨道目标物深度学习检测方法[J]. 宇航学报,2021,42(10):1 283-1 292

[9] 周佳玮,涂理林,陈洪建,等. 融合空间和时序遥感信息的深度学习水稻提取[J]. 地理空间信息,2022,20(2):39-44

[10] Kathiravan M, Reddy N A, Prakash V, et al. Ship Detection from Satellite Images Using Deep Learning[C]. IEEE 7th International Conference on Communication and Electronics Systems,2022:1 044-1 050

[11] Chen L C, 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,2017,40(4):834-848

[12] Ioffe S. Batch Renormalization: Towards Reducing Minibatch Dependence in Batch-normalized Models[J]. Advances in Neural Information Processing Systems, 2017(30):239-248