

引用格式:郑志强,刘妍妍,潘长城,等.改进YOLO V3遥感图像飞机识别应用[J].电光与控制,2019,26(4):28-32. ZHENG Z Q, LIU Y Y, PAN C C, et al. Application of improved YOLO V3 in aircraft recognition of remote sensing images[J]. Electronics Optics & Control, 2019, 26(4):28-32.

改进YOLO V3遥感图像飞机识别应用

郑志强¹, 刘妍妍¹, 潘长城¹, 李国宁²

(1. 长春理工大学, 长春 130022; 2. 中国科学院长春光学精密机械与物理研究所, 长春 130033)

摘要: 为了准确识别遥感图像中的飞机,基于YOLO V3算法,通过使用K-means算法对数据集进行聚类分析,借鉴Densenet网络的思想,将YOLO V3网络中的两个残差网络模块替换为两个密集网络模块,改进为一种Dense-YOLO深度卷积神经网络结构。对改进前与改进后的网络进行训练,分别选出使两个网络识别效果最好的权重文件,针对高质量遥感图像与过度曝光、云雾遮挡等低质量遥感图像分别进行测试与分析。实验结果表明,新改进的深度卷积神经网络应用在两种图像上效果均有提升。其中,改进的算法在高质量的遥感图像中准确率高达99.72%,比原始算法准确率提升了0.85%,召回率高达98.34%,召回率提升了1.94%。在低质量遥感图像中准确率高达96.12%,比原始算法准确率提升了5.07%,召回率高达93.10%,召回率提升了19.75%。

关键词: 卷积神经网络; YOLO V3; K-means; Densenet; 飞机识别; 遥感图像

中图分类号: TP753 **文献标志码:** A **doi:**10.3969/j.issn.1671-637X.2019.04.006

Application of Improved YOLO V3 in Aircraft Recognition of Remote Sensing Images

ZHENG Zhi-qiang¹, LIU Yan-yan¹, PAN Chang-cheng¹, LI Guo-ning²

(1. Changchun University of Science and Technology, Changchun 130022, China; 2. Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China)

Abstract: In order to accurately identify the aircrafts in remote sensing images, K-means algorithm is used to carry out a clustering analysis on the dataset based on YOLO V3 algorithm. By referring to Densenet theory, the two residual network modules in YOLO V3 are replaced by two dense network modules, and a Dense-YOLO Deep Convolution Neural Network (DCNN) is developed. The networks before and after improvement are trained respectively, and the weight files with which the two networks have the best recognition results are selected. The high-quality remote sensing images and the low-quality remote sensing images having the problems of over-exposure and cloud occlusions are tested and analyzed respectively. The results show that the application of the improved DCNN to the two kinds of images improves the recognition performance. In high-quality remote sensing images, the accuracy rate of the improved algorithm is as high as 99.72%, which is improved by 0.85%; the recall rate of the improved algorithm is as high as 98.34%, which is improved by 1.94%. In low-quality remote sensing images, the accuracy rate of the improved algorithm is as high as 96.12%, which is improved by 5.07%; the recall rate of the improved algorithm is as high as 93.10%, which is improved by 19.75%.

Key words: CNN; YOLO V3; K-means; Densenet; aircraft recognition; remote sensing image

0 引言

近几年,随着航天、遥感与传感器等相关技术的不断

进步,遥感图像中所包含的数据和信息越来越丰富,这也为研究人员对遥感图像的分析工作提供了便利。飞机在民用生活和军事作战中都扮演着相当重要的角色,通过遥感图像识别世界各地的飞机是一项很重要的技术。

随着深度学习技术的不断发展,多种多样的目标识别与检测算法被提出,逐渐取代传统算法。随着科

收稿日期:2018-05-23

修回日期:2019-02-26

基金项目:国家“八六三”项目(863-2-5-1-13B)

作者简介:郑志强(1995—),男,吉林吉林人,硕士生,研究方向为目标检测技术及嵌入式技术。

技的发展,硬件设施的处理速度也不断提高,使深度卷积神经网络在硬件上的实时实现成为可能。利用深度卷积神经网络对图像进行识别与检测的方式应用的范围越来越广,效果也越来越好。随着深度残差网络和深度密集网络等网络的提出,深度卷积神经网络的层数越来越深,随着网络加深所带来的过拟合现象也大大减少,识别效果也越来越准确。

深度卷积神经网络可以自主地对要识别的目标进行学习,完善自己的模型,整个过程不需要人工干预。目前应用较为广泛的卷积神经网络目标识别与检测的方法可以分为两类:第一类是基于区域的目标识别方法,如 Faster R-CNN^[1], Mask R-CNN^[2] 等,这类方法在小目标的检测上效果较好,但是检测速度很慢;另一类是基于回归的目标识别方法,如 SSD^[3], YOLO 等。其中,基于回归的目标识别方法采用的是端到端的目标检测和识别,速度相对于基于区域的目标识别方法快很多,可以基本满足实时性的要求。

本文以基于回归的目标识别方法的深度卷积神经网络 YOLO V3 为基础,将遥感图像中的飞机作为目标,对 YOLO V3 网络的结构进行修改,并对层级结构中的参数进行调整,改进为一种带有密集卷积块的 Dense-YOLO 神经网络结构,其不仅能准确识别遥感图像中拍摄到的正常情况下的飞机,而且对不同卫星拍摄出的曝光强度不同甚至过度曝光、云雾遮挡等情况下的遥感图像中的飞机都有很好的识别效果。实验结果表明:对正常情况下的遥感图像中飞机的检测准确率在 99% 以上,召回率在 98% 以上,虚警率在 1% 以下;对过度曝光和云雾遮挡等复杂条件下的检测准确率在 96% 以上,召回率在 93% 以上,虚警率在 4% 以下。改进后的网络相比原网络各项指标有大幅提高,并且在电脑上仿真使用 Nvidia 1070 Ti 显卡时处理速度可达 14 帧/s。

1 基本概念

卷积神经网络^[4-5]是一种特殊的人工神经网络,卷积神经网络与传统的 BP 神经网络有很大差异,BP 神经网络在层与层之间采用全连接的方式,而卷积神经网络采用的是局部连接,与局部感受野十分类似,并且卷积神经网络同一层中的神经元与上一层连接的权值是共享的。卷积神经网络可以将一整幅图像作为数据的输入,图像的局部区域作为最小的输入单位输入到网络,通过局部连接和权值共享大大减少训练的参数,使训练的难度降低。

1.1 YOLO V3 网络介绍

YOLO(You Only Look Once)算法最初是 REDMON^[6]

等在 2016 年提出的一种基于回归的目标识别方法,到 2018 年已经发展到了第 3 代 YOLO V3^[7]。正如它的英文名一样,它只需要做一遍前向运算就可以对多种物体进行检测,所以 YOLO 系列算法检测速度很快。

YOLO V3 网络依旧保持了 YOLO V2^[8] 网络的快速检测的优点,并且识别的正确率大大提高,尤其在小目标的检测与识别上,准确率有很大提升。YOLO V3 网络借鉴了残差神经网络的思想,引入多个残差网络模块和使用多尺度预测的方式改善了 YOLO V2 网络在对小目标识别上的缺陷,因为检测的准确率高并且时效性好,该算法是目前目标检测领域最优秀的算法之一。该模型使用了很多有良好表现的 3×3 和 1×1 的卷积层,在后面进行多尺度预测时也使用了一些残差网络的结构,最终它有 53 个卷积层,因此也可把它们称作 Darknet-53。

YOLO V3 网络引入 Faster R-CNN 中使用 anchor boxes^[9] 的思想,针对 COCO 数据集和 VOC 数据集都采用了 3 个尺度进行预测,每一个尺度有 3 个 anchor boxes,尺度大的特征图使用小的先验框,因此可以根据自己要识别的目标选择适当的先验框 anchor boxes,并且根据准备预测的尺度对网络结构进行修改。

1.2 Densenet 网络介绍

2017 年 HUANG 等提出了新的网络结构 Densenet^[10],这个网络借鉴了 Resnet^[11] 思想,从图像的特征入手,通过对特征的极致利用达到更好的效果和更好的参数。继续延续残差神经网络的思路,并且将其网络结构变得更加紧密,在保证网络中层与层之间最大程度的信息传输前提下,直接将每个密集卷积块中的所有层连接起来。

Densenet 网络结构中包含 3 个密集卷积块,其中,每个密集卷积块中包含 4 个卷积层,在每个密集卷积块中,每一个卷积层可以得到所有之前的卷积层的输出作为输入,每个密集卷积块的结构如图 1 所示,相邻的卷积层之间通过卷积层和池化层相连接。

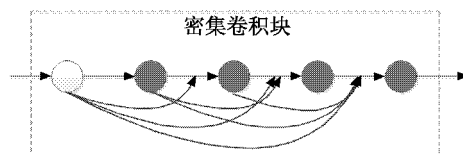


图 1 密集卷积块结构图

Fig. 1 Architecture diagram of dense convolution block

Densenet 网络引入这样的密集卷积块有很多优点:它可以像残差神经网络一样减轻梯度消失的情况;使用这样的紧密连接结构可以加强图片特征的传递,更加有效地利用了图片的特征,并且在一定程度上减少了参数的数量。本文设计的优化方案正是利用密集

卷积块的这些特点,引入密集神经网络加强图片特征的传递,使网络可以更好地识别遥感图像中的小飞机。

2 算法改进

相对于整幅遥感图像,飞机是一个小目标,所以原网络定义的 anchor boxes 与层级结构并不适用于本文。首先对数据集进行聚类分析,再针对小目标的检测,借鉴 Densenet 的思想,对网络中的层级结构进行改进,满足对飞机目标识别的要求。

2.1 基于数据集的聚类的改进

YOLO V3 网络使用 anchor boxes 作为先验框对图像中的目标进行检测。在 Faster R-CNN 和 SSD 中需要手工设置先验框,这样会导致选择的主观性比较强,如果可以选择一个维度合适的先验框,那么深度卷积神经网络将会更容易学习,从而做出更好的预测,所以 YOLO V3 网络中采用 K-means 算法对数据集的目标框大小进行聚类。

K-means 算法中通常使用欧氏距离、曼哈顿距离、切比雪夫距离等方式作为距离量度对两点间的距离进行计算。由于设置先验框的主要目的是为了使得预测框与 ground truth 的 IOU 更好,这样就导致使用这些常用距离在样本上不能产生很好的效果,如使用欧氏距离会让大的 bounding boxes 比小的 bounding boxes 产生更多的错误,而本文希望能通过 anchor boxes 获得好的 IOU,并且 IOU 是与 box 的尺寸无关的。所以本文使用的新的距离公式为

$$d(box, centroid) = 1 - IOU(box, centroid) \quad (1)$$

式中:centroid 表示簇的中心;box 表示样本;IOU(box, centroid) 表示簇中心框和聚类框的交并比。交并比 IOU 表示预测框的准确程度,其算式为

$$IOU(b_{gt}, b_{dt}) = \frac{b_{gt} \cap b_{dt}}{b_{gt} \cup b_{dt}} \quad (2)$$

式中:b_{gt}表示真实框;b_{dt}表示预测框。

本文选取 K = 1, ..., 9, 分别对训练集进行 K-means 聚类分析,得到的结果 K 的大小与距离之间的关系如图 2 所示。

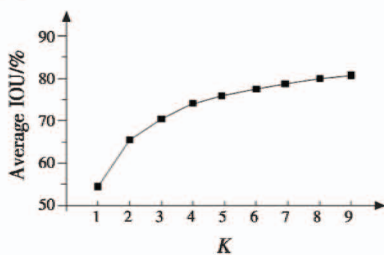


图 2 K-means 聚类结果图

Fig. 2 K-means clustering result

根据坐标图中的折线图得知,在 K = 4 之后,曲线逐渐开始平缓,所以选择 anchor boxes 的数量为 4,对应的预测框的大小设置为 4 个聚类的中心,在本训练集上分别是 (29, 56), (14, 20), (22, 36), (43, 82)。

根据文献[7],YOLO V3 网络在 COCO 和 VOC 数据集上都采用了 3 个尺度分别进行预测,具体方式如图 3 所示。

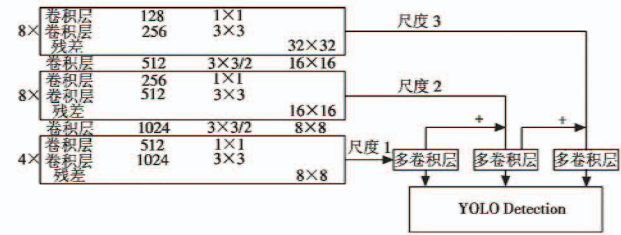


图 3 YOLO V3 网络多尺度预测方式

Fig. 3 YOLO V3 multiscale prediction method

由图 3 可以看到,YOLO V3 网络结构中对 32 × 32, 16 × 16, 8 × 8 大小的分辨率分别进行了预测,其中小的分辨率对应大的尺度。在 COCO 数据集中,8 × 8 大小的输出对应的 anchor boxes 的 3 个尺度分别是 (116, 90), (156, 198), (373, 326)。根据本文聚类的结果,在飞机的训练集中,最大的尺度大小为 (43, 82)。如果使用 8 × 8 分辨率进行预测,那么会导致虚警的现象,如图 4 所示。所以本文方法在 8 × 8 的分辨率上不进行预测,可删除图 3 中尺度 1 的预测层来降低虚警的现象。



图 4 YOLO V3 识别虚警现象

Fig. 4 False alarm in YOLO V3 recognizing process

2.2 Dense-YOLO 网络设计

在采集遥感图像过程中发现,由于受光照、天气、相机质量等影响,拍摄的遥感图像的质量参差不齐,有些过度曝光到人眼都难以识别。为了使网络在这些图像中依旧有好的准确率,本文借鉴 Densenet 网络的思想,将 YOLO V3 网络中对尺度 2 与尺度 3 进行预测输入的两组残差神经网络模块替换为自定义的密集神经网络模块。构建一个带有紧密连接的 Dense-YOLO 网络,使尺度 2 与尺度 3 在进行预测之前能够接收密集连接块输出的多层卷积特征,实现特征的复用与融合, Dense-YOLO 网络具体结构如图 5 所示。

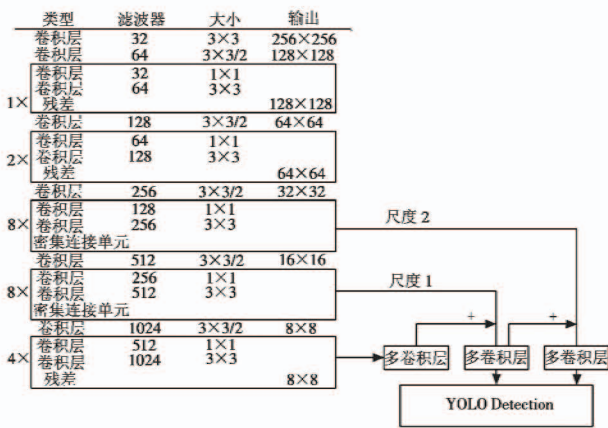


图5 Dense-YOLO网络结构

Fig. 5 Dense-YOLO network structure

这两个密集网络块具体的结构如图6所示,若设 x_0 为模块输入的特征图, x_n 表示第 n 块的输出, $[x_0, x_1, \dots, x_{n-1}]$ 表示对 x_0, x_1, \dots, x_{n-1} 的拼接。 $H(\cdot)$ 表示 BN-ReLU-Conv(1, 1) 与 BN-ReLU-Conv(3, 3) 的连接, 密集模块的算式为

$$x_n = H_n([x_0, x_1, \dots, x_{n-1}]) \quad n = 1, \dots, 8. \quad (3)$$

例如在尺度1的第3层组合中,由特征图 x_0 、第1层输出的特征图 x_1 和第2层输出的特征图 x_2 拼接成1024个通道特征图,以此类推,可以得出经过更改后将会有2304个特征图输入尺度1的预测网络,有4608个特征图输入到尺度2的网络。

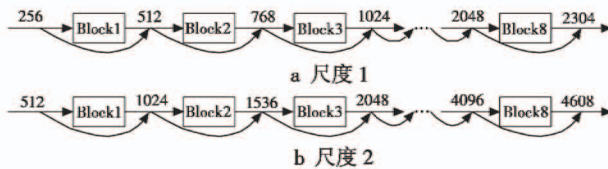


图6 密集网络块结构图

Fig. 6 Dense network block

3 实验对比

本文实验的软硬件平台配置如下。Intel(R) Core(TM) i7-8700 CPU@3.20 GHz * 12; GPU: NVIDIA GeForce GTX 1070Ti; 操作系统: ubuntu 16.04 LTS; 深度学习框架: darknet。

使用 Google Earth 软件截取了1000幅世界不同区域的遥感图像,包括过度曝光、云雾遮挡等质量不同的遥感图像,其中包含飞机7428架;随机抽取850幅图像作为训练集,其中包含6412架飞机;剩余150幅图像1016架飞机作为测试集。

3.1 网络的训练

对 YOLO V3 网络和 Dense-YOLO 网络分别进行训练,训练阶段采用动量项为 0.9 的异步随机梯度下降,每一个 batch 包含 32 张图像,权值的初始学习率为

0.001,衰减的系数设置为 0.0005。由于不同卫星拍摄出的遥感图像质量不同,飞机的角度不同、高度不同等因素,在训练时采用旋转训练集的图像,增加图像的对比度和曝光度,每迭代训练 1 个 batch 就进行 1 次尺度变换等策略来实现对数据集的扩充。

本文对 YOLO V3 网络和 Dense-YOLO 网络分别训练 15 000 次,经过实验测试,如果训练网络的迭代次数超过一定数量会出现过拟合的现象,为方便以后测试其最佳权重文件,训练时每 1000 次迭代保存一次网络的权重文件。

选择测试集的前 50 张图像,使用两种网络分别对其所有的权重进行测试,计算召回率为

$$R = T_p / (T_p + F_N) \quad (4)$$

式中: T_p 表示正确检测出来的飞机架数; F_N 表示没有被检测出来的飞机架数。部分测试结果如表 1 所示。

表1 召回率与迭代次数关系对比

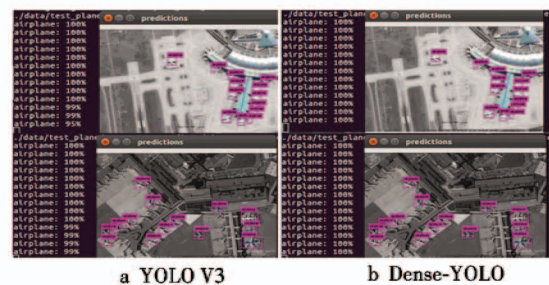
Table 1 Relationship between recalling rate and iteration times

迭代次数	YOLO V3/%	Dense-YOLO/%
8000	85.15	94.94
9000	86.58	95.12
10 000	90.33	94.53
11 000	87.71	93.81

从表1中可以看出,YOLO V3 网络的最佳迭代次数在 10 000 左右, Dense-YOLO 网络的最佳迭代次数在 9000 左右,并且 Dense-YOLO 网络在飞机识别上的优势已经体现出来了。下面就分别选用这两种网络的最佳迭代次数对两种网络进行对比与测试。

3.2 应用于高质量遥感图像

本文对 50 幅质量较高的遥感图像进行测试,发现对于质量较高的遥感图像,飞机识别的准确率很高,其检测效果与检测精度如图 7 所示。这两幅图像中,两种网络均能准确地识别出图像中的所有飞机,且精度很高。本文新改进的 Dense-YOLO 网络的精度比 YOLO V3 的原网络要高一些。



a YOLO V3

b Dense-YOLO

图7 高质量遥感图像检测结果

Fig. 7 Detection results on high-quality remote sensing image

这 50 幅图像共包含 362 架飞机,将两种网络在数据集上进行测试,分别计算准确率和召回率,准确率为

$$P = T_p / (T_p + F_p) \quad (5)$$

式中: T_p 表示正确检测出来的飞机架数; F_p 表示被误检的飞机架数。

经过计算得出 YOLO V3 网络误检架数为 4, 准确率为 98.87%, 漏检架数为 13, 召回率为 96.41%; Dense-YOLO 网络误检架数为 1, 准确率为 99.72%, 漏检架数为 6, 召回率为 98.34%。

可以看出, 两种方法在高质量的遥感图像上都有很好的效果, 本文改进的 Dense-YOLO 网络效果更佳, 其中准确率提升了 0.85%, 召回率提升了 1.94%。

3.3 应用于低质量遥感图像

本文对测试集中 50 幅质量较低的遥感图像进行测试, 发现本文改进的 Dense-YOLO 网络的优势体现得更加明显, 如图 8 所示, 在过度曝光的情况下飞机的信息损失严重, YOLO V3 原网络的漏检情况也很严重, 并且会出现虚警的现象, 而本文改进的 Dense-YOLO 网络有效地改善了这些现象。

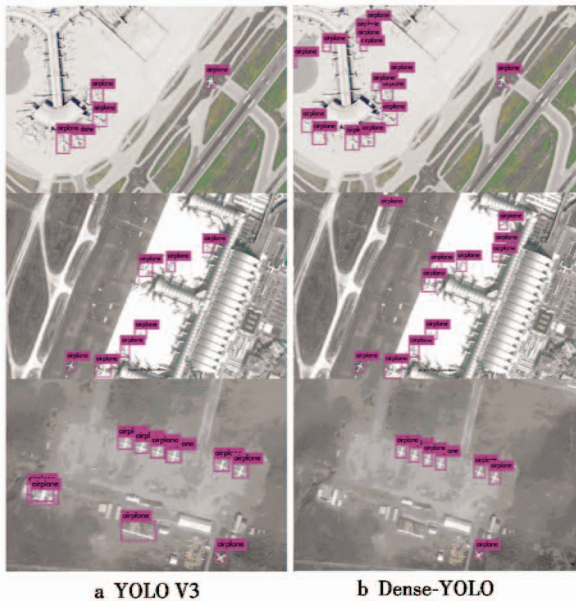


图 8 低质量遥感图像检测结果

Fig. 8 Detection results on low-quality remote sensing image

这 50 幅图像中共有 319 架飞机, 使用这两种网络分别在数据集上测试, 分别计算准确率和召回率, 得出 YOLO V3 网络误检架数为 23, 准确率为 91.05%, 漏检架数为 85, 召回率为 73.35%; Dense-YOLO 网络误检架数为 12, 准确率为 96.12%, 漏检架数为 22, 召回率为 93.10%。

可以看出, 在质量较低的遥感图像中, 本文改进的 Dense-YOLO 网络相比于 YOLO V3 网络, 准确率提升了 5.07%, 召回率提升了 19.75%, 效果显著。

4 总结

本文以识别遥感图像中的飞机目标为目的, 基于 YOLO V3 深度卷积神经网络, 对数据集进行了聚类分析, 根据自己数据集中的尺度对网络进行改进, 并参考密集神经网络的特点, 改进为一种 Dense-YOLO 网络, 分别对 YOLO V3 网络与 Dense-YOLO 网络进行训练与测试。经实验表明, 本文改进网络的效果提升显著: 对高质量的遥感图像有识别率高并且漏检率低的特点; 在质量较低的遥感图像下效果提升更加显著。

参考文献

- [1] GIRSHICK R. Fast R-CNN [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1440-1448.
- [2] HE K M, GKIOXARI G, DOLLAR P, et al. Mask R-CNN [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2018: 1-12.
- [3] LIU W, ANQUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2016: 21-37.
- [4] 周飞燕, 金林鹏, 董军. 卷积神经网络研究综述[J]. 计算机学报, 2017, 40(6): 1229-1251.
- [5] 唐小佩, 杨小冈, 刘云峰, 等. 基于深度卷积神经网络的飞机识别研究[J]. 电光与控制, 2018, 25(5): 68-72.
- [6] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real time object detection [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [7] REDMON J, FARHADI A. YOLO V3: an incremental improvement [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2018: 1-6.
- [8] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2017: 6517-6525.
- [9] DAI J F, LI Y, HE K M, et al. R-FCN: object detection via region-based fully convolutional networks [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2016: 379-387.
- [10] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2016: 4700-4708.
- [11] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.