行人属性识别:基于元学习的概率集成方法

王文龙[1],张　磊[1],张誉馨[1],吴晓富[1],张索非[2]
(1. 南京邮电大学 通信与信息工程学院,江苏 南京 210003;
2. 南京邮电大学 物联网学院,江苏 南京 210003)

摘　要:行人属性识别(pedestrian attribute recognition,PAR)的目的是从输入图像中挖掘行人的属性信息。近年来,卷积神经网络(convolution neural network,CNN)的兴起在行人属性识别中获得了广泛的应用。现有的方法多采用属性不可知的视觉注意或启发式的身体部位定位机制来增强局部特征表达,而忽略了多模型集成所能够带来的提升,因此该领域内鲜少有集成算法的提出。为了进一步提高行人属性识别的性能,该文从 CNN 模型的预测概率角度入手,基于元学习提出了一种行人属性识别的概率集成算法(probabilistic ensemble learning method,PEM)。在行人属性识别数据集 RAP 上的实验结果表明,该算法随着模型有效数的增加表现出递增的平均准确度(mean accuracy)和 F1 值(F1-score),且评测结果均优于多个典型的行人属性识别算法。

关键词:卷积神经网络;行人属性识别;元学习;概率集成;行人检索
中图分类号:TP391　　**文献标识码**:A　　**文章编号**:1673-629X(2022)03-0071-05
doi:10.3969/j.issn.1673-629X.2022.03.012

Pedestrian Attribute Recognition: Probabilistic Ensemble Learning Method Based on Meta-learning

WANG Wen-long[1], ZHANG Lei[1], ZHANG Yu-xin[1], WU Xiao-fu[1], ZHANG Suo-fei[2]
(1. School of Communication and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China;
2. School of Internet of Things, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: Pedestrian attribute recognition (PAR) aims to mine pedestrian attribute information from input images. In recent years, convolution neural network (CNN) has been widely used in PAR. Existing methods mostly use visual attention with unknown attributes or heuristic body part localization mechanism to enhance local feature expression, but ignore the improvement brought by multi-model integration. Therefore, there are few integration algorithms proposed in this field. In order to further improve the performance of PAR, we propose a probabilistic ensemble learning method based on meta-learning from the perspective of CNN model's prediction probability. The experimental results on RAP dataset of pedestrian attribute recognition show that the proposed algorithm achieves improved mean accuracy (mA) and F1 score with the increase of models, which are better than several existing pedestrian attribute recognition algorithms.

Key words: convolution neural network; pedestrian attribute recognition; meta-learning; probabilistic ensemble; pedestrian retrieval

0　引　言

行人属性识别(pedestrian attribute recognition,PAR)的目的是在行人图像中挖掘目标人物的属性信息。传统的行人属性识别方法通常侧重于从人工生成的特征、强大的分类器或属性关系的角度来获得更鲁棒的特征。在过去的几年里,深度学习在利用多层非线性变换进行自动特征提取方面取得了令人瞩目的成绩,特别是在计算机视觉、语音识别和自然语言处理等方面。基于这些突破,研究者提出了几种基于深度学习的属性识别算法,如 ACN[1]、DeepMAR[2]。虽然行人属性特征的提取由于卷积神经网络的应用得到大幅的提升,但行人属性识别的性能在实际应用中仍不能满足需求。该任务的困难在于 PAR 中不同类别的属性所属的粒度不同,如发型、颜色、帽子、眼睛等信息只是局部图像块的低层属性信息,而年龄、性别等信息却是全局的高层语义信息。并且,在视角、光线等信息变

收稿日期:2021-03-30　　**修回日期**:2021-07-30
基金项目:国家自然科学基金(61701252)
作者简介:王文龙(1997-),男,硕士,研究方向为行人识别;吴晓富,博士,教授,研究方向为信息论与编码、机器学习与计算机视觉。

化时,采样到的图像变化可能很大,但这些属性信息却不会改变。最近不少研究者试图利用属性间的空间关系和语义关系来进一步提高识别性能,所提出的方法可分为三个基本类别:

基于语义关系的方法:利用语义关系来辅助属性识别。Wang 等人[3]利用属性间的依赖性和相关性提出了一个基于 CNN-RNN 的框架。Sudowe[1]提出一个整体的 CNN 模型来共同学习不同的属性。然而,这些方法需要人工定义规则,如预测顺序、属性组等,在实际应用中很难确定。

基于注意力机制的方法:利用视觉注意机制来提高属性识别。Liu 等人[4]提出了一种多方向注意模型,用于行人多尺度注意特征的学习分析。虽然识别精度有所提高,但这些方法都是属性不可知的,没有考虑到属性的具体信息。

基于局部特征提取的方法:利用行人的身体部位特征来提高属性识别。Zhu 等人[5]将整个图像分割成 15 个 Rigid 块,融合不同块的特征。Yang 等人[6]利用外部姿态估计模块定位身体部位。

以上这些方法从语义关系、特征提取、注意力机制等角度来提升模型的性能。该文主要从集成学习的角度来提高行人属性识别性能。不同于该领域中广泛应用的多模型集成方案,该文从单个模型的属性预测输出概率入手,利用少量数据训练元学习器的集成规则,最终通过概率集成的方式获得性能上的提升。通过对比最新的行人属性识别算法,如 RDEM[7]、VAC[8]、ALM[9],PEM 算法均能够获得可观的性能提升。

1 文中算法

1.1 算法思想

整体的算法流程如图 1 所示。该算法需要有监督数据对元学习器进行训练,因此将训练数据分割为训练集和验证集。训练集用于基本分类器的训练,验证集用于元学习器的训练。实验结果表明数据集的较佳分割比例为 4∶1(见 2.4)。

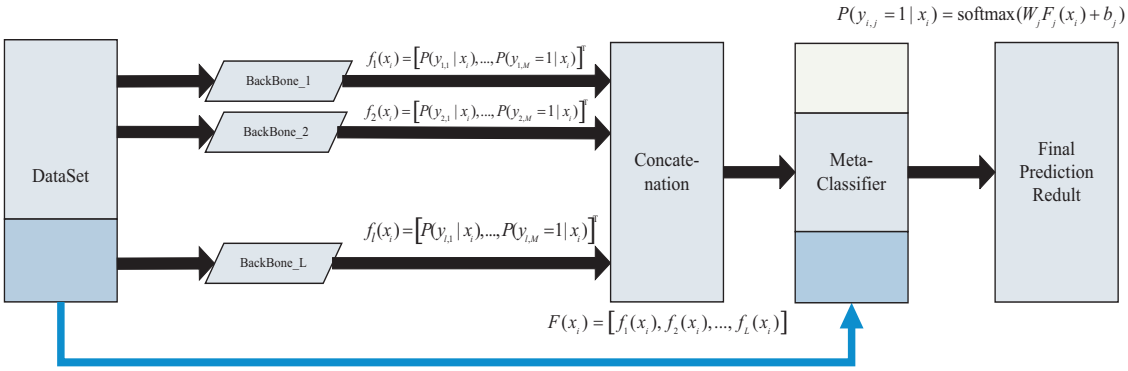

图 1 算法流程

PEM 算法首先通过 RDEM 算法[7]的框架,用不同的模型结构得到 L 个基本分类器,然后将各个分类器的输出概率进行拼接。$F(x_i)$ 是基本分类器预测概率拼接的结果,它是一个 $L\times M$ 的向量,M 代表属性的个数。元学习器会对 $F(x_i)$ 中各个分类器的输出概率赋予不同的权重和偏置,经过优化器的调优后得到最佳的参数,最后将该参数用于加权集成得到最终的预测概率。

1.2 分类器叠加

给定一个数据集:

$$D=\{(x_i,y_i),y_i\in\{0,1\}^M,i=1,2,\cdots,N\} \quad (1)$$

其中,y_i 表示第 i 张图片的正确的标签,而 N,M 分别表示训练的图像个数和属性个数,x_i 表示第 i 个行人图片。行人属性识别是一个多标签识别任务,对给定的行人图片 x_i 要求判断属性 y_i 是否存在。在标签向量 y 和预测向量 y_i 中,对应位置的 0 或 1 代表了该属性是否出现在行人图片当中。

在对属性进行预测时,将每个属性都看作一个二分类的问题,那么其二元交叉熵损失函数可以用式(5)来表示,其中 $w_j(y_i)=e^{y_{i,j}(1-r_j)+(1-y_{i,j})r_j}$ 是样本的权重,用来缓解样本不平衡分布带来的问题[10-11]。r_j 表示第 j 个属性的正样本率。

给定 L 个基本分类器,其中对 M 个属性进行预测的第 l 个分类器的预测概率表达式如下:

$$f_l(x_i)=[P(y_{i,1}\mid x_i),\cdots,P(y_{i,M}=1\mid x_i)]^{\mathrm{T}}=[y_{i,1}^l,\cdots,y_{i,M}^l]^{\mathrm{T}} \quad (2)$$

通过将每个属性分类器的 Softmax 层的输出堆叠为长度为 M 的列向量,最终的叠加概率特征可以表达为如下形式:

$$F(x_i)=[f_1(x_i),f_2(x_i),\cdots,f_L(x_i)] \quad (3)$$

1.3 元分类器

将 $f_l(x_i),l=1,2,\cdots,L$ 作为最终的概率特征向量,由元学习器(或集成学习器)将不同的预测结果进行拟合,得到最终的预测。该文提出的元分类器采用

最简单的单层全连接网络,使用 $F(x_i)$ 作为输入,通过一个全连接层、Softmax 层输出最终预测,也即:

$$P(y_{i,j}=1 \mid x_i)=\text{softmax}(W_jF_j(x_i)+b_j)$$
$$j=1,2,\cdots,M \tag{4}$$

其中, W_j 和 b_j 是权重和偏置。以上参数的获得需要通过一些有标注的训练数据进行训练获得,因此需要对训练集进行分割。对于训练集的分割,应当将其分割成为两部分:一部分用于模型各自分类器的训练,另一部分用于元分类器的训练。

2　实　验

2.1　数据集

实验所使用的数据集来自于 PRCV2020 大型行人检索竞赛数据集 RAP[12] 的子集,RAP 数据来源是从一个室内购物中心的高清晰度(1 280×720)监控网络中采集的。RAP 数据集涵盖了 2 589 个行人身份标签。该数据集中关于行人属性识别的数据共由 68 071 张图片组成,其中 42 085 张用于训练(训练集+验证集),25 986 张用于测试。对于每个样本都标注了 72 个细粒度属性,例如性别发型、鞋子类型以及附属物类型等,其中 54 个属性用于本次竞赛。54 个属性标签大致可分为七大类,分别为:人物属性、头部、上衣、下衣、鞋子、附属物、行为,具体如下:

$$L_{\text{BCE}}=-\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}\{w_j(y_i)\{y_{i,j}\log\hat{y}_{i,j}+(1-y_{i,j})\log(1-\hat{y}_{i,j})\}\}=-\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}w_j(y_i)<y_{i,j},\hat{y}_{i,j}> \tag{5}$$

· 人物属性:性别、年龄 16、年龄 30、年龄 45、年龄 60、体型微胖、体型标准、体型偏瘦、顾客、店员。

· 头部:光头、长发、黑色头发、戴帽、眼镜。

· 上衣:衬衣、毛衣、马甲、T 恤、棉服、夹克、西服、卫衣、短袖、其他。

· 下衣:长裤、裙子、短裙、连衣裙、牛仔裤、包腿裤。

· 鞋子:皮鞋、运动鞋、靴子、布鞋、休闲鞋、其他。

· 附属物:双肩包、单肩包、手提包、箱子、塑料袋、纸袋、购物车、其他。

· 行为:打手机、交谈、聚集、抱东西、推东西、拉拽东西、夹带东西、拎东西、其他。

据统计,人的属性数量在 4 到 26 之间,平均值为 12.2。此外,大多数二元属性类别中,RAP 具有严重的不平衡分布。如表 1 所示,用正样本占整个数据集的比例来衡量,正样本率较少(小于 10%)的属性数量占到了总数量的 60%,这使得开发高质量的识别算法成为一个具有挑战性的问题。

表 1　正样本率分布

正样本率	属性个数
(0,0.1)	32
[0.1,0.2)	8
[0.2,0.3)	5
[0.3,1)	9

2.2　评测指标

对于行人属性识别任务,将每个属性看作一个二分类问题,并将识别结果分为以下四种情况:TP(True Positive),识别正确的正样本;FP(False Positive),识别错误的负样本;TN(True Negative),识别正确的负样本;FN(False Negative),识别错误的正样本。

评价属性识别算法的第一个性能指标是平均准确度[13](mA)。考虑到属性分布的不均衡性,对于每个属性,mA 分别计算正样本和负样本的分类精度,然后取它们的平均值作为属性的结果。之后,mA 取所有属性的平均值作为最终结果。计算方式如下:

$$\text{mA}=\frac{1}{2N}\sum_{i=1}^{L}\left(\frac{\text{TP}_i}{P_i}+\frac{TN_i}{N_i}\right) \tag{6}$$

其中, L 表示第 L 个属性; P_i 和 N_i 分别为当前属性总的正、负样本数。

评价属性识别算法的第二个性能指标是 F1 值[14](F1-score),它是精准率和召回率的调和平均数,公式如下:

$$\text{F1}=2*\frac{\text{Prec}*\text{Rec}}{\text{Prec}+\text{Rec}} \tag{7}$$

式中,Prec(Precision)是准确率,可以反映模型仅返回相关图片的能力,Rec(Recall)是召回率,可以反映模型识别所有相关图片的能力。其计算方式为:

$$\text{Prec}=\frac{\text{TP}}{\text{TP}+\text{FP}} \tag{8}$$

$$\text{Rec}=\frac{\text{TP}}{\text{TP}+\text{FN}} \tag{9}$$

从公式的计算方式可以看出,准确率和召回率会相互影响。一般情况下当准确率越高时,召回率就越低,反之亦然。

2.3　实验结果

本实验中,待集成的模型分类器的获取来自于 RDEM 框架[7]。该方法用 PyTorch 实现,并进行了端到端的训练。该文采用多个网络(例如 ResNet、OSNet 等)作为主干提取行人图像特征。将行人图像调整为 256×192,并对其采用随机水平镜像的处理。训练优化器采用随机梯度下降(stochastic gradient descent, SGD),动量为 0.9,权重衰减为 0.000 5。初始学习率等于 0.01,批次大小设置为 64。训练的总迭代次数为 30 次。

实验主要集中于 PRCV2020 的竞赛数据集 RAP 上,将其中的 8 417 张图片固定为测试集。对于该方法,将剩余的 33 668 张图片按比例拆分成训练集和验证集,拆分比例为 4∶1(见 2.4)。训练集用于训练多个模型的基本分类器,验证集用于训练元分类器。由于对比算法不包含元分类器,且输出概率为单模型的输出,因此将分割后的训练集和验证集全部用于训练其所用模型的分类器。最终的实验结果如表 2 所示。

表 2 实验结果

所用模型	mA	F1-score
RDEM[7](arXiv20)	0.790 8	0.793 9
ALM[9](ICCV19)	0.783 2	0.779 1
VAC[8](CVPR19)	0.773 6	0.804 8
PEM(7 模型)	0.801 8	0.814 0
PEM(10 模型)	0.809 6	0.815 6
PEM(14 模型)	0.811 8	0.816 3

实验选取了一些主流的网络结构,包括:ResNet[15]、OSNet[16]、EfficientNet[17]、DenseNet[18]、Inception[19]、Xception[20]等(包括其改进结构),以上模型的部分改进结构命名规则如下:SE(Squeeze-and-Excitation Networks)表示该网络采用了 SE 模块[21],该模块能够调整通道间的权重,并将重要特征强化,以取得更好的效果;ibn 表示该网络采用了 Instance-batch Normalization(IBN)模块[22],该模块将 Instance Normalization(IN)和 Batch Normalization(BN)组合使用,提高了模型的范化能力。它有两种结构:ibn-a 和 ibn-b,区别在于 IN 的位置有所不同。对于 EfficientNet 而言,其网络配置参数为 b0-b8(该文采用 b0 和 b4 的配置),不同配置的网络,其宽度、深度、Dropout 参数均不相同。对于 ResNeXt 网络[23],它是 ResNet 网络的升级,使用一种平行堆叠相同拓扑结构的 blocks 代替原来 ResNet 的三层卷积的 block,能够在不明显增加参数量级的情况下提升模型的准确率。PEM 算法各个配置所使用的模型如下:

· PEM(7 模型):由 ResNet50,SE-ResNet101-ibn-a,ResNet101-ibn-a,OSNet-x1.0,OSNet-ibn-x1.0,EfficientNet-b0,DenseNet161 等模型组成。

· PEM(10 模型):在 7 模型的基础上新增了以下模型:DenseNet201, EfficientNet_b4, Inceptionv4。

· PEM(14 模型):在 10 模型的基础上新增了以下模型:ResNet152, SE-ResNet50, SE-ResNeXt50-32x4d,Xception。

从表 2 中可以看出,通过增加基本分类器的个数,其性能也在不断的上升。对比 RDEM 算法,14 个模型的堆叠使得其在 mA 评测中上升了 1.17%,在 F1 值评测中上升了 2.24%。对比 ALM 算法和 VAC 算法,该方法在 mA 上各自提升 2.86% 和 3.82%,在 F1 值上各自提升了 3.72% 和 1.15%。

2.4 对比实验

2.4.1 数据集分割对比实验

该算法需要对数据集进行分割,其分割比例不可避免地会对训练过程产生影响。为了探究不同分割比例对于模型最终性能的影响,采取了三种不同的分割比例(训练参数与 PEM 算法的实验参数保持一致),并在 7 模型集成结果中进行了展示。表 3 给出了三次实验的分割比例和数据集总图片数的对应关系,测试图片为固定的 8 417 张图片。

表 3 数据集的分割与实验结果

分割比例	具体数量	测试集	mA	F1-score
5∶1	28 024∶5 644	8 417	0.801 2	0.812 9
4∶1	26 935∶6 733	8 417	0.801 8	0.814 0
3∶1	25 251∶8 417	8 417	0.798 7	0.811 6

从表 3 的集成结果上来看,数据集的分割比例保持在 4∶1 是一个较优的分割选择,在此分割比例下,模型在最终的两项评测中均优于其他分割比例。

2.4.2 集成方式对比实验

在行人属性识别任务中,目前尚未有集成方法的提出。PEM 算法作为基于元学习的概率加权集成算法,其实质是一个非等概加权的过程,另一种容易想到的加权方式为等概加权集成,为了探究两种方式在集成性能之间的不同,对两种方式分别进行了实验。

在等概加权集成(equal weight ensemble,EWE)实验中,将集成方式更改为等概加权方式。实验中,PEM 算法与 EWE 算法所使用的基本分类器相同,在得到最终预测结果时,PEM 算法将多个基本分类器预测结果进行非等概加权,加权参数则是利用部分训练集训练得到;EWE 算法则直接将基本分类器预测结果相加后平均。注意,为了保证公平对比,等概加权集成实验的训练数据为分割后的训练集加验证集。最终其实验结果如表 4 所示。

表 4 集成方式对比结果

集成方式	mA	F1-score	Pos_Rec
PEM(7 模型)	0.801 8	0.814 0	0.661 7
EWE(7 模型)	0.790 6	0.814 8	0.637 7
PEM(14 模型)	0.811 8	0.816 3	0.680 3
EWE(14 模型)	0.790 1	0.817 5	0.635 5

通过表 4 的结果可以看出,EWE 方法在 mA 评测中大幅落后于 PEM 算法,在 F1 值的评测领先 PEM 算法。为了探究该现象产生的原因,又对正样本召回率(Pos_Rec)进行了测试,结果表明 PEM 算法具有更好

的正样本识别能力。

从表 4 中可知,在 14 模型的集成实验中,EWE 算法在正样本召回率评测中落后 PEM 算法约 6.59%,根据 mA 的计算公式不难得出如下结论:EWE 算法相对于 PEM 算法,无法更加有效地辨别正样本,从而导致 mA 评测始终无法提升。由于 RAP 数据集具有严重的不平衡分布(正样本率过低),即使出现模型预测全部为负样本的情况,其总体准确率依旧很高,而正样本识别能力的增强虽然会提高召回率(Rec),但由于正样本数量过少,其提升对 F1 值评测贡献有限,进而使得 EWE 算法在 F1 值评测上略高于 PEM 算法。

综上所述,PEM 算法在保证 F1 值评测的情况下,大幅提升了对正样本的识别能力,克服了 EWE 算法无法提升 mA 评测的困难。

3　结束语

该文提出了一种行人属性识别的概率集成算法(PEM),通过元学习器对多个模型的输出概率进行拟合,最终在行人属性识别数据集 RAP 上获得了更好的行人属性预测结果。在此基础上,该文进一步对数据集分割比例、模型数量对算法性能的影响进行了研究,实验结果表明:对于 RAP 数据集而言,数据集分割比例在 4 ∶ 1 时较为合适;PEM 算法随着集成模型的有效个数增多性能得到稳步提升,表现出优异的模型集成能力。最终该算法在 PRCV2020 大规模行人检索竞赛行人属性识别任务中获得第三名(具体见网址:https://lsprc.github.io/leaderboard.html)。

参考文献:

[1] SUDOWE P,SPITZER H,LEIBE B. Person attribute recognition with a jointly-trained holistic CNN model[C]//2015 IEEE international conference on computer vision workshop. Santiago,Chile:IEEE,2015:329-337.

[2] LI D,CHEN X,HUANG K. Multi-attribute learning for pedestrian attribute recognition in surveillance scenarios[C]//2015 3rd IAPR Asian conference on pattern recognition (ACPR). Kuala Lumpur,Malaysia:IEEE,2015:111-115.

[3] WANG X,ZHENG S,YANG R,et al. Pedestrian attribute recognition:a survey[J]. arXiv:1901.07474,2019.

[4] LIU X,ZHAO H,TIAN M,et al. Hydraplus-net:attentive deep features for pedestrian analysis[C]//2017 IEEE international conference on computer vision. Venice,Italy:IEEE,2017:350-359.

[5] ZHU J,LIAO S,YI D,et al. Multi-label CNN based pedestrian attribute learning for soft biometrics[C]//2015 international conference on biometrics (ICB). Phuket,Thailand:IEEE,2015:535-540.

[6] YANG L,ZHU L,WEI Y,et al. Attribute recognition from adaptive parts[J]. arXiv:1607.01437,2016.

[7] JIA J,HUANG H,YANG W,et al. Rethinking of pedestrian attribute recognition:realistic datasets with efficient method[J]. arXiv:2005.11909v1,2020.

[8] GUO H,ZHENG K,FAN X,et al. Visual attention consistency under image transforms for multi-label image classification[C]//2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR). Long Beach,CA:IEEE,2019:729-739.

[9] TANG C,SHENG L,ZHANG Z,et al. Improving pedestrian attribute recognition with weakly-supervised multi-scale attribute-specific localization[C]//IEEE international conference on computer vision. Seoul,South Korea:IEEE,2019:4997-5006.

[10] CAO K,WEI C,GAIDON A,et al. Learning imbalanced datasets with label-distribution-aware margin loss[C]//Advances in neural information processing systems. Vancouver,Canada:NeurIPS,2019:1567-1578.

[11] BUDA M,MAKI A,MAZUROWSKI M A. A systematic study of the class imbalance problem in convolutional neural networks[J]. Neural Networks,2018,106:249-259.

[12] LI D,ZHANG Z,CHEN X,et al. A richly annotated pedestrian dataset for person retrieval in real surveillance scenarios[J]. IEEE Transactions on Image Processing,2018,28(4):1575-1590.

[13] DENG Y,LUO P,LOY C C,et al. Pedestrian attribute recognition at far distance[C]//22nd ACM international conference on multimedia. Florida,Orlando,FL:ACM,2014:789-792.

[14] ZHANG M L,ZHOU Z H. A review on multi-label learning algorithms[J]. IEEE Transactions on Knowledge and Data Engineering,2013,26(8):1819-1837.

[15] HE K,ZHANG X,REN S,et al. Deep residual learning for image recognition[C]//2016 IEEE conference on computer vision and pattern recognition. Seattle,WA:IEEE,2016:770-778.

[16] ZHOU K,YANG Y,CAVALLARO A,et al. Omni-scale feature learning for person re-identification[C]//IEEE international conference on computer vision. Seoul,South Korea:IEEE,2019:3701-3711.

[17] TAN M,LE Q. Efficientnet:rethinking model scaling for convolutional neural networks[C]//International conference on machine learning. Da Lat,Vietnam:PMLR,2019:6105-6114.

[18] HUANG G,LIU Z,VAN DER MAATEN L,et al. Densely connected convolutional networks[C]//2017 IEEE conference on computer vision and pattern recognition. Honolulu,HI:IEEE,2017:2261-2269.

[19] SZEGEDY C,LIU W,JIA Y,et al. Going deeper with convolutions[C]//2015 IEEE conference on computer vision

(下转第 83 页)