

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2020 年 9 月 10 日 (10.09.2020)



(10) 国际公布号
WO 2020/177217 A1

- (51) 国际专利分类号:
G06K 9/00 (2006.01) **G06N 3/04** (2006.01)
G06K 9/34 (2006.01)
- (21) 国际申请号: PCT/CN2019/087164
- (22) 国际申请日: 2019 年 5 月 16 日 (16.05.2019)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201910161808.0 2019年3月4日 (04.03.2019) CN
- (71) 申请人: 东南大学 (SOUTHEAST UNIVERSITY)
[CN/CN]; 中国江苏省南京市玄武区四牌楼2号, Jiangsu 210096 (CN)。
- (72) 发明人: 李旭 (LI, Xu); 中国江苏省南京市玄武区四牌楼2号, Jiangsu 210096 (CN)。 郑智勇 (ZHENG, Zhiyong); 中国江苏省南京市玄武区四牌楼2号, Jiangsu 210096 (CN)。 韦坤

(WEI, Kun); 中国江苏省南京市玄武区四牌楼2号, Jiangsu 210096 (CN)。

(74) 代理人: 南京众联专利代理有限公司 (NANJING ZHONGLIAN PATENT AGENCY CO., LTD.); 中国江苏省南京市建邺区福园街129号万达商务楼7层, Jiangsu 210017 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ,

(54) **Title:** METHOD OF SEGMENTING PEDESTRIANS IN ROADSIDE IMAGE BY USING CONVOLUTIONAL NETWORK FUSING FEATURES AT DIFFERENT SCALES

(54) 发明名称: 基于变尺度多特征融合卷积网络的路侧图像行人分割方法

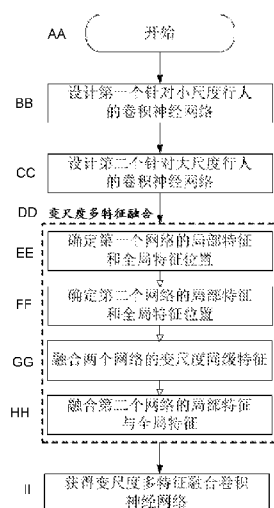


图 1

(57) **Abstract:** A method of segmenting pedestrians in roadside image by using a convolutional network fusing features at different scales. To resolve an issue in which pedestrians in an image captured by a roadside smart terminal appear at markedly different scales, the method employs two parallel convolutional neural networks to extract local features and global features of pedestrians at various scales. The local features and global features extracted by the first network and the local features and global features extracted by the second network are fused on a same-level basis, and then secondary fusion is performed on the fused local and global features to obtain a convolutional neural network implementing fusion of features at multiple scales. The network undergoes training, and a roadside pedestrian image is input thereto to realize pedestrian segmentation. The above method effectively solves the problems of unclear and incomplete segmentation which occur easily in the majority of existing pedestrian segmentation methods based on a single network structure, thereby enhancing the accuracy and robustness of pedestrian segmentation.

- AA Start
BB Design a first convolutional neural network for pedestrians at a small scale
CC Design a second convolutional neural network for pedestrians at a large scale
DD Determine local features and global features extracted by the first network
EE Fuse the features at different scales
FF Determine local features and global features extracted by the second network
GG Fuse the features at different scales of the two networks on a same-level basis
HH Fuse the local features and the global features extracted by the second network
II Obtain a convolutional network implementing fusion of features at multiple scales



NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布：

— 包括国际检索报告(条约第21条(3))。

(57) 摘要：一种基于变尺度多特征融合卷积网络的路侧图像行人分割方法，该方法针对智能路侧终端图像中行人尺度变化显著的情况，设计了并行的两个卷积神经网络来提取不同尺度行人的局部特征及全局特征，进而将第一个网络提取的局部特征及全局特征与第二个网络提取的局部特征及全局特征进行同级融合，接着对融合后的局部特征与全局特征进行二次融合，得到变尺度多特征融合卷积神经网络，而后对该网络进行训练并输入路侧行人图像实现行人分割。上述方法有效地解决了当前大部分基于单一网络结构的行人分割方法容易出现分割边界模糊以及漏分割的问题，进一步提高了行人分割的准确性和鲁棒性。

基于变尺度多特征融合卷积网络的路侧图像行人分割方法

技术领域

本发明属于计算机视觉及智能路侧感知技术领域，涉及智能路侧终端图像行人分割方法，尤其涉及一种基于变尺度多特征融合卷积网络的路侧图像行人分割方法。

背景技术

随着汽车保有量的持续快速增长，我国道路交通压力凸显，交通事故尤其是汽车与行人碰撞事故的发生对行人的生命安全造成了严重威胁。为了保护行人安全，在制定严格交通法规的同时，提高车辆的智能化水平实现对行人准确地感知，进而辅助驾驶员安全行驶，可有效防止碰撞事故的发生。然而，当前大部分智能车辆仅利用自车的传感器进行行人感知，其视野有限并且存在视距盲区，对行人易受到障碍物遮挡的情况适应性不足，存在极大的安全隐患。在这种背景下，智能路侧行人感知技术的发展为行人安全保护提供了新的解决思路，借助路侧感知设备作用范围广的优势，将感知到的行人信息以通信的手段广播给附近车辆，提醒驾驶员谨慎驾驶以避让行人。

考虑到成本、可行性等影响因素，目前大部分行人感知方法是基于计算机视觉的，其中，基于视觉的行人分割作为行人感知的重要手段之一得到了广泛研究。大部分传统基于手工设计特征的行人分割方法只考虑了单个行人的情况，利用行人个体与背景的差异来得到分割结果。然而，由于路侧成像设备视野广阔，再加上透视投影的影响，

所获得的图像中行人尺度变化显著，即近距离的行人变“大”，而远距离的行人变得尤其“小”，此时传统的行人分割方法由于其有限的特征表达能力而难以获得满意的结果。

近年来，深度学习技术的不断发展为智能路侧终端图像行人分割任务提供了一种新的解决方案，与传统方法相比，深度学习的突出优势是具有强大的特征表达能力。基于深度神经网络的行人分割方法对复杂的交通场景有良好的适应性，能够获得更准确的分割性能。然而，目前利用深度神经网络进行行人分割的方法主要采用单一的网络结构，仅凭网络深度难以准确地提取智能路侧终端图像中大尺度行人的边界局部特征以及小尺度行人的全局特征，容易造成边界模糊甚至漏分割，限制了行人分割精度的进一步提升，无法取得令人满意的效果。

发明内容

为解决上述问题，本发明公开了一种基于变尺度多特征融合卷积网络的路侧图像行人分割方法，有效地解决了当前大部分基于单一网络结构的行人分割方法难以适用于变尺度行人的问题，进一步提高了行人分割的准确性和鲁棒性。

为了达到上述目的，本发明提供如下技术方案：

基于变尺度多特征融合卷积网络的路侧图像行人分割方法，包括以下步骤：

- (1) 建立行人分割数据集；
- (2) 构建变尺度多特征融合卷积神经网络架构，首先设计并行

的两个卷积神经网络来提取图像中不同尺度行人的局部特征和全局特征，第一个网络针对小尺度行人设计了精细的特征提取结构，第二个网络针对大尺度行人扩大了网络在浅层处的感受野；进而提出两级融合策略对所提取的特征进行融合，首先对不同尺度的同级特征进行融合，得到适用于变尺度行人的局部和全局特征，然后构建跳跃连接结构将融合后的局部特征与全局特征进行二次融合，获取完备的变尺度行人局部细节信息和全局信息，最终得到变尺度多特征融合卷积神经网络，包括以下子步骤：

子步骤 1：设计第一个针对小尺度行人的卷积神经网络，具体包括：

① 设计池化层，池化层数量为 2，均采用最大池化操作，采样尺寸均为 2×2 ，步长均为 2；

② 设计标准卷积层，标准卷积层数量为 18，其中 8 层卷积核大小均为 3×3 ，卷积核数量分别为 64、64、128、128、256、256、256、2，步长均为 1，剩下 10 层卷积核大小均为 1×1 ，卷积核数量分别为 32、32、64、64、128、128、128、128、128、128，步长均为 1；

③ 设计反卷积层，反卷积层数量为 2，卷积核大小均为 3×3 ，步长均为 2，卷积核数量分别为 2、2；

④ 确定网络架构，根据步骤（2）的子步骤 1 中①~③涉及的网络层参数建立不同的网络模型，然后利用步骤（1）所建立的数据集对这些模型进行验证，从中筛选出兼顾准确性和实时性的网络结构，得到最优网络架构如下：

标准卷积层 1_1: 用 64 个 3×3 的卷积核与 $A \times A$ 像素的输入样本做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为 $A \times A \times 64$ 的特征图;

标准卷积层 1_1_1: 用 32 个 1×1 的卷积核与标准卷积层 1_1 输出的特征图做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为 $A \times A \times 32$ 的特征图;

标准卷积层 1_1_2: 用 32 个 1×1 的卷积核与标准卷积层 1_1_1 输出的特征图做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为 $A \times A \times 32$ 的特征图;

标准卷积层 1_2: 用 64 个 3×3 的卷积核与标准卷积层 1_1_2 输出的特征图做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为 $A \times A \times 64$ 的特征图;

池化层 1: 用 2×2 的核为标准卷积层 1_2 输出的特征图做最大池化, 步长为 2, 得到维度为 $\frac{A}{2} \times \frac{A}{2} \times 64$ 的特征图;

标准卷积层 2_1: 用 128 个 3×3 的卷积核与池化层 1 输出的特征图做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为 $\frac{A}{2} \times \frac{A}{2} \times 128$ 的特征图;

标准卷积层 2_1_1: 用 64 个 1×1 的卷积核与标准卷积层 2_1 输出的特征图做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为 $\frac{A}{2} \times \frac{A}{2} \times 64$ 的特征图;

标准卷积层 2_1_2: 用 64 个 1×1 的卷积核与标准卷积层 2_1_1 输出的特征图做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为 $\frac{A}{2} \times \frac{A}{2} \times 64$ 的特征图;

标准卷积层 2_2: 用 128 个 3×3 的卷积核与标准卷积层 2_1_2 输出的特征图做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为 $\frac{A}{2} \times \frac{A}{2} \times 128$ 的特征图;

池化层 2: 用 2×2 的核对标准卷积层 2_2 输出的特征图做最大池化, 步长为 2, 得到维度为 $\frac{A}{4} \times \frac{A}{4} \times 128$ 的特征图;

标准卷积层 3_1: 用 256 个 3×3 的卷积核与池化层 2 输出的特征图做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为 $\frac{A}{4} \times \frac{A}{4} \times 256$ 的特征图;

标准卷积层 3_1_1: 用 128 个 1×1 的卷积核与标准卷积层 3_1 输出的特征图做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为 $\frac{A}{4} \times \frac{A}{4} \times 128$ 的特征图;

标准卷积层 3_1_2: 用 128 个 1×1 的卷积核与标准卷积层 3_1_1 输出的特征图做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为 $\frac{A}{4} \times \frac{A}{4} \times 128$ 的特征图;

标准卷积层 3_2: 用 256 个 3×3 的卷积核与标准卷积层 3_1_2 输出的特征图做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为 $\frac{A}{4} \times \frac{A}{4} \times 256$ 的特征图;

标准卷积层 3_2_1: 用 128 个 1×1 的卷积核与标准卷积层 3_2 输出的特征图做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为 $\frac{A}{4} \times \frac{A}{4} \times 128$ 的特征图;

标准卷积层 3_2_2: 用 128 个 1×1 的卷积核与标准卷积层 3_2_1 输出的特征图做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为

$\frac{A}{4} \times \frac{A}{4} \times 128$ 的特征图;

标准卷积层 3_3: 用 256 个 3×3 的卷积核与标准卷积层 3_2_2 输出的特征图做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为 $\frac{A}{4} \times \frac{A}{4} \times 256$ 的特征图;

标准卷积层 3_3_1: 用 128 个 1×1 的卷积核与标准卷积层 3_3 输出的特征图做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为 $\frac{A}{4} \times \frac{A}{4} \times 128$ 的特征图;

标准卷积层 3_3_2: 用 128 个 1×1 的卷积核与标准卷积层 3_3_1 输出的特征图做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为 $\frac{A}{4} \times \frac{A}{4} \times 128$ 的特征图;

标准卷积层 3_4: 用 2 个 3×3 的卷积核与标准卷积层 3_3_2 输出的特征图做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为 $\frac{A}{4} \times \frac{A}{4} \times 2$ 的特征图;

反卷积层 4: 用 2 个 3×3 的卷积核与标准卷积层 3_4 输出的特征图做反卷积, 步长为 2, 得到维度为 $\frac{A}{2} \times \frac{A}{2} \times 2$ 的特征图;

反卷积层 5: 用 2 个 3×3 的卷积核与反卷积层 4 输出的特征图做反卷积, 步长为 2, 得到维度为 $A \times A \times 2$ 的特征图;

子步骤 2: 设计第二个针对大尺度行人的卷积神经网络, 具体包括:

① 设计池化层, 池化层数量为 2, 均采用最大池化操作, 采样尺寸均为 2×2 , 步长均为 2;

② 设计扩张卷积层, 扩张卷积层数量为 7, 扩张率分别为 2、4、

8、2、4、2、4，卷积核大小均为 3×3 ，步长均为1，卷积核数量分别为128、128、256、256、256、512、512；

③ 设计标准卷积层，标准卷积层数量为4，卷积核大小均为 3×3 ，步长均为1，卷积核数量分别为64、64、512、2；

④ 设计反卷积层，反卷积层数量为2，卷积核大小均为 3×3 ，步长均为2，卷积核数量分别为2、2；

⑤ 确定网络架构，根据步骤（2）的子步骤2中①~④涉及的网络层参数建立不同的网络模型，然后利用步骤（1）所建立的数据集对这些模型进行验证，从中筛选出兼顾准确性和实时性的网络结构，得到最优网络架构如下：

标准卷积层 1_1：用64个 3×3 的卷积核与 $A \times A$ 像素的输入样本做卷积，步长为1，再经过 ReLU 激活，得到维度为 $A \times A \times 64$ 的特征图；

标准卷积层 1_2：用64个 3×3 的卷积核与标准卷积层 1_1 输出的特征图做卷积，步长为1，再经过 ReLU 激活，得到维度为 $A \times A \times 64$ 的特征图；

池化层 1：用 2×2 的核对标准卷积层 1_2 输出的特征图做最大池化，步长为2，得到维度为 $\frac{A}{2} \times \frac{A}{2} \times 64$ 的特征图；

扩张卷积层 2_1：用128个 3×3 的卷积核与池化层 1 输出的特征图做卷积，步长为1，扩张率为2，再经过 ReLU 激活，得到维度为 $\frac{A}{2} \times \frac{A}{2} \times 128$ 的特征图；

扩张卷积层 2_2：用128个 3×3 的卷积核与扩张卷积层 2_1 输出的特征图做卷积，步长为1，扩张率为4，再经过 ReLU 激活，得到维

度为 $\frac{A}{2} \times \frac{A}{2} \times 128$ 的特征图；

池化层 2：用 2×2 的核对扩张卷积层 2_2 输出的特征图做最大池化，步长为 2，得到维度为 $\frac{A}{4} \times \frac{A}{4} \times 128$ 的特征图；

扩张卷积层 3_1：用 256 个 3×3 的卷积核与池化层 2 输出的特征图做卷积，步长为 1，扩张率为 8，再经过 ReLU 激活，得到维度为 $\frac{A}{4} \times \frac{A}{4} \times 256$ 的特征图；

扩张卷积层 3_2：用 256 个 3×3 的卷积核与扩张卷积层 3_1 输出的特征图做卷积，步长为 1，扩张率为 2，再经过 ReLU 激活，得到维度为 $\frac{A}{4} \times \frac{A}{4} \times 256$ 的特征图；

扩张卷积层 3_3：用 256 个 3×3 的卷积核与扩张卷积层 3_2 输出的特征图做卷积，步长为 1，扩张率为 4，再经过 ReLU 激活，得到维度为 $\frac{A}{4} \times \frac{A}{4} \times 256$ 的特征图；

标准卷积层 3_4：用 512 个 3×3 的卷积核与扩张卷积层 3_3 输出的特征图做卷积，步长为 1，再经过 ReLU 激活，得到维度为 $\frac{A}{4} \times \frac{A}{4} \times 512$ 的特征图；

扩张卷积层 3_5：用 512 个 3×3 的卷积核与标准卷积层 3_4 输出的特征图做卷积，步长为 1，扩张率为 2，再经过 ReLU 激活，得到维度为 $\frac{A}{4} \times \frac{A}{4} \times 512$ 的特征图；

扩张卷积层 3_6：用 512 个 3×3 的卷积核与扩张卷积层 3_5 输出的特征图做卷积，步长为 1，扩张率为 4，再经过 ReLU 激活，得到维度为 $\frac{A}{4} \times \frac{A}{4} \times 512$ 的特征图；

标准卷积层 3_7：用 2 个 3×3 的卷积核与扩张卷积层 3_6 输出的

特征图做卷积，步长为 1，再经过 ReLU 激活，得到维度为 $\frac{A}{4} \times \frac{A}{4} \times 2$ 的特征图；

反卷积层 4：用 2 个 3×3 的卷积核与标准卷积层 3_7 输出的特征图做反卷积，步长为 2，得到维度为 $\frac{A}{2} \times \frac{A}{2} \times 2$ 的特征图；

反卷积层 5：用 2 个 3×3 的卷积核与反卷积层 4 输出的特征图做反卷积，步长为 2，得到维度为 $A \times A \times 2$ 的特征图；

子步骤 3：提出两级融合策略对两路网络提取的特征进行融合，具体包括：

① 确定第一个卷积神经网络的局部特征和全局特征所在位置，局部特征位于从左至右第 9 个卷积层，全局特征位于从左至右第 18 个卷积层；

② 确定第二个卷积神经网络的局部特征和全局特征所在位置，局部特征位于从左至右第 5 个卷积层，全局特征位于从左至右第 11 个卷积层；

③ 融合两个网络的变尺度同级特征，将第一个网络第 9 个卷积层提取的局部特征与第二个网络第 5 个卷积层提取的局部特征融合，再将第一个网络第 18 个卷积层提取的全局特征与第二个网络第 11 个卷积层提取的全局特征融合；

融合第二个网络的局部特征与全局特征，使用 1×1 卷积对第二个网络浅层所包含的变尺度行人局部特征进行降维，使其具有与深层全局特征相同的维度，然后构建跳跃连接结构将局部特征与全局特征融合，得到变尺度多特征融合卷积神经网络架构；

(3) 训练设计的变尺度多特征融合卷积神经网络，获得网络参数；

(4) 使用变尺度多特征融合卷积神经网络进行行人分割。

与现有技术相比，本发明具有如下优点和有益效果：

(1) 本发明提出了一种适用于智能路侧终端图像的行人分割方法，借助路侧设备感知范围广的优势，实现行人的超视距全局感知，克服了当前大部分智能车辆主要依靠自车传感器进行行人感知从而容易出现视距盲点的不足，有效降低了行人漏检率；

(2) 本发明设计了并行的两个针对不同尺度行人的卷积神经网络来提取智能路侧终端图像中的行人特征，进而提出两级融合策略对所提取的特征进行融合，首先通过同级特征融合获得变尺度行人的局部特征和全局特征，进而对融合后的局部特征与全局特征进行二次融合，得到变尺度多特征融合卷积神经网络。该网络不仅大大降低了行人尺度差异化对分割精度的影响，而且兼顾了不同尺度行人的局部细节信息和全局信息，与当前大部分基于单一网络结构的行人分割方法相比，有效地解决了分割边界模糊以及漏分割问题，提高了行人分割的准确性和鲁棒性。

附图说明

图 1 是本发明的变尺度多特征融合卷积神经网络设计流程图；

图 2 是本发明设计的变尺度多特征融合卷积神经网络结构示意图；

图 3 是本发明设计的变尺度多特征融合卷积神经网络训练流程图。

具体实施方式

以下将结合具体实施例对本发明提供的技术方案进行详细说明，应理解下述具体实施方式仅用于说明本发明而不适用于限制本发明的范围。

本发明公开了一种基于变尺度多特征融合卷积网络的路侧图像行人分割方法。该方法设计了并行的两个卷积神经网络来提取图像中不同尺度行人的局部特征和全局特征，进而提出两级融合策略对所提取的特征进行融合，首先对不同尺度的同级特征进行融合，得到适用于变尺度行人的局部和全局特征，然后再将上一步得到的局部特征与全局特征融合，获取完备的变尺度行人局部细节信息和全局信息，最终得到变尺度多特征融合卷积神经网络，有效地解决了当前大部分基于单一网络结构的行人分割方法难以适用于变尺度行人的问题，进一步提高了行人分割的准确性和鲁棒性。

具体的说，本发明提供的基于变尺度多特征融合卷积网络的路侧图像行人分割方法，包括如下步骤：

(1) 建立行人分割数据集，对智能路侧终端获取的行人样本进行标注或者使用已有数据样本，然后将样本尺寸调整为 227×227 像素并记为 D_k 。

(2) 设计变尺度多特征融合卷积神经网络架构，该架构由并行的两个卷积神经网络组成，第一个网络针对小尺度行人设计了精细的特征提取结构，第二个网络针对大尺度行人扩大了网络在浅层处的感受野，然后将第一个网络提取的局部特征以及全局特征与第二网络提

取的局部特征以及全局特征进行同级融合,进而构建跳跃连接结构对融合后的局部特征与全局特征进行二次融合,设计流程如图1所示,包括以下子步骤:

子步骤1:设计第一个针对小尺度行人的卷积神经网络,具体包括:

① 设计池化层,在面向语义分割的卷积神经网络中,池化层一方面能够缩小特征图的尺寸以减小计算量,另一方面能够扩大感受野以捕捉更完整的行人信息,然而频繁地池化操作容易造成行人空间位置信息的损失,阻碍了分割精度的提高。相反,虽然无池化操作保留了尽量多的空间位置信息,但加大了计算负担。因此,在设计时综合考虑这两方面的影响,

设池化层数量为 n_{p1} ,取值范围为 $2\sim 3$,均采用最大池化操作,采样尺寸均为 2×2 ,步长均为2;

② 设计标准卷积层,为了准确地提取图像中小尺度行人的特征,设计了精细的特征提取结构,该结构由级联的卷积核大小均为 1×1 的两个标准卷积层组成,可用来提取小尺度行人的局部细节特征。此外,为了充分发挥卷积神经网络的局部感知优势,网络中也使用了 3×3 大小的卷积核。一般而言,网络的特征表达能力随着卷积层数量的增加而增强,但堆叠较多的卷积层加大了计算量,而卷积层数量少则难以提取到表达能力强的行人特征。鉴于此,设卷积核为 1×1 的标准卷积层数量为 n_f ,取值范围为 $2\sim 12$,卷积核数量为 $n_b(b=1,2,\dots,n_f)$, n_b 一般取值为2的整数次幂,步长均为1;设卷积核为 3×3 的标准卷积层

数量为 n_{s1} ，取值范围为 $5 \sim 10$ ，卷积核数量为 $n_{a1}(a1=1,2,...,n_{s1})$ ， n_{a1} 一般取值为 2 的整数次幂，步长均为 1；

③ 设计反卷积层，由于在步骤（2）子步骤 1 的中执行了 n_{p1} 次池化操作，使得特征图缩小了 $1/n_{p1}$ 倍，为了使特征图恢复到原始图片大小，同时避免引入大量噪声，采用 n_{p1} 个参数可学习的反卷积层对特征图中包含的行人特征进行解耦，由于行人分割任务是对每个像素进行二分类，所以反卷积层的卷积核数量均为 2，卷积核大小均为 3×3 ，步长均为 2。

④ 确定网络架构，根据步骤（2）子步骤 1 中各变量的取值范围，建立不同的网络模型，然后利用步骤（1）所建立的数据集对这些模型进行验证，从中筛选出兼顾准确性和实时性的最优网络架构。其中，池化层数量 $n_{p1}=2$ ；卷积核为 1×1 的标准卷积层数量 $n_f=10$ ，对应的卷积核数量 n_b 分别为 32、32、64、64、128、128、128、128、128、128；卷积核为 3×3 的标准卷积层数量 $n_{s1}=8$ ，对应的卷积核数量 n_{a1} 分别为 64、64、128、128、256、256、256、2。第一个卷积神经网络具体结构表述如下：

标准卷积层 1_1：用 64 个 3×3 的卷积核与 227×227 像素的输入样本做卷积，步长为 1，再经过 ReLU 激活，得到维度为 $227 \times 227 \times 64$ 的特征图；

标准卷积层 1_1_1：用 32 个 1×1 的卷积核与标准卷积层 1_1 输出的特征图做卷积，步长为 1，再经过 ReLU 激活，得到维度为 $227 \times 227 \times 32$ 的特征图；

标准卷积层 1_1_2: 用 32 个 1×1 的卷积核与标准卷积层 1_1_1 输出的特征图做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为 $227 \times 227 \times 32$ 的特征图;

标准卷积层 1_2: 用 64 个 3×3 的卷积核与标准卷积层 1_1_2 输出的特征图做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为 $227 \times 227 \times 64$ 的特征图;

池化层 1: 用 2×2 的核对标准卷积层 1_2 输出的特征图做最大池化, 步长为 2, 得到维度为 $113 \times 113 \times 64$ 的特征图;

标准卷积层 2_1: 用 128 个 3×3 的卷积核与池化层 1 输出的特征图做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为 $113 \times 113 \times 128$ 的特征图;

标准卷积层 2_1_1: 用 64 个 1×1 的卷积核与标准卷积层 2_1 输出的特征图做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为 $113 \times 113 \times 64$ 的特征图;

标准卷积层 2_1_2: 用 64 个 1×1 的卷积核与标准卷积层 2_1_1 输出的特征图做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为 $113 \times 113 \times 64$ 的特征图;

标准卷积层 2_2: 用 128 个 3×3 的卷积核与标准卷积层 2_1_2 输出的特征图做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为 $113 \times 113 \times 128$ 的特征图;

池化层 2: 用 2×2 的核对标准卷积层 2_2 输出的特征图做最大池化, 步长为 2, 得到维度为 $56 \times 56 \times 128$ 的特征图;

标准卷积层 3_1: 用 256 个 3×3 的卷积核与池化层 2 输出的特征图做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为 $56 \times 56 \times 256$ 的特征图;

标准卷积层 3_1_1: 用 128 个 1×1 的卷积核与标准卷积层 3_1 输出的特征图做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为 $56 \times 56 \times 128$ 的特征图;

标准卷积层 3_1_2: 用 128 个 1×1 的卷积核与标准卷积层 3_1_1 输出的特征图做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为 $56 \times 56 \times 128$ 的特征图;

标准卷积层 3_2: 用 256 个 3×3 的卷积核与标准卷积层 3_1_2 输出的特征图做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为 $56 \times 56 \times 256$ 的特征图;

标准卷积层 3_2_1: 用 128 个 1×1 的卷积核与标准卷积层 3_2 输出的特征图做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为 $56 \times 56 \times 128$ 的特征图;

标准卷积层 3_2_2: 用 128 个 1×1 的卷积核与标准卷积层 3_2_1 输出的特征图做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为 $56 \times 56 \times 128$ 的特征图;

标准卷积层 3_3: 用 256 个 3×3 的卷积核与标准卷积层 3_2_2 输出的特征图做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为 $56 \times 56 \times 256$ 的特征图;

标准卷积层 3_3_1: 用 128 个 1×1 的卷积核与标准卷积层 3_3 输

出的特征图做卷积，步长为 1，再经过 ReLU 激活，得到维度为 $56 \times 56 \times 128$ 的特征图；

标准卷积层 3_3_2：用 128 个 1×1 的卷积核与标准卷积层 3_3_1 输出的特征图做卷积，步长为 1，再经过 ReLU 激活，得到维度为 $56 \times 56 \times 128$ 的特征图；

标准卷积层 3_4：用 2 个 3×3 的卷积核与标准卷积层 3_3_2 输出的特征图做卷积，步长为 1，再经过 ReLU 激活，得到维度为 $56 \times 56 \times 2$ 的特征图；

反卷积层 4：用 2 个 3×3 的卷积核与标准卷积层 3_4 输出的特征图做反卷积，步长为 2，得到维度为 $113 \times 113 \times 2$ 的特征图；

反卷积层 5：用 2 个 3×3 的卷积核与反卷积层 4 输出的特征图做反卷积，步长为 2，得到维度为 $227 \times 227 \times 2$ 的特征图。

子步骤 2：设计第二个针对大尺度行人的卷积神经网络，具体包括：

① 设计池化层，由步骤（2）子步骤 1 的①中所知，频繁地使用池化层造成了行人空间位置信息极大的损失，极易引起分割精度的下降，虽然无池化操作能够保留更多的空间位置信息，但却加大了计算资源消耗。因此，在设计时同时考虑这两方面的影响，设池化层数量为 n_{p2} ，取值范围为 $2 \sim 3$ ，均采用最大池化操作，采样尺寸均为 2×2 ，步长均为 2；

② 设计扩张卷积层，利用扩张卷积在不改变特征图尺寸的前提下能够扩大感受野的优点，分别在网络的浅层以及深层处使用扩张卷

积代替标准卷积以完整地捕捉浅层处大尺度行人的边界特征和深层处的全局特征。虽然堆叠扩张卷积层以及使用大扩张率增大了局部感受野，但引入了噪声，而且感受野太大使得网络忽略了行人的局部细节信息，造成分割不连续甚至漏分割。与之相反，感受野过小则使得卷积层难以感知行人的全局信息。基于上述考虑，设扩张卷积层数量为 n_d ，取值范围为 $6 \sim 10$ ，扩张率为 $d_r (r=1,2,\dots,n_d)$ ， d_r 为偶数，取值范围为 $2 \sim 10$ ，卷积核数量为 $n_e (e=1,2,\dots,n_d)$ ， n_e 一般取值为 2 的整数次幂，卷积核大小均为 3×3 ，步长均为 1；

③ 设计标准卷积层，一般而言，网络的特征表达能力随着卷积层数量的增加而增强，但堆叠较多的卷积层加大了计算负担，而卷积层数量少则难以提取到表达能力强的行人特征。考虑到步骤 (2) 子步骤 2 的②中已经设计了扩张卷积层，设标准卷积层数量为 n_{s2} ，取值范围为 $2 \sim 7$ ，卷积核数量为 $n_{a2} (a2=1,2,\dots,n_{s2})$ ， n_{a2} 一般取值为 2 的整数次幂，卷积核大小均为 3×3 ，步长均为 1；

④ 设计反卷积层，由于在步骤 (2) 子步骤 2 的①中执行了 n_{p2} 次池化操作，特征图缩小了 $1/n_{p2}$ 倍，为了将其恢复到原始图片大小，同时避免引入大量噪声，采用 n_{p2} 个参数可学习的反卷积层对特征图中包含的行人特征进行解耦，反卷积层的卷积核数量均为 2，卷积核大小均为 3×3 ，步长均为 2。

⑤ 确定网络架构，根据步骤 (2) 子步骤 2 中各变量的取值范围，建立不同的网络模型，然后利用步骤 (1) 所建立的数据集对这些模型进行验证，从中筛选出兼顾准确性和实时性的最优网络架构。其中，

池化层数量 $n_{p2}=2$ ；扩张卷积层数量 $n_d=7$ ，扩张率 d_r 分别为 2、4、8、2、4、2、4，对应的卷积核数量 n_e 分别为 128、128、256、256、256、512、512；标准卷积层数量 $n_{s2}=4$ ，对应的卷积核数量 n_{a2} 分别为 64、64、512、2。第二个卷积神经网络具体结构表述如下：

标准卷积层 1_1：用 64 个 3×3 的卷积核与 227×227 像素的输入样本做卷积，步长为 1，再经过 ReLU 激活，得到维度为 $227\times 227\times 64$ 的特征图；

标准卷积层 1_2：用 64 个 3×3 的卷积核与标准卷积层 1_1 输出的特征图做卷积，步长为 1，再经过 ReLU 激活，得到维度为 $227\times 227\times 64$ 的特征图；

池化层 1：用 2×2 的核对标准卷积层 1_2 输出的特征图做最大池化，步长为 2，得到维度为 $113\times 113\times 64$ 的特征图；

扩张卷积层 2_1：用 128 个 3×3 的卷积核与池化层 1 输出的特征图做卷积，步长为 1，扩张率为 2，再经过 ReLU 激活，得到维度为 $113\times 113\times 128$ 的特征图；

扩张卷积层 2_2：用 128 个 3×3 的卷积核与扩张卷积层 2_1 输出的特征图做卷积，步长为 1，扩张率为 4，再经过 ReLU 激活，得到维度为 $113\times 113\times 128$ 的特征图；

池化层 2：用 2×2 的核对扩张卷积层 2_2 输出的特征图做最大池化，步长为 2，得到维度为 $56\times 56\times 128$ 的特征图；

扩张卷积层 3_1：用 256 个 3×3 的卷积核与池化层 2 输出的特征图做卷积，步长为 1，扩张率为 8，再经过 ReLU 激活，得到维度为

56×56×256 的特征图；

扩张卷积层 3_2：用 256 个 3×3 的卷积核与扩张卷积层 3_1 输出的特征图做卷积，步长为 1，扩张率为 2，再经过 ReLU 激活，得到维度为 56×56×256 的特征图；

扩张卷积层 3_3：用 256 个 3×3 的卷积核与扩张卷积层 3_2 输出的特征图做卷积，步长为 1，扩张率为 4，再经过 ReLU 激活，得到维度为 56×56×256 的特征图；

标准卷积层 3_4：用 512 个 3×3 的卷积核与扩张卷积层 3_3 输出的特征图做卷积，步长为 1，再经过 ReLU 激活，得到维度为 56×56×512 的特征图；

扩张卷积层 3_5：用 512 个 3×3 的卷积核与标准卷积层 3_4 输出的特征图做卷积，步长为 1，扩张率为 2，再经过 ReLU 激活，得到维度为 56×56×512 的特征图；

扩张卷积层 3_6：用 512 个 3×3 的卷积核与扩张卷积层 3_5 输出的特征图做卷积，步长为 1，扩张率为 4，再经过 ReLU 激活，得到维度为 56×56×512 的特征图；

标准卷积层 3_7：用 2 个 3×3 的卷积核与扩张卷积层 3_6 输出的特征图做卷积，步长为 1，再经过 ReLU 激活，得到维度为 56×56×2 的特征图；

反卷积层 4：用 2 个 3×3 的卷积核与标准卷积层 3_7 输出的特征图做反卷积，步长为 2，得到维度为 113×113×2 的特征图；

反卷积层 5：用 2 个 3×3 的卷积核与反卷积层 4 输出的特征图做

反卷积，步长为 2，得到维度为 $227 \times 227 \times 2$ 的特征图。

子步骤 3：提出两级融合策略对两路网络提取的特征进行融合，具体包括：

① 确定第一个卷积神经网络的局部特征和全局特征所在位置，根据深度学习分层提取

特征的特点，即局部特征一般位于网络浅层，全局特征一般位于深层，初步确定局部特征的位置即位于第几个卷积层，记为 s_{11} ，取值范围为 $6 \sim 10$ ，然后通过特征可视化的手段确定 s_{11} 的具体数值。一般将最后一个标准卷积层所提取的特征作为全局特征以获取更加充分的行人全局信息，即全局特征位于从左至右第 18 个卷积层；

② 确定第二个卷积神经网络的局部特征和全局特征所在位置，根据步骤（2）子步骤 3 的①中所述的方法来确定局部特征和全局特征所在位置，其中局部特征的位置记为 s_{12} ，取值范围为 $3 \sim 6$ ，全局特征位于从左至右第 11 个卷积层；

③ 融合两个网络的变尺度同级特征，在 s_{11} 和 s_{12} 的取值范围内，通过特征可视化方法得到 s_{11} 取值为 9、 s_{12} 取值为 5 后，将第一个网络第 9 个卷积层提取的局部特征与第二个网络第 5 个卷积层提取的局部特征融合，再将第一个网络第 18 个卷积层提取的全局特征与第二个网络第 11 个卷积层提取的全局特征融合；

④ 融合第二个网络的局部特征与全局特征，为了减小特征融合时额外引入的网络参数数量，使用卷积核大小为 1×1 卷积对第二个网络的浅层所包含的变尺度行人局部特征进行降维，使其具有和深层全局

特征相同的维度，然后构建跳跃连接结构将局部特征与全局特征融合，得到的变尺度多特征融合卷积神经网络架构，结构示意图如图 2 所示。

(3) 训练设计的变尺度多特征融合卷积神经网络，通过随机梯度下降法迭代并优化网络参数。训练过程包括前向传播和反向传播两个阶段，在前向传播阶段，将样本集 $(\mathbf{x}, \mathbf{y})$ 输入网络，其中 $\mathbf{x}$ 为输入图像， $\mathbf{y}$ 为对应的标签。通过网络逐层运算得到实际输出 $f(\mathbf{x})$ ，使用带有 L2 正则化项的交叉熵代价函数来衡量理想输出 $\mathbf{y}$ 与实际输出 $f(\mathbf{x})$ 之间的误差：

$$J(\boldsymbol{\theta}) = -\frac{1}{M} \left[\sum_{i=1}^M \sum_{j=1}^N \sum_{q=1}^Q \mathbf{1}\{y_i^j = q\} \cdot \log p_q(x_i^j) \right] + \frac{\lambda}{2M} \boldsymbol{\theta}^T \cdot \boldsymbol{\theta} \quad (1)$$

式 (1) 中，第一项为交叉熵代价函数，第二项为 L2 正则化项，用以防止过拟合， $\boldsymbol{\theta}$ 表示卷积神经网络模型要学习的参数， M 表示训练样本数量， N 表示每幅图像的像素个数， Q 表示样本中语义类别的数量，对于道路分割而言， $Q=2$ ， $\mathbf{1}\{y=q\}$ 是指示函数，当 $y=q$ 时取值为 1，否则取值为 0， λ 是正则化系数， x_i^j 表示第 i 个样本中第 j 个像素的灰度值， y_i^j 表示 x_i^j 对应的标签， $p_q(x_i^j)$ 表示 x_i^j 属于第 q 个类别的概率，定义为：

$$p_q(x_i^j) = \frac{e^{f_q(x_i^j)}}{\sum_{q=1}^Q e^{f_q(x_i^j)}} \quad (2)$$

式 (2) 中， $f_q(x_i^j)$ 表示最后一个反卷积层的第 q 个特征图在 x_i^j 处的输出，定义为：

$$f_q(x_i^j) = \boldsymbol{\theta}_q^T \cdot \mathbf{x}_i^j \quad (3)$$

在反向传播阶段,通过随机梯度下降算法从后往前逐层更新网络参数以达到极小化实际输出与理想输出之间误差的目的,参数更新公式如下:

$$\theta \leftarrow (1 - \alpha \frac{\lambda}{M})\theta - \alpha \nabla_{\theta} J_0(\theta) \quad (4)$$

式(4)中, α 为学习率, $J_0(\theta)$ 为交叉熵代价函数, $\nabla_{\theta} J_0(\theta)$ 为计算的梯度。

当选定代价函数、正则化方法以及优化算法后,使用深度学习框架训练设计好的卷积神经网络。为了使训练结果更准确,正式训练前先进行预训练,然后再对预训练得到的参数进行微调,训练流程如图3所示,具体包括以下子步骤:

子步骤 1: 选择与自动驾驶相关的数据集,如 ApolloScape、Cityscapes、CamVid,对其进行处理使其只包含行人类别,然后将样本尺寸调整为 227×227 像素并记为 D_c ,接着使用 D_c 对设计好的两个卷积神经网络进行预训练,分别设置预训练超参数,其中最大迭代次数分别为 I_{c1} 、 I_{c2} ,学习率分别为 α_{c1} 、 α_{c2} ,权重衰减分别为 λ_{c1} 、 λ_{c2} ,最终将预训练得到的网络参数保存下来;

子步骤 2: 使用步骤(1)中所建立的数据集 D_k 对步骤(3)的子步骤 1 中预训练得到的两个网络的参数进行微调,设置最大迭代次数分别为 I_{k1} 、 I_{k2} ,学习率分别为 α_{k1} 、 α_{k2} ,权重衰减分别为 λ_{k1} 、 λ_{k2} ,然后根据训练损失曲线和验证损失曲线的变化情况,得到网络参数最优时的两个卷积神经网络模型;

子步骤 3: 使用步骤(1)中所建立的数据集 D_k 对步骤(2)的子

步骤 3 中得到的变尺度多特征融合卷积神经网络进行训练, 重新设置最大迭代次数为 I_{k3} , 学习率分别为 α_{k3} , 权重衰减分别为 λ_{k3} , 然后根据训练损失曲线和验证损失曲线的变化情况即当训练损失曲线缓慢下降趋于收敛而验证损失曲线处于上升的临界点时, 得到参数最优变尺度多特征融合卷积神经网络模型。

(4) 使用变尺度多特征融合卷积神经网络进行行人分割, 将智能路侧终端获取的行人样本尺寸调整为 227×227 像素并将其输入到已训练的变尺度多特征融合卷积神经网络中, 得到行人分割结果。

权 利 要 求

1、基于变尺度多特征融合卷积网络的路侧图像行人分割方法，其特征在于，包括以下步骤：

(1) 建立行人分割数据集；

(2) 构建变尺度多特征融合卷积神经网络，首先设计并行的两个卷积神经网络来提取图像中不同尺度行人的局部特征和全局特征，第一个网络针对小尺度行人设计了精细的特征提取结构，第二个网络针对大尺度行人扩大了网络在浅层处的感受野；进而提出两级融合策略对所提取的特征进行融合，首先对不同尺度的同级特征进行融合，得到适用于变尺度行人的局部特征和全局特征，然后构建跳跃连接结构将融合后的局部特征与全局特征进行二次融合，获取完备的变尺度行人局部细节信息和全局信息，最终得到变尺度多特征融合卷积神经网络，包括以下子步骤：

子步骤 1：设计第一个针对小尺度行人的卷积神经网络，具体包括：

① 设计池化层，池化层数量为 2，均采用最大池化操作，采样尺寸均为 2×2 ，步长均为 2；

② 设计标准卷积层，标准卷积层数量为 18，其中 8 层卷积核大小均为 3×3 ，卷积核数量分别为 64、64、128、128、256、256、256、2，步长均为 1，剩下 10 层卷积核大小均为 1×1 ，卷积核数量分别为 32、32、64、64、128、128、128、128、128、128，步长均为 1；

③ 设计反卷积层，反卷积层数量为 2，卷积核大小均为 3×3 ，步

长均为 2，卷积核数量分别为 2、2；

④ 确定网络架构，根据步骤（2）的子步骤 1 中①~③涉及的网络层参数建立不同的网络模型，然后利用步骤（1）所建立的数据集对这些模型进行验证，从中筛选出兼顾准确性和实时性的网络结构，得到最优网络架构如下：

标准卷积层 1_1：用 64 个 3×3 的卷积核与 $A \times A$ 像素的输入样本做卷积，步长为 1，再经过 ReLU 激活，得到维度为 $A \times A \times 64$ 的特征图；

标准卷积层 1_1_1：用 32 个 1×1 的卷积核与标准卷积层 1_1 输出的特征图做卷积，步长为 1，再经过 ReLU 激活，得到维度为 $A \times A \times 32$ 的特征图；

标准卷积层 1_1_2：用 32 个 1×1 的卷积核与标准卷积层 1_1_1 输出的特征图做卷积，步长为 1，再经过 ReLU 激活，得到维度为 $A \times A \times 32$ 的特征图；

标准卷积层 1_2：用 64 个 3×3 的卷积核与标准卷积层 1_1_2 输出的特征图做卷积，步长为 1，再经过 ReLU 激活，得到维度为 $A \times A \times 64$ 的特征图；

池化层 1：用 2×2 的核对标准卷积层 1_2 输出的特征图做最大池化，步长为 2，得到维度为 $\frac{A}{2} \times \frac{A}{2} \times 64$ 的特征图；

标准卷积层 2_1：用 128 个 3×3 的卷积核与池化层 1 输出的特征图做卷积，步长为 1，再经过 ReLU 激活，得到维度为 $\frac{A}{2} \times \frac{A}{2} \times 128$ 的特征图；

标准卷积层 2_1_1：用 64 个 1×1 的卷积核与标准卷积层 2_1 输出

的特征图做卷积,步长为 1,再经过 ReLU 激活,得到维度为 $\frac{A}{2} \times \frac{A}{2} \times 64$ 的特征图;

标准卷积层 2_1_2: 用 64 个 1×1 的卷积核与标准卷积层 2_1_1 输出的特征图做卷积,步长为 1,再经过 ReLU 激活,得到维度为 $\frac{A}{2} \times \frac{A}{2} \times 64$ 的特征图;

标准卷积层 2_2: 用 128 个 3×3 的卷积核与标准卷积层 2_1_2 输出的特征图做卷积,步长为 1,再经过 ReLU 激活,得到维度为 $\frac{A}{2} \times \frac{A}{2} \times 128$ 的特征图;

池化层 2: 用 2×2 的核对标准卷积层 2_2 输出的特征图做最大池化,步长为 2,得到维度为 $\frac{A}{4} \times \frac{A}{4} \times 128$ 的特征图;

标准卷积层 3_1: 用 256 个 3×3 的卷积核与池化层 2 输出的特征图做卷积,步长为 1,再经过 ReLU 激活,得到维度为 $\frac{A}{4} \times \frac{A}{4} \times 256$ 的特征图;

标准卷积层 3_1_1: 用 128 个 1×1 的卷积核与标准卷积层 3_1 输出的特征图做卷积,步长为 1,再经过 ReLU 激活,得到维度为 $\frac{A}{4} \times \frac{A}{4} \times 128$ 的特征图;

标准卷积层 3_1_2: 用 128 个 1×1 的卷积核与标准卷积层 3_1_1 输出的特征图做卷积,步长为 1,再经过 ReLU 激活,得到维度为 $\frac{A}{4} \times \frac{A}{4} \times 128$ 的特征图;

标准卷积层 3_2: 用 256 个 3×3 的卷积核与标准卷积层 3_1_2 输出的特征图做卷积,步长为 1,再经过 ReLU 激活,得到维度为 $\frac{A}{4} \times \frac{A}{4} \times 256$ 的特征图;

标准卷积层 3_2_1: 用 128 个 1×1 的卷积核与标准卷积层 3_2 输出的特征图做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为 $\frac{A}{4} \times \frac{A}{4} \times 128$ 的特征图;

标准卷积层 3_2_2: 用 128 个 1×1 的卷积核与标准卷积层 3_2_1 输出的特征图做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为 $\frac{A}{4} \times \frac{A}{4} \times 128$ 的特征图;

标准卷积层 3_3: 用 256 个 3×3 的卷积核与标准卷积层 3_2_2 输出的特征图做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为 $\frac{A}{4} \times \frac{A}{4} \times 256$ 的特征图;

标准卷积层 3_3_1: 用 128 个 1×1 的卷积核与标准卷积层 3_3 输出的特征图做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为 $\frac{A}{4} \times \frac{A}{4} \times 128$ 的特征图;

标准卷积层 3_3_2: 用 128 个 1×1 的卷积核与标准卷积层 3_3_1 输出的特征图做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为 $\frac{A}{4} \times \frac{A}{4} \times 128$ 的特征图;

标准卷积层 3_4: 用 2 个 3×3 的卷积核与标准卷积层 3_3_2 输出的特征图做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为 $\frac{A}{4} \times \frac{A}{4} \times 2$ 的特征图;

反卷积层 4: 用 2 个 3×3 的卷积核与标准卷积层 3_4 输出的特征图做反卷积, 步长为 2, 得到维度为 $\frac{A}{2} \times \frac{A}{2} \times 2$ 的特征图;

反卷积层 5: 用 2 个 3×3 的卷积核与反卷积层 4 输出的特征图做反卷积, 步长为 2, 得到维度为 $A \times A \times 2$ 的特征图;

子步骤 2：设计第二个针对大尺度行人的卷积神经网络，具体包括：

① 设计池化层，池化层数量为 2，均采用最大池化操作，采样尺寸均为 2×2 ，步长均为 2；

② 设计扩张卷积层，扩张卷积层数量为 7，扩张率分别为 2、4、8、2、4、2、4，卷积核大小均为 3×3 ，步长均为 1，卷积核数量分别为 128、128、256、256、256、512、512；

③ 设计标准卷积层，标准卷积层数量为 4，卷积核大小均为 3×3 ，步长均为 1，卷积核数量分别为 64、64、512、2；

④ 设计反卷积层，反卷积层数量为 2，卷积核大小均为 3×3 ，步长均为 2，卷积核数量分别为 2、2；

⑤ 确定网络架构，根据步骤（2）的子步骤 2 中①~④涉及的网络层参数建立不同的网络模型，然后利用步骤（1）所建立的数据集对这些模型进行验证，从中筛选出兼顾准确性和实时性的网络结构，得到最优网络架构如下：

标准卷积层 1_1：用 64 个 3×3 的卷积核与 $A \times A$ 像素的输入样本做卷积，步长为 1，再经过 ReLU 激活，得到维度为 $A \times A \times 64$ 的特征图；

标准卷积层 1_2：用 64 个 3×3 的卷积核与标准卷积层 1_1 输出的特征图做卷积，步长为 1，再经过 ReLU 激活，得到维度为 $A \times A \times 64$ 的特征图；

池化层 1：用 2×2 的核为标准卷积层 1_2 输出的特征图做最大池化，步长为 2，得到维度为 $\frac{A}{2} \times \frac{A}{2} \times 64$ 的特征图；

扩张卷积层 2_1: 用 128 个 3×3 的卷积核与池化层 1 输出的特征图做卷积, 步长为 1, 扩张率为 2, 再经过 ReLU 激活, 得到维度为 $\frac{A}{2} \times \frac{A}{2} \times 128$ 的特征图;

扩张卷积层 2_2: 用 128 个 3×3 的卷积核与扩张卷积层 2_1 输出的特征图做卷积, 步长为 1, 扩张率为 4, 再经过 ReLU 激活, 得到维度为 $\frac{A}{2} \times \frac{A}{2} \times 128$ 的特征图;

池化层 2: 用 2×2 的核对扩张卷积层 2_2 输出的特征图做最大池化, 步长为 2, 得到维度为 $\frac{A}{4} \times \frac{A}{4} \times 128$ 的特征图;

扩张卷积层 3_1: 用 256 个 3×3 的卷积核与池化层 2 输出的特征图做卷积, 步长为 1, 扩张率为 8, 再经过 ReLU 激活, 得到维度为 $\frac{A}{4} \times \frac{A}{4} \times 256$ 的特征图;

扩张卷积层 3_2: 用 256 个 3×3 的卷积核与扩张卷积层 3_1 输出的特征图做卷积, 步长为 1, 扩张率为 2, 再经过 ReLU 激活, 得到维度为 $\frac{A}{4} \times \frac{A}{4} \times 256$ 的特征图;

扩张卷积层 3_3: 用 256 个 3×3 的卷积核与扩张卷积层 3_2 输出的特征图做卷积, 步长为 1, 扩张率为 4, 再经过 ReLU 激活, 得到维度为 $\frac{A}{4} \times \frac{A}{4} \times 256$ 的特征图;

标准卷积层 3_4: 用 512 个 3×3 的卷积核与扩张卷积层 3_3 输出的特征图做卷积, 步长为 1, 再经过 ReLU 激活, 得到维度为 $\frac{A}{4} \times \frac{A}{4} \times 512$ 的特征图;

扩张卷积层 3_5: 用 512 个 3×3 的卷积核与标准卷积层 3_4 输出的特征图做卷积, 步长为 1, 扩张率为 2, 再经过 ReLU 激活, 得到维

度为 $\frac{A}{4} \times \frac{A}{4} \times 512$ 的特征图；

扩张卷积层 3_6：用 512 个 3×3 的卷积核与扩张卷积层 3_5 输出的特征图做卷积，步长为 1，扩张率为 4，再经过 ReLU 激活，得到维度为 $\frac{A}{4} \times \frac{A}{4} \times 512$ 的特征图；

标准卷积层 3_7：用 2 个 3×3 的卷积核与扩张卷积层 3_6 输出的特征图做卷积，步长为 1，再经过 ReLU 激活，得到维度为 $\frac{A}{4} \times \frac{A}{4} \times 2$ 的特征图；

反卷积层 4：用 2 个 3×3 的卷积核与标准卷积层 3_7 输出的特征图做反卷积，步长为 2，得到维度为 $\frac{A}{2} \times \frac{A}{2} \times 2$ 的特征图；

反卷积层 5：用 2 个 3×3 的卷积核与反卷积层 4 输出的特征图做反卷积，步长为 2，得到维度为 $A \times A \times 2$ 的特征图；

子步骤 3：提出两级融合策略对两路网络提取的特征进行融合，具体包括：

① 确定第一个卷积神经网络的局部特征和全局特征所在位置，局部特征位于从左至右第 9 个卷积层，全局特征位于从左至右第 18 个卷积层；

② 确定第二个卷积神经网络的局部特征和全局特征所在位置，局部特征位于从左至右第 5 个卷积层，全局特征位于从左至右第 11 个卷积层；

③ 融合两个网络的变尺度同级特征，将第一个网络第 9 个卷积层提取的局部特征与第二个网络第 5 个卷积层提取的局部特征融合，再将第一个网络第 18 个卷积层提取的全局特征与第二个网络第 11 个卷

积层提取的全局特征融合；

④ 融合第二个网络的局部特征与全局特征，使用 1×1 卷积对第二个网络浅层所包含的变尺度行人局部特征进行降维，使其具有与深层全局特征相同的维度，然后构建跳跃连接结构将局部特征与全局特征融合，得到变尺度多特征融合卷积神经网络架构；

(3) 训练设计的变尺度多特征融合卷积神经网络，获得网络参数；

(4) 使用变尺度多特征融合卷积神经网络进行行人分割。

附图

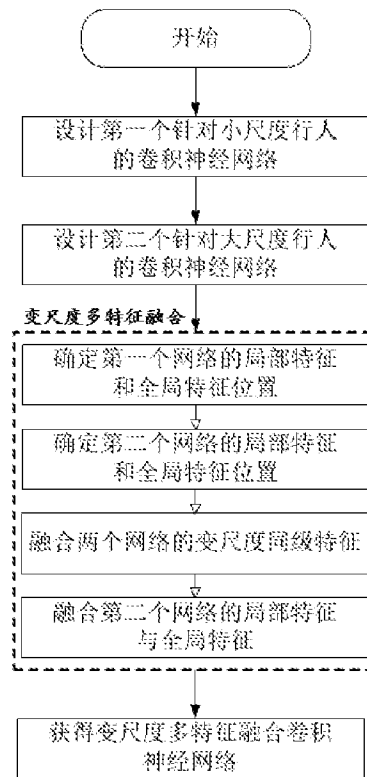


图 1

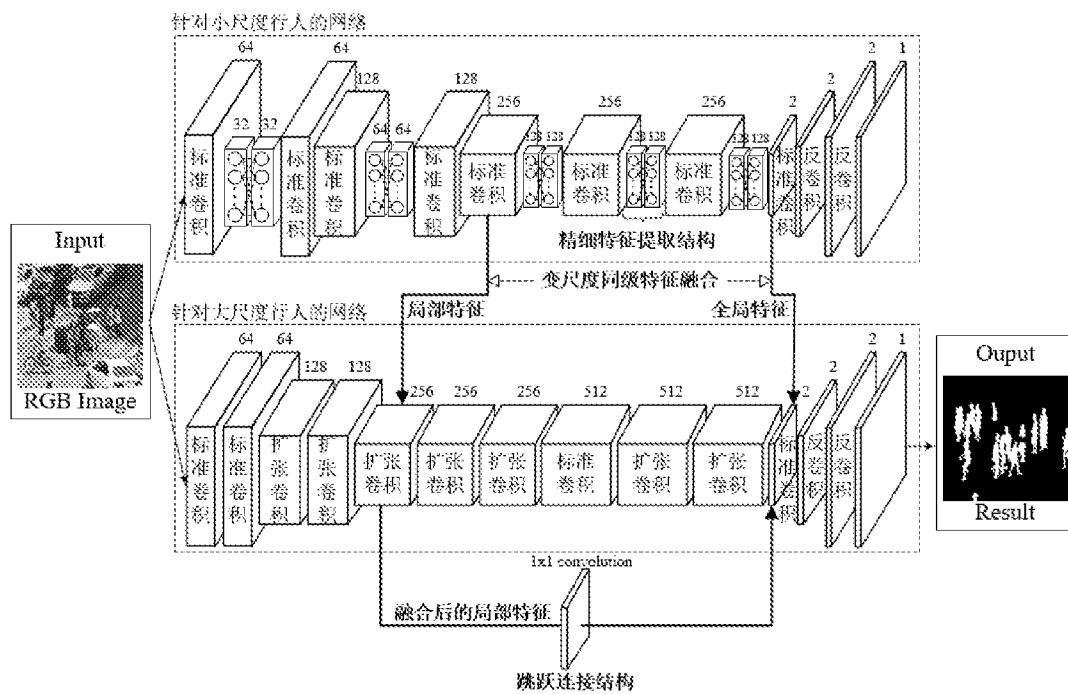


图 2

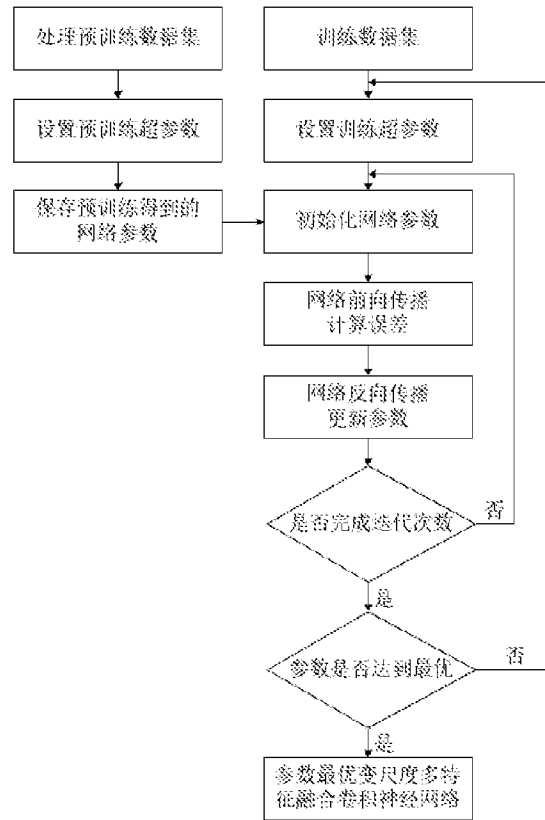


图 3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/087164

A. CLASSIFICATION OF SUBJECT MATTER

G06K 9/00(2006.01)i; G06K 9/34(2006.01)i; G06N 3/04(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06K, G06N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, EPODOC, WPI, GOOGLE, CNKI: 图像, 识别, 检测, 分割, 行人, 卷积, 神经网络, 尺度, 变, 多, 并行, 融合, image, recognize, segment, pedestrian, convolution, network, CNN

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 108520219 A (TAIZHOU ZHIBIAN TECHNOLOGY CO., LTD.) 11 September 2018 (2018-09-11) description, paragraphs [0068]-[0088], and figure 2	1
A	CN 107688786 A (NANJING UNIVERSITY OF SCIENCE & TECHNOLOGY) 13 February 2018 (2018-02-13) entire document	1
A	WO 2018003212 A1 (CLARION CO., LTD.) 04 January 2018 (2018-01-04) entire document	1
A	LI, Jianan et al. "Scale-Aware Fast R-CNN for Pedestrian Detection" <i>IEEE TRANSACTIONS ON MULTIMEDIA</i> , Vol. 20, No. 4, 30 April 2018 (2018-04-30), pp. 985-996	1



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

13 November 2019

Date of mailing of the international search report

28 November 2019

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/087164

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)		Publication date (day/month/year)
CN	108520219	A	11 September 2018	None		
CN	107688786	A	13 February 2018	None		
WO	2018003212	A1	04 January 2018	JP	2018005520	A 11 January 2018

国际检索报告

国际申请号

PCT/CN2019/087164

A. 主题的分类

G06K 9/00(2006.01)i; G06K 9/34(2006.01)i; G06N 3/04(2006.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06K, G06N

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CNPAT, EPDOC, WPI, G00GLE, CNKI: 图像, 识别, 检测, 分割, 行人, 卷积, 神经网络, 尺度, 变, 多, 并行, 融合, image, recognize, segment, pedestrian, convolution, network, CNN

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
A	CN 108520219 A (台州智必安科技有限责任公司) 2018年 9月 11日 (2018 - 09 - 11) 说明书第[0068]-[0088]段, 图2	1
A	CN 107688786 A (南京理工大学) 2018年 2月 13日 (2018 - 02 - 13) 全文	1
A	WO 2018003212 A1 (CLARION CO., LTD.) 2018年 1月 4日 (2018 - 01 - 04) 全文	1
A	LI, Jianan 等. "Scale-Aware Fast R-CNN for Pedestrian Detection" IEEE TRANSACTIONS ON MULTIMEDIA, 第20卷, 第4期, 2018年 4月 30日 (2018 - 04 - 30), 第985-996页	1

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2019年 11月 13日

国际检索报告邮寄日期

2019年 11月 28日

ISA/CN的名称和邮寄地址

中国国家知识产权局(ISA/CN)

中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

受权官员

李娜

电话号码 86-(10)-53961403

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/087164

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	108520219	A	2018年 9月 11日	无	
CN	107688786	A	2018年 2月 13日	无	
WO	2018003212	A1	2018年 1月 4日	JP 2018005520 A	2018年 1月 11日