



(12) 发明专利

(10) 授权公告号 CN 110458844 B

(45) 授权公告日 2022.02.15

(21) 申请号 201910659062.6

G06N 3/04 (2006.01)

(22) 申请日 2019.07.22

(56) 对比文件

(65) 同一申请的已公布的文献号

CN 104850865 A, 2015.08.19

申请公布号 CN 110458844 A

US 2019171223 A1, 2019.06.06

CN 109410129 A, 2019.03.01

(43) 申请公布日 2019.11.15

Dengxin Dai et al..Dark Model

(73) 专利权人 大连理工大学

Adaptation: Semantic Image Segmentation

地址 116024 辽宁省大连市甘井子区凌工
路2号

from Daytime to Nighttime.《http:arXiv:

1810.02575v1》.2018,

(72) 发明人 杨鑫 朱锦程 王昊然 魏小鹏

李良福等.基于生成式对抗网络的细小桥梁

张强 尹宝才

裂缝分割方法.《激光与光电子学进展》.2019,第

56卷(第10期),

(74) 专利代理机构 大连理工大学专利中心

审查员 郑岩

21200

代理人 梅洪玉 刘秋彤

(51) Int. Cl.

G06T 7/10 (2017.01)

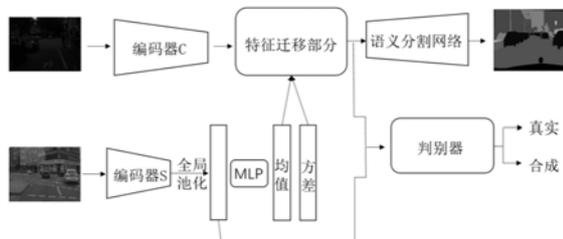
权利要求书2页 说明书5页 附图1页

(54) 发明名称

一种低光照场景的语义分割方法

(57) 摘要

一种低光照场景的语义分割方法,属于计算机视觉技术领域。本发明将正常图像的语义分割问题当做源域问题,低光照图像的语义分割问题当做目标域问题,利用迁移学习中的特征迁移的方法,充分发挥正常场景图像的信息充足的优势,将正常场景中的有用信息提取出来,并将这些有用信息与低光照图像的特征信息进行转换的结合,获取更多的有助于语义分割的图像信息,从而训练深度神经网络。基于这一思想,在生成对抗网络的基础之上,利用迁移学习的方法,设计并实现了用于低光照场景直接语义分割的网络模型。利用这一模型,能够有效解决低光照图片的语义分割任务。



1. 一种低光照场景的语义分割方法,其特征在于,步骤如下:

(一) 基于迁移学习的语义分割网络架构

所述基于迁移学习的语义分割网络包括以下四个部分:

编码器部分:语义分割网络包含两个编码器,编码器C和编码器S;编码器C用于提取低光照图像的低光照图像特征信息,编码器C的前半部分由采用了ResNet的网络结构前半部分,包含50个卷积层以及残差和Bottleneck结构,用来减小参数加速收敛,并能够获得包含2048个通道的特征图;随后将获得的2048个通道的特征图进行实例标准化,并对特征进行正则化,然后通过 1×1 的卷积层将特征通道数缩小到256个,得到低光照图像的特征图 F_c ;编码器S用于提取正常场景图像风格特征的网络部分,其包括不同卷积核大小卷积层,不包括池化层,使用四个步长为2的卷积层,能达到缩小特征维度、减少网络参数的作用;随着卷积的进行,特征的通道数也在不断的增加到256个通道,然后通过一个全局池化层一步缩小特征维度,最后使用一个全连接层得到一个代表正常场景图像风格特征的256维的列向量 F_s ;

特征迁移网络部分:黑天场景的特征图和白天场景的特征图在网络的处理下整体上都满足正态分布;使用的特征迁移的方法是为了改变低光照图像特征的正态分布的参数:均值和方差;在正常图像的均值和方差的获取上,使用网络学习的方式获取,对于输入的正常图像的特征 F_s ,利用两个多层感知机MLP分别学习两个列向量来分别代表各个特征层每层的均值 s_mean 和方差 s_val ;对于低光照图像部分,输入的特征 F_c 则直接计算均值 c_mean 和方差 c_val ,然后通过公式(1)进行特征的转换,得到转换后的特征 \hat{F}_c ;

$$\hat{F}_c = s_val * ((F_c - c_mean)/c_val) + s_mean \quad (1)$$

输入的低光照图像特征 F_c 计算完均值 c_mean 和方差 c_val 后经过公式(1)进行两次转换特征转换,将得到的特征与转换前的特征 F_c 相加得到结果的过程称作一次特征适应,共进行了3-5次特征适应的过程,得到最终转换后的低光照特征 F_{c_s} ;

判别器部分:判别器的结构是一个二分类的分类网络结构,分类为0代表夜晚场景,分类为1代表正常白天场景;判别器用来判别转换后的特征反编码后的结果属于夜晚场景还是白天场景,使得判别器无法判别特征转换后的特征属于哪种场景,或者使得判别器对转换后的特征评价尽可能的接近于1,说明转换比较成功,其损失函数用于监督网络的训练;判别器网络主要由多个步长为2的卷积层组成,不断缩小维度来获取最终的预测概率;

语义分割部分:对于语义分割部分,使用DeepLabv3网络的中提出的金子塔池化ASPP结构;

(二) 数据集采集

数据集方面,在Cityscapes数据集的基础之上利用Photoshop软件和Matlab人工合成低光照的场景数据集,并结合原有的公开数据组成完整的训练数据;

(三) 训练步骤

首先需要对编码器C和编码器S分别输入低光照的图像和用于提取特征信息的正常场景的图像;编码器C部分提取低光照场景图像中的特征后输入到特征迁移网络部分;编码器S部分提取正常场景的特征后经过两个多层感知机MLP,将最终得到的特征与编码器C部分输出的特征经过特征迁移部分进行一个特征的融合并迁移;低光照场景图像的特征经过特

征迁移部分之后得到一个近似正常场景图像的特征分布值,然后与正常场景图像的特征值同时输入到判别器部分,来判断特征迁移是否成功,迁移成功,图像则从低光照转变为较为正常的场景图像,迁移学习后的特征就通过最后的语义分割部分产生最终的结果,即低光照图像对应的语义分割结果;

训练的分割网络的优化过程要分为两步,分别对生成器G和判别器D进行优化;判别器输出的预测概率只用于训练过程中损失函数的计算以达到网络监督的作用,在测试过程中判别器部分将不再进行计算;其中生成误差 L_G 由两个损失函数组成:分割的预测误差 L_{seg} 和转换后特征的判别误差 D_g ;对于误差 D_g ,使用最小二乘损失,其函数表示如公式(2):

$$D_g = |\text{pre}_g - 1| \quad (2)$$

其中, pre_g 表示判别器网络对输入的转换后的特征的概率预测,而 L_{seg} 依然使用交叉熵损失函数,如公式(3)所示:

$$L_{seg} = -\sum_{c=1}^M y_c \log(p_c) \quad (3)$$

其中, M 表示类别的数量, y_c 表示变量的类别判断,类别相同则为1,不同则为0, p_c 表示对于观测样本属于类别 c 的预测概率;完整的生成误差 L_G 的函数表示如公式(4)所示:

$$L_G = L_{seg} + D_g \quad (4)$$

判别误差 L_D 同样也是由两部分组成,分别为正常图像的特征的判别器的预测损失和转换后特征的判别器预测损失,其函数表示如公式(5)所示:

$$L_D = |\text{pre}_n - 1| + \text{pre}_g \quad (5)$$

其中, pre_n 表示正常图像的特征的判别器的预测损失;

将训练数据多线程分批输送到待训练的网络模型中,并按照公式(4)和公式(5)来分别计算网络的生成器误差和判别器误差;最后使用反向传播的梯度下降方法和Adam优化器迭代更新网络参数,直到达到预定迭代次数或者是多轮迭代后训练误差不再发生变化后停止,完成网络的训练;

网络的初始参数要导入ResNet和DeepLabV3网络对应层的参数进行网络的初始化,加速网络收敛;对于生成损失 L_G 和判别损失 L_D ,使用优化器Adam分别进行反向传播的优化,每次循环分别优化一次,使网络不断的收敛。

一种低光照场景的语义分割方法

技术领域

[0001] 本发明属于计算机视觉技术领域,尤其涉及基于深度学习的图像语义分割技术,针对低光照场景这一最常见的非理想场景,获取较高精确度的黑暗场景语义分割结果。

背景技术

[0002] 语义分割是经典的计算机视觉问题,其将图像等种类的原始数据作为输入,并根据原始输入,输出其对应的感兴趣的区域掩模。全像素语义分割以单个像素为分类的基本单位,这与人类的场景理解的感知有很大的相似,对比早期的只关注图像边缘与渐变的计算机视觉问题有很大的优势。语义分割能够将图片中属于同一部分的像素聚集在一起,能够很好的解决场景理解问题,与其他图像任务对比来说,语义分割是不同且领先的技术,这也提高了其研究价值,扩展了应用领域。语义分割常见的应用领域主要包括自动驾驶、智能医疗以及人脸识别的常见领域。对于低光照的场景的语义分割问题,由于场景图片亮度过低,导致场景图像的物体颜色、轮廓和纹理等图像中物体的细节隐藏在黑暗中,对于人以及机器系统对图像的理解和处理都产生了很大的障碍。本发明则主要关注场景理解中的对精度要求最高的语义分割子任务,主要针对非理想环境的黑暗场景,利用深度学习的方法,计算出低光照场景的精确的语义分割结果。

[0003] 语义分割作为一种经典的计算机视觉算法,一直以来受到学者的广泛的关注。在国内学者的共同研究下,越来越多的算法在不断的出现,算法的精确度和完整度也在不断的提高。传统的语义分割算法没有使用神经网络和深度学习的相关算法,主要结合大量的先验特征知识,并结合概率图进行计算,一些算法后期还需要人工参与才能完成最终的语义分割任务。传统语义分割主要分为两大类,基于无监督的分割方法和基于概率图模型分类方法。基于无监督的分割算法一直以来都存在一定的争议,因为其后期需要人工参与才能完成最终分割过程,严格意义上来说是一种非语义分割算法。但无监督分割算法可以当作监督分割算法的一种有效的信息来源,所以一直以来都有学者不断地进行研究。无监督的分割算法尝试去检测一致的区域或者区域的边界,通过各个区域的分类来完成所有像素的分类,常结合聚类算法进行计算。基于概率图的方法则是利用一些分类方法对每个像素进行分类,概率高的代表所属的分类类别。图像分割可以选择使用多种不同的概率图模型,常见的概率图模型有支持向量机、条件随机场、马尔可夫随机场以及随机决策森林等。由于需要对所有像素进行概率计算,因此基于概率图模型的分割算法效率比较低,且需要以下条件假设和先验知识完成概率图模型的生成。

[0004] 随着机器学习与深度学习技术的飞速发展,其活跃在包括语义分割计算机视觉的各个领域,并使得计算机视觉取得飞跃式的发展。深度学习算法能够建立起输入与输出之间的多维映射,对于语义分割来说,它能够建立起输入图像与输出类别间的像素级别的映射,并且整个过程快速、独立、完整,完全不需要人为的参与。深度学习分割方法相对于传统方法能够取得显著效果提升的根本原因是深度学习能够学习到分割问题合适的特征表示,目前,大多数的基于深度学习技术的语义分割算法来源属于一个共同的先驱分割模型FCN。

FCN被认为是予以分割问题上的一个里程碑,因为它展示了如何针对语义分割问题训练端到端的深度卷积网络,能够高效的对任意大小的输入进行密集的语义分割预测。随后,在FCN的基础上,出现了许多的改进的分割网络,Vi jay等人提出了 SegNet网络,其在特征提取的池化操作的时存储池化索引,并利用池化索引进行上采样来扩大特征分辨率,有效的提升了分割的准确率。Ronneberger等人对 FCN的改进主要是通过扩展解码器的容量来完成的,所提出的U-Net由捕获上下文关系的收缩路径和支撑精确定位的扩展路径组成,这种简单的体系结构已经在各种分割方法中流行起来。PSPNet提出了一种适用于语义分割的金字塔网络结构,通过不同尺度的池化层获取更多的上下文信息,并连接在一起改进分割结果。随着扩张卷积的应用和发展,DeepLab系列相继出现。DeepLab使用扩张卷积获取更多的特征信息,并在网络架构上不断的改进,提出ASPP等网络结构来获取更多的图片信息从而提升分割的准确度。

[0005] 近年来基于深度学习技术,出现了很多语义分割方法,但由于数据集缺乏等原因,这些方法的作用场景单一,需要理想的充足亮度和关照的条件,一旦亮度不足,准确度将严重下降。因此,可以看出当前阶段,语义分割方法依然需要进一步提高其鲁棒性。尤其是针对亮度较低等常见非理想场景,需要获得更多的关注这也是本发明的目的所在。

发明内容

[0006] 本发明针对低光照的场景的语义分割问题,由于场景图片亮度过低,导致场景图像的物体颜色、轮廓和纹理等图像中物体的细节隐藏在黑暗中,对于人以及机器系统对图像的理解和处理都产生了很大的障碍。为了解决这些问题,本发明将正常图像的语义分割问题当做源域问题,低光照图像的语义分割问题当做目标域问题,利用迁移学习中的特征迁移的方法,充分发挥正常场景图像的信息充足的优势,将正常场景中的有用信息提取出来,并将这些有用信息与低光照图像的特征信息进行转换的结合,获取更多的有助于语义分割的图像信息,从而训练深度神经网络。基于这一思想,在生成对抗网络的基础之上,利用迁移学习的方法,设计并实现了用于低光照场景直接语义分割的网络模型。利用这一模型,能够有效解决低光照图片的语义分割任务。

[0007] 本发明的技术方案:

[0008] 为了实现直接的端到端的语义分割网络,本发明提出基于迁移学习的语义分割算法。整体思想是利用正常图像的特征图中包含的图像亮度信息对低光照的特征信息进行优化,优化后的特征将包含更全面的正常场景的亮度、颜色等有助于分割的特征信息,在此特征基础上再进行后续的上采样和像素的语义分类。因此本发明利用生成对抗网络的思想,将低光照图像的特征转换的过程看作是生成器网络生成正常图像特征的过程,利用一个判别器网络对转换后的结果进行评价,并利用这个评价的预测概率的损失函数来进一步监督特征转换以及特征提取部分的网络收敛过程。下面对该方法的网络架构以及实施方案进行详细的介绍。

[0009] 一种低光照场景的语义分割方法,步骤如下:

[0010] (一) 基于迁移学习的语义分割网络架构

[0011] 所述基于迁移学习的语义分割网络包括以下四个部分:

[0012] 编码器部分:语义分割网络包含两个编码器,编码器C和编码器S;编码器C用于提

取低光照图像的低光照图像特征信息,编码器C的前半部分由采用了ResNet的网络结构前半部分,包含50个卷积层以及残差和Bottleneck结构,用来减小参数加速收敛,并能够获得包含2048个通道的特征图;随后将获得的 2048个通道的特征图进行实例标准化,并对特征进行正则化,然后通过 1×1 的卷积层将特征通道数缩小到256个,得到低光照图像的特征图 F_c ;编码器S用于提取正常场景图像风格特征的网络部分,其包括不同卷积核大小卷积层,不包括池化层,使用四个步长为2的卷积层,能达到缩小特征维度、减少网络参数的作用。随着卷积的进行,特征的通道数也在不断的增加到256个通道,然后通过一个全局池化层一步缩小特征维度,最后使用一个全连接层得到一个代表正常场景图像风格特征的256维的列向量 F_s 。

[0013] 特征迁移网络部分:黑天场景的特征图和白天场景的特征图在网络的处理下整体上都满足正态分布。使用的特征迁移的方法为了改变低光照图像特征的正态分布的参数:均值和方差。在正常图像的均值和方差的获取上,使用网络学习的方式获取,对于输入的正常图像的特征 F_s ,利用两个多层感知机MLP分别学习两个列向量来分别代表各个特征层每层的均值 s_mean 和方差 s_val 。对于低光照图像部分,输入的特征 F_c ,则直接计算均值 c_mean 和方差 c_val ,然后通过公式(1)进行特征的转换,得到转换后的特征 \hat{F}_c 。

$$[0014] \quad \hat{F}_c = s_val * ((F_c - c_mean)/c_val) + s_mean \quad (1)$$

[0015] 输入的低光照图像特征 F_c 计算完均值 c_mean 和方差 c_val 后经过公式(1)进行两次转换特征转换,将得到的特征与转换前的特征 F_c 相加得到结果的过程称作一次特征适应,共进行了3-5次特征适应的过程,得到最终转换后的低光照特征 F_{c_s} 。

[0016] 判别器部分:判别器的结构是一个二分类的分类网络结构,分类为0代表夜晚场景,分类为1代表正常白天场景。判别器用来判别转换后的特征反编码后的结果属于夜晚场景还是白天场景,使得判别器无法判别特征转换后的特征属于哪种场景,或者使得判别器对转换后的特征评价尽可能的接近于1,说明转换比较成功,其损失函数用于监督网络的训练。判别器网络主要由多个步长为2的卷积层组成,不断缩小维度来获取最终的预测概率。

[0017] 语义分割部分:对于语义分割部分,使用DeepLabv3网络的中提出的金字塔池化ASPP结构。

[0018] (二)数据集采集

[0019] 在数据集方面,本发明在Cityscapes数据集的基础之上利用Photoshop软件和Matlab人工合成低光照的场景数据集,并结合原有的公开数据组成完整的训练数据。

[0020] (三)训练步骤

[0021] 首先需要对编码器C和编码器S分别输入低光照的图像和用于提取特征信息的正常场景的图像。编码器C部分提取低光照场景图像中的特征后输入到特征迁移网络部分;编码器S部分提取正常场景的特征后经过两个多层感知机MLP,将最终得到的特征与编码器C部分输出的特征经过特征迁移部分进行一个特征的融合并迁移。低光照场景图像的特征经过特征迁移部分之后得到一个近似正常场景图像的特征分布值,然后与正常场景图像的特征值同时输入到判别器部分,来判断特征迁移是否成功,迁移成功,图像则从低光照转变为较为正常的场景图像,迁移学习后的特征就通过最后的语义分割部分产生最终的结果,即低光照图像对应的语义分割结果。

[0022] 训练的分割网络的优化过程要分为两步,分别对生成器G和判别器D进行优化。判别器输出的预测概率只用于训练过程中损失函数的计算以达到网络监督的作用,在测试过程中判别器部分将不再进行计算。其中生成误差 L_G 由两个损失函数组成:分割的预测误差 L_{seg} 和转换后特征的判别误差 D_g 。对于误差 D_g ,使用最小二乘损失,其函数表示如公式(2):

$$[0023] \quad D_g = |\text{pre}_g - 1| \quad (2)$$

[0024] 其中, pre_g 表示判别器网络对输入的转换后的特征的概率预测,而 L_{seg} 本发明依然使用交叉熵损失函数,如公式(3)所示:

$$[0025] \quad L_{seg} = -\sum_{c=1}^M y_c \log(p_c) \quad (3)$$

[0026] 其中,M表示类别的数量, y_c 表示变量的类别判断(0或1),如果类别相同则为1,不同则为0, p_c 表示对于观测样本属于类别c的预测概率。因此完整的生成误差 L_G 的函数表示如公式(4)所示:

$$[0027] \quad L_G = L_{seg} + D_g \quad (4)$$

[0028] 判别误差 L_D 同样也是由两部分组成,分别为正常图像的特征的判别器的预测损失和转换后特征的判别器预测损失,其函数表示如公式(5)所示:

$$[0029] \quad L_D = |\text{pre}_n - 1| + \text{pre}_g \quad (5)$$

[0030] 其中, pre_n 表示正常图像的特征的判别器的预测损失。

[0031] 将训练数据多线程分批输送到待训练的网络模型中,并按照公式(4)和公式(5)来分别计算网络的生成器误差和判别器误差。最后使用反向传播的梯度下降方法和Adam优化器迭代更新网络参数,直到达到预定迭代次数或者是多轮迭代后训练误差不再发生变化后停止,完成网络的训练。

[0032] 网络的初始参数要导入ResNet和DeepLabV3网络对应层的参数进行网络的初始化,加速网络收敛。对于生成损失 L_G 和判别损失 L_D ,使用优化器Adam分别进行反向传播的优化,每次循环分别优化一次,使网络不断的收敛。

[0033] 本发明的有益效果:本发明提出了一种基于迁移学习的低光照场景语义分割模型。该模型利用迁移学习的思想,通过迁移学习中的特征迁移的方法,结合正常光照场景图片提取出的包含亮度的有用信息的图片特征来对低光照图像的特征进行修改,目的是使得低光照图片的特征包含更多的场景信息,同时提高特征中的亮度表示。为了给网络提供更多的监督信息,促进网络特征迁移和分割的参数的收敛,获得更好的实验效果,本发明使用生成对抗网络的思想,利用判别器网络的预测的损失函数进一步监督网络的训练,加速网络收敛。实验结果表明,本发明方法能够取得更优的实验效果,说明了方法的可行性和有效性。

附图说明

[0034] 图1为方法流程架构图。低光照图片的语义分割网络是一个生成对抗网络结构,生成器用于图片特征的转换与分割,判别器部分用于对特征转换后的特征进行种类的判断,提高网络的收敛效果。

具体实施方式

[0035] (一)网络训练

[0036] 首先导入ResNet和DeepLabV3网络对应层的参数进行一个网络的初始化,加速网络之后的训练收敛,也就是编码器C、编码器S和最后的语义分割部分的预训练。将采集好的数据集进行一个随机分组,使其每一组有一张低光照场景图像和正常场景的图像,分别输入到两个编码器进行相应的特征提取。此过程是在导入ResNet预训练模型后的再训练过程。编码器C部分提取低光照场景图像中的特征后输入到特征迁移网络部分;编码器S部分提取正常场景的特征后经过两个多层感知机(MLP),将最终得到的特征与编码器C部分输出的特征经过特征迁移部分进行一个特征的融合并迁移。经过特征迁移网络部分后得到相应的特征用来对判别器部分进行训练,对于生成损失 L_G 和判别损失 L_D ,使用优化器Adam分别进行反向传播的优化,每次循环分别优化一次,使网络不断的收敛。当判别器部分训练到感知到输入的特征90%概率判别为正常场景时可以停止训练判别器,说明特征迁移网络部分以达到特征迁移目的。到此训练编码器、特征迁移网络和判别器全部训练完毕。

[0037] 在网络训练时,本发明使用的是Ubuntu操作系统的服务器,配备Intel(R) Xeon(R) CPU E5-2650 V4的2.2GHz的CPU处理器,系统内存为128GB,使用英伟达Titan X显卡,显存为12GB。代码使用Python3编程语言,在Tensorflow深度学习框架下进行实验模型的训练及测试。在训练时需要加入想要将此发明应用到的特定低光照场景的情况下的数据集,以此来提升网络对特定场景的分割效果。

[0038] (二)测试与应用

[0039] 本发明主要针对低光照场景的语义分割问题,对输入的低光照场景图像的分辨率大小没有限制,不需要对输入图像进行预处理便可直接将图片输入到网络中。由于正常场景图像只需要进行图像特征的提取一遍获得亮度等特征信息,因此输入的低光照图像和正常场景图像不需要在图像内容上进行精确的匹配,可以为任意的两张不同场景的图像。

[0040] 在测试和应用时,首先创建目标场景图像(低光照)文件夹和正常场景图像文件夹,修改代码中的文件路径参数,将所有训练参数调为False,同时因为判别器在测试和应用时无需进行判断,所以将判别器部分功能关闭以减少网络资源占用和提升运行速度。最后将语义分割部分与特征迁移部分连接,以此得到一个最终低光照场景图像的语义分割效果。除了语义分割图像的视觉效果可供参考外,也可以查看网络运行后保存的MIoU指标,本发明一般在0.3~0.4左右。

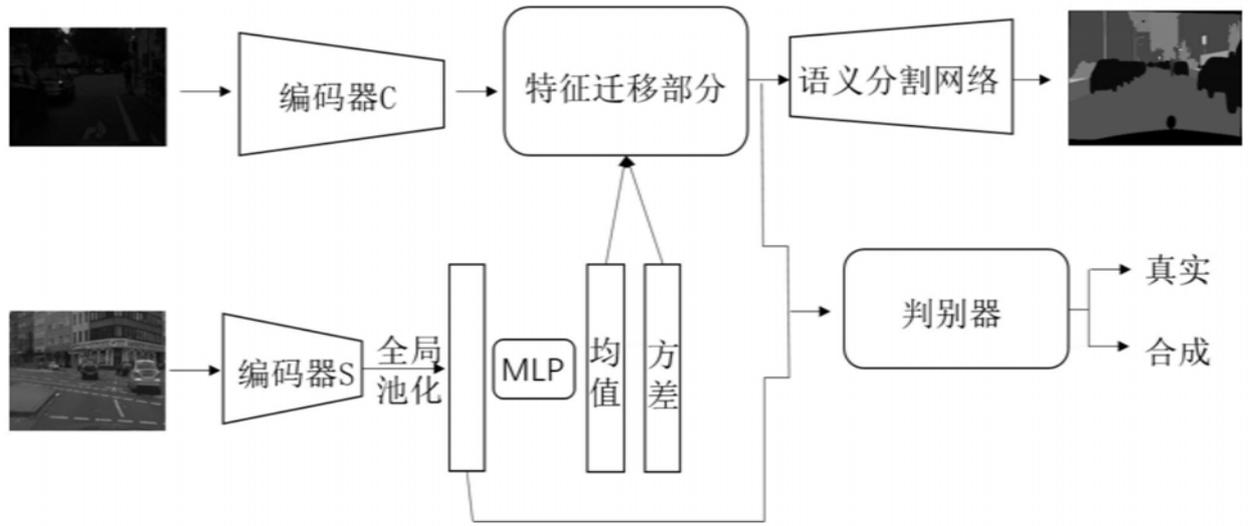


图1