(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 113538457 A
(43)申请公布日 2021.10.22

(21)申请号 202110718738.1

(22)申请日 2021.06.28

(71)申请人 杭州电子科技大学
地址 310018 浙江省杭州市下沙高教园区2号大街

(72)发明人 李平 陈俊杰 王然 徐向华

(74)专利代理机构 杭州君度专利代理事务所(特殊普通合伙) 33240
代理人 陈炜

(51)Int.Cl.
G06T 7/11(2017.01)
G06K 9/00(2006.01)
G06K 9/62(2006.01)
G06N 3/04(2006.01)
G06N 3/08(2006.01)

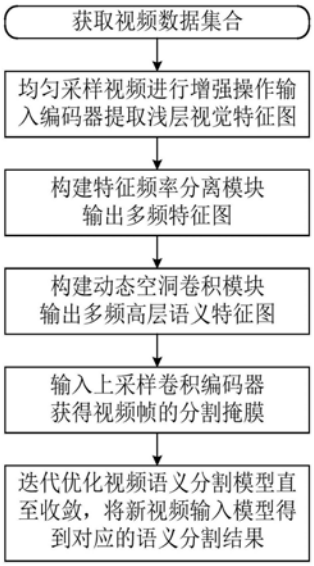
权利要求书3页 说明书8页 附图1页

(54)发明名称

利用多频动态空洞卷积的视频语义分割方法

(57)摘要

本发明公开了利用多频动态空洞卷积的视频语义分割方法。本发明方法首先对视频数据的采样帧图像进行增强处理，并通过编码器提取浅层视觉特征图；然后构建特征频率分离模块获得视频帧对应的多频特征图，并将其输入动态空洞卷积模块，得到对应的多频高层语义特征图，再通过上采样卷积编码器获得视频帧的分割掩膜；利用随机梯度下降算法迭代训练模型直至收敛，将新视频输入模型得到语义分割结果。本发明方法对视频帧的特征图按不同频率分离以刻画不同视觉区域变化，能够减少低频视觉空间冗余信息、降低计算复杂度，通过动态空洞卷积自适应地扩大多频特征图的感受野，提升对视频不同语义类的判别能力，从而获得更优视频语义分割结果。



CN 113538457 A

1.利用多频动态空洞卷积的视频语义分割方法,其特征在于,该方法首先获取视频数据集合,然后进行如下操作:

步骤(1)对视频采样获得视频帧,并进行增强操作,然后输入至编码器,即深度卷积神经网络,获得对应的浅层视觉特征图;

步骤(2)构建特征频率分离模块,输入为浅层视觉特征图,输出多频特征图;

步骤(3)构建动态空洞卷积模块,输入为多频特征图,输出多频高层语义特征图;

步骤(4)将多频高层语义特征图输入解码器即上采样卷积模块,获得视频帧的分割掩膜;

步骤(5)迭代训练由编码器、特征频率分离模块、动态空洞卷积模块、解码器组成的视频语义分割模型直至收敛,然后将新视频输入至该模型得到对应的语义分割结果。

2.如权利要求1所述的利用多频动态空洞卷积的视频语义分割方法,其特征在于,步骤(1)具体是:

(1-1)对单个视频进行均匀采样获得视频帧,采样率为10～15帧/秒,并对其进行增强操作得到数量为N的视频帧序列I,记为$\mathbf{I}=\left\{\mathbf{I}_i\middle|\mathbf{I}_i\in\mathbb{R}^{3\times H\times W},i=1,\ldots,N\right\}$,其中$I_i$表示第i个视频帧,$\mathbb{R}$表示实数域,3表示RGB通道数量,H表示视频帧高度,W表示视频帧宽度;

(1-2)利用大型图像库ImageNet上预训练的卷积神经网络ResNet对视频帧序列I依次提取浅层视觉特征图$\mathbf{f}_i\in\mathbb{R}^{C_f\times H_f\times W_f}$,$C_f$表示特征图的通道数,$H_f$表示特征图高度,$W_f$表示特征图宽度;ResNet具有多个由卷积层组成的模块,f_i为第i个视频帧经过RestNet前三个由多个卷积层组成的模块得到的特征图。

3.如权利要求2所述的利用多频动态空洞卷积的视频语义分割方法,其特征在于,步骤(2)具体是:

(2-1)构建特征频率分离模块,利用图像具有频率可分离的特点,对浅层视觉特征图进行三次高低频特征分离操作获得多频特征图;其中,高频特征刻画特征图的轮廓区域,低频特征刻画特征图的平面区域,中频特征刻画特征图的内容区域;

(2-2)高低频特征分离的具体操作如下:

首先对浅层视觉特征图f_i做快速傅里叶变换,将空域信号转换为频域信号得到f_i的频谱图$\widehat{\mathbf{f}}_i\in\mathbb{R}^{C_f\times H_f\times W_f}$,将$\widehat{\mathbf{f}}_i$中低频信号部分平移到中间得到平移频谱图$\breve{\mathbf{f}}_i\in\mathbb{R}^{C_f\times H_f\times W_f}$,确定$\breve{\mathbf{f}}_i$的中心位置向量(P,Q);其中,$\breve{\mathbf{f}}_i$通道中心点的横坐标值组成的向量$\mathbf{P}=\left\{\mathbf{P}_r\middle|\mathbf{P}_r\in\mathbb{R},r=1,\ldots,C_f\right\}$,纵坐标值组成的向量$\mathbf{Q}=\left\{\mathbf{Q}_r\middle|\mathbf{Q}_r\in\mathbb{R},r=1,\ldots,C_f\right\}$,下标r表示$\breve{\mathbf{f}}_i$的通道索引;

然后将$\breve{\mathbf{f}}_i$中每个元素与低频转移函数$H_l(u_{r,a},,v_{r,b})$作乘法运算得到低频平移频谱图$\breve{\mathbf{f}}_i^l\in\mathbb{R}^{C_f\times H_f\times W_f}$;高斯低通滤波器的转移函数$\mathrm{H}_l\left(u_{r,a},v_{r,b}\right)=\exp\left(-D^2\left(u_{r,a},v_{r,b}\right)/2D_0^2\right)$,l表示低频信号,a表示像素点横轴坐标值,b表示像素点纵轴坐标值,{0≤a≤H_f,0≤b≤W_f},exp(•)表示指数函数,D_0是设定的标准差;其中,$D\left(u_{r,a},v_{r,b}\right)=\left(u_{r,a}^2,v_{r,b}^2\right)^{\frac{1}{2}}$,表示$\breve{\mathbf{f}}_i$中第r个通道

像素点(a,b)距离坐标点(P_r,Q_r)的欧式距离,$u_{r,a}$是$\breve{\mathbf{f}}_i$中第r个通道频谱位置(a,0)距离P_r的欧式距离,$v_{r,b}$是$\breve{\mathbf{f}}_i$中第r个通道频谱位置(0,b)距离Q_r的欧式距离;

同理,将$\breve{\mathbf{f}}_i$中每个元素与高频转移函数$H_h(u_{r,a},,v_{r,b})$作乘法运算得到高频平移频谱图$\breve{\mathbf{f}}_i^h \in \mathbb{R}^{C_f \times H_f \times W_f}$,其中h表示高频信号,$H_h\left(u_{r,a},v_{r,b}\right)=1-\exp(-D^2\left(u_{r,a},v_{r,b}\right)/2D_0^2)$;

分别将频谱图$\breve{\mathbf{f}}_i^l$与$\breve{\mathbf{f}}_i^h$中的低频信号从中间平移回到原始位置,得到低频频谱图$\widehat{\mathbf{f}}_i^l \in \mathbb{R}^{C_f \times H_f \times W_f}$和高频频谱图$\widehat{\mathbf{f}}_i^h \in \mathbb{R}^{C_f \times H_f \times W_f}$;

最后将$\widehat{\mathbf{f}}_i^l$和$\widehat{\mathbf{f}}_i^h$分别做快速傅里叶逆变换将频域信号转换为空域信号,得到弱低频特征图$\mathbf{f}_i^l \in \mathbb{R}^{C_f \times H_f \times W_f}$和弱高频特征图$\mathbf{f}_i^h \in \mathbb{R}^{C_f \times H_f \times W_f}$;

(2-3)按照(2-2),对弱高频特征图$\mathbf{f}_i^h$进行第二次高低频特征分离操作,得到强高频特征图$\mathbf{f}_i^{hh} \in \mathbb{R}^{C_f \times H_f \times W_f}$和中高频特征图$\mathbf{f}_i^{hl} \in \mathbb{R}^{C_f \times H_f \times W_f}$,hh表示特征图经过两次高频信号过滤,hl表示特征图先经过一次高频信号过滤,再经过一次低频信号过滤;

按照(2-2),对弱低频特征图$\mathbf{f}_i^l$进行第二次高低频特征分离操作,得到强低频特征图$\mathbf{f}_i^{ll} \in \mathbb{R}^{C_f \times H_f \times W_f}$和中低频特征图$\mathbf{f}_i^{lh} \in \mathbb{R}^{C_f \times H_f \times W_f}$,ll表示特征图经过两次低频信号过滤,lh表示特征图先经过一次低频信号过滤,再经过一次高频信号过滤;

(2-4)将中高频特征图$\mathbf{f}_i^{hl}$和中低频特征图$\mathbf{f}_i^{lh}$进行一次拼接,经过一次大小为1×1的卷积操作得到压缩后的特征图,再经过步长为2的最大池化操作进行下采样得到中频特征图$\hat{\mathbf{f}}_i^m \in \mathbb{R}^{C_f^m \times \frac{H_f}{2} \times \frac{W_f}{2}}$,其中m表示中频信号,$C_f^m$为中频特征图的通道维度;

(2-5)将强低频特征图$\mathbf{f}_i^{ll}$经过一次大小为1×1卷积操作得到压缩后的特征图,再经过步长为4的最大池化操作进行下采样得到低频特征图$\hat{\mathbf{f}}_i^l \in \mathbb{R}^{C_f^l \times \frac{H_f}{4} \times \frac{W_f}{4}}$;将强高频特征图$\mathbf{f}_i^{hh}$经过一次大小为1×1卷积操作得到压缩后的高频特征图$\hat{\mathbf{f}}_i^h \in \mathbb{R}^{C_f^h \times H_f \times W_f}$;其中,$C_f^m + C_f^h + C_f^l = C_f$,$C_f^h$和$C_f^l$分别表示高频特征图和低频特征图的通道维度。

4.如权利要求3所述的利用多频动态空洞卷积的视频语义分割方法,其特征在于,步骤(3)具体是:

(3-1)构建由一个权重计算器、K个并行的空洞卷积核组成的动态空洞卷积模块,将多频特征图分别输入到动态空洞卷积模块,得到多频高层语义特征图,包括低频高层语义特征图、中频高层语义特征图和高频高层语义特征图;

(3-2)动态空洞卷积的具体操作如下:将低频特征图$\hat{\mathbf{f}}_i^l$输入到权重计算器得到输出K个权重$\{w_t \in \mathbb{R} \mid t=1,...,K\}$,$w_t$表示第t个空洞卷积的权重,0≤$w_t$<1,$\sum_t^K w_t = 1$;权重计算器由一次全局平均池化操作、一个全连接层、一个Relu函数、一个全连接层、一个Softmax函数组

成;K个并行的空洞卷积核$\{\mathbf{K}_t \in \mathbb{R}^{C_f^l \times 3 \times 3} \mid t=1,...,K\}$,$K_t$表示第t个空洞率为2的3×3空洞卷积;$K_t$分别与对应的权重$w_t$做点乘运算,再将K个并行的空洞卷积相加得到集成空洞卷积核$\widehat{\mathbf{K}} \in \mathbb{R}^{C_f^l \times 3 \times 3}$;低频特征图$\hat{\mathbf{f}}_i^l$再与综合空洞卷积核$\widehat{\mathbf{K}}$进行卷积操作得到低频高层语义特征图$\hat{\mathbf{t}}_i^l \in \mathbb{R}^{2C_f^l \times \frac{H_f}{8} \times \frac{W_f}{8}}$,$2C_f^l$表示通道数$C_f^l$的两倍;

(3-3)对动态空洞卷积模块串行叠加,第一个动态空洞卷积模块的输出作为第二个动态空洞卷积模块的输入;按照(3-2),中频特征图$\hat{\mathbf{f}}_i^m$经过两个串行的动态空洞卷积模块得到中频高层语义特征图$\hat{\mathbf{t}}_i^m \in \mathbb{R}^{4C_f^m \times \frac{H_f}{8} \times \frac{W_f}{8}}$,$4C_f^m$表示通道数$C_f^m$的四倍;高频特征图$\hat{\mathbf{f}}_i^h$经过四个串行的动态空洞卷积模块得到高频高层语义特征图$\hat{\mathbf{t}}_i^h \in \mathbb{R}^{8C_f^h \times \frac{H_f}{8} \times \frac{W_f}{8}}$,$8C_f^h$表示通道数$C_f^h$的八倍。

5.如权利要求4所述的利用多频动态空洞卷积的视频语义分割方法,其特征在于,步骤(4)具体是:

(4-1)构建由三个转置卷积层组成的解码器,转置卷积即卷积的逆向过程,通过与输入的小尺寸特征图进行卷积操作得到大尺寸特征图;

(4-2)将低频高层语义特征图$\hat{\mathbf{t}}_i^l$、中频高层语义特征图$\hat{\mathbf{t}}_i^m$和高频高层语义特征图$\hat{\mathbf{t}}_i^h$进行通道维度上的拼接得到集成高层语义特征图$\mathbf{t}_i=[\hat{\mathbf{t}}_i^l;\hat{\mathbf{t}}_i^m;\hat{\mathbf{t}}_i^h] \in \mathbb{R}^{\left(2C_f^l+4C_f^m+8C_f^h\right) \times \frac{H_f}{8} \times \frac{W_f}{8}}$;

(4-3)将集成语义特征图t_i输入解码器得到分割掩膜$\hat{\mathbf{y}}_i \in \mathbb{R}^{C \times H \times W}$,C表示语义类别总数,视频帧中每个像素对应的类别为所有类别中概率最大的类别。

6.如权利要求5所述的利用多频动态空洞卷积的视频语义分割方法,其特征在于,步骤(5)具体是:

(5-1)建立由编码器、特征频率分离模块、动态空洞卷积模块、解码器组成的视频语义分割模型;

(5-2)将视频帧序列依次输入到语义分割模型得到分割掩膜$\hat{\mathbf{y}}_i \in \mathbb{R}^{C \times H \times W}$, $i=1,...,N$,根据交叉熵损失,通过梯度反向传播方法调整模型参数,迭代优化模型直至收敛;

(5-3)将新视频的每一帧输入到已训练好的模型中,依据(5-2)依次输出相应的分割结果$\mathbf{y} \in \mathbb{R}^{1 \times H \times W}$;其中,第一个维度表示语义类别。

利用多频动态空洞卷积的视频语义分割方法

技术领域

[0001] 本发明属于计算机视觉技术领域，尤其是视频处理中的语义分割领域，涉及一种利用多频动态空洞卷积的视频语义分割方法。

背景技术

[0002] 随着各类车辆的与日俱增，驾驶安全成为政府和民众非常关心的方面。一般来说，连续驾驶较长时间会使人疲劳注意力分散，同时大型车辆的驾驶员容易存在视觉盲区，给驾驶安全带来极大隐患。近年来，自动驾驶技术引起业界对自动驾驶技术的浓厚兴趣，越来越多的研究力量被投入到这一领域。高效的视觉理解能为自动驾驶的安全提供保障，视频语义分割是其核心技术之一。视频语义分割旨在对存在时序关联的视频帧进行像素级别的类别标记，得到与原始视频帧同等尺寸的逐像素类别掩膜矩阵，可广泛应用在机器视觉、视频监控、无人机侦察、自动驾驶等领域。例如，在自动驾驶环境中，对车辆视觉场景中的道路、行人或其他车辆等物体进行像素级分割，能够获得比边界框更为精确的物体区域信息，从而为自动驾驶系统提供更为准确的视觉感知内容，有利于规避行人、车辆等障碍物并确保司乘安全。目前，视频语义分割领域的主要挑战包括模型的计算复杂度高、处理高分辨率视频帧耗时长、模型难以部署在实时环境中。

[0003] 传统语义分割方法主要分为阈值、边缘、超像素聚类等几类。其中，阈值分割方法将图像每个像素点的灰度值与阈值比较，灰度值大于阈值的像素被判断成前景，其他为背景，但只适用灰度图像；边缘分割方法先对图像进行边缘检测，同一边缘内的像素代表同一物体，缺点是分割精度受限于边缘检测算法；超像素聚类方法将近似的超像素块聚合以刻画相同物体，缺点是超像素的形成受限于像素的颜色和像素区域的纹理，且易将同一物体的不同部分分成多个超像素，导致分割错误。近年来，深层神经网络由于其强大的特征提取能力而广受欢迎，典型的方法均利用卷积神经网络作为编码器提取视频帧的抽象语义信息，通过解码器的逐层上采样操作获得语义分割掩膜。然而，卷积层仅能提取帧图像的局部语义信息，难以刻画全局场景特征。为此，空间金字塔池化技术被用于语义分割，其显著特点是：对从编码器获取的特征图做多次并行池化操作得到不同大小的压缩特征图，以捕获多个尺寸感受野的全局场景特征，再经过上采样恢复成与初始特征图大小相同的特征图并与其拼接得到总体特征图，最后经解码器得到语义分割掩膜，据此获得视频语义分割结果。

[0004] 现有的语义分割方法仍然存在许多缺点：1)空间金字塔池化技术同时考虑了局部和全局的时空结构信息使得分割结果更加可靠，但是对高分辨率的特征图使用最大平均池化操作会造成容错性不佳、泛化能力差、计算复杂度高等不足；2)利用注意力机制虽然加强了特征图之间的长期语义依赖关系，但是模型臃肿、内存占用多，不利于模型的实时部署；3)Transformer编码器，作为特征抽取器被广泛用于自然语言处理，以二维图像的一维嵌入特征表示序列为输入，利用自注意力机制、多层感知机堆叠捕获视频帧之间的长期依赖关系，但是模型缺乏权值共享导致参数量巨大，且自注意力的计算复杂度高使得实时性难以保障。同时，大多数分割方法的精度和实时性无法做到有效平衡，导致不能有效地满足实际

分割任务的需求。因此，针对分割模型的计算复杂度高、泛化能力差等问题，迫切需要一种既能保障分割模型的实时性又能达到较高语义分割精度的方法。

发明内容

[0005] 本发明的目的就是针对现有技术的不足，提供一种利用多频动态空洞卷积的视频语义分割方法，通过傅里叶变换对特征图进行多种频率分离，多频特征图能够刻画不同视觉区域的不同灰度值变化，以减少低频视觉空间冗余信息并降低计算复杂度；同时设计动态空洞卷积自适应扩大多频特征图的感受野，从全局和局部角度提升模型对视频不同语义类的判别能力，从而提高视频语义分割精度。

[0006] 本发明方法首先获取视频数据集合，然后进行如下操作：

[0007] 步骤(1)对视频采样获得视频帧，并进行增强操作，然后输入至编码器，即深度卷积神经网络，获得对应的浅层视觉特征图；

[0008] 步骤(2)构建特征频率分离模块，输入为浅层视觉特征图，输出多频特征图；

[0009] 步骤(3)构建动态空洞卷积模块，输入为多频特征图，输出多频高层语义特征图；

[0010] 步骤(4)将多频高层语义特征图输入解码器即上采样卷积模块，获得视频帧的分割掩膜；

[0011] 步骤(5)迭代训练由编码器、特征频率分离模块、动态空洞卷积模块、解码器组成的视频语义分割模型直至收敛，然后将新视频输入至该模型得到对应的语义分割结果。

[0012] 进一步，步骤(1)具体是：

[0013] (1-1)对单个视频进行均匀采样获得视频帧，采样率为10～15帧/秒，并对其进行增强操作得到数量为N的视频帧序列I，记为$\mathbf{I}=\left\{\mathbf{I}_i\middle|\mathbf{I}_i\in\mathbb{R}^{3\times H\times W},i=1,\ldots,N\right\}$，其中$I_i$表示第i个视频帧，$\mathbb{R}$表示实数域，3表示RGB通道数量，H表示视频帧高度，W表示视频帧宽度；

[0014] (1-2)利用大型图像库ImageNet上预训练的卷积神经网络ResNet对视频帧序列I依次提取浅层视觉特征图$\mathbf{f}_i\in\mathbb{R}^{C_f\times H_f\times W_f}$，$C_f$表示特征图的通道数，$H_f$表示特征图高度，$W_f$表示特征图宽度；ResNet具有多个由卷积层组成的模块，f_i为第i个视频帧经过RestNet前三个由多个卷积层组成的模块得到的特征图。

[0015] 更进一步，步骤(2)具体是：

[0016] (2-1)构建特征频率分离模块，利用图像具有频率可分离的特点，对浅层视觉特征图进行三次高低频特征分离操作获得多频特征图；其中，高频特征刻画特征图的轮廓区域，低频特征刻画特征图的平面区域，中频特征刻画特征图的内容区域；

[0017] (2-2)高低频特征分离的具体操作如下：

[0018] 首先对浅层视觉特征图f_i做快速傅里叶变换，将空域信号转换为频域信号得到f_i的频谱图$\widehat{\mathbf{f}}_i\in\mathbb{R}^{C_f\times H_f\times W_f}$，将$\widehat{\mathbf{f}}_i$中低频信号部分平移到中间得到平移频谱图$\breve{\mathbf{f}}_i\in\mathbb{R}^{C_f\times H_f\times W_f}$，确定$\breve{\mathbf{f}}_i$的中心位置向量(P，Q)；其中，$\breve{\mathbf{f}}_i$通道中心点的横坐标值组成的向量$\mathbf{P}=\left\{\mathbf{P}_r\middle|\mathbf{P}_r\in\mathbb{R},r=1,\ldots,C_f\right\}$，纵坐标值组成的向量$\mathbf{Q}=\left\{\mathbf{Q}_r\middle|\mathbf{Q}_r\in\mathbb{R},r=1,\ldots,C_f\right\}$，下标r表示$\breve{\mathbf{f}}_i$的通道索引；

[0019] 然后将$\breve{\mathbf{f}}_i$中每个元素与低频转移函数$H_l(u_{r,a},v_{r,b})$作乘法运算得到低频平移频谱图$\breve{\mathbf{f}}_i^l \in \mathbb{R}^{C_f \times H_f \times W_f}$；高斯低通滤波器的转移函数$H_l\left(u_{r,a},v_{r,b}\right)=\exp\left(-D^2\left(u_{r,a},v_{r,b}\right)/2D_0^2\right)$，l表示低频信号，a表示像素点横轴坐标值，b表示像素点纵轴坐标值，{0≤a≤H_f，0≤b≤W_f}，exp(•)表示指数函数，D_0是设定的标准差；其中，$D\left(u_{r,a},v_{r,b}\right)=\left(u_{r,a}^2,v_{r,b}^2\right)^{\frac{1}{2}}$，表示$\breve{\mathbf{f}}_i$中第r个通道像素点(a,b)距离坐标点($P_r$,$Q_r$)的欧式距离，$u_{r,a}$是$\breve{\mathbf{f}}_i$中第r个通道频谱位置(a,0)距离$P_r$的欧式距离，$v_{r,b}$是$\breve{\mathbf{f}}_i$中第r个通道频谱位置(0,b)距离$Q_r$的欧式距离；

[0020] 同理，将$\breve{\mathbf{f}}_i$中每个元素与高频转移函数$H_h(u_{r,a},v_{r,b})$作乘法运算得到高频平移频谱图$\breve{\mathbf{f}}_i^h \in \mathbb{R}^{C_f \times H_f \times W_f}$，其中h表示高频信号，$H_h\left(u_{r,a},v_{r,b}\right)=1-\exp(-D^2\left(u_{r,a},v_{r,b}\right)/2D_0^2)$；

[0021] 分别将频谱图$\breve{\mathbf{f}}_i^l$与$\breve{\mathbf{f}}_i^h$中的低频信号从中间平移回到原始位置，得到低频频谱图$\widehat{\mathbf{f}}_i^l \in \mathbb{R}^{C_f \times H_f \times W_f}$和高频频谱图$\widehat{\mathbf{f}}_i^h \in \mathbb{R}^{C_f \times H_f \times W_f}$；

[0022] 最后将$\widehat{\mathbf{f}}_i^l$和$\widehat{\mathbf{f}}_i^h$分别做快速傅里叶逆变换将频域信号转换为空域信号，得到弱低频特征图$\mathbf{f}_i^l \in \mathbb{R}^{C_f \times H_f \times W_f}$和弱高频特征图$\mathbf{f}_i^h \in \mathbb{R}^{C_f \times H_f \times W_f}$；

[0023] (2-3)按照(2-2)，对弱高频特征图$\mathbf{f}_i^h$进行第二次高低频特征分离操作，得到强高频特征图$\mathbf{f}_i^{hh} \in \mathbb{R}^{C_f \times H_f \times W_f}$和中高频特征图$\mathbf{f}_i^{hl} \in \mathbb{R}^{C_f \times H_f \times W_f}$，hh表示特征图经过两次高频信号过滤，hl表示特征图先经过一次高频信号过滤，再经过一次低频信号过滤；

[0024] 按照(2-2)，对弱低频特征图$\mathbf{f}_i^l$进行第二次高低频特征分离操作，得到强低频特征图$\mathbf{f}_i^{ll} \in \mathbb{R}^{C_f \times H_f \times W_f}$和中低频特征图$\mathbf{f}_i^{lh} \in \mathbb{R}^{C_f \times H_f \times W_f}$，ll表示特征图经过两次低频信号过滤，lh表示特征图先经过一次低频信号过滤，再经过一次高频信号过滤；

[0025] (2-4)将中高频特征图$\mathbf{f}_i^{hl}$和中低频特征图$\mathbf{f}_i^{lh}$进行一次拼接，经过一次大小为1×1的卷积操作得到压缩后的特征图，再经过步长为2的最大池化操作进行下采样得到中频特征图$\hat{\mathbf{f}}_i^m \in \mathbb{R}^{C_f^m \times \frac{H_f}{2} \times \frac{W_f}{2}}$，其中m表示中频信号，$C_f^m$为中频特征图的通道维度；

[0026] (2-5)将强低频特征图$\mathbf{f}_i^{ll}$经过一次大小为1×1卷积操作得到压缩后的特征图，再经过步长为4的最大池化操作进行下采样得到低频特征图$\hat{\mathbf{f}}_i^l \in \mathbb{R}^{C_f^l \times \frac{H_f}{4} \times \frac{W_f}{4}}$；将强高频特征图$\mathbf{f}_i^{hh}$经过一次大小为1×1卷积操作得到压缩后的高频特征图$\hat{\mathbf{f}}_i^h \in \mathbb{R}^{C_f^h \times H_f \times W_f}$；其中，$C_f^m+C_f^h+C_f^l=C_f$，$C_f^h$和$C_f^l$分别表示高频特征图和低频特征图的通道维度。

[0027] 再进一步，步骤(3)具体是：

[0028] (3-1)构建由一个权重计算器、K个并行的空洞卷积核组成的动态空洞卷积模块，将多频特征图分别输入到动态空洞卷积模块，得到多频高层语义特征图，包括低频高层语

义特征图、中频高层语义特征图和高频高层语义特征图；

[0029] (3-2)动态空洞卷积的具体操作如下：将低频特征图$\hat{\mathbf{f}}_i^l$输入到权重计算器得到输出K个权重$\{w_t \in \mathbb{R} \mid t=1,...,K\}$，$w_t$表示第t个空洞卷积的权重，$0 \leqslant w_t < 1$，$\sum_t^K w_t = 1$；权重计算器由一次全局平均池化操作、一个全连接层、一个Relu函数、一个全连接层、一个Softmax函数组成；K个并行的空洞卷积核$\{\mathbf{K}_t \in \mathbb{R}^{C_f^l \times 3 \times 3} \mid t=1,...,K\}$，$K_t$表示第t个空洞率为2的3×3空洞卷积；$K_t$分别与对应的权重$w_t$做点乘运算，再将K个并行的空洞卷积相加得到集成空洞卷积核$\widehat{\mathbf{K}} \in \mathbb{R}^{C_f^l \times 3 \times 3}$，用于利用多个并行空洞卷积的参数以捕获不同感受野；低频特征图$\hat{\mathbf{f}}_i^l$再与综合空洞卷积核$\widehat{\mathbf{K}}$进行卷积操作得到低频高层语义特征图$\hat{\mathbf{t}}_i^l \in \mathbb{R}^{2C_f^l \times \frac{H_f}{8} \times \frac{W_f}{8}}$，$2C_f^l$表示通道数$C_f^l$的两倍；(3-3)对动态空洞卷积模块串行叠加，第一个动态空洞卷积模块的输出作为第二个动态空洞卷积模块的输入；按照(3-2)，中频特征图$\hat{\mathbf{f}}_i^m$经过两个串行的动态空洞卷积模块得到中频高层语义特征图$\hat{\mathbf{t}}_i^m \in \mathbb{R}^{4C_f^m \times \frac{H_f}{8} \times \frac{W_f}{8}}$，$4C_f^m$表示通道数$C_f^m$的四倍；同理，高频特征图$\hat{\mathbf{f}}_i^h$经过四个串行的动态空洞卷积模块得到高频高层语义特征图$\hat{\mathbf{t}}_i^h \in \mathbb{R}^{8C_f^h \times \frac{H_f}{8} \times \frac{W_f}{8}}$，$8C_f^h$表示通道数$C_f^h$的八倍。

[0030] 又进一步，步骤(4)具体是：

[0031] (4-1)构建由三个转置卷积层组成的解码器，转置卷积即卷积的逆向过程，通过与输入的小尺寸特征图进行卷积操作得到大尺寸特征图；

[0032] (4-2)将低频高层语义特征图$\hat{\mathbf{t}}_i^l$、中频高层语义特征图$\hat{\mathbf{t}}_i^m$和高频高层语义特征图$\hat{\mathbf{t}}_i^h$进行通道维度上的拼接得到集成高层语义特征图$\mathbf{t}_i=[\hat{\mathbf{t}}_i^l;\hat{\mathbf{t}}_i^m;\hat{\mathbf{t}}_i^h] \in \mathbb{R}^{\left(2C_f^l+4C_f^m+8C_f^h\right) \times \frac{H_f}{8} \times \frac{W_f}{8}}$；

[0033] (4-3)将集成语义特征图t_i输入解码器得到分割掩膜$\hat{\mathbf{y}}_i \in \mathbb{R}^{C \times H \times W}$，C表示语义类别总数，视频帧中每个像素对应的类别为所有类别中概率最大的类别。

[0034] 还进一步，步骤(5)具体是：

[0035] (5-1)建立由编码器、特征频率分离模块、动态空洞卷积模块、解码器组成的视频语义分割模型；

[0036] (5-2)将视频帧序列依次输入到语义分割模型得到分割掩膜$\hat{\mathbf{y}}_i \in \mathbb{R}^{C \times H \times W}$，i＝1，…，N，根据交叉熵损失，通过梯度反向传播方法调整模型参数，迭代优化模型直至收敛；

[0037] (5-3)将新视频的每一帧输入到已训练好的模型中，依据(5-2)依次输出相应的分割结果$\mathbf{y} \in \mathbb{R}^{1 \times H \times W}$；其中，第一个维度表示语义类别。

[0038] 本发明方法利用特征频率分离机制和动态空洞卷积模块对视频进行语义分割，该方法具有以下几个特点：1)不同于已有方法通过对高分辨率特征图进行统一处理，本发明所设计的特征频率分离模块将特征图分离出不同频率的特征，高频特征代表变化幅度大的区域，低频特征代表变化幅度小的区域，中频特征代表变化幅度适中的区域，对不同频率的

特征区分处理，使得网络学习到更有针对性的语义特征；2)通过构建动态空洞卷积模块，在不需要增加网络深度和宽度的情况下，根据输入特征的特点动态地给多个并行的空洞卷积分配不同的权重，能有效地将多个空洞卷积融合在一起，有利于提取更有效的语义特征；3)现有的大多数方法都通过叠加修正模块和增加网络深度来提高分割精度，而忽视了模型的冗余性和分割速度慢等问题，本发明方法通过对不同频率的特征进行不同深度的操作，进而有效地减少计算复杂度，提高分割速度。

[0039] 本发明方法适用于对实时性要求严格的视频语义分割，有益效果包括：1)利用特征频率分离模块能够有效地将特征图中不同频率的特征分离开来并区别处理，能够提高处理效率；2)通过构建动态空洞卷积模块能够在不显著增加网络复杂度的情况下，将多个空洞卷积进行融合，捕获特征图中更有效的语义信息，获得更加精确的分割结果；3)对于不同频率的特征，经过不同深度的动态空洞卷积模块，不仅可以有针对性地处理不同频率的特征，而且还能大量减少模型的计算量，提高了模型对视频的语义分割速度。本发明可应用于智能监控、无人机侦察、机器视觉与自动驾驶等实际任务中。

附图说明

[0040] 图1是本发明方法的流程图。

具体实施方式

[0041] 以下结合附图对本发明作进一步说明。

[0042] 如图1，利用多频动态空洞卷积的视频语义分割方法，首先对给定的视频进行采样并输入由卷积神经网络组成的编码器得到视频帧的浅层视觉特征图；然后利用由傅里叶变换、高斯滤波器、傅里叶逆变换构成的特征频率分离模块从浅层视觉特征图中分离出多频特征图；再利用由一个权重计算器、多个并行的空洞卷积核所构成的动态空洞卷积根据多频特征图进行不同深度的处理得到多频高层语义特征图；最后，将多频高层语义特征图进行拼接并输入解码器进行上采样得到语义分割结果。该方法将图像频率可分离的思想推广到浅层视觉特征图中，能对特征图的不同频率视觉区域进行区分，并利用不同深度的动态空洞卷积处理不同频率的特征图，在扩大特征图感受野的同时减少了模型的计算复杂度，能够实时地获得较高的语义分割精度。

[0043] 该方法首先获取视频数据集合，然后进行以下操作：

[0044] 步骤(1)对视频采样获得视频帧，并进行增强操作，然后输入至编码器，即深度卷积神经网络，获得对应的浅层视觉特征图；具体是：

[0045] (1-1)对单个视频进行均匀采样获得视频帧，采样率为10帧/秒，并对其进行增强操作得到数量为N的视频帧序列I，记为$\mathbf{I}=\left\{\mathbf{I}_i \middle| \mathbf{I}_i \in \mathbb{R}^{3\times H\times W}, i=1,\ldots,N\right\}$，其中$I_i$表示第i个视频帧，$\mathbb{R}$表示实数域，3表示RGB通道数量，H表示视频帧高度，W表示视频帧宽度；

[0046] (1-2)利用大型图像库ImageNet上预训练的卷积神经网络ResNet对视频帧序列I依次提取浅层视觉特征图$\mathbf{f}_i \in \mathbb{R}^{C_f\times H_f\times W_f}$，$C_f$表示特征图的通道数(本实施例取1024)，$H_f$表示特征图高度，$W_f$表示特征图宽度；ResNet具有多个由卷积层组成的模块，f_i为第i个视频帧经过RestNet前三个由多个卷积层组成的模块得到的特征图。

[0047] 步骤(2)构建特征频率分离模块，输入为浅层视觉特征图，输出多频特征图；具体是：

[0048] (2-1)构建特征频率分离模块，利用图像具有频率可分离的特点，对浅层视觉特征图进行三次高低频特征分离操作获得多频特征图；其中，高频特征刻画特征图的轮廓区域，低频特征刻画特征图的平面区域，中频特征刻画特征图的内容区域；

[0049] (2-2)高低频特征分离的具体操作如下：

[0050] 首先对浅层视觉特征图f_i做快速傅里叶变换，将空域信号转换为频域信号得到f_i的频谱图$\widehat{\mathbf{f}}_i \in \mathbb{R}^{C_f \times H_f \times W_f}$，将$\widehat{\mathbf{f}}_i$中低频信号部分平移到中间得到平移频谱图$\breve{\mathbf{f}}_i \in \mathbb{R}^{C_f \times H_f \times W_f}$，确定$\breve{\mathbf{f}}_i$的中心位置向量(P，Q)；其中，$\breve{\mathbf{f}}_i$通道中心点的横坐标值组成的向量$\mathbf{P}=\left\{\mathbf{P}_r \middle| \mathbf{P}_r \in \mathbb{R}, r=1,\ldots,C_f\right\}$，纵坐标值组成的向量$\mathbf{Q}=\left\{\mathbf{Q}_r \middle| \mathbf{Q}_r \in \mathbb{R}, r=1,\ldots,C_f\right\}$，下标r表示$\breve{\mathbf{f}}_i$的通道索引；

[0051] 然后将$\breve{\mathbf{f}}_i$中每个元素与低频转移函数$H_l(u_{r,a},v_{r,b})$作乘法运算得到低频平移频谱图$\breve{\mathbf{f}}_i^l \in \mathbb{R}^{C_f \times H_f \times W_f}$；高斯低通滤波器的转移函数$H_l\left(u_{r,a},v_{r,b}\right)=\exp\left(-D^2\left(u_{r,a},v_{r,b}\right)/2D_0^2\right)$，l表示低频信号，a表示像素点横轴坐标值，b表示像素点纵轴坐标值，{0≤a≤H_f，0≤b≤W_f}，exp(·)表示指数函数，D_0是设定的标准差(本实施例设为10)；其中，$D\left(u_{r,a},v_{r,b}\right)=\left(u_{r,a}^2,v_{r,b}^2\right)^{\frac{1}{2}}$，表示$\breve{\mathbf{f}}_i$中第r个通道像素点(a,b)距离坐标点($P_r$,$Q_r$)的欧式距离，$u_{r,a}$是$\breve{\mathbf{f}}_i$中第r个通道频谱位置(a,0)距离$P_r$的欧式距离，$v_{r,b}$是$\breve{\mathbf{f}}_i$中第r个通道频谱位置(0,b)距离$Q_r$的欧式距离；

[0052] 同理，将$\breve{\mathbf{f}}_i$中每个元素与高频转移函数$H_h(u_{r,a},v_{r,b})$作乘法运算得到高频平移频谱图$\breve{\mathbf{f}}_i^h \in \mathbb{R}^{C_f \times H_f \times W_f}$，其中h表示高频信号，$H_h\left(u_{r,a},v_{r,b}\right)=1-\exp(-D^2\left(u_{r,a},v_{r,b}\right)/2D_0^2)$；

[0053] 分别将频谱图$\breve{\mathbf{f}}_i^l$与$\breve{\mathbf{f}}_i^h$中的低频信号从中间平移回到原始位置，得到低频频谱图$\widehat{\mathbf{f}}_i^l \in \mathbb{R}^{C_f \times H_f \times W_f}$和高频频谱图$\widehat{\mathbf{f}}_i^h \in \mathbb{R}^{C_f \times H_f \times W_f}$；

[0054] 最后将$\widehat{\mathbf{f}}_i^l$和$\widehat{\mathbf{f}}_i^h$分别做快速傅里叶逆变换将频域信号转换为空域信号，得到弱低频特征图$\mathbf{f}_i^l \in \mathbb{R}^{C_f \times H_f \times W_f}$和弱高频特征图$\mathbf{f}_i^h \in \mathbb{R}^{C_f \times H_f \times W_f}$；

[0055] (2-3)按照(2-2)，对弱高频特征图$\mathbf{f}_i^h$进行第二次高低频特征分离操作，得到强高频特征图$\mathbf{f}_i^{hh} \in \mathbb{R}^{C_f \times H_f \times W_f}$和中高频特征图$\mathbf{f}_i^{hl} \in \mathbb{R}^{C_f \times H_f \times W_f}$，hh表示特征图经过两次高频信号过滤，hl表示特征图先经过一次高频信号过滤，再经过一次低频信号过滤；

[0056] 按照(2-2)，对弱低频特征图$\mathbf{f}_i^l$进行第二次高低频特征分离操作，得到强低频特征图$\mathbf{f}_i^{ll} \in \mathbb{R}^{C_f \times H_f \times W_f}$和中低频特征图$\mathbf{f}_i^{lh} \in \mathbb{R}^{C_f \times H_f \times W_f}$，ll表示特征图经过两次低频信号过滤，lh表示特征图先经过一次低频信号过滤，再经过一次高频信号过滤；

[0057] (2-4)将中高频特征图$\mathbf{f}_i^{hl}$和中低频特征图$\mathbf{f}_i^{lh}$进行一次拼接,经过一次大小为1×1的卷积操作得到压缩后的特征图,再经过步长为2的最大池化操作进行下采样得到中频特征图$\hat{\mathbf{f}}_i^m \in \mathbb{R}^{C_f^m \times \frac{H_f}{2} \times \frac{W_f}{2}}$,其中m表示中频信号,$C_f^m$为中频特征图的通道维度;

[0058] (2-5)将强低频特征图$\mathbf{f}_i^{ll}$经过一次大小为1×1卷积操作得到压缩后的特征图,再经过步长为4的最大池化操作进行下采样得到低频特征图$\hat{\mathbf{f}}_i^l \in \mathbb{R}^{C_f^l \times \frac{H_f}{4} \times \frac{W_f}{4}}$;将强高频特征图$\mathbf{f}_i^{hh}$经过一次大小为1×1卷积操作得到压缩后的高频特征图$\hat{\mathbf{f}}_i^h \in \mathbb{R}^{C_f^h \times H_f \times W_f}$;其中,$C_f^m + C_f^h + C_f^l = C_f$,$C_f^h$和$C_f^l$分别表示高频特征图和低频特征图的通道维度。

[0059] 步骤(3)构建动态空洞卷积模块,输入为多频特征图,输出多频高层语义特征图;具体是:

[0060] (3-1)构建由一个权重计算器、K个并行的空洞卷积核组成的动态空洞卷积模块,将多频特征图分别输入到动态空洞卷积模块,得到多频高层语义特征图,包括低频高层语义特征图、中频高层语义特征图和高频高层语义特征图;

[0061] (3-2)动态空洞卷积的具体操作如下:将低频特征图$\hat{\mathbf{f}}_i^l$输入到权重计算器得到输出K个权重$\{w_t \in \mathbb{R} \mid t = 1, ..., K\}$,$w_t$表示第t个空洞卷积的权重,0≤$w_t$<1,$\sum_t^K w_t = 1$;权重计算器由一次全局平均池化操作、一个全连接层、一个Relu函数、一个全连接层、一个Softmax函数组成;K个并行的空洞卷积核$\{\mathbf{K}_t \in \mathbb{R}^{C_f^l \times 3 \times 3} \mid t = 1, ..., K\}$,$K_t$表示第t个空洞率为2的3×3空洞卷积;$K_t$分别与对应的权重$w_t$做点乘运算,再将K个并行的空洞卷积相加得到集成空洞卷积核$\widehat{\mathbf{K}} \in \mathbb{R}^{C_f^l \times 3 \times 3}$,用于利用多个并行空洞卷积的参数以捕获不同感受野;低频特征图$\hat{\mathbf{f}}_i^l$再与综合空洞卷积核$\widehat{\mathbf{K}}$进行卷积操作得到低频高层语义特征图$\hat{\mathbf{t}}_i^l \in \mathbb{R}^{2C_f^l \times \frac{H_f}{8} \times \frac{W_f}{8}}$,$2C_f^l$表示通道数$C_f^l$的两倍;(3-3)对动态空洞卷积模块串行叠加,第一个动态空洞卷积模块的输出作为第二个动态空洞卷积模块的输入;按照(3-2),中频特征图$\hat{\mathbf{f}}_i^m$经过两个串行的动态空洞卷积模块得到中频高层语义特征图$\hat{\mathbf{t}}_i^m \in \mathbb{R}^{4C_f^m \times \frac{H_f}{8} \times \frac{W_f}{8}}$,$4C_f^m$表示通道数$C_f^m$的四倍;同理,高频特征图$\hat{\mathbf{f}}_i^h$经过四个串行的动态空洞卷积模块得到高频高层语义特征图$\hat{\mathbf{t}}_i^h \in \mathbb{R}^{8C_f^h \times \frac{H_f}{8} \times \frac{W_f}{8}}$,$8C_f^h$表示通道数$C_f^h$的八倍。

[0062] 步骤(4)将多频高层语义特征图输入解码器即上采样卷积模块,获得视频帧的分割掩膜;具体是:

[0063] (4-1)构建由三个转置卷积层组成的解码器,转置卷积即卷积的逆向过程,通过与输入的小尺寸特征图进行卷积操作得到大尺寸特征图;

[0064] (4-2)将低频高层语义特征图$\hat{\mathbf{t}}_i^l$、中频高层语义特征图$\hat{\mathbf{t}}_i^m$和高频高层语义特征图$\hat{\mathbf{t}}_i^h$

进行通道维度上的拼接得到集成高层语义特征图$\mathbf{t}_i=[\hat{\mathbf{t}}_i^l;\hat{\mathbf{t}}_i^m;\hat{\mathbf{t}}_i^h]\in\mathbb{R}^{\left(2C_f^l+4C_f^m+8C_f^h\right)\times\frac{H_f}{8}\times\frac{W_f}{8}}$；

[0065] (4-3)将集成语义特征图t_i输入解码器得到分割掩膜$\hat{\mathbf{y}}_i\in\mathbb{R}^{C\times H\times W}$，C表示语义类别总数，视频帧中每个像素对应的类别为所有类别中概率最大的类别。

[0066] 步骤(5)迭代训练由编码器、特征频率分离模块、动态空洞卷积模块、解码器组成的视频语义分割模型直至收敛，然后将新视频输入至该模型得到对应的语义分割结果；具体是：

[0067] (5-1)建立由编码器、特征频率分离模块、动态空洞卷积模块、解码器组成的视频语义分割模型；

[0068] (5-2)将视频帧序列依次输入到语义分割模型得到分割掩膜$\hat{\mathbf{y}}_i\in\mathbb{R}^{C\times H\times W}$，i＝1,…,N,根据交叉熵损失，通过梯度反向传播方法调整模型参数，迭代优化模型直至收敛；

[0069] (5-3)将新视频的每一帧输入到已训练好的模型中，依据(5-2)依次输出相应的分割结果$\mathbf{y}\in\mathbb{R}^{1\times H\times W}$；其中，第一个维度表示语义类别。

[0070] 本实施例所述的内容仅仅是对发明构思的实现形式的列举，本发明的保护范围的不应当被视为仅限于实施例所陈述的具体形式，本发明的保护范围也及于本领域技术人员根据本发明构思所能够想到的等同技术手段。

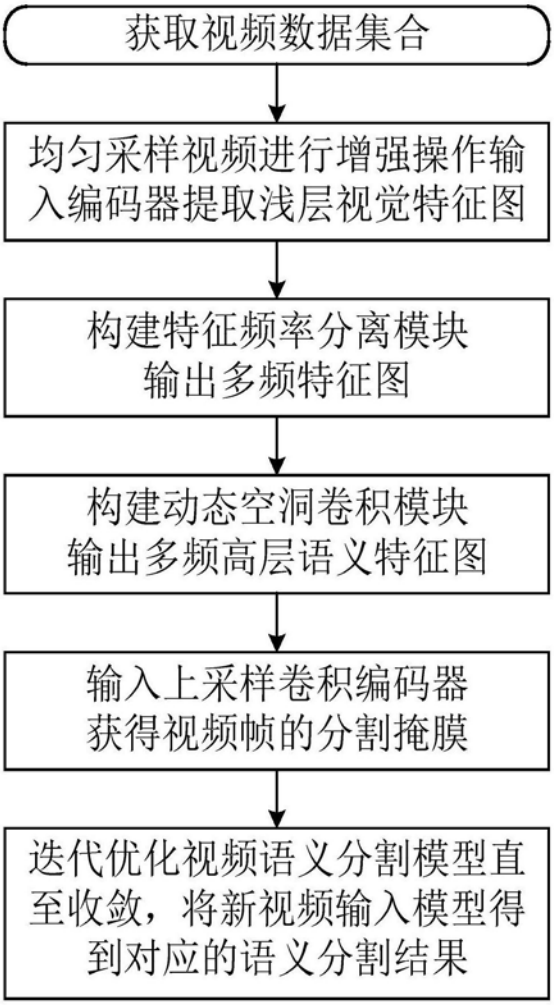

获取视频数据集合
均匀采样视频进行增强操作输入编码器提取浅层视觉特征图
构建特征频率分离模块输出多频特征图
构建动态空洞卷积模块输出多频高层语义特征图
输入上采样卷积编码器获得视频帧的分割掩膜
迭代优化视频语义分割模型直至收敛，将新视频输入模型得到对应的语义分割结果

图1