



(12) 发明专利申请

(10) 申请公布号 CN 117036711 A

(43) 申请公布日 2023. 11. 10

(21) 申请号 202311064941.7

G06N 3/0895 (2023.01)

(22) 申请日 2023.08.23

G06N 3/0455 (2023.01)

G06N 3/082 (2023.01)

(71) 申请人 南京信息工程大学

地址 210032 江苏省南京市江北新区宁六路219号

(72) 发明人 苏京峰 李军侠

(74) 专利代理机构 南京经纬专利商标代理有限公司 32200

专利代理师 周科技

(51) Int. Cl.

G06V 10/26 (2022.01)

G06V 20/70 (2022.01)

G06V 10/82 (2022.01)

G06N 3/042 (2023.01)

G06N 3/084 (2023.01)

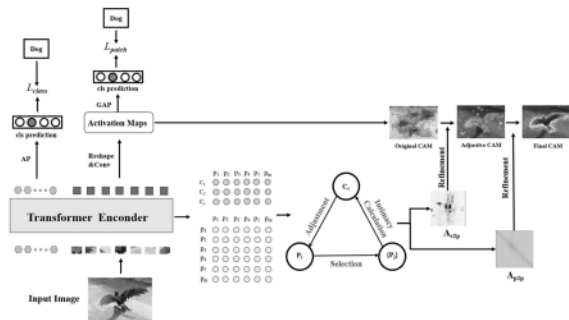
权利要求书2页 说明书5页 附图2页

(54) 发明名称

一种基于注意力调节的弱监督语义分割方法

(57) 摘要

本发明公开了一种基于注意力调节的弱监督语义分割方法,探索Transformer在弱监督语义分割任务中的应用。基于Transformer的方法会使用注意力对类激活图进行优化,然而由于部分类与块之间的注意力存在错误,导致优化之后得到的类激活图存在不完全激活问题。针对此问题,本发明提出了一种新颖的弱监督语义分割框架,在该框架中设计了一个注意力调节策略,根据块与块之间的注意力来调节类与块之间的注意力,调节后的注意力可以激活更多的目标区域。在PASCAL VOC 2012数据集和MS COCO 2014数据集上与最新的方法进行比较,本发明方法均取得了最优的结果。



1. 一种基于注意力调节的弱监督语义分割方法,其特征在于,包括以下步骤:

步骤1,数据准备:获取标注图像数据集,并将数据集划分为训练集、验证集和测试集;

步骤2,数据预处理:对图像进行随机水平翻转以及颜色抖动处理,对图像进行归一化处理,并进行随机裁剪,将裁剪后的图像作为弱监督语义分割模型的输入;

步骤3,模型搭建:采用在ImageNet上预训练的DeiT-S作为模型的主干,构建弱监督语义分割模型;

步骤4,模型训练:使用Adam优化器优化弱监督语义分割模型,同时使用训练集对模型进行设定周期的训练,损失函数使用多标签交叉熵损失,训练好的模型生成类激活图;

步骤5,根据类激活图的值对每个像素位置分配一个类别生成像素级伪标签,然后使用像素级伪标签对语义分割网络DeeplabV2进行训练;将验证集和测试集中的图片输入到训练好的模型中得到最终的分割图。

2. 根据权利要求1所述的基于注意力调节的弱监督语义分割方法,其特征在于,步骤3中模型搭建包括:

步骤3.1,搭建基于注意力融合的弱监督语义分割框架,将预处理后的图像分割成N个不重叠的块,然后通过线性映射构造N个块令牌,并将C个类令牌与N个块令牌拼接得到框架的输入令牌;

步骤3.2,将输入令牌输入到框架中的Transformer编码层,得到输出令牌;然后从输出令牌中提取最后N个块令牌组成输出块令牌 $T_{p\_out}$ ,并对其进行重组以及卷积操作得到初始类激活图Original-CAM;

步骤3.3,输入令牌经过Transformer编码层时注意力模块对输入令牌进行注意力计算产生注意力Attention,计算公式如下:

$$\text{Attention} = \frac{QK^T}{\sqrt{d_k}}$$

其中Q和K分别表示输入令牌在经过Transformer编码层时通过线性投影得到的Query矩阵及Key矩阵,T表示矩阵转置, $d_k$ 表示缩放因子;

步骤3.4,Attention进一步划分为类到块注意力 $A_{c2p}$ 和块到块注意力 $A_{p2p}$ ,然后通过块与块之间的注意力 $A_{p2p}$ 对类与块之间的注意力 $A_{c2p}$ 进行调节;

步骤3.5,使用类到块注意力 $A_{c2p}$ 和块到块注意力 $A_{p2p}$ 对初始类激活图进行优化。

3. 根据权利要求2所述的基于注意力调节的弱监督语义分割方法,其特征在于,类到块注意力 $A_{c2p}$ 和块到块注意力 $A_{p2p}$ 表示如下:

$$A_{c2p} = \text{Attention}[1:C, C+1:C+N]$$

$$A_{p2p} = \text{Attention}[C+1:C+N, C+1:C+N]$$

调节类c与块i之间的注意力,过程如下:

首先,根据各个块与块i之间的注意力对各个块按照注意力值从大到小的顺序进行排序,并选择排序后排名前p%的块;

然后,将类c与选择的块之间的注意力取出并进行计算,获得类c与块i之间的注意力调节因子:

$$r(c, i) = \frac{1}{S} \sum_{j=1}^U A_{c2p}(c, j)$$

其中,  $r(c, i)$  表示  $A_{c2p}$  中类  $c$  与块  $i$  之间的注意力调节因子,  $c \in \{1, 2, \dots, C\}$  表示数据集类别总数,  $i, j$  表示块,  $i \in \{1, 2, \dots, N\}$ ,  $j \in U$ ,  $U$  表示与块  $i$  之间注意力最大的前  $p\%$  个块的集合,  $S$  表示  $U$  中块的数量;  $A_{c2p}(c, j)$  表示表示类  $c$  与块  $j$  间的注意力;

接下来将注意力调节因子  $r(c, i)$  加到类  $c$  与块  $i$  之间的注意力中进行调节:

$$A_{c2p}(c, i) = A_{c2p}(c, i) + \alpha * r(c, i)$$

其中,  $A_{c2p}(c, i)$  表示类  $c$  与块  $i$  之间的注意力,  $\alpha$  代表注意力调节因子系数。

4. 根据权利要求2所述的基于注意力调节的弱监督语义分割方法, 其特征在于, 步骤3.5中使用类到块注意力  $A_{c2p}$  和块到块注意力  $A_{p2p}$  对初始类激活图进行优化, 包括:

首先将初始类激活图  $Original-CAM$  与类到块注意力相乘, 得到初步优化后的调节类激活图;

然后再通过将块到块注意力与调节类激活图之间进行矩阵乘法来进一步优化, 得到最终类激活图。

5. 根据权利要求1-4任一所述的基于注意力调节的弱监督语义分割方法, 其特征在于, 步骤4中模型训练过程如下:

步骤4.1, 设置弱监督语义分割模型超参数: 模型训练次数  $Epoch$ 、初始学习率和模型训练批次  $batch\_size$ , 训练时使用的优化器为  $Adam$  优化器, 损失函数是多标签交叉熵损失;

步骤4.2, 对弱监督语义分割模型进行多轮训练, 将训练  $mIoU$  值最高的一轮结果对应的参数进行保存;

步骤4.3, 弱监督语义分割模型训练完成之后, 将保存的最好参数加载到模型中, 然后将训练集数据输入到模型中, 训练好的模型生成完整的类激活图。

## 一种基于注意力调节的弱监督语义分割方法

### 技术领域

[0001] 本发明属于图像分割技术领域,具体涉及一种基于注意力调节的弱监督语义分割方法。

### 背景技术

[0002] 语义分割是计算机视觉领域一个基础而又具有挑战性的任务之一,其研究目的是对图像中的每个像素点进行分类,并将其分配给特定的语义类别。语义分割在许多领域具有广泛的应用,如图像识别、自动驾驶、医学图像分析、场景理解和视频分析等,它可以帮助计算机更好地理解图像中的内容,从而实现自动化的场景理解和决策。近年来由于深度学习方法的蓬勃发展,语义分割也取得了显著的进展,其中全监督的语义分割模型被广泛应用并取得了优秀的性能。然而训练全监督的语义分割模型往往需要大规模的像素级标注数据,而像素级标注数据的获取往往难度大且耗时耗力。为了解决这个问题,许多工作开始采用弱监督语义分割技术。旨在使用边界框标注、点标注、涂鸦标注或图像级标注等弱标注训练语义分割网络。其中图像级标注是最方便获取的标注,在弱监督语义分割中得到了广泛的研究。

[0003] 虽然图像级标注的获取非常方便,但是图像级标注存在一个问题,即它不能提供足够的位置监督信息,因为它仅给出一张图像中包含的对象类别信息,并没有指出对象类别在图像中具体的位置信息。类激活图(CAM)的发展提供了一种只使用图像级标注来获取位置信息的有效方法。对于图像级标注的弱监督语义分割,大多数现有方法通常使用以下流程来解决:1)使用图像级标注训练卷积神经网络(CNN),从中生成类激活图以获得种子区域;2)对种子区域进行一定约束的扩展以获得伪标签;3)使用伪标签作为真实标签来训练全监督语义分割网络。但是卷积神经网络产生的类激活图存在一个问题,即它倾向于激活一个局部的有辨别力的区域,而忽略了完整的对象范围,导致不完全激活问题。目前有研究证明这是由于卷积神经网络的固有特性导致的,即卷积神经网络中的卷积操作只能捕获小范围的特征依赖性,无法探索全局特征关系,这导致激活对象区域过小,从而影响生成的伪标签质量,最终导致难以取得理想的弱监督语义分割结果。

[0004] 目前,Transform在许多计算机视觉任务中取得了巨大的成功,这主要得益于其本身的注意力机制。Transformer的注意力机制可以对全局特征关系进行建模,并克服卷积神经网络的上述缺点。故一些研究者们开始使用Transformer进行弱监督语义分割的研究,它们通常利用Transformer结构进行图像特征提取并生成类激活图,之后使用注意力对类激活图进行优化得到较为完整的类激活图。虽然现有基于Transformer的弱监督语义分割方法通常使用注意力对类激活图进行优化,但是由于Transformer生成的注意力中部分类与块之间注意力存在误差,导致类激活图在经过注意力优化后仍然不能够完整地激活对象区域。

## 发明内容

[0005] 本发明的目的是解决弱监督语义分割中目标区域不能够被完全激活的问题,提供了一种基于注意力融合的弱监督语义分割方法,生成更加完整且准确的覆盖目标区域的类激活图,通过类激活图来生成较高质量的像素级伪标签进行语义分割网络的训练,提高网络的语义分割性能。

[0006] 技术方案:为实现本发明的目的,本发明所采用的技术方案是:一种基于注意力调节的弱监督语义分割方法,包括以下步骤:

[0007] 步骤1,数据准备:获取标注图像数据集,并将数据集划分为训练集、验证集和测试集;

[0008] 步骤2,数据预处理:对图像进行随机水平翻转以及颜色抖动处理,对图像进行归一化处理,并进行随机裁剪,将裁剪后的图像作为弱监督语义分割模型的输入;

[0009] 步骤3,模型搭建:采用在ImageNet上预训练的DeiT-S作为模型的主干,构建弱监督语义分割模型;

[0010] 步骤4,模型训练:使用Adam优化器优化弱监督语义分割模型,同时使用训练集对模型进行设定周期的训练,损失函数使用多标签交叉熵损失,训练好的模型生成类激活图;

[0011] 步骤5,根据类激活图的值对每个像素位置分配一个类别生成像素级伪标签,然后使用像素级伪标签对语义分割网络DeeplabV2进行训练;将验证集和测试集中的图片输入到训练好的模型中得到最终的分割图。

[0012] 进一步的,步骤3中模型搭建包括:

[0013] 步骤3.1,搭建基于注意力融合的弱监督语义分割框架,将预处理后的图像分割成N个不重叠的块,然后通过线性映射构造N个块令牌,并将C个类令牌与N个块令牌拼接得到框架的输入令牌;

[0014] 步骤3.2,将输入令牌输入到框架中的Transformer编码层,得到输出令牌;然后从输出令牌中提取最后N个块令牌组成输出块令牌Tp\_out,并对其重组以及卷积操作得到初始类激活图Original-CAM;

[0015] 步骤3.3,输入令牌经过Transformer编码层时注意力模块对输入令牌进行注意力计算产生注意力Attention,计算公式如下:

$$[0016] \quad \text{Attention} = \frac{QK^T}{\sqrt{d_k}}$$

[0017] 其中Q和K分别表示输入令牌在经过Transformer编码层时通过线性投影得到的Query矩阵及Key矩阵,T表示矩阵转置, $d_k$ 表示缩放因子;

[0018] 步骤3.4,Attention进一步划分为类到块注意力 $A_{c2p}$ 和块到块注意力 $A_{p2p}$ ,然后通过块与块之间的注意力 $A_{p2p}$ 对类与块之间的注意力 $A_{c2p}$ 进行调节;

[0019] 步骤3.5,使用类到块注意力 $A_{c2p}$ 和块到块注意力 $A_{p2p}$ 对初始类激活图进行优化。

[0020] 进一步的,类到块注意力 $A_{c2p}$ 和块到块注意力 $A_{p2p}$ 表示如下:

$$[0021] \quad A_{c2p} = \text{Attention}[1:C, C+1:C+N]$$

$$[0022] \quad A_{p2p} = \text{Attention}[C+1:C+N, C+1:C+N]$$

[0023] 调节类c与块i之间的注意力,过程如下:

[0024] 首先,根据各个块与块i之间的注意力对各个块按照注意力值从大到小的顺序进

行排序,并选择排序后排名前p%的块;

[0025] 然后,将类c与选择的块之间的注意力取出并进行计算,获得类c与块i之间的注意力调节因子:

$$[0026] \quad r(c, i) = \frac{1}{S} \sum_{j=1}^U A_{c2p}(c, j)$$

[0027] 其中, $r(c, i)$ 表示 $A_{c2p}$ 中类c与块i之间的注意力调节因子, $c \in \{1, 2, \dots, C\}$ 表示数据集类别总数, $i, j$ 表示块, $i \in \{1, 2, \dots, N\}$ , $j \in U$ , $U$ 表示与块i之间注意力最大的前p%个块的集合, $S$ 表示 $U$ 中块的数量; $A_{c2p}(c, j)$ 表示表示类c与块j间的注意力;

[0028] 接下来将注意力调节因子 $r(c, i)$ 加到类c与块i之间的注意力中进行调节:

$$[0029] \quad A_{c2p}(c, i) = A_{c2p}(c, i) + \alpha * r(c, i)$$

[0030] 其中, $A_{c2p}(c, i)$ 表示类c与块i之间的注意力, $\alpha$ 代表注意力调节因子系数。

[0031] 进一步的,步骤3.5中使用类到块注意力 $A_{c2p}$ 和块到块注意力 $A_{p2p}$ 对初始类激活图进行优化,包括:

[0032] 首先将初始类激活图Original-CAM与类到块注意力相乘,得到初步优化后的调节类激活图;

[0033] 然后再通过将块到块注意力与调节类激活图之间进行矩阵乘法来进一步优化,得到最终类激活图。

[0034] 进一步的,步骤4中模型训练过程如下:

[0035] 步骤4.1,设置弱监督语义分割模型超参数:模型训练次数Epoch、初始学习率和模型训练批次batch\_size,训练时使用的优化器为Adam优化器,损失函数是多标签交叉熵损失;

[0036] 步骤4.2,对弱监督语义分割模型进行多轮训练,将训练mIoU值最高的一轮结果对应的参数进行保存;

[0037] 步骤4.3,弱监督语义分割模型训练完成之后,将保存的最好参数加载到模型中,然后将训练集数据输入到模型中,训练好的模型生成完整的类激活图。

[0038] 有益效果:与现有技术相比,本发明的技术方案具有以下有益的技术效果:

[0039] 本发明主要解决弱监督语义分割中类激活图的不完全激活问题。以Transformer为基本网络结构,提出了一个简单而有效的弱监督语义分割框架。在该框架中,首先设计了一个注意力调节策略,根据块和块之间的注意力对类和块之间的注意力进行调节,有效降低了类和块之间产生错误关联的概率,然后使用调节后的注意力对类激活图进行优化,此时得到的类激活图中目标区域能够被更加完整且准确地激活,能够较好地解决类激活图的不完全激活问题。

## 附图说明

[0040] 图1是基于注意力融合的弱监督语义分割总体框架图。

[0041] 图2是PASCAL VOC 2012验证集上的分割结果示例图。

[0042] 图3是MS COCO 2014验证集上的分割结果示例图。

## 具体实施方式

[0043] 下面结合附图和实施例对本发明的技术方案作进一步的说明。

[0044] 本发明所述的一种基于注意力调节的弱监督语义分割方法,提出了一种新颖的基于Transformer的框架,用于图像级标注下的弱监督语义分割任务,该框架总体结构如图1所示,主要由三部分组成:1)利用Transformer进行特征提取并生成初始类激活图;2)注意力调节模块,根据块与块之间的注意力来调节类与块之间的注意力,有效提高类与块之间注意力的准确性;3)利用注意力来优化类激活图,得到更加完整且准确的类激活图。本发明方法具体实现包括以下步骤:

[0045] 步骤1:数据准备。

[0046] 在本发明中,使用PASCAL VOC 2012数据集和MS COCO 2014数据集。其中Pascal VOC 2012数据集有21个类别,包括20个对象类和一个背景类;MS COCO 2014数据集有81个类别,包括80个对象类和一个背景类。PASCAL VOC 2012数据集可以分为三部分:训练集(包括1464幅图像)、验证集(包括1449幅图像)和测试集(包括1456幅图像),其中训练集通常采用额外数据扩充后的10582张图像;MS COCO 2014数据集可以分为两部分:训练集(包括82081幅图像)和验证集(包括40137幅图像)。

[0047] 步骤2:数据预处理。

[0048] 对图像进行随机水平翻转以及颜色抖动处理,并将图像的亮度、对比度、饱和度值均设置0.3。使用transforms.Normalize对图像进行归一化处理,将其大小调整为 $256 \times 256$ ,然后使用transforms.RandomCrop对图像进行随机裁剪,将图像大小调整为 $224 \times 224$ 。裁剪后的图像输入到模型中。

[0049] 步骤3:模型搭建。

[0050] 步骤3.1:搭建基于注意力融合的弱监督语义分割框架,将步骤2预处理后的图像分割成N个不重叠的块,然后通过线性映射构造N个块令牌,并将C个类令牌与N个块令牌拼接得到框架的输入令牌。

[0051] 步骤3.2:将输入令牌输入到框架中的Transformer编码层,得到输出令牌。然后从输出令牌中提取最后N个块令牌组成输出块令牌 $T_{p\_out}$ ,并对其进行重组(Reshape)以及卷积(Conv)操作得到初始类激活图Original-CAM。

[0052]  $Original-CAM = Conv(Reshape(T_{p\_out}))$

[0053] 步骤3.3:输入令牌经过Transformer编码层时注意力模块会对输入令牌进行注意力计算产生注意力Attention,形状为 $[C+N, C+N]$ ,计算公式如下:

$$[0054] \quad Attention = \frac{QK^T}{\sqrt{d_k}}$$

[0055] 其中Q,K表示输入令牌在经过transformer编码层时通过线性投影得到的Query矩阵及Key矩阵,T表示矩阵转置, $d_k$ 表示缩放因子。

[0056] 步骤3.4:Attention可以进一步划分为类到块注意力 $A_{c_{2p}}$ 和块到块注意力 $A_{p_{2p}}$ ,其中 $A_{c_{2p}} = Attention[1:C, C+1:C+N]$ ,  $A_{p_{2p}} = Attention[C+1:C+N, C+1:C+N]$ 。然后通过块与块之间的注意力 $A_{p_{2p}}$ 来对类与块之间的注意力 $A_{c_{2p}}$ 进行调节。假如要调节类c与块i之间的注意力,首先根据各个块与块i之间的注意力对各个块按照注意力值从大到小的顺序进行排序,接下来选择排序后排名前30%的一些块,然后对这些块之间的注意力进行计算,获得类c与

块i之间的注意力调节因子:

$$[0057] \quad r(c, i) = \frac{1}{S} \sum_{j=1}^U A_{c2p}(c, j)$$

[0058] 其中 $r(c, i)$ 表示 $A_{c2p}$ 中类 $c$ 与块 $i$ 间的注意力调节因子,  $c \in \{1, 2, \dots, C\}$ 表示数据集类别总数,  $i, j$ 表示块,  $i \in \{1, 2, \dots, N\}$ ,  $j \in U$ ,  $U$ 表示与块 $i$ 之间注意力较大的块的集合,  $S$ 表示 $U$ 中块的数量。接下来将注意力调节因子 $r(c, i)$ 加到类 $c$ 与块 $i$ 之间的注意力中进行调节:

$$[0059] \quad A_{c2p}(c, i) = A_{c2p}(c, i) + \alpha * r(c, i)$$

[0060] 其中,  $A_{c2p}(c, i)$ 表示类 $c$ 与块 $i$ 之间的注意力,  $\alpha$ 代表注意力调节因子系数。

[0061] 步骤3.5:使用类到块注意力 $A_{c2p}$ 和块到块注意力 $A_{p2p}$ 来对初始类激活图进行优化。首先将初始类激活图Original-CAM与类到块注意力相乘,得到初步优化后的调节类激活图,然后再通过将块到块注意力与调节类激活图之间进行矩阵乘法来进一步优化,得到最终类激活图。

[0062] 步骤4:模型训练。

[0063] 步骤4.1:设置弱监督语义分割模型相关超参数,将模型训练次数Epoch设置为60,将模型训练批次batch\_size设置为64,训练时使用的优化器为Adam优化器,损失函数是多标签交叉熵损失,初始学习率设置为 $5e-4$ 。

[0064] 步骤4.2:对弱监督语义分割模型进行多轮训练,通过对训练集结果进行观察,将训练的最好一轮结果(训练mIoU值最高)对应的参数进行保存。

[0065] 步骤4.3:弱监督语义分割模型训练完成之后,将保存的最好参数加载到模型中,然后将训练集数据输入到模型中,训练好的模型能够生成较为完整的类激活图。

[0066] 步骤5:根据类激活图的值对每个像素位置分配一个类别生成像素级伪标签,然后使用像素级伪标签对现有的语义分割网络DeepLabV2进行训练。将验证集和测试集中的图片输入到训练好的模型中得到最终的分割图,如图2、图3所示,第二列是真实分割图,第三列是本发明的预测分割图,可以发现,本发明的模型预测的分割图跟真实的分割图很接近。



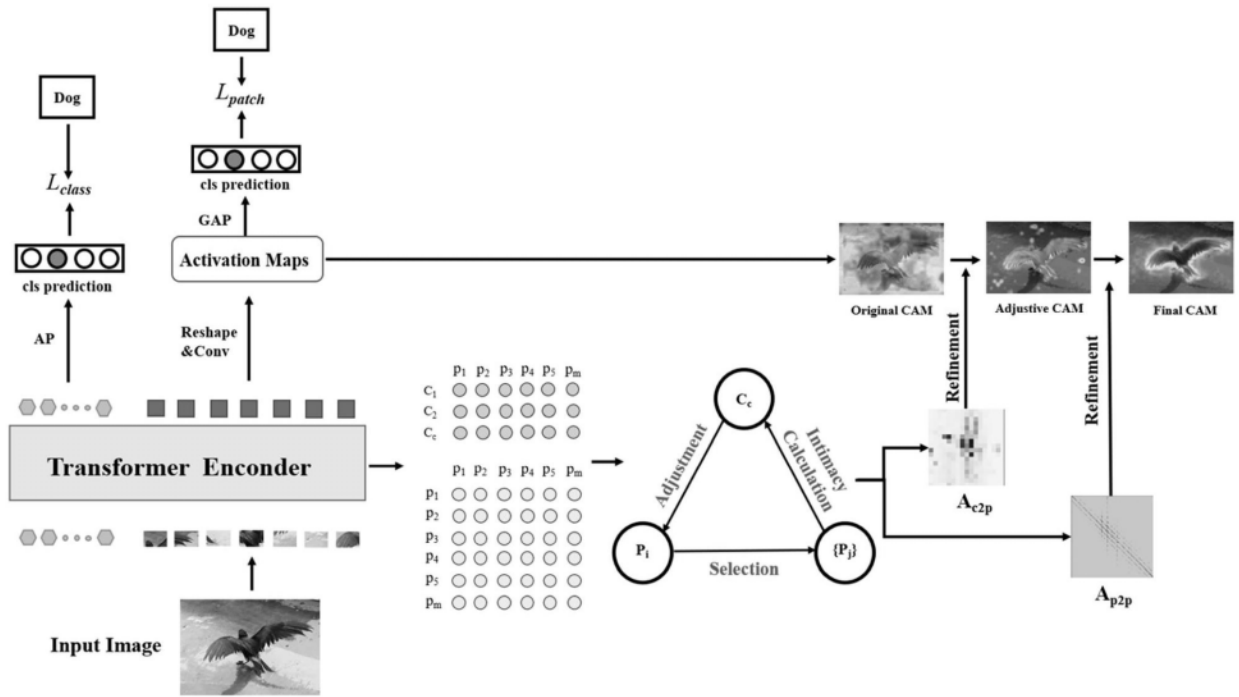


图1



图2

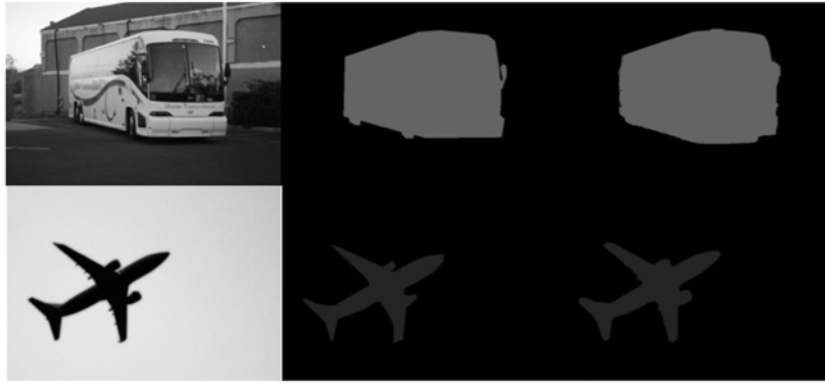


图3