

基于不确定性校准的烟雾语义分割^①



刘志宏, 杨海波, 贾军营, 卢 鑫

(沈阳工业大学 信息科学与工程学院, 沈阳 110870)

通信作者: 杨海波, E-mail: yanghb@sut.edu.cn

摘 要: 烟雾检测在早期火灾预警当中非常重要. 现有检测算法基本是基于确定性的卷积神经网络来进行的, 然而确定性的神经网络往往会给出非常自信的预测结果, 即使它完全不知道某些区域当中是否有目标对象, 尤其是烟雾边缘区域有着更加透明的效果, 致使该区域和周围环境极易混淆, 因此检测算法对该区域并不能进行很好的判断, 进而造成大量的假阳性. 因此, 本文提出一种改进的 DeepLabV3+ 算法, 首先, 该算法基于贝叶斯思想优化 DeepLabV3+ 从而输出非确定性的特征编码, 以量化预测图像中不确定性的量, 校准模型的学习过程. 其次基于预处理思想对特征编码进行预处理, 降低无关干扰特征信息量, 并且强化 DeepLabV3+ 网络中特征融合能力, 充分利用网络提取到的多尺度特征信息. 最后将 DeepLabV3+ 网络中上采样算子优化为 CARAFE 算子, 降低上采样过程中重要信息的丢失. 模型在公开的 SMOKE5K 数据集上取得良好的性能, *MIoU* 指标达到了 92.41%.

关键词: DeepLabV3+; 烟雾检测; 不确定性校准; 特征预处理; CARAFE

引用格式: 刘志宏, 杨海波, 贾军营, 卢鑫. 基于不确定性校准的烟雾语义分割. 计算机系统应用, 2025, 34(5): 31-40. <http://www.c-s-a.org.cn/1003-3254/9835.html>

Smoke Semantic Segmentation Based on Uncertainty Calibration

LIU Zhi-Hong, YANG Hai-Bo, JIA Jun-Ying, LU Xin

(School of Information Science and Engineering, Shenyang University of Technology, Shenyang 110870, China)

Abstract: Smoke detection is very important in early fire warning. The existing detection algorithms are basically based on deterministic convolutional neural networks. However, deterministic neural networks tend to give very confident prediction results, even though they do not know whether there is a target object in some regions at all. In particular, the smoke edge region is more transparent, making it extremely easy for these areas to be confused with the surrounding environment. Therefore, the detection algorithm cannot make a good judgment on this region, producing a large number of false positives. So, an improved DeepLabV3+ algorithm is proposed. First, the algorithm optimizes DeepLabV3+ based on Bayesian ideas to output non-deterministic feature coding, so as to quantify the measurement of uncertainty in the predicted image and calibrate the learning process of the model. Secondly, feature coding is preprocessed based on the preprocessing idea to reduce the amount of information of irrelevant interfering features, and the feature fusion capability in the DeepLabV3+ network is strengthened to make full use of the multi-scale feature information extracted from the network. Finally, the upsampling operator in the DeepLabV3+ network is optimized as the CARAFE operator to reduce the loss of important information in the upsampling process. The model achieves good performance on the open SMOKE5K dataset, with the *MIoU* index reaching 92.41%.

Key words: DeepLabV3+; smoke detection; uncertainty calibration; feature preprocessing; CARAFE

① 基金项目: 辽宁省应用基础研究项目 (2022JH2/101300243); 2022 年度沈阳市科学技术计划 (22316107)

收稿时间: 2024-10-21; 修改时间: 2024-11-12; 采用时间: 2024-11-28; csa 在线出版时间: 2025-03-24

CNKI 网络首发时间: 2025-03-25

火灾的发生不仅给人们的财产带来极大的损失,甚至危及人的生命安全.如果能在火灾发生初期通过对烟雾进行识别及时给人们预警可以有效遏止火灾的扩散,降低损失.传统烟雾检测是采用光电和离子感烟器等传感器,适合对室内场景进行监控.这种检测方法灵敏度高,但监控范围窄,响应速度慢,不宜应用在高大空间和室外等场合.近年来,随着深度学习在图像处理领域的发展,基于深度学习的烟雾检测技术已经崭露头角.这种技术检测速度快,监控范围广,成本低,特别适用于森林和仓库等室外场景及大空间场景的火焰和烟雾检测.此外,由于其反应速度快于传统火焰和烟雾检测,因而能在火焰和烟雾形成的初期就及时进行预警.但是目前的烟雾检测算法存在一些问题,首先,烟雾的半透明特殊外观特征导致检测算法很容易受到周围环境的干扰,会导致大量的误检.其次,为了提升模型的泛化能力,现有的烟雾检测算法往往只针对数据进行预处理,对于后续的特征融合,现有的检测算法对融合前的特征不作任何处理而直接进行融合,使得融合后的特征冗余度太高,引入大量与烟雾不相关的干扰特征信息,反而会导致模型泛化能力的降低.最后,现有的烟雾检测算法在进行最后的上采样时往往采用简单的线性插值方法,线性插值的方法是不可学习的,因此其可扩展性低,且线性插值方法会丢失太多特征信息,严重影响模型的检测性能.因此,本文针对现存的问题进行相关研究和改进,主要涉及以下3个方面.

(1) 检测算法对于烟雾半透明特征(尤其是烟雾边缘区域)很难学习识别,因此本文使用非确定性的贝叶斯神经网络来对烟雾区域进行不确定性量化,进而校准模型学习方向,提升模型对烟雾特征的学习能力.

(2) 基于预处理思想,引入特征编码预处理,降低算法提取的特征编码中存在的干扰冗余,基于此增强算法的特征融合能力,进一步提高算法检测烟雾能力.

(3) 为优化检测算法上采样时重要信息的丢失,采用更先进、可学习的CARAFE上采样算子,增加算法在预测时的可利用信息以优化算法检测烟雾的性能.

1 相关工作

1.1 烟雾检测综述

现有方法依赖于深度卷积神经网络(CNN),并且主要关注以下一些内容:(1)多尺度预测,由于烟雾大小在图像中的变化,因此有学者将输入图像进行多尺

度变换以达到该目的^[1];(2)高阶-低阶水平特征融合,该方法本质上仍然要实现多尺度烟雾图像的预测,只不过将多尺度信息从多尺度图像转换为多尺度特征信息^[2];(3)扩大感受视野,有学者设计一个显著性检测网络,通过不同级别的特征融合来突出最具信息量的烟雾^[3].还有学者引入循环神经网络与分类代理来进一步细化^[4].还有学者提出Attention U-Net方法,用于融合粗层和细层特征进行多尺度预测^[5].还有学者依赖于ASPP网络^[6]结构来获得较大的感受视野^[7].

不同于现有方法利用语义分割架构和策略进行分割的方法,本文基于烟雾特殊的半透明特点,引入不确定性校准因子的损失函数来建模非刚性形状和半透明外观.

1.2 不确定量化综述

从统计学的角度来审视神经网络,一般有两种观点,一种是基于确定性的频率学派,另一种是基于不确定性的贝叶斯学派.前者将神经网络的学习过程视作求解参数的极大似然估计过程,最终得到确定性的参数大小.后者则认为参数的值不应当是确定的,而是一个概率分布,因此神经网络应当学习参数的概率分布,而不仅是一个确定的数值.

理论上讲,贝叶斯神经网络可以学习到参数的概率分布,意味着相比经典的确定性神经网络,贝叶斯神经网络有更好的泛化能力,可以适应不同复杂场景.

而实际上,求解参数的后验分布在以百万计的神经网络中是难以实现的,对此有关学者提出变分推断^[8]来解决该问题,因此我们可以得到一个新的目标函数:

$$L(\theta, \omega; x) = \mathbb{E}_{z \sim P_{\omega}(z|x)} [-\log_e (P_{\theta}(y|x, z))] + D_{\text{KL}}(P_{\omega}(z|x)|P(z)) \quad (1)$$

可以看到,该目标函数分为两部分,前者为预测损失项,后者为KL散度损失项.其中预测损失项含有积分,因此实际上很难求解.为解决该问题,我们可以使用蒙特卡洛采样来求解此积分,因此,预测损失项就可以表示为式(2):

$$P(y|x, D) \approx \frac{1}{B} \sum_{b=1}^B P(y|x, \theta_b) \quad (2)$$

烟雾图像经常在采样和标注时被噪声破坏,传统的确定性网络很容易给出确定性的,过于自信的判断,导致较高的假阳性和假阴性,使得网络的泛化能力较差,难以适应烟雾存在的多变和复杂环境,因此降低模

型的自信程度对于烟雾检测至关重要,引入不确定性的检测能力是实现该目标的基础^[8]。

本文基于 MobileNetV3 来输出不确定性特征,具体方法是将 MobileNetV3 中的第 4、11 和 17 层转换为 Bayesian 卷积层,使网络可以输出非确定性特征编码,基于非确定性特征编码得到最终烟雾预测图像。

1.3 轻量化特征编码器综述

从 SqueezeNet^[9]开始模型的参数量就不断下降,为进一步减少模型的实际操作数,MobileNets^[10]利用深度可分离卷积提高计算效率,而 MobileNetV2^[11]则加入线性 Bottleneck 和反转残差模块构成的高效基本模块。相较于 MobileNets 和 MobileNetV2, MobileNetV3^[12]较大的改变在于引入了 squeeze-and-excitation 结构的轻量级注意力模型,但其根本核心还是深度可分离卷积,包括 depthwise 卷积和 pointwise 卷积两个步骤来完成以前一次卷积完成的操作,depthwise 卷积在特征图每个通道上分别作卷积操作,pointwise 卷积在 depthwise 卷积之后对特征图的空间信息作卷积操作,两步骤卷积使得参数量得到大幅度降低,为模型可以在轻量化设备上运行创造了良好的条件。

1.4 上采样方法综述

编码器-解码器体系结构中上采样操作对于保留有用信息至关重要。最流行的上采样方法是双线性插值和转置卷积。双线性插值根据相邻像素的加权距离对特征图进行上采样,是一种不可学习的上采样方法。还有一个重要问题是,双线性插值只提高特征图的分辨率,而不改变特征图的深度。因此,双线性插值通常是通过一个卷积操作来改变特征映射的数量。

转置卷积是可学习的,对特征图进行上采样的方法。在转置卷积的过程中,特征映射中的每个元素都与内核中的每个元素倍增,最终的输出大小由转置卷积操作的步幅决定。例如,若转置卷积的步长为 2,那么 2×2 的输入特征图将转换为 4×4 的输出特征图。但是转置卷积的参数量巨大,会降低网络检测性能。

为此需要寻求一个速度较快且有效的上采样算子。CARAFE 算子^[13]则非常符合要求,其足够轻量化,可以满足速度的要求,且该算子上采样性能强大。CARAFE 解码主要分为两个步骤,第 1 个步骤是上采样核预测,第 2 个步骤是特征重组。其中上采样核预测的具体过程为:首先将输入的特征图使用卷积压缩其通道数;然后使用卷积层来生成上采样预测核,并且将该预测核

的通道维度在空间维度中展开;最后使用 Softmax 来生成最终的归一化上采样预测核。第 2 个特征重组则将输出的特征图和预测的上采样在局部区域作点积操作,得到最终的上采样特征图。

2 本文方法

在生成模型的领域当中,VAE^[14]极为重要,其最重要的成果在将随机性引入到特征编码当中。VAE 的核心思想是将编码的特征映射到有标准正态分布的随机变量上,而解码时则只需要将从标准正态分布中随机采样得到数据输入到解码器当中解码即可。

在烟雾的检测分割当中,烟雾模糊半透明的特点常会使得基于偏置归纳的确定性神经网络在不确定的区域上输出过于自信的特征编码(尤其是烟雾的边缘区域),使得特征编码过于受限,导致模型的泛化能力较低,产生大量的假阳性。因此,本文的核心思想基于网络不应该只是确定性的,而应该是确定性和不确定性的统一体。

受到 VAE 的启发,本文以 DeepLabV3+^[15]为基础,使 DeepLabV3+ 输出不确定性的编码特征,围绕不确定性特征这一核心,构建不确定性特征预处理模块以及解码模块,最终形成 UnCertainty_DeepLabV3+ 模型,整体架构如图 1 所示。

(1) 编码模块。如图 1 中 encoder 实线框所示。该模块实现不确定性特征编码,本文考虑到模型轻量化,在编码模块以 MobileNetV3 特征编码器为核心,将其第 4、11 和 17 层改造为 Bayesian 卷积层得到 Bayesian_MobileNetV3 (蓝色椭圆形框),实现不确定性特征编码输出,量化不确定性大小,基于该量化校准模型学习方向,提升模型学习烟雾模糊特征能力。

(2) 特征预处理模块。如图 1 中绿色椭圆框所示。该模块实现特征冗余降低以及特征融合。编码特征往往有着大量冗余信息,尤其对于烟雾这类单一目标的检测,可能存在大量背景干扰。为避免冗余特征对检测烟雾的影响,对于编码器得到的编码特征,本文引入注意力机制,分别对不同层次输出的不确定性特征进行处理,以期降低特征中存在的大量冗余信息,处理冗余之后融合不同层次特征,可以有效扩大感受视野,且可以建立不同层次特征之间的关系,有效提升网络的特征提取能力,减少信息量丢失,进而增强网络整体性能。

(3) 解码模块. 如图 1 中为 decoder 实线框所示. 该模块实现更快速且更有效的特征解码. 原 DeepLabV3+ 采用双线性插值上采样, 其优点是速度较快, 但其很明显的缺陷为不可学习, 因此在上采样倍数过大的

情况下结果并不理想. 为此需要寻求一个速度较快且有效的上采样算子. 而 CARAFE 算子 (黄色椭圆形框) 非常符合要求, 其足够轻量化可以满足速度要求, 且该上采样可学习, 因此有着较好的性能.

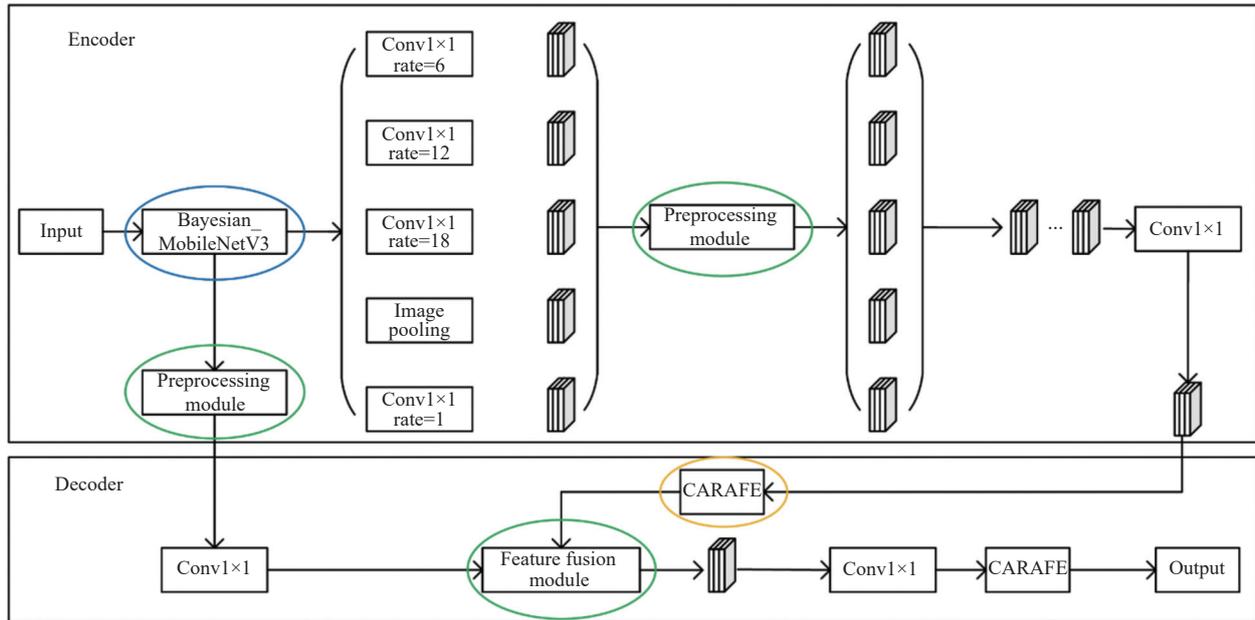


图 1 Uncertainty_DeepLabV3+

2.1 Bayesian_MobileNetV3

Bayesian_MobileNetV3 实现不确定性特征的编码. 同时, 本文也考虑到模型的轻量化, 因此在编码模块以 MobileNetV3 特征编码器为核心, 将其卷积层改造为 Bayesian 卷积层 (B_Bneck) 以实现输出不确定性的特征编码, 为实现不确定性量化提供基础. 然后基于不确定性的特征图, 校准模型的学习过程, 以提升模型学习烟雾特征及边界特征的能力. Bayesian_MobileNetV3 整体结构如表 1 所示, 其中 Input 表示输入数据的大小; Operator 表示该层对输入所作的卷积操作类型, 均基于 3×3 大小卷积核构建; EXP size 表示该卷积操作将输入数据升维后的维度大小; Out 表示该层的输出大小; SE 表示该层是否使用 SE 注意力模块; NL 表示该层所使用的激活函数类型, HS 表示 HardWish 激活函数, RE 表示 ReLU 激活函数; S 表示步长大小, S=2 时表示输出长宽为输入的一半, S=1 时则表示输出长宽和输入长宽是一致的. 从表 1 中可以看到卷积操作分为两种, 一种是确定性卷积操作 Bottleneck (Bneck), 另一种则是非确定性卷积操作 Bayesian_Bottleneck (B_Bneck), 整个网络是确定性和非确定性卷积的统一体.

表 1 Bayesian_MobileNetV3

Input	Operator	EXP size	Out	SE	NL	S
512 ² ×3	Conv2D	—	16	—	HS	2
256 ² ×3	Bneck	16	16	—	RE	1
256 ² ×3	Bneck	64	24	—	RE	2
128 ² ×3	B_Bneck	72	24	—	RE	1
128 ² ×3	Bneck	72	40	√	RE	2
64 ² ×3	Bneck	120	40	√	RE	1
64 ² ×3	Bneck	120	40	√	RE	1
64 ² ×3	Bneck	240	80	—	HS	2
32 ² ×3	Bneck	200	80	—	HS	1
32 ² ×3	Bneck	184	80	—	HS	1
32 ² ×3	Bneck	184	80	—	HS	1
32 ² ×3	B_Bneck	480	112	√	HS	1
32 ² ×3	Bneck	672	112	√	HS	1
32 ² ×3	Bneck	672	160	√	HS	2
16 ² ×3	Bneck	960	160	√	—	1
16 ² ×3	Bneck	960	160	√	HS	1
16 ² ×3	B_Bneck	—	960	—	—	1

MobileNetV3 转化为 Bayesian_MobileNetV3 的具体思路为: 首先, 考虑到将整个 MobileNetV3 的卷积层转化为 Bayesian 卷积层会使得到的模型参数量翻倍, 模型会变得异常复杂, 训练难度会大幅度上升, 鉴于此

本文采取局部卷积层转化的策略。其次,为得到模型在不同阶段提取到的低阶、中阶和高阶特征,以完成后续不同阶段的特征融合,因此 Bayesian 卷积层应当位于模型的不同阶段以编码不确定性的低阶、中阶和高阶特征;一般来讲,特征编码的尺寸越小,其所包含的抽象信息更加丰富,且通道数较大;反之其包含的具体信息越多;基于这个规则,本文需要的 Bayesian 卷积层将从 MobileNetV3 的输出特征编码尺寸为 128×128 、 32×32 、 16×16 这 3 类卷积层选取转化,以此得到不确定性的低阶、中阶和高阶特征编码,且选取的 128×128 卷积层通道数应尽可能少, 32×32 卷积层通道数应适中, 16×16 卷积层通道数应尽可能多。最后,实际应用需要模型尽可能在改进后依然轻量化,而通道数很大程度影响模型参数量,所以转化上述 3 类卷积层时应当考虑其输出的通道数大小,避免大通道数影响模型在实际场景下的性能表现。综合考虑这 3 个因素,本文将 MobileNetV3 第 4、11 和 17 层确定性卷积层转化为非确定性 B_Bneck, 得到 Bayesian_MobileNetV3。从表 1 看出,第 4 层输出的不确定性特征编码由 24 个通道构

成,尺寸大小为 128×128 ,通道数较少而尺寸较大,可以很好代表低阶特征编码;第 11 层输出的不确定性特征编码由 112 个通道构成,尺寸大小为 32×32 ,通道数和尺寸大小均较为适中,可以很好代表中阶特征编码;第 17 层输出的不确定性特征编码由 960 个通道构成,尺寸为 16×16 ,通道数较大而尺寸较小,可以很好代表高阶特征编码;其余层次则为确定性 Bottleneck 卷积算子。依靠不确定性 B_Bneck, Bayesian_MobileNetV3 可以实现不确定性特征编码的输出。

2.2 Feature_Preprocessing

特征预处理模块专注于降低特征冗余和特征融合。编码器得到的特征大多有大量冗余特征,尤其是烟雾这类单一目标的检测,可能存在大量背景干扰。为降低冗余特征对检测的影响,对于编码器得到的特征,本文引入注意力机制处理模型输出的不确定性特征,以期降低特征中存在的大量冗余特征。本模块具体由 3 种注意力模块组成,分别为通道注意力模块、空间注意力模块和兼具通道和空间注意力的 CBAM^[16],分别如图 2-图 4 所示。

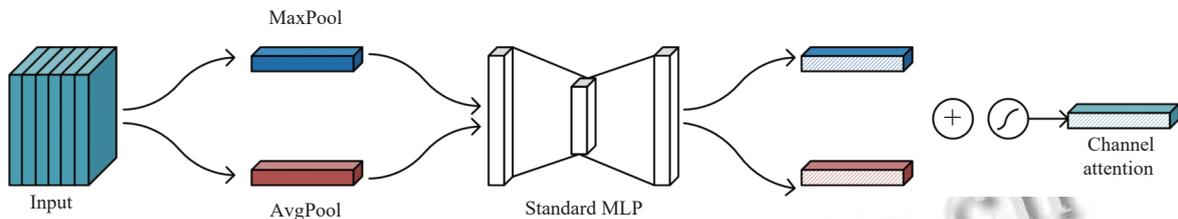


图 2 Channel attention

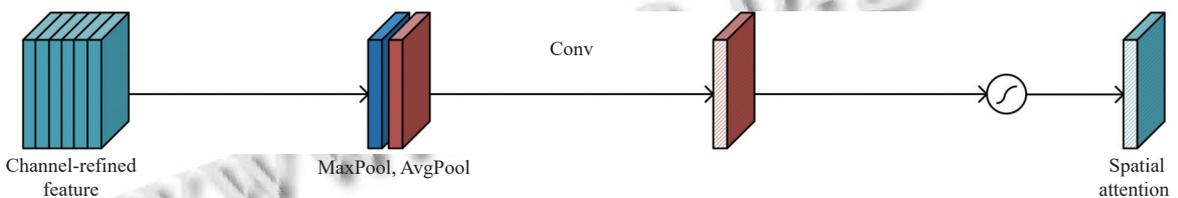


图 3 Spatial attention

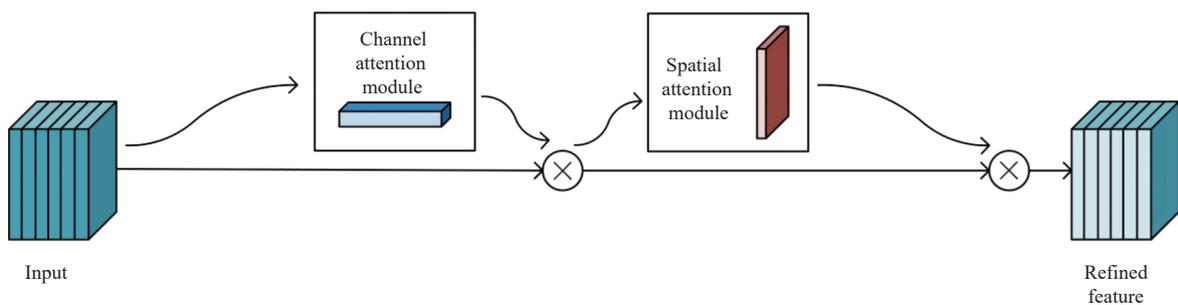


图 4 CBAM

具体来讲,第4层输出特征图通道数较少,但分辨率相对较大,因此使用空间注意力模块对其处理,目的是减少空间位置冗余特征;第17层输出特征图通道数较多,而分辨率相对较小,因此使用通道注意力模块对其处理,目的是减少通道冗余特征;第11层输出的特征图通道数和分辨率没有明显数量级差距,因此使用CBAM模块进行处理,目的是降低空间和通道冗余特征.具体过程为:将编码器提取的第4、11和17层不确定性特征分别输入到空间注意力模块、通道注意力模块和CBAM模块中,降低特征图空间和通道上与烟雾不相关的冗余特征,输出冗余度较低的特征图.进一步地,为有效扩大感受野且建立不同层次特征关系,在降低特征图冗余后,本文将融合这些低冗余特征,具体过程为首先将低冗余的第17层和第11层特征相融合得到新的输出特征,后将该新特征和低冗余的第4层特征融合得到需要预测解码的特征.

2.3 CARAFE_Upsampling

基于编码器-解码器体系结构的语义分割算法中,最终预测阶段必须进行上采样操作,若检测算法当中需要特征融合,那么也需要上采样操作的支持.可以看

到上采样操作在检测算法中必不可少,且由于其存在于检测算法的各个阶段,若上采样过程中丢失太多的信息,那么会较大影响算法最终的检测性能.

本文对算法上采样算子进行优化.其中原始DeepLabV3+采用双线性插值上采样,其最大优点是速度快,但缺陷是不可学习,且上采样倍数过大情况下信息丢失量巨大.为此需要寻求一个速度较快且保存较多信息的上采样算子.CARAFE算子非常符合要求,其足够轻量化,可以满足速度要求,且该上采样是可学习的,因此可以保存较多信息,其结构如图5所示.CARAFE上采样分两个步骤,第1个步骤是上采样核预测,第2个步骤是特征重组.其中上采样核预测具体过程为:首先将输入特征图用卷积压缩通道数,然后用卷积层来生成上采样预测核,并且将该预测核通道维度在空间维度中展开,最后使用Softmax来生成最终的归一化的上采样预测核.特征重组具体过程则将输出的特征图和预测的上采样核在局部区域作点积操作,得到最终的上采样图.上采样操作CARAFE在重组时可以有较大的感受野,会根据输入特征来指导重组过程,同时整个算子比较轻量级.

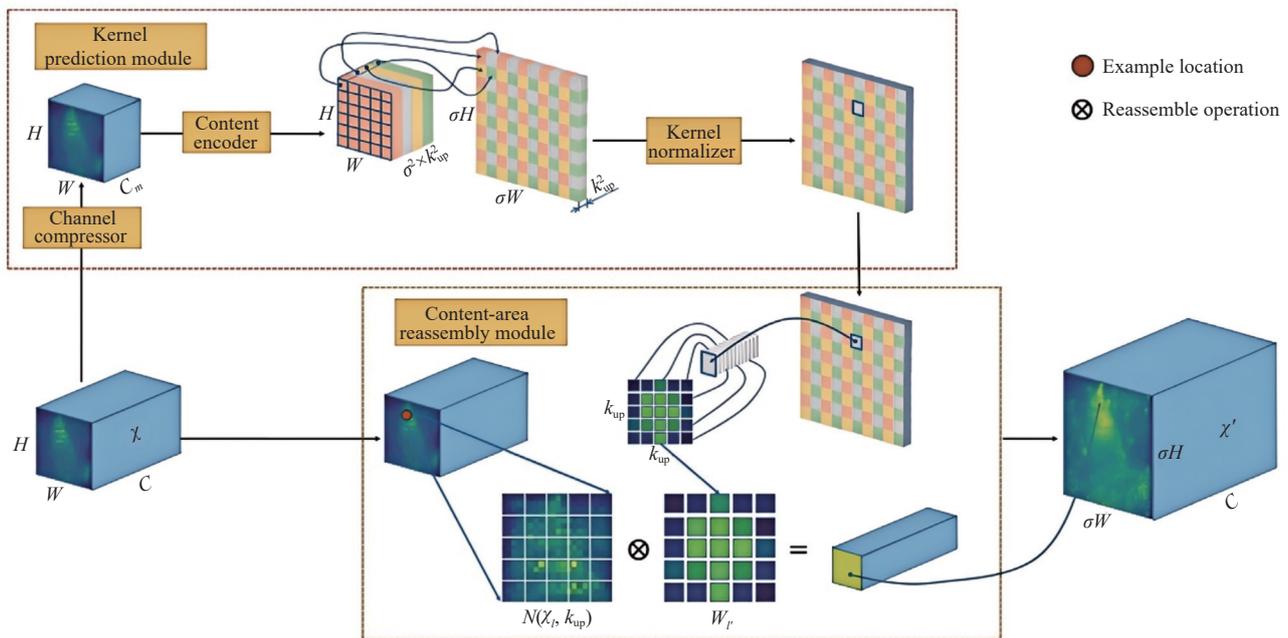


图5 CARAFE

2.4 Uncertainty_Calibration_Loss Function

最终的不确定性输出可以实现不确定性大小的量化,本文使用预测熵来量化不确定性的大小^[17-20],预测熵的计算公式为:

$$H(X) = - \sum_{x \in X} P(x) \log_2 P(x) \quad (3)$$

预测熵的可视化描述如图6,可以看到,当某个像素点的预测概率值为0.5时,该像素点的不确定大小达

到最大值 1. 本文取不确定性阈值 $\mu=0.8$, 当某个像素点的不确定性大小大于该阈值时, 说明算法对该像素点并不能很好地认知, 此时将赋予更大的损失值在该像素点上, 此过程可以校准算法对不确定性区域的学习, 提升模型的性能和泛化能力.

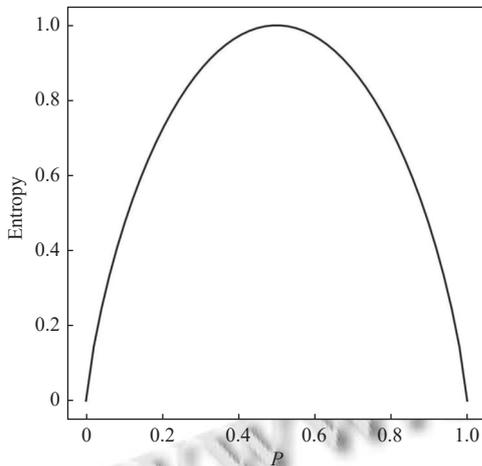


图6 Predictive entropy

由于高分辨率下对每个像素计算不确定性大小计算量巨大, 因此本文采用局部不确定性大小的量化, 具体思路为: 将得到的预测标签图划分为 4 个大小相同的区域, 计算每个区域不确定性大小的平均值; 设定阈值 $\mu=0.8$, 若某个区域的不确定性大小平均值大于阈值 μ , 那么说明网络对该区域所作出的预测具有较大的不确定性, 则对该区域的损失赋予一定大小的权重 λ_i , 其中 i 表示分割的区域数量. 因此本文最终需要优化的目标函数形式为:

$$\begin{aligned} Loss = & \lambda_1 L(\theta, \omega|x) + \lambda_2 L(\theta, \omega|x) + \lambda_3 L(\theta, \omega|x) \\ & + \lambda_4 L(\theta, \omega|x) \end{aligned} \quad (4)$$

3 实验分析

本文训练时所使用实验系统环境为 Windows 11, 硬件平台 CPU 为 Intel 酷睿 I9-13900HK, 32 GB 内存, GPU 为 GeForce GTX4070, 8 GB 显存, 使用的深度学习框架为 PyTorch 2.2.0, CUDA 11.8, CUDNN 8.9.7.

3.1 数据集和评价指标

本文将在 SMOKE5K 数据集上进行充分实验以证明模型的有效性. SMOKE5K 数据集包含 5 400 张图像, 包括真实烟雾场景和合成烟雾场景, 该数据集较为庞大, 因此模型在该数据集上可以很好地学习烟雾的特征编码信息, 且模型在大容量数据集上不容易产生过拟合现象, 也使得学习到的模型可以更好地适应实

际中存在的不同烟雾检测环境.

在模型进行训练之前, 需要将数据集划分得到训练集和验证集. 数据集的划分包括随机数种子的设置、训练集和验证集比例的设置; 随机数种子设置为 0, 以保证每次划分的训练集和验证集相同, 避免训练集和验证集发生变化导致实验不可复现. 训练集和验证集的比例设置为 9:1.

SMOKE5K 数据集中所有图像都有其对应的真值标签, 其标签像素为 0 和 255, 其中背景非烟雾区域为 0, 烟雾区域为 255. 由于 DeepLabV3+ 训练时使用的数据集格式为 VOC 格式, 在 VOC 格式中, 标签背景像素值为 0, 而标签前景像素值从 1 开始, 因此需要将 SMOKE5K 数据集真值标签中的前景像素值从 255 转化为 1.

评价指标为模型参数量、模型浮点计算量及模型的平均交并比. 其中模型平均交并比最为重要. 模型参数量和浮点计算量衡量了模型的计算复杂度, 平均交并比定义了两个集合真实值与预测值的交集和并集重合比例, 用来衡量模型的真实性能和泛化能力, 公式为:

$$MIoU = \frac{1}{K+1} \sum_{H=0}^K \frac{P_{HH}}{\sum_{J=0}^K P_{HJ} + \sum_{J=0}^K P_{JH} - P_{HH}} \quad (5)$$

其中, K 表示除背景以外的语义类别总数; P_{HJ} 是指原本属于第 H 类却被分为第 J 类的像素数量; P_{JH} 是指原本属于第 J 类却被分为第 H 类的像素数量; P_{HH} 是指原本属于第 H 类且仍被分为第 H 类的像素数量; 该指标是衡量语义分割性能中最重要的指标.

3.2 训练细节

首先是数据集的 Resize 和增强. Resize 可以使所有图像大小保持一致. 这样不仅可以降低模型复杂度, 而且直接影响模型对多尺度特征信息的提取. 图像增强使用随机图像旋转、翻转和高斯模糊来增加数据集多样性. 图像增强方法可以提升模型泛化能力已经成为共识, 由于烟雾数据集的特殊性, 其较难标注且数量稀少, 因此烟雾数据集的图像增强是必不可少的.

其次是训练数据的模型选择. 本文所使用的特征编码器是 ImageNet 数据集上预训练的 MobileNetV3. 使用在大规模数据集上预训练的模型不仅可以有效地加快模型在具体下游任务上的收敛速度, 更重要的是没有在大规模数据集上预训练的初始化模型在具体任务上进行训练时几乎难以收敛, 因此使用预训练模型是非常重要的环节, 也是迁移学习的核心思想.

最后是训练数据时所需要的优化器及优化策略选

择. 在确定所使用的模型之后, 模型的优化器及优化策略选择很大程度上影响模型的收敛方向, 好的优化器和优化策略可以使得模型向更好的方向优化, 避免陷入局部最优解. 本文所选用的优化器为带动量的 SGD 优化器, 相比于 SGD 优化器, 其引入动量因子来预防模型陷入局部最优解. 为了进一步避免模型陷入局部最优解, 本文使用了额外的学习率优化策略, 即经典的 Poly 学习策略, 该优化策略可以随学习的轮次对基础学习率进行动态赋值, 使得学习率的调整不再受到人为因素的影响. 具体表达式为:

$$LR = BASE_LR \times \left(1 - \frac{ITER}{MAX_ITER}\right)^{POWER} \quad (6)$$

其中, LR 表示当前迭代轮次的学习率大小; $BASE_LR$ 为初始学习率, 设为 0.001; MAX_ITER 表示最大迭代轮次, 设为 500. $POWER$ 为衰减指数, 设为 0.9.

3.3 模型性能分析和消融实验

本节实验分为 3 部分, 第 1 部分实验验证 MobileNetV3 编码器、特征处理模块和 CARAFE 上采样算子的有效性, 该部分并未将 MobileNetV3 编码器转换为 Bayesian_MobileNetV3 编码器. 第 2 部分实验验证 Bayesian_MobileNetV3 编码器相比 MobileNetV3 编码器的有效性. 第 3 部分实验分析本文模型与其他经典语义分割模型在 SMOKE5K 数据集上的表现.

3.3.1 消融实验 1

本节为第 1 部分实验, 验证 MobileNetV3 编码器、特征处理模块和 CARAFE 上采样算子的有效性. 本实验中优化原 DeepLabV3+ (Baseline) 编码器为 MobileNetV3, 引入特征处理模块和 CARAFE 上采样算子, 此实验 MobileNetV3 没有转换为 Bayesian_MobileNetV3, 是确定性神经网络, 记该模型为 Certain_DeepLabV3+. 将该模型与 Baseline (Xception) 和 Baseline (ResNet101) 进行比较, 括号内为原 DeepLabV3+ 所用编码器. 图 7 和表 2 分别展示了不同模型的检测结果和性能. 分析图 7 可以看出 Certain_DeepLabV3+ 相比较于 Baseline (Xception) 和 Baseline (ResNet101) 有较为明显提升. 分析表 2 可以得到结论: 具有 MobileNetV3 编码器、特征处理模块以及 CARAFE 上采样算子的 Certain_DeepLabV3+ 相较于 Baseline 有所提升. 单独引入特征预处理模块, $MIoU$ 提升到 91.54%; 单独改进上采样模块, $MIoU$ 提升到 91.49%; 可以看到仅改进上采样模块时提升较小, 但是结合特征预处理模块则有较大的提升, 说明两者级联使用效果更好.

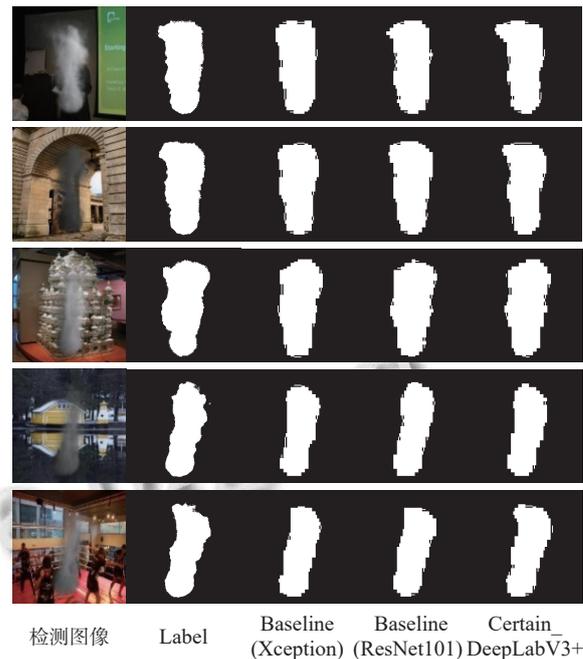


图 7 检测的烟雾结果图

表 2 消融实验 1

模型	参数量 (M)	浮点计算量 (G)	$MIoU$ (%)
Baseline (Xception)	54.709	166.841	89.90
Baseline (ResNet101)	58.037	89.862	90.86
Baseline (MobileNetV3)	11.725	25.515	91.43
Baseline (MobileNetV3)+Fp	11.801	27.153	91.54
Baseline (MobileNetV3)+CARAFE	12.236	33.232	91.49
Certain_DeepLabV3+	12.312	34.870	91.77

3.3.2 消融实验 2

本节为第 2 部分实验, 该实验主要目的在于验证 Bayesian_MobileNetV3 的有效性, 因此本节实验将 Certain_DeepLabV3+ 中的 MobileNetV3 编码器替换为 Bayesian_MobileNetV3 编码器, 并且记该检测模型为 Uncertain_DeepLabV3+. 引入 Bayesian_MobileNetV3 的主要原因在于该编码器可以实现不确定性大小的量化, 引入 Bayesian_MobileNetV3 后可以预测特征图中各个区域甚至像素点的不确定性大小, 然后基于不确定性大小来优化模型学习方向. 因此, 验证 Bayesian_MobileNetV3 有效性的关键在于验证目标函数中不确定性项校准的有效性, 即式 (4) 中超参数 λ_i 的有效性. 图 8 和表 3 分别为不同模型的检测结果和性能. 从烟雾检测结果图可以看到, 相较于 Certain_DeepLabV3+, Uncertain_DeepLabV3+ 的检测效果有所提升, 且对于烟雾的边缘有着更好的处理. 从表 3 中可以看出: 在引入不确定性因素之后, 依靠包含不确定性参数校准的

损失函数, 模型的检测性能获得了一定提升, 参数量稍有变大, 但是相对于该提升而言付出的代价较小. 实验证明 Bayesian_MobileNetV3 相比于 MobileNetV3 有着更好的表现. 另一方面也充分说明对于烟雾这类具有模糊、半透明特点的特殊对象而言, 使用不确定性来校准模型是非常有效的, 可以避免模型对不确定性的烟雾区域 (尤其是烟雾边缘区域) 作出过于自信的预测, 从而得到更好的预测结果.

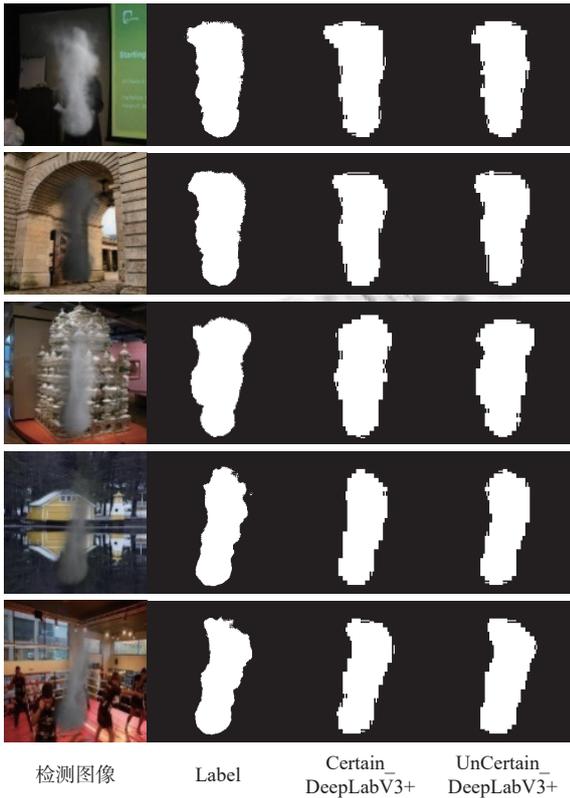


图8 确定性与不确定性模型检测结果对比

表3 消融实验2

模型	参数量 (M)	浮点计算量 (G)	<i>MIoU</i> (%)
UnCertain_DeepLabV3+	12.665	36.653	92.41
Certain_DeepLabV3+	12.312	34.870	91.77

3.3.3 与其他模型进行比较

为进一步评估 UnCertain_DeepLabV3+ 的有效性, 本节将 UnCertain_DeepLabV3+ 与其他经典语义分割网络在 SMOKE5K 数据集上比较, 包括 FCN、FCNV2、DenseASPP、PSPNet、CGNet、BiSeNet 和 ICNet 等. 图9和表4分别为不同模型检测结果和性能. 可以看出本文方法 UnCertain_DeepLabV3+ 在所有模型中综合性能最好. 图9表明 UnCertain_DeepLabV3+ 可以更好地检测烟雾区域, 且对烟雾边缘的检测效果更加出色. 表4显示, UnCertain_DeepLabV3+ 在重要指标 *MIoU*

上达到 92.41%, 相比于其他模型有较为明显提升. 综上, 该对比实验进一步验证了本文对于具体烟雾检测任务所提出改进的有效性, 尤其是将不确定性校准应用于烟雾检测分割中的有效性, 也充分说明了不确定性校准在烟雾这类特殊目标的检测任务中具有重要作用.

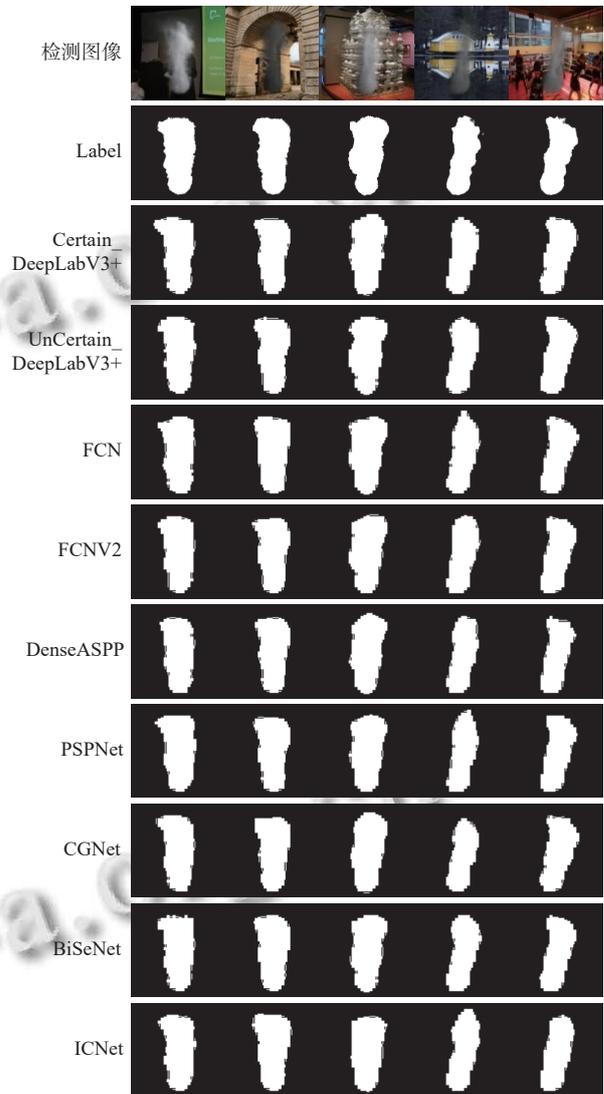


图9 与其他经典语义分割模型检测结果对比

表4 对比实验

模型	参数量 (M)	浮点计算量 (G)	<i>MIoU</i> (%)
Certain_DeepLabV3+	12.312	34.870	91.77
UnCertain_DeepLabV3+	12.665	36.653	92.41
FCN ^[1]	30.021	641.731	89.02
FCNV2	35.308	52.839	90.00
DenseASPP ^[7]	9.173	31.397	89.87
PSPNet ^[21]	46.583	52.946	90.71
CGNet ^[22]	491.956	7.110	89.10
BiSeNet ^[23]	12.795	26.094	90.61
ICNet ^[24]	26.119	20.343	90.67

4 结论与展望

本文提出一种基于不确定性校准的烟雾语义分割算法。该语义分割算法核心思想在于特征编码器输出的非确定性特征编码,围绕非确定性特征编码这一核心,进一步构建特征处理模块以期降低特征编码器输出特征的冗余度,提高编码特征的利用率,在降低特征图冗余后,本文将融合这些低冗余特征,有效扩大感受野并建立不同层次特征关系,之后使用可学习的 CARAFE 上采样算子,减少上采样过程中的信息丢失率,进一步提升算法的检测性能。最后,在目标函数当中引入不确定性校准项,指导和强化算法在非确定性区域的学习。本文算法在公开 SMOKE5K 数据集上进行了大量对比实验以验证本文提出的烟雾分割算法的有效性,实验结果可以表明,改进后的模型在关键指标 *MIoU* 上达到 92.41%,并且检测结果显示,相比于其他的检测分割算法,本文算法性能最优。

参考文献

- Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015. 3431–3440.
- Ding H, Jiang X, Shuai B, *et al.* Context contrasted feature and gated multi-scale aggregation for scene segmentation. Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 2393–2402.
- Xu G, Zhang Y, Zhang Q, *et al.* Video smoke detection based on deep saliency network. Fire Safety Journal, 2019, 105: 277–285.
- Yuan F, Zhang L, Xia X, *et al.* A gated recurrent network with dual classification assistance for smoke semantic segmentation. IEEE Transactions on Image Processing, 2021, 30: 4409–4422.
- Zhang J, Zhu H, Wang P, *et al.* ATT squeeze U-Net: A lightweight network for forest fire detection and recognition. IEEE Access, 2021, 9: 10858–10870.
- Li X, Chen Z, Wu QMJ, *et al.* 3D parallel fully convolutional networks for real-time video wildfire smoke detection. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 30(1): 89–103.
- Yang M, Yu K, Zhang C, *et al.* DenseASPP for semantic segmentation in street scenes. Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 3684–3692.
- Yan SY, Zhang J, Barnes N. Transmission-guided Bayesian generative model for smoke segmentation. Proceedings of the 36th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI, 2022. 3009–3017.
- Iandola FN, Han S, Moskewicz MW, *et al.* SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size. arXiv:1602.07360, 2016.
- Howard AG, Zhu ML, Chen B, *et al.* MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv:1704.04861, 2017.
- Sandler M, Howard A, Zhu ML, *et al.* MobileNetV2: Inverted residuals and linear bottlenecks. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 4510–4520.
- Howard A, Sandler M, Chen B, *et al.* Searching for MobileNetV3. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019. 1314–1324.
- Wang JQ, Chen K, Xu R, *et al.* CARAFE: Content-aware reassembly of features. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019. 3007–3016.
- Kingma DP, Welling M. Auto-encoding variational Bayes. arXiv:1312.6114v10, 2014.
- Chen LC, Zhu YK, Papandreou G, *et al.* Encoder-decoder with atrous separable convolution for semantic image segmentation. Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 833–851.
- Woo S, Park J, Lee JY, *et al.* CBAM: Convolutional block attention module. Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 3–19.
- Gawlikowski J, Tassi CRN, Ali M, *et al.* A survey of uncertainty in deep neural networks. Artificial Intelligence Review, 2023, 56(Suppl 1): 1513–1589.
- Kendall A, Gal Y. What uncertainties do we need in Bayesian deep learning for computer vision? Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 5580–5590.
- Depeweg S, Hernández-Lobato JM, Doshi-Velez F, *et al.* Uncertainty decomposition in Bayesian neural networks with latent variables. arXiv:1706.08495, 2017.
- Gal Y, Ghahramani Z. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. Proceedings of the 33rd International Conference on Machine Learning. New York: JMLR.org, 2016. 1050–1059.
- Zhao H, Shi J, Qi X, *et al.* Pyramid scene parsing network. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 2881–2890.
- Wu T, Tang S, Zhang R, *et al.* CGNet: A light-weight context guided network for semantic segmentation. arXiv:1811.08201v2.
- Yu C, Wang J, Peng C, *et al.* BiSeNet: Bilateral segmentation network for real-time semantic segmentation. Proceedings of the 2018 European Conference on Computer Vision (ECCV). Munich: Springer, 2018. 325–341.
- Zhao H, Qi X, Shen X, *et al.* ICNet for real-time semantic segmentation on high-resolution images. Proceedings of the 2018 European Conference on Computer Vision (ECCV). Munich: Springer, 2018. 405–420.

(校对责编:王欣欣)