



(12)发明专利申请

(10)申请公布号 CN 111582316 A

(43)申请公布日 2020.08.25

(21)申请号 202010281081.2

(22)申请日 2020.04.10

(71)申请人 天津大学

地址 300072 天津市南开区卫津路92号

(72)发明人 雷建军 祝新鑫 范晓婷 石雅南  
李奕

(74)专利代理机构 天津市北洋有限责任专利代  
理事务所 12201

代理人 李林娟

(51) Int. Cl.

G06K 9/62(2006.01)

G06K 9/46(2006.01)

G06N 3/04(2006.01)

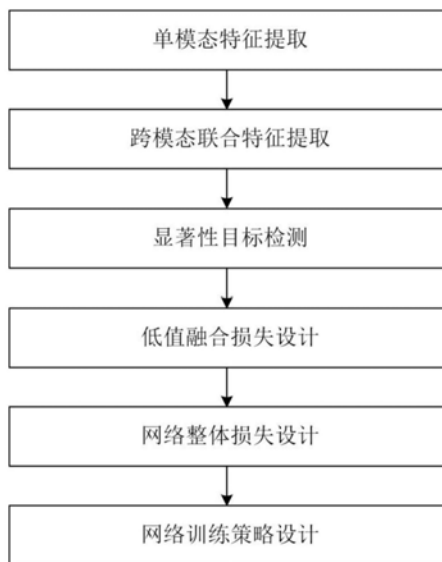
权利要求书1页 说明书6页 附图4页

(54)发明名称

一种RGB-D显著性目标检测方法

(57)摘要

本发明公开了一种RGB-D显著性目标检测方法,所述方法包括以下步骤:通过单模态特征提取,分别获取RGB图和深度图的每一级的单模态显著性特征;通过跨模态联合特征提取对每一级的RGB和深度单模态显著性特征进行选择,逐级捕获RGB图和深度图的互补特征,生成跨模态联合特征;将跨模态联合特征和单模态显著性特征输入到显著性目标检测部分;通过设计低值融合损失和网络整体损失,实现RGB流和深度流检测结果的融合和网络的监督学习,输出最终的显著性检测结果。本发明通过跨模态联合特征提取捕获有价值的跨模态联合特征,并通过设计的低值融合损失使网络关注显著性图的低值显著性区域,提高显著性值的下界。



1. 一种RGB-D显著性目标检测方法,其特征在于,所述方法包括以下步骤:

通过单模态特征提取,分别获取RGB图和深度图的每一级的单模态显著性特征;

通过跨模态联合特征提取对每一级的RGB和深度单模态显著性特征进行选择,逐级捕获RGB图和深度图的互补特征,生成跨模态联合特征;

将跨模态联合特征和单模态显著性特征输入到显著性目标检测部分;

通过设计低值融合损失和网络整体损失,实现RGB流和深度流检测结果的融合和网络的监督学习,输出最终的显著性检测结果。

2. 根据权利要求1所述的一种RGB-D显著性目标检测方法,其特征在于,所述单模态特征提取基于VGG-16模型,包含五组卷积块,每组卷积块由相应的卷积层和池化层组成;

去掉第五组卷积块中最后一层池化层,并在后面加入了四个并行的空洞卷积捕获全局多感受野信息,四个空洞卷积输出的特征进行级联,组成第五组卷积块的输出。

3. 根据权利要求1所述的一种RGB-D显著性目标检测方法,其特征在于,所述跨模态联合特征提取具体为:

在单模态特征提取过程中,先对RGB和深度单模态显著性特征进行空间特征选择,然后结合形成跨模态联合特征表示;跨模态特征表示经过通道注意力模块重新加权,生成加权后的RGB-D特征;加权后的RGB-D特征、RGB特征和深度特征进行结合,生成每一级的跨模态联合特征。

4. 根据权利要求3所述的一种RGB-D显著性目标检测方法,其特征在于,所述跨模态联合特征具体为:

$$Y_C^i = Up\left(Conv\left(Y_W^i + X_R^i + X_D^i\right)\right)$$

其中,Up(·)表示利用双线性插值的上采样操作, $Y_C^i$ 即为跨模态联合特征提取模块的输出, $Y_W^i$ 为经过一个通道注意力模块产生加权后的RGB-D特征, $X_R^i$ 为RGB特征, $X_D^i$ 为深度特征,Conv(·)表示卷积操作。

5. 根据权利要求1所述的一种RGB-D显著性目标检测方法,其特征在于,所述显著性目标检测部分包括RGB和深度显著性目标检测,两部分采用相同的结构,

每一部分都设计了五个融合块为了整合单模态的多尺度特征和跨模态的联合特征,提高每流显著性目标检测的精度;

单模态的多尺度特征由两个卷积层和上采样层捕获,在每个融合块中,每级捕获的多尺度特征和联合特征通过通道级联和卷积进行整合,生成更新后的显著性特征。

6. 根据权利要求1所述的一种RGB-D显著性目标检测方法,其特征在于,所述低值融合损失具体为:

$$L_{low-bound-fusion} = \frac{1}{2N} \sum_{k=1}^N \left\| \min(P_k^R, P_k^D) - G_k \right\|_2^2$$

其中,k表示像素的索引, $P_k^R$ 和 $P_k^D$ 分别表示RGB流和深度流预测的显著性值, $G_k$ 表示真值, $\|\cdot\|_2$ 为2范数,N为图像的像素个数。

## 一种RGB-D显著性目标检测方法

### 技术领域

[0001] 本发明涉及图像处理、立体视觉技术领域,尤其涉及一种RGB-D显著性目标检测方法。

### 背景技术

[0002] 在面对复杂的自然场景时,人类视觉系统具有快速搜索并定位感兴趣区域和目标的能力。通过将视觉注意力机制引入到计算机视觉中,可以优化计算资源,使视觉信息的处理更符合人眼视觉特征。显著性目标检测旨在自动识别不同场景中的显著性区域,已广泛地应用于分割、重定向、检索、编码、分类等任务。近年来,图像显著性目标检测取得了令人瞩目的进展。实际上,人类视觉系统还具有感知场景中深度信息的能力,深度信息可以作为彩色图的互补信息,应用到显著性目标检测任务中。随着相机设备的快速发展,深度图的获取变得越来越便捷,促进了RGB-D显著性目标检测的研究。

[0003] 传统的RGB-D显著性目标检测的方法将深度信息作为彩色信息的补充,通过设计手工特征来捕获信息。Peng等人将RGB-D图像对作为四通道的输入,送入多阶段显著性推理模型。Song等人结合四通道的数据和传统技术计算多尺度的显著性值。Liang等人利用对比度和深度背景先验知识,提出了立体显著性检测模型。然而,这些算法受限于无法有效地捕获图像的高级语义特征,并且直接将不同模态的特征进行整合,可能会造成特征不匹配或带来冗余信息。

[0004] 最近,基于卷积神经网络(Convolutional Neural Network, CNN)的方法成为RGB-D显著性检测任务的主流。与传统的方法相比,基于CNN的方法能够提取并融合高级语义特征和低级纹理特征。大多数基于CNN的方法采用两流架构,以自底向上的方式分别提取RGB和深度特征,然后集成到后处理步骤中进行融合。Han等人提出了一个双流网络来提取RGB和深度特征,然后利用全连接层整合这些特征预测最终的显著性图。Ding等人利用两个不同的CNN来分别预测RGB流和深度流的显著性图,并在显著性融合网络中进行融合。

[0005] 发明人在实现本发明的过程中,发现现有技术中至少存在以下缺点和不足:

[0006] 首先,现有技术大多利用两个独立的显著性检测模型来提取单模态的特征,这种方式忽略了RGB特征和深度特征之间的互补性。因此,如何在特征提取的过程中逐步结合跨模态特征仍需要进一步探索。其次,在后期融合策略中,现有技术大多通过像素级相加或相乘的方式,融合RGB流和深度流的显著性预测,这种简单的融合方式没有充分利用两流的预测,无法有效地突出完整的显著性目标、抑制背景区域。

### 发明内容

[0007] 本发明提供了一种RGB-D显著性目标检测方法,通过跨模态联合特征提取捕获有价值的跨模态联合特征,并通过设计的低值融合损失使网络关注显著性图的低值显著性区域,提高显著性值的下界,详见下文描述:

[0008] 一种RGB-D显著性目标检测方法,所述方法包括以下步骤:

[0009] 通过单模态特征提取,分别获取RGB图和深度图的每一级的单模态显著性特征;

[0010] 通过跨模态联合特征提取对每一级的RGB和深度单模态显著性特征进行选择,逐级捕获RGB图和深度图的互补特征,生成跨模态联合特征;

[0011] 将跨模态联合特征和单模态显著性特征输入到显著性目标检测部分;

[0012] 通过设计低值融合损失和网络整体损失,实现RGB流和深度流检测结果的融合和网络的监督学习,输出最终的显著性检测结果。

[0013] 其中,所述单模态特征提取基于VGG-16模型,包含五组卷积块,每组卷积块由相应的卷积层和池化层组成;

[0014] 去掉第五组卷积块中最后一层池化层,并在后面加入了四个并行的空洞卷积捕获全局多感受野信息,四个空洞卷积输出的特征进行级联,组成第五组卷积块的输出。

[0015] 进一步地,所述跨模态联合特征提取具体为:

[0016] 在单模态特征提取过程中,先对RGB和深度单模态显著性特征进行空间特征选择,然后结合形成跨模态联合特征表示;跨模态特征表示经过通道注意力模块重新加权,生成加权后的RGB-D特征;加权后的RGB-D特征、RGB特征和深度特征进行结合,生成每一级的跨模态联合特征。

[0017] 其中,所述跨模态联合特征具体为:

$$[0018] \quad Y_C^i = \text{Up}\left(\text{Conv}\left(Y_W^i + X_R^i + X_D^i\right)\right)$$

[0019] 其中,Up(·)表示利用双线性插值的上采样操作, $Y_C^i$ 即为跨模态联合特征提取模块的输出, $Y_W^i$ 为经过一个通道注意力模块产生加权后的RGB-D特征, $X_R^i$ 为RGB特征, $X_D^i$ 为深度特征,Conv(·)表示卷积操作。

[0020] 进一步地,所述显著性目标检测部分包括RGB和深度显著性目标检测,两部分采用相同的结构,

[0021] 每一部分都设计了五个融合块为了整合单模态的多尺度特征和跨模态的联合特征,提高每流显著性目标检测的精度;

[0022] 单模态的多尺度特征由两个卷积层和上采样层捕获,在每个融合块中,每级捕获的多尺度特征和联合特征通过通道级联和卷积进行整合,生成更新后的显著性特征。

[0023] 其中,所述低值融合损失具体为:

$$[0024] \quad L_{low-bound-fusion} = \frac{1}{2N} \sum_{k=1}^N \left\| \min(P_k^R, P_k^D) - G_k \right\|_2^2$$

[0025] 其中,k表示像素的索引, $P_k^R$ 和 $P_k^D$ 分别表示RGB流和深度流预测的显著性值, $G_k$ 表示真值, $\|\cdot\|_2$ 为2范数,N为图像的像素个数。

[0026] 本发明提供的技术方案的有益效果是:

[0027] 1、本发明通过设计跨模态联合特征提取部分和显著性目标检测部分,有效地整合了RGB图像和深度图的互补信息,提高了每流显著性检测的精度;

[0028] 2、本发明通过设计低值融合损失,有效建立了两流显著性检测之间的关系,使网络关注显著性值较低的区域,提高显著性值的下界,本发明取得了良好的显著性目标检测性能。

## 附图说明

- [0029] 图1为一种RGB-D显著性目标检测方法的流程图；  
[0030] 图2为一种RGB-D显著性目标检测方法的具体实施框架；  
[0031] 图3为跨模态特征提取模块 (CFM) 和RGB显著性检测部分中融合块 (FB) 的结构图；  
[0032] 图4为检测结果的示意图。

## 具体实施方式

[0033] 为使本发明的目的、技术方案和优点更加清楚，下面对本发明实施方式作进一步地详细描述。

### [0034] 实施例1

[0035] 本发明实施例提出了一种基于跨模态联合特征提取和低值融合损失的RGB-D显著性目标检测方法。通过设计跨模态联合特征提取部分，有效地捕获了RGB特征和深度特征之间的互补性；通过设计显著性目标检测部分，有效地整合了单模态的多尺度特征和跨模态的联合特征，提高了每流显著性检测的精度；通过设计低值融合损失，有效地提高显著性值的下界，促进了不同检测结果之间的融合。

[0036] 整个流程分为六部分：1) 单模态特征提取；2) 跨模态联合特征提取；3) 显著性目标检测；4) 低值融合损失设计；5) 网络整体损失设计；6) 网络训练策略设计，具体步骤如下：

### [0037] 一、单模态特征提取

[0038] 单模态特征提取部分由RGB特征提取和深度特征提取组成，由于RGB图像和深度图中具有相同的显著性目标，两部分采用相同的结构。特征提取部分基于VGG-16模型，包含五组卷积块，每组卷积块由相应的卷积层和池化层组成。

[0039] 为了更好地适应该任务，该部分去掉了第五组卷积块中最后一层池化层，并在后面加入了四个并行的空洞卷积来捕获全局多感受野信息，四个空洞卷积输出的特征进行级联，组成第五组卷积块的输出。在本发明中，四个空洞卷积的空洞率分别被设置为1、2、4、6，每个空洞卷积输出特征的尺寸（高度×宽度×通道数）为 $14 \times 14 \times 128$ 。输入图像的大小为 $224 \times 224$ ，五组卷积块输出特征的尺寸依次为 $112 \times 112 \times 64$ 、 $56 \times 56 \times 128$ 、 $28 \times 28 \times 256$ 、 $14 \times 14 \times 512$ 、 $14 \times 14 \times 512$ 。以下，将RGB特征提取部分中每组卷积块的输出记为 $X_R^i (i=1,2,\dots,5)$ ，深度特征提取部分中每组卷积块的输出记为 $X_D^i$ 。

### [0040] 二、跨模态联合特征提取

[0041] 跨模态联合特征提取部分由设计的五个跨模态联合特征提取模块 (CFM) 组成，其目的是为了在单模态特征提取过程中逐步捕获RGB图像和深度图的互补信息，生成每一级的联合特征。在跨模态联合特征提取模块中，RGB特征和深度特征首先进行空间特征选择，然后再结合形成跨模态特征表示。为了强调显著性特征的重要性，加入了现有的通道注意力模块（该模块为本领域技术人员所公知，本发明实施例对此不做赘述）来加权不同通道上的响应。最后，加权后的特征、RGB特征和深度特征结合在一起，通过卷积和上采样操作形成最终的跨模态联合特征。与此同时，各级联合特征又相继送入到单模态的显著性目标检测部分，为单模态的显著性检测提供有价值的互补信息。

### [0042] 1) 单模态特征的选择

[0043] 每个CFM  $i$ 模块的结构如图3(a)所示，每个模块都有2个输入，1个输出，两个输入

分别为每一级对应的RGB特征  $X_R^i$  和深度特征  $X_D^i$ 。

[0044] 即,在跨模态联合特征提取模块中,首先将RGB特征和深度特征沿着通道分为J组,每组特征分别表示为  $X_{R_j}^i$  和  $X_{D_j}^i$ ;然后在每组通道上进行分组卷积,得到RGB流和深度流经过选择的特征  $Y_{R_j}^i$  和  $Y_{D_j}^i$ ,其表示为:

$$[0045] \quad Y_{R_j}^i = \begin{cases} X_{R_j}^i & j=1 \\ \text{Conv}(X_{R_j}^i) & j=2 \\ \text{Conv}(X_{R_j}^i + Y_{R_{j-1}}^i) & 2 < j \leq J \end{cases} \quad (1)$$

[0046] 和

$$[0047] \quad Y_{D_j}^i = \begin{cases} X_{D_j}^i & j=1 \\ \text{Conv}(X_{D_j}^i) & j=2 \\ \text{Conv}(X_{D_j}^i + Y_{D_{j-1}}^i) & 2 < j \leq J \end{cases} \quad (2)$$

[0048] 其中,Conv(•)表示 $3 \times 3$ 的卷积操作,J=4为分组的个数。

[0049] 2) 跨模态联合特征的生成

[0050] 对于RGB流和深度流经过选择的每组特征,通过元素级相加进行整合,得到每组跨模态特征表示  $Y_j^i$ ,其公式表达为:

$$[0051] \quad Y_j^i = Y_{R_j}^i + Y_{D_j}^i \quad (3)$$

[0052] 每组跨模态的特征表示通过通道级联整合在一起,送入 $1 \times 1$ 的卷积层进行融合。然后,经过一个通道注意力模块产生加权后的RGB-D特征  $Y_w^i$ 。

[0053] 最后,加权后的特征、RGB特征和深度特征进行元素级相加,通过 $3 \times 3$ 的卷积层和上采样层产生最终的跨模态联合特征  $Y_C^i$  ( $224 \times 224 \times 64$ ),其公式表达为:

$$[0054] \quad Y_C^i = \text{Up}\left(\text{Conv}\left(Y_w^i + X_R^i + X_D^i\right)\right) \quad (4)$$

[0055] 其中,Up(•)表示利用双线性插值的上采样操作,公式(4)的结果  $Y_C^i$  即为跨模态联合特征提取模块的输出。跨模态联合特征和单模态的多尺度特征一同送入下面的显著性目标检测部分进行整合。

[0056] 三、显著性目标检测

[0057] 显著性目标检测部分包括RGB显著性目标检测和深度显著性目标检测,两部分采用相同的结构。其中,在每一部分都设计了五个融合块(FB),其目的是为了整合单模态的多尺度特征和跨模态的联合特征,提高每流显著性目标检测的精度。其中,单模态的多尺度特征由两个卷积层和上采样层捕获。在每个融合块中,每级捕获的多尺度特征和联合特征通过通道级联和卷积进行整合,生成更新后的显著性特征。

[0058] RGB显著性目标检测部分的融合块(FB)如图3(b)所示。首先,通过两个卷积层和一个上采样层,捕获单模态的多尺度特征;然后,将捕获的多尺度特征、跨模态联合特征以及

前一个融合块输出的特征进行级联,送入 $3 \times 3$ 的卷积层进行特征整合。

[0059] RGB流生成的整合特征  $F_R^i$  表示为:

$$[0060] \quad F_R^i = \text{Conv}(\text{Concat}(F_R^{i-1}, S_R^{5-i}, Y_C^i))$$

[0061] 其中,  $F_R^{i-1}$  表示RGB流中第 $i-1$ 个融合块输出的特征,  $S_R^i$  表示RGB特征提取部分捕获的多尺度特征,  $\text{Concat}(\cdot)$  表示级联操作。特别地,第一个融合块的输入为  $S_R^5$ 、 $S_R^4$  和  $Y_C^1$ ; 第五个融合块的输入为  $F_R^4$  和  $Y_C^5$ 。

[0062] 同样,深度流生成的整合特征  $F_D^i$  表示为:

$$[0063] \quad F_D^i = \text{Conv}(\text{Concat}(F_D^{i-1}, S_D^{5-i}, Y_C^i))$$

[0064] 其中,  $F_D^{i-1}$  表示深度流中第 $i-1$ 个融合块的输出,  $S_D^i$  表示深度特征提取部分捕获的多尺度特征。

[0065] 在RGB流和深度流的最后一个融合块后,分别加入一个 $1 \times 1$ 的卷积进行显著性目标检测,生成各自预测的显著性图 ( $224 \times 224 \times 1$ )。

[0066] 四、低值融合损失设计

[0067] 在得到两流的显著性检测结果之后,为了使两流的检测结果得到进一步的融合,本发明设计了一种低值融合损失来约束不同预测之间的关系,提高显著性值的下界。设计的低值融合损失  $L_{\text{low-bound-fusion}}$  定义为:

$$[0068] \quad L_{\text{low-bound-fusion}} = \frac{1}{2N} \sum_{k=1}^N \left\| \min(P_k^R, P_k^D) - G_k \right\|_2^2$$

[0069] 其中,  $k$  表示像素的索引,  $P_k^R$  和  $P_k^D$  分别表示RGB流和深度流预测的显著性值,  $G_k$  表示真值,  $\|\cdot\|_2$  为2范数,  $N$  为图像的像素个数。

[0070] 通过引入低值融合损失,使网络迭代更新每个像素的低显著值。换句话说,通过在RGB流和深度流的显著性目标检测之间建立相关性,本发明设计的低值融合损失提供了一种在后期提升显著性图融合性能的新方法。

[0071] 五、网络整体损失设计

[0072] 在低值融合损失的基础上,加入常见的交叉熵损失来监督每一流的显著性检测。

RGB流的交叉熵损失  $L_{\text{cross-entropy}}^R$  和深度流的交叉熵损失  $L_{\text{cross-entropy}}^D$  定义为:

$$[0073] \quad L_{\text{cross-entropy}}^R = - \sum_k G_k \log P_k^R + (1 - G_k) \log(1 - P_k^R)$$

[0074] 和

$$[0075] \quad L_{\text{cross-entropy}}^D = - \sum_k G_k \log P_k^D + (1 - G_k) \log(1 - P_k^D)$$

[0076] 最后,网络整体损失函数  $L$  定义为:

$$[0077] \quad L = L_{\text{low-bound-fusion}} + L_{\text{cross-entropy}}^R + L_{\text{cross-entropy}}^D$$

[0078] 六、网络训练策略设计

[0079] 整个网络的训练分为两个阶段,即预训练阶段和联合训练阶段。在预训练阶段,RGB流和深度流分别进行训练,采用交叉熵损失对每一流的显著性检测进行监督。每一流的训练收敛后,将这一阶段的参数作为下一阶段训练的初始化权重。在联合训练阶段,加入跨模态联合特征提取部分,与单模态显著性检测部分一起进行联合训练。利用上一阶段两个流的权值初始化整个网络的参数,跨模态联合特征提取部分的权重采用随机初始化的方式。同时,加入设计的低值融合损失,其与交叉熵损失一起作为整个网络的监督。

[0080] 实施例2

[0081] 图1给出了本发明的技术流程图,主要包括:单模态特征提取、跨模态联合特征提取、显著性目标检测、低值融合损失设计、网络整体损失设计和网络训练策略设计六个部分。

[0082] 图2给出了本发明的具体实施框图。

[0083] 图3给出了跨模态特征提取模块(CFM)和RGB显著性检测部分融合块(FB)的结构图。

[0084] 图4给出了RGB-D显著性目标检测的例子。其中,第一列为RGB图像,第二例为深度图,第三列为显著性目标检测的真值图,第四列为本发明方法得到的结果。

[0085] 从结果可以看出,本发明方法有效地融合了RGB图像和深度图的信息,检测到的显著性目标结构完整、轮廓清晰。

[0086] 本领域技术人员可以理解附图只是一个优选实施例的示意图,上述本发明实施例序号仅仅为了描述,不代表实施例的优劣。

[0087] 以上所述仅为本发明的较佳实施例,并不用以限制本发明,凡在本发明的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。



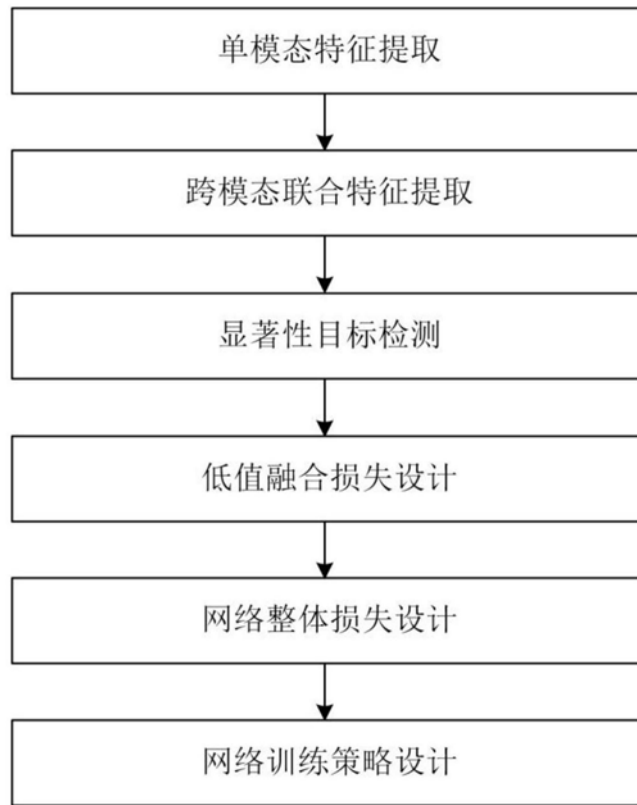


图1

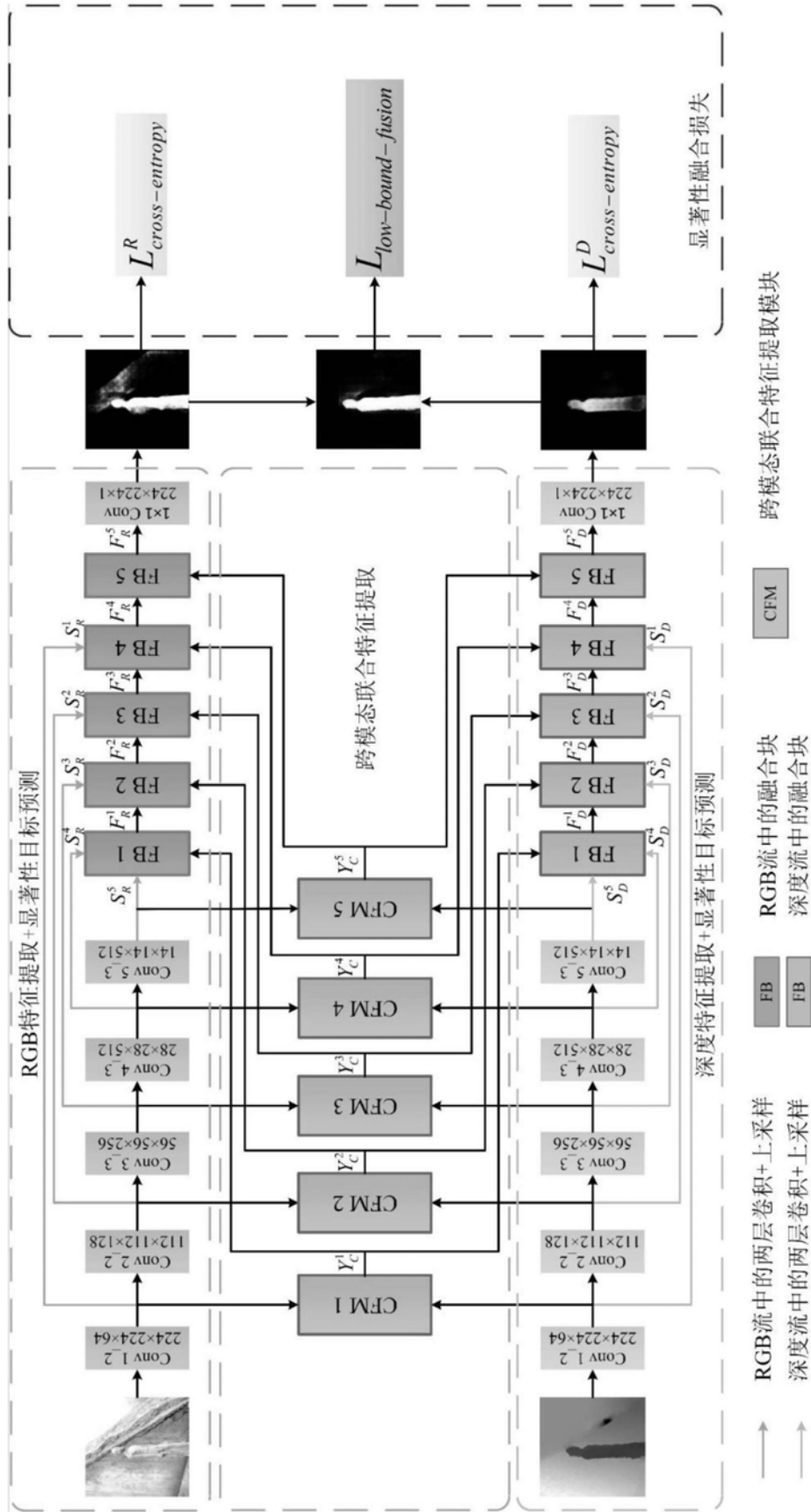


图2

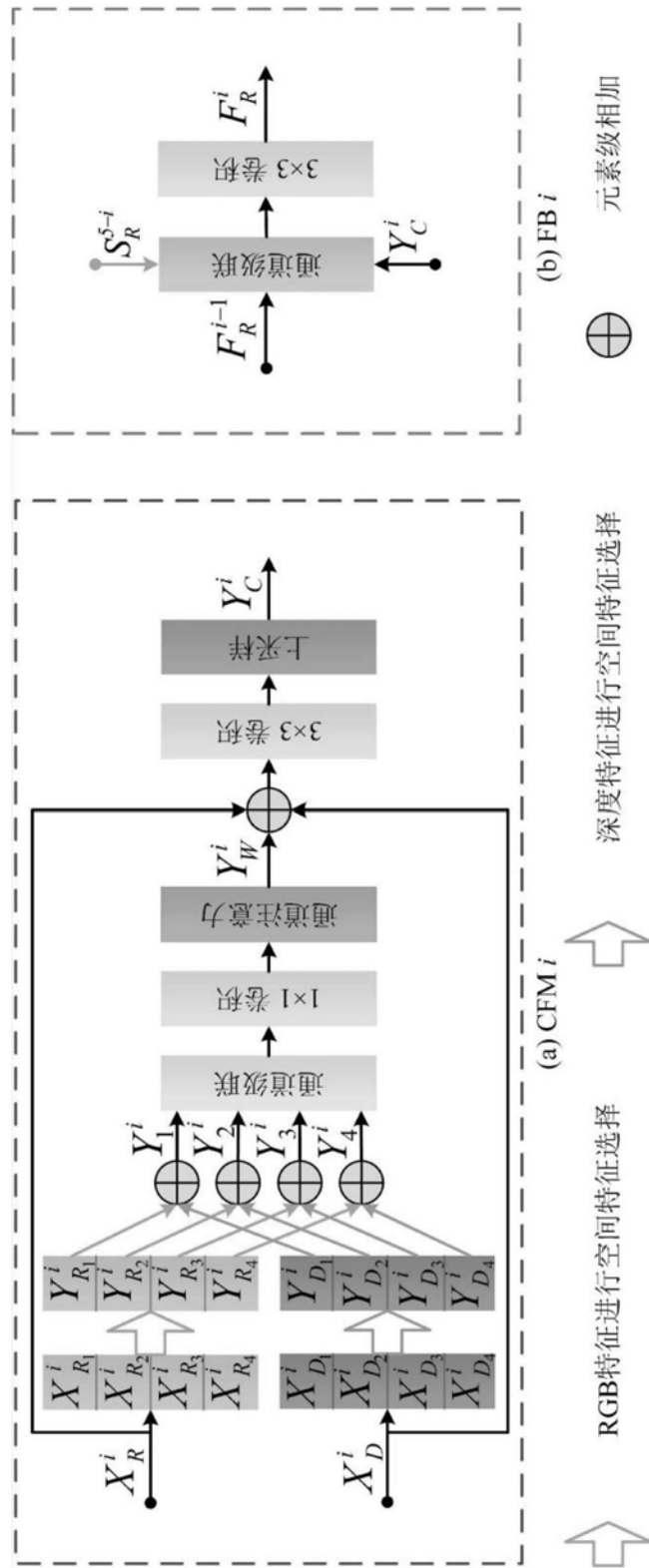


图3

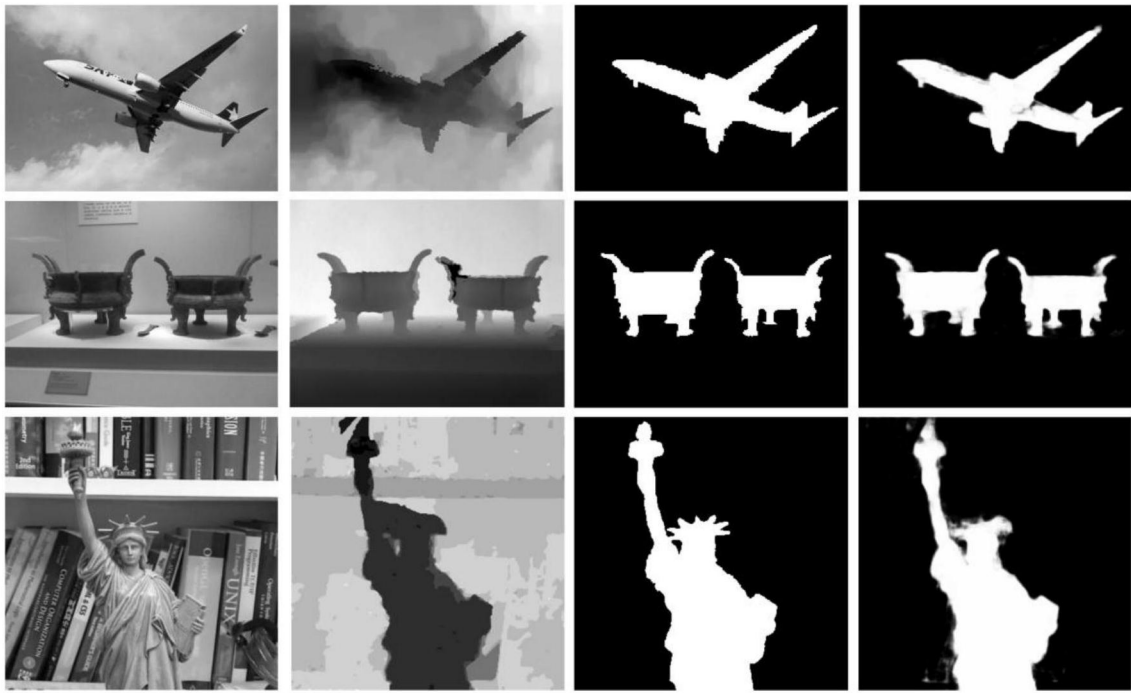


图4