



(12) 发明专利申请

(10) 申请公布号 CN 118887484 A

(43) 申请公布日 2024. 11. 01

(21) 申请号 202410959643.2

(22) 申请日 2024.07.17

(71) 申请人 广东工业大学

地址 510062 广东省广州市东风东路729号

(72) 发明人 周湘鑫 陈一嘉 陈绪行 陈平华
邵长城

(74) 专利代理机构 长沙轩荣专利代理有限公司
43235

专利代理师 齐超

(51) Int. Cl.

G06V 10/77 (2022.01)

G06V 10/764 (2022.01)

G06V 10/56 (2022.01)

G06V 10/774 (2022.01)

G06V 10/82 (2022.01)

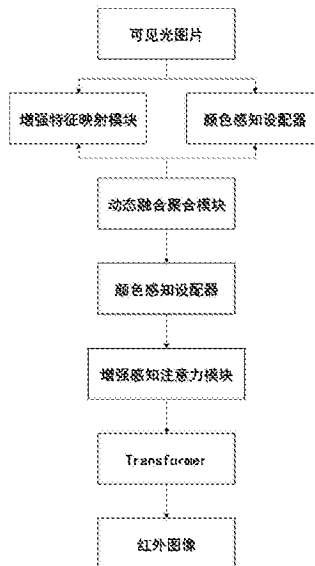
权利要求书2页 说明书6页 附图3页

(54) 发明名称

一种端到端的轻量级可见光图像到红外图像转换方法

(57) 摘要

本发明提出了一种端到端的轻量级可见光图像到红外图像转换方法,包括:获取可见光图片,通过增强特征映射模块和颜色感知适配器协同工作,从可见光图片中提取纹理和颜色特征;动态融合聚合模块将上述提取到的纹理和颜色特征进行融合;将融合后的特征通过颜色感知适配器和增强感知注意力机制的协同作用,将红外图像的转换细化,生成红外图像。与现存方法相比,本申请实例,是首个基于Transformer的端到端可见光图像转红外图像的方法,本模型体积小、速度快、可部署到移动设备,能有效缓解了现有技术容易出现捏造、不规则的视觉伪影等技术问题。



1. 一种端到端的轻量级可见光图像到红外图像转换方法,其特征在于包含以下步骤:

S1: 获取待转换的可见光图像数据;

S2: 将上述S1中采集到的待转换的可见光图像输入到增强特征映射模块和颜色感知适配器,从而提取可见光图片中的纹理和颜色特征;

S3: 通过动态融合模块,将S2中提取到的纹理和颜色特征进行融合;

S4: 将S3中融合后的特征通过颜色感知适配器转置到红外像素域;

S5: 将S4的结果通过增强感知注意力模块,以保留和突出关键的纹理细节,最终形成具有较好清晰度和信息内容的红外图像;

S6: 利用Transformer架构来增强先前模块外推的信息的集成,对于完成和完善红外图像重建过程比较有利。

2. 根据权利要求1所述的一种端到端的轻量级可见光图像到红外图像转换方法,其特征在于,所述步骤S2具体包括:

S21: 通过颜色感知适配器,从可见光图片中提取红外颜色特征,颜色感知适配器的工作过程是将输入特征分成四个不同的部分即可见光的三个RGB通道以及可见光图片中蕴含的红外分量,然后将RGB通道通过卷积提取特征后与红外特征进行融合;

S22: 通过增强特征映射模块,来增强细节和粒度;增强特征映射模块通过采用各种尺度的下采样来获得多尺度特征表示。

3. 根据权利要求1所述的一种端到端的轻量级可见光图像到红外图像转换方法,所述步骤S3中强调的动态融合聚合模块,包括:

该模块负责融合可见光的特征并将其投射到潜在空间中,我们构建一个注意力矩阵,该矩阵有矩阵集的最大值和平均值提供信息,随后的特征映射通过这种注意力驱动机制进行,该过程通过如下公式进行:

$$f = [\text{MaxPool}_{d0}(x); \text{AvgPool}_{d0}(x)]$$

$$y = x \times \sigma(f(x))$$

上式中, f 代表注意力矩阵, σ 表示逻辑函数。

4. 根据权利要求1所述的一种端到端的轻量级可见光图像到红外图像转换方法,步骤S4中强调的颜色感知适配器,包括:

颜色感知适配器的将输入特征分成四个不同的部分即可见光的三个RGB通道以及可见光图片中蕴含的红外分量,然后将RGB通道通过卷积提取特征后与红外特征进行融合;将上个过程提取到的特征转置到红外像素域。

5. 根据权利要求1所述的一种端到端的轻量级可见光图像到红外图像转换方法,所述步骤S5强调的增强感知注意力模块,其特征在于利用由通道和像素注意力策略组成的双重注意力系统,完善和增强潜在空间的特征集;这种双重机制确保每个像素都经过调整,以保留和突出关键的纹理细节,最终形成具有较好清晰度和信息内容的红外图像,具体包括:

S51: 通过增强特征表示,来补偿由低光照条件或遮挡引起的信息丢失,采用公式如下:

$$Y = x + \text{Conv}(X1 \odot X2) \times W_1$$

上式中, Conv 表示卷积运算,而 $X1$ 和 $X2$ 是通过拆分 X 的通道获得的两个相同的特征图, W_1 表示一组自定义可学习的权重;

S52: 利用通道注意力机制(CA)能够捕捉全局信息,它首先将空间信息压缩到通道维

度,然后应用多层感知器计算通道注意力,用于加权特征图,采用公式如下:

$$CA(X) = X \times \sigma(W2 \max(0, W1pool(X)))$$

上式中, σ 为非线性激活函数,Sigmoid、W1、W2为全连接层;

S53:像素注意力机制(PA)可以自适应调整每个像素的权重,使模型更好地聚焦于重要像素,从而捕捉图像中的微小变化和细节,并提取更多的信息特征,采用公式如下:

$$PA(X) = X \times \sigma(W4 \max(0, W3(X)))$$

上式中,式中 σ 为非线性激活函数,Sigmoid、W3、W4为全连接层。

6. 根据权利要求1所述的一种端到端的轻量级可见光图像到红外图像转换方法,为了更好地保存结构信息并增强图像对比度,我们的模型采用了双损失函数策略,将Smooth L1损失函数的优势与结构相似性指数测量LSSIM协同作用,采用公式如下:

$$L_{smooth}(x) = \begin{cases} 0.5 \times x^2 & |x| < 1 \\ |x| - 0.5 & otherwise \end{cases}$$

$$L_{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

$$L(x, y) = L_{smooth}(x) + L_{SSIM}(x, y)$$

上式中, μ_x 和 μ_y 代表x和y的平均值, σ_x^2 和 σ_y^2 代表x和y的方差, σ_{xy} 是x和y的协方差, $C_1 = (k_1L)^2$ 、 $C_2 = (k_2L)^2$ 是两个常数,其中L表示图像数据的动态范围,L表示总损失函数。

一种端到端的轻量级可见光图像到红外图像转换方法

技术领域

[0001] 本发明属于图像处理技术领域,涉及一种图像转换方法,具体涉及一种基于Transformer的轻量级可见光图像到红外图像转换方法。

背景技术

[0002] 在可见光成像过程中,在某些情况下,所获取的图像对比度较低,或者某些物体在可见光谱中的反射率降低。红外图像(IR)可以缓解这些问题,提供增强的对比度和物体的可检测性,无论其可见光反射率如何。例如,在弱光条件下获得的可见光图像容易出现对比度低、边缘模糊和细节丢失等问题。对边缘信息有较高需求的领域——包括行人检测、自动驾驶、机器人导航、和精准农业——可能会看到实质性的改进。来自红外成像提供的卓越边缘清晰度。此类增强功能对于自动驾驶等安全关键型应用尤其重要。然而,与广泛使用的可见光相机相比,红外成像设备的成本仍然很高。

[0003] 相关技术在转换中存在不足,比如GAN在图像生成任务中表现出比卷积神经网络(CNN)更优越的性能。利用GAN进行图像翻译的应用,例如分割、语义到视觉转换、超分辨率、图像合成和风格转移已经有论文表现出显著的成功。但将GAN用于VIS-IR(可见转红外)图像转换时,结果可能并不理想。虽然GAN模型能够有效地从红外图像中提取对比度特征,但往往容易误判与光度相关的特征,导致在没有物体存在的区域出现捏造的情况和不规则的视觉伪影等问题。

[0004] 基于此,目前亟需一种一种端到端的轻量级可见光图像到红外图像转换方法来解决上述技术问题。

发明内容

[0005] 本发明的目的在于克服上述现有技术存在的缺陷,提出了一种一种端到端的轻量级可见光图像到红外图像转换方法,用于解决现有技术存在的在没有物体存在的区域出现捏造的情况和不规则的视觉伪影等技术问题。

[0006] 为实现上述目的,本发明包括如下步骤:

[0007] 步骤S1:获取待转换的可见光图像数据;

[0008] 步骤S2:将上述S1中采集到的待转换的可见光图像输入到增强特征映射模块和颜色感知适配器,从而提取可见光图片中的纹理和颜色特征;

[0009] 步骤S3:通过动态融合模块,将S2中提取到的纹理和颜色特征进行融合;

[0010] 步骤S4:将S3中融合后的特征通过颜色感知适配器转置到红外像素域;

[0011] 步骤S5:将S4的结果通过增强感知注意力模块,以保留和突出关键的纹理细节,最终形成具有较好清晰度和信息内容的红外图像;

[0012] 步骤S6:利用Transformer架构来增强先前模块外推的信息的集成,对于完成和完善红外图像重建过程比较有利。

[0013] 进一步地,所述步骤S2具体包括:

[0014] 步骤S21:通过颜色感知设配器,从可见光图片中提取红外颜色特征,颜色感知设配器的工作过程是将输入特征分成四个不同的部分即可见光的三个RGB通道以及可见光图片中蕴含的红外分量,然后将RGB通道通过卷积提取特征后与红外特征进行融合;

[0015] 步骤S22:通过增强特征映射模块,来增强细节和粒度。增强特征映射模块通过采用各种尺度的下采样来获得多尺度特征表示。

[0016] 进一步地,步骤S3中强调的动态融合聚合模块,其特征就在于,该模块负责融合可见光的特征并将其投射到潜在空间中,我们构建一个注意力矩阵,该矩阵有矩阵集的最大值和平均值提供信息,随后的特征映射通过这种注意力驱动机制进行,该过程通过如下公式进行:

$$[0017] \quad f = [\text{MaxPool}d0(x); \text{AvgPool}d0(x)]$$

$$[0018] \quad y = x \times \sigma(f(x))$$

[0019] 式中f代表注意力矩阵, σ 表示逻辑函数

[0020] 进一步地,步骤S4中强调的颜色感知设配器,将输入特征分成四个不同的部分即可见光的三个RGB通道以及可见光图片中蕴含的红外分量,然后将RGB通道通过卷积提取特征后与红外特征进行融合;将上个过程提取到的特征转置到红外像素域。

[0021] 进一步地,步骤S5强调的增强感知注意力模块,其特征就在于利用由通道和像素注意力策略组成的双重注意力系统,完善和增强潜在空间的特征集。这种双重机制确保每个像素都经过调整,以保留和突出关键的纹理细节,最终形成具有较好清晰度和信息内容的红外图像,具体包括:

[0022] S51:通过增强特征表示,来补偿由低光照条件或遮挡引起的信息丢失,采用公式如下:

$$[0023] \quad Y = x + \text{Conv}(X1 \odot X2) \times W_1$$

[0024] 式中Conv表示卷积运算,而X1和X2是通过拆分X的通道获得的两个相同的特征图, W_1 表示一组自定义可学习的权重。

[0025] S52:利用通道注意力机制(CA)能够捕捉全局信息。它首先将空间信息压缩到通道维度,然后应用多层感知器计算通道注意力,用于加权特征图,采用公式如下:

$$[0026] \quad CA(X) = X \times \sigma(W2 \max(0, W1\text{pool}(X)))$$

[0027] 式中 σ 为非线性激活函数,Sigmoid、W1、W2为全连接层

[0028] S53:像素注意力机制(PA)可以自适应调整每个像素的权重,使模型更好地聚焦于重要像素,从而捕捉图像中的微小变化和细节,并提取更多的信息特征。采用公式如下:

$$[0029] \quad PA(X) = X \times \sigma(W4 \max(0, W3(X)))$$

[0030] 式中 σ 为非线性激活函数,Sigmoid、W3、W4为全连接层

[0031] 进一步地,为了更好地保存结构信息并增强图像对比度,我们的模型采用了双损失函数策略,将Smooth L1损失函数的优势与结构相似性指数测量LSSIM协同作用,采用公式如下:

$$[0032] \quad L_{\text{smooth}}(x) = \begin{cases} 0.5 \times x^2 & |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases}$$

$$L_{\text{SSIM}}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

[0033] 式中 μ_x 和 μ_y 代表x和y的平均值, σ_x^2 和 σ_y^2 代表x和y的方差, σ_{xy} 是x和y的协方差, $C_1 = k_1 L)^2$ 、 $C_2 = (k_2 L)^2$ 是两个常数,其中L表示图像数据的动态范围,L表示总损失函数。

附图说明

[0034] 为了更清楚地说明本申请实施例中的技术方案,下面将对实施例描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本申请的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0035] 图1为本发明的实现流程图;

[0036] 图2的为本发明构建的模型的整体结构示意图;

[0037] 图3为本发明采用的颜色感知适配器的结构示意图;

[0038] 图4为本发明采用增强的特征映射模块的结构示意图;

[0039] 图5为本发明采用动态融合聚合模块的结构示意图;

[0040] 图6为本发明采用增强感知注意力模块的结构示意图。

具体实施方式

[0041] 为使本申请实施例的目的、技术方案和优点更加清楚,下面将结合本申请实施例中的附图,对本申请实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例是本申请的一部分实施例,而不是全部的实施例。基于本申请中的实施例,本领域普通技术人员在没有做出创造性劳动的前提下所获得的所有其他实施例,都属于本申请保护的范围。

[0042] 如图1所示,本发明提供了一种一种端到端的轻量级可见光图像到红外图像转换方法,包括如下步骤:

[0043] S1:获取待转换的可见光图像数据,如手机、摄像机等移动设备上获取可见光图像,后将图片输入到模型中;

[0044] S2:提取可见光图片中的纹理和颜色特征:将上述S1中采集到的待转换的可见光图像输入到增强特征映射模块和颜色感知适配器,其中颜色感知适配器从可见光图像中提取红外颜色特征,增强特征映射模块增强了细节和粒度;

[0045] S3:融合可见光的特征并将其投射到潜在空间:通过动态融合模块,将S2中提取到的纹理和颜色特征进行融合,该模块经过精心设计,可确保视觉信息的无缝过渡,保持原始特征的完整性;

[0046] S4:将特征转置到红外像素域:将S3中融合后的特征通过颜色感知适配器转置到红外像素域,通过这个过程,模块学会了适应这些特征,增强了模型解释和重建输入红外方面的能力;

[0047] S5:保留和突出关键的纹理细节:将S4的结果通过增强感知注意力模块,它侧重于补偿在弱光环境或视野受阻的区域出现的信息丢失,保留和突出关键的纹理细节,最终形成具有较好清晰度和信息内容的红外图像;

[0048] S6:生成红外图像:利用Transformer架构来增强先前模块外推的信息的集成,对于完成和完善红外图像重建过程比较有利。

[0049] 在一些实施方式中,上述步骤S2的具体过程为:

[0050] S21:参见图3,通过颜色感知设配器,从可见光图片中提取特征存在于原始图片中但人眼无法察觉红外颜色特征,颜色感知设配器的工作过程是将输入特征分成四个不同的部分即可见光的三个RGB通道以及可见光图片中蕴含的红外分量,然后将RGB通道通过卷积提取特征后与红外特征进行融合;

[0051] S22:参见图4,通过增强特征映射模块,来增强细节和粒度,这种丰富的特征本质上更具辨别力和信息性,从而为后续任务(如目标检测和图像分类)提供了实质性的好处。增强特征映射模块通过采用各种尺度的下采样来获得多尺度特征表示。

[0052] 在一些实施方式中,上述步骤S3所述的动态融合聚合模块的具体过程为:

[0053] 参见图5,动态融合聚合模块负责融合可见光的特征并将其投射到潜在空间中,我们构建一个注意力矩阵,该矩阵有矩阵集的最大值和平均值提供信息,随后的特征映射通过这种注意力驱动机制进行,可确保视觉信息的无缝过渡,保持原始特征的完整性,同时为准确的红外表示做好准备,该过程通过如下公式进行:

[0054] $f = [\text{MaxPool}d0(x); \text{AvgPool}d0(x)]$

[0055] $y = x \times \sigma(f(x))$

[0056] 式中 f 代表注意力矩阵, σ 表示逻辑函数

[0057] 在一些实施方式中,上述步骤S4所述的颜色感知设配器的具体过程为:

[0058] 将输入特征分成四个不同的部分即可见光的三个RGB通道以及可见光图片中蕴含的红外分量,然后将RGB通道通过卷积提取特征后与红外特征进行融合;将上个过程提取到的特征转置到红外像素域。通过这个过程,模块学会了适应这些特征,增强了模型解释和重建输入红外方面的能力。

[0059] 在一些实施方式中,上述步骤S5所述的增强感知注意力模块,它侧重于补偿在弱光环境或视野受阻的区域出现的信息丢失。该模块利用由通道和像素注意力策略组成的双重注意力系统,完善和增强潜在空间的特征集。这种双重机制确保每个像素都经过调整,以保留和突出关键的纹理细节,最终形成具有较好清晰度和信息内容的红外图像,具体过程为:

[0060] S51:选择从潜在空间转换红外图像以生成更高质量的红外图像的特征。该模块旨在增强特征表示,以补偿由低光照条件或遮挡引起的信息丢失,采用公式如下:

[0061] $Y = x + \text{Conv}(X1 \odot X2) \times W_1$

[0062] 式中Conv表示卷积运算,而 $X1$ 和 $X2$ 是通过拆分 X 的通道获得的两个相同的特征图, W_1 表示一组自定义可学习的权重。

[0063] S52:利用通道注意力机制(CA)能够捕捉全局信息,并且计算效率高。它首先将空间信息压缩到通道维度,然后应用多层感知器计算通道注意力,用于加权特征图,采用公式如下:

[0064] $CA(X) = X \times \sigma(W2 \max(0, W1 \text{pool}(X)))$

[0065] 式中 σ 为非线性激活函数,Sigmoid、 $W1$ 、 $W2$ 为全连接层

[0066] S53:像素注意力机制(PA)可以自适应调整每个像素的权重,使模型更好地聚焦于重要像素,从而捕捉图像中的微小变化和细节,并提取更多的信息特征。采用公式如下:

[0067] $PA(X) = X \times \sigma(W4 \max(0, W3(X)))$

[0068] 式中 σ 为非线性激活函数,Sigmoid、 $W3$ 、 $W4$ 为全连接层

[0069] CA中的特征通过池化进行处理,将特征大小从(C,H,W)减小到(C,1,1)。最终卷积输出特征的大小为(C,1,1)。PA中的特征通过卷积进行处理,将(C,H,W)特征转换为(1,H,W)

[0070] 在一些实施方式中,为了更好地保存结构信息并增强图像对比度,我们的模型采用了双损失函数策略,将Smooth L1损失函数的优势与结构相似性指数测量LSSIM协同作用,该指标旨在通过考虑图像的亮度、对比度和结构相似性来评估图像的质量,采用公式如下:

$$L_{smooth}(x) = \begin{cases} 0.5 \times x^2 & |x| < 1 \\ |x| - 0.5 & otherwise \end{cases}$$

$$LSSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

$$L(x,y) = L_{smooth}(x) + LSSIM(x,y)$$

[0073] 式中 μ_x 和 μ_y 代表x和y的平均值, σ_x^2 和 σ_y^2 代表x和y的方差, σ_{xy} 是x和y的协方差, $C_1 = (k_1L)^2$ 、 $C_2 = (k_2L)^2$ 是两个常数,其中L表示图像数据的动态范围,L表示总损失函数。

[0074] 下面结合仿真结果,对本发明的技术效果作进一步说明:

[0075] 1.仿真条件与内容:

[0076] 我们的模型利用PyTorch神经网络框架来处理可见光和红外光谱中的输入图像,每个图像的分辨率为 256×256 像素。为了微调网络的超参数,我们从 2×10^{-4} 的初始学习率开始,并使用AdamW优化器

[0077] 为了在整个训练过程中动态调整学习率,使用了CosineAnnealingLR调度器

[0078] 训练进行了400多轮,以确保模型的彻底学习和收敛。我们应用批量大小为1,这可能会增加计算时间,但它会导致更精确的梯度计算,从而提高学习效率和模型稳定性。

[0079] 将来自五个独立数据集的训练集合包括:LLVIP、RoadScene、M3FD、FLIR、MCubeS并到一个全面的训练方案中。这种集成旨在扩大数据的多样性并提高模型的泛化性。使用峰值信噪比(PSNR)和结构相似性指数测量(SSIM)作为评估指标。PSNR通过计算原始图像与其失真图像(例如压缩版本)之间的均方误差(MSE)来确定图像质量

[0080] 为了评估我们方法的有效性,我们使用上述五个数据集进行了比较实验。我们的方法已经针对一些红外图像生成算法进行了基准测试:CycleGAN、Pix2Pix、UNIT、MUNIT、ThermalGAN、BCI、InfraGAN和ClawGAN。

[0081] 表I对实验结果进行了广泛而客观的评估。该表收集并量化了每种算法的性能指标,从而有助于对其有效性进行更全面和定量的分析。

[0082] 表I

Method	LLVIP		RoadScene		M3FD		FLIR		MCubeS		MSE (G1)	Params(M)
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM		
CycleGAN	4.34	0.12	4.19	0.01	3.87	0.07	3.35	0.03	1.30	0.17	14.39	1.31
Pix2Pix	1.63	0.01	4.68	0.02	0.74	0.12	4.19	0.05	2.58	0.12	18.15	0.72
UNIT	3.38	0.08	4.22	0.00	5.28	0.07	3.13	0.01	0.79	0.13	61.82	1.33
MUNIT	4.06	0.12	4.70	0.03	4.53	0.06	3.66	0.02	0.53	0.14	96.31	1.68
ThermalGAN	33.90	0.06	38.43	0.00	35.43	2.97	36.00	3.20	32.89	0.00	18.15	0.72
BCI	6.84	0.11	7.67	0.00	3.33	0.13	13.14	0.21	3.23	0.06	18.15	16.15
InfraGAN	39.52	0.00	39.26	0.00	39.63	0.00	39.86	1.05	32.38	3.47	18.15	5.94
ClawGAN	3.25	0.13	5.07	0.01	7.20	0.03	5.91	0.04	3.73	0.16	14.07	0.51
Ours	14.01	0.38	14.38	0.25	14.36	0.36	17.74	0.38	8.96	0.79	2.43	0.03

[0084] 本发明在不同测试样本集上的峰值信噪比和结构相似度相对于现有的技术,峰值信噪比和结构相似度有显著提高,且仅使用0.04M参数和2.41G MA,保持较低的计算开销

[0085] 综合上述仿真实验中的结果分析,本发明提出一种基于Transformer的模型,该模

型可以有效地将可见光图像转换为红外图像。本发明提出的动态融合聚合模块和增强感知注意力模块协同工作,有效地捕捉和保留关键的纹理和颜色特征,然后将其准确地转换为红外域。通过广泛的基准实验证明了我们模型的有效性,这些实验在产生定性和定量优越的红外图像方面明显优于现有方法。

[0086] 本说明书中的各个实施例均采用相关的方式描述,各个实施例之间相同相似的部分互相参见即可,每个实施例重点说明的都是与其他实施例的不同之处。尤其,对于系统实施例而言,由于其基本相似于方法实施例,所以描述的比较简单,相关之处参见方法实施例的部分说明即可。

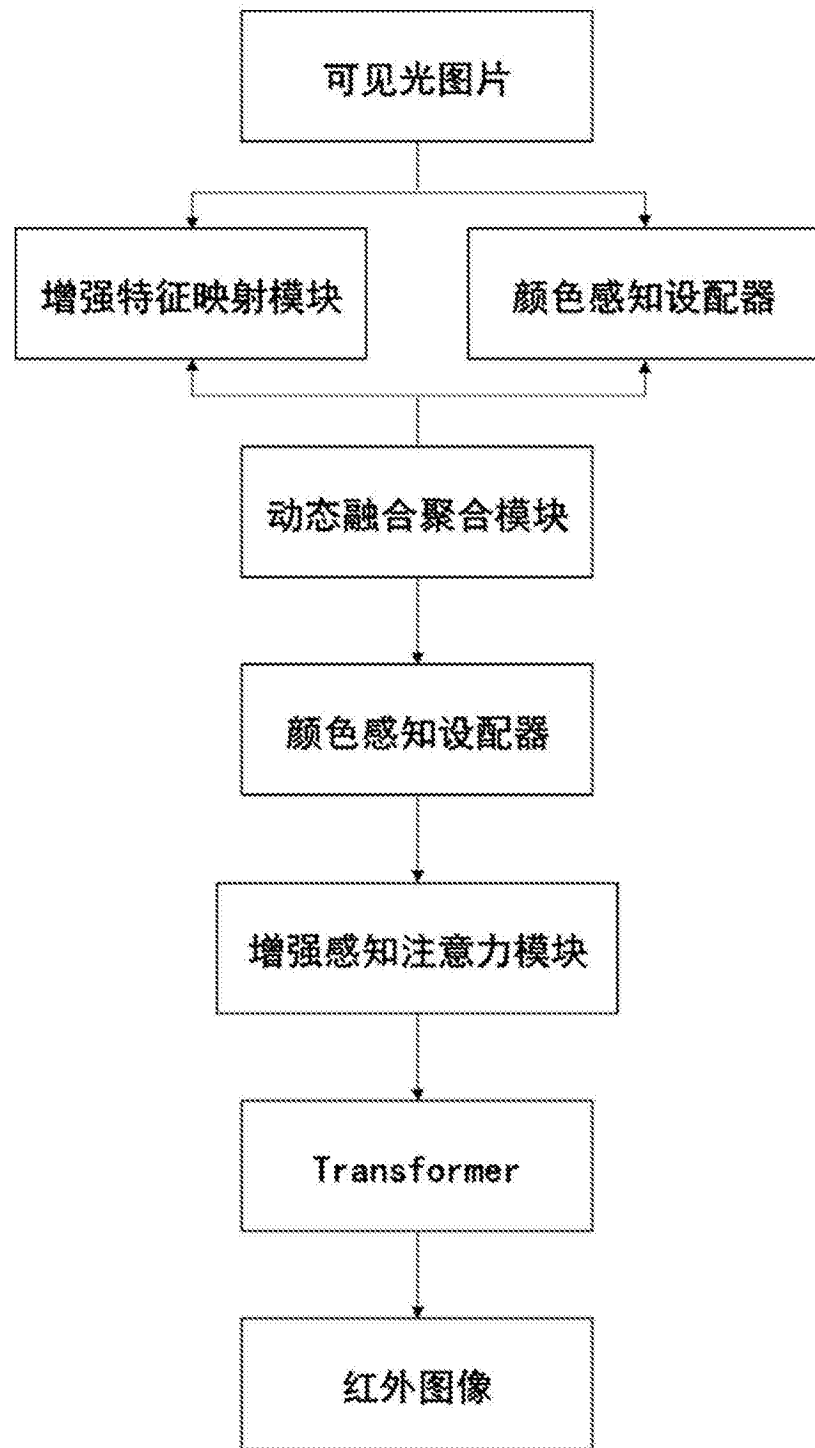


图1

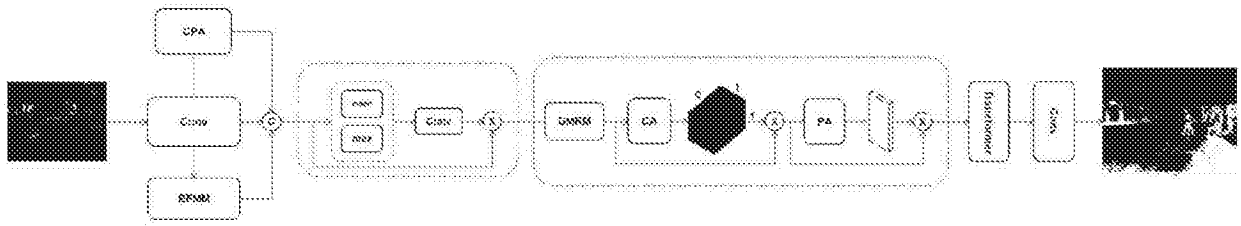


图2

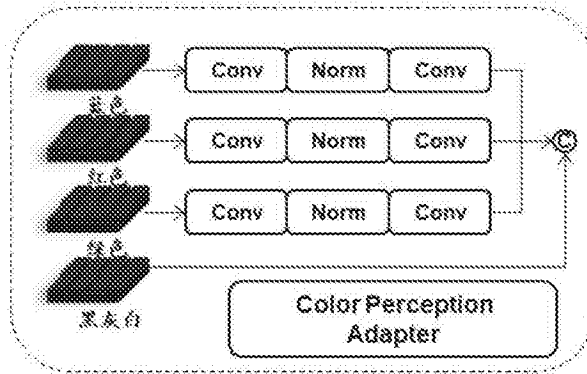


图3

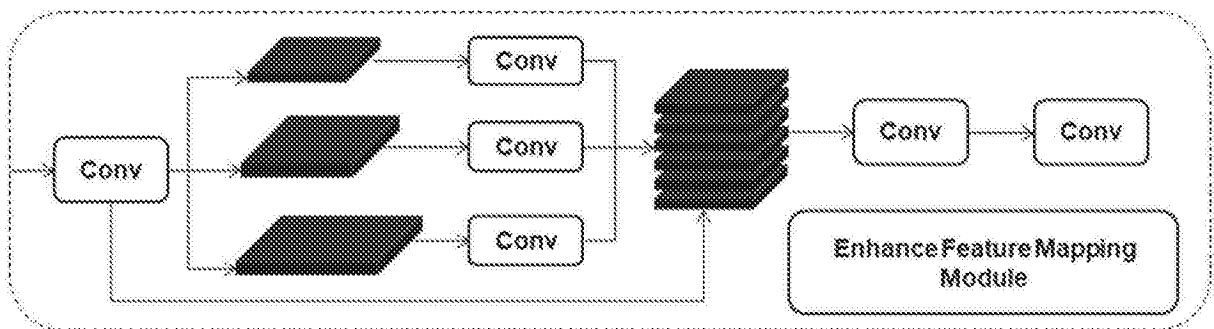


图4

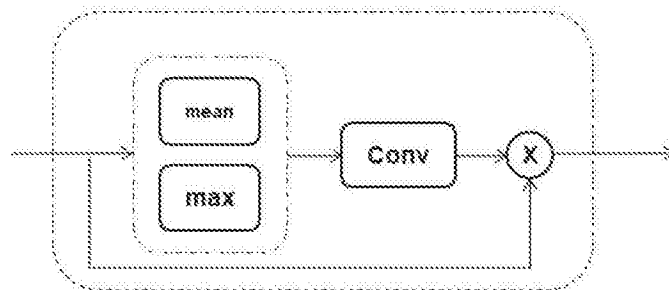


图5

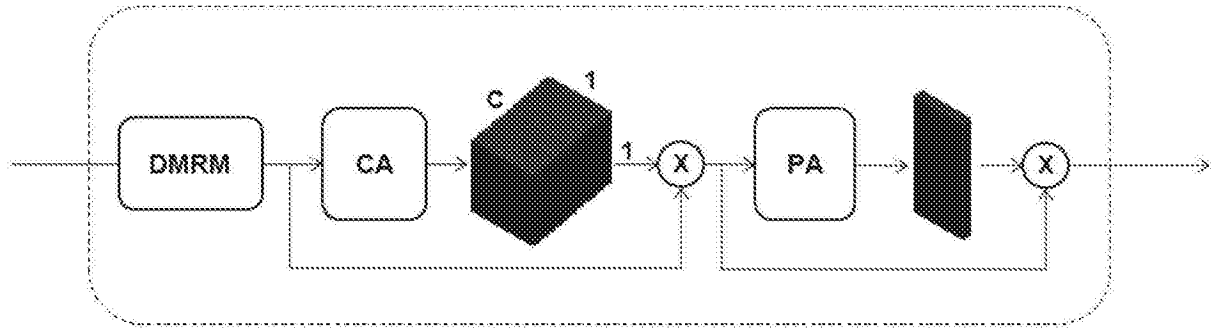


图6