



(12) 发明专利申请

(10) 申请公布号 CN 114266814 A

(43) 申请公布日 2022. 04. 01

(21) 申请号 202111488926.6

G06V 10/44 (2022.01)

(22) 申请日 2021.12.07

G06V 10/82 (2022.01)

(71) 申请人 中国科学院深圳先进技术研究院

地址 518000 广东省深圳市南山区深圳大学
学城学苑大道1068号

(72) 发明人 程俊 王飞

(74) 专利代理机构 深圳中一联合知识产权代理有限公司 44414

代理人 张瑞志

(51) Int. Cl.

G06T 7/50 (2017.01)

G06T 9/00 (2006.01)

G06T 3/40 (2006.01)

G06N 3/08 (2006.01)

G06N 3/04 (2006.01)

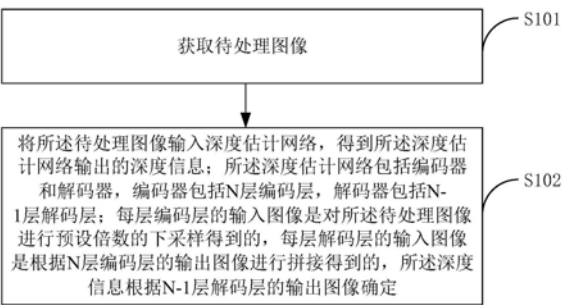
权利要求书2页 说明书13页 附图4页

(54) 发明名称

图像的深度估计方法、装置、电子设备及存储介质

(57) 摘要

本申请适用于图像处理领域,提供了图像的深度估计方法、装置、电子设备及存储介质。图像的深度估计方法包括获取待处理图像,将待处理图像输入深度估计网络,得到深度估计网络输出的深度信息。深度估计网络包括编码器和解码器,编码器包括N层编码层,解码器包括N-1层解码层;每层编码层的输入图像是对待处理图像进行预设倍数的下采样得到的,每层解码层的输入图像是根据N层编码层的输出图像进行拼接得到的,深度信息根据N-1层解码层的输出图像确定,N表示大于0的整数。因此,每层解码层的输入信息融合了编码层的多级语义信息,降低了编码器和解码器的特征图之间的语义代沟,提高了恢复出的场景深度信息的精度。



1. 一种图像的深度估计方法,其特征在于,包括:

获取待处理图像;

将所述待处理图像输入深度估计网络,得到所述深度估计网络输出的深度信息;所述深度估计网络包括编码器和解码器,所述编码器包括N层编码层,所述解码器包括N-1层解码层;每层编码层的输入图像是对所述待处理图像进行预设倍数的下采样得到的,每层解码层的输入图像是根据N层编码层的输出图像进行拼接得到的,所述深度信息根据N-1层解码层的输出图像确定,N表示大于0的整数。

2. 根据权利要求1所述的方法,其特征在于,第N-1层解码层的输入图像由第一输入图像和第二输入图像进行拼接得到,所述第一输入图像由第N层编码层的输出图像依次经过精调处理、上采样得到,所述第二输入图像由第i层编码层的输出图像依次进行 2^{N-1-i} 倍下采样、精调处理得到, $i=0,1,2,\dots,N-1$,所述精调处理用于对图像的局部特征信息和全局特征信息进行拼接。

3. 根据权利要求1所述的方法,其特征在于,第k层解码层的输入图像由第三输入图像和第四输入图像进行拼接得到,所述第三输入图像由第k+1层解码层的输出图像依次经过精调处理、上采样得到,所述第四输入图像由第i层编码层的输出图像依次经过 2^{k-i} 倍下采样、精调处理得到,所述精调处理用于对图像的局部特征信息和全局特征信息进行拼接, $k=0,1,2,\dots,N-2,k-i\geq 0$ 。

4. 根据权利要求2或3所述的方法,其特征在于,图像的精调处理过程包括:

提取待精调处理的图像的局部特征图,所述局部特征图包括局部特征信息;

对所述待精调处理的图像进行不重叠采样,得到互不重叠的子特征图;

根据所述子特征图确定全局特征图,所述全局特征图包括全局特征信息;

对所述待精调处理的图像、所述局部特征图、所述全局特征图进行拼接,得到精调处理后的图像。

5. 根据权利要求4所述的方法,其特征在于,根据所述子特征图确定全局特征图,包括:

确定与子特征图对应的特征向量;

对所述特征向量依次进行特征向量变换、线性变换、分辨率恢复、膨胀处理,得到所述全局特征图。

6. 根据权利要求2或3所述的方法,其特征在于,图像的上采样过程包括:

按照通道维度将待上采样的图像分割为第一低分辨率图像和第二低分辨率图像;

采用双线性差值方式对所述第一低分辨率图像进行上采样,得到第一高分辨率图像;

对所述第二低分辨率图像进行膨胀处理和精调处理,得到第二高分辨率图像;

对所述第一高分辨率图像和所述第二高分辨率图像进行拼接,得到上采样后的图像。

7. 根据权利要求1所述的方法,其特征在于,所述深度信息由第0层解码层的输出图像进行上采样得到。

8. 一种图像的深度估计装置,其特征在于,包括:

获取模块,用于获取待处理图像;

处理模块,用于将所述待处理图像输入深度估计网络,得到所述深度估计网络输出的深度信息;所述深度估计网络包括编码器和解码器,所述编码器包括N层编码层,所述解码器包括N-1层解码层;每层编码层的输入图像是对所述待处理图像进行预设倍数的下采样

得到的,每层解码层的输入图像是根据N层编码层的输出图像进行拼接得到的,所述深度信息根据N-1层解码层的输出图像确定,N表示大于0的整数。

9.一种电子设备,包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机程序,其特征在于,所述处理器执行所述计算机程序时实现如权利要求1至7任一项所述的方法。

10.一种计算机可读存储介质,所述计算机可读存储介质存储有计算机程序,其特征在于,所述计算机程序被处理器执行时实现如权利要求1至7任一项所述的方法。

图像的深度估计方法、装置、电子设备及存储介质

技术领域

[0001] 本申请属于图像处理领域,尤其涉及图像的深度估计方法、装置、电子设备及存储介质。

背景技术

[0002] 相机拍摄的图像一般都是二维图像,丢失了图像所在场景的深度信息。从二维图像中恢复出图像的场景深度信息,有助于更好的理解场景的三维结构以及图像中物体之间的三维关系,从而更好地完成视觉任务。现有的恢复图像的场景深度信息的方法一般是通过神经网络模型进行恢复。神经网络模型包括编码器和解码器,编码器通过对图像逐级下采样的方式来增大特征图的感受野以及降低计算的开销,解码器用于从下采样的图像中恢复出具有原始图像分辨率的场景深度信息。但是目前用于深度估计的神经网络模型中,只是简单的将编码器特征与解码器特征进行混合,忽略了编码器与解码器特征图之间的语义代沟,从而降低了恢复出的场景深度信息的精度。

发明内容

[0003] 有鉴于此,本申请实施例提供了图像的深度估计方法、装置、电子设备及存储介质,解决了恢复出的场景深度信息的精度不高的问题。

[0004] 本申请实施例的第一方面提供了一种图像的深度估计方法,包括:

[0005] 获取待处理图像;

[0006] 将所述待处理图像输入深度估计网络,得到所述深度估计网络输出的深度信息;所述深度估计网络包括编码器和解码器,所述编码器包括N层编码层,所述解码器包括N-1层解码层;每层编码层的输入图像是对所述待处理图像进行预设倍数的下采样得到的,每层解码层的输入图像是根据N层编码层的输出图像进行拼接得到的,所述深度信息根据N-1层解码层的输出图像确定,N表示大于0的整数。

[0007] 在一种可能的实现方式中,第N-1层解码层的输入图像由第一输入图像和第二输入图像进行拼接得到,所述第一输入图像由第N层编码层的输出图像依次经过精调处理、上采样得到,所述第二输入图像由第i层编码层的输出图像依次进行 2^{N-1-i} 倍下采样、精调处理得到, $i=0,1,2,\dots,N-1$,所述精调处理用于对图像的局部特征信息和全局特征信息进行拼接。

[0008] 在一种可能的实现方式中,第k层解码层的输入图像由第三输入图像和第四输入图像进行拼接得到,所述第三输入图像由第k+1层解码层的输出图像依次经过精调处理、上采样得到,所述第四输入图像由第i层编码层的输出图像依次经过 2^{k-i} 倍下采样、精调处理得到,所述精调处理用于对图像的局部特征信息和全局特征信息进行拼接, $k=0,1,2,\dots,N-2,k-i\geq 0$ 。

[0009] 在一种可能的实现方式中,图像的精调处理过程包括:

[0010] 提取待精调处理的图像的局部特征图,所述局部特征图包括局部特征信息;

- [0011] 对所述待精调处理的图像进行不重叠采样,得到互不重叠的子特征图;
- [0012] 根据所述子特征图确定全局特征图,所述全局特征图包括全局特征信息;
- [0013] 对所述待精调处理的图像、所述局部特征图、所述全局特征图进行拼接,得到精调处理后的图像。
- [0014] 在一种可能的实现方式中,根据所述子特征图确定全局特征图,包括:
- [0015] 确定与子特征图对应的特征向量;
- [0016] 对所述特征向量依次进行特征向量变换、线性变换、分辨率恢复、膨胀处理,得到所述全局特征图。
- [0017] 在一种可能的实现方式中,图像的上采样过程包括:
- [0018] 按照通道维度将待上采样的图像分割为第一低分辨率图像和第二低分辨率图像;
- [0019] 采用双线性差值方式对所述第一低分辨率图像进行上采样,得到第一高分辨率图像;
- [0020] 对所述第二低分辨率图像进行膨胀处理和精调处理,得到第二高分辨率图像;
- [0021] 对所述第一高分辨率图像和所述第二高分辨率图像进行拼接,得到上采样后的图像。
- [0022] 在一种可能的实现方式中,所述深度信息由第0层解码层的输出图像进行上采样得到。
- [0023] 本申请实施例的第二方面提供了一种图像的深度估计装置,包括:
- [0024] 获取模块,用于获取待处理图像;
- [0025] 处理模块,用于将所述待处理图像输入深度估计网络,得到所述深度估计网络输出的深度信息;所述深度估计网络包括编码器和解码器,所述编码器包括N层编码层,所述解码器包括N-1层解码层;每层编码层的输入图像是对所述待处理图像进行预设倍数的下采样得到的,每层解码层的输入图像是根据N层编码层的输出图像进行拼接得到的,所述深度信息根据N-1层解码层的输出图像确定,N表示大于0的整数。
- [0026] 本申请实施例的第三方面提供了一种电子设备,包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机程序,所述处理器执行所述计算机程序时实现如上述第一方面所述的图像的深度估计方法。
- [0027] 本申请实施例的第四方面提供了一种计算机可读存储介质,所述计算机可读存储介质存储有计算机程序,所述计算机程序被处理器执行时实现如上述第一方面所述的图像的深度估计方法。
- [0028] 本申请实施例的第五方面提供了一种计算机程序产品,当计算机程序产品在电子设备上运行时,使得电子设备执行上述第一方面中任一项所述的图像的深度估计方法。
- [0029] 本申请实施例与现有技术相比存在的有益效果是:深度估计网络包括编码器和解码器,编码器包括N层编码层,解码器包括N-1层解码层;每层编码层的输入图像是对待处理图像进行预设倍数的下采样得到的,每层解码层的输入图像是根据N层编码层的输出图像进行拼接得到的,因此,每层解码层的输入信息与N层解码层的输出信息均存在关联,从而使得每层解码层的输入信息融合了编码层的多级语义信息,进而降低了编码器和解码器的特征图之间的语义代沟。之后,再将待处理图像输入深度估计网络,根据N-1层解码层的输出图像确定深度信息,提高了恢复出的场景深度信息的精度。

附图说明

[0030] 为了更清楚地说明本申请实施例中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍。

[0031] 图1是本申请一实施例提供的图像的深度估计方法的实现流程示意图;

[0032] 图2是本申请实施例提供的深度估计网络的示意图;

[0033] 图3是本申请实施例提供的精调处理函数的计算流程示意图;

[0034] 图4是本申请实施例提供的上采样函数的计算流程示意图;

[0035] 图5是本申请实施例提供的图像的深度估计装置示意图;

[0036] 图6是本申请实施例提供的电子设备的结构示意图。

具体实施方式

[0037] 以下描述中,为了说明而不是为了限定,提出了诸如特定系统结构、技术之类的具体细节,以便透彻理解本申请实施例。然而,本领域的技术人员应当清楚,在没有这些具体细节的其它实施例中也可以实现本申请。在其它情况中,省略对众所周知的系统、装置、电路以及方法的详细说明,以免不必要的细节妨碍本申请的描述。

[0038] 应当理解,当在本说明书和所附权利要求书中使用时,术语“包括”指示所描述特征、整体、步骤、操作、元素和/或组件的存在,但并不排除一个或多个其它特征、整体、步骤、操作、元素、组件和/或其集合的存在或添加。

[0039] 还应当理解,在此本申请说明书中所使用的术语仅仅是出于描述特定实施例的目的而并不意在限制本申请。如在本申请说明书和所附权利要求书中所使用的那样,除非上下文清楚地指明其它情况,否则单数形式的“一”、“一个”及“该”意在包括复数形式。

[0040] 另外,在本申请的描述中,术语“第一”、“第二”、“第三”等仅用于区分描述,而不能理解为指示或暗示相对重要性。

[0041] 目前用于深度估计的神经网络模型中,只是简单的将编码器特征与解码器特征进行混合,忽略了编码器与解码器特征图之间的语义代沟,从而降低了恢复出的场景深度信息的精度。

[0042] 为此,本申请提供一种图像的深度估计方法,将待处理图像输入深度估计网络,深度估计网络包括编码器和解码器,编码器包括N层编码层,解码器包括N-1层解码层。每层解码层的输入图像是根据N层编码层的输出图像进行拼接得到的,从而使得每层解码层的输入信息均与各层编码层的输出信息关联,使得每层解码层的输入信息融合了编码层的多级语义信息,降低了编码器和解码器的特征图之间的语义代沟。再根据N-1层解码层的输出图像确定深度信息,提高了恢复出的场景深度信息的精度。

[0043] 下面对本申请提供的图像的深度估计方法进行示例性说明。

[0044] 请参阅附图1,本申请一实施例提供的图像的深度估计方法包括:

[0045] S101:获取待处理图像。

[0046] 其中,待处理图像是二维图像,用于恢复出场景深度信息。

[0047] 在一实施例中,待处理图像是对原始图像进行预处理后得到的。预处理包括随机翻转、随机裁剪、数据归一化等处理过程。示例性的,预处理后得到维度为 $C \times H \times W$ 的张量数据,也即待处理图像的维度为 $C \times H \times W$,C表示通道数量,H表示图像高度,W表示图像宽度。

[0048] 在一实施例中,原始图像是自动驾驶数据集KITTI的单目图像。

[0049] S102:将所述待处理图像输入深度估计网络,得到所述深度估计网络输出的深度信息;所述深度估计网络包括编码器和解码器,所述编码器包括N层编码层,所述解码器包括N-1层解码层;每层编码层的输入图像是对所述待处理图像进行预设倍数的下采样得到的,每层解码层的输入图像是根据N层编码层的输出图像进行拼接得到的,所述深度信息根据N-1层解码层的输出图像确定,N表示大于0的整数。在一实施例中,待处理图像是第0层编码层的输入图像,i大于0时,第i层编码层输入图像是对第i-1层编码层的输入图像依次进行特征提取和2倍的下采样,得到的。例如,对第0层编码层的输入图像依次进行特征提取和2倍的下采样,得到第1层编码层的输入图像;对第1层编码层的输入图像依次进行特征提取和2倍的下采样,得到第2层编码层的输入图像;对第2层编码层的输入图像依次进行特征提取和2倍的下采样,得到第3层编码层的输入图像。

[0050] 在另一实施例中,待处理图像是第0层编码层的输入图像,对待处理图像分别依次进行特征提取和不同倍数的下采样,得到每层编码层的输入图像。例如,对待处理图像依次进行特征提取和2倍的下采样所得到的图像是第1层编码层的输入图像,对待处理图像依次进行特征提取和4倍的下采样所得到的图像是第2层编码层的输入图像…对待处理图像依次进行特征提取和 2^N 倍下采样所得到的图像是第N层编码层的输入图像。

[0051] 每层解码层对应一个拼接规则,根据拼接规则对N层编码层的输出图像进行拼接,得到每层解码层的输入图像。对每层解码层的输出图像进行处理,得到深度信息。

[0052] 在一实施例中,第N-1层解码层的输入图像由第一输入图像和第二输入图像进行拼接得到,所述第一输入图像由第N层编码层的输出图像依次经过精调处理、上采样得到,所述第二输入图像由第i层编码层的输出图像依次进行 2^{N-1-i} 倍下采样、精调处理得到, $i=0,1,2,\dots,N-1$,所述精调处理用于对图像的局部特征信息和全局特征信息进行拼接。将第一输入图像和第二输入图像进行拼接后输入第N-1层解码层,可以使得第N-1层解码层的输入信息融合N层编码层的语义信息,降低了编码器和解码器的特征图之间的语义代沟。

[0053] 在一实施例中,第k层解码层的输入图像由第三输入图像和第四输入图像进行拼接得到,所述第三输入图像由第k+1层解码层的输出图像依次经过精调处理、上采样得到,所述第四输入图像由第i层编码层的输出图像依次经过 2^{k-i} 倍下采样、精调处理得到,所述精调处理用于对图像的局部特征信息和全局特征信息进行拼接, $k=0,1,2,\dots,N-2,k-i\geq 0$ 。将第三输入图像和第四输入图像进行拼接后输入第k层解码层,可以使得每层解码层的输入信息融合N层编码层的语义信息,降低了编码器和解码器的特征图之间的语义代沟。

[0054] 在一实施例中,第一输入图像和第二输入图像进行拼接之后,对拼接后的图像进行精调处理,得到第N-1层解码层的输入图像。第三输入图像和第四输入图像进行拼接之后,对拼接后的图像进行精调处理,得到第k层解码层的输入图像。

[0055] 示例性的,根据公式

$$[0056] \quad X_{dec}^k = \begin{cases} F_{refine}(F_{up}(F_{refine}(X_{enc}^N)) + \sum_{i=0}^{N-1} X_{enc}^{(i,N-1-i)}), & k = N-1 \\ F_{refine}(F_{up}(F_{refine}(X_{dec}^{k+1})) + \sum_{i=0}^k X_{enc}^{(i,k-i)}), & 0 \leq k \leq N-2 \end{cases}$$

[0057] 确定每层解码层的输入图像,其中, X_{dec}^k 表示第k层解码层的输入图像, F_{refine} 表示精调处理函数, F_{up} 表示上采样函数, X_{enc}^N 表示第N层编码层的输出图像, X_{dec}^{k+1} 表示第k+1层解码层的输出图像; $F_{down}^{(i,j)}$ 表示对第i阶段输入特征进行 2^j 倍下采样,由步长为2,核大小为 3×3 的卷积核与激活函数组成; $X_{enc}^{(i,j)} = F_{refine}(F_{down}^{(i,j)}(X_{enc}^i))$ 表示将第i阶段编码器层特征图像依次进行 2^j 倍下采样和信息精调后所得特征图像,

[0058] $j=N-1-i$ 或 $j=k-i$ 。

[0059] 在一实施例中,在对图像进行精调处理之前,先对图像进行通道变换,再对通道变换得到的图像进行精调处理,从而对图像的通道进行压缩,使得图像的信息更加紧凑,降低计算开销。

[0060] 在一实施例中,深度信息由第0层解码层的输出图像进行上采样得到。在各层解码层的输出图像中,第0层解码层的输出图像的分辨率最高,根据第0层解码层的输出图像确定深度信息,提高了确定出的深度信息的准确度。

[0061] 在一实施例中, $N=4$,深度估计网络的结构如图2所示。对原始图像进行预处理得到待处理图像,也即第0层编码层enc0的输入图像,对第0层编码层的输入图像依次进行特征提取和2倍的下采样,得到第1层编码层enc1的输入图像,对第1层编码层的输入图像依次进行特征提取和2倍的下采样,得到第2层编码层enc2的输入图像,对第2层编码层的输入图像依次进行特征提取和2倍的下采样,得到第3层编码层enc3的输入图像,对第3层编码层的输入图像依次进行特征提取和2倍的下采样,得到第4层编码层enc4的输入图像。对第4层编码层的输出图像依次进行通道变换、信息精调、上采样,得到第一输入图像。对第0层编码层的输出图像依次进行8倍下采样、精调处理,对第1层编码层的输出图像依次进行特征提取、4倍下采样、精调处理,对第2层编码层的输出图像依次进行特征提取、2倍下采样、精调处理,对第3层编码层的输出图像依次进行特征提取、1倍下采样、精调处理,得到第二输入图像,对第一输入图像和第二输入图像进行拼接,对拼接后的图像再依次进行通道变换、精调处理,得到第3层解码层的输入图像。其中,图像进行1倍下采样所得到的图像还是采样前的图像。

[0062] 对第3层解码层dec3的输出图像依次进行通道变换、信息精调、上采样,得到第2层解码层的第三输入图像。对第0层编码层的输出图像依次进行特征提取、4倍下采样、精调处理,对第1层编码层的输出图像依次进行特征提取、2倍下采样、精调处理,对第2层编码层的输出图像依次进行特征提取、1倍下采样、精调处理,得到第2层解码层的第四输入图像,对第2层解码层的第三输入图像和第四输入图像进行拼接,对拼接后的图像再依次进行通道变换、精调处理,得到第2层解码层的输入图像。

[0063] 对第2层解码层dec2的输出图像依次进行通道变换、信息精调、上采样,得到第1层解码层的第三输入图像。对第0层编码层的输出图像依次进行特征提取、2倍下采样、精调处理,对第1层编码层的输出图像依次进行特征提取、1倍下采样、精调处理,得到第1层解码层的第四输入图像,对第1层解码层的第三输入图像和第四输入图像进行拼接,对拼接后的图像再依次进行通道变换、精调处理,得到第1层解码层的输入图像。

[0064] 对第1层解码层dec1的输出图像依次进行通道变换、信息精调、上采样,得到第0层

解码层的第三输入图像。对第0层编码层的输出图像依次进行特征提取、1倍下采样、精调处理,得到第0层解码层的第四输入图像,对第0层解码层的第三输入图像和第四输入图像进行拼接,对拼接后的图像再依次进行通道变换、精调处理,得到第0层解码层的输入图像。对第0层解码层的输出图像进行上采样,得到深度信息。

[0065] 上述实施例中,将待处理图像输入深度估计网络,深度估计网络包括编码器和解码器,编码器包括N层编码层,解码器包括N-1层解码层;每层编码层的输入图像是待处理图像进行预设倍数的下采样得到的,每层解码层的输入图像是根据N层编码层的输出图像进行拼接得到的,因此,每层解码层的输入信息与N层解码层的输出信息均存在关联,从而使得每层解码层的输入信息融合了编码层的多级语义信息,进而降低了编码器和解码器的特征图之间的语义代沟。之后,再将待处理图像输入深度估计网络,根据N-1层解码层的输出图像确定深度信息,提高了恢复出的场景深度信息的精度。

[0066] 在一实施例中,图像的精调处理过程包括:提取待精调处理的图像的局部特征图,所述局部特征图包括局部特征信息;对所述待精调处理的图像进行不重叠采样,得到互不重叠的子特征图;根据所述子特征图确定全局特征图,所述全局特征图包括全局特征信息;对所述待精调处理的图像、所述局部特征图、所述全局特征图进行拼接,得到精调处理后的图像。

[0067] 通过精调处理的过程,可以对特征图的像素长范围依赖能力进行建模,得到同时具有局部特征信息和全局特征信息的特征图像,为图像中的像素提供更好的语义信息,更好的推理出像素的深度信息。

[0068] 在一实施例中,根据子特征图确定全局特征图的流程为:确定与子特征图对应的特征向量,对特征向量依次进行特征向量变换、非线性变换、分辨率恢复、膨胀处理,得到全局特征图,从而提高了得到的全局特征图的非线性拟合能力。

[0069] 在其他实施例中,也可以采用预先训练的深度卷积神经网络模型提取待精调处理的图像的局部特征图以及全局特征图。

[0070] 示例性地,精调处理函数 F_{refine} 的计算流程如图3所示。

[0071] 设定 $X_{\text{refine}} = F_{\text{refine}}(X)$, X 表示待精调处理的特征图, $X_{\text{refine}} \in \mathbb{R}^{C \times H \times W}$ 表示精调处理后的图像。本申请实施例中,待精调处理的特征图可以是第N层编码层的输出图像、第i层编码层的输出图像、第k+1层解码层的输出图像等。

[0072] 首先,根据公式

$$[0073] \quad \begin{cases} X_1 = F_{\text{squeeze}}(X), X \in \mathbb{R}^{C \times H \times W}, X_1 \in \mathbb{R}^{\frac{C}{4} \times H \times W} \\ X_2 = F_{\text{local}}(X_1), X_2 \in \mathbb{R}^{\frac{C}{4} \times H \times W} \\ X_3 = F_{\text{unsqueeze}}(X_2), X_3 \in \mathbb{R}^{C \times H \times W} \end{cases} \quad \text{确定待精调处理的特征图X}$$

的局部特征图 X_3 。其中,函数 F_{squeeze} 由 1×1 的卷积核和激活函数构成,用于对特征图进行压缩处理,函数 F_{local} 由 3×3 的卷积核和激活函数构成,表示局部特征提取器,用于感知特征图的局部上下文信息,函数 $F_{\text{unsqueeze}}$ 由 1×1 卷积核和激活函数组成,表示特征图的膨胀处理操

作,用于通道维度的恢复, \mathfrak{R} 表示特征图空间, C 表示通道数量, H 表示图像高度, W 表示图像宽度,待精调处理的特征图 X 的维度为 $C \times H \times W$,局部特征图 X_3 的维度也为 $C \times H \times W$ 。

[0074] 然后,根据公式

$$[0075] \quad \begin{cases} x = F_{embed}(F_{sample}(X_1)), x \in \mathfrak{R}^{\frac{H}{H_{sub}} * \frac{W}{W_{sub}} \times C_{embed}} \\ Q, K, V = F_{split}(F_{reshape}(F_{linear}(x))), Q, K, V \in \mathfrak{R}^{N_{att} \times \frac{H}{H_{sub}} * \frac{W}{W_{sub}} \times \frac{C_{embed}}{N}} \\ x_1 = F_{linear}(F_{reshape}(Softmax(QK^T * s)V)), x_1 \in \mathfrak{R}^{\frac{H}{H_{sub}} * \frac{W}{W_{sub}} \times C_{embed}} \\ x_2 = x + x_1 + F_{mlp}(x + x_1), x_2 \in \mathfrak{R}^{\frac{H}{H_{sub}} * \frac{W}{W_{sub}} \times C_{embed}} \\ X_4 = F_{rrl}(x_2), X_4 \in \mathfrak{R}^{\frac{C}{4} \times H \times W} \\ X_5 = F_{unsqueeze}(X_4), X_5 \in \mathfrak{R}^{C \times H \times W} \end{cases}$$

[0076] 确定待精调处理的特征图 X 的全局特征图。其中, X_1 表示特征图 X 压缩处理后的特征图,函数 F_{sample} 表示特征图的不重叠采样操作,用于将特征图采样为互不重叠的子特征图,由卷积核长度和宽度分别为子特征图的长度和宽度、步长的长度和宽度分别为子特征图的长度和宽度的卷积核,以及激活函数组成,函数 F_{embed} 用于将二维的子特征图变换成一维特征向量,再将一维特征向量维度与通道维度进行交换, x 表示将子特征图嵌入成的特征向量, $H_{sub}, W_{sub}, C_{embed}$ 分别表示子特征图的高度、子特征图的宽度、得到的特征向量的数量。函数 F_{linear} 表示线性变换函数, $F_{linear}(x)$ 用于将特征向量 x 线性变换成三个相同数量的特征向量, $F_{reshape}$ 表示特征向量的维度变换函数,用于将特征向量调整成指定形状, F_{split} 用于将特征向量划分为 Q (Query), K (Key), V (Value) 三个子特征向量,其中 Q 作为查询特征向量, K 作为键值特征向量, V 作为值特征向量, N_{att} 表示注意力头的个数, s 表示标量缩放因子,“ T ”表示转置运算,“ $*$ ”表示乘法运算, $Softmax$ 表示逻辑回归函数。对逻辑回归处理后的特征向量再进行维度变换以及线性变换,即是对逻辑回归处理后的特征向量进行逆变换,得到逆变换后的特征向量 x_1 。 F_{mlp} 表示由两层线性变换层和激活函数组成的多层感知器,用于增强特征向量的非线性拟合能力。 F_{rrl} 表示分辨率恢复函数,由一个 1×1 的卷积核,以及卷积核长度和宽度分别为子特征图的长度和宽度、步长的长度和宽度分别为子特征图的长度和宽度的反卷积组成, $F_{rrl}(x_2)$ 表示将特征向量 x_2 变换成与特征图 X_1 相同维度的全局特征图 X_4 。 $F_{unsqueeze}$ 表示特征图的膨胀处理操作, $F_{unsqueeze}(X_4)$ 用于将全局特征图 X_4 变换成与特征图 X 具有相同维度的全局特征图 X_5 。

[0077] 在得到局部特征图和全局特征图之后,根据公式

[0078] $X_{refine} = X + X_3 + X_5, X_{refine} \in \mathfrak{R}^{C \times H \times W}$ 对待精调处理的特征图、局部特征图、全局特征图进行拼接,得到精调处理后的图像 $X_{refine} \in \mathfrak{R}^{C \times H \times W}$ 。

[0079] 在一实施例中,图像的上采样过程包括:按照通道维度将待上采样的图像分割为第一低分辨率图像和第二低分辨率图像;采用双线性差值方式对所述第一低分辨率图像进行上采样,得到第一高分辨率图像;对所述第二低分辨率图像进行膨胀处理和精调处理,得到第二高分辨率图像;对所述第一高分辨率图像和所述第二高分辨率图像进行拼接,得到上采样后的图像。

[0080] 对图像进行上采样处理,可以弥补双线性插值上采样无法从低分辨率特征图中恢复出细粒度高分辨率特征图的缺点,提高了从低分辨率特征图中恢复出细粒度的高分辨率特征图像的能力。

[0081] 在其他实施例中,也可以采用最邻近插值算法、转置卷积算法或者上池化等方法对图像进行上采样。

[0082] 示例性地,上采样函数 F_{up} 的计算流程如图4所示。

[0083] 设定 $X_{high} = F_{up}(X_{low})$, $X_{low} \in \mathbb{R}^{C \times H \times W}$ 表示待精调上采样的特征图, $X_{high} \in \mathbb{R}^{C \times 2H \times 2W}$ 表示精调上采样后的图像。

[0084] 根据公式 $X_{low}^1, X_{low}^2 = F_{split}(X_{low})$, $X_{low}^1, X_{low}^2 \in \mathbb{R}^{\frac{C}{2} \times H \times W}$ 将待精调上采样处理的特征图 $X_{low} \in \mathbb{R}^{C \times H \times W}$ 按照通道维度均等划分成两个子特征图 X_{low}^1 和 X_{low}^2 , X_{low}^1 为第一低分辨率特征图,即低分辨率特征图A, X_{low}^2 为第二低分辨率特征图,即低分辨率特征图B, 待精调上采样处理的特征图 $X_{low} \in \mathbb{R}^{C \times H \times W}$ 的维度为 $C \times H \times W$, X_{low}^1 和 X_{low}^2 的维度均为 $X_{low}^1, X_{low}^2 = F_{split}(X_{low})$, $X_{low}^1, X_{low}^2 \in \mathbb{R}^{\frac{C}{2} \times H \times W}$ 。

[0085] 对于第一低分辨率特征图 X_{low}^1 ,采用双线性插值方式对 X_{low}^1 进行上采样,得到第一高分辨率特征图,即高分辨率特征图A。

[0086] 具体地,根据公式 $X_{high}^1 = F_{bilinear}(X_{low}^1)$, $X_{high}^1 \in \mathbb{R}^{\frac{C}{2} \times H \times W}$ 计算第一高分辨率特征图, 其中, X_{high}^1 表示第一高分辨率特征图, $F_{bilinear}$ 表示双线性插值函数。

[0087] 对于第二低分辨率特征图 X_{low}^2 ,根据公式

$X_{low}^3 = F_{expand}(X_{low}^2)$, $X_{low}^3 \in \mathbb{R}^{2C \times H \times W}$ 对 X_{low}^2 进行膨胀处理,得到膨胀处理后的特征图 X_{low}^3 ,其中, F_{expand} 表示膨胀函数,由核大小为 1×1 的卷积核和激活函数组成。

[0088] 之后,根据公式 $X_{low}^4 = F_{refine}(X_{low}^3)$, $X_{low}^4 \in \mathbb{R}^{2C \times H \times W}$ 对 X_{low}^3 进行精调处理,得到精调处理后的特征图 X_{low}^4 。 X_{low}^4 的维度为 $2C \times H \times W$,根据公式

[0089] $X_{high}^2 = F_{pixelshuffle}(X_{low}^4)$, $X_{high}^2 \in \mathbb{R}^{\frac{C}{2} \times 2H \times 2W}$ 将 X_{low}^4 重新排列成维度为

$\frac{C}{2} \times 2H \times 2W$ 的第二高分辨率特征图, X_{high}^2 表示第二高分辨率特征图, 即高分辨率特征图B,

$F_{\text{pixelshuffle}}$ 表示元素重排列函数。

[0090] 之后, 将第一高分辨率特征图和第二高分辨率特征图B按照通道维度进行拼接, 得到表示精调上采样后的图像 $X_{high} \in \mathbb{R}^{C \times 2H \times 2W}$ 。具体地, 根据公式

$$X_{high} = F_{\text{concat}}(X_{high}^1, X_{high}^2), X_{high} \in \mathbb{R}^{C \times 2H \times 2W} \text{ 得到精调上采样后的图像}$$

$X_{high} \in \mathbb{R}^{C \times 2H \times 2W}$ 。其中, F_{concat} 表示拼接函数。

[0091] 在一实施例中, 在估计待处理图像的深度信息之前, 需要对深度估计网络进行训练, 以得到深度估计网络的最优参数。其中, 训练深度估计网络的训练样本为样本图像序列, 样本图像序列包括目标帧图像和参考帧图像, 将样本图像序列输入深度估计网络, 得到深度估计网络输出的估样本图像序列的深度信息。将样本图像序列输入相机运动网络, 得到相机运动网络输出的样本图像序列之间的相对位姿。之后, 根据参考帧图像的深度信息、相机内参将参考帧图像仿射变换到三维空间, 得到三维空间中的参考帧图像。根据相机运动网络输出的参考帧图像与目标帧图像之间的相对位姿以及相机内参, 将三维空间中的参考帧图像仿射变换到对应的目标帧图像三维空间, 然后将对应的目标帧图像三维空间投影到二维平面, 得到仿射变换后的图像坐标。之后, 根据仿射变换后的图像坐标合成目标帧图像, 将训练样本中的目标帧图像与合成的目标帧图像之间的差异作为指导信号来联合指导深度估计网络和相机运动网络的参数更新。直到获得深度估计网络与相机运动网络的最优参数或达到预设的迭代次数。

[0092] 其中, 在训练阶段, 对深度估计网络的第0层至第N-2层解码层的输出图像依次进行上采样和视差输出层得到分别与待处理图像至第N-2层解码层的输出图像相同分辨率的深度图像, 根据深度图像确定深度信息。示例性地, $N=4$, 对第0层至第2层解码层的输出图像依次经过上采样、视差输出层, 得到分别与待处理图像、第0层解码层的输出图像、第2层解码层的输出图像相同分辨率的深度图像, 共3种尺度的深度图像, 根据深度图像重建多尺度目标图, 通过待处理图像与重建出的多尺度目标图像之间的差异, 来指导网络权重的更新。

[0093] 在得到深度估计网络的最优参数后, 深度估计网络可以估计待处理图像的深度信息, 相机运动网络可以估计待处理图像所在的图像序列中, 相邻帧之间的相对位姿。

[0094] 在一实施例中, 训练样本中的图像的维度为 $C \times H \times W$, $H=256$, $W=832$, 输入深度估计网络的图像中, $C=3$, 输入相机运动网络的图像中, $C=9$ 。

[0095] 本申请实施例中, 深度估计网络中, 每层解码层的输入图像是根据N层编码层的输出图像进行拼接得到的, 因此, 每层解码层的输入信息与N层解码层的输出信息均存在关联, 从而使得每层解码层的输入信息融合了编码层的多级语义信息, 进而降低了编码器和解码器的特征图之间的语义代沟。

[0096] 第一输入图像由第N层编码层的输出图像依次经过精调处理、上采样得到, 第二输入图像由第i层编码层的输出图像依次进行 2^{N-1-i} 倍下采样、精调处理得到, 将第一输入图像和第二输入图像进行拼接后输入第N-1层解码层, 可以使得第N-1层解码层的输入信息融合

N层编码层的语义信息,降低了编码器和解码器的特征图之间的语义代沟。

[0097] 第三输入图像由第k+1层解码层的输出图像依次经过精调处理、上采样得到,第四输入图像由第i层编码层的输出图像依次经过 2^{k-i} 倍下采样、精调处理得到,将第三输入图像和第四输入图像进行拼接后输入第k层解码层,可以使得每层解码层的输入信息融合N层编码层的语义信息,降低了编码器和解码器的特征图之间的语义代沟。

[0098] 通过精调处理的过程,可以对特征图的像素长范围依赖能力进行建模,得到同时具有局部特征信息和全局特征信息的特征图像,为图像中的像素提供更好的语义信息,更好的推理出像素的深度信息。

[0099] 通过对图像进行上采样处理,可以弥补双线性插值上采样无法从低分辨率特征图中恢复出细粒度高分辨率特征图的缺点,提高了从低分辨率特征图中恢复出细粒度的高分辨率特征图像的能力。

[0100] 应理解,上述实施例中各步骤的序号的大小并不意味着执行顺序的先后,各过程的执行顺序应以其功能和内在逻辑确定,而不对本申请实施例的实施过程构成任何限定。

[0101] 对应于上文实施例所述的图像的深度估计方法,图5示出了本申请实施例提供的图像的深度估计装置的结构框图,为了便于说明,仅示出了与本申请实施例相关的部分。

[0102] 如图5所示,图像的深度估计装置包括,

[0103] 获取模块51,用于获取待处理图像;

[0104] 处理模块52,用于将所述待处理图像输入深度估计网络,得到所述深度估计网络输出的深度信息;所述深度估计网络包括编码器和解码器,所述编码器包括N层编码层,所述解码器包括N-1层解码层;每层编码层的输入图像是对所述待处理图像进行预设倍数的下采样得到的,每层解码层的输入图像是根据N层编码层的输出图像进行拼接得到的,所述深度信息根据N-1层解码层的输出图像确定,N表示大于0的整数。

[0105] 在一种可能的实现方式中,第N-1层解码层的输入图像由第一输入图像和第二输入图像进行拼接得到,所述第一输入图像由第N层编码层的输出图像依次经过精调处理、上采样得到,所述第二输入图像由第i层编码层的输出图像依次进行 2^{N-1-i} 倍下采样、精调处理得到, $i=0,1,2,\dots,N-1$,所述精调处理用于对图像的局部特征信息和全局特征信息进行拼接。

[0106] 在一种可能的实现方式中,第k层解码层的输入图像由第三输入图像和第四输入图像进行拼接得到,所述第三输入图像由第k+1层解码层的输出图像依次经过精调处理、上采样得到,所述第四输入图像由第i层编码层的输出图像依次经过 2^{k-i} 倍下采样、精调处理得到,所述精调处理用于对图像的局部特征信息和全局特征信息进行拼接, $k=0,1,2,\dots,N-2,k-i\geq 0$ 。

[0107] 在一种可能的实现方式中,处理模块52具体用于:

[0108] 提取待精调处理的图像的局部特征图,所述局部特征图包括局部特征信息;

[0109] 对所述待精调处理的图像进行不重叠采样,得到互不重叠的子特征图;

[0110] 根据所述子特征图确定全局特征图,所述全局特征图包括全局特征信息;

[0111] 对所述待精调处理的图像、所述局部特征图、所述全局特征图进行拼接,得到精调处理后的图像。

[0112] 在一种可能的实现方式中,处理模块52具体还用于:

[0113] 确定与子特征图对应的特征向量;

[0114] 对所述特征向量依次进行特征向量变换、非线性变换、分辨率恢复、膨胀处理,得到所述全局特征图。

[0115] 在一种可能的实现方式中,处理模块52具体还用于:

[0116] 按照通道维度将待上采样的图像分割为第一低分辨率图像和第二低分辨率图像;

[0117] 采用双线性差值方式对所述第一低分辨率图像进行上采样,得到第一高分辨率图像;

[0118] 对所述第二低分辨率图像进行膨胀处理和精调处理,得到第二高分辨率图像;

[0119] 对所述第一高分辨率图像和所述第二高分辨率图像进行拼接,得到上采样后的图像。

[0120] 在一种可能的实现方式中,所述深度信息由第0层解码层的输出图像进行上采样得到。

[0121] 需要说明的是,上述装置/单元之间的信息交互、执行过程等内容,由于与本申请方法实施例基于同一构思,其具体功能及带来的技术效果,具体可参见方法实施例部分,此处不再赘述。

[0122] 图6是本申请实施例提供的电子设备的结构示意图。所述电子设备可以是桌上型计算机、笔记本、掌上电脑等计算设备。

[0123] 如图6所示,该实施例的电子设备包括:处理器61、存储器62以及存储在所述存储器62中并可在所述处理器61上运行的计算机程序63。所述处理器61执行所述计算机程序63时实现上述图像的深度估计方法实施例中的步骤,例如图1所示的步骤S101至S102。或者,所述处理器61执行所述计算机程序63时实现上述各装置实施例中各模块/单元的功能,例如图5所示获取模块51至处理模块52的功能。

[0124] 示例性的,所述计算机程序63可以被分割成一个或多个模块/单元,所述一个或者多个模块/单元被存储在所述存储器62中,并由所述处理器61执行,以完成本申请。所述一个或多个模块/单元可以是能够完成特定功能的一系列计算机程序指令段,该指令段用于描述所述计算机程序63在所述电子设备中的执行过程。

[0125] 本领域技术人员可以理解,图6仅仅是电子设备的示例,并不构成对电子设备的限定,可以包括比图示更多或更少的部件,或者组合某些部件,或者不同的部件,例如所述电子设备还可以包括输入输出设备、网络接入设备、总线等。

[0126] 所述处理器61可以是中央处理单元(Central Processing Unit,CPU),还可以是其他通用处理器、数字信号处理器(Digital Signal Processor,DSP)、专用集成电路(Application Specific Integrated Circuit,ASIC)、现场可编程门阵列(Field-Programmable Gate Array,FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。

[0127] 所述存储器62可以是所述电子设备的内部存储单元,例如电子设备的硬盘或内存。所述存储器62也可以是所述电子设备的外部存储设备,例如所述电子设备上配备的插接式硬盘,智能存储卡(Smart Media Card,SMC),安全数字(Secure Digital,SD)卡,闪存

卡 (Flash Card) 等。进一步地,所述存储器62还可以既包括所述电子设备的内部存储单元也包括外部存储设备。所述存储器62用于存储所述计算机程序以及所述电子设备所需的其他程序和数据。所述存储器62还可以用于暂时地存储已经输出或者将要输出的数据。

[0128] 所属领域的技术人员可以清楚地了解到,为了描述的方便和简洁,仅以上述各功能单元、模块的划分进行举例说明,实际应用中,可以根据需要而将上述功能分配由不同的功能单元、模块完成,即将所述装置的内部结构划分成不同的功能单元或模块,以完成以上描述的全部或者部分功能。实施例中的各功能单元、模块可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中,上述集成的单元既可以采用硬件的形式实现,也可以采用软件功能单元的形式实现。另外,各功能单元、模块的具体名称也只是为了便于相互区分,并不用于限制本申请的保护范围。上述系统中单元、模块的具体工作过程,可以参考前述方法实施例中的对应过程,在此不再赘述。

[0129] 在本申请所提供的实施例中,应该理解到,所揭露的装置/电子设备和方法,可以通过其它的方式实现。例如,以上所描述的装置/电子设备实施例仅仅是示意性的,例如,所述模块或单元的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,例如多个单元或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另一点,所显示或讨论的相互之间的耦合或直接耦合或通讯连接可以是通过一些接口,装置或单元的间接耦合或通讯连接,可以是电性,机械或其它的形式。

[0130] 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

[0131] 集成的模块/单元如果以软件功能单元的形式实现并作为独立的产品销售或使用时,可以存储在一个计算机可读取存储介质中。基于这样的理解,本申请实现上述实施例方法中的全部或部分流程,也可以通过计算机程序来指令相关的硬件来完成,所述的计算机程序可存储于一计算机可读存储介质中,该计算机程序在被处理器执行时,可实现上述各个方法实施例的步骤。其中,所述计算机程序包括计算机程序代码,所述计算机程序代码可以为源代码形式、对象代码形式、可执行文件或某些中间形式等。所述计算机可读介质可以包括:能够携带所述计算机程序代码的任何实体或装置、记录介质、U盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器 (ROM, Read-Only Memory)、随机存取存储器 (RAM, Random Access Memory)、电载波信号、电信信号以及软件分发介质等。

[0132] 本领域普通技术人员可以意识到,结合本文中所公开的实施例描述的各示例的单元及算法步骤,能够以电子硬件、或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行,取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能,但是这种实现不应认为超出本申请的范围。

[0133] 以上所述实施例仅用以说明本申请的技术方案,而非对其限制;尽管参照前述实施例对本申请进行了详细的说明,本领域的普通技术人员应当理解:其依然可以对前述各实施例所记载的技术方案进行修改,或者对其中部分技术特征进行等同替换;而这些修改或者替换,并不使相应技术方案的本质脱离本申请各实施例技术方案的精神和范围,均应

包含在本申请的保护范围之内。

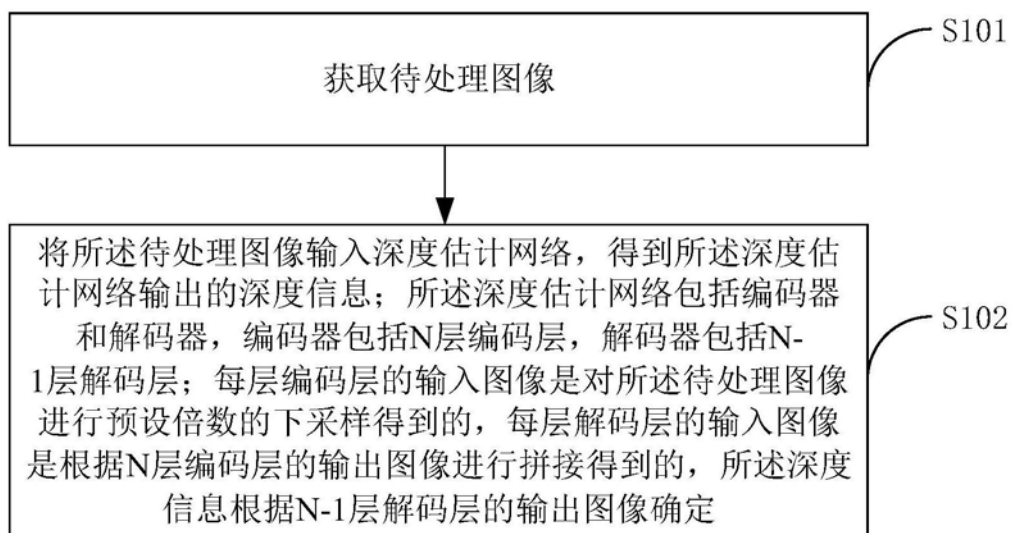


图1

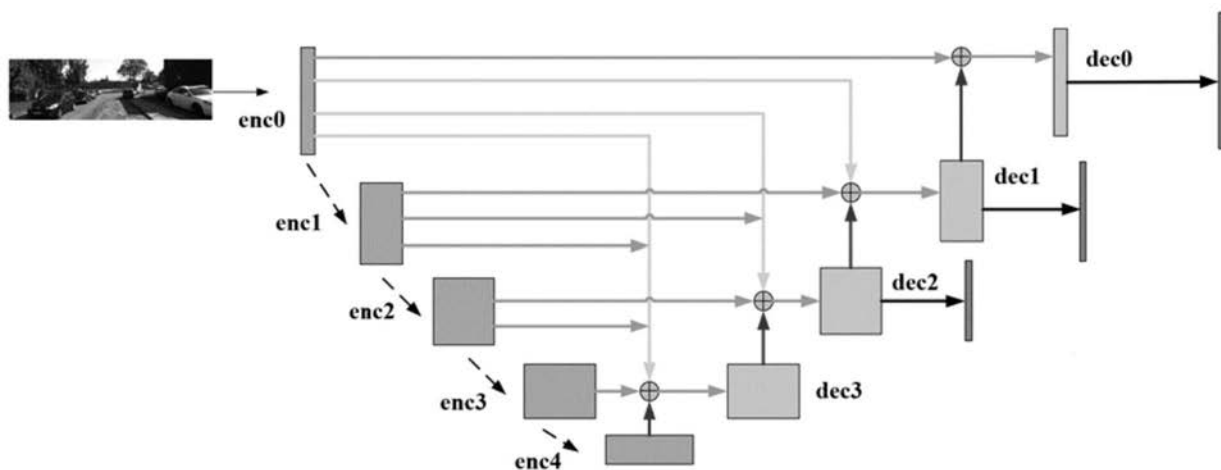


图2

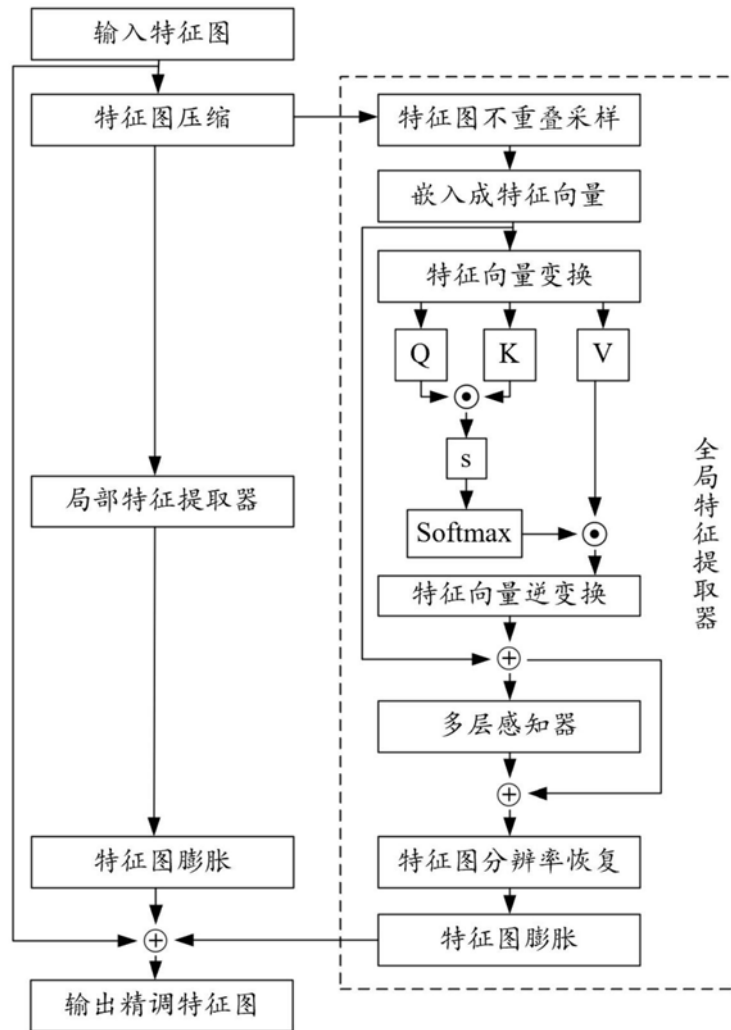


图3

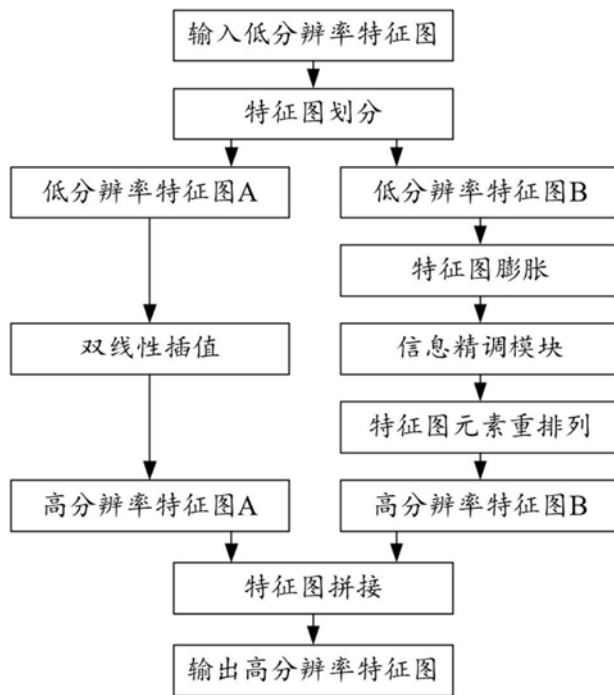


图4

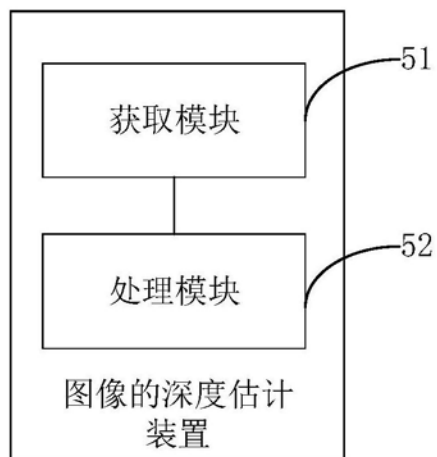


图5

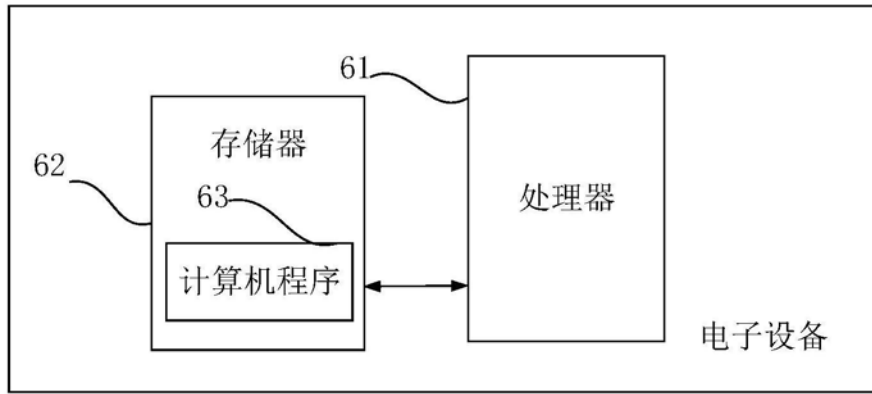


图6