



(12)发明专利

(10)授权公告号 CN 107767413 B

(45)授权公告日 2020.02.18

(21)申请号 201710850577.5

(22)申请日 2017.09.20

(65)同一申请的已公布的文献号

申请公布号 CN 107767413 A

(43)申请公布日 2018.03.06

(73)专利权人 华南理工大学

地址 510640 广东省广州市天河区五山路381号

(72)发明人 李格 余翔宇

(74)专利代理机构 广州市华学知识产权代理有限公司 44245

代理人 李斌

(51)Int.Cl.

G06T 7/50(2017.01)

(56)对比文件

CN 105488534 A,2016.04.13,

CN 105657402 A,2016.06.08,

CN 106981080 A,2017.07.25,

CN 106991646 A,2017.07.28,

CN 106157307 A,2016.11.23,

CN 106599805 A,2017.04.26,

Simon Je'gou 等.The One Hundred Layers Tiramisu:Fully Convolutional DenseNets for Semantic Segmentation.《CVPRW 2017》.2017,

许路 等.基于深层卷积神经网络的单目红外图像深度估计.《光学学报》.2016,第36卷(第7期),

Saurav Sharma 等.DenseNet with pre-activated deconvolution for estimating depth map from single image.《AMMDS 2017》.2017,

姚广顺 等.基于红外与雷达的夜间无人车场景深度估计.《激光与光电子学进展》.2017,

审查员 黄文琪

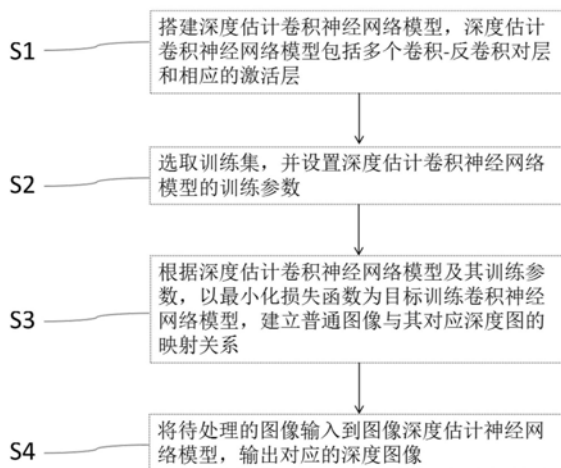
权利要求书1页 说明书6页 附图1页

(54)发明名称

一种基于卷积神经网络的图像深度估计方法

(57)摘要

本发明公开了一种基于卷积神经网络的图像深度估计方法,包括以下步骤:搭建卷积-反卷积神经网络模型,所述卷积-反卷积神经网络模型包括多个不同的卷积层、多个卷积-反卷积层对和激活层;选取训练集,并设置卷积-反卷积神经网络模型的训练参数;根据卷积-反卷积神经网络模型及其训练参数,以最小化损失函数为目标训练卷积-反卷积神经网络模型形成图像深度估计神经网络模型;将待处理的图像输入到图像深度估计神经网络模型,输出对应的深度图。利用本发明的基于卷积-反卷积神经网络的图像深度估计方法获得的深度图灰度值较精确,深度图层次感更强。



1. 一种基于卷积神经网络的图像深度估计方法,其特征在于,所述方法包括以下步骤:

S1、搭建卷积-反卷积对神经网络模型,所述卷积-反卷积对神经网络模型包括多个不同的卷积层、多个卷积-反卷积层对和激活层;

所述搭建的卷积-反卷积对神经网络模型中前四层卷积层的卷积核大小分别为 7×7 、 7×7 、 5×5 和 5×5 ,从第15层卷积层到最后一层卷积层共有7对卷积-反卷积层对,由一个卷积核为 3×3 的卷积层和一个卷积核为 4×4 的反卷积层组成,其余第5层至第14层的卷积层的卷积核大小都是 3×3 ;

所述7对卷积-反卷积层对中,最后4个卷积-反卷积层对中的每个卷积层后均添加有一层输出层,与下一个卷积-反卷积层对中的反卷积层连结,所述4层输出层的输出通道分别是128、256、512和512;当输出层与下一个卷积-反卷积层对中的反卷积层连结时,先对输出层通过双线性采样进行上采样,把输出层的特征图的大小扩展至下一个卷积-反卷积层对中的反卷积层特征图的大小;

且卷积-反卷积对神经网络模型中的每个卷积层后面都添加有相应的ReLU激活层,用作深度预测的卷积层除外;

S2、选取训练集,并设置卷积-反卷积对神经网络模型的训练参数;

S3、根据卷积-反卷积对神经网络模型及其训练参数,以最小化损失函数为目标训练卷积-反卷积对神经网络模型形成图像深度估计神经网络模型;

S4、将待处理的图像输入到图像深度估计神经网络模型,输出对应的深度图。

2. 根据权利要求1所述的一种基于卷积神经网络的图像深度估计方法,其特征在于,步骤S3中的损失函数 $L(\theta)$ 为交叉熵损失函数:

$$L(\theta) = -\sum_{i=1}^m y^{(i)} \log(h_{\theta}(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_{\theta}(x^{(i)}))$$

其中, $x^{(i)}$ 和 $y^{(i)}$ 分别表示选取的训练集中的原始图像和对应的深度图像, $y^{(i)}$ 的值取0或1, θ 表示权重, m 表示图像的个数, $h_{\theta}(x^{(i)})$ 表示训练出的原始彩色图像到深度图像的映射。

3. 根据权利要求1所述的一种基于卷积神经网络的图像深度估计方法,其特征在于:步骤S3在训练卷积-反卷积对神经网络模型的过程中,所述卷积-反卷积对神经网络模型的权重初始值由高斯函数随机生成,最小化损失函数采用Adam优化方法。

4. 根据权利要求1所述的一种基于卷积神经网络的图像深度估计方法,其特征在于:步骤S3中所述图像深度估计神经网络模型是根据最小化损失函数获得的卷积层的权重来建立的。

5. 根据权利要求1所述的一种基于卷积神经网络的图像深度估计方法,其特征在于:步骤S2中所述的训练集包括NYU DEPTH数据集和KITTI数据集,先通过NYU DEPTH数据集做预训练,再用KITTI数据集做精细化调整。

一种基于卷积神经网络的图像深度估计方法

技术领域

[0001] 本发明涉及计算机视觉与图像处理领域,具体涉及一种基于卷积-反卷积对神经网络的图像深度估计方法。

背景技术

[0002] 深度估计方法用于估计待处理的图像中各像素点的深度信息,获取待处理的图像的全局深度图,在计算机视觉和计算机图形学应用领域发挥着重要的作用。当前深度估计方法从摄像头数量上可分为单目和双目。

[0003] 双目立体视觉使用两台摄像机从两个视点观测同一物体,获取在物体不同视角下的感知图像,通过三角测量的方法将匹配点的视差信息转换为深度。一般的双目视觉方法都是利用对极几何将问题变换到欧式几何条件下,然后再使用三角测量的方法估计深度信息。双目立体视觉法的优点是方法成熟,能够稳定地获得较好的深度估计结果,应用于三维重建也优于其他基于视觉的三维重建方法;不足的是运算量仍然偏大,而且在基线距离较大的情况下重建效果明显降低。

[0004] 单目图像的深度估计方法可以细分为四种思路:1. 聚焦法:如果一张图片拍摄时聚焦在近景那么远景就会模糊,因此可以通过判断图片中的清晰和模糊区域来估计近景还是远景,从而得到深度图。2. 形状匹配法:这种方法假设形状和深度有一定的对应关系,通过预先建立的形状和深度对应的关系库,来找出每个形状对应的深度从而得到深度图。3. 零平面法:利用场景里的物体和零平面的关系来估计深度图。4. 人工标记法:这种方法通过大量的人工交互来完成图像各区域的深度估计。上述各种思路下估算图像的深度信息,均有多种实现方法。

[0005] 然而,现有的多种估计深度信息的方法,通常仅依据图像中像素点的位置信息确定深度信息,依据自底向上的原则,将图像底部的物体视为近景,图像顶部的物体视为远景,确定得到图像的深度信息。这样的估计方法,通常估计的深度值不精确,深度图层次感不强。同时未考虑到图像中物体边缘的影响,得到的深度图边缘模糊。

发明内容

[0006] 本发明的目的是针对现有技术的不足,提供了一种基于卷积神经网络的图像深度估计方法,利用所述方法得到的深度灰度值更精确,深度图层次感也更强。

[0007] 本发明的目的可以通过如下技术方案实现:

[0008] 一种基于卷积神经网络的图像深度估计方法,所述方法包括以下步骤:

[0009] S1、搭建卷积-反卷积对神经网络模型,所述卷积-反卷积对神经网络模型包括多个不同的卷积层、多个卷积-反卷积层对和激活层;

[0010] S2、选取训练集,并设置卷积-反卷积对神经网络模型的训练参数;

[0011] S3、根据卷积-反卷积对神经网络模型及其训练参数,以最小化损失函数为目标训练卷积-反卷积对神经网络模型形成图像深度估计神经网络模型;

[0012] S4、将待处理的图像输入到图像深度估计神经网络模型，输出对应的深度图。

[0013] 进一步地，步骤S1的卷积-反卷积层对是由一个卷积核为 3×3 的卷积层和一个卷积核为 4×4 的反卷积层组成。

[0014] 进一步地，步骤S1搭建的卷积-反卷积对神经网络模型中前四层卷积层的卷积核大小分别为 7×7 、 7×7 、 5×5 和 5×5 ，从第15层卷积层到最后一层卷积层共有7对卷积-反卷积层对，由一个卷积核为 3×3 的卷积层和一个卷积核为 4×4 的反卷积层组成，其余第5层至第14层的卷积层的卷积核大小都是 3×3 。

[0015] 进一步地，步骤S1的卷积-反卷积对神经网络模型中的每个卷积层后面都添加有相应的ReLU激活层，用作深度预测的卷积层除外。

[0016] 进一步地，步骤S1的卷积-反卷积对神经网络模型包括7个卷积-反卷积层对，其中最后4个卷积-反卷积层对中的每个卷积层后均添加有一层输出层，与下一个卷积-反卷积层对中的反卷积层连结，所述4层输出层的输出通道分别是128、256、512和512。

[0017] 进一步地，当输出层与下一个卷积-反卷积层对中的反卷积层连结时，先对输出层通过双线性采样进行上采样，把输出层的特征图的大小扩展至下一个卷积-反卷积层对中的反卷积层特征图的大小。

[0018] 进一步地，步骤S3中的损失函数 $L(\theta)$ 为交叉熵损失函数：

$$[0019] \quad L(\theta) = -\sum_{i=1}^m y^{(i)} \log(h_{\theta}(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_{\theta}(x^{(i)}))$$

[0020] 其中， $x^{(i)}$ 和 $y^{(i)}$ 分别表示选取的训练集中的原始图像和对应的深度图像， $y^{(i)}$ 的值取0或1， θ 表示权重， m 表示图像的个数， $h_{\theta}(x^{(i)})$ 表示训练出的原始彩色图像到深度图像的映射。

[0021] 进一步地，步骤S3在训练卷积-反卷积对神经网络模型的过程中，所述卷积-反卷积对神经网络模型的权重初始值由高斯函数随机生成，最小化损失函数采用Adam优化方法。

[0022] 进一步地，步骤S3中所述图像深度估计神经网络模型是根据最小化损失函数获得的卷积层的权重来建立的。

[0023] 进一步地，步骤S2中所述的训练集包括NYU DEPTH数据集和KITTI数据集，先通过NYU DEPTH数据集做预训练，再用KITTI数据集做精细化调整。

[0024] 本发明与现有技术相比，具有如下优点和有益效果：

[0025] 本发明的图像深度估计方法基于深度网络的学习，通过引入卷积-反卷积层对，极大地增强了神经网络的学习能力，建立起原始图像到深度图像的准确映射；将图像深度估计过程分为模型训练过程和估计过程，能够显著提高图像深度估计的PNSR和视觉效果，减少了深度估计时间，应用在三维图像的建模和场景生成中的深度估计环节能有效地提升实验结果的效率和质量。另外，本发明搭建的卷积-反卷积对神经网络模型中的卷积-反卷积层对选用合适大小卷积核的卷积层和反卷积层，使得不需要引入池化层就能够使网络整体有足够的容量很好地估计图像深度，从而避免了因为引入池化层使得参数减少而导致的模型不精确、效果变差等问题。

附图说明

[0026] 图1为本发明实施例一种基于卷积神经网络的图像深度估计方法的流程图。

[0027] 图2为本发明实施例的卷积-反卷积对神经网络模型的内部构造示意图。

具体实施方式

[0028] 下面结合实施例及附图对本发明作进一步详细的描述,但本发明的实施方式不限于此。

[0029] 实施例:

[0030] 本实施例提供了一种基于卷积神经网络的图像深度估计方法,所述方法的神经网络通过引入卷积-反卷积层对、卷积层和激活层,借助卷积层的学习能力和激活层的筛选能力获取好的特征,极大地增强了神经网络的学习能力,准确地学习出从原始图像到深度图像的映射以建立输入到输出的映射,从而能够通过学习到的映射进行深度图像的预测和估计。流程图如图1所示,包括以下步骤:

[0031] S1、搭建卷积-反卷积对神经网络模型,所述卷积-反卷积对神经网络模型包括多个不同的卷积层、多个卷积-反卷积层对和激活层;

[0032] 本实施例中采用的是监督学习的方法,训练集的图像有RGBD四通道,即红绿蓝和深度四通道,即是含有深度值标签或任何对应的深度图。为了对图像深度值做出有效预测,采用一种互相关联的、多尺度预测的卷积-反卷积对神经网络模型设计结构,其内部构造示意图如图2所示,把合成深度图像的结果作为监督信号,把实地深度值作为参考标准计算损失函数LOSS。本实施例的卷积-反卷积对神经网络模型包括一层输入层,14层卷积层,7对卷积-反卷积层对,以及每个卷积层后面跟着的激活层(除了用作深度预测的卷积层);其中前四层卷积层的卷积核大小分别为 7×7 、 7×7 、 5×5 和 5×5 ,其余卷积层的卷积核都是 3×3 ,卷积-反卷积层对中的卷积层的卷积核也为 3×3 ,反卷积层的是 4×4 ,该大小的卷积核引入的参数不会太多故计算量不会太大,但是却能够提取到相对精确的特征;另外,本实施例中每个卷积层后的激活层选用修正线性单元ReLU函数:

$$[0033] \quad f(x) = \frac{1}{(\alpha * \text{sigmoid}(x) + \beta)}$$

[0034] 其中 $\alpha=10$, $\beta=0.1$,以此来限制其输出深度值处于一个合理的范围,使得训练后的网络更具稀疏性。相比其它sigmoid激活函数,ReLU函数在数据量大且复杂的网络结构的表现更高效,更快速。也正是因为ReLU这种非线性函数能够很好地传递深度值的误差梯度,越底层的卷积收敛地越快,高层的特征需要在充分训练后才能收敛,与深度网络的梯度弥散现象正好相反。

[0035] 最后4对卷积-反卷积层对用作多尺度估计图像深度,是一个从粗糙到精确的估计过程。每一组中的反卷积层对图像深度做出估计,并对估计深度图进行上采样然后再与反卷积层连结。上述的连结操作指:需要连结的两张特征图必须是维数相等的,即同宽同高,如分别是 $N_1 \cdot H \times W$ 和 $N_2 \cdot H \times W$,那么它们连结后的特征图为 $(N_1+N_2) \cdot H \times W$,这个操作有效增加了样本数,提高了神经网络模型的准确率。其中,反卷积层的作用尤为重要,它把与该反卷积层相连结的卷积层的特征图对应的矩阵进行转置后用作反卷积层的滤波器参数,相当于进行卷积的逆运算。它连接输入激活层的单个特征值和多个输出特征,是一个扩大了

的密级的激活特征图。然后裁剪该扩大后的特征图至上一层ReLu激活层的大小。反卷积层里经过学习后的卷积核与图像基础大小相对应,以此来重建输入图像对应深度图。

[0036] 基于反卷积层的功能作用,本实施例设计了多级的卷积层-反卷积层网络结构用来捕获不同级别的深度信息细节特征。底层的反卷积对捕捉输入图像中景象的粗糙的大概深度值,靠近顶层的反卷积对捕捉并编码特定位置的景物的深度细节,如边缘深度等。反卷积层善于放大激活层中与边缘深度相关的特征而有效压制其它区域深度信息的干扰,再在相连的卷积层输出深度估计图。因此通过该方法,输出深度图里的边缘特征会比其它卷积神经网络更清晰分明。因为连结的前提条件是特征图等高等宽,所以在连结深度估计图和下一卷积层对中的反卷积层时,先对深度估计图像进行上采样,把特征图的大小扩展至下一层反卷积层特征图的大小,此处的上采样采取双线性采样的方法,具体过程为:

[0037] 根据源图像的像素点 p_s 深度值附近4个像素点的深度值的线性加权估计出新插入的深度值,分别是左上,右上,右下,左下四个像素点,具体计算公式如下:

$$[0038] \quad \hat{I}_s(p_t) = I_s(p_s) = \sum_{i \in \{t,b\}, j \in \{l,r\}} w^{ij} \cdot I_s(p_s^{ij})$$

[0039] 其中, $I_s(p_s)$ 表示像素点 p_s 对应的图像深度值, i, j 表示像素点的横坐标、纵坐标, $i-5 \leq t < b \leq i+5, j-5 \leq l < r \leq j+5$,权重 w^{ij} 与 p_s 到 p_s^{ij} 的距离成正比,且 $\sum_{i,j} w^{ij} = 1$,从而重构出超出原有维度大小的深度图图像。然后把深度估计图作为新的输入图像先输入到下一次级反卷积层继续训练。最后经过4级卷积层-反卷积层的由粗到精的提取、采样、训练后,输出一幅精度最高的深度估计图。

[0040] 通过本实施例中建立的卷积神经网络模型中选取的卷积层的总层数和卷积核大小,在保证神经网络的能力的基础上,无需引入池化层便足以实现很好地深度值估计,从而避免了引入池化层导致参数减少而引起的深度结果不精确、效果变差等问题。

[0041] S2、选取训练集,并设置卷积-反卷积对神经网络模型的训练参数;

[0042] 本实施例中选取在图像识别领域常用的、高质量的NYU Depth数据集中的一部分和KITTI数据集作为训练集,每张图像分别有对应的景物图像和深度图像。在采用基于卷积神经网络估计图像深度的方法前,需要对该卷积神经网络进行预训练,然后设置该卷积神经网络模型的训练参数,包括每次进行训练的图像数量、输入图像的大小和输出图像的大小、学习速率等。学习速率设为0.002,每次训练时的衰减速率设为0.9;该网络模型的权重参数的初始值由高斯随机函数去生成,足够的随机性可以增强网络的鲁棒性;通过Adam优化方法更新每个卷积层的参数;输入的训练图片剪裁成128*436的大小,但其实整个方法对测试图片的大小不做要求;训练的最小批取4,即每次训练取4个样本进行训练。每训练2000次后进行一次测试,观察目前模型的效果以更改模型的相关参数。在本模型中经过约15000次迭代后收敛,学习速率降为0。具体涉及参数如表1所示:

[0043]

参数名	参数值
图片大小	128*436
批大小(batch_size)	4
最大迭代次数	50000
学习率	0.002
学习率递减迭代次数	1000

学习率迭代因子	0.9
---------	-----

[0044] 表1

[0045] S3、根据卷积-反卷积对神经网络模型及其训练参数,以最小化损失函数为目标训练卷积-反卷积对神经网络模型形成图像深度估计神经网络模型;

[0046] 其中损失函数 $L(\theta)$ 为交叉熵损失函数:

$$[0047] \quad L(\theta) = -\sum_{i=1}^m y^{(i)} \log(h_{\theta}(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_{\theta}(x^{(i)}))$$

[0048] 其中, $x^{(i)}$ 和 $y^{(i)}$ 分别表示选取的训练集中的原始图像和对应的深度图像, $y^{(i)}$ 的值取0或1, θ 表示权重, m 表示图像的个数, $h_{\theta}(x^{(i)})$ 表示训练出的原始彩色图像到深度图像的映射。

[0049] 卷积核的权重在迭代中不断更新,最后达到最优解。在本实施例中采用Adam优化方法优化全局参数,最小化损失函数。过程可归纳为:Adam每时间步长迭代一次,利用梯度的样本均值估计和样本平方均值估计动态调整每个参数的学习率,公式为:

$$[0050] \quad m_t = \mu * m_{t-1} + (1 - \mu) * g_t$$

$$[0051] \quad n_t = \nu * n_{t-1} + (1 - \nu) * g_t^2$$

$$[0052] \quad \hat{m}_t = \frac{m_t}{1 - \mu^t}$$

$$[0053] \quad \hat{n}_t = \frac{n_t}{1 - \nu^t}$$

$$[0054] \quad \Delta \theta_t = -\frac{\hat{m}_t}{\sqrt{\hat{n}_t + \varepsilon}} * \eta$$

[0055] 其中 m_t , n_t 分别是对梯度的一阶矩估计和二阶矩估计,可以看作对期望 $E|g_t|$ 、 $E|g_t^2|$ 的估计; \hat{m}_t , \hat{n}_t 是对 m_t , n_t 的校正,这样可以近似为对期望的无偏估计, θ_t 是学习率。

可以看出, $-\frac{\hat{m}_t}{\sqrt{\hat{n}_t + \varepsilon}}$ 对学习率形成一个动态约束,经过偏置校正后,每一次迭代学习率都有个确定范围,使得参数比较平稳。

[0056] 以最小化损失函数 $L(\theta)$ 为目标训练该网络模型,优化各个卷积层的权重从而建立有效的图像深度估计神经网络模型,该模型应用无监督的学习方法达到了普遍监督学习模型的效果,并能更好地保存物体边缘的深度值信息、更好地估计细长物体的深度值,如树和路灯等。

[0057] S4、将待处理的图像输入到图像深度估计神经网络模型,输出对应的深度图。

[0058] 本实施例的图像深度估计方法,通过引入卷积-反卷积层对,极大地增强了神经网络的学习能力,建立起NYU数据集图像、KITTI数据集图像到深度图的准确映射。卷积-反卷积层对中的卷积核分别选用 3×3 、 4×4 的效果很好,所述大小的卷积核引入的参数不会太多故计算量不会太大,但是却能够提取到相对精确的特征;除了卷积-反卷积层对的引入,本实施例还在每个卷积层后面增加了Adam激活函数的隐藏层,搭建起需要学习的深度估计卷积神经网络模型后,通过不断减小损失函数的数值来训练网络模型的参数,损失函数选

用交叉熵代价函数,减小交叉熵代价能够增大PSNR,从而提高图像的质量;用不同的数据集如KITTI、Make3D或Stereo等训练本深度估计卷积神经网络模型以构造普通彩色图像到深度图像的映射,最终都可以获得接近实地的深度图,证明了本发明具有良好的通用性。

[0059] 以上所述,仅为本发明专利较佳的实施例,但本发明专利的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本发明专利所公开的范围内,根据本发明专利的技术方案及其发明专利构思加以等同替换或改变,都属于本发明专利的保护范围。

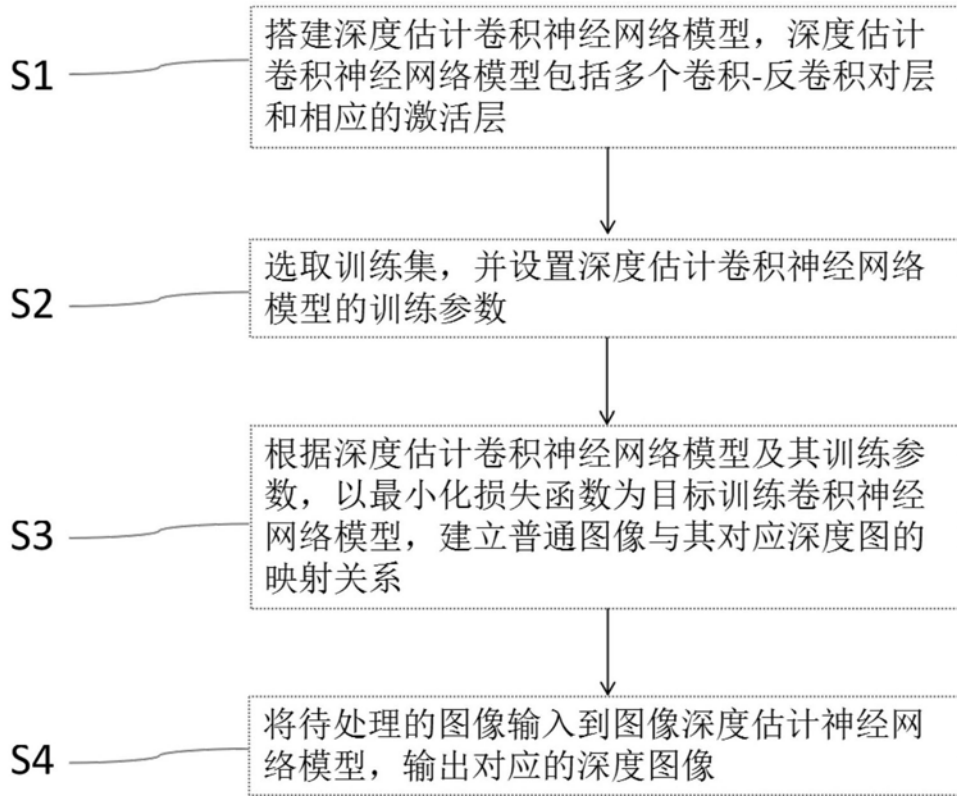


图1

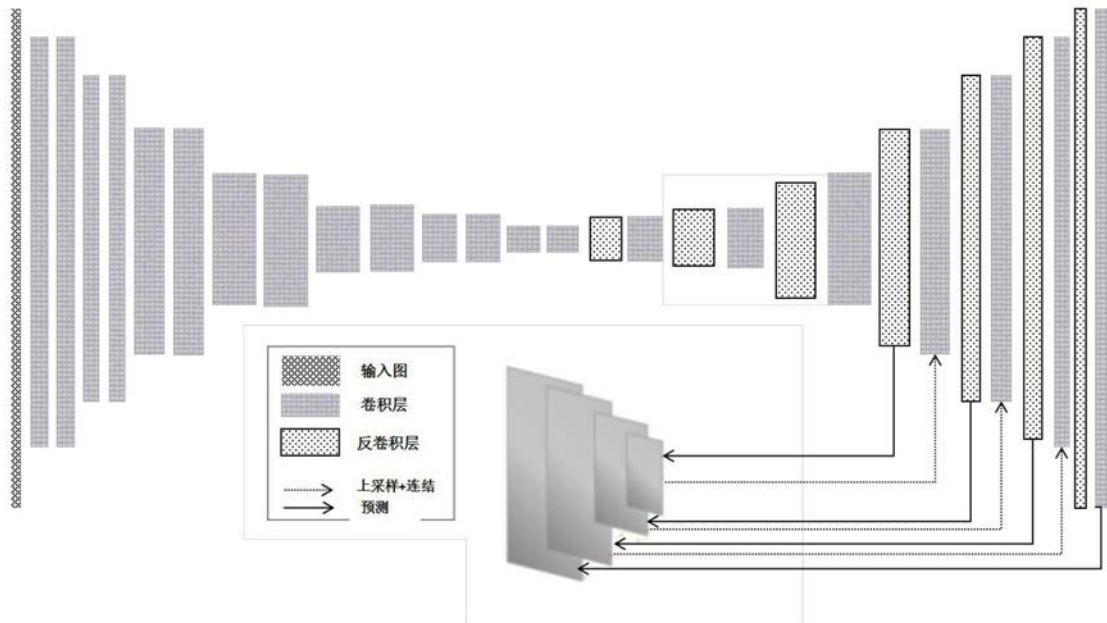


图2