Journal of Computer Applications
计算机应用,2020,40(8):2351－2357
ISSN 1001-9081
CODEN JYIIDU
2020-08-10
http://www.joca.cn

文章编号:1001-9081(2020)08-2351-07
DOI:10.11772/j.issn.1001-9081.2020010070

深度学习在单图像三维模型重建的应用

张　豪[1,2*],张　强[2],邵思羽[2],丁海斌[3]
(1.空军工程大学 研究生院,西安 710038;　2.空军工程大学 防空反导学院,西安 710038;
3.陆军工程大学 训练基地,江苏 徐州 221004)
(*通信作者电子邮箱421821467@qq.com)

摘　要:针对基于单图像重建的三维模型具有高度不确定性问题,提出了一种基于深度图像估计、球面投影映射、三维对抗生成网络相结合的网络模型算法。首先,通过深度估计器得到输入图像的深度图像,这有利于对图像进一步的分析;其次,将得到的深度图像通过球面投影映射转换为三维模型;最后,利用三维对抗生成网络对重建的三维模型的真实性进行判断,建立更逼真的三维模型。理论分析和仿真实验表明,与学习先验知识生成三维模型的算法LVP相比,所提模型在真实三维模型与重建三维模型的交并比(IoU)上提高了20.1%,倒角距离(CD)缩小了13.2%。实验结果表明,所提模型在单视图三维模型重建中具有良好的泛化能力。

关键词:深度图像;深度估计;三维重建;对抗生成网络;球面投影
中图分类号:TP391　　**文献标志码:**A

Application of deep learning to 3D model reconstruction of single image

ZHANG Hao[1,2*], ZHANG Qiang[2], SHAO Siyu[2], DING Haibin[3]
(1. *Graduate School, Air Force Engineering University, Xi'an Shannxi* 710038, *China*;
2. *Air and Missile Defense College, Air Force Engineering University, Xi'an Shannxi* 710038, *China*;
3. *Training Base, Army Engineering University of PLA, Xuzhou Jiangsu* 221004, *China*)

Abstract: To solve the problem that the reconstructed 3D model of a single image has high uncertainty, a network model based on depth image estimation, spherical projection mapping and 3D generative adversarial network was proposed. Firstly, the depth image of the input image was obtained by the depth estimator, which was helpful for the further analysis of the image. Secondly, the obtained depth image was converted into a 3D model by spherical projection mapping. Finally, 3D generative adversarial network was utilized to judge the authenticity of the reconstructed 3D model, so as to obtain 3D model closer to reality. In the comparison experiments with LVP algorithm which learning view priors for 3D reconstruction, the proposed model has the Intersection-over-Union (IoU) increased by 20.1% and the Charmfer Distance (CD) decreased by 13.2%. Theoretical analysis and simulation results show that the proposed model has good generalization ability in the 3D model reconstruction of a single image.

Key words: depth image; depth estimation; 3D reconstruction; Generative Adversarial Network (GAN); spherical projection

0　引言

近年来,计算机视觉和机器学习领域的研究者们在单图像的三维重建方面取得了令人瞩目的进展。按照模型重建方法的不同,可以分为传统重建方法和基于深度学习的重建方法两类。传统的三维重建方法主要基于第三方模型库,这些第三方库内部封装了大量的三维图形处理函数,能够轻易绘制三维图形。但这种方法的先验知识在模型的设计阶段就已经被设定好,很难扩展到其他类别物体上,生成的模型只能沿着特定类别的变化而变化。随着深度学习在三维领域的发展,单幅图像三维重建的方法开始逐渐增多。周继来等[1]提出了一种基于曲度特征的三维模型算法,解决了复杂曲面三维模型的问题。朱俊鹏等[2]运用卷积神经网络处理深度图像,重建出了三维视差图。3D ShapeNets[3]是较早提出的一种基于深度学习的三维重建网络,通过输入深度图像[4-5],利用深度卷积网络提取信息,将三维几何模型表示为三维体素上的二值分布,预测外形类型并填补未知的体素来完成三维重建,取得了一定的效果。

目前基于深度学习的三维重建方法大多都是将预测的形状与真实的三维模型进行比较,通过损失函数最小化使得预测的形状越来越精确。由于模型的多维性,网络在训练的过程中很难学习到有用的细节[6]。以前的工作大多使用了不同的损失函数,在预测形状的模块上添加了先验知识,或是使用额外的监督训练使预测的三维模型接近真实形状。例如最近

收稿日期:2020-01-22;**修回日期:**2020-03-31;**录用日期:**2020-03-31。
基金项目:江苏省普通高校学术学位研究生科研创新计划项目(KYCX18_0072)。
作者简介:张豪(1995—),男,福建福州人,硕士研究生,主要研究方向:深度学习、模式识别;　张强(1973—),男,陕西汉中人,副教授,博士,主要研究方向:深度学习、电力系统及其自动化;　邵思羽(1991—),女,山东邹城人,讲师,博士,主要研究方向:基于深度学习、迁移学习的机电设备健康状态检测与故障诊断;　丁海斌(1985—),男,山东荣成人,讲师,硕士,主要研究方向:防空反导模拟训练。

的两种方法：DRC（Differentiable Ray Consistency）算法[7]和AtalsNet算法[8]，使用了不同的表示、损失函数以及训练策略来解决网络学习的问题。图1是两种模型在椅子类别的训练集上进行训练和测试的结果，可以看出当输入的图像为椅子类别时，两个模型建立的三维模型效果都比较好。但是，如果输入的图像不是来自椅子训练集，这两种方法都无法给出合理的预测，而是给出与训练的数据集相似的输出。Henderson等[9]使用LSI模型将网格作为输出表示，运用光照参数和阴影信息从图像中重构三维模型，效果优于以往大多数工作。Kato等[10]通过LVP算法训练鉴别器来学习视图的先验知识，成功解决了生成三维形状模糊性的问题。但从某种程度上看，现有的三维重建技术只是基于训练的三维数据集进行筛选，挑选出数据集中最接近输入图像的三维模型，泛化能力弱。

本文针对三维模型重建问题提出了一种基于深度图像估计、球面投影映射、三维对抗生成网络相结合的算法模型。该算法模型的主要特点有：1）采用模块化设计，强制网络的每个模块使用前一个模块的特性，而不是直接记住训练数据集中的形状；每个模块预测的结果与输入的图像具有相同的尺寸，这样可以得到更规则的映射。2）现有的三维重建算法在输入图像分辨率低时很难重建出精细的三维模型，本文模型通过引入超分辨率模块能很好地解决这个问题。

仿真实验结果表明，与目前主流的三维重建技术相比，本文算法模型在训练类别外的单图像三维形状重建中取得了理想的效果，解决了当前模型存在的泛化能力弱的问题。

(a) DRC模型 (b) AlatsNet模型

图1 不同输入下DRC与AtlasNet模型测试结果

Fig. 1 DRC and Atlasnet model test results with different inputs

1 网络模型结构

单图像三维重建算法通过学习某种函数，将二维图像映射到三维形状。本文针对目前三维重建技术泛化能力差的问题进行研究，以MarrNet[11]网络模型作为基础，提出了一个改进算法模型。如图2，首先通过一个单视图深度估计器从输入的二维图像中预测其深度信息；然后深度图像投影到球形网格中，并对球形图进行精修；最后引入一个体素重建网络估计三维形状。该算法的神经网络模块只需要通过球形图来重建对象的几何模型，而不必学习几何投影，增强了模型的通用性；同时通过在判别模块引入正则化参数来解决模型泛化能力差的问题。下面介绍每个模块的具体实现细节。

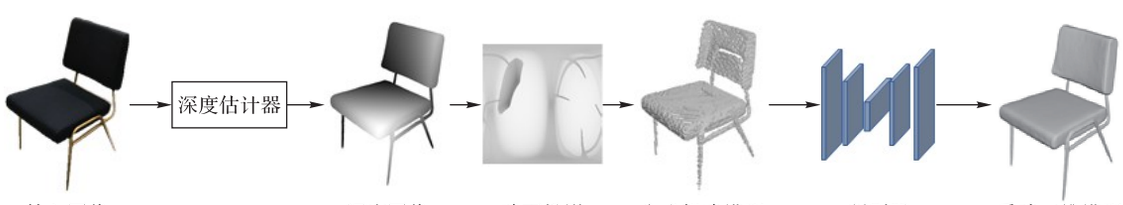


图2 本文模型结构示意图

Fig. 2 Structure diagram of the proposed model

1.1 深度图像估计器

深度图像的来源可以追溯到计算机视觉的早期，类似于灰度图像，它的每个像素值是传感器距离物体的实际距离。多年来，研究人员一直在探索如何从纹理、阴影或彩色图像中恢复2.5D草图。近年来，随着深度学习的发展，很多人开始用神经网络来估计图像的深度、表面法线。Wu等[11]提出了一种名为MarrNet的单图像三维重构模型，将深度图像估计器作为重建网络中的一个模块。但MarrNet的深度估计器的缺点在于当输入的图像清晰度低时，提取的深度图像效果不佳，进而会影响图像三维重建的效果。本文采用的深度图像估计器与现有的深度图像估计方法相比，能解决输入图像分辨率低导致深度信息丢失以及训练困难的问题。

深度估计器模块由超分辨率重构网络及编码器-解码器（AutoEncoder，AE）结构组成。由于拍摄的二维图像可能由于对焦、摇晃等外界因素导致分辨率不够清晰，从而对深度图像的提取造成影响，导致建立的三维模型效果欠佳[12]。在整个网络结构的最前端加入一个超分辨率重构的网络模块（Enhanced Super-Resolution Generative Adversarial Network，ESRGAN）[13]有助于提高图像清晰度，使深度图像提取效果更佳，建立的三维模型更真实。AE的目标是找到随机映射，使输入输出差异最小。通过编码解码的过程使模型学习到数据的分布和特征，能够有效捕捉图像的有效特征，提取出效果较好的深度图像。

1.1.1 超分辨率重构模块

增强型超分辨率生成对抗网络ESRGAN是本文估计器的核心结构，这是一种单图像超分辨率改进算法结构。如图3所示，此模型由卷积层、采样层和4个残差块构成。

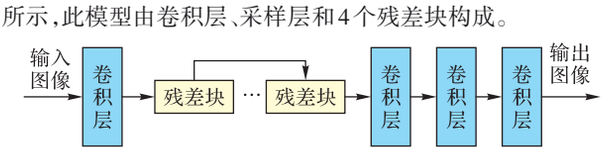


图3 ESRGAN结构示意图

Fig. 3 Structure diagram of ESRGAN

与其他超分辨率模块不同的是，ESRGAN在残差块中移除了归一化（Batch Normalization，BN）层，图4是传统残差块与ESRGAN中残差块对比。BN层通过在训练中使用一批数据的均值和方差规范化特征并且在测试时通过使用在整个训练

集上预估后的均值和方差规范化测试数据。当训练集和测试集的统计结果相差甚远时,BN层常常趋向于引入一些伪影并且限制了模型的泛化能力。通过移除BN层能够有效地提高性能并且减少模型的计算复杂度。低分辨率的图像经过卷积、采样等操作能得到分辨率较高的图像。

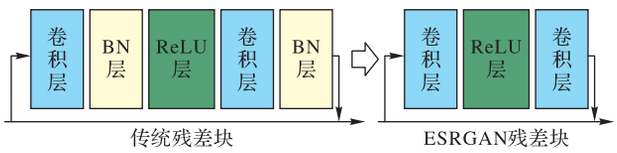


图4　传统残差块和ESRGAN残差块的对比

Fig. 4　Comparison of traditional residual block and ESRGAN residual block

1.1.2　编码器–解码器

在最近的深度图像提取器研究中,编码器主要采用18层的深度残差网络(Residual neural Network, ResNet)。这是微软实验室何凯明团队在2015年提出的一种深度卷积网络。ResNet相较普通网络,在每两层间增加了短路机制,形成残差学习,ResNet的提出解决了网络深度变深以后性能退化的问题,为一些复杂的特征提取和分类提供了可行性[14]。

但是在实际的训练中,ResNet存在着训练较为困难、图像分割细节不够好的问题,本文在ResNet的基础上引入U型网络结构(nested U-Net architecture, UNet++),在网络中间添加更多的跳转连接,以更好地结合图像的信息进行分割[15]。

式(1)是UNet++网络结构的核心思想,其中:$H()$表示一个卷积和一个激活函数;$u()$表示一个上采样层;[]表示连接层。例如$x^{1,2}$就是由$x^{1,0}$、$x^{1,1}$和上采样后的$x^{2,1}$拼接之后,再经过一次卷积和线性激活函数层(Rectified Linear Unit, ReLU)得到的。

$$x^{i,j}=\begin{cases}H(x^{i-1,j}), & j=0\\ H([x^{i,k}]_{k=0}^{j-1},u(x^{i+1,j-1})), & j>0\end{cases} \tag{1}$$

解码器包含四组5×5全卷积层和激活函数ReLU层,然后是四组1×1卷积层和ReLU层。解码器输出相应的深度图像信息。

1.2　球面投影模块

球面投影这一概念近几年被证明在三维形状的重构上起到了不错的效果。文献[16-17]中研究了球面投影上可微的球面卷积,目的是使模块中神经网络输出的投影能够保持旋转不变性。这项技术需要在有限频带的频谱域中进行卷积,这样就会导致高频信息的混叠和丢失,而形状重建的质量很大程度上依赖于高频分量[18]。目前,将球面投影这项技术应用于单图像三维重构领域还处于初步阶段,还有很大的发展空间。

本文将深度图像作为输入,如图5使用摄像机参数将深度信息转换成点云的形式,然后用立方体算法将它们转换成表面信息。立方体算法是体素重建中的经典算法,该算法的基本思想是逐个处理三维模型数据中的立方体,找出与等值面相交的立方体,采用线性插值计算出等值面与立方体边的交点。根据立方体每一顶点与等值面的相对位置,将等值面与立方体边上的交点按一定方式连接生成等值面,作为等值面在该立方体内的一个逼近表示。

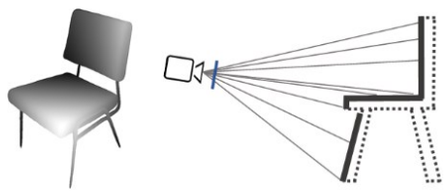


图5　球面投影示意图

Fig. 5　Schematic diagram of spherical projection

假设$P(x,y,z)$是立方体中的任意一点,根据线性插值运算,可得该点处的函数值,如式(2):

$$f(x,y,z)=a_0+a_1x+a_2y+a_3z+a_4xy+a_5yz+a_6zx+a_7xyz \tag{2}$$

其中:系数$a_i(i=0,1,\cdots,7)$为立方体中8个定点的函数值,若等值面阈值为c,则联立方程组为式(3),就可以计算出等值面与立方体边界上的交线。

$$\begin{cases}f(x,y,z)=c\\ f(x,y,z)=a_0+a_1x+a_2y+a_3z+a_4xy+a_5yz+a_6zx+a_7xyz\end{cases} \tag{3}$$

将得到的表面信息通过单位球面的每个U轴、V轴投射到球体中心来生成球面表示,整个过程是不可微的。图6展示了整个球面投影模块的结构,深度估计器模块输出的深度图像作为输入,通过球面投影转换为三维模型。

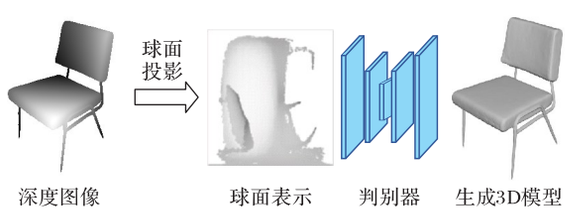


图6　球面投影模块结构

Fig. 6　Structure of spherical projection moduale

1.3　判别模块

由于球面投影的局限性,当图像中的物体发生自遮挡时,曲面信息会丢失,通过深度图像投影产生的三维模型在细节上会存在严重的丢失。引入三维对抗生成网络中的判别模块能较好地解决这个问题。

判别器由5个带有ReLU激活函数的卷积层构成,将球面投影得到的三维形状与实际形状区分开来。为了提升模型的泛化能力,在损失函数中加入了正则化惩罚项。这个模块借鉴了三维对抗生成网络(3D-Generative Adversarial Network, 3D-GAN)的思想,使判别器能够对合成的三维模型的真实性进行判断,建立更逼真的三维模型[19]。

式(4)是判别模块的损失函数,其中:P_g、P_x分别代表不同图像数据集生成的三维形状,P_r则为对应的真实三维形状,D表示判别器,λ是惩罚项。判别模块通过训练使损失函数值减小,恢复重建的三维模型更多细节[20-21]。

$$L=\underset{\tilde{x}\sim P_g}{E}[D(\tilde{x})]-\underset{x\sim P_r}{E}[D(x)]+\lambda\underset{\hat{x}\sim P_x}{E}\left[\left(\left\|\nabla_{\hat{x}}D(\hat{x})\right\|_2-1\right)^2\right] \tag{4}$$

2 实验与结果分析

2.1 评价指标

2.1.1 深度估计器

在深度估计器模块，本文将 MarrNet 网络中的深度提取器作为基准，分别对不同二维图像提取出的深度图像进行对比分析。

在评价指标方面，使用真实图像和输出图像二维交并比（Intersection-over-Union ，IoU）与均方误差（Mean Squared Error，MSE）作为衡量指标。计算方法如式(5)、(6)所示：

$$IoU = \frac{A \cap B}{A \cup B} \tag{5}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y^i - RUL^i)^2 \tag{6}$$

其中：A 表示预测图像，B 表示真实图像。IoU 值越高则预测效果越好。

2.1.2 三维重建指标

在评价三维重建效果方面，本文以目前单视图重建效果最优的两种网络 DRC 和 AtalsNet 作为基准，采用 IoU 值和倒角距离（Chamfer Distance，CD）值作为衡量三维重建效果的指标。式(7)是三维体素重建模型与实际三维模型之间的体素相交部分，其中：i、j、k 表示体素位置，$I()$ 表示一个指标函数，t 为体素化阈值。

$$IoU = \frac{\sum_{i,j,k}[I(p_{(i,j,k)} > t)I(y_{i,j,k})]}{I(I(p_{(i,j,k)} > t) + I(y_{(i,j,k)}))} \tag{7}$$

由于深度图像和球面图都无法提供形状内部信息，只通过 IoU 来评价重建效果是片面的，因此引入倒角距离作为另一种评价指标。式(8)为倒角距离的计算公式，其中：S_1、S_2 分别表示从预测和真实三维形状的表面采样点的集合，通过计算预测点到真实点的平均距离衡量三维模型的重建效果。

$$CD(S_1,S_2) = \frac{1}{|S_1|}\sum_{x\in S_1}\min_{y\in S_2}\|x-y\|_2 + \frac{1}{|S_2|}\sum_{y\in S_2}\min_{x\in S_1}\|x-y\|_2 \tag{8}$$

2.2 训练数据集

本文使用 Shapenet 数据集进行网络培训和测试，此数据集包含了 55 种常见的物品类别以及对应的 51 300 个三维模型。在此数据集的基础上，本文还增加了部分类别的真实深度图像用于深度估计器的训练。对于所有的模型，本文挑选了汽车、椅子、飞机这三种类别进行训练。将训练完成的模型在其他类别上进行测试，用来评估模型的泛化能力。

2.3 结果分析

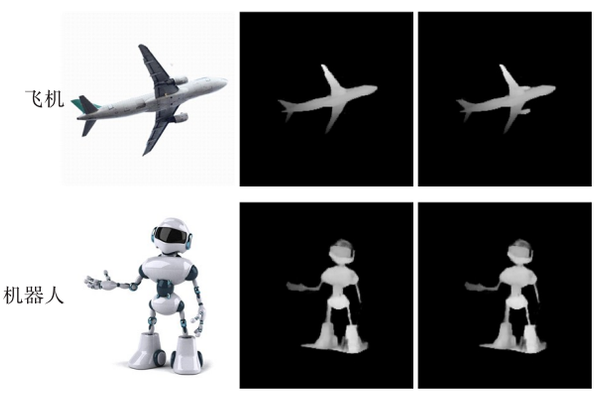
2.3.1 深度估计器

本文分别将分辨率为 72 和分辨率为 300 的图像作为输入，对椅子等四个类别的图像来测试本文深度估计器提取效果。从表 1 可以看出，当输入图像为低分辨率时，本文模型 IoU 指标较 Marrnet 模型提高了 0.189。图 7 是低分辨率图像提取的深度图像对比，可以明显看出本文的模型提取的深度图像细节更多。如图 8 所示，当输入图像为高分辨率时，本文模型提取的深度图像效果也略优于 Marrnet。从结果上看，本文采用的深度估计器模块在提取效果上优于目前主流的模型。图 9 是各类别预测结果的 MSE 值，从中可以看出在汽车、飞机等数据集上训练深度估计器，本文模型训练结果可以推广到预测新的类别深度图像。训练集与新测试类别的 MSE 值区别并不是特别明显，足以证明本文深度估计器的优秀泛化能力。

表 1 不同分辨率图像的 IoU 值对比

Tab. 1 Comparison of IoU values of images with different resolutions

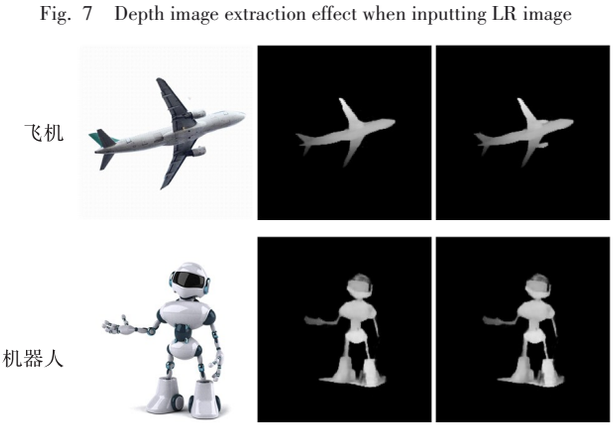
分辨率	模型	IoU 值				
		椅子	汽车	飞机	沙发	平均
低分辨率	Marrnet	0.678	0.657	0.585	0.623	0.636
	本文模型	0.823	0.819	0.838	0.821	**0.825**
高分辨率	Marrnet	0.785	0.701	0.705	0.714	0.726
	本文模型	0.878	0.923	0.855	0.749	**0.851**



(a) 输入图像 (b) MarrNet (c) 本文深度估计器

图 7 输入为低分辨率图像的深度图像提取效果

Fig. 7 Depth image extraction effect when inputting LR image



(a) 输入图像 (b) MarrNet (c) 本文深度估计器

图 8 输入为高分辨率图像的深度图像提取效果

Fig. 8 Depth image extraction effect when inputting HR image

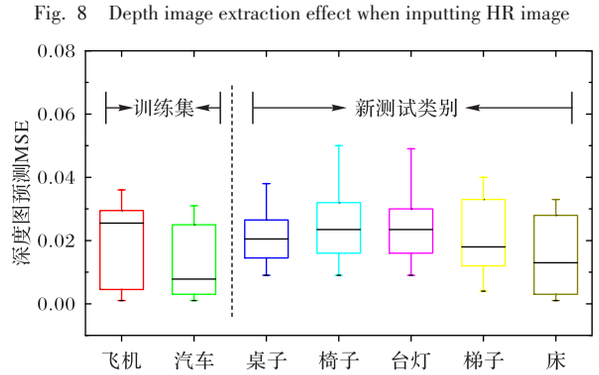


图 9 各类别预测结果的 MSE 值

Fig. 9 MSE values of prediction results of different classes

2.3.2　三维模型重建

为了验证本文提出的模型具有较好的泛化能力,整个训练过程中只用到了飞机、汽车、椅子三个数据集进行训练。在最后的测试中,分为了两个部分:一是对训练数据集进行测试,二是对训练数据集以外的图像进行测试,以验证模型的泛化能力。

表2是在训练集类别模型测试的倒角距离。从表中可以看出,在训练集(椅子、汽车、飞机)上测试,AtlasNet模型在倒角距离的表现略优于本文模型,但从图10的模型重构的实际结果上看,AtlasNet模型并没有预测出符合实际的结果形状。经过分析,在训练集的CD值表现弱于AtlasNet是由于此模型是从训练库中挑选与输入图像最为接近的三维形状作为预测结果,而CD值计算的是预测点和真实点的平均距离,因此AtlasNet表现较好。但从结果上看,AtlasNet部分预测结果偏离实际,没有参考价值。

在训练集之外,将本文模型与目前主流模型进行了实验对比,从表3的结果可以看出在训练集外,本文模型所有类别的倒角距离表现都优于对比模型。预测的部分三维形状如图11所示。从各模型输出的三维形状结果可以看出,现有模型在训练集外三维重建上效果不佳,重建的形状与真实形状相差甚远,泛化能力弱。而本文模型重建的效果相较于现有模型有明显提升,尽管在细节方面还需要改进,但本文模型生成的三维形状最符合客观事实,验证了模型的高泛化能力。

表2　训练集类别各模型倒角距离测试结果

Tab. 2　Test results of chamfer distance of training set classes obtained by different models

模型	训练集类别			平均
	椅子	汽车	飞机	
DRC	0.089	0.113	0.140	0.112
AtlasNet	0.083	**0.106**	**0.101**	**0.097**
本文模型	**0.081**	0.114	0.109	0.101

表3　训练集外类别各模型倒角距离测试结果

Tab. 3　Test results of chamfer distance of classes outside training set obtained by different models

模型	训练集外部类别							平均
	沙发	书架	浴缸	桌子	梯子	台灯	床	
DRC	0.093	0.151	0.118	0.134	0.152	0.192	0.149	0.141
AtlasNet	0.088	0.140	0.112	0.126	0.145	0.188	0.142	0.134
LSI	0.084	0.124	0.103	**0.104**	0.132	0.147	0.129	0.118
LVP	0.087	0.116	0.109	0.114	0.144	0.158	0.114	0.121
本文模型	**0.079**	**0.109**	**0.082**	0.105	**0.118**	**0.135**	**0.108**	**0.105**

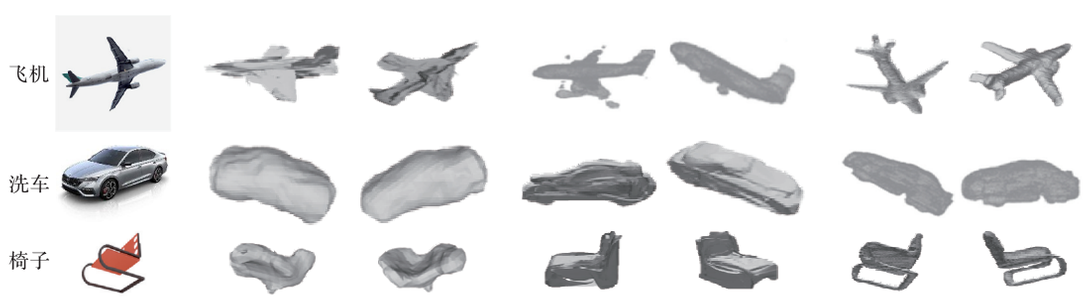


图10　输入为训练集图像的三维模型预测结果

Fig. 10　3D model prediction results with inputting training set images

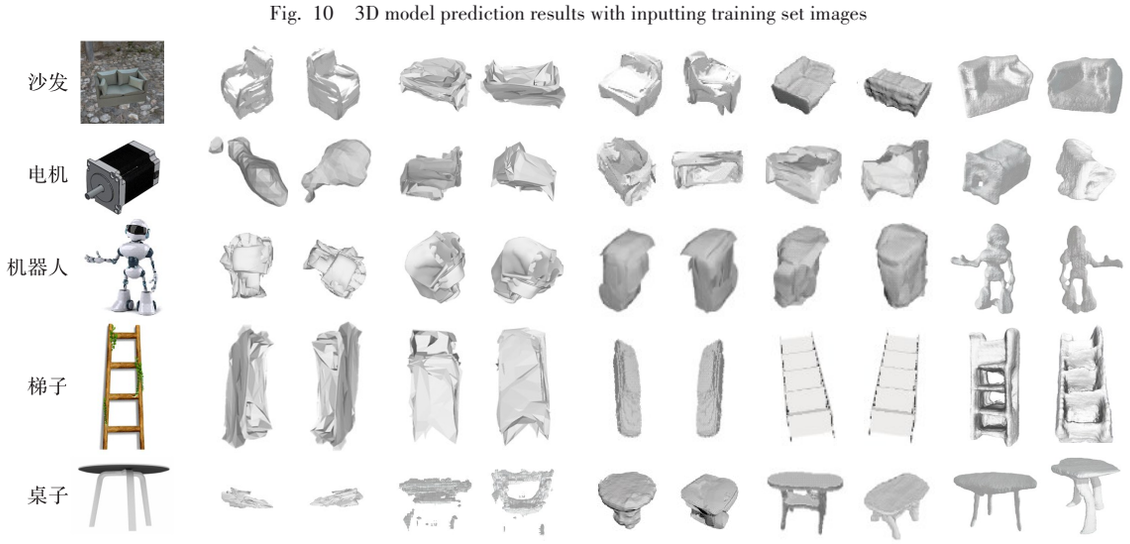


图11　输入为训练集外图像的三维模型预测结果

Fig. 11　3D model prediction results with inputting images outside training set

图12是现有主流模型与本文模型在六种常见物体的三维重建IoU评价指标对比。从图中可以看出,不论是在训练集还是在非训练集上测试,本文模型预测结果和真实三维模型的IoU指标都优于对比的模型。但本文模型在部分物体的

IoU 指标上提升并不明显,这是由于 IoU 指标的局限性导致的,单一的 IoU 指标并不能直观地表现三维重建的客观真实性,所以本文在评价指标上还引入了 CD 值作为参考。从图 11 的重建效果也可以看出本文模型的优势。总的来说,本文模型三维重建的效果更优,更加接近真实三维形状,尤其是在非训练集上的测试,更是证明了本文模型具有较强的泛化能力。

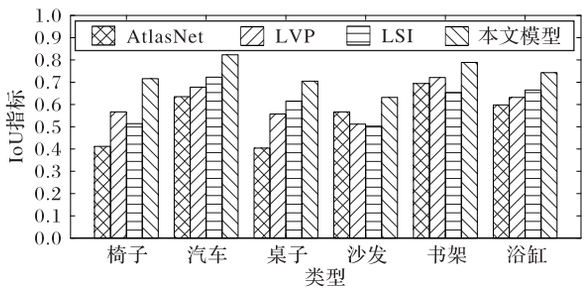


图 12 现有模型与本文模型重建 IoU 值对比

Fig. 12 Reconstruction IoU comparison of existing models and the proposed model

3 结语

针对目前单视图三维重建泛化能力差,以及当输入图像分辨率低时,重建的三维模型细节丢失的问题,本文提出了一个基于深度学习的三维重构框架,通过引入超分辨率模块,AutoEncoder 网络结构很好地对深度图像进行了提取。为了保证重构三维模型的细节和泛化能力,模型加入球面投影以及判别模块,使得训练结果不会太过依赖数据集,解决了过拟合的问题。从实验结果可以看出,本文模型在重构细节和泛化能力上优于目前主流的三维模型重构方法;并且当输入图像清晰度低时,本文模型依旧能通过提取较好深度图像,建立符合客观现实的三维重建模型。但由于三维模型数据库还不够完善,且三维模型的维度较高、训练难度较大,本文模型只在简单的物体类别上进行了训练和验证,如何通过单图像建立复杂的三维模型是目前需要解决的一个问题,也是我们今后进一步研究的方向。

参考文献(References)

[1] 周继来,周明全,耿国华,等. 基于曲度特征的三维模型检索算法[J]. 计算机应用, 2016, 36(7):1914-1917, 1922. (ZHOU J L, ZHOU M Q, GENG G H, et al. 3D model retrieval algorithm based on curvedness feature[J]. Journal of Computer Applications, 2016, 36(7):1914-1917, 1922.)

[2] 朱俊鹏,赵洪利,杨海涛. 基于卷积神经网络的视差图生成技术[J]. 计算机应用, 2018, 38(1):255-259, 289. (ZHU J P, ZHAO H L, YANG H T. Disparity map generation technology based on convolutional neural network [J]. Journal of Computer Applications, 2018, 38(1):255-259, 289.)

[3] CHANG A X, FUNKHOUSER T, GUIBAS L, et al. ShapeNet: an information-rich 3D model repository [EB/OL]. [2019-11-21]. https://arxiv.org/pdf/1512.03012.pdf.

[4] DENG X, SONG P, RODRIGUES M R D. RADAR: robust algorithm for depth image super resolution based on FRI theory and multimodal dictionary learning[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019(Early Access):1-1.

[5] DREWS P L J, NASCIMENTO E R, BOTELHO S S C, et al. Underwater depth estimation and image restoration based on single images[J]. IEEE Computer Graphics and Applications, 2016, 36(2):24-35.

[6] 陈加,张玉麒,宋鹏,等. 深度学习在基于单幅图像的物体三维重建中的应用[J]. 自动化学报, 2019, 45(4): 657-668. (CHEN J, ZHANG Y Q, SONG P, et al. Application of deep learning in 3D object reconstruction based on single image [J]. Acta Automatica Sinica, 2019, 45(4): 657-668.)

[7] TULSIANI S, ZHOU T, EFROS A, et al. Multi-view supervision for single-view reconstruction via differentiable ray consistency[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2019(Early Access):1-1.

[8] GROUEIX T, FISHER M, KIM V G, et al. AltasNet: a Papier-Mâché approach to learning 3D surface generation[C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 216-224.

[9] HENDERSON P, FERRARI V. Learning single-image 3D reconstruction by generative modelling of shape, pose and shading [J]. International Journal of Computer Vision, 2020, 128(4): 835-854.

[10] KATO H, HARADA T. Learning view priors for single-view 3D reconstruction [C]// Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 9770-9779.

[11] WU J, WANG Y, XUE T, et al. MarrNet: 3D shape reconstruction via 2.5D sketches [C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc., 2017:540-550.

[12] 李伟,张旭东. 基于卷积神经网络的深度图像超分辨率重建方法[J]. 电子测量与仪器学报, 2017, 31(12):1918-1928. (LI W, ZHANG X D. Depth image super-resolution reconstruction based on convolutional neural network [J]. Journal of Electronic Measurement and Instrumentation, 2017, 31(12): 1918-1928)

[13] CHEN Y, SHI F, CHRISTODOULOU A G, et al. Efficient and accurate MRI super-resolution using a generative adversarial network and 3D multi-level densely connected network [C]// Proceedings of the 2018 International Conference on Medical Image Computing and Computer-Assisted Intervention, LNCS 11070. Cham: Springer, 2018: 91-99.

[14] ZHANG K, SUN M, HAN T X, et al. Residual networks of residual networks: multilevel residual networks [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 28(6): 1303-1314.

[15] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation [C]// Proceedings of the 2015 International Conference on Medical Image Computing and Computer-Assisted Intervention, LNCS 9351. Cham: Springer, 2015:234-241.

[16] XIA Y, XIAO J, WANG Y. A fast registration algorithm of rock point cloud based on spherical projection and feature extraction [J]. Frontiers of Computer Science, 2019, 13(1):170-182.

[17] RAN L, ZHANG Y, ZHANG Q, et al. Convolutional neural network-based robot navigation using uncalibrated spherical images [J]. Sensors, 2017, 17(6): No. 1341.

[18] GORBACHEV V A, OSOKIN I V. Detection and removal of foreground objects in spherical images for the synthesis of photorealistic intermediate images [J]. Pattern Recognition and Image Analysis, 2019, 29(3):471-485.

[19] BASHMAL L, BAZI Y, ALHICHRI H, et al. Siamese-GAN: learning invariant representations for aerial vehicle image categorization[J]. Remote Sensing, 2018, 10(3): No. 351.

[20] CUI Z, ZHANG M, CAO Z, et al. Image data augmentation for SAR sensor via generative adversarial nets [J]. IEEE Access, 2019, 7: 42255-42268.

[21] 曹仰杰,贾丽丽,陈永霞,等. 生成式对抗网络及其计算机视觉应用研究综述[J]. 中国图象图形学报, 2018, 23(10): 1433-1449. (CAO Y J, JIA L L, CHEN Y X, et al. Review of computer vision based on generative adversarial networks [J]. Journal of Image and Graphics, 2018, 23(10): 1433-1449.)

This work is partially supported by the Scientific Research Innovation Plan for Graduate Students of Academic Degree in Colleges and Universities in Jiangsu Province (KYCX18_0072).

ZHANG Hao, born in 1995, M. S. candidate. His research interests include deep learning, pattern recognition.

ZHANG Qiang, born in 1973, Ph. D., associate professor. His research interests include deep learning, power system and its automation.

SHAO Siyu, born in 1991, Ph. D., lecturer. Her research interests include health status detection and fault diagnosis of electromechanical equipment based on deep learning and transfer learning.

DING Haibin, born in 1985, M.S., lecturer. His research interests include air and missile defense simulation training.