

硕 士 研 究 生 读 书 报 告



题目 High-Quality Textured 3D Shape Reconstruction

with Cascaded Fully Convolutional Networks

作者姓名 邹其锋

作者学号 22051068

指导教师 李启雷

学科专业 电子信息-软件工程

所在学院 软件学院

提交日期 二○二○ 年 十二 月

Monster Mash: High-Quality Textured 3D Shape Reconstruction

with Cascaded Fully Convolutional Networks

Paper accepted in IEEE

Transaction on

Visualization and Computer Graphics, Vol. 27, No. 1.

January 2021.

A Dissertation Submitted to

Zhejiang University

in partial fulfillment of the requirements for

the degree of

Master of Engineering

Major Subject: Software Engineering

Advisor: Qilei Li

By

Zou Qifeng

Zhejiang University, P.R. China

2020．12

摘要

我们提出了一种基于学习的方法来重建具有详细几何形状和高保真纹理的高分辨率三维形状。尽管进行了广泛的研究，但从多视图深度和颜色进行3D重建的算法（RGB-D)扫描仍然容易产生测量噪声和阻塞；有限的扫描和捕获角度通常还会导致重建不完整。在3D深度学习技术的最新进展的推动下，本文介绍了一种新颖的计算和内存高效级联3D卷积网络体系结构，该体系结构学习了嘈杂和不完美的RGB-D图。所提出的3D神经网络以渐进和从粗到细的方式进行重构，从而实现了前所未有的输出分辨率和保真度。同时，开发了一种对所提出的级联结构进行端到端训练的算法。我们进一步介绍了Human10，这是一个新创建的数据集，其中包含详细的和纹理化的全身重建以及对应的10个对象的原始RGB-D扫描。

在合成和真实数据集上的定性和定量实验结果表明，所提出的方法在重建模型的视觉质量和准确性方面优于现有的最新技术。

**关键词：**高保真重建，3D视觉，纹理重建，级联架构

# Abstract

We present a learning-based approach to reconstructing high-resolution three-dimensional (3D) shapes with detailed geometry and high-fidelity textures. Albeit extensively studied, algorithms for 3D reconstruction from multi-view depth-and-color(RGB-D) scans are still prone to measurement noise and occlusions; limited scanning or capturing angles also often lead to incomplete reconstructions. Propelled by recent advances in 3D deep learning techniques, in this paper, we introduce a novel computation- and memory-efficient cascaded 3D convolutional network architecture, which learns to reconstruct implicit surface representations as well as the corresponding color information from noisy and imperfect RGB-D maps. The proposed 3D neural network performs reconstruction in a progressive and coarse-to-fine manner, achieving unprecedented output resolution and fidelity. Meanwhile, an algorithm for end-to-end training of the proposed cascaded structure is developed. We further introduce Human10, a newly created dataset containing both detailed and textured full-body reconstructions as well as corresponding raw RGB-D scans of 10 subjects. Qualitative and quantitative experimental results on both synthetic and real-world dataset demonstrate that the presented approach outperforms existing state-of-the-art work regarding visual quality and accuracy of reconstructed models.

**Keywords：**High-fidelity reconstruction, texture reconstruction, 3D vision, cascaded architecture

**1问题提出与研究动机**

**1.1 需求分析与研究动机**

高保真重建3D对象和场景是3D环境理解，混合现实应用以及下一代机器人技术的关键，并且多年来一直是计算机视觉和计算机间隙论研究的主要领域之一。同时，诸如微软Kinect和Intel RealSense之类的消费级RGB-D传感器的可用性使更多的新手用户能够扫描周围的3D环境，从而需要强大的重建算法，例如能够容忍错误的输入数据（噪声、失真和缺失区域）。

尽管最近在3D环境重构方面取得了进展，但是从休闲扫描程序和消费级RGB-D传感器中获取不完美数据的高保真3D形状仍然是一个特别具有挑战性的问题。自从KinectFusion的开创性工作以来，已经提出了许多3D重建系统，这些系统经常使用场景几何的体积表示，即TSDF。但是，消费级深度相机获取的深度测量包含大量噪声，加上有限的扫描角度会导致缺失区域，从而使vanilla深度融合有表面细节模糊和几何形状不完整的问题。另外有研究专注于从嘈杂和稀疏采样的点云中重建完整的几何体，但无法处理丢失数据比例很高的点云，并且可能会产生凸起的伪像。

大型3D模型资料库的更广泛可用性，刺激了数据驱动的形状重建和完成方法的发展。

在这项工作中，我们提出了一种使用3D级联完全卷积网络（3D-CFCN)架构，从嘈杂和不完整的输入中对3D形状进行高保真体积重构的从粗到细的替代方案。我们的方法选择了最近推出的基于八叉树的高效3D深度学习数据结构作为基本构建块，但是我们建议使用多级网络级联来进行详细的形状信息重构，而不是使用标准的单级卷积神经网络。在该级联中，通过一系列子网逐步预测和优化对象的几何形状。选择级联结构的原理是双重的。首先，要预测高分辨率（5123，10243或更高）的几何信息，可能必须部署更深的3D神经网络，即使使用内存效率高的数据也可能显著增加内存需求表示形式。其次，通过将几何推论划分为多个阶段，我们还简化了学习任务，因为每个子网现在只需要学习以一定分辨率重建形状。

训练级联的体系结构是一项艰巨的任务，特别是在采用基于八叉树的数据表示形式时，需要预测输出八叉树的结构和值。因此，我们设计了子网，以学习在何处改进输入量的3D空间分区，并且相同的信息也用于指导连续阶段之间的数据传播，这通过避免穷举性使得端到端训练变得可行传播每个体积块。

尽管高分辨率形状重建提供的几何信息可以进行形状分析和物理模拟等应用，但获得正确的几何形状只是3D重建和建模的开始；恢复准确的色彩信息和3D形状的外观对于人和机器的感知也至关重要。因此，基于我们的级联网络架构，我们进一步介绍了一种用于纹理重构的集成方法。共同学习目标对象的几何和颜色信息，以提供一致性和效率。但是，即使具有高输出空间分辨率，对于神经网络来说，生成高频纹理细节仍然很困难。因此，我们建议将预测的颜色和投影纹理融合，从而生成具有丰富细节的完整纹理贴图。

我们工作的主要贡献是一种基于学习的渐进方法，可以从不完整的数据中进行高精度3D形状重建，并且还提供了一种用于恢复高保真形状纹理的混合方法。为了训练和定量评估，我们在真实世界3D形状上的模型，我们还提供了一个数据集，其中包含详细的全身重建和10个对象的原始RGB-D扫描。然后，我们在模拟数据集和实际数据集上进行仔细地实验，将建议的框架与各种最新的替代方案进行比较。这些实验表明，在处理嘈杂和不完整的输入时，我们的方法所生成的3D形状比其他现有方法具有更高的准确性和质量。

在本文中，我们将进一步解决色彩信息重建问题，旨在恢复完整而详细的纹理图。此外，我们提供了更多的定量评估，包括与最新方法的比较以及对所提出的级联体系结构的计算消耗分析。

**2 研究内容**

在这项工作中，我们提出了一种使用3D级联完全卷积网络（3D-CFCN)架构，从嘈杂和不完整的输入中对3D形状进行高保真体积重构的从粗到细的替代方案。我们的方法选择了最近推出的基于octree的高效3D深度学习数据结构作为基本构建块，但是我们建议使用多级网络级联来进行详细的形状信息重构，而不是使用标准的单级卷积神经网络。在该级联中，通过一系列子网逐步预测和优化对象的几何形状。选择级联结构的原理是双重的。首先，要预测高分辨率（5123，10243或更高）的几何信息，可能必须部署更深的3D神经网络，即使使用内存效率高的数据也可能显著增加内存需求表示形式。其次，通过将几何推论划分为多个阶段，我们还简化了学习任务，因为每个子网现在只需要学习以一定分辨率重建形状。

**2.1 系统架构**

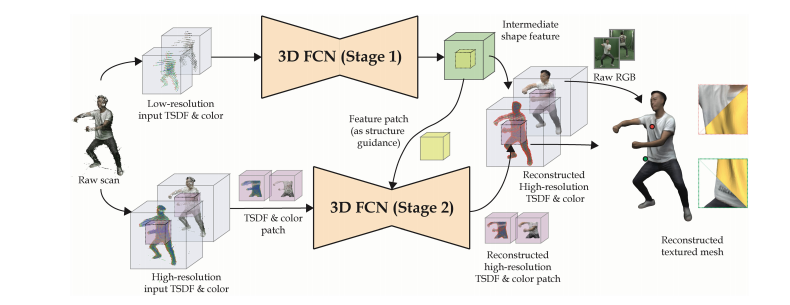
****

图1 两阶段3D-CFCN体系结构的图示。给定部分且嘈杂的原始RGB-D扫描作为输入，将融合的低分辨率TSDF彩色体积馈送到第一阶段3D全卷积网络（3D-FCN），产生中间表示。利用这一中间特征，网络接着回归分辨率低但完整的TSDF颜色体积，并预测应进一步完善的体积斑块。对于每个需要进一步优化的色块，将从融合的高分辨率TSDF颜色输入中裁剪出相应的快，然后在第二阶段的3D-FCN使用它来推断详细的高分辨率本地ISDF颜色体积，以代替前面提到的TSDF颜色体积中的相应区域，从而提高输出的分辨率。请注意，全局中间表示的补丁也进入阶段2，以提供结构指导。最后，将预测的颜色与原始RGB图像混合以生成最终的纹理贴图。最右边的列显示了高质量的重建。特写镜头可准确显示重构后的几何形状和外观细节，例如衣服上的皱纹和文字。请注意，输入扫描是从两个角度进行融合的。

我们的3D-CFCN是级联的体积重建模块，它们是基于OctNet的全卷积子网，旨在推断缺失的表面积并优化几何细节。每个模块在给定的体素分辨率和空间范围内运行。在我们的实验中，我们发现5123体素分辨率和相应的两阶段架构足以满足日常的3D扫描任务，因此在本文的其余部分中，我们将重点介绍这种架构。首先，虽然可以部署单级深度3D神经网络来推断高分辨的形状和颜色信息，但是使用内存高效的数据结构，计算和内存成本也将急剧增加，因为网络需要以每种分辨率全局处理体积特征。其次，我们通过将形状和颜色预测分为多个阶段来简化学习任务，同时每个子网仅负责以一定的分辨率和比例进行重构。

在我们的实现中，对于两个子网，我们都采用了U-net架构，同时用来自OctNet的相应操作来代替卷积和池化层。在相应的编码器和解码器层之间也采用跳过连接，以确保在推断的输出预测中保留输入量的结构。为了完成部分输入数据并优化其网格八叉树结构，我们避免使用OctNet的解池化操作，而是建议使用结构优化模块，该模块可以学习预测是否需要拆分八分圆图才能恢复更精细的几何细节。

第一个子网M0接收已编码的低分辨率（即1283）TSDF卷Vl,该分辨率是从3D对象的原始深度扫描结构优化模块。由于OctNet的解池操作限制了即时修改octree结构的可能性，我们采用用结构细化模块替换为合并的层。与从重构中隐式地推断出新的八叉树结构不同的是，我们使用33卷积滤波器，直接从特征图中预测是否应进一步分割八分圆。相反地，OGN预测三个状态掩码使用13滤波器，然后是三向softmax。为了确定是否需要将3D局部补丁馈送到Ml，我们取得此补丁中所有八分圆的平均分裂函数，并将其与置信度阀值p（=0.5）比较。通过采用自适应分区和传播方案，我们可以事先高分辨率的体积重建，同时将计算和存储成本保持在最低水平。

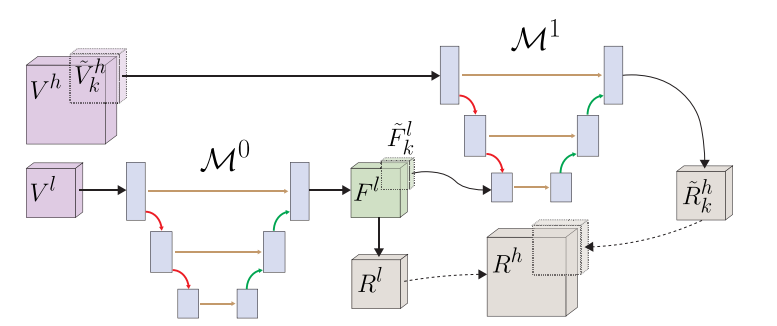


图2 两阶段3D-CFCN的体系架构。在这种情况下，网络会采用一对低噪声和高分辨率（即1283和5123）的嘈杂且不完整的TSDF卷Vl和Vh作为输入，并以5123体素分辨率生成精炼的TSDF体积。

**2.2 算法**

**2.3训练**

在TSDF上以受监督的方式对3D-CFCN进行了训练。数据集分两个阶段，其中Vl和Vh表示在低和高体素分辨率下不完整的输入TSDF，而G和G分别是低分辨率和高分辨率地面真实TSDF。

在第一个阶段中，M0是单独训练的，混合了l,二进制交叉熵和结构损失。

此外，我们采用了多层次的监督减轻潜在的梯度消失。具体来说，在每次合并操作之后，将要素图连接在一起，在相应的位置使用降采样的输入TSDF音量分辨率，我们评估每个结构细化层的缩减混合损耗。

在第二阶段，训练M1，同时，M0是在微调。为了减轻过度拟合并加快训练过程，在所有预计将被馈送到M1的局部补丁中，我们随机保留其中的k个并丢弃其余部分。

**2.3.1 综合训练集**

我们的第一个数据集建立在合成3D形状存储库ModelNet40上。我们选择10个类别的子集，总共有4051个形状实例（用于训练的是3245个，用于测试的是806个）。与现有方法类似，我们设置了虚拟围绕物体的摄像机并绘制深度图，然后模拟体积融合过程以生成真实的TSDF。为了产生噪声和局部训练样本，先前的方法是在深度图上添加随机噪声和空洞，以模拟传感器噪声。但是用这种方法在现的合成噪声通常不符合实际噪声分布。因此，改为使用合成立体深度相机。在实验中，生成训练样本时随机选择2或4个虚拟视点。

从本质上讲，除了形状完成外，学习体积深度融合还在于寻找函数将原始深度扫描映射到无噪声的TSDF。因此，为了保留来自所有输入深度扫描的信息，论文采用基于直方图的TSDF表示作为输入训练样本的编码。为每个体素分配一个100平滑直方图，改图使用5个负数的仓位和5个正数距离的仓位，其中第一个和最后一个仓位为截短的距离保留。深度观察的贡献线性地分布在两个最接近的仓之间。对于输出，我们只需选择普通的一维TSDF作为表示。由于我们在网络训练中采用了级联架构并使用了多尺度监督，因此我们需要以多种分辨率生成训练和真实样本对。

**2.3.2 真实数据集**

论文构建高质量动态3D重构（或自由视点视频，FVV）系统，并收集10个人类动作的4D序列，每个序列捕获一个不同的主题。然后随机采样了9746帧，从序列中按4:1的比例分为训练和测试集，并将其命名为Human10.对于每一帧，论文将2或4个随机选择的原始扫描融合在一起，并获得训练样本的TSDF-Hist编码；而真实的TSDF通过虚拟扫描产生相应的输出三角形FVV系统的网格。

**2.4 论文成果**

我们已经在ModelNet40和Human10上评估了我们的3D-CFCN架构，并将我们的方法的不同方面与其他最新的替代方案进行了比较。

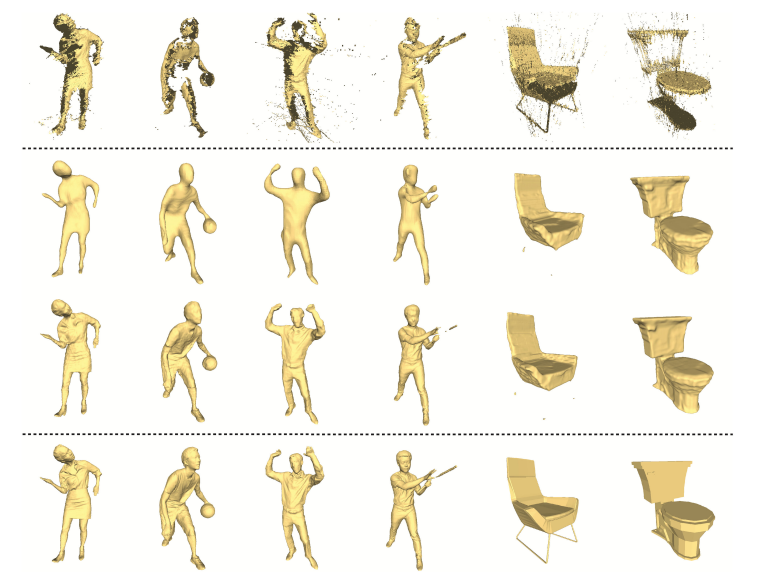


图3 高精度形状重建的结果库。第一行：从2个随机选择的视点融合的输入扫描。第二行：3D-CFCN第一阶段的重建结果。第三行：两阶段3D-CFCN体系结构的全分辨率重建结果。下排：真实的参考。

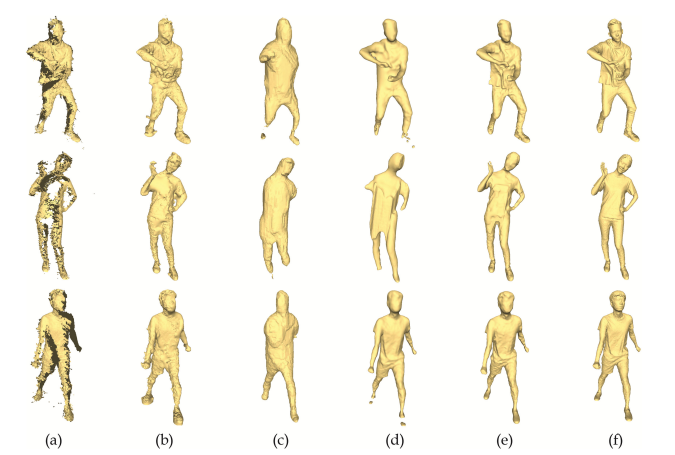
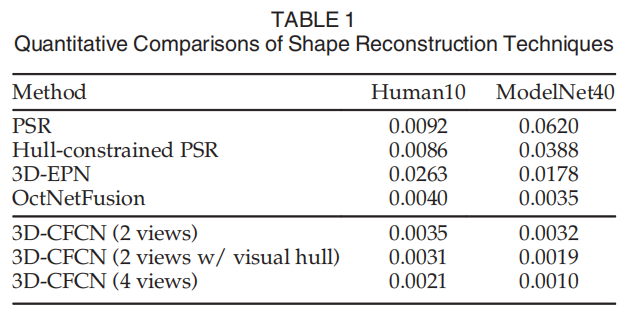


图4 论文形状重建结果与其他最新替代方案的比较 a：输入扫描 b：PSR

c：3D-EPN d：OctNetFusion e：论文 f：真实的参考。



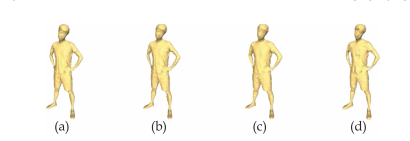
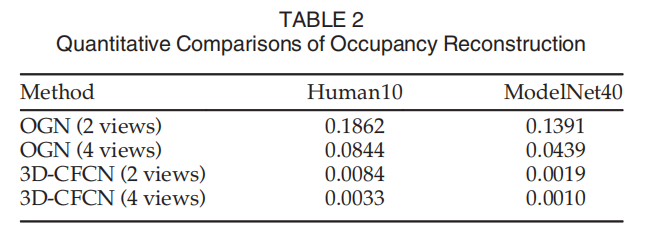


图5 在不同级别的校准误差下，提出的3D-CFCN的重建结果。a：没有错误 b：2.5% c：5% d：10%



由于输入为二进制占用量，因此我们使用平均Hamming距离来测量重建精度。我们的方法在两个数据集上的性能均优于OGN。



图6 我们的纹理重建结果与替代方法的比较。插入框突出显示了重建的3D模型上的瑕疵或细节。 a：输入扫描 b：经过筛选的PSR c：使用3D-CFCN预测的几何形状，并使用投影映射进行纹理化。 d：使用投影贴图进行纹理化，并使用PatchMatch进行修补。 e：使用投影贴图进行纹理处理，并使用Tizuka等人的方法进行修补。 f：论文结果 g：真实的参考。每个主题的第一行从2个视图融合，而第二行从4个视图中融合。

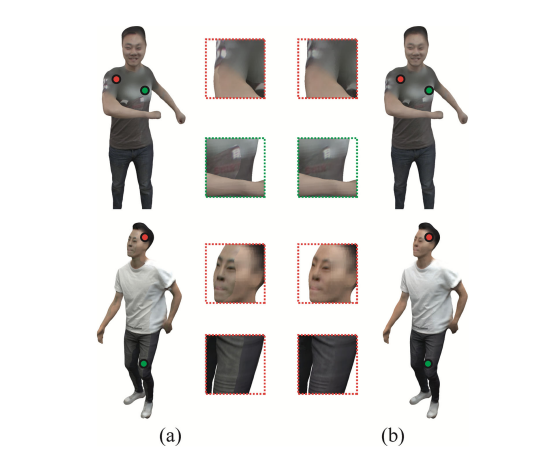


图7 纹理混合的效果。 a：没有融合 b：与泊松混合。插图框突出显示了重建模型的详细信息。

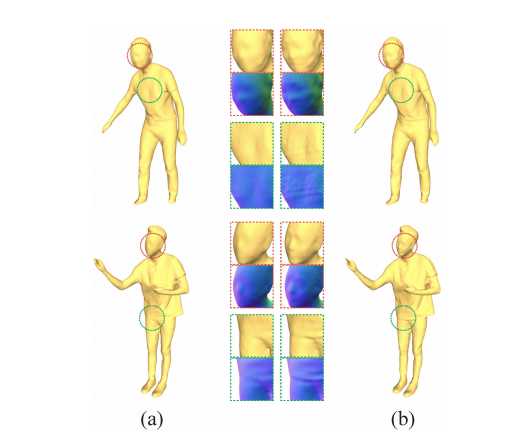
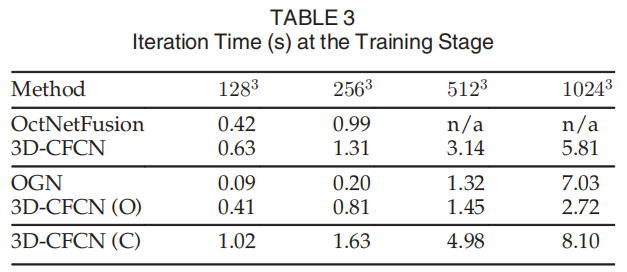


图8 带有辅助颜色预测任务的改进形状重构。 a：结果没有颜色重建损失。b：具有颜色重建损失的结果。插图框突出显示了改进。



3D-CFCN(C)预测TSDF和颜色量，而3D-CFCN(O)预测占用图。

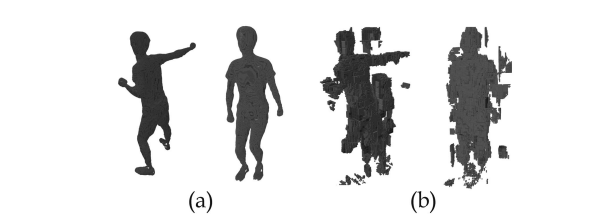


图9 与OGN的比较。 a：3D-CFCN重建的占用图 b：使用3D-CFCN学习的功能，由OGN解码的占用图。

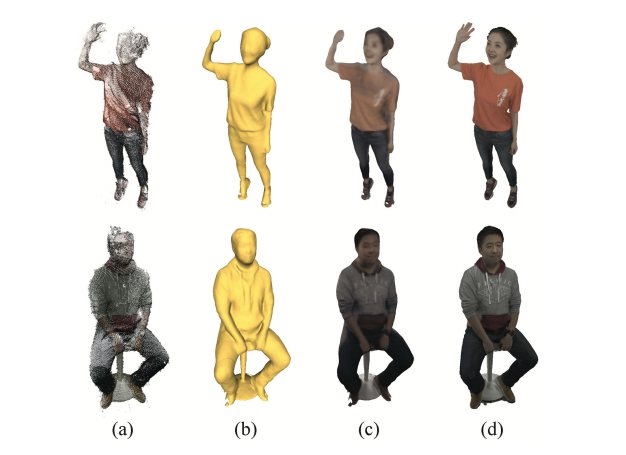
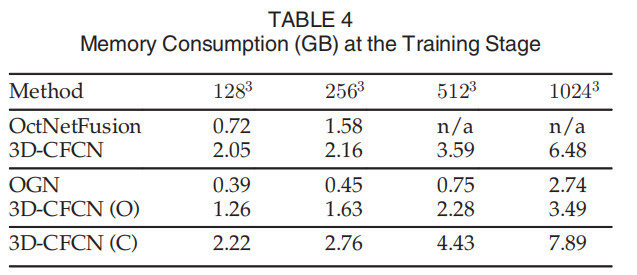


图10 3D-CFCN在看不见的数据上的泛化能力。 a：输入扫描（从2个角度融合）

b：使用3D-CFCN预测的几何形状。 c：使用3D-CFCN共同重建形状和颜色。 d：真实的参考。



3D-CFCN(C)预测TSDF和颜色量，而3D-CFCN(O)预测占用图。

**2.4.1 高质量的形状重建**

在我们的实验中，我们使用具有0.0001学习率的ADAM优化器，在每个数据集上分别训练20 epochs（阶段1为12个，阶段2为8个）得到3D-CFCN，这需要花费80个小时来收敛。图1、图3和图4说明了该3D CFCN体系结构所实现的高质量重建结果。

图3显示了Human10和ModelNet40数据集中的各种测试用例。从2个角度使用深度图对所有输入的TSDF-Hist进行融合，并应用相同的TSDF截断阀值。尽管存在大量的噪声和数据丢失，但此方法仍能够减少噪声并推断出丢失的结构，从而产生干净且详细的重建结果。比较第二和第三列，对于Human10模型，3D-CFCN的第2阶段通过为输出网格带来更多的几何细节，显著提高了质量。另一方面，ModelNet40的分辨率为1283.因此，在这些情况下，第二阶段并未显示出明显的改进。

1） 辅助视觉船体信息

实际上，大多数RGB-D传感器都可以捕获同步的深度和颜色图像，这为获得辅助分割蒙版提供了可能性。给定每个视图的分割蒙版，可以提取相应的可视外壳，该外壳实际上是一个占用空间。视觉船体提供了有关占用空间和空白空间分布的其他信息，这对于可靠的形状完成至关重要。因此，当视觉船体信息可用时，我们评估了3D-CFCN的性能。为了实现这一目标，我们在两个阶段都添加了相应的可视外壳输入分支，在两个33卷积层之后将它们与中间特征连接在一起。表1报告了预测的3D网格物体与真实的3D网格物体之间的平均Hausdorff RMS距离，表明使用额外的可视船体体积作为输入可带来约11%的性能提升。本实验中的TSDF-Hist和视觉船体体积均使用2个视点视点生成。为了进一步证明提出的视觉船体分支的有效性，我们还比较乱我们的方法和受船体约束的PSR的重构精度（表1，第二行），显示了所提出的3D-CFCN体系结构的显著优势。

2） 视点数

在这里，我们评估了输入TSDF-Hist的完整性（即用于融合原始深度扫描的视点数量）对重建质量的影响。我们使用融合了2个视点和4个视点的TSDF-Hist训练并测试了3D-CFCN体系结构，结果列于表1.正如预期的那样，由于输入TSDF-Hist更完整，使用更多深度扫描可提高输出网格的准确性。

3） 校准和跟踪误差的鲁棒性

除传感器噪声外，校准和跟踪误差方法是可能导致扫描模型破裂的另一个主要因素。为了评估所提出的校准和跟踪误差方法的鲁棒性，我们添加了随机扰动姿势（从2.5%到10%）到真实相机姿势，生成相应的测试样本以及使用3D-CFCN的预测重建结果。如图5所示，尽管尚未建立网络，经过校准误差样本的训练，仍然可以推断几何结构合理。

**2.4.2 纹理重建**

在这一部分中，我们仅考虑在Human10上进行的实验数据集。因为ModelNet40不提供真实数据纹理。λc在所有实验中均设置为0.5.剩下的实验设置与形状重建相同。

我们将我们的纹理重建方法与朴素的投影纹理映射，Screened Poisson Surface Reconstruction,PatchMatch和基于学习的图像完成方法进行了比较。与Screened Poisson Surface Reconstruction相比，3D-CFCN体系结构能够以更好的完整性和准确性来重建几何形状，从而减少投影纹理上的变形和重影伪影。虽然PatchMatch能够用合理的颜色填充缺失的区域。但有时无法修补大的孔；同时，在精细结构周围进行预测可能会有问题。饭冢等的方法也无法推断出大的缺失区域，并且可能产生重影效应。相比之下，我们的方法能够完成内容风一致的缺失区域。

我们在图7中进一步评估了纹理融合的效果。由于不一致的分辨率和不同视点的光照条件，我们可以在可见区域和遮挡区域之间的边界上观察到一些纹理接缝。这些对应的区域在混合后呈现出更自然的外观。

此外，图8说明了辅助RGB输入和颜色预测分支对几何重构的影响。比较带有和不带有RGB输入的形状重建结果，可以观察到颜色信息有助于增强精细几何细节的重建。从多任务学习的角度来看，颜色重建分支可提供互补的高频信号，而颜色重建分支则有助于改善训练和通用性。

**2.4.3 与现有的基于学习的方法的比较**

图4和表1比较了我们的3D-CFCN架构和3D形状重构的三种基于学习的最新技术，即OctNetFusion,3D-EPN和OGN，以及广泛使用的几何方法泊松曲面重建。

**2.4 泛化能力**

在图10中，我们演示了未包含在人体扫描中的方法的重建结果.

**2.5 计算效率**

表3和表4比较了OctNetFusion，OGN和建议的3D-CFCN的运行时长和内存消耗。

**3 结论和讨论**

我们提出了一种级联的3D卷积网络架构，以实现高效，高保真的形状和纹理高分辨率的重建。我们的方法以渐进和自适应的方式改进了部分输入模型和嘈杂输入模型的体积表示，从而大大简化了学习任务并降低了计算成本。实验结果表明，该方法可以产生具有精确几何细节和视觉上看似合理的纹理的高质量重建。我们还认为，将提出的方法扩展到重构动态序列是一个有前途的方向。

局限性。所提出的混合纹理形状重建方法的局限性在于，尽管它可以产生整体的高保真重建，但是在较大的遮挡区域中，以较低分辨率推断的纹理可能仍会引起明显的模糊伪影（参见图7b，第一行）。为了减少模糊效果，我们可以设计网络体系结构以预测不同空间分辨率下的几何形状和纹理（例如，添加更多的级联以进行颜色优化）。同时，预测颜色梯度而不是值可以帮助减少解空间，从而改善颜色重建。此外，在跨接缝混合纹理时，我们没有设置约束，这在某些情况下可能导致可见的接缝。

# 参考文献

1. A. Dai, M. Nießner, M. Zollh€ofer, S. Izadi, and C. Theobalt,“BundleFusion: Real-time globally consistent 3D reconstruction using on-the-fly surface reintegration,” ACM Trans. Graph.,vol. 36, no. 3, pp. 24:1–24:18, May 2017.
2. M. Kazhdan and H. Hoppe, “Screened poisson surface reconstruction,” ACM Trans. Graph., vol. 32, no. 3, pp. 29:1–29:13,Jul. 2013.
3. G. Riegler, A. O. Ulusoy, H. Bischof, and A. Geiger,“OctNetFusion: Learning depth fusion from data,” in Proc. Int.Conf. 3D Vis., 2017, pp. 57–66.
4. R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J.Davison, P. Kohi, J. Shotton, S. Hodges, and A. Fitzgibbon,“KinectFusion: Real-time dense surface mapping and tracking,”in Proc. 10th IEEE Int. Symp. Mixed Augmented Reality, Oct. 2011,pp. 127–136.
5. S. Choi, Q. Y. Zhou, and V. Koltun, “Robust reconstruction of indoor scenes,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit.,Jun. 2015, pp. 5556–5565.
6. F. Steinbr€ucker, J. Sturm, and D. Cremers, “Real-time visual odometry from dense RGB-D images,” in Proc. IEEE Int. Conf.Comput. Vis. Workshops, Nov. 2011, pp. 719–722.