第 10 卷 第 6 期
2021 年 11 月
集 成 技 术
JOURNAL OF INTEGRATION TECHNOLOGY
Vol. 10 No. 6
Nov. 2021

引文格式：
梁炎兴, 王映辉. 基于可见光单图像三维结构恢复方法综述 [J]. 集成技术, 2021, 10(6): 34-57.
Liang YX, Wang YH. 3D structure reconstruction methods based on visible light single image: a survey [J]. Journal of Integration Technology, 2021, 10(6): 34-57.

基于可见光单图像三维结构恢复方法综述

梁炎兴 王映辉*

（江南大学人工智能与计算机学院 无锡 214122）

摘 要 基于可见光单图像的三维重构方法一直是计算机视觉领域的研究热点，该文从光照物体的材质和结构差异，以及成像过程中信息损失等因素着手，对基于光照模型、基于几何图元以及基于深度学习策略的三维重建方法进行了分类和概述，并分析讨论各类方法的优缺点以及未来的研究方向。

关键词 三维恢复；单图像；计算机视觉；光照模型；深度学习
中图分类号 TP 391.4 **文献标志码** A doi: 10.12146/j.issn.2095-3135.20210618001

3D Structure Reconstruction Methods Based on Visible Light Single Image: A Survey

LIANG Yanxing WANG Yinghui*

(*School of Artificial Intelligence and Computer Scinece*, *Jiangnan University*, *Wuxi* 214122, *China*)
**Corresponding Author*: *wangyh@jiangnan.edu.cn*

Abstract Single image based three dimensional structure reconstruction is a classical and important topic in computer vision domain. This survey focus on image acquisition, such as surface material, surface shape and the information loss of target, and classified the single image based 3D reconstruction methods into three categories, i.e. illumination model, geometric element distribution, and deep learning. By analyzing and concluding the advantage and disadvantage of different methods, potential research direction is also suggested.

Keywords 3D reconstruction; single image; computer vision; illumination model; deep learning
Funding This work is supported by National Natural Science Foundation of China (61872291, 62172190)

收稿日期：2021-06-18 **修回日期**：2021-09-22
基金项目：国家自然科学基金资助项目（61872291, 62172190）
作者简介：梁炎兴，博士研究生，研究方向为三维计算机视觉；王映辉（通讯作者），教授，研究方向为三维计算机视觉，E-mail：wangyh@jiangnan.edu.cn。

1　引　言

从二维图像重建出三维结构作为计算机视觉的一个重要研究领域，已取得丰富的成果[1-2]。其主要任务是通过相机获取物体的二维图像信息，利用三维重建的相关理论分析处理、恢复真实物体的表面形貌。该技术广泛应用于人工智能、机器人、无人驾驶、虚拟现实、航空遥感测量、工业自动化等重要领域。目前，许多基于多图像的三维恢复方法[3-4]已被提出并得到广泛应用。虽然基于单图像的三维恢复方法因其病态性而更具难度，但其方便性一直受到业界和学术界的关注。尤其是基于可见光(波段在 380～760 nm 区间的肉眼可见光)而非结构光、红外激光、超声波等方式的单图像，其自身已经丢失很多关键几何信息，需要通过一些假设、先验知识，或借助基于已有模型的深度学习方法实现三维恢复。总的来说，基于可见光单图像的三维恢复存在以下影响因素和困难：

(1) 物体自身的材质差异。不同材质的物体因微观分子结构不同，呈现出不同的表面特性，如金属、白纸、玻璃等。如果只考虑某种特定材质的物体，往往会使三维恢复系统缺乏泛化性和鲁棒性[5]，而针对多个类别会因较大的类内差异和较小的类间差异导致重建精度下降[6]。

(2) 物体表面的几何结构差异。点、线、面代表了不同维度的几何结构，这些基本几何结构元素的组合构成了物体的表面形貌。同一个物体的不同区域，因物体表面凹凸、高低程度的不同，造成表面结构、轮廓的差异[7]。即使是同种材质的不同物体，也会因制造工艺、设计外形等因素导致物体表面的几何结构有较大的差异。

(3) 图像信息采集的损失。真实世界中的物体往往受到环境的影响，存在高光、阴影、遮挡、非刚性变形等现象[8]，加上相机拍摄角度、距离、镜头畸变、投影等因素，导致图像本身的信息损失甚至错误，干扰三维恢复的数据输入。

基于单图像的三维结构恢复是一个不确定性问题，即病态性问题，仅靠单幅图像无法得到唯一确定的三维恢复结果，如何利用一定的先验知识和预标定数据集，来指导和约束三维重建是一大难点。

基于以上困难，国内外许多研究成果给出了不同的解决方案和方法，概括起来包括：基于光照模型的方法、基于几何图元展布规律的方法和基于深度学习的方法。

2　基于光照模型的方法

2.1　基于纹理的形状恢复方法

从纹理恢复形状(Shape from Texture，SfT)的方法，是由 Gibson 于 1950 年首次提出[9]，它是一种根据物体表面纹理变化来推算表面形变情况，进而恢复出物体三维结构的方法。为了简化模型使其可计算，通常假定物体表面在一个水平面上，此时该方法将问题转变为估算物体所在平面的法向量。之后，该方法逐渐从平面扩展到光滑连续曲面[10]。

应用 SfT 方法必须满足以下先验条件：(1)纹理由规则的纹理单元组成[11]，并假定这些单元具有完全一致的固定形状(通常只有人工构造的规则图案才满足该要求)；(2)纹理分布具有均匀性[12]，即纹理密度相同；(3)纹理图像能够转换成基于频域的表示[13]；(4)纹理具有各向同性特性或随机相位特性[13]。

由于该方法限制条件严格，通用性弱，且纹理图案极易受到光照、阴影的影响，导致其准确性较低。该方法逐渐被基于明暗的形状恢复方法所代替。

2.2　基于明暗的形状恢复方法

基于明暗的形状恢复(Shape from Shading，SfS)方法是计算机视觉领域中三维结构恢复的重

要方法之一。该方法最早由 Horn[14]于 1986 年提出，其基本过程是借助一定的成像模型，从单幅图像的明暗变化出发，根据表面点的亮度取决于入射光线和表面法线之间的角度这一物理定理，通过施加约束条件求解物体表面的梯度场，进而由积分梯度的方式得到表面起伏高度值。基于 SfS 方法的系统具有设备简单、分辨率高、适用性强等优势，在工业生产过程检测[15]、医学图像分析与重建[16-17]、人脸与指纹等生物特征识别[18-19]、星球表面形貌重建[20]等领域得到广泛的应用。

2.2.1 经典的 SfS 方法

由于物体表面的明暗极易受到光源、形状、材质特性，以及相机或视点的角度、距离、参数等因素的影响，因此经典的 SfS 方法需满足以下前提假设[14]：(1)表面微观结构需要抽象为一种朗伯特反射模型；(2)物体表面各点的光照反射特性一致，且反射系数已知；(3)光源为无限远处的点光源；(4)物体表面与相机距离较远，成像几何关系满足正交投影。

如图 1 所示，由朗伯特反射模型可知，反射光的强度与入射光的强度，以及入射光与体表面法向量之间夹角的余弦值成正比[21]，如公式(1)所示：

$$E(x,y)=I(x,y)\cdot\rho\cdot\cos\theta \tag{1}$$

其中，x 和 y 为图像的二维坐标；$E(x,y)$为漫反射光强度；$I(x,y)$为光源强度；ρ 为表面反射系数；θ 为入射光与表面法向量之间的夹角。

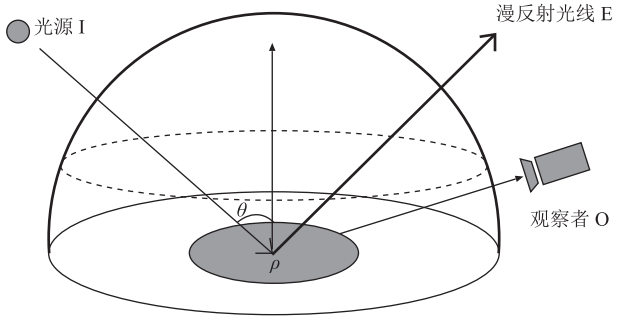


图 1 朗伯特反射模型示意图

Fig. 1 Lambertian reflection model

若以相机坐标系为参照系，设物体表面起伏高度为 $Z=Z(x,y)$，则物体表面的法方向可通过表面各点的法向量 $\boldsymbol{n}=(Z_x,Z_y,-1)$和表面梯度$(\boldsymbol{p},\boldsymbol{q})$表示，它们之间的关系如公式(2)~(3)所示：

$$\boldsymbol{p}=\boldsymbol{p}(x,y)=\frac{\partial Z}{\partial x}=Z_x \tag{2}$$

$$\boldsymbol{q}=\boldsymbol{q}(x,y)=\frac{\partial Z}{\partial y}=Z_y \tag{3}$$

由公式(1)～(3)可知，朗伯特反射模型可由公式(4)表示：

$$E(x,y)=R(\boldsymbol{p},\boldsymbol{q})=\frac{\boldsymbol{p}_0\boldsymbol{p}+\boldsymbol{q}_0\boldsymbol{q}+1}{\sqrt{\boldsymbol{p}_0^2+\boldsymbol{q}_0^2+1}\cdot\sqrt{\boldsymbol{p}^2+\boldsymbol{q}^2+1}} \tag{4}$$

其中，$E(x,y)$为归一化的图像亮度；$R(\boldsymbol{p},\boldsymbol{q})$为反射函数；$\boldsymbol{p}_0$ 和 $\boldsymbol{q}_0$ 为反射点沿光源方向的向量。通常，仅由该模型无法确定其唯一解，因此，必须建立联合表面反射模型和表面微观结构模型的正则化模型，对上式进行进一步约束和求解。

根据建立正则化模型方式的不同，SfS 算法大致可分为最小值方法、演化方法、线性化方法和局部方法等 4 类典型算法。

(1) 最小值方法

最小值方法就是将物体表面反射模型推导出的亮度方程和物体表面微观结构模型联合表示成一个能量函数的泛函极值问题或最优化问题，以求得最小值解或近似解。由于二维图像数据与由反射模型所确定的物体表面亮度之间存在误差，该方法首先将亮度方程转化为误差函数的形式；然后结合不同的约束条件(如光滑性约束[22]、可积性约束[22]、图像梯度约束[23]等)，联立得到新的泛函极值函数，并应用交错网格方法[22]或三角面元逼近方法[24]将其离散化；最后利用 Gauss-Seidel 迭代方法得到物体表面梯度$(\boldsymbol{p},\boldsymbol{q})$和表面起伏高度 Z 的网格点值。

(2) 演化方法

演化方法的核心是利用动力学思想，将 SfS

的泛函求解问题看作是一个 Hamilton 系统方程问题。当给定初值或边界条件时，该方程就转化为一个柯西初值问题或狄利克雷边界问题。这类问题通常可以利用特征线方法[25-26]、Viscosity 方法[27]、Level Sets 方法[28-29]等方法进行求解。其中，确定图像中唯一形状的特征点是关键，该点是演化过程开始的起点。演化过程从起点开始，搜索邻近点，找出其中远离光源方向的所有点，并从中筛选出离光源方向最近的点，再沿着该方向构成的演化路径计算图像中每一点的高度值，从而得到整个表面的高度[30]。由于演化过程是关于时间可微的，故应用演化方法求解 SfS 问题，实际上也隐含地利用了物体表面微观结构模型。

(3)线性化方法

线性化方法是指通过对反射函数作泰勒展开后，舍去其非线性项，将其转化为线性问题进行求解。该方法认为在反射函数中，低阶项占主要成份，舍去高阶项后的结果与真实情况接近，且满足泰勒展开的要求，从而间接要求物体表面的高低变化满足连续缓慢的特性。因此，先将表面反射函数表示为表面梯度的函数，并作泰勒展开，只保留常数项和一次项，两边同时进行傅里叶变换，然后根据光源方向的倾角和偏角对其进行改写，再进行逆傅里叶变换，即可得到物体表面的高度值[31]。

(4) 局部化方法

上述方法的求解过程是全局的，不能独立得到物体表面的局部形状表示。而局部化方法首先根据先验知识假定物体表面微观结构是一个特定的形状(如球形)；然后将反射模型与物体微观结构模型联合构成形状参数的线性偏微方程组，通过寻找图像特征点，旋转图像使其与光源方向在图像平面上的投影方向一致，计算在该坐标系下表面点的倾角 γ 和偏角 θ；最后利用边界条件迭代求解，即可直接确定物体的局部三维表面形状[32]。

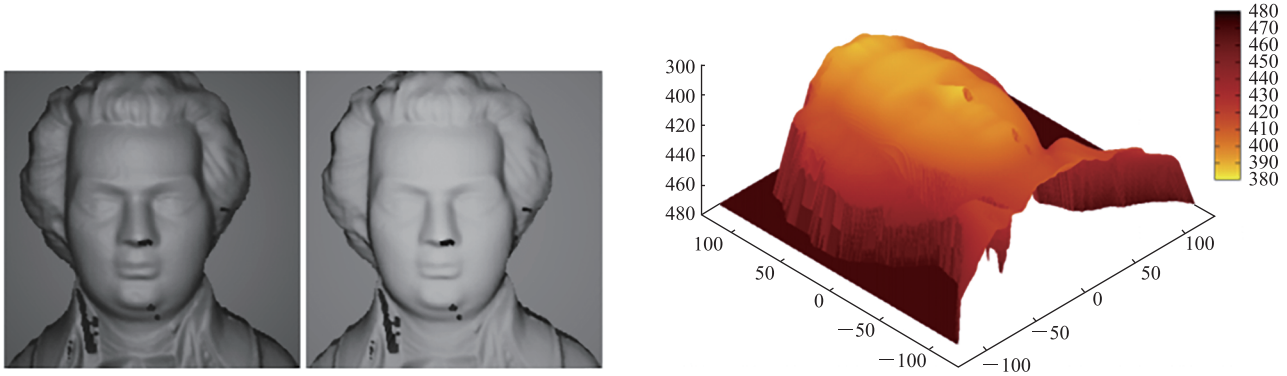
综上所述，经典 SfS 方法的算法复杂度高，对初始环境尤其是光照条件限制严格。朗伯特反射模型本身存在缺陷——理想漫反射的条件在现实中几乎无法满足，以此为基础的各种计算方法必然存在较大误差。但是，该类方法为其他方法奠定了许多光学和计算机渲染的理论基础，如后改进的 SfS 方法。

2.2.2 后改进的 SfS 方法

经典的 SfS 方法使用简化的成像模型(如假设物体表面满足理想反射、光源位于无限远处、相机遵循正交投影模型等)，尽管降低了 SfS 方法的复杂性，但也直接导致了三维恢复结果的误差较大。这是因为实际物体的表面并非理想的漫反射表面，而是既含有漫反射又含有镜面反射的混合表面。尤其是当相机距离物体表面较近时，相机不再满足正交投影，而接近于透视投影，甚至还会发生阴影、遮挡等现象，从而对表面各点的亮度产生更大的干扰。同时，实际物体的表面材质是非均匀、各向异性的，使得物体表面各点的光照反射特性不一致，反射系数也会随着表面起伏高度和凹凸发生变化。近年来，国内外诸多学者对经典的 SfS 方法展开了不同方面的研究和改进，衍生出许多突破前提假设的后改进的 SfS 方法。

(1) 基于表面微观结构的方法

经典的 SfS 算法中一个最重要的前提假设是物体表面的光反射模型遵循朗伯特反射模型，该模型是一个高度简化的理想模型，忽略了许多实际情况。因此，采用不同的表面微观结构模型和反射模型，尽可能地覆盖多种光照情况，可以大大提高三维恢复结果的精确性。Ahmed 等[33-34]首次建立了 Ward 模型[35]下的 SfS 图像辐照度方程，并利用 Lax-Friedrichs 算法[36]进行了求解。Vogel 等[37-38]提出了透视投影下基于 Phong 模型[39]的混合表面 SfS 方法的研究，如图 2 所示，同样使用了 Lax-Friedrichs 算法进行求解。Archinal 等[40]基于数字表面模型[41]利用月球轨道观察相机捕捉



(a) 输入图片 (Phong 模型，Oren-Nayar 模型)　　(b) 真值

图 2　使用 SfS 方法恢复莫扎特脸模[38]

Fig. 2　Shape from shading on the Mozart face[38]

到窄视角图像，通过光电映射增强技术，改进了月球表面重建模型的细节。O'Hara 等[42]使用朗伯特反射模型和 Oren-Nayar 反射模型[43]的混合模型，基于小孔成像相机模型，实现了单图像的火星地表重建。Yang 等[44]提出摒弃简单的反射模型，将基于径向基函数的模型拟合到数据中，其实验结果相比于朗伯特反射模型有明显提升。Camilli 等[45]研究了如何使一些非朗伯特模型应用在 SfS 方法的适配性问题上，拓展了 SfS 方法的普适性。王国珲等[46]提出一种基于牛顿-拉弗森法的 Blinn-Phong 混合表面模型的三维恢复快速 SfS 算法，相比于其他方法提高了求解效率。

(2) 基于光照反射率的方法

物体表面的凹凸和高度，可根据表面点的亮度变化，通过不同的反射模型计算得到。均匀的光照反射率假定物体表面是光滑的，忽略了现实的非均匀性和各向异性。对不同情况的反射率进行分类处理，有利于提高三维恢复的精度。Samaras 等[47]建立了具有分段恒定反射率的多视点 SfS 模型，并将其应用于人脸重建，提高了人脸模型的精细程度。Capanna 等[48]使用最大似然估计方法来降低噪声对不同材质的反射率的敏感性，并将其应用于重建 Lutetia 小行星中。Wu 等[49]使用单幅图像和不同的约束条件，从低分辨率表面模型中恢复出不同的反射率对应的不同形貌，结果表明可以达到和使用相对高分辨率图像一样的重建效果。

(3) 基于光源或相机与物体距离的方法

相机距离物体远近的不同直接决定后续计算使用正交投影还是透视投影，从而影响三维恢复的精度。Herbort 等[50]基于非朗伯特模型和可变反射率，通过主动距离扫描技术，不断改变相机和物体之间的距离，实现三维物体恢复，同时增加距离惩罚项进行优化约束，保证其精度接近原始曲面，以提高三维恢复模型的细节。Liu 等[51]仔细分析了光照方向和光源与物体的距离对三维结构恢复的影响，提出一种误差预测模型。该模型揭示了光源与物体表面的距离和方位角如何影响三维恢复精度。实验结果表明，在窄视角高分辨率相机采集的图像中，其三维恢复结果优于其他同时期的方法。

相比于经典的 SfS 方法，基于光源或相机与物体距离的方法在三维结构恢复的结果上有明显提升，可以根据不同的场景适应不同的重建要求。但良好的重建结果依赖于准确的先验知识，包括对光照情况的综合考虑、物体表面微观结构的精确建模、相机与视点的角度关系等。对于小范围的室内近距离单个物体，或结构简单的星球宏观地貌等，SfS 方法的三维恢复效果较好，而对于大范围的复杂室外场景，恢复效

果较差。为了提高室外场景的三维恢复效果，SfS 方法逐渐被以多视图几何理论为基础的运动结构恢复(Structure from Motion，SfM)方法和同时定位与地图生成(Simultaneous Localization And Mapping，SLAM)方法所取代，但这类方法不属于基于单图像范畴的三维结构恢复方法。

3　基于几何图元展布规律的方法

自然界中的部分物体，尤其是人造物体具有明显的几何规律[52-53]，如重复的纹理、对称的结构、规则的几何拼接图形、人造 CAD 模型等。借助几何规律这一重要特性，通过对单幅图像局部建模和全局拓展，就可以恢复出完整的三维模型。具体可分为利用二维几何特征的方法和利用三维构造模型的方法。

3.1　基于二维几何特征的方法

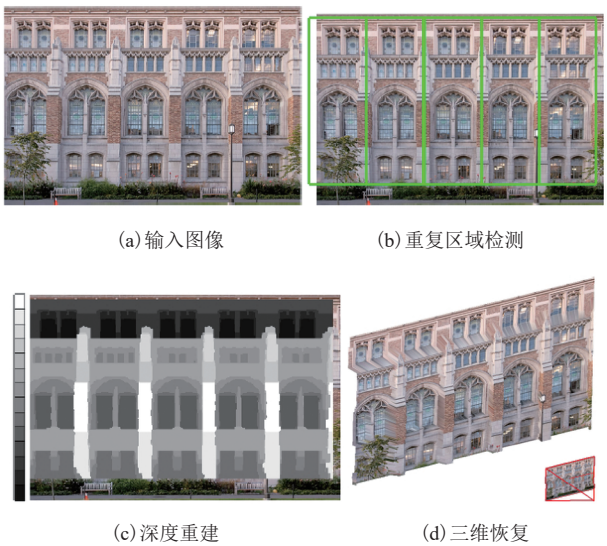
基于二维几何特征的方法是指一个三维模型映射在二维平面上的几何图形具有诸如对称、重复等规律，通过将一个单位图元旋转、平移或缩放就可以反推出整个三维模型。

该方法的第一步是定义和检测这种规律，即需要对目标形状或预先训练的模型进行强约束[54]。Wei 等[55]对此提出了一般对称性的概念(包括平移对称、旋转对称和反射对称)，Chertok[56]、Lee[57]和 Loy[58]在二维图像的对称性检测方面也做了许多工作。这些定义和方法针对特定的目标类(如人脸[54]，人体[59]和汽车[60])或某些特定场景(如具有平面墙、天花板和地板的室内场景[61]，具有重复图案的平面场景)取得了良好的效果。

第二步要针对邻近像素进行强制光度匹配，使二维单位图元重复拓展拼接形成三维模型的过程中，图元之间的拼接处更加平滑自然。通常使用基于马尔可夫随机场(Markov Random Field，MRF)的立体优化来强制匹配像素之间的光度一致性，使用一个平滑项来惩罚像素邻域之间的不一致性[62-64]。

第三步为了使图元之间具有相互一致的深度值，还需要对三维模型的深度图进行建模。Zabih 等[65]定义多个图像之间的交互集并强制可见性约束，Sun 等[66]使用遮挡项来惩罚遮挡，这间接地使深度贴图保持了一致。

基于上述 3 个重建步骤，许多学者提出了系统性框架。Wu 等[67]提出一种侧重于利用图元重复性的框架，该框架能通过输入单幅图像，自动检测重复区域，并将其以图像中稠密像素匹配的形式恢复出三维模型，如图 3 所示。该匹配关系由一个区间图表示，区间图表示图像中每个像素与其匹配像素之间的距离。为了获得稠密的重复结构，该方法还提出了一个图割来平衡高层次的几何重复约束、低层次的光度一致性和空间平滑性约束，以消除重复拼接处的不一致性。Xue 等[68]提出一种侧重于利用图元对称性缩小搜索空间的框架，通过输入一个对称分段平面物体的单幅图像，寻找所有的对称线匹配对，然后基于对称线和平面线，通过 MRF 恢复出深度图，相比于其他方法计算效率更高。



(a)输入图像　(b)重复区域检测　(c)深度重建　(d)三维恢复

图 3　基于重复图元的单图像重建[67]

Fig. 3　Repetition-based dense single-view reconstruction[67]

相较于其他三维模型，中国古代建筑是一种典型的同时具备轴对称和中心对称特点的三维模型。王映辉教授团队针对此类问题进行了详细的研究[69]，并提出了一种中国唐朝风格的古建筑建模方法[70]。该方法只需要已知建筑物一个角的图像，就可以根据其几何特征规则恢复出完整的唐朝建筑三维模型，相比于其他方法具有数据量少、鲁棒性强的特点。基于上述建模方法，团队更进一步提出一种基于构件提取的室内场景重建方法[71]。该方法对几何图元规律进行了延伸和拓展，提出了模型构件理论。首先，利用形状检测和平面分解方法提取室内场景中基本形状构件，用基于边界检测方法及基于有向包围盒的方法实现室内场景中基本形状构件的拟合；然后，选择基本形状构件集中最大的构件作为基准构件，以基准构件为中心寻找最佳的组合构件，对组合构件与标准模型库的标准模型逐一匹配，寻找匹配度最高的构件组合，识别最佳组合构件组成的物体，并利用标准模型库中的对应标准模型进行替换；最后，完成室内场景的重建。该方法丰富了二维几何特征的种类和表达方式，保证了场景物体构件提取的准确性和场景物体的形状完整性，并提高了室内场景重建的准确性。

重复性和对称性是一种简单明确的先验条件，只需知道一个图元就可以根据规律重建出所有表面，大大减少了三维结构恢复的难度。但是，特定在一个三维模型上的图元无法用于处理另一个三维模型。理想的约束条件应尽可能广义，以适应更多的对象，但是也应尽可能严格，使问题收敛。

3.2 基于三维构造模型的方法

现实中有许多物体是具有简单几何构型的，如立方体、圆柱体等，也有许多物体是具有特殊固定形状的，如人脸是由眼睛、鼻子、嘴巴、耳朵和面部构成的，汽车是由底盘、车轮、车壳构成的。这些模型的三维结构清楚明确，只需通过对基本几何体拼接组合即可得到一个更大的复杂几何体。因此，在三维恢复过程中，使用特定的三维构造模型代替通用的光照反射模型可以大大提高重建精度。基于三维构造模型的方法由待表示对象的参数模型组成，通过寻找最佳拟合时的输入图像和三维模型投影之间的参数来实现重建。

最早 Pentland 针对自然界中的常见物体提出了超二次曲面模型[72]，为基于三维构造模型的方法奠定了基础。随后 Jia[73]提出了广义柱体的概念，并对所有柱类外形进行详细分类描述。Gupta 等[74]提出了针对方形物体的建模规则，如图 4 所示，将模型针对不同的应用场景进一步细化分类，提高了重建精度。Xiao 等[75]随后也提出了类似的建模规则。这些模型虽然都能对某种特定外形物体进行描述，但各模型的局限性太大，导致其适用面较为狭窄。王映辉等[76]提出一种实现三维网格细化的可调多边形方法，该方法首先通过将三角形的中心点映射到切线平面来生成映射点；然后将映射点按一定比例移动，逆时针连接，得到切平面上的可调多边形；最后形成可调三角形和四边形来填充可调多边形之间的间隙。该方法生成的细分曲面可以根据不同的运动系数灵活调整，相较于传统超二次曲面模型具有较强的鲁棒性和有效性。

随着 CAD 技术的不断成熟，基于 CAD 模型的方法[77-78]逐渐涌现。该类方法通过建立一组对应点描述模型，可以有效地确定物体的近似视点，从而粗略表示任意物体的近似外形。此外，还有基于 CAD 模型的非参数化重建的方法[79]，但是该方法仅限于对预先分割好的在线商品图像进行三维恢复，其局限性较大，究其原因是因为没有对模型的各个组成部分进行有效分割和内部特征表示。王映辉等[80-81]提出一套多域物质体数据内部分界面提取方法和多域物质体数据内部结构特征表达方法。该方法通过构建有向骨架树、

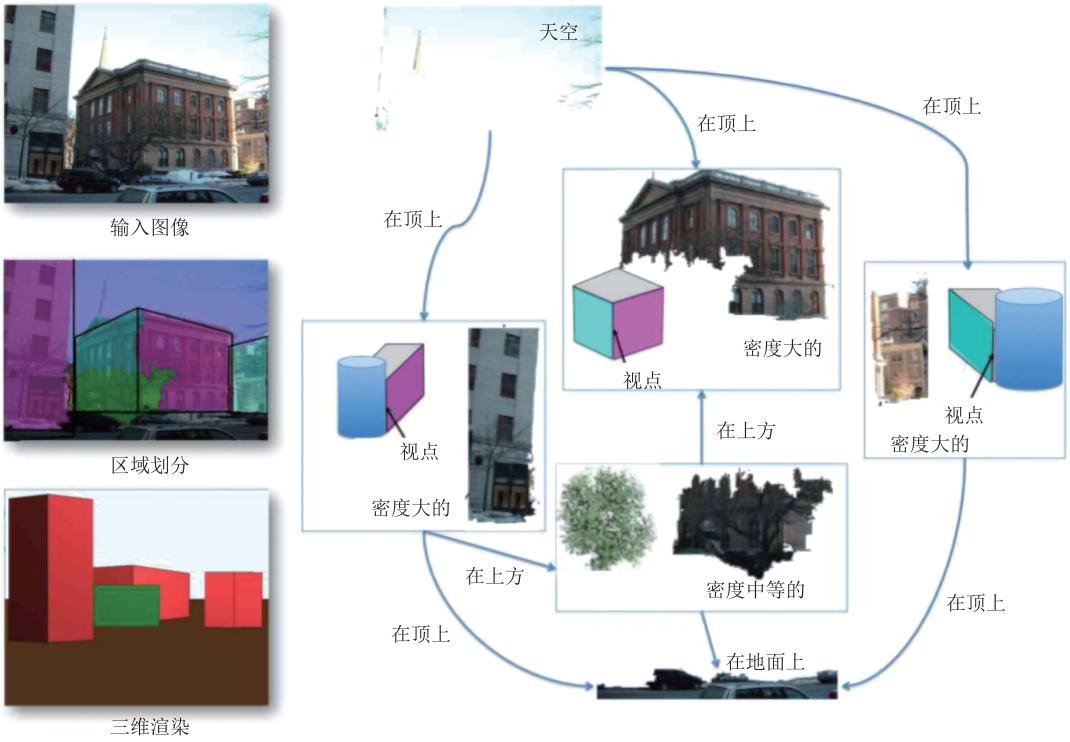


图 4　基于三维模型解析图的几何重建[74]

Fig. 4　Reconstruction based on 3D parse graph[74]

提取骨架形状特征和脊骨特征，借助树形结构拓扑进行向量表示，实现了体数据分界面形状特征的完整描述。实验结果表明，该方法不仅能够准确表达三维恢复模型，同时还能清晰地分割和描述模型内外的结构关系，增强了模型细节的精确性。

总体来看，基于几何图元展布规律的方法的先验知识，在图元或模型设计阶段就已经被设定好，可针对特定物体提供更多的先验信息，因此能取得较好的重建效果。虽然这类方法很难扩展到其他物体上，但因其应用面广泛，成为继 SfS 方法之后又一个重要的三维结构恢复方法。

4　基于深度学习的方法

深度学习(Deep Learning)源于对人工神经网络(Artificial Neural Network，ANN)进一步发展。本质上它是一种特征学习方法，负责把低层次的原始数据通过一些简单的、非线性的网络模型转化成为高层次的表达[82]。1986 年 Rumelhart 等[83]提出反向传播(Back Propagation，BP)算法，但由于该算法在梯度下降时会陷入局部极值，加之存在梯度消失、硬件算力不足等问题，未被大规模应用。直到 2006 年，Hinton 等[84]提出一种新的深度神经网络模型，该模型利用预训练的方法缓解了局部极值问题，降低了深度神经网络的优化难度和对计算机算力的要求，才使该类方法得以重新应用。2012 年，在 ImageNet 图像识别大赛中，Krizhevsky 等[85]采用深度学习模型 AlexNet 一举夺冠。从此，深度学习受到国内外业界学者的广泛关注和应用。随着一些新的网络结构、训练模型、训练数据集的出现，深度学习在语音识别[86-88]、自然语言处理[89-91]、图像识别和分割[92-93]等多个领域都取得了显著的效果。自 AlexNet 网络发布以来，深度学习在三维数据的分类、识别和重建上也取得了较

大的进展[94-95]。目前，广泛应用的深度学习模型主要包括深度置信网络(Deep Belief Network，DBN)[84,96]、卷积神经网络(Convolutional Neural Networks，CNN)[97]、循环神经网络(Recurrent Neural Networks，RNN)[98]、生成对抗网络(Generative Adversarial Networks，GAN)[99]等。

相较于二维图像领域，深度学习在三维重建上的研究起步较晚，但自 2012 年以来也取得了较大进展。其中，基于语义标签的方法是三维恢复深度学习方法得以应用的重要前提，也是实现通过数据集训练三维恢复深度网络的重要基础。场景的语义理解对于尺度和三维结构的感知起重要作用。基于语义标签的三维恢复方法是指从带有几何信息提示(如地平线、消失点、表面边界等)的单幅图像中生成空间上合理的场景三维恢复[100]。该方法通过了解像素或区域的语义类，可以很容易地实现深度和几何约束(如“天空”距离较远，“地面”是水平的)，从而建立局部二维图像和整体三维模型之间的映射关系。但是，要唯一确定绝对深度，还需要诸如纹理、相对深度、相机参数等额外信息。特别的，该方法非常依赖语义类的初始定义，语义类训练集的精准与否直接影响最终的重建效果。

目前，国际上公开的数据集包括 PASCAL3D+[101]、ObjectNet3D[102]和 IKEA[103]等。这些数据集对多个类别的物体语义和位姿信息进行预先人工标注。公开数据集为各大深度学习算法提供了一个相同的训练起点和参考标准，但是这些数据集也有其自身的局限性：(1)样本数量不足，仅限于很少的对象类别和样本；(2)只能从有限的标签字典中选择一个标签来标注模型，即使语义不够准确，也不能创造发明新的标签；(3)图像和三维模型因为拍摄视角、相机畸变等因素导致不能完全匹配；(4)数据集之间对标签的尺度定义不统一，存在线段、平面、CAD 模型等多种尺度。以上问题造成了深度学习方法在监督程度上的差异，从而直接影响三维恢复质量。根据实际应用需要，深度学习方法通常分为有监督学习、半监督学习和无监督学习。

4.1 有监督学习

Wu 等[104]建立 3D ShapeNets 网络，将三维几何外形标签表示为三维体素上二值变量的概率分布，通过吉布斯采样预测外形类型，实现填补未知空洞来完成重建。Kar 等[105]提出立体学习机系统，使用逆投影变换，将二维图像特征投影到三维模型网格中，利用单视点语义线索进行三维恢复。该系统在简化特征匹配过程的同时仍能保持较好的泛化性。Wu 等[106]提出 MarrNet 网络模型，在端到端生成重建结果的网络结构中加入 2.5D 草图，增强了重建效果，使网络可以针对不同类别的物体进行三维重建。Tulsiani 等[107]利用射线一致性约束构造了一个通用检测器，通过学习单视点的三维结构来训练多视点的几何一致性，使得普通 CNN 网络可以测量不同三维物体之间的外观一致性。Kato 等[108]提出一种近似梯度渲染网格渲染器，并将其集成到神经网络中，经过渲染器处理，使得神经网络可以通过输入单幅二维轮廓图像来监督三维结构重建过程。

特别的，对于一些具有固定形貌的三维物体，有监督学习可以极大帮助深度网络快速收敛，提高三维重聚的精确性。下面具体以人脸模型和人体模型为例进行简单介绍。

人脸具有明确的五官和高度的对称性[109-110]，且眼睛、眉毛、鼻子、嘴巴和耳朵等相对位置是固定的，深度学习网络只需根据输入的二维人脸图像，进行参数调整和模型变形，就能得到对应的三维人脸模型。3D 主动形变模型(3D Morphable Models，3DMM)[111-113]正是对应该思路的一种三维参数化模型，该模型通过利用原型人脸的大数据集进行人脸识别和图像编码，寻求构建基于图像的二维人脸线性表示。实现该模型的最直接思路就是在线性空间中嵌入所有三

维面部[114-116]，或从大量的三维激光扫描图像公开数据集中学习面部的密度函数参数[117-118]。借助 3DMM 人脸模型，Romdhani[119]提出了一种基于多特征的方法，该方法使用了非线性最小二乘优化拟合，提高了恢复精度。Jourabloo[120]使用 CNN 回归来估计和更新 3DMM 模型参数。虽然这些方法可以实现针对人脸的高精度模板生成和精确的单图像人脸重建，但是非常依赖图像与模板模型之间详细准确的逐点匹配和复杂的参数拟合过程，以及大量的人脸数据的支持。

为了简化模型训练和参数拟合的复杂度，Castelan 等[121]和 Dovgard 等[122]利用面部特征对称性，将所有模型的表面形状和亮度融合到一个单一的耦合统计模型中，简化了参数拟合的过程。这种方法可以生成更加精确的面部曲面轮廓，且当新面孔和存储的模板面孔之间形状差异很小时，可以将新面孔表示为存储的三维面孔的线性组合。但是在差异较大的情况下，需要调整模板以适应特定形状(如输入的是笑脸时，数据库应包括各种笑脸形状)。同时，该模型不能显式地对表面亮度进行建模，当图像明暗发生变化时会匹配失败，特别是针对肤色变化时，这种失效十分普遍。

Kemelmacher 等[123]提出并解决了一个用于正面图像的非凸优化问题，该方法使用深度图和反射率图代替普通的光照图，并针对深度值和反射率值增加了对应的损失函数，提高深度学习网络在不同亮度下的重建效果。Deng 等[124]提出一种利用 3DMM 模型的 R-Net、C-Net 的联合网络框架，如图 5 所示。该网络首先通过约束人脸表情、纹理、方位、光照等信息，利用鲁棒的混合损失函数进行弱监督学习，同时使用感知水平的信息作为置信度，结合图像与模型的互补信息进行形状聚集，最终实现人脸重建。Xu 等[125]使用 3DMM 模型以及其他头部区域的深度图作为输入，提出一种双层网络来重建头部模型。该模型首先使用自重建方法在单个图像上学习人脸形状，然后使用立体图像学习头发和耳朵的几何形状，不仅提高了精度，而且保证了整体头部几何形状的一致性。

同理，人体也是一种具有固定特征的模型，人体三维恢复的任务是从单幅图像中分析二维人体姿态[126-129]，估计一个简单的三维人体骨架[130-131]，从而实现完整的三维姿态和三维人体模型的恢复。虽然这个问题在多相机和多视图几何理论下得到了很好的解决[132-133]，但是对于单幅图像，不确定的成像条件和有限的数据集使得该任务变得非常复杂。传统基于优化的方法[134-136]为单目姿态

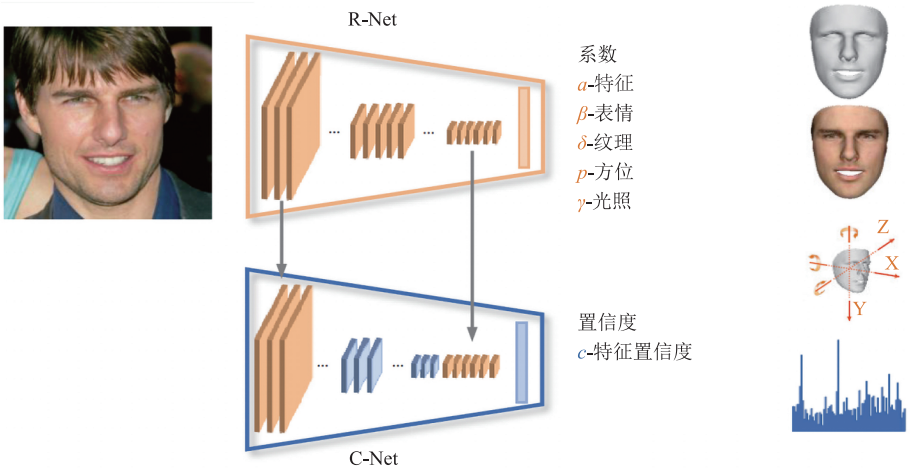


图 5　基于 R-Net、C-Net 的三维人脸精确重建[124]

Fig. 5　Accurate 3D Face Reconstruction With R-Net、C-Net[124]

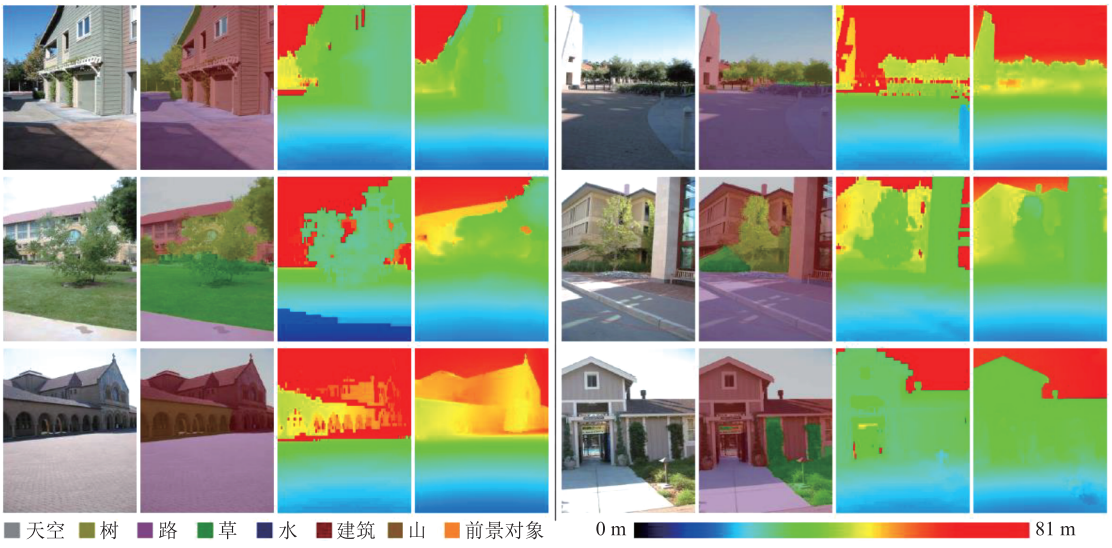
和形状恢复提供了最可靠的解决方案。然而，由于运行时间慢、对初始化条件的依赖，以及陷入局部极小值等问题使得效果并不显著。借助人体参数模型(Skinned Multi-Person Linear，SMPL)可直接从图像中回归姿势和形状，甚至特征点[135]、骨架点[137]、轮廓[137]、语义分割[138]或原始像素[139]。以 Kolotouros 等[140]的方法为例，该方法首先使用 SMPL 作为人体模型的模板引入网格；然后引入 GraphCNN[141]直接处理输入的单幅图像并提取特征点，随后直接附着在 SMPL 模型的顶点坐标图结构中以便继续处理；最后每个顶点都将其在 SMPL 模型变形网格中的三维位置作为最终的输出结果。该方法能直接恢复出人体的完整三维几何模型，而无需显式地求解预先指定的参数化空间。同时，在得到每个顶点的三维坐标后，如果需要适配并预测符合特定的模型，只需要从当前模型中反向回归其参数即可。Jiang 等[142]提出一种基于 SMPL 参数模型和距离场的深度学习网络，能够同时利用两种损失函数参与网络训练，生成更加准确的人体姿态模型。Zhu 等[143]提出一种结合参数模型与自由形变的深度学习网络，该网络利用身体关节、轮廓和每个像素着色信息的约束信息进行分层网格变形优化，不仅能恢复出完整人体模型，而且能实现精准的纹理贴图匹配。

4.2 半监督学习

不同于直接使用三维模型数据集或三维参数模型数据集训练深度网络求解绝对深度信息的有监督学习方法，半监督学习方法使用三维空间上的特征(如特征点、特征线段、特征面)作为语义标签，建立标签和深度信息的关联性，从而实现三维模型恢复。

Delage 等[144]利用室内场景中的几何线索(如天花板和墙壁的接缝)，使用 MRF 重建墙壁、天花板和地板的相对位置。Hedau 等[145]利用相似的几何线索恢复了杂乱房间的空间布局。这两种方法对于简单的室内场景效果明显，但是对房间结构和房间布局有严格的要求，应用十分有限。Gould 等[146]提出的场景分解模型证明了户外场景中几何信息和语义之间的强相关性。Hoiem 等[147]提出了一套语义松散的几何集，定义诸如建筑物是垂直的，道路、草和水是水平的等概念，并构建了一种简单的三维恢复模型与之匹配，该模型可以通过“弹出”垂直区域来恢复结构。Russell 等[148]采用一种更具语义动机的方法——利用详细的人工标定数据集，来分割和推断区域和区域边缘的几何类别(如天空总是在尽可能远的深度，草地和道路形成支撑其他对象的地平面等)，并且通过建立相对于地平面的支撑和附着关系来完成深度推断。

除了单纯地使用数据集训练网络之外，与传统算法相结合的方法也可以帮助网络更快地收敛。Haines 等[149]利用深度学习预测预分割区域的连续三维方向，并将区域平面检测作为 MRF 模型的优化问题。Fouhey 等[150]首先检测凸/凹边、遮挡边界、超像素及其方向，然后将分组问题表述为二元二次规划问题。Heitz 等[151]将目标检测、多类图像标记和深度感知相结合。Liu 等[152]基于 Hetiz 的方法，将 MRF 和机器学习相结合。该方法首先使用一个学习好的多类图像标签集来推断图像中每个像素的语义类，该标签集设置为：天空、树、路、草、水、建筑、山和前景对象(前 7 类覆盖了室外场景中的大部分背景区域，而最后一类负责标记一组前景对象)。然后使用基于像素和超像素的机器学习网络，结合全局深度优先、全局结构特征等规则约束，实现了较好的重建结果，如图 6 所示。Yang 等[153]将复杂的分割问题转化为深度预测问题，不再显示区分各个标签，并提出了一种不需要区分真实地面的深度学习网络。然而，该方法受网络架构的影响，限制了预测平面的总数量，导致其在复杂场景中的性能下降。Liu 等[154]在 Yang 的方法基础



注：从左到右分别为原始图像、语义覆盖、深度真值、三维恢复深度真值(图例显示语义颜色标签和深度比例尺)

图 6　室外场景语义分类集[152]

Fig. 6　Outdoor scene semantic classification set[152]

上，提出基于 Mask R-CNN[155]的实例分割框架，解决了这个问题。

4.3　无监督学习

虽然基于有监督学习和半监督学习的方法效果显著，但构建大规模全覆盖的监督训练数据集十分困难，而且重建结果特别依赖数据集的标签质量。本质上讲，语义标签还是某种特定的人工图像特征，实际过程离不开求解从图像特征到深度的映射。尽管网络可以隐式推理出上下文语义，但是重建结果的优劣严重依赖语义集的设定，导致网络的泛化性不足，使用场景有限。随着研究的不断深入，一些无监督学习的方法逐渐被提出。

Rezende 等[2]首次提出一种无监督学习的三维重建网络结构。该网络实现了无需三维模型形貌标签，就可以直接通过二维图像进行端到端的无监督学习训练。虽然只适用于立方体和圆柱体等简单形状，但其证明了无监督学习三维表征的可能。Choy 等[156]提出一种基于标准长短期记忆网络(Long Short Term Memory，LSTM)的扩展网络结构——三维循环重建神经网络(3D Recurrent Reconstruction Neural Network，3D-R2N2)，并建立了大型 CAD 模型数据集 ModelNet。该网络无需利用图像分类标签进行训练，就能很好地适应缺乏纹理特征和宽基线特征的问题。虽然该网络在重建细节方面存在缺失，但由于实现了在单个架构中同时支持单视图和多视图重建，且实验结果均优于传统方法，使其具有十分重要的意义。Girdhar 等[157]提出的 TL-Embedding Network 网络首先在训练自编码器时利用像素网格学习三维模型嵌入，然后通过 ConvNets 输入二维图像找到对应的模型嵌入，最后经过解码器得到体素表示的三维重建模型。Yan 等[158]提出的透视变换网络(Perspective Tansformer Nets)在传统卷积神经网络中加入透视变换，同时将在不同特定视角下的二维物体轮廓和对应体素轮廓的距离作为新的损失函数，因此在无监督学习下取得了较好的泛化能力。Li 等[159]提出一种通过二维图像和轮廓的集合来预测目标对象的三维网格形状和纹理的深度学习网络，该网络将建模对象表示为可变形构件图像的集合，通过对大量可变性构件图像的分割，有效地加强了重建网格和原始图像之间的

语义一致性。由于该网络不需要三维监督、手动注释关键点、对象的多视图图像或 3D 参数化模板，因此很容易推广到没有此类标签的各种对象类别。

为了更好地利用二维图像和三维模型之间的着色信息，同时减少二维图像和三维模型之间匹配误差导致的“块状重叠”问题，Chang[160] 和 Hao[161] 都提出直接从带有纹理的合成 CAD 模型出发，使用合成图像训练深度模型以估计相机位姿和重建三维形状。纹理 CAD 模型能够表示任意方向和尺寸的曲面，并且借助纹理着色能够捕捉到更加精细的细节。其关键在于深度学习网络首先要训练无标签的二维图片集，然后训练与之对应的无标签的三维模型集，最后通过一定的惩罚函数将二者联立，并在输入一幅新图像时匹配判断。这类方法有两个优势：(1)避免了人工定义模型和人工标注可能带来的错误，同时纹理 CAD 模型之间可以任意组合，生成几乎无限量的具有精确真实姿态和三维模型的渲染训练图像[162-163]；(2)深度学习网络可以应对大量的外观变化，对复杂建模的效果尤为明显[164-165]。虽然纹理 CAD 模型在合成图像(即人工构造的纹理图像，或人工构造的纹理模型对应映射的二维图像)上有明显的效果，但在应用于自然图像(即非合成图像)时，性能有明显下降[164]。为了克服这个问题，个别学者[166]尝试在训练集中添加少量人工标记的自然图像来微调网络参数，但是人工标签又会引入由于标注错误带来的误差。

此外，一些学者尝试利用 GAN 网络进行三维恢复[167-170]。其中，具有代表性的是 Wu 等[169] 提出的 3D-VAE-GAN 网络。该网络首先通过变分自编码网络得到输入二维图像的潜在向量，然后通过 GAN 网络的生成器得到重建物体。其优点是可以从概率表征空间中采样新的三维对象，并且判别器带有三维物体识别的信息特征。实验表明，与 TL-Embedding Network 的重建精度相比，3D-VAE-GAN 网络取得了更好的效果。

综上所述，相较于传统方法，深度学习具有无需人工描述规则和设定参数、数据处理量大等诸多优势，并取得了明显成果。但深度学习也存在以下问题：(1)公共数据集较少。与目前千万级的二维图像数据集相比，三维模型公共数据集规模小、种类少，早期具有代表性的公开数据集如 PASCAL3D+[101] 和 ObjectNet3D[102] 已无法满足实际需要。(2)重建分辨率及精度问题。网络支持的重建物体分辨率通常是 32×32×32，且重建结果与真实模型对比，精度未达到 95% 以上，存在细节部分缺失严重的问题。但是三维相比于二维多了一个维度，若盲目增加分辨率会导致数据量呈指数级增长，极大降低计算效率。(3)单幅图像重建的不确定问题。与传统方法一样，基于深度学习的方法在利用单幅图像进行三维恢复时，一幅图像往往对应多个不同的三维模型。这种不确定性反映在训练集中就是两幅看起来相似的图像可能导致完全不同的重建结果。目前，只能通过尽可能准确的定义损失函数和外加约束条件来限制其结果的不确定性。

5 总结与展望

基于可见光单图像的三维结构恢复本身是一个不确定性问题。自 20 世纪 90 年代以来，国内外许多学者提出了各种方法，如表 1 所示。基于光照模型的方法通过图像的纹理和明暗关系，假设和建立物体表面的微观结构模型，构建二维图像和三维深度之间的对应关系，实现三维结构恢复。该方法在已知材质反射率(即消除了材质差异因素)的前提下，试图从几何结构差异作为切入点进行求解，但该方法极易受到实际环境的光线情况、相机视点和光照模型类型的影响，且计算量较大。基于几何图元展布规律的方法利用二维图像或三维模型存在的几何规律代替光照模

表 1　基于可见光单图像三维结构恢复方法对比

Table 1　Comparison of 3D structure reconstruction methods based on visible light single image

方法分类			基本原理	适用前提	特点
基于光照模型	基于纹理		纹理变化推算	纹理单元完全一致且形状固定	对纹理图案限制多，极易受到光照、阴影的影响
	基于明暗	经典方法	最小值	已知反射模型、表面微观结构模型、约束条件	理论成熟，计算方法多；受反射模型限制，忽略了许多前提条件以简化计算
			演化	已知反射模型、表面唯一特征点、边界约束条件	
			线性化	已知反射模型、表面微观结构模型、光源方向，且表面起伏缓慢连续	
			局部化	已知反射模型、表面微观结构模型、光源方向	
		后改进方法	表面微观结构	已知某种特定物体的表面微观结构模型	对某一种特定物体效果较好，但适用面较窄
			光照反射率	已知某种特定材料的反射率	需要提前测定材料的反射率
			光源或相机与物体的距离	已知光源或相机与物体表面的距离，且连续变化	需要严格控制和精确测量到光源、相机、物体三者之间的距离和倾角
基于几何图元展布			二维几何特征	已知某种特定图元和该图元在三维物体上各个面的分布情况	对规则三维几何体的重建效果显著，算法高效；受限于基本几何图元，适用面窄
			三维构造模型	已知某些基本三维几何构件及其组合方式	
基于深度学习			有监督	具有对被重建物体的先验知识和标签完备的数据集支持	依赖于网络架构的设计和训练数据集的构建
			半监督	标签完备的数据集支持	
			无监督	海量数据集支持	

型，通过平移、旋转、缩放、重复等操作实现三维恢复，从而回避了求解物体表面几何结构差异带来的误差问题，对于人造纹理和模型有明显的优势，但正是这种先验规律限制了该方法在其他不规则物体上的应用，导致其适用面较窄。基于深度学习的方法利用深度网络避免了传统方法中人工定义关系和人工设定参数的局限性，配合有监督、半监督或无监督的方法，实现了利用特征点、特征线段、特征面、特征模型等多维度的空间信息，根据输入图像直接得到对应三维深度点的求解过程。并且基于几何图元展布规律的方法依赖海量数据的支撑，有效减小了图像采集过程中可能带来的误差。但其缺点也显而易见：非常依赖网络架构设计和训练数据集的质量。虽然基于深度学习的方法比传统方法有了明显进步，但是完全依赖深度学习方法的效果仍不尽如人意。对于病态性问题，只有将单幅图像扩展到多幅图像，利用多视图几何理论才能尽可能地减小误差。其中，SfM 和 SLAM 是多视图几何理论的两种代表性方法，由于已经超出了单幅图像的讨论范围，敬请读者自行查阅相关资料。

从影响因素的角度来看，物体自身的材质差异和几何结构差异是决定三维恢复结果优劣的根本原因，而图像信息采集损失带来的不确定性是其外部原因。从现有方法来看，无法通过数学计算来精准求解三维结构，只能通过构造合理的光照模型或寻找规则的几何图元纹理来近似描述物体表面的微观结构，在误差允许的范围内缩小或忽略差异，亦或通过深度学习的方式，在网络训练的时候，通过增加大量高精度、高分辨率图像，以减少信息损失和不确定性，从而逼近真实物体的表面形貌。此外，采用多方法的融合统一框架将是解决上述问题的一种新的趋势。2020 年，Henderson 等[171]提出一种传统方法和深度学

习相结合的新型网络框架。该框架解决了从单幅图像中进行三维恢复，以及生成新的三维形状样本的问题。框架算法中不仅结合了传统的光照模型、先验模型的方法，而且同时支持无标注数据集的学习和带有语义标签的有监督学习。结果表明，该算法能适应单色光以及白光环境，可以自动调整阴影和轮廓在网络中的权重，生成的模型具有更精细的表面细节和较强的鲁棒性。这种集成优势是前文所述任何单一算法所无法实现的。

综上所述，基于可见光单图像三维结构恢复问题未来可以从以下几方面发展和突破：

(1)传统方法与基于深度学习方法相结合

现有基于深度学习方法相较于传统方法已经取得了明显的效果，但是深度学习网络的训练非常依赖数据集(数据集的好坏直接影响网络效果)。而基于可见光单图像三维结构恢复问题缺少相应的海量标准数据集，使得网络缺少泛化能力。传统方法虽然计算复杂度高，但由于其通用性强，目前仍然发挥不可缺少的作用。二者相结合，可以最大发挥其方法的优势，达到更好的重建效果。

(2)基于 GAN 网络或组合 GAN 网络实现三维重建

尽管 GAN 网络本身的特性导致在训练过程中引入噪声，使得训练结果不稳定，但是这种方法对于缺乏大型标准数据集的情况仍然显示出良好的潜力。此外，将 GAN 网络视作形状或轮廓先验知识模型的一部分，可以很好地帮助网络快速收敛，使其满足特定问题场景的需要。

(3)建立真实场景的大型标准训练数据集

多数研究者选用纯白背景或 CAD 模型渲染出的人工合成数据集进行训练。这些数据集环境复杂，标准不一，与真实场景差异较大，且每个物体的外形复杂程度差异很大，不利于网络的训练和最终实验数据的对比，致使其网络在真实环境中效果较弱。当下迫切需要参照二维图像领域构建一些大型的标准数据集供大家测试和对比使用。

总体而言，每种基于可见光单图像三维结构恢复方法在各自特定问题领域都取得了明显的成果，但是每种方法的普适性较弱，对问题的初始条件要求严格。单纯依靠某一种方法来解决恢复问题已经变得越发困难，未来基于多种方法以适用于更加广泛的通用场景的融合解决方案，特别是结合深度学习的途径，是一个亟待研究的重点方向。

参 考 文 献

[1] Shen SH. Accurate multiple view 3D reconstruction using patch-based stereo for large-scale scenes [J]. IEEE Transactionson Image Processing, 2013, 22(5): 1901-1914.

[2] Rezende DJ, Ali Eslami SM, Mohamed S, et al. Unsupervised learning of 3D structure from images [C] // Proceedings of the Conferenceon Neural Information Processing Systems, 2016: 4996-5004.

[3] Lhuillier M, Quan L. A quasi-dense approach to surface reconstruction from uncalibrated images [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(3): 418-433.

[4] Habbecke M, Kobbelt L. A surface-growing approach to multi-view stereo reconstruction [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2007: 1-8.

[5] Blanz V, Vetter T. A morphable model for the synthesis of 3D faces [C] // Proceedings of the Annual Conferenceon Computer Graphics and Interactive Techniques, 1999: 187-194.

[6] Bakshi S, Yang YH. Shape from shading for non-Lambertian surfaces [C] // Proceedings of the International Conference on Image Processing, 1994: 130-134.

[7] Jin HL, Soatto S, Yezzi AJ. Multi-view stereo reconstruction of dense shape and complex appearance [J]. International Journal of Computer Vision, 2005, 63(3): 175-189.

[8] Ahmed A, Farag A. Shape from shading for

hybrid surfaces [C] // Proceedings of the IEEE International Conference on Image Processing, 2007: 525-528.

[9] Gibson JJ. The perception of vision surfaces [J]. The American Journal of Psychology, 1950, 63(3): 367-384.

[10] 廖熠, 赵荣椿. 从明暗恢复形状(SFS)的几类典型算法分析与评价 [J]. 中国图象图形学报, 2001, 6(10): 11-19.
LiaoY, Zhao RC. Analysis and evaluation of several typical algorithms for shape from shading (SFS) [J]. Chinese Journal of Image and Graphics, 2001, 6(10): 11-19.

[11] Aloimonos J, Swain J. Shape from texture [C] // Proceedings of International Joint Conference Artificial Intelligence, 1985: 926-931.

[12] Kittler J, Illingworth J. On threshold selection using clustering criteria [J]. IEEE Transactions on Systems, Man and Cybernetics, 1985, 15(5): 652-655.

[13] Cheng HD, Jiang XH, Sun Y, et al. Color image segmentation: advances and prospects [J]. Pattern Recognition, 2001, 34(12): 2259-2281.

[14] Horn BKP, Brooks MJ. The variational approach to shape from shading [J]. Computer Vision, Graphics, and Image Processing, 1986, 33(2): 174-208.

[15] Yang L, Li E, Long T, et al. A welding quality detection method for arc welding robot based on 3D reconstruction with SFS algorithm [J]. The International Journal of Advanced Manufacturing Technology, 2018, 94(1-4): 1209-1220.

[16] Lei L, Li J, Liu M, et al. Shape from shading and optical flow used for 3D reconstruction of endoscope image [J]. Acta Oto-Laryngological, 2016, 136(11): 1190-1192.

[17] Turan M, Pilavci YY, Ganiyusufoglu I, et al. Sparse-then-dense alignment-based 3D map reconstruction method for endoscopic capsule robots [J]. Machine Vision and Applications, 2018, 29(2): 345-359.

[18] Hu JF, Zheng WS, Xie X, et al. Sparse transfer for facial shape-from-shading [J]. Pattern Recognition, 2017, 68: 272-285.

[19] Kumar A, Kwong C. Towards contactless, low-cost and accurate 3D fingerprint identification [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3): 681-696.

[20] Liu WC, Wu B. An integrated photogrammetric and photo clinometric approach for illumination-invariant pixel-resolution 3D mapping of the lunar surface [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 159: 153-168.

[21] Pentland A. Shape information from shading: a theory about human perception [C] // Proceedings of International Conferenceon Computer Vision, 1988: 404-413.

[22] Horn BKP. Height and gradient from shading [J]. International Journal of Computer Vision, 1990, 5(1): 37-75.

[23] Zheng Q, Chellappa R. Estimation of illuminant direction, albedo, and shape from shading [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1991, 13(7): 680-702.

[24] Lee KM, Kuo CJ. Shape from shading with a linear triangular element surface model [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1993, 15(8): 815-822.

[25] Oliensis J. Uniqueness in shape from shading [J]. International Journal of Computer Vision, 1991, 6(2): 75-104.

[26] Bruckstein AM. On shape from shading [J]. Computer Vision, Graphics, and Image Processing, 1988, 44(2): 139-154.

[27] Rouy E, Tourin A. A viscosity solutions approach to shape-from-shading [J]. SIAM Journal of Numerical Analysis, 1992, 29(3): 867-884.

[28] Kimmel R, Bruckstein AM. Tracking level sets by level sets: a method for solving the shape from shading problem [J]. Computer Vision and Image Understanding, 1995, 62(1): 47-58.

[29] Osher S. A level set formulation for the solution of the dirichlet problem for hamilton-jacobi equation

[J]. SIAM Journal on Mathematical Analysis, 1993, 24(5): 1145-1152.

[30] Bichsel M, Pentland AP. A simple algorithm for shape from shading [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 1992: 459-465.

[31] Pentland AP. Local shading analysis [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1984, 6(2): 170-187.

[32] Lee CH, Rosenfeld A. Improved methods of estimating shape from shading using the light source coordinate system [J]. Artificial Intelligence, 1985, 26(2): 125-143.

[33] Ahmed AH, Farag AA. Shape from shading under various imaging conditions [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2007: 1-8.

[34] Ahmed AH, Farag AA. Shape from shading for hybrid surfaces [C] // Proceedings of IEEE International Conference on Image Processing, 2007: 525-528.

[35] Ward GJ. Measuring and modeling anisotropic reflection [J]. Proceedings of the 19th Annual Conference on Computer Graphics and Interactive Techniques, 1992, 26(2): 265-272.

[36] Kao CY, Osher S, Qian J. Lax-Friedrichs sweeping scheme for static Hamilton-Jacobi equations [J]. Journal of Computational Physics, 2004, 196(1): 367-391.

[37] Vogel O, Breu M, Weickert J. Perspective shape from shading with non-Lambertian reflectance [C] // Proceedings of Symposium of the German Association for Pattern Recognition, 2008: 517-526.

[38] Vogel O, Cristiani E. Numerical schemes for advanced reflectance models for shape from shading [C] // Proceedings of the IEEE International Conference on Image Processing, 2011: 5-8.

[39] Phong BT. Illumination for computer generated pictures [J]. Communications of the ACM, 1975, 18(6): 311-317.

[40] Archinal BA, Gaddis LR, Hare TM, et al. Progress on high resolution mapping of the lunar south pole-aitken basin interior [C] // Proceedings of Lunar and Planetary Science Conference, 2011: 2316.

[41] Wu C, Wilburn B, Matsushita Y, et al. High-quality shape from multi-view stereo and shading under general illumination [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2011: 969-976.

[42] O'Hara R, Barnes D. A new shape from shading technique with application to mars express HRSC images [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2012, 67: 27-34.

[43] Oren M, Nayar SK. Generalization of the Lambertian model and implications for machine vision [J]. International Journal of Computer Vision, 1995, 14(3): 227-251.

[44] Yang ZM, Zhao HD. A new RBF reflection model for shape from shading [J]. 3D Research, 2017, 8(3): 1-10.

[45] Camilli F, Tozza S. A unified approach to the well-posedness of some non-Lambertian models in shape-from-shading theory [J]. SIAM Journal on Imaging Sciences, 2017, 10(1): 26-46.

[46] 王国珲, 张璇. 一种透视投影下混合表面 3D 重建的快速 SfS 算法 [J]. 光学学报, 2021, 41(12): 1-9.
Wang GH, Zhang X. A fast SfS algorithm for 3D reconstruction of mixed surfaces under perspective projection [J]. Journal of Optics, 2021, 41(12): 1-9.

[47] Samaras D, Metaxas D, Fua P, et al. Variable albedo surface reconstruction from stereo and shape from shading [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2000, 480-487.

[48] Capanna C, Gesquière G, Jorda L, et al. Three-dimensional reconstruction using multi resolution photoclinometry by deformation [J]. The Visual Computer, 2013, 29(6): 825-835.

[49] Wu B, Liu WC, Grumpe A, et al. Shape and albedo from shading (SAfS) for pixel-level DEM generation from monocular images constrained by low-resolution DEM [J]. International Archives of

the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2016, 41: 521-527.

[50] Herbort S, Grumpe A, Whler C. Reconstruction of non-Lambertian surfaces by fusion of shape from shading and active range scanning [C] // Proceedings of IEEE International Conference on Image Processing, 2011: 17-20.

[51] Liu WC, Wu B, Wöhler C. Effects of illumination differences on photometric stereo shape-and-albedo-from-shading for precision lunar surface reconstruction [J]. ISPRS Journal of Photogrammetry & Remote Sensing, 2018, 136: 58-72.

[52] Park M, Brocklehurst K, Collins RT, et al. Deformed lattice detection in real-world images using mean-shift belief propagation [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2009, 31(10): 1804-1816.

[53] Wu C, Frahm JM, Pollefeys M. Detecting large repetitive structures with salient boundaries [C] // Proceedings of the European Conference on Computer Vision, 2010: 142-155.

[54] Blanz V, Vetter T, Rockwood A. A morphable model for the synthesis of 3D faces [C] // Proceedings of the Conference on Computer Graphics and Interactive Techniques, 1999: 187-194.

[55] Wei H, Yang AY, Huang K, et al. On symmetry and multiple-view geometry: structure, pose, and calibration from a single image [J]. International Journal of Computer Vision, 2004, 60(3): 241-265.

[56] Chertok M, Keller Y. Spectral symmetry analysis [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 32(7): 1227-1238.

[57] Lee S, Liu Y. Curved glide-reflection symmetry detection [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2011, 34(2): 266-278.

[58] Loy G, Eklundh JO. Detecting symmetry and symmetric constellations of features [C]// Proceedings of the European Conference on Computer Vision, 2006: 508-521.

[59] Sigal L, Balan AO, Black MJ. Combined discriminative and generative articulated pose and non-rigid shape estimation [C] // Proceedings of the Conference on Neural Information Processing Systems, 2007: 1337-1344.

[60] Leotta MJ, Mundy JL. Predicting high resolution image edges with a generic, adaptive, 3D vehicle model [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2009: 1311-1318.

[61] Lee DC, Hebert M, Kanade T. Geometric reasoning for single image structure recovery [C] // Proceedings of IEEE Conference on Computer Vision & Pattern Recognition, 2009: 2136-2143.

[62] Boykov Y, Veksler O, Zabih R. Fast approximate energy minimization via graph cuts [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2001, 23(11): 1222-1239.

[63] Scharstein D, Szeliski R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms [J]. International Journal of Computer Vision, 2002, 47(1): 7-42.

[64] Tappen MF, Freeman WT. Comparison of graph cuts with belief propagation for stereo, using identical MRF parameters [C] // Proceedings of IEEE International Conference on Computer Vision, 2003: 900-907.

[65] Zabih R, Kolmogorov V, Gortler S. Generalized multi-camera scene reconstruction using graph cuts [C] // Proceedings of the International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition, 2003: 501-516.

[66] Sun J, Li Y, Bing S, et al. Symmetric stereo matching for occlusion handling [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2005: 399-406.

[67] Wu C, Frahm JM, Pollefeys M. Repetition-based dense single-view reconstruction [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2011: 3113-3120.

[68] Xue T, Liu J, Tang X. Symmetric piecewise planar

object reconstruction from a single image [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2011: 2577-2584.

[69] Ning XJ, Wang YH. 3D reconstruction of architecture appearance: a survey [J]. Journal of Computational Information Systems, 2013, 10(5): 3837-3848.

[70] 王映辉, 唐婧, 杨芳, 等. 中国唐朝风格古建筑的建模方法: 中国, CN103279983A [P]. 2016-01-27.
Wang YH, Tang J, Yang F, et al. Modeling method of ancient chinese tang dynasty architecture: China, CN103279983A [P]. 2016-01-27.

[71] 宁小娟, 巨晨阳, 王怡轩, 等. 一种基于构件提取的室内场景重建方法: 中国, CN109102535A [P]. 2018-12-18.
Ning XJ, Ju CY, Wang YX, et al. An indoor scene reconstruction method based on component extraction: China, CN109102535A [P]. 2018-12-18.

[72] Pentland AP. Perceptual organization and the representation of natural form [J]. Artificial Intelligence, 1986, 28(3): 293-331.

[73] Jia Y. Description and recognition of curved objects [C] // Proceedings of the 11th IAPR International Conference on Pattern Recognition, 1992: 464-467.

[74] Gupta A, Efros AA, Hebert M. Blocks world revisited: image understanding using qualitative geometry and mechanics [C] // Proceedings of European Conference on Computer Vision, 2010: 482-496.

[75] Xiao JX, Russell BC, Torralba A. Localizing 3D cuboidsin single-view images [C] // Proceedings of the 25th International Conference on Neural Information Processing Systems, 2012: 746-754.

[76] Wang YH, Hao W, Ning XJ, et al. An adjustable polygon connecting method for 3D mesh refinement [C]. Proceedings of 2014 International Conference on Virtual Reality and Visualization, 2014: 202-207.

[77] Satkin S, Rashid M, Lin J, et al. 3DNN: 3D nearest neighbor [J]. International Journal of Computer Vision, 2015, 111(1): 69-97.

[78] Pepik B, Stark M, Gehler P, et al. 3D object class detection in the wild [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop, 2015: 1-10.

[79] Huang QX, Wang H, Koltun V. Single-view reconstruction via joint analysis of image and shape collections [J]. ACM Transactions on Graphics, 2015, 34(4): 1-10.

[80] 王映辉, 张缓缓, 薛香莲. 一种多域物质体数据内部分界面提取方法: 中国, CN111369683A [P]. 2020-07-03.
Wang YH, Zhang HH, Xue XL. Method for extracting partial interfaces in multi-domain material volume data: China, CN111369683A [P]. 2020-07-03.

[81] 王映辉, 张缓缓, 薛香莲. 一种多域物质体数据内部结构特征表达方法: 中国, CN111341392A [P]. 2020-06-26.
Wang YH, Zhang HH, Xue XL. Method for expressing internal structure characteristics of multi-domain material body data: China, CN111341392A [P]. 2020-06-26.

[82] Lecun Y, Bengio Y, Hinton G. Deep learning [J]. Nature, 2015, 521(7553): 436.

[83] Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back propagating errors [J]. Nature, 1986, 323(6088): 533-536.

[84] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks [J]. Science, 2006, 313(5786): 504-507.

[85] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks [J]. Advances in Neural Information Processing Systems, 2012, 25: 1097-1105.

[86] 焦李成, 杨淑媛, 刘芳, 等. 神经网络七十年: 回顾与展望 [J]. 计算机学报, 2016, 39(8): 1697-1716.
Jiao LC, Yang SY, Liu F, et al. Seventy years beyond neural networks: retrospect and prospect [J]. Chinese Journal of Computers, 2016, 39(8): 1697-1716.

[87] Graves A, Mohamed AR, Hinton G. Speech

recognition with deep recurrent neural networks [C] // Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2013: 6645-6649.

[88] Feng X, Zhang YD, Glass J. Speech feature denoising and dereverberation via deep autoencoders for noisy reverber-ant speech recognition [C] // Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2014: 1759-1763.

[89] Collobert R, Weston J. A unified architecture for naturallanguage processing: deep neural networks with multi-task learning [C] // Proceedings of the International Conference on Machine Learning, 2008: 160-167.

[90] Huang EH, Socher R, Manning CD, et al. Improving word representations via global context and multiple word prototypes [C] // Proceedings of the Annual Meeting of the Association for Computational Linguistics, 2012: 873-882.

[91] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space [C] // Proceeding of the International Conference on Learning Representations, 2013: 1301–3781.

[92] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks [C] // Proceedings of the International Conference on Neural Information Processing Systems, 2012: 1097-1105.

[93] Le QV. Building high-level features using large scale unsupervised learning [C] // Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2013: 8595-8598.

[94] Socher R, Huval B, Bath B, et al. Convolutional-recursive deep learning for 3D object classification [C] // Proceedings of the International Conference on Neural Information Processing Systems, 2012: 656-664.

[95] Gupta S, Girshick R, Arbel´aez P, et al. Learning richfeatures from RGB-D images for object detection and segmentation [C] // Proceedings of the European Conference on Computer Vision, 2014: 345-360.

[96] Hinton GE, Osindero S, Teh YW. A fast learning algorithm for deep belief nets [J]. Neural Computation, 2006, 18(7): 1527-1554.

[97] Lecun Y, Bottou L. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[98] Williams RJ, Zipser D. A learning algorithm for continually running fully recurrent neural networks [J]. Neural Computation, 1989, 1(2): 270-280.

[99] Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks [J]. Advances in Neural Information Processing Systems, 2014, 3: 2672-2680.

[100] Criminisi A, Reid I, Zisserman A. Single view metrology [J]. International Journal of Computer Vision, 2000, 40(2): 123-148.

[101] Xiang Y, Mottaghi R, Savarese S. Beyond pascal: a benchmark for 3D object detection in the wild [C] // Proceedings of IEEE Winter Conference on Applications of Computer Vision, 2014: 75-82.

[102] Yu X, Kim W, Wei C, et al. ObjectNet3D: alarge scale database for 3D object recognition [C] // Proceedings of European Conference on Computer Vision, 2016: 160-176.

[103] Lim JJ, Pirsiavash H, Torralba A. Parsing IKEA objects: fine pose estimation [C] // Proceedings of the IEEE International Conference on Computer Vision, 2013: 2992-2999.

[104] Wu ZR, Song SR, Khosla A, et al. 3D ShapeNets: a deep representation for volumetric shapes [C] // Proceedings of the IEEE Conferenceon Computer Vision and Pattern Recognition, 2015: 1912-1920.

[105] Kar A, Hane C, Malik J. Learning a multi-view stereo machine [C] // Proceedings of the International Conferenceon Neural Information Processing Systems, 2017: 364-375.

[106] Wu JJ, Wang YF, Xue TF, et al. MarrNet: 3D shape reconstructionvia 2.5D sketches [C] // Proceedings of the International Conference on Neural Information Processing Systems, 2017: 540-550.

[107] Tulsiani S, Zhou TH, Efros AA, et al. Multi-view supervision for single-view reconstruction via differentiable ray consistency [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2626-2634.

[108] Kato H, Ushiku Y, Harada T. Neural 3D mesh renderer [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 3907-3916.

[109] Shimshoni I, Moses Y, Lindenbaum M. Shape reconstruction of 3D bilaterally symmetric surfaces [J]. International Journal of Computer Vision, 2000, 39(2): 97-110.

[110] Zhao WY, Chellappa R. Symmetric shape-from-shading using self-ratio image [J], International Journal of Computer Vision, 2001, 45(1): 55-75.

[111] Edwards GJ, Lanitis A, Taylor CJ, et al. Modelling the variability in face images [C] // Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, 1996: 328-333.

[112] Cootes TF, Edwards GJ, Taylor CJ. Active appearance models [C] // Proceedings of European Conference on Computer Vision, 2001, 23(6): 681-685.

[113] Lanitis A. Automatic interpretation and coding of face images using flexible models [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2002, 19(7): 743-756.

[114] Atick JJ, Griffin PA, Redlich AN. Statistical approach to shape from shading: reconstruction of 3D face surfaces from single 2D images [J]. Neural Computation, 1996, 8(6): 1321-1340.

[115] Zhou SK, Chellappa R, Jacobs DW. Characterization of human faces under illumination variations using rank, integrability, and symmetry constraints [C] // Proceedings of European Conference on Computer Vision, 2004: 588-601.

[116] Smith WA, Hancock ER. Recovering facial shape and albedo using a statistical model of surface normal direction [C] // Proceedings of International Conference on Computer Vision, 2005, 1(1): 588-595.

[117] Sim T, Kanade T. Combining models and exemplars for face recognition: an illuminating example [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2001: 1.

[118] Zhang L, Samaras D. Face recognition under variable lighting using harmonic image exemplars [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognitio, 2003: 1-1.

[119] Romdhani S, Vetter T. Estimating 3D shape and texture using pixel intensity, edges, specular highlights, texture constraints and a prior [C] // Proceedings of IEEE Computer Society Conference on Computer Vision & Pattern Recognition, 2005, 2: 986-993.

[120] Jourabloo A, Liu X. Large-pose face alignment via CNN-based dense 3D model fitting [C] // Proceedings of IEEE Computer Society Conference on Computer Vision & Pattern Recognition, 2016: 4188-4196.

[121] Castelan M, Smith W, Hancock ER. A coupled statistical model for face shape recovery from brightness images [J]. IEEE Transactions on Image Processing, 2007, 16: 1139.

[122] Dovgard R, Basri R. Statistical symmetric shape from shading for 3D structure recovery of faces [C] // Proceedings of European Conference on Computer Vision, 2006: 99-113.

[123] Kemelmacher-Shlizerman I, Basri R. 3D face reconstruction from a single image using a single reference face shape [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 33(2): 394-405.

[124] Deng Y, Yang J, Xu S, et al. Accurate 3D face reconstruction with weakly-supervised learning: from single image to image set [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019.

[125] Xu SC, Yang JL, Chen D, et al. Deep 3D portrait from a single image [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern

Recognition, 2020: 7710-7720.

[126] Newell A, Yang K, Jia D. Stacked hourglass networks for human pose estimation [C] // Proceedings of European Conference on Computer Vision, 2016: 483-499.

[127] Wei SE, Ramakrishna V, Kanade T, et al. Convolutional pose machines [C] // Proceedings of IEEE Computer Society Conference on Computer Vision & Pattern Recognition, 2016: 4724-4732.

[128] Pishchulin L, Insafutdinov E, Tang S, et al. Deepcut: joint subset partition and labeling for multi person pose estimation [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2016: 4929-4937.

[129] Zhe C, Simon T, Wei SE, et al. Realtime multi-person 2D pose estimation using part affinity fields [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7291-7299.

[130] Martinez J, Hossain R, Romero J, et al. A simple yet effective baseline for 3D human pose estimation [C] // Proceedings of IEEE International Conference on Computer Vision, 2017: 2659-2668.

[131] Mehta D, Sridhar S, Sotnychenko O, et al. VNect: real-time 3D human pose estimation with a single RGB camera [J]. ACM Transactions on Graphics, 2017, 36(4): 1-14.

[132] Huang Y, Bogo F, Lassner C, et al. Towards accurate marker-less human shape and pose estimation over time [C] // Proceedings of International Conference on 3D Vision, 2018: 421-430.

[133] Joo H, Simon T, Sheikh Y. Total capture: a 3D deformation model for tracking faces, hands, and bodies [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8320-8329.

[134] Bogo F, Kanazawa A, Lassner C, et al. Keep it SMPL: automatic estimation of 3D human pose and shape from a single image [C] // Proceedings of European Conference on Computer Vision, 2016: 561-578.

[135] Lassner C, Romero J, Kiefel M, et al. Unite the people: closing the loop between 3D and 2D human representations [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2017: 6050-6059.

[136] Andrei Z, Elisabeta M, Cristian S. Monocular 3D pose and shape estimation of multiple people in natural scenes: the importance of multiple scene constraints [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2018: 2148-2157.

[137] Pavlakos G, Zhu L, Zhou X, et al. Learning to estimate 3D human pose and shape from a single-color image [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2018: 459-468.

[138] Omran M, Lassner C, Pons-Moll G, et al. Neural body fitting: unifying deep learning and model-based human pose and shape estimation [C] // Proceedings of International Conference on 3D Vision, 2018: 484-494.

[139] Kanazawa A, Black MJ, Jacobs DW, et al. End-to-end recovery of human shape and pose [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7122-7131.

[140] Kolotouros N, Pavlakos G, Daniilidis K. Convolutional mesh regression for single-image human shape reconstruction [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2019: 4496-4505.

[141] Scarselli F, Gori M, Tsoi AC, et al. Computational capabilities of graph neural networks [J]. IEEE Transactions on Neural Networks, 2009, 20(1): 81-102.

[142] Jiang W, Kolotouros N, Pavlakos G, et al. Coherent reconstruction of multiple humans from a single image [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2020: 5578-5587.

[143] Zhu H, Zuo XX, Yang HT, et al. Detailed avatar

recovery from single image [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2021, DOI: 10.1109/TPAMI.2021.3102128.

[144] Delage E, Lee H, Ng AY. A dynamic bayesian network model for autonomous 3D reconstruction from a single indoor image [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2006, 2: 2418-2428.

[145] Hedau V, Hoiem D, Forsyth D. Recovering the spatial layout of cluttered rooms [C] // Proceedings of IEEE International Conference on Computer Vision, 2010: 1849-1856.

[146] Gould S, Fulton R, Koller D. Decomposing a scene into geometric and semantically consistent regions [C] // Proceedings of IEEE International Conference on Computer Vision, 2009: 1-8.

[147] Hoiem D, Efros AA, Hebert M. Recovering surface layout from an image [J]. International Journal of Computer Vision, 2007, 75(1): 151-172.

[148] Russell BC, Torralba A. Building a database of 3D scenes from user annotations [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2009: 2711-2718.

[149] Haines O, Calway A. Recognising planes in a single image [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 37(9): 1849-1861.

[150] Fouhey DF, Gupta A, Hebert M. Unfolding an indoor origami world [C] // Proceedings of the European Conference on Computer Vision, 2014: 687-702.

[151] Heitz G, Gould S, Saxena A, et al. Cascaded classification models: combining models for holistic scene understanding [C] // Proceedings of Conference on Neural Information Processing Systems, 2008: 641-648.

[152] Liu B, Gould S, Koller D. Single image depth estimation from predicted semantic labels [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2010: 1253-1260.

[153] Yang F, Zhou Z. Recovering 3D planes from a single image via convolutional neural networks [C] // Proceedings of the European Conference on Computer Vision, 2018: 85-100.

[154] Liu C, Kim K, Gu J, et al. PlaneR-CNN: 3D plane detection and reconstruction from a single image [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 4450-4459.

[155] He K, Gkioxari G, Dollár P, et al. Mask R-CNN [C] // Proceedings of IEEE International Conference on Computer Vision, 2017: 2961-2969.

[156] Choy CB, Xu DF, Gwak JY, et al. 3D-R2N2: a unified approach for single and multi-view 3D object reconstruction [C] // Proceedings of the European Conference on Computer Vision, 2016: 628-644.

[157] Girdhar R, Fouhey DF, Rodriguez M, et al. Learning a predictable and generative vector representation for objects [C] // Proceedings of the European Conference on Computer Vision, 2016: 484-499.

[158] Yan XC, Yang JM, Yumer E, et al. Perspective transformer nets: learning single-view 3D object reconstruction without 3D supervision [C] // Proceedings of the Conference on Neural Information Processing Systems, 2016: 1696-1704.

[159] Li XT, Liu SF, Kim K, et al. Self-supervised single-view 3D reconstruction via semantic consistency [C] // Proceedings of the European Conference on Computer Vision, 2020: 677-693.

[160] Chang AX, Funkhouser T, Guibas L, et al. ShapeNet: an information-rich 3D model repository [J]. Computer Science, 2015.

[161] Hao S, Qi CR, Li Y, et al. Render for CNN: viewpoint estimation in images using CNNs trained with rendered 3D model views [C] // Proceedings of the IEEE International Conference on Computer Vision, 2015: 2686-2694.

[162] Bansal A, Russell B, Gupta A. Marr revisited: 2D-3D alignment via surface normal prediction [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 5965-5974.

[163] Dosovitskiy A, Springenberg JT, Tatarchenko M, et al. Learning to generate chairs, tables and cars with convolutional networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(4): 692-705.

[164] Massa F, Russell BC, Aubry M. Deep exemplar 2D-3D detection by adapting from real to rendered views [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 6024-6033.

[165] Wu J, Zhang C, Xue T, et al. Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling [C] // Proceedings of Conference on Neural Information Processing System, 2016: 82-90.

[166] Sajjadi M, Scholkopf B, Hirsch M. EnhanceNet: single image super-resolution through automated texture synthesis [C] // Proceedings of the IEEE International Conference on Computer Vision, 2017: 4491-4500.

[167] Zhu R, Galoogahi H, Wang C, et al. Rethinking reprojection: closing the loop for pose-aware shape reconstruction from a single image [C] // Proceedings of the IEEE International Conference on Computer Vision, 2017: 57-65.

[168] Liu J, Yu F, Funkhouser T. Interactive 3D modeling with a generative adversarial network [C] // Proceedings of the International Conference on 3D Vision, 2018: 126-134.

[169] Wu JJ, Zhang CK, Xue TF, et al. Learning a probabilistic latent space of object shapes via 3D generative adversarial modeling [C] // Proceedings of the Conference on Neural Information Processing Systems, 2016: 82-90.

[170] Gadelha M, Maji S, Wang R. 3D shape induction from 2D views of multiple objects [C] // Proceedings of the International Conference on 3D Vision, 2017: 402-411.

[171] 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.