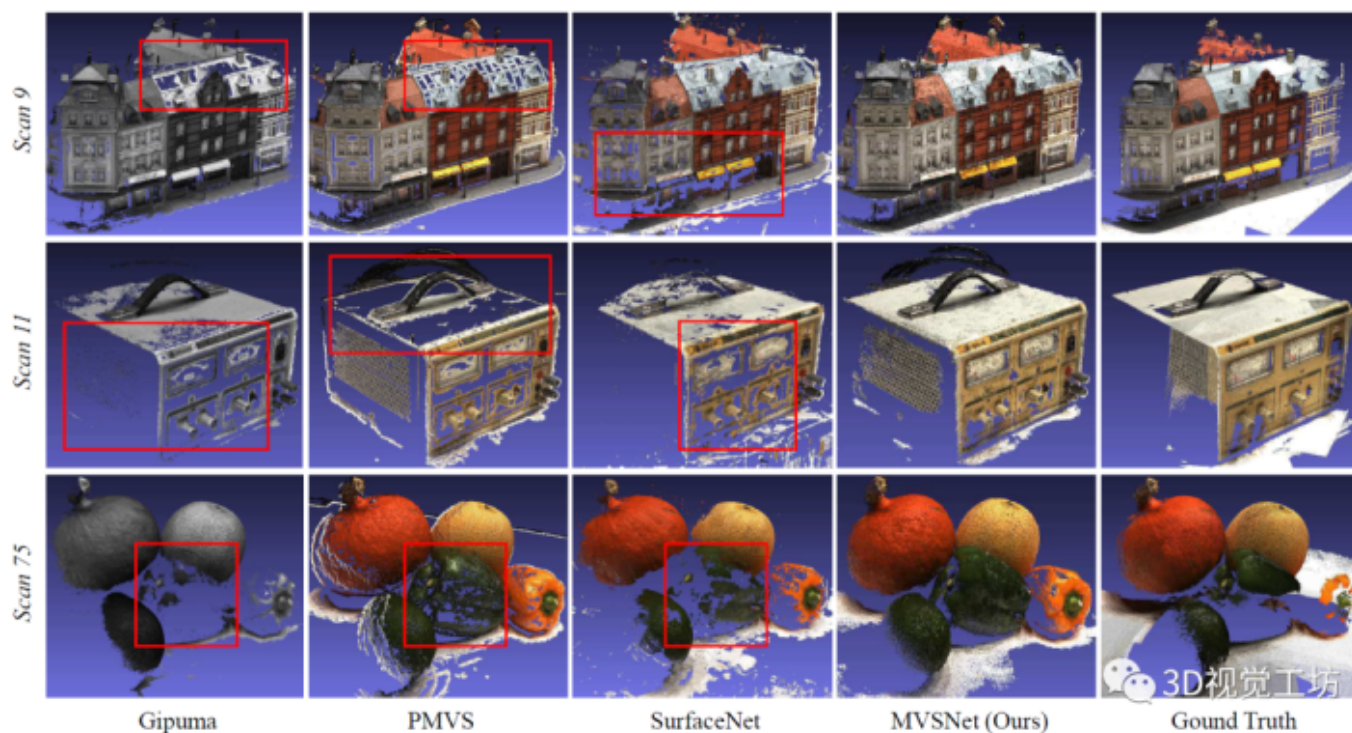


多视图立体匹配MVSNet网络解析：保姆级教程



Date: 2021-5-4

作者: Todd-Qi

首发: [公众号【3D视觉工坊】](#)

注1: 文末附有【三维重建】交流群加入方式哦~

注2: [视觉三维重建系统教程及实现](#)

论文题目: MVSNet: Depth Inference for Unstructured Multi-view Stereo, (ECCV2018 Oral)

摘要

作者提出MVSNet，基于多视图图像的深度估计网络。在网络中，首先在2D图像上进行特征提取得到特征图，然后通过可微分的单应变换，基于参考视图的相机视锥体构建3D代价体。然后使用3D卷积对代价体进行正则化，回归得到初始的深度图；初始的深度图通过参考图像优化得到最后的深度图。

MVSNet可以灵活地处理任意数量视图，因为基于方差的代价指标可将多个特征代价融合为一个特征代价体。MVSNet在大规模室内数据集DTU上进行验证。通过简单的后处理，不仅有显著的性能，还有更快的运行速度。同时也在更加复杂的室外数据集Tanks and Temples上进行评测。无需任何微调，MVSNet在排行榜上位列第一，证明了网络具有良好的泛化性能。

关键词： 多视图立体匹配；深度估计；深度学习

一、引言

多视图立体匹配（Multi-view Stereo, MVS）是计算机领域中的一个核心问题。多视图立体匹配重建可以当做对确定场景进行拍摄照片的逆过程。拍照通过相机把三维场景映射为二维，而多视图立体匹配重建则刚好相反，其目的是通过不同视点拍摄的图像，恢复出真实的三维场景。

传统的方法使用手工设计的相似性度量指标和正则化方法计算场景的稠密对应关系（比如使用归一化互相关Normalized Cross-Correlation和半全局匹配semi-global matching）。这些方法在非朗伯体表面、无弱纹理区域的场景可以达到很好的效果。但是在弱纹理区域，人工设计的相似性指标变得不可信，因此导致重建结果不完整。由MVS数据集的排行榜可知，这些方法具有很高的精度，然而在重建的完整度方法还有很大的空间可以提升。

卷积神经网络的研究的最新进展引发了人们对改善立体匹配重建的兴趣。从概念上讲，基于学习的方法可以学习到全局的语义信息，比如说有高光和反射的先验可以得到更加鲁棒的匹配效果。目前已经有一些两视图立体匹配的尝试，用神经网络替换手工设计的相似性度量或正则化方法。这些方法展现出更好的结果，并且逐步超过立体匹配领域的传统方法。事实上，立体匹配任务完全适合使用CNN，因为图像对是已经过修正过的，因此立体匹配问题转化为水平方向上逐像素的视差估计。

与双目立体匹配不同的是，MVS的输入是任意数目的视图，这是深度学习方法需要解决的一个棘手的问题。而且只有很少的工作意识到该问题，比如SurfaceNet事先重建彩色体素立方体，将所有像素的颜色信息和相机参数构成一个3D代价体，所构成的3D代价体即为网络的输入。然而受限于3D代价体巨大的内存消耗，SurfaceNet网络的规模很难增大：SurfaceNet运用了一个启发式的“分而治之”的策略，对于大规模重建场景则需要花费很长的时间。

为此，作者提出端到端的多视图深度预测网络MVSNet，和其他基于深度图的MVS方法类似，MVSNet将一张参考图像和多张源图像作为输入，为参考图像预测深度图，而不是整个3D场景。网络的关键之处在于可微分的单应变换操作，在从2D特征图构建3D代价体的过程中，网络将相机参数隐式地将编码入网络。为使网络能够适应任意数目的输入视图数，提出基于方差的指标，该指标将多个特征体映射为一个代价体。对代价体进行多尺度的3D卷积可以回归出一个初始的深度图。最后使用参考图像对深度图进行优化以提升边界区域的精度。我们的方法与之前基于学习方法有两个明显的不同。第一点，基于深度图估计的目的，本文的3D代价体的构建是基于相机视锥体而不是规整的欧式空间。第二点，本文方法将MVS重建划分为逐视图的深度估计问题，使得大规模的重建成为可能。

本文在DTU数据集上对所提出的MVSNet进行训练和评测。实验表明通过简单的后处理，MVSNet在完整度和总分上都要比其他方法要好。除此之外，还在室外数据集*Tanks and Temples*上验证了网络的泛化性能，MVSNet在榜单上排名第一（截止至2018年4月18号），无需任何调整，MVSNet的效果比开源MVS方法（COLMAP）和商业软件（Pixel4D）要好。同样值得注意的是，在运行速度方面MVSNet比之前最好的方法要快几倍甚至几个数量级。

二、相关工作

多视图立体匹配重建可以当做对确定场景进行拍摄照片的逆过程。拍照通过相机把三维场景映射到二维，而多视图立体匹配重建正好相反，其目标是借助从不同视点拍摄的图像，恢复真实的三维场景。

通过匹配不同视点拍摄的图像中的相似像素，可以恢复出物体的三维位置。假设空间有一个点 A 在一个物体的表面上，它投影在相机 1 拍摄的图像上的位置为 a_1 ，投影到相机 2 拍摄的图像上的位置为 a_2 。在得知两个相机的参数后（比如说它们的光心位置 c_1, c_2 ，可以求得射线 c_1a_1 和 c_2a_2 的交点，即空间点 A 的位置。此方法称之为三角测量。

 3D视觉工坊

在立体匹配算法研究中关键的一个组成部分是匹配不同视点拍摄的图像中相似点，即找到某一空间点 A 在不同图像中的投影位置，也就是找到所对应的点。该匹配的准则在多视图立体匹配重建领域及其他计算机视觉相关领域中都是至关重要的存在。其中有一个简单的准则，该准则根据比较像素的亮度和颜色来确定其相似度，然而这样的准则对光照的改变十分敏感。为了使图像匹配更加鲁棒，一些新的度量评价指标逐渐被提出：归一化互相关（Normalized Cross-Correlation, NCC），关键点描述符（SIFT描述符）。基于这些基本方法，立体匹配算法逐步进行改进。

匹配图像相似点的最终目的是得到图像的一致性。图像一致性的评定是多视图立体匹配的关键点之一，评定图像一致性的方法是比较图像中的像素点，比较的过程中可能用到三维的信息，包括体元，物体表面或是三角面片。评定匹配程度的指标

一般有如下几种：亮度平方差之和（Sum of Squared difference, SSD）；亮度绝对值差之和（Sum of absolute difference, SAD）等。

基于投影的技术也在深度图的重建中广泛应用。其中代表的方法就是平面扫描算法（Plane-Sweep）。平面扫描算法的发明者是Collins，他在1996年提出了这个算法，并且用于在输入图片上过滤检测到的边缘，以便于用这些灰度图片对场景进行立体三维重建。平面扫描的输入数据是一系列经过校准的照片。与其他重建方法一样，平面扫描方法也假设场景中所有物体只有漫反射。平面扫描虚构了一个虚拟的摄像机，在面对这个虚拟摄像机的空间中选定一个远平面以及一个近平面，需要重建的场景的物体要位于这两个平面之间。

三、MVSNet

本章节对所提出的MVSNet进行详细的描述，MVSNet的设计遵循了相机几何原理，并且借鉴了传统MVS方法的一些观点。在接下来的部分，将网络中的每个模块与传统的MVS方法进行对比，以此验证基于深度学习MVS方法的优势。MVSNet的网络架构如图1所示。

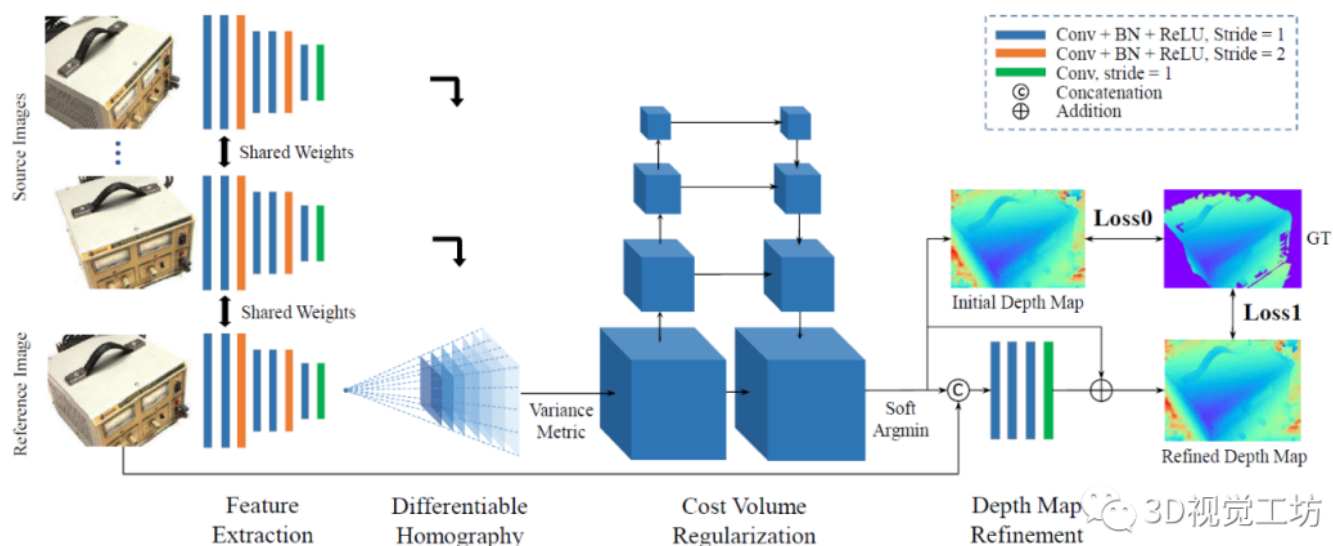


图 1 MVSNet网络架构图

3.1 图像特征提取

MVSNet 的第一步是在 N 张输入图像 $\{I_i\}_{i=1}^N$ 上进行特征提取，得到 N 个深度特征图 $\{F_i\}_{i=1}^N$ ，这些特征图用于后续的稠密匹配。特征提取网络使用的是 8 层的 2D 的卷积神经网络，其第 3 层、第 6 层的卷积步长设为 2，以此得到 3 个尺度的特征图。与一般的匹配网络相同，特征提取网络的参数共享权重。

2D卷积神经网络的输出为 N 个 32 通道的特征图。与原图相比，在每个维度上进行了 4 倍的下采样。值得注意的是，虽然特征图经过了下采样，但被保留下来像素的邻域信息已经被编码保存至 32 通道的特征描述子中，这些特征描述子为匹配提供了丰富的语义信息。与在原始图像上进行特征匹配对比，使用提取的特征图进行匹配显著提高了重建质量。

3.2 代价体

接下来是基于提取的特征图和相机参数构建3D代价体。之前的空间划分的方法是基于规整的网格，但是对于深度预测的任务而言，我们是基于参考相机视锥体进行代价体的构建。

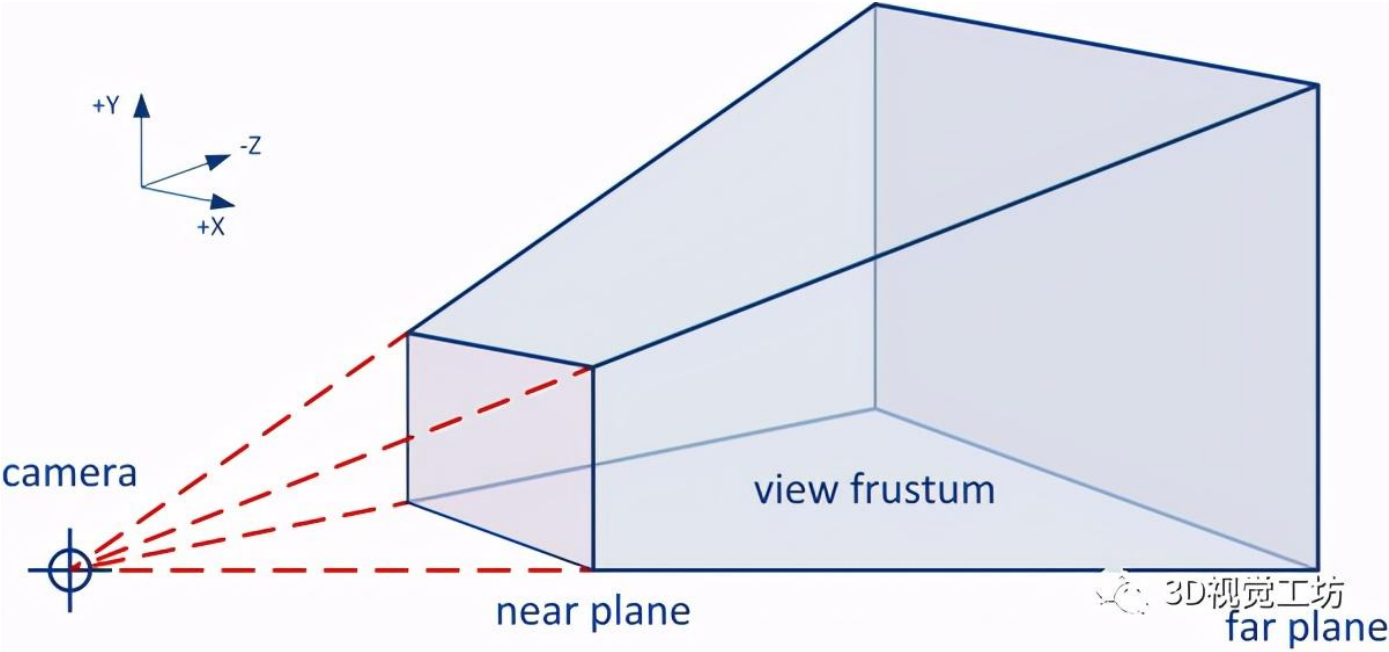


图 2 相机视锥体示意图

在后面的说明中，我们将参考图像记为 I_1 ，源图像记为 $\{I_i\}_{i=2}^N$ ， $\{K_i, R_i, t_i\}_{i=1}^N$ 为特征图对应的相机内参、旋转矩阵和平移向量。
3D视觉工坊

3.2.1 单应矩阵

单应矩阵（Homography） H 用于描述两个平面之间的映射关系。若场景中的特征点都落在同一平面上（比如墙，地面等），则可以通过单应矩阵来进行运动估计。

单应矩阵通常描述处于共同平面上的一些点，在两张图像之间的变换关系。考虑在图像 I_1 和 I_2 有一对匹配好的特征点 p_1 和 p_2 。这些特征点落在某一平面上。设平面满足的方程为：

$$n^T P + d = 0$$

即：

$$-\frac{n^T P}{d} = 1$$

可得：

$$\begin{aligned} p_2 &= K(RP + t) \\ &= K(RP + t(-\frac{n^T P}{d})) \\ &= K(R - \frac{tn^T}{d})P \\ &= K(R - \frac{tn^T}{d})K^{-1}p_1 \end{aligned}$$

于是我们得到一个直接描述图像坐标 p_1 和 p_2 之间的变换，将中间这部分记为H（单应矩阵），于是有

$$p_2 = Hp_1$$

单应矩阵是一个 3×3 的矩阵。

3.2.2 可微分的单应变换

N 张特征图 (feature maps) 投影至参考图像下的若干个平行平面上构成 N 个特征体 (feature volumes) $\{V_i\}_{i=1}^N$ 。平面单应变换决定了从特征图到位于深度值 d 处代价体的坐标变换。 $x' \sim H_i(d) \cdot x$ ，其中 \sim 表示投影等式， $H_i(d)$ 表示在深度值 d 处 F_i ($i = 1, 2, \dots, N$) 与 F_1 的单应变换矩阵。 n_1 表示为参考相机的主轴方向，单应矩阵可表示为 3×3 的矩阵：

$$H_i(d) = K_i \cdot R_i \cdot \left(I - \frac{(t_1 - t_i) n_1^T}{d} \right) \cdot R_1^T \cdot K_1^T$$

单应变换投影的过程类似于经典的平面扫描算法，唯一的区别在于采样点来自特征图 $\{F_i\}_{i=1}^N$ 而不是图像 $\{I_i\}_{i=1}^N$ 。

3D视觉工坊

3.2.3 代价指标 (Cost Metric)

接下来我们将多个特征体 $\{V_i\}_{i=1}^N$ 聚合为一个代价体 C 。为了适应任意数目的视图输入，MVSNet 使用了一个基于方差的代价指标 M ，该指标用于衡量 N 张视图间的相似性。 W, H, D, F 分别代表输入图像的宽度，深度，采样数和特征图的通道数，故而特征体的尺寸为： $V = \frac{W}{4} \cdot \frac{H}{4} \cdot D \cdot F$ 。我们的代价指标定义了映射关系 M ：

$$C = M(V_1, \dots, V_N) = \frac{\sum_{i=1}^N (V_i - \bar{V}_i)^2}{N}$$

3D视觉工坊

3.2.4 代价体正则化

从图像特征图计算得到的初始代价体很有可能是包含噪声的（由于非朗伯体表面的存在或者视线遮挡的问题），为了预测深度图还需要进行光滑处理。正则化的步骤是对上面的代价体进行优化(refine)以得到概率体。受SurfaceNet等工作的启发，我们使用一个多尺度的3D-CNN网络用于代价体正则化。4个尺度的网络类似于3D版本的UNet，使用编码-解码的结构方式，以相对较小的存储/计算代价在一个大的感受野范围内进行

邻域信息聚合。为了减轻网络的计算代价，在第一个3D卷积层后，我们将32通道的代价体缩减为8通道，将每个尺度的卷积从3层降为2层。最后卷积层的输出为1通道的体

(Volume)。最终在深度方向上使用 $softmax$ 操作进行概率值的归一化。

产生的**概率体**可很好适合于深度值预测，不仅可以用于逐像素的深度预测，还可以用来衡量估计的置信度。

3.3 深度图

3.3.1 初始深度估计

从**概率体**中恢复深度图最简单的方式是采用“赢者通吃”的策略（比如 $argmax$ ）。但是 $argmax$ 操作不能够产生亚像素级的估计，而且由于不可微分不能在网络中进行反向传播。所以我们采用在深度方向上计算期望，即所有假设深度值的加权和：

$$D = \sum_{d=d_{min}}^{d_{max}} d \times P(d)$$

其中 $P(d)$ 为在深度值 d 处的所估计的概率值。该操作在文献中被称为 $softmax$ 操作，该操作可微分并且能够产生 $argmax$ 操作的结果。在代价体构建的过程中深度假设值是在范围 $[d_{min}, d_{max}]$ 内均匀采样得到的，所以预测的深度值是连续的。输出深度图的尺寸与2D特征图相同，即在每个维度上是原始图像的1/4。

3.3.2 概率图

沿深度方向上的概率分布还可以反映了深度估计的质量。尽管多尺度的3D CNN有一定的正则化能力，可以将概率分布调整为单峰分布，我们也应注意到由于错误匹配的出现，使得深度的概率分布散乱，不能集中在一个峰值处。如图3所示，参考视图(a)的概率概率率为图(d)。

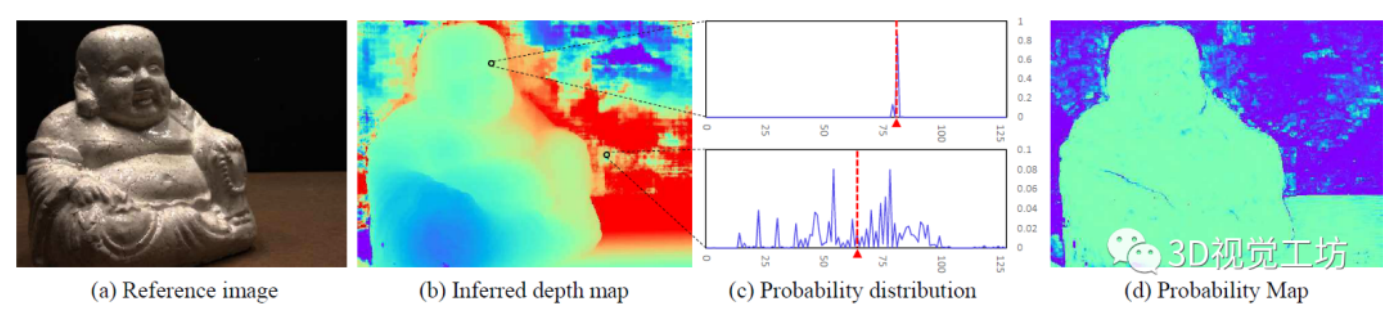


图 3 参考图像、预测的深度图、概率分布、概率图

因为深度假设在相机视锥体内离散采样得到的，简单地将四个最近深度假设的概率求和得到最后的估计质量。注意其他的统计指标（比如标准差或熵）也可以用在里，但在我们的实验中我们观察到这些指标对于后续的深度图过滤没有显著的提升。除此之外，我们的概率和公式可以在后续的外点剔除过程中起到阈值参数的作用。

3.3.3 深度图优化

尽管从**概率体** 恢复出深度图是一个直接而合理的过程，但还是会存在一些问题。由于正则化过程中较大的感受野造成重建深度图边界过平滑的现象，这些问题也是在语义分割和图像抠图（Image Matting）中存在的。

我们注意到参考图像是包含图像边界信息的，因此可以将参考图像作为引导来优化预测的深度图。受到图像抠图算法的启发，我们在MVSNet的末端添加了一个深度残差学习的网络。

预测的深度图和被resized的参考图像被连接在一起构成一个4通道的输入，被送入三个32通道的2D卷积层和一个1通道的卷积层，来学习得到深度的残差值。为了可以学习到负的残差，最后一层网络没有使用BN层和ReLU单元。同样地，为避免在某一特定深度值上存在偏差，将初始深度幅值缩放至范围[0, 1]内，在深度图优化完成之后转换回来。

3.4 损失函数

初始深度值和优化深度值的损失都被考虑在内了。我们使用真实深度图与所估计深度的L1损失作为我们的训练损失。考虑到真实深度图并不是每个像素点都有值的，因此我们只考虑那些有效的像素点，即存在Ground Truth标签的像素。

损失函数定义如下：

$$Loss = \sum_{p \in \mathbf{p}_{valid}} \underbrace{\|d(p) - \hat{d}_i(p)\|_1}_{Loss0} + \lambda \cdot \underbrace{\|d(p) - \hat{d}_r(p)\|_1}_{Loss1}$$

四、网络的实现

4.1 训练

4.1.1 数据准备

目前MVS数据集提供的Ground Truth数据是以点云或三角面片(mesh)的格式呈现的, 所以我们需要自己使用这些数据产生深度图。DTU数据集包含了128个场景, 数据集提供了带有法向信息的点云。我们使用泊松表面重建获得三角网格表面, 将三维网格投影至每个视点以得到我们需要的深度图。

为获取更好质量的三维网格, 我们将泊松表面重建中树的深度(*depth-of-tree*) 设为11; 同时, 为减轻表面边缘处的人工痕迹, 我们将修剪因子(*trimming-factor*) 设为9.5。为了公平地与其他深度学习的方法进行对比, DTU数据集的划分同SurfaceNet一样。

根据文献[7], 将DTU数据集划分训练集、验证集和测试集。验证集中场景序列号包括{3, 5, 17, 21, 28, 35, 37, 38, 40, 43, 56, 59, 66, 67, 82, 86, 106, 117}; 测试集中场景序列号包括{1, 4, 9, 10, 11, 12, 13, 15, 23, 24, 29, 32, 33, 34, 48, 49, 65, 75, 77, 110, 114, 118}; 训练集为其余79个场景。

训练集包括79个场景, 每个场景包含49张图, 有7组不同光照的图。将每张图作为参考图像, DTU数据集总共有27097个训练样本。

4.1.2 视图选择

在训练中我们使用 3 张视图，即一张参考图像和两张源图像。根据稀疏点，我们为每一图像对计算一个评分 $s(i,j) = \sum_p G(\theta_{ij}(\mathbf{p}))$ ， \mathbf{p} 为视图 i 与视图 j 的特征点对。式中

$$\theta_{ij}(\mathbf{p}) = \frac{180}{\pi} \arccos((\mathbf{c}_i - \mathbf{p}) \cdot (\mathbf{c}_j - \mathbf{p}))$$

为 \mathbf{p} 的基线夹角， \mathbf{c} 为相机光心。 G 是以 θ_0 为中心的分段高斯函数：

$$G(\theta) = \begin{cases} \exp\left(-\frac{(\theta - \theta_0)^2}{2\sigma_1^2}\right), & \theta \leq \theta_0 \\ \exp\left(-\frac{(\theta - \theta_0)^2}{2\sigma_2^2}\right), & \theta > \theta_0 \end{cases}$$

实验过程中，将 θ_0 设为 5， σ_1 与 σ_2 分别设为 1 和 10。

 3D视觉工坊


注意到图像尺寸在特征提取阶段会被下采样，再加上3D正则化部分的4层编码-解码结构，输入的图像尺寸必须能被32整除。同时考虑到GPU显存的限制，我们将图像分辨率从1600×1200降至800×600，然后在图像中心处截取640×480大小的图像块作为网络训练的输入。输入的相机参数也随之改变。深度假设在425mm~935mm范围内以2mm的分辨率均匀采样（采样数D=256）。

4.2 后处理

4.2.1 深度图过滤

经过上述步骤，我们为参考图像的每个像素估计一个深度值。在将预测结果转为稠密点云之前，非常有必要将背景区域和遮挡区域进行过滤。我们使用光度一致性和几何一致性来进行鲁棒的深度图过滤。

光度一致性衡量了匹配质量，如 3.3 节所提及的，我们用概率图来衡量深度估计的质量。在实验中，将概率值低于 0.8 的点视为外点。几何一致性用于衡量多视图之间的深度连续性。与双目立体匹配中左右视差校验相似，将参考图像像素点 p_1 处的深度值 d_1 投影至邻域视图的 p_i 点处，然后将 p_i 处的深度值 d_i 重投影至参考图像 p_{reproj} ，其深度值 d_{reproj} ，如果满足 $|p_{reproj} - p_1| < 1$ 并且 $|d_{reproj} - d_1|/d_1 < 0.01$ ，我们称像素 p_1 处的深度估计值 d_1 是两视图连续的。在实验中，所有的深度值至少是三视图连续的。简单的两步过滤策略在滤除不同类型的外点中显现出很强的鲁棒性。

 3D视觉工坊

4.2.2 深度图融合

与其他多视图立体重建方法类似，我们运用深度融合的方式将多视图的深度图融合为三维空间中的点云。在重建过程中使用基于可见性的融合，使得不同视图间的深度遮挡最小化。为进一步抑制重建噪声，在过滤步骤中，为每个像素计算可见视图，将所有重投影深度值的平均值 $\overline{d_{reproj}}$ 作为像素的最终深度值估计。融合的深度图被反投影回 3D 空间得到最后的点云。

 3D视觉工坊

五、实验结果

基于深度学习的多视图立体重建一般由两部分组成，深度图重建、深度图融合为点云。

5.1 DTU数据集评测

在DTU数据集中22个测试场景进行评测，输入视图数，图像宽度，图像高度和深度采样数分别设置为N=5，W=1600，H=1184，D=256。基于DTU提供的MATLAB脚本对重建点云的精度 (*accuracy*) 和完整度 (*completeness*) 进行评测。同时也定义了一个“总分”的评价指标用于衡量重建的质量，即精度和完整度的平均数。实验结果的定量结果如表1所示。点云的重建结果如图4所示。

	Mean Distance (mm)			Percentage (<1mm)			Percentage (<2mm)		
	Acc.	Comp.	overall	Acc.	Comp.	f-score	Acc.	Comp.	f-score
Camp [3]	0.835	0.554	0.695	71.75	64.94	66.31	84.83	67.82	73.02
Furu [7]	0.613	0.941	0.777	69.55	61.52	63.26	78.99	67.88	70.93
Tola [35]	0.342	1.190	0.766	90.49	57.83	68.07	93.94	63.88	73.61
Gipuma [8]	0.283	0.873	0.578	94.65	59.93	70.64	96.42	63.81	74.16
SurfaceNet[14]	0.450	1.04	0.745	83.8	63.38	69.95	87.15	67.99	74.4
MVSNet (Ours)	0.396	0.527	0.462	86.46	71.13	75.69	91.06	75.31	80.25

表 1 DTU数据集实验结果

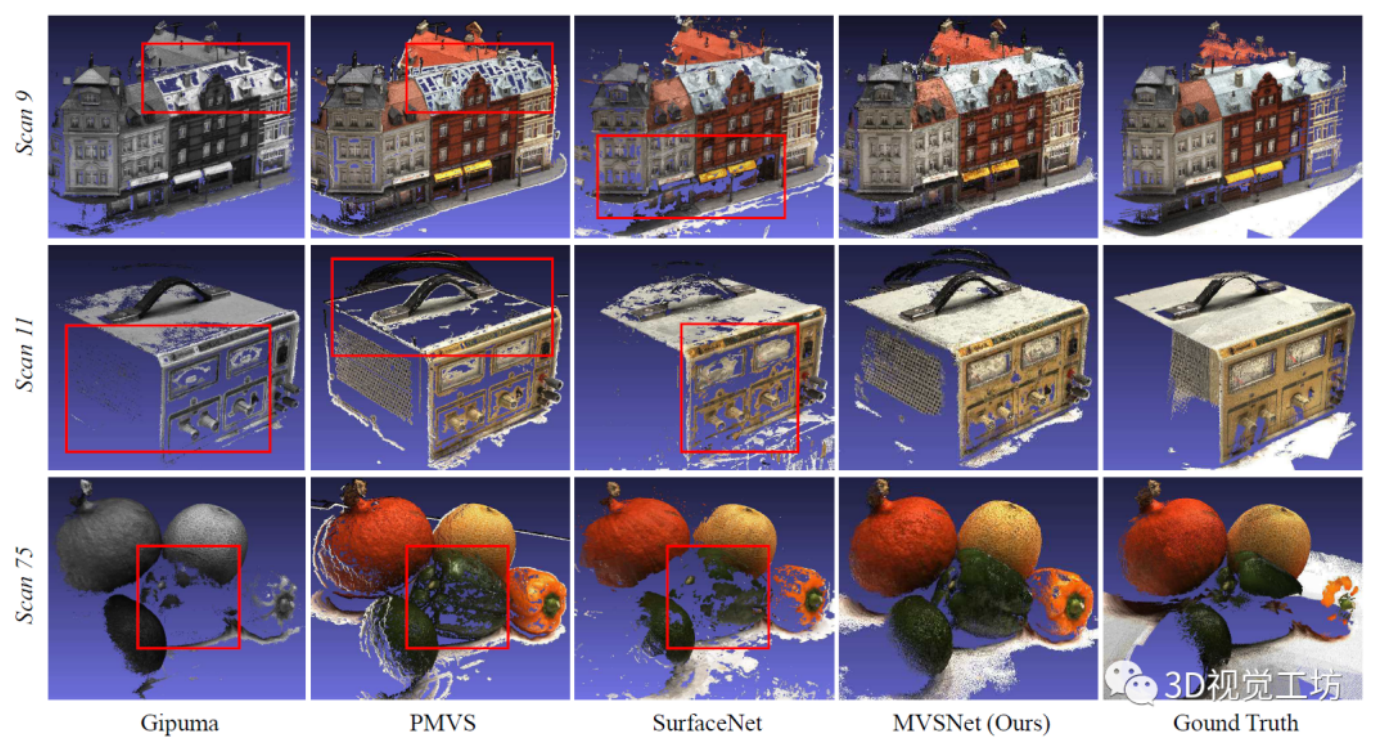


图4 点云重建对比图

5.2 消融分析 (Ablations Study)

本节对MVSNet网络中的若干组成部分进行分析，在接下来的实验中，我们使用验证集的深度值损失来衡量重建质量。

5.2.1 视图数量

首先研究输入视图数N对重建结果的影响，此实验也同时说明了模型可以适用于任意视图的输入。如4.1小节说明的，训练时视图数N=3，本实验测试时模型视图使用了N=2, 3, 5。结果如图5所示。

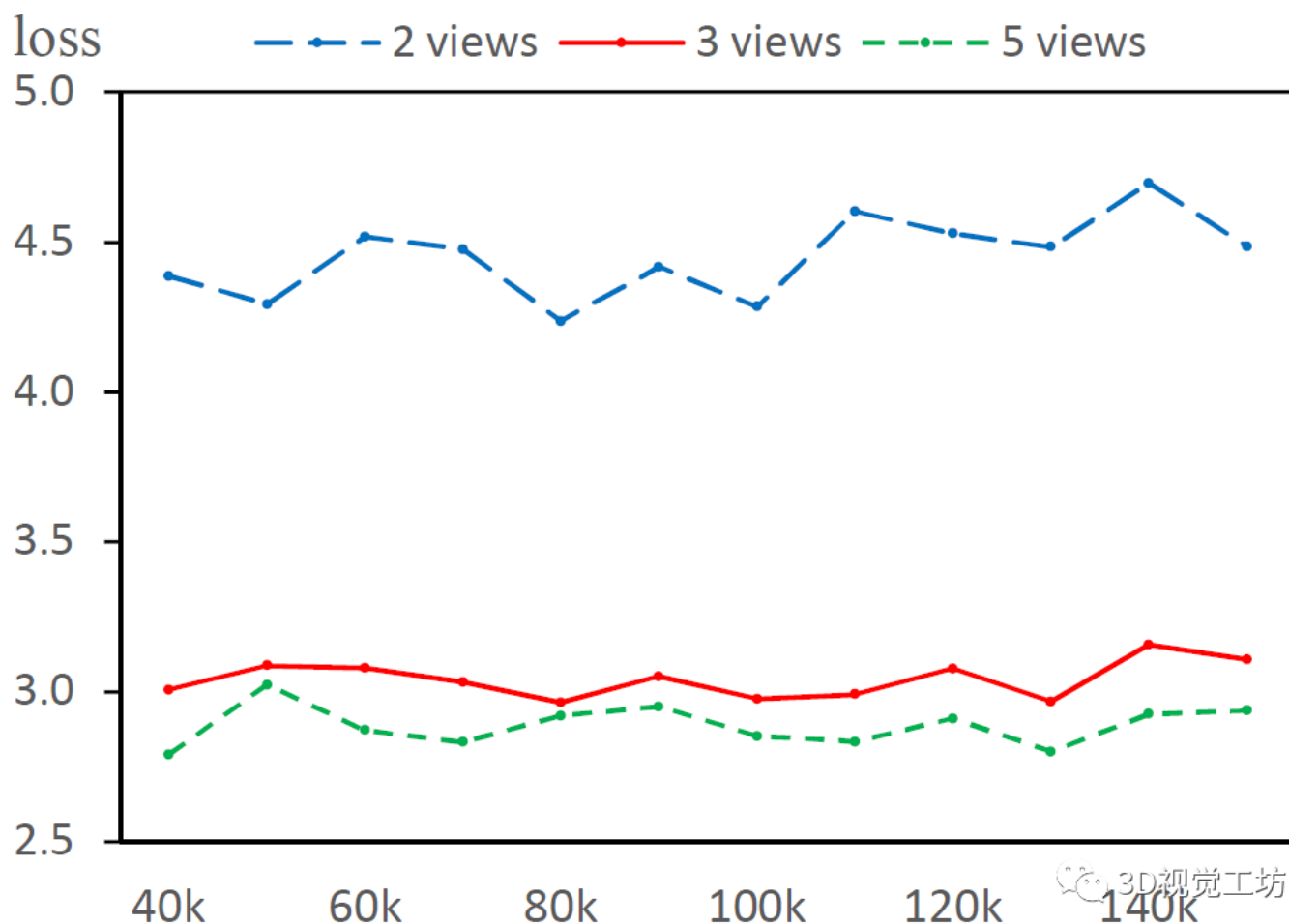


图 5 视图数对验证损失的影响

5.2.2 图像特征

我们研究发现，基于深度学习的特征可以显著提升MVS重建质量，为了在MVSNet建模出传统的基于图像块的图像特征，我们将网络中的特征提取部分替换为一32通道的卷积层。卷积核的大小设定为 7×7 ，步长为4。由图6可知，2D特征提取的网络在验证损失上比单层网络效果更好。

5.2.3 代价指标

我们还比较了基于平均的代价指标和基于方差的代价指标。如图6所示，逐元素的方差操作可以达到更快的收敛速度，同时也验证了使用显式差异指标去计算多视图间的相似性是合理的。

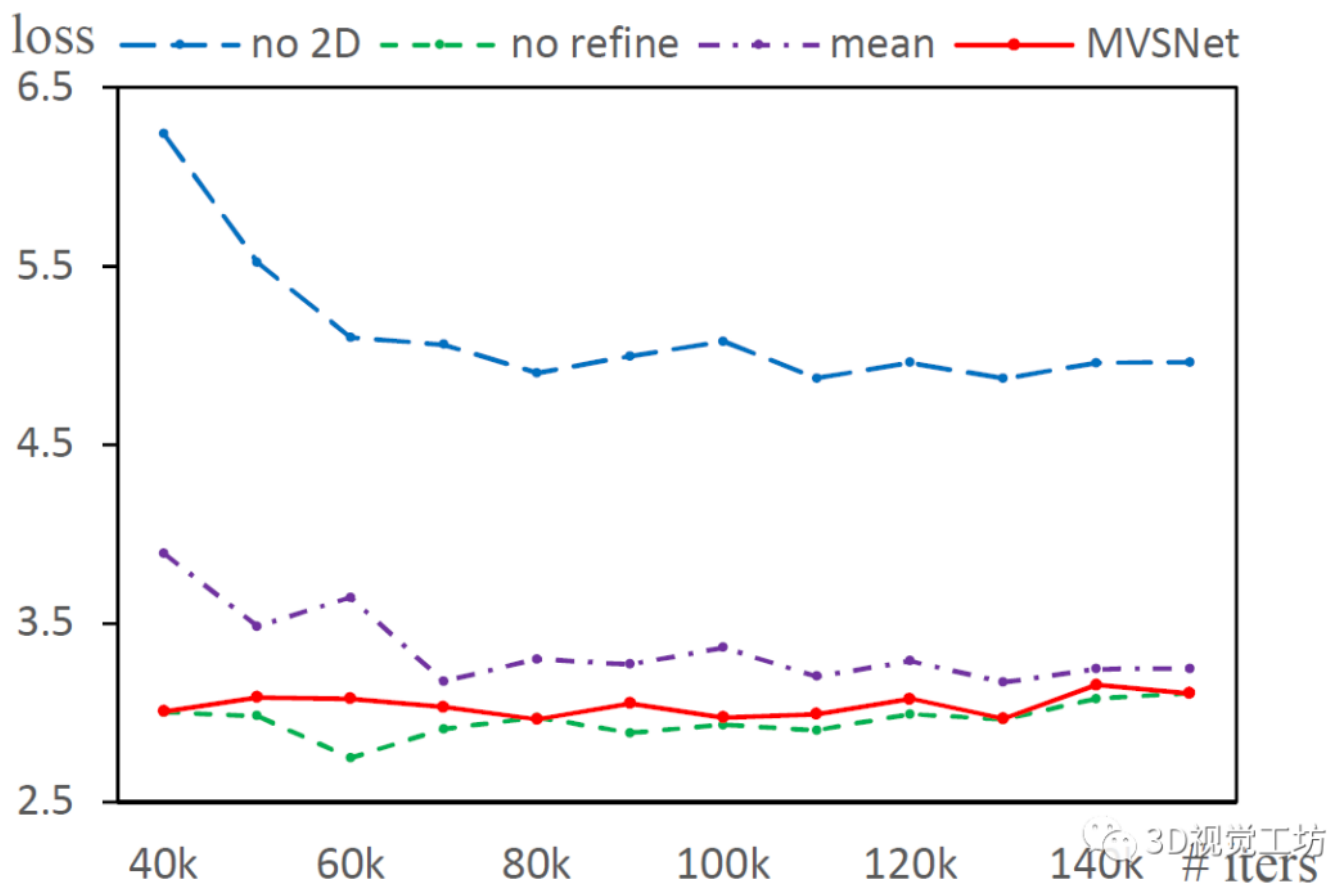


图6 对2D图像特征、代价指标和深度优化的消融分析图

5.2.4 深度图优化

对有深度图优化、没有深度图优化的MVSNet两种情况进行训练，深度图优化基本上对验证集损失没有影响，但是深度图优化步骤可以提高点云重建结果。在小于1mm条件下 f -score从75.58提高为75.69；2mm条件下 f -score从79.98提高为80.25。

5.3 部分深度图重建结果

部分场景的重建深度图如图7~10所示。在每一幅图中，左子图为参考图像，右子图为重建的深度图，深度图以彩虹图的形式呈现。

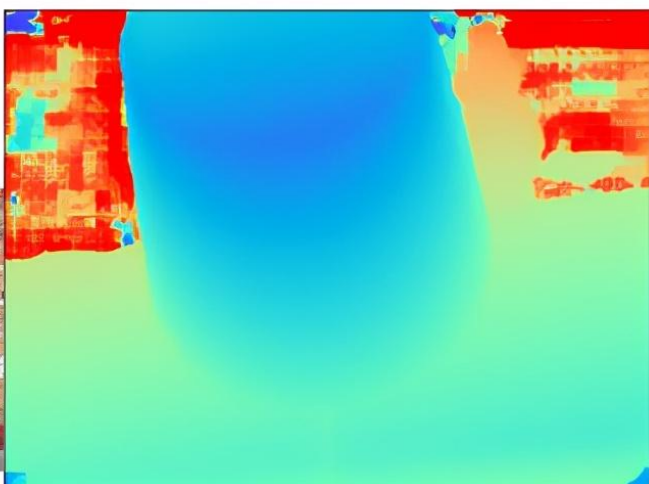


图 7 DTU 数据集 Scan1 重建深度图

3D视觉工坊

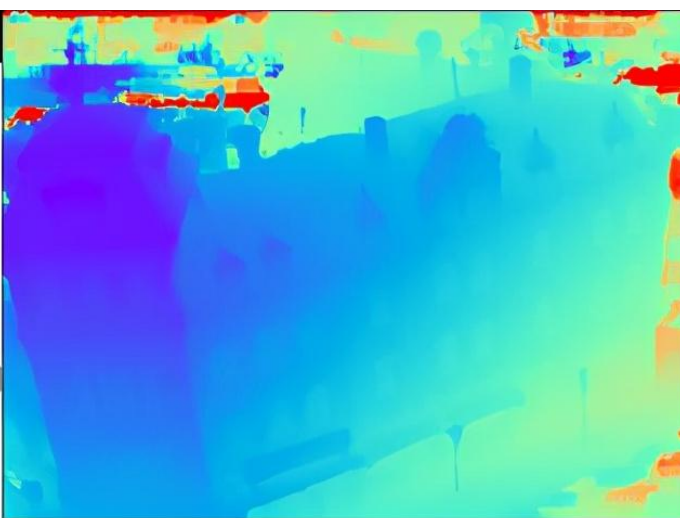


图 8 DTU 数据集 Scan9 重建深度图

3D视觉工坊

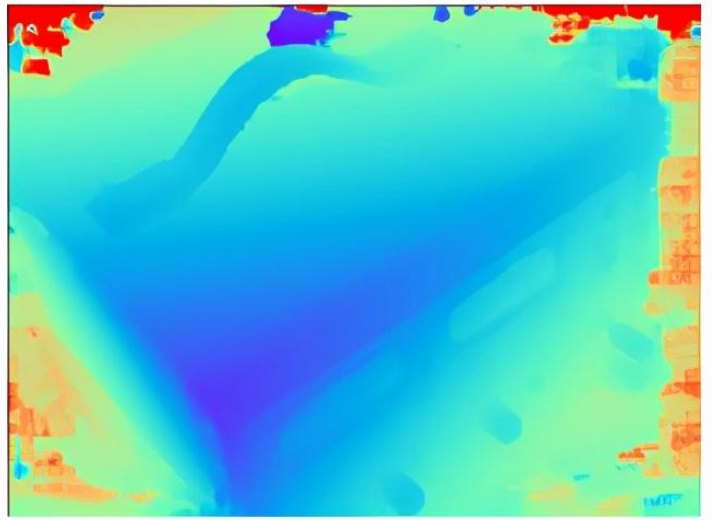


图 9 DTU 数据集 Scan11 重建深度图

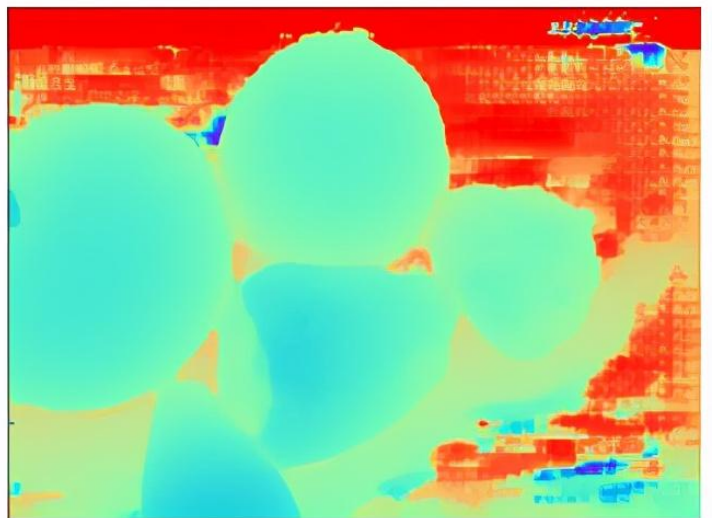


图 10 DTU 数据集 Scan75 重建深度图

3D视觉工坊

参考文献

1. Furukawa Y, Hernández C. Multi-view stereo: A tutorial[J]. Foundations and Trends® in Computer Graphics and Vision, 2015, 9(1-2): 1-148.

- 2.Yao Y, Luo Z, Li S, et al. Mvsnet: Depth inference for unstructured multi-view stereo[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 767-783.
- 3.Yao Y, Luo Z, Li S, et al. Recurrent mvsnet for high-resolution multi-view stereo depth inference[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 5525-5534.
- 4.Cui Z, Gu J, Shi B, et al. Polarimetric multi-view stereo[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 1558-1567.
- 5.Schönberger J L, Zheng E, Frahm J M, et al. Pixelwise view selection for unstructured multi-view stereo[C]//European Conference on Computer Vision. Springer, Cham, 2016: 501-518.
- 6.Xu Q, Tao W. Multi-Scale Geometric Consistency Guided Multi-View Stereo[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 5483-5492.

7.Ji M, Gall J, Zheng H, et al. Surfacenet: An end-to-end 3d neural network for multiview stereopsis[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 2307-2315.

8.王跃嵩. 多视图立体匹配三维重建方法[D].华中科技大学,2015.

9.Kendall A, Martirosyan H, Dasgupta S, et al. End-to-end learning of geometry and context for deep stereo regression[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 66-75.

10.Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234-241.

更多干货

欢迎加入【3D视觉工坊】交流群，方向涉及3D视觉、计算机视觉、深度学习、vSLAM、激光SLAM、立体视觉、自动驾驶、点云处理、三维重建、多视图几何、结构光、多传感器融合、VR/AR、学术交流、求职交流等。工坊致力于干货输出，为3D领域贡献自己的力量！欢迎大家一起交流成长~

添加小助手微信：CV_LAB，备注学校/公司+姓名+研究方向即可加入工坊一起学习进步。