**三维动画与交互技术期末作业**

22451034 刘洋

我的期末作业是关于三维重建技术的读书报告。实时三维重建技术主要强调如何从图像、深度、激光扫描等传感数据中恢复出三维模型。该项技术广泛应用于自动驾驶、机器人交互、VR、AR、游戏、电影等领域。按照输入图像的类型可分为单目三维重建和多视角三维重建。按照三维模型的表现形式可以划分为基于体素的三维重建、基于点云的三维重建和基于网格的三维重建。按照目标类型可以划分为针对单个物体的重建、针对静态或者动态的三维场景的重建[2]。

单目三维重建仅使用单一视角的图像，其重要的方法包括单目深度估计，在自动驾驶领域应用广泛，代表技术有Monodepth2[1]。Monodepth2 是一种改进的自监督单目深度估计方法，其使用了一个全连接的U-Net来预测深度，另一个单独的神经网络来预测相邻帧之间相机的位置变化，如下图1所示。

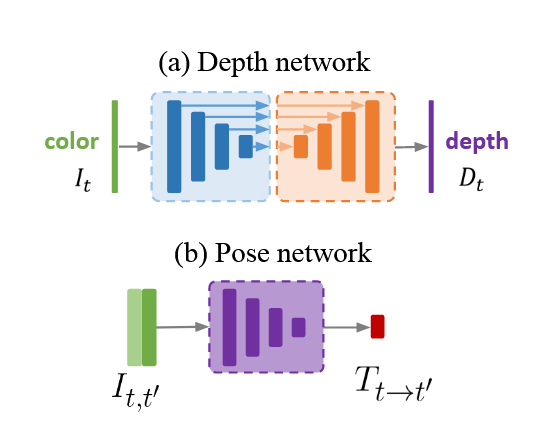


图1 Moonodepth2的神经网络

Monodepth2还提出了每个像素的最小重投影作为损失函数。先前的方法是使用每个像素在不同连续图像间的平均重投影误差作为损失函数，Monodepth2使用最小的重投影误差作为损失函数使得最终的深度图中的物体边缘变得不再模糊。如下图2和图3所示。

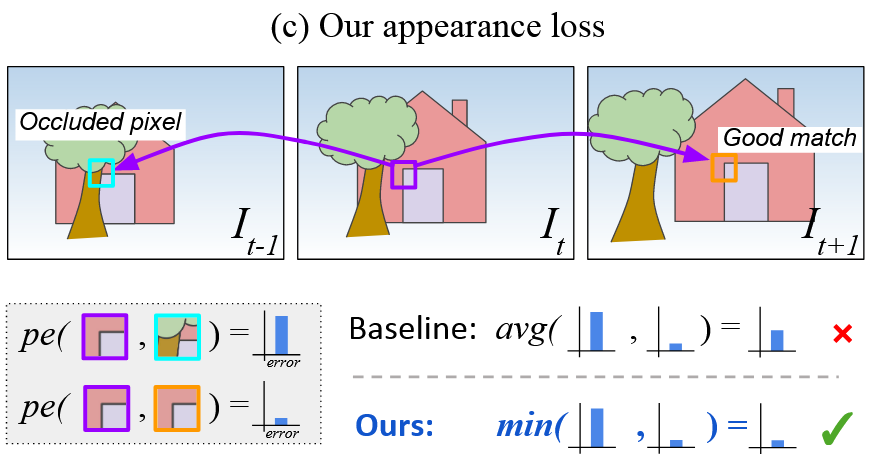


图2 Monodepth2使用最小重投影做为损失函数

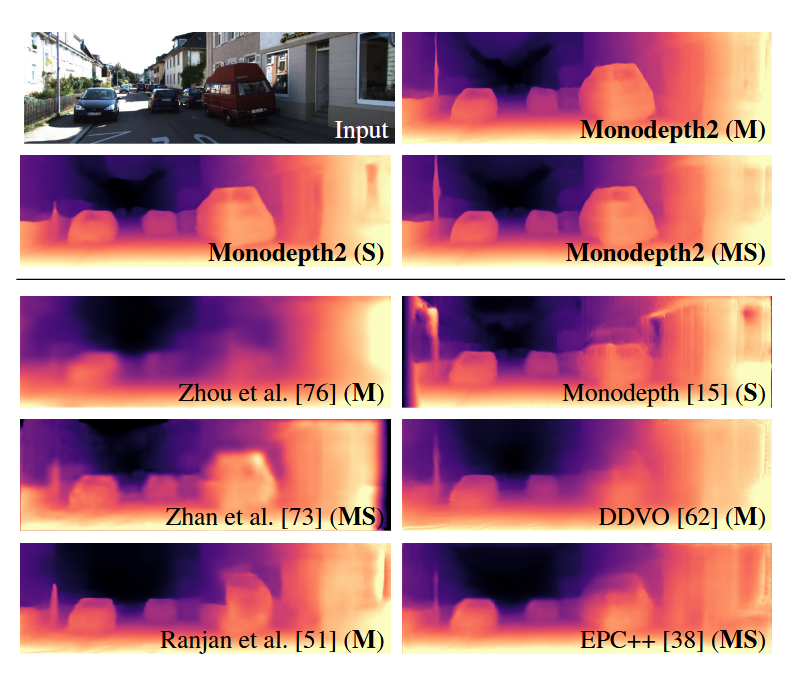


图3 Monodepth2的深度图中有清晰的物体边缘

此外，他们还提出了一种提升到全分辨率的计算损失的方法。现有方法会将原图像在只在低分辨率下进行深度估计和计算误差，而他们会上采样到全分辨率，这样保证利用到了高分辨率的图像信息。如下图4所示。

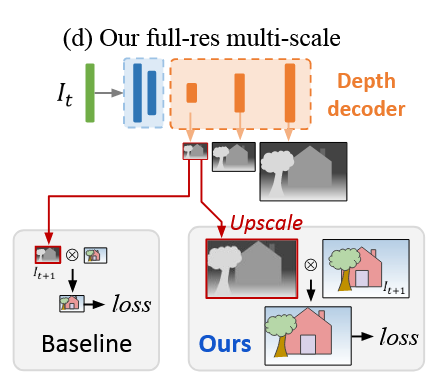


图4 Monodepth2的全分辨率采样

单目三维重建的常用的数据集包括KITTI[3]、TUM RGB-D[4]、NYU Depth Dataset V2[5]。

多目三维重建可以会使用不同角度的图像推算出物体的深度形状位置信息，构建出三维模型。Cheng等人提出了一种单目和多目融合的深度估计系统[6]。他们的系统中主要包含三个部分：单目分支、多目分支和自适应融合模块。如下图5所示。

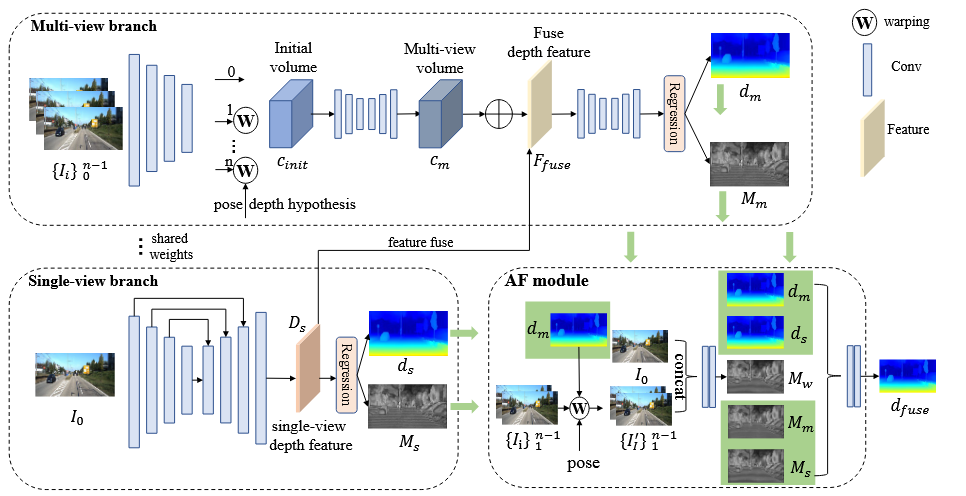


图5 Cheng等人的单目多目融合深度估计系统

他们的单目和多目分支共享了相同的骨干网络和权重来提取图像特征，单目的分支会得到图像的深度特征，多目的分支会融合多个图像包括单目的深度特征，最终经过回归计算后得到单目和多目的深度预测以及置信度Map。最终，融合模块会选择最准确的作为最终深度。在评估最准确深度时使用了三个置信度Map，其中两个分别是单目和多目的结果，还有一个置信度Map引入了不同的相机位姿和多目深度估计来计算，可用于判断多目深度结果的可靠性。因此，他们工作在不同的相机姿态下具有鲁棒性。

3DFIRES这篇文章也是多目的三维重建，与前两篇论文的不同之处在于其侧重于构建场景而不是深度估计[7]。他们的主要贡献在于从少量的且已知姿态的图像中进行场景级别的三维重建，在构建出其几何结构时还考虑了物体的隐藏面，如下图6所示，其中着色部分即为隐藏面。

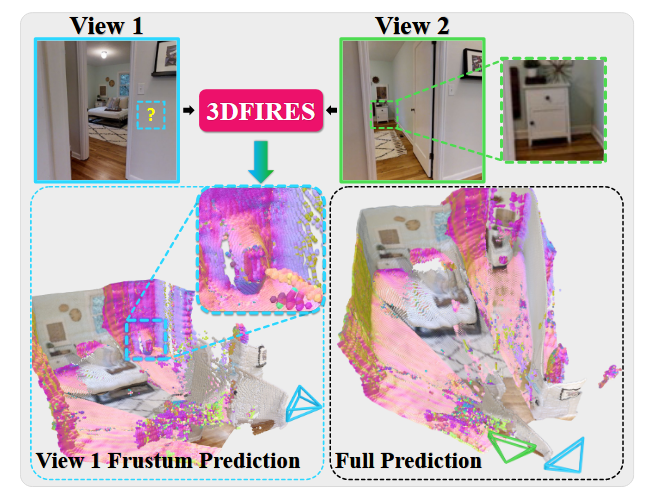


图6 3DFIRES的三维重建

简单来说，3DFIRES利用了DRDF（Directed Ray Distance Function，定向射线距离函数）和基于注意力机制的预测DRDF模型，学习到了一些场景的先验知识和几何关系，从而去合理的推断隐藏面的位置和形状吗，如下图7所示。DRDF是先前的一种用于三维场景重建的隐式表示方法，通过测量点到最近表面沿射线的距离来描述场景的几何结构，这是一种隐式的表示，与网格和点云的显示表示方法不同，不需要直接存储物体表面的几何信息。这个函数接受定点和射线方向作为输入，就可以输出距离当前定点最近表面的距离信息，适用于多目的融合。

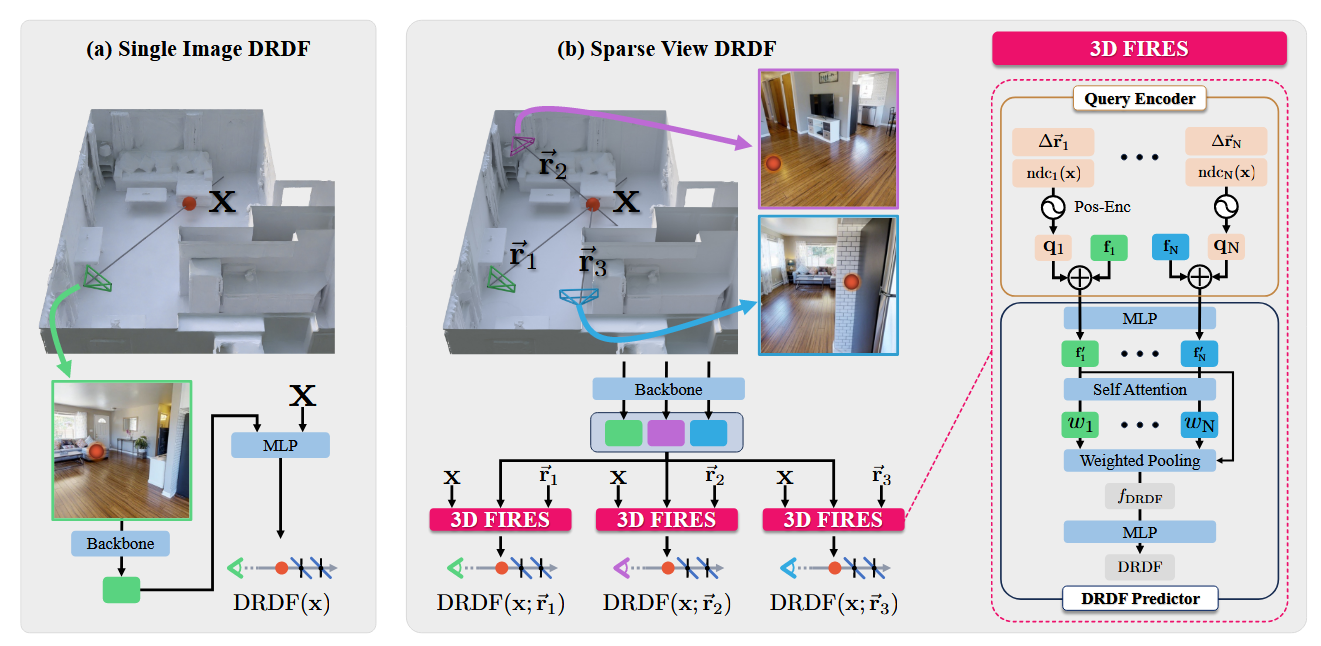


图7 3DFIRES的框架

Xu等人提出了一种贝叶斯扩散模型的三维重建方法，侧重于三维模型的点云生成，与Cheng等人文章的思想相同之处在于，他们也强调了融合不同方法来增强最后的结果[8]。他们的方法结合自上而下的先验信息和自下而上的数据驱动过程进行贝叶斯推断，如下图8所示，其中红色的部分就是自上而下，蓝色部分就是自下而上。所谓自上而下就是从模型的整体到局部，从先验的全局信息出发逐步细化到具体细节的过程，这篇文章中主要利用模型的形状先验信息知道自上而下的重建。所谓自下而上就是从局部到整体，这篇文章利用率输入图像中的局部特征作为数据驱动，逐步构建出三维模型。

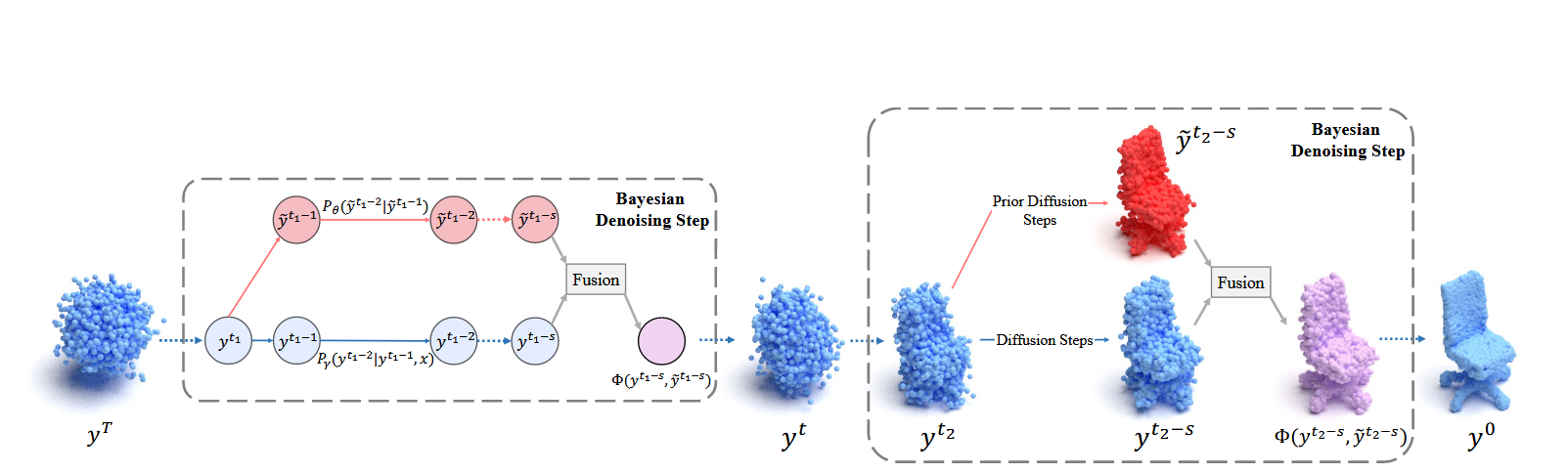


图8 Xu等人的贝叶斯扩散三维重建过程

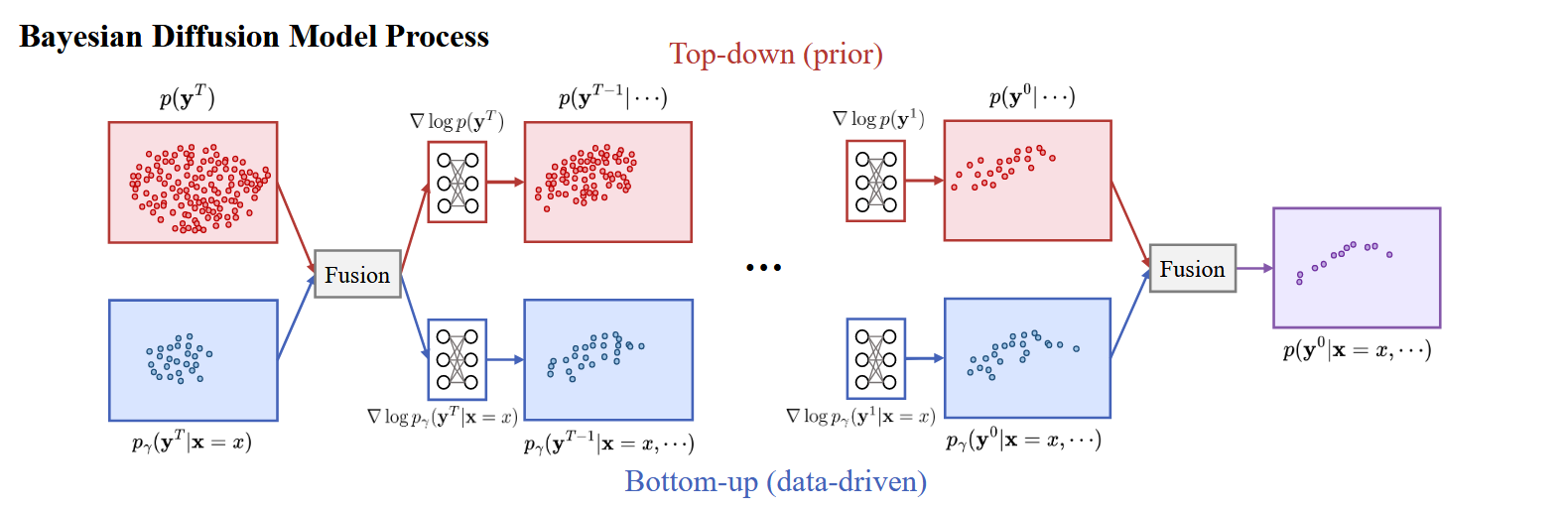


图9 Xu等人贝叶斯扩散模型中的自上而下和自下而上

此外，他们还提出了两种信息融合策略BDM-M和BDM-B，进而有效整合先验模型和重建模型的信息。BDM-M强调使用两个扩散模型，一个用于先验模型，一个用于重建模型，然后再模型的解码器层面进行特征融合输出。BDM-B强调在点云层面进行融合，根据点云中点的分配概率来进行选择。如下图10所示。

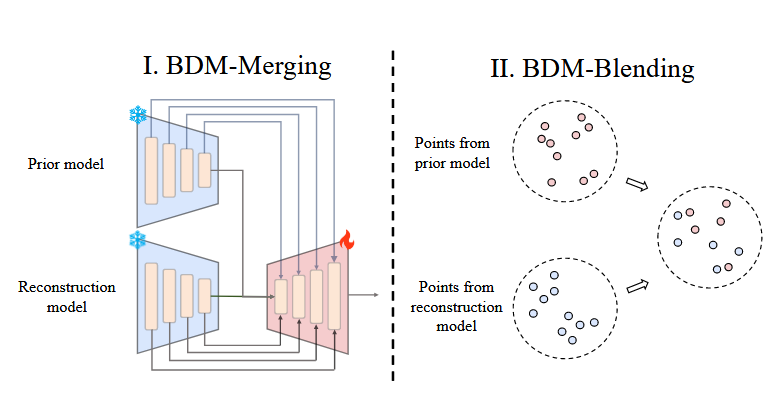


图10 Xu等人的信息融合策略

UFORecon这篇文章于3DFIRES类似，同样专注于从有限的多目图像中进行看不见表面的三维重建[9]。同时，与Cheng等人的文章类似，他们也强调了不利于重建的相机位姿干扰，如下图11所示。

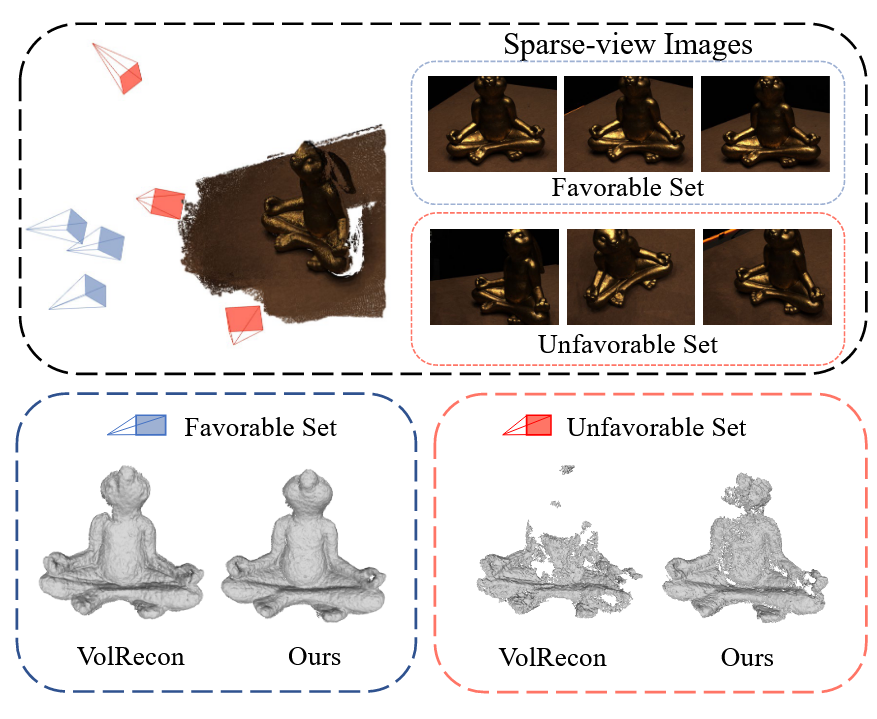


图11 UFORecon使用不同的稀疏相机视角组合进行重建

UFORecon的关键组件包括：跨视图的特征匹配Transformer、3D相机视锥相关性提取和2D几何感知相似性特征编码、三维重建Transformer，如下图12所示。首先，使用特征金字塔网络分别提取各个视角图像的多尺度特征，再通过跨视图的特征匹配Transformer处理特征生成成对的匹配特征，为后续的重建过程提供了视图间的相关信息，增强了模型对不同视图组合的适应能力。然后，3D相机视锥相关性，编码了不同视图之间的全局几何关系；2D的几何感知相似性为模型重建提供了图像上的先验信息。最后，三维重建Transformer融合这些信息生成用于渲染的体积表示。

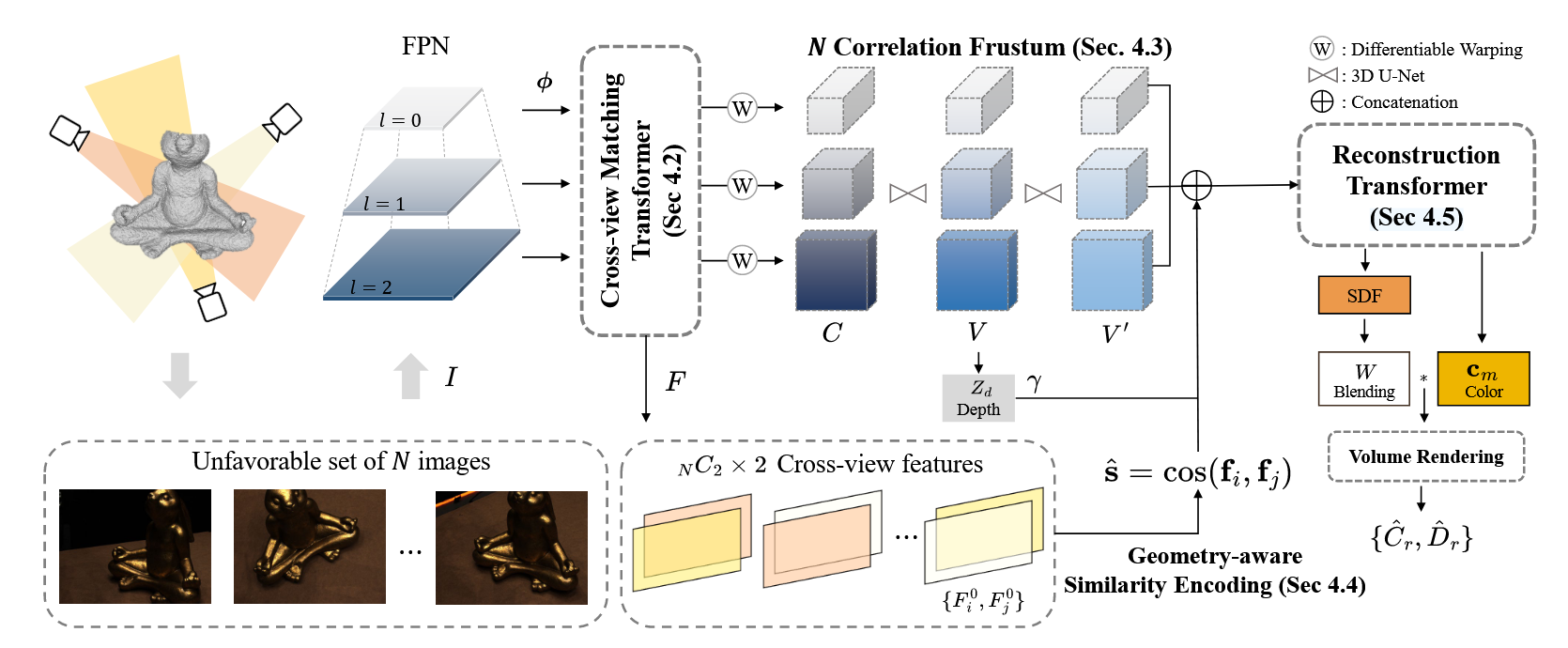


图12 UFORecon的pipeline

G3DR这篇文章则提出了在文字或者单目情况下去生成三维物体，而不是仅仅依赖于现实中的物体和场景，如下图13所示[10]。

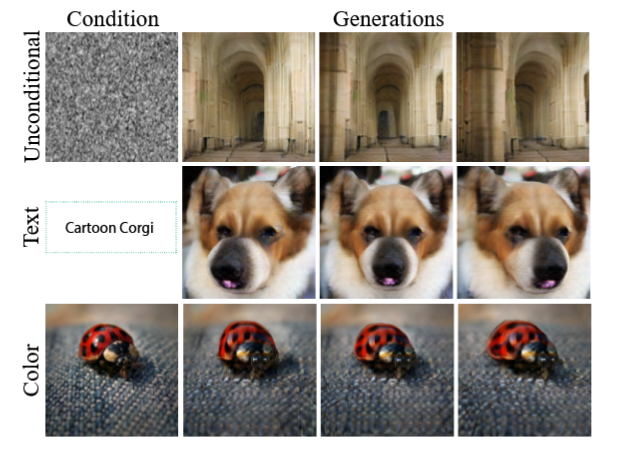


图13 G3DR生成三维物体

G3DR的框架如下图14所示。在该框架下，他们利用了一个潜在的扩散模型进行来生成带有二维图像和深度信息。然后，他们进一步将其转化为三维表示，将其特征映射到三平面上，通过体积渲染生成三维场景。他们提出了深度正则化技术来优化三平面生成的训练过程，防止出现体积崩溃，确保生成的几何形状和深度图的一致性。他们还采用了多分辨率采样来提高生成图像的质量。

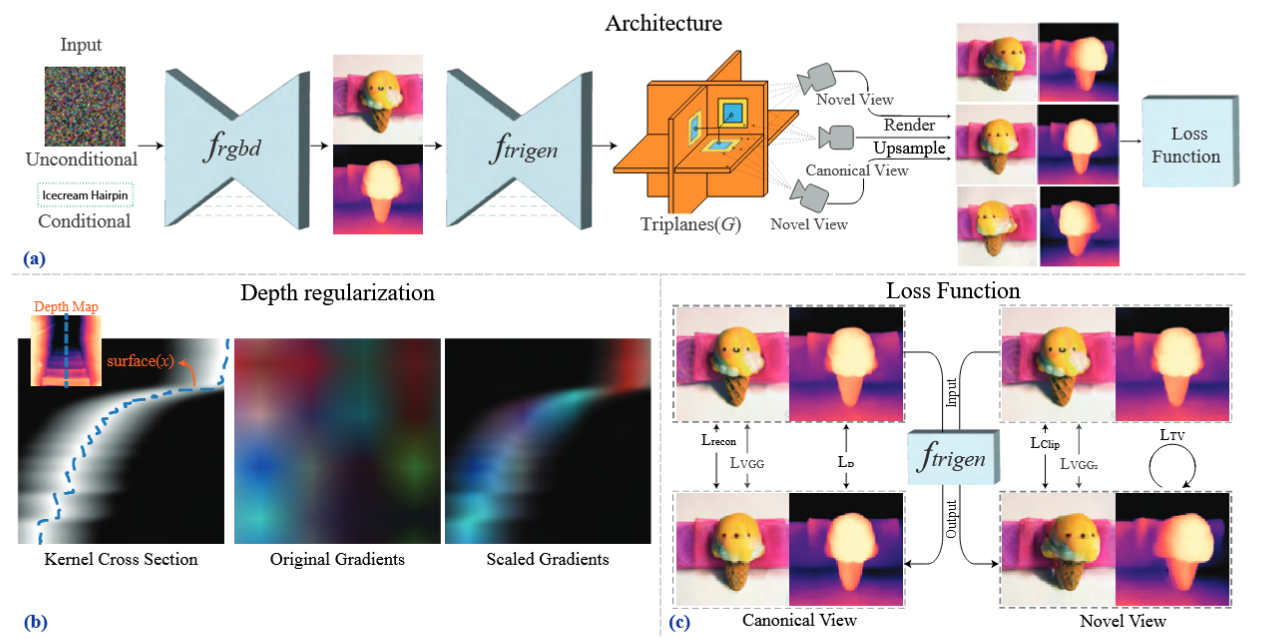


图14 G3DR的框架

三维重建还有一个比较关键的技术是高斯溅射，最早在1999年提出，其将每个点云数据视为离散的高斯核，将点云数据的信息更光滑地传播到图像空间中，从而产生真实的三维外观。

DNGuassian这篇文章就是基于3D高斯溅射场，通过深度正则化来恢复准确的场景几何结构，同时保持了三维物体的精细的颜色外观。他们提出了硬和软深度正则化和全局与局部的深度归一化技术，框架如下图15所示。

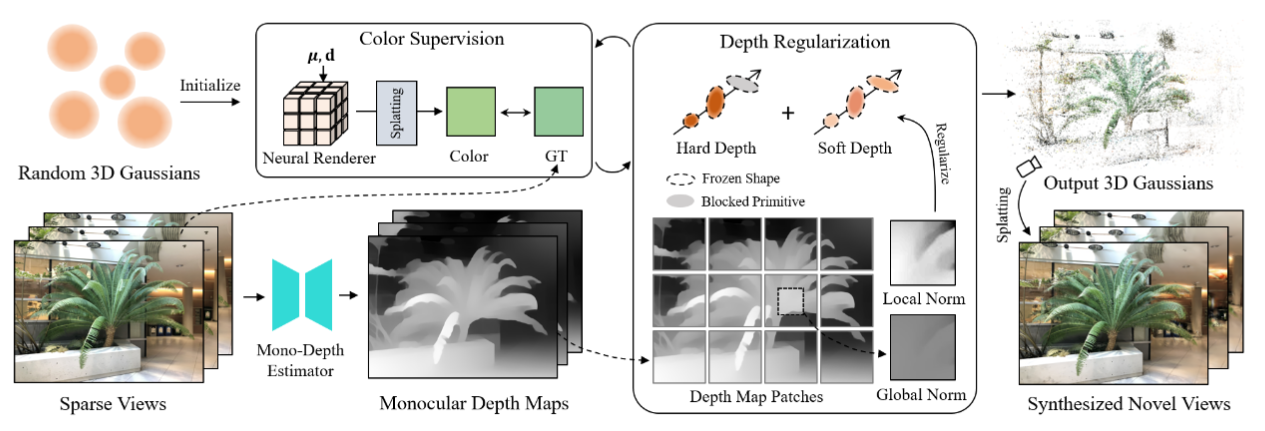


图15 DNGuassian的框架

简单来说，硬深度正则化主要关注最近的高斯基元，而软深度正则化则调整高斯的不透明度，以确保表面的完整性和细节。全局与局部深度归一化就是在局部和全局尺度上对深度进行归一化，使模型能够更好地关注小的局部深度变化，从而改善细节重建。

总的来说，这些文章都关注于三维重建，但是又各有千秋。我将他们的特点总结如下表1所示。

表1文章特点总结

|  |  |
| --- | --- |
| **文章** | **特点** |
| Monodepth2 | 自监督单目深度估计，使用U-Net网络进行深度预测，采用最小重投影误差损失函数，利用全分辨率深度估计提升图像细节 |
| Cheng等人的单目多目融合深度估计系统 | 融合单目和多目深度估计，采用共享骨干网络进行图像特征提取，通过置信度Map选择最准确的深度预测 |
| 3DFIRES | 强调从少量已知姿态图像进行场景级别三维重建，使用DRDF隐式表示和注意力机制推断隐藏面。 |
| Xu等人的贝叶斯扩散三维重建方法 | 结合自上而下的先验信息和自下而上的数据驱动过程，使用贝叶斯推断进行三维点云重建，采用信息融合策略提升模型性能 |
| UFORecon | 聚焦于有限的多目图像进行三维重建，采用Transformer处理跨视图特征匹配，增强不同视角的适应能力 |
| G3DR | 基于文字或单目图像生成三维物体，利用潜在扩散模型和三平面生成技术，优化深度一致性并提高生成质量 |
| DNGaussian | 基于3D高斯溅射场，采用深度正则化技术恢复准确的场景几何和细节，结合全局与局部深度归一化优化三维重建效果 |

从这些文章中，我们可以看到融合多种数据来源和处理方法的有效性，可以看到神经网络在三维重建领域的广泛应用，可以看到好的损失函数设计的重要性。我相信随着深度学习和相关传感技术的持续进步，三维重建方法将能实现更高精度和细节的重建，能够适应多模态的输入数据进行重建，并且有实时性和效率的提升，进而推动其在智能驾驶、游戏、电影等工业场景中发挥重要作用。

**参考文献**

1. Godard C, Aodha O M, Firman M, et al. Digging Into Self-Supervised Monocular Depth Estimation. 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South): IEEE, 2019. 3827~3837
2. Zhu J, Gao C, Sun Q, et al. A Survey of Indoor 3D Reconstruction Based on RGB-D Cameras. IEEE Access, 2024, 12: 112742~112766
3. Geiger A, Philip L, Raquel U. Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite. Conference on Computer Vision and Pattern Recognition (CVPR), 2012.
4. Sturm J, Engelhard N, Endres F, et al. A Benchmark for the Evaluation of RGB-D SLAM Systems. Proc. of the International Conference on Intelligent Robot Systems (IROS), 2012
5. Silberman N, Hoiem D, Kohli P, et al. Indoor Segmentation and Support Inference from RGBD Images
6. Cheng J, Yin W, Wang K, et al. Adaptive Fusion of Single-View and Multi-View Depth for Autonomous Driving. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2024. 10138~10147
7. Jin L, Kulkarni N, Fouhey D F. 3DFIRES: Few Image 3D REconstruction for Scenes with Hidden Surfaces. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2024. 9742~9751
8. Xu H, Lei Y, Chen Z, et al. Bayesian Diffusion Models for 3D Shape Reconstruction. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2024. 10628~10638
9. Na Y, Kim W J, Han K B, et al. UFORecon: Generalizable Sparse-View Surface Reconstruction from Arbitrary and Unfavorable Sets. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2024. 5094~5103
10. Reddy P, Elezi I, Deng J. G3DR: Generative 3D Reconstruction in ImageNet. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2024. 9655～9665
11. 1 Li J, Zhang J, Bai X, et al. DNGaussian: Optimizing Sparse-View 3D Gaussian Radiance Fields with Global-Local Depth Normalization. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2024. 20775～20785