

# MVSNNet系列



旷视研究院SLAM组实习生 黄百川

多视角  
图像

特征提  
取匹配

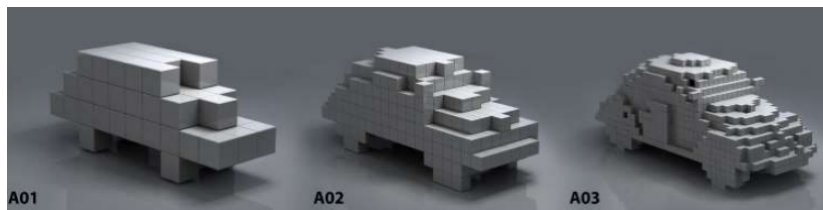
稀疏重  
建SFM

稠密重  
建MVS

## 稠密重建MVS

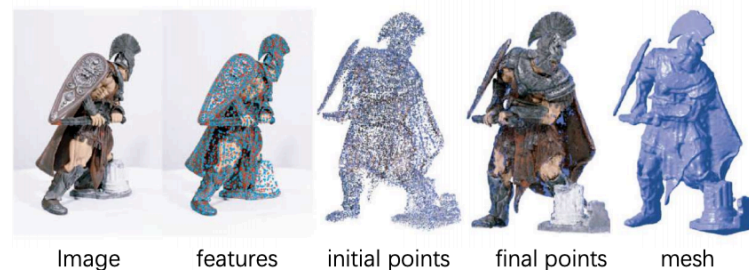
基于体素

MVS等价为一个3D空间Voxel的标记问题



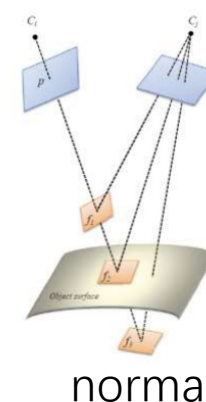
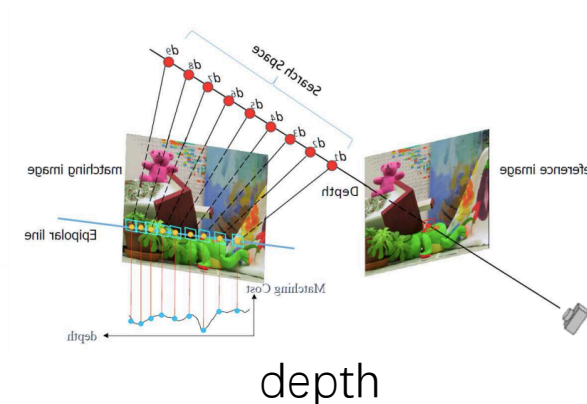
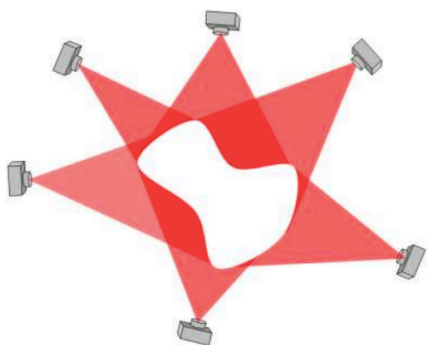
基于特征点扩散

利用特征点生成初始点云再扩散和过滤



基于深度图融合

为每一幅图片选取立体图像对来计算每一幅图的深度再进行深度图的融合得到点云模型

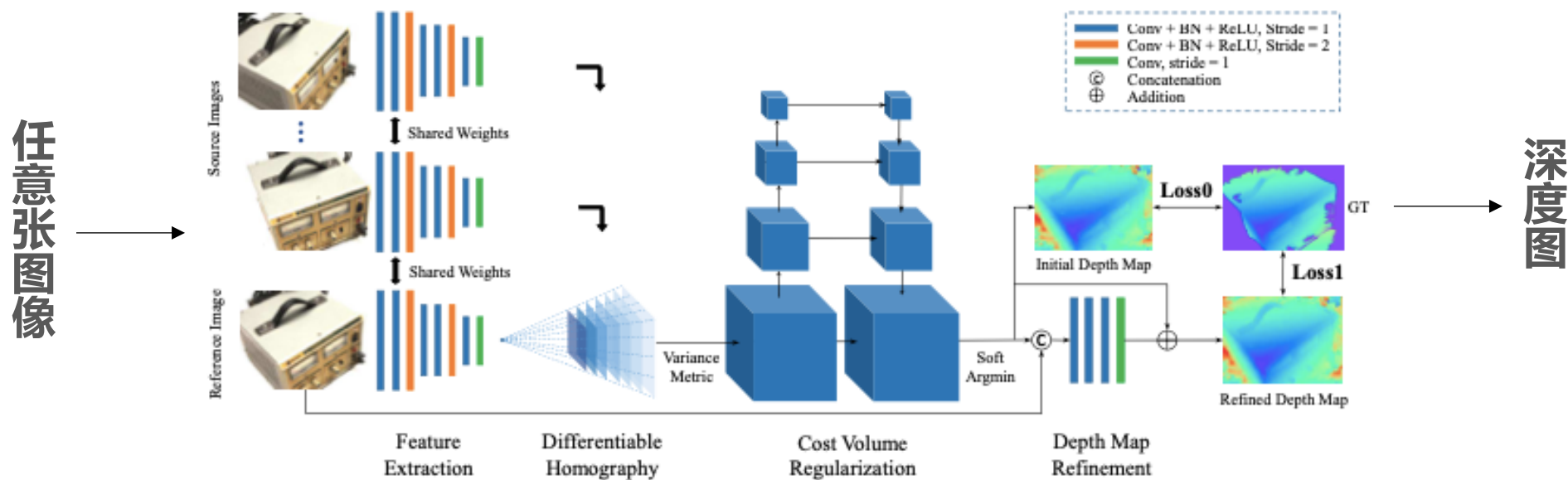


**传统MVS** 基于手工设计的相似性算法和鲁棒性矫正方法去计算稠密匹配并恢复3D点

虽然精度较高，但是由于弱纹理，镜面或者反射效应的表面的原因使得在大场景下的点云完整度待提高

**双目立体** 双目图像对是矫正对齐过的图片，不用考虑相机参数，只需估计图像对水平方向上像素点的视差

**网络MVS** 输入是任意位姿的多张图像，且多张图片之间的关系需要被整体考虑

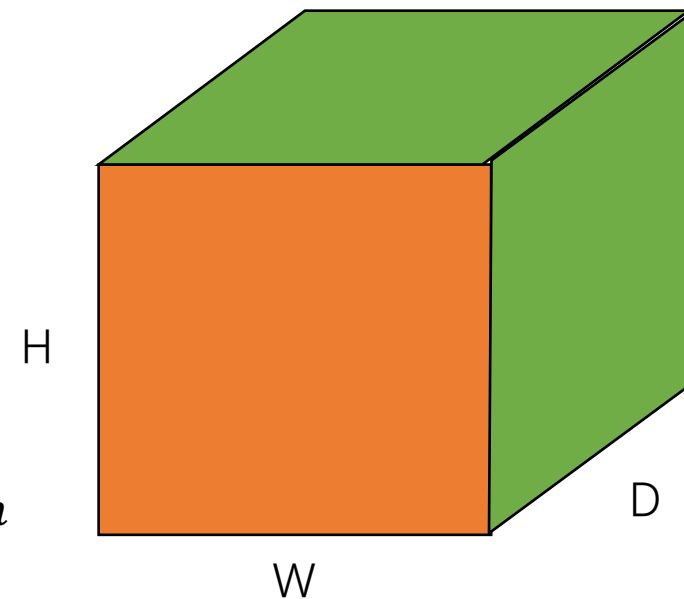
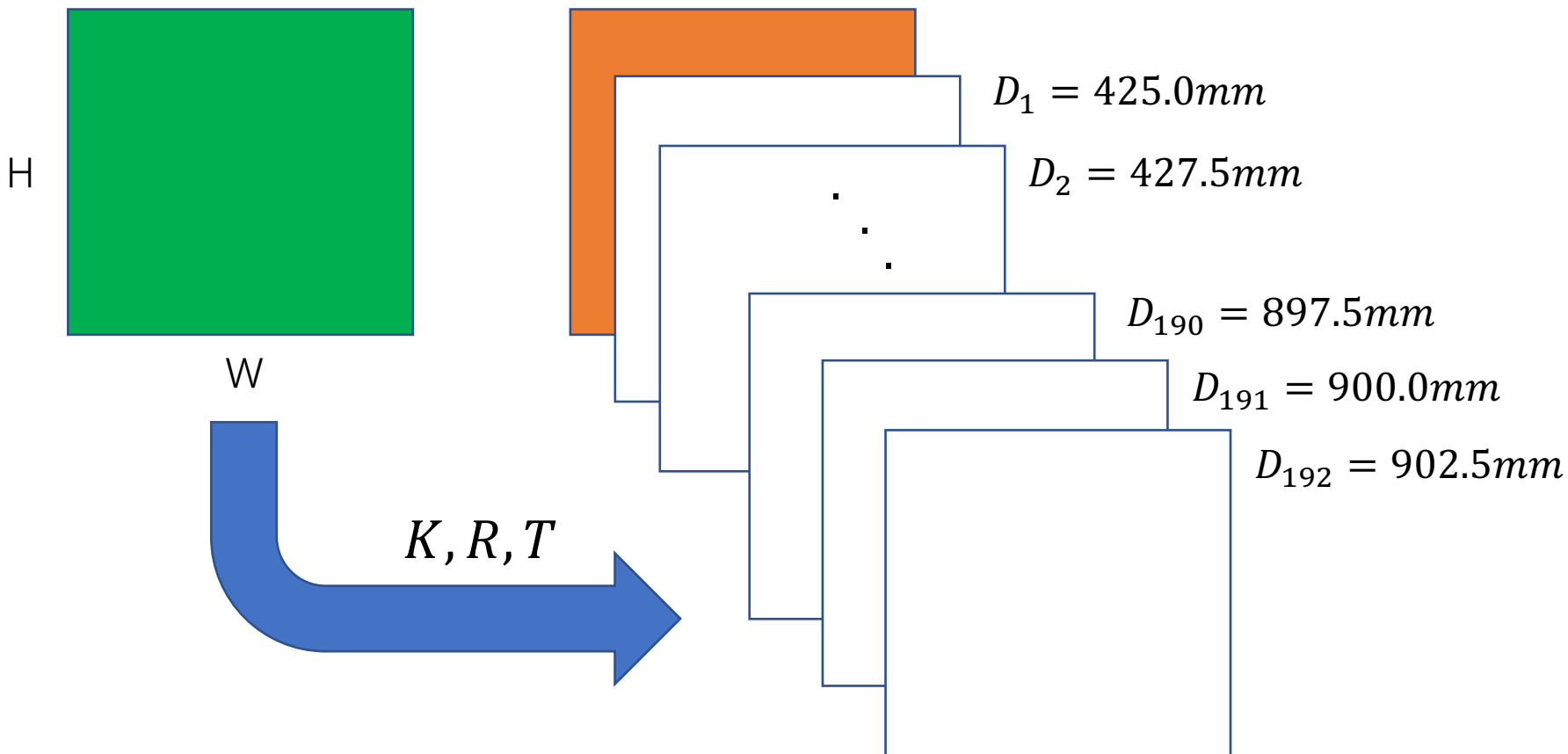


# Homography Warping

对比图

参考图

Cost Volume



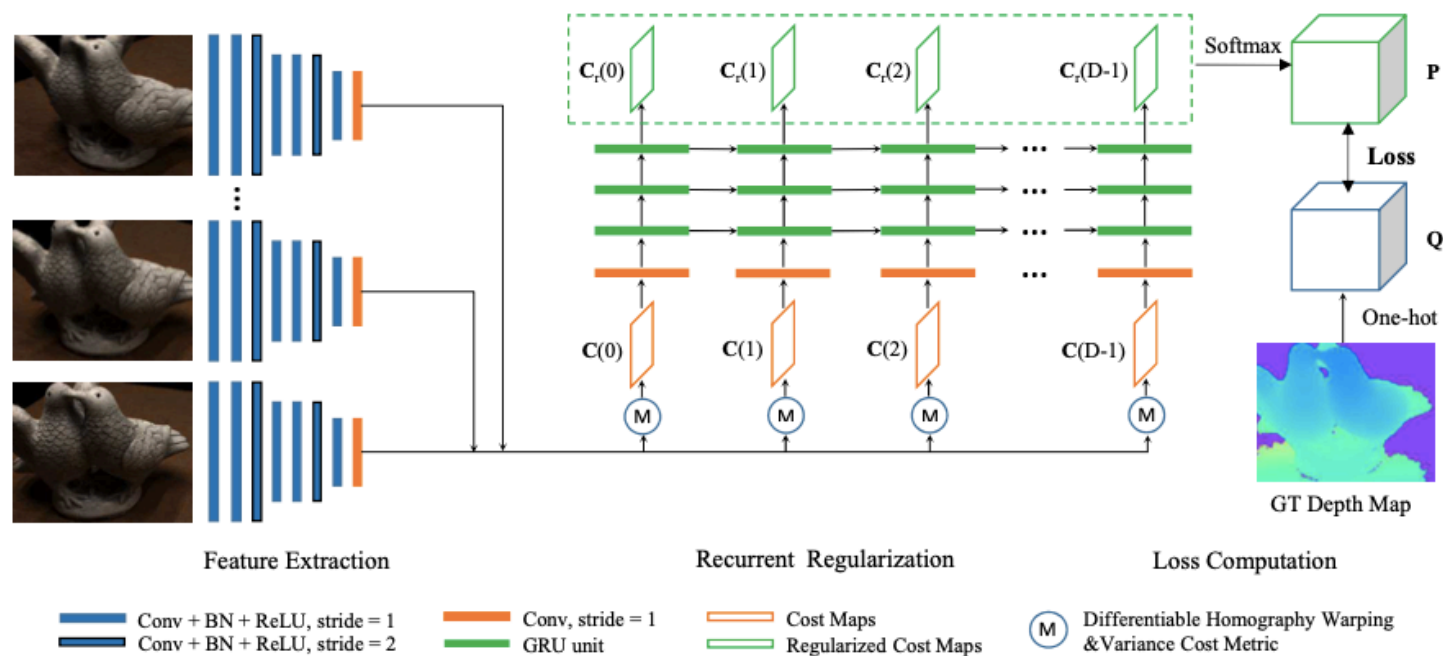
耗费显存

$$H \times W \times D \times F = 1600 \times 1184 \times 192 \times 32 = 11G$$

优点

1. 在保证准确度的情况下，点云恢复**完整性**最好
2. 运行速度是Gipuma的5倍，colmap的**100倍**

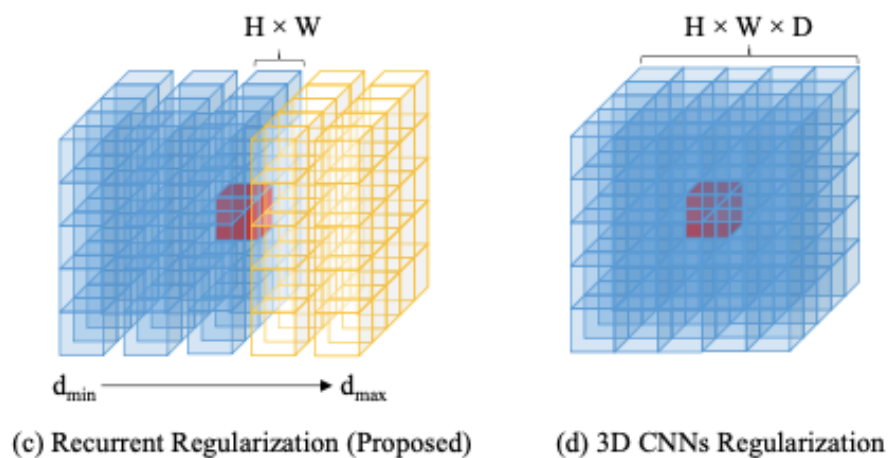
**网络结构** 门控循环单元代替了3D CNN去做cost volume的Regularization  
GRU可以更好地捕捉depth序列中depth相隔较大的整体关系



## 方法对比

3D CNN的方法占用的显存立方性增长

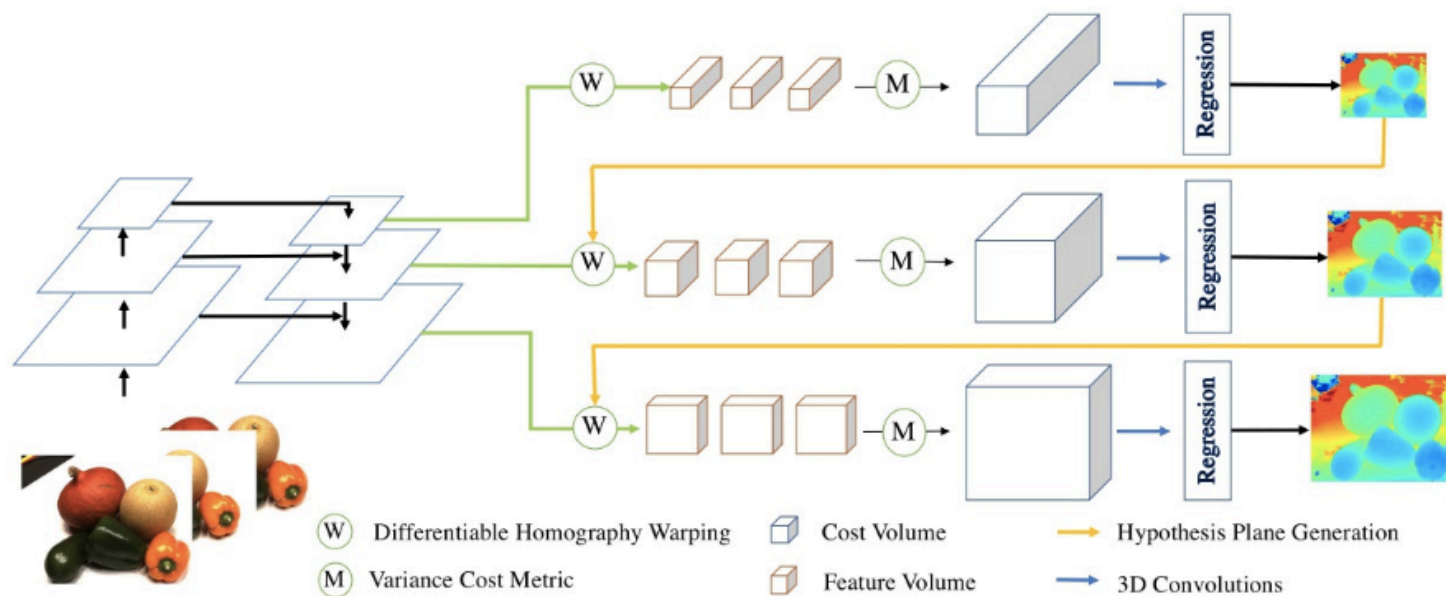
基于GRU的方法明显更节约显存



**优点** 低显存占用可以重建更大depth范围的场景和更高的重建精度, 且8倍效率相对于MVSNet

12G显存GPU最大可以处理分辨率为3072x2048的图像, 胜任大多数的MVS数据集除了ETH3D的高分辨率数据集

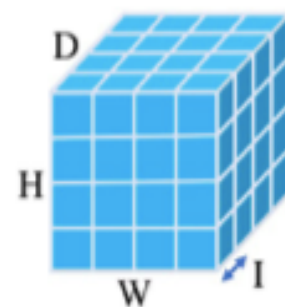
**网络结构** 采用特征图像金字塔结构对不同尺度的特征进行编码  
coarse-to-fine和thin adaptive volume



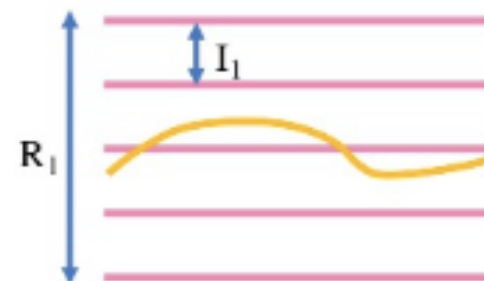
## Coarse-to-Fine

在步进的stage中，构造的cost volume：

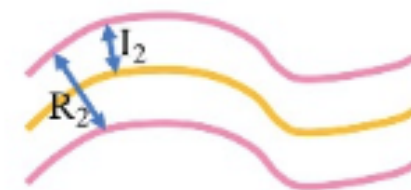
- $D$ 的范围 ( $R_1$ ) 会逐步变小
- 分辨率 ( $I_1$ ) 会逐步变小



	Plane Num.	Plane Interv.	Spatial Res.
Efficiency	Negative	Positive	Negative
Accuracy	Positive	Negative	Positive



Stage 1



Stage 2

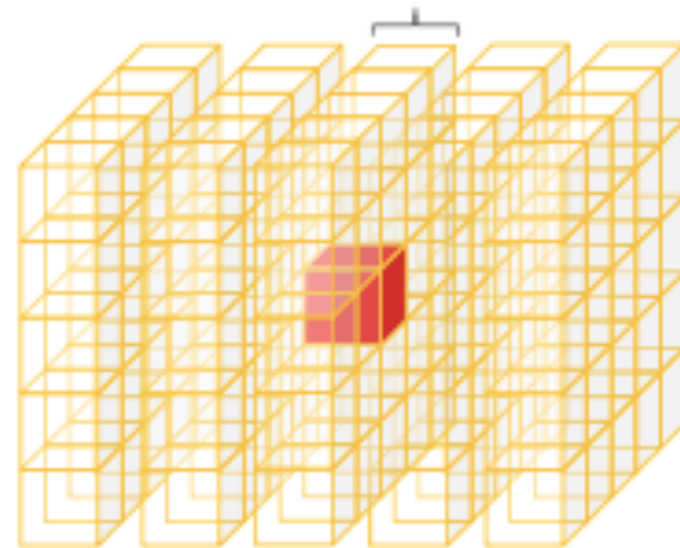
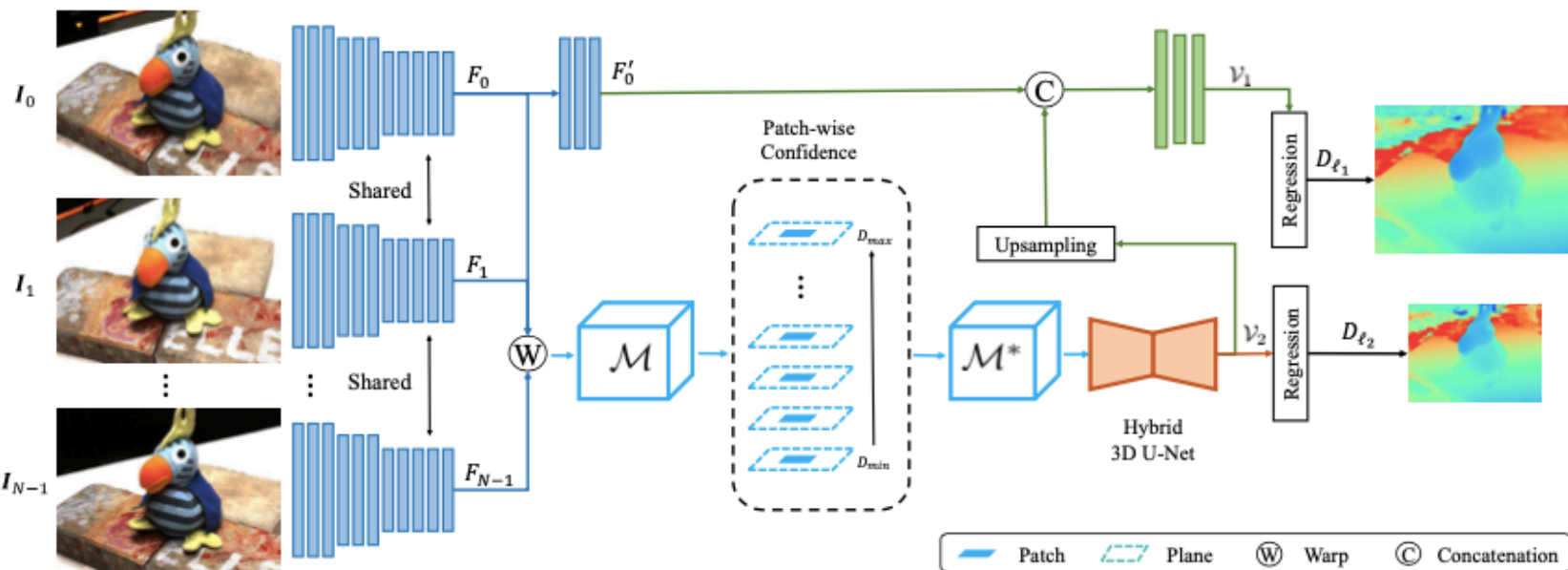
**优点** 50.6%的显存占用， 59.3%的运行时间和23.1%的精度提高相对于MVSNet



**网络结构** 从pixel-wise拓展到patch-wise cost volume  
Hybrid 3D U-Net + depth修正高分辨率输出

## Pixel-wise

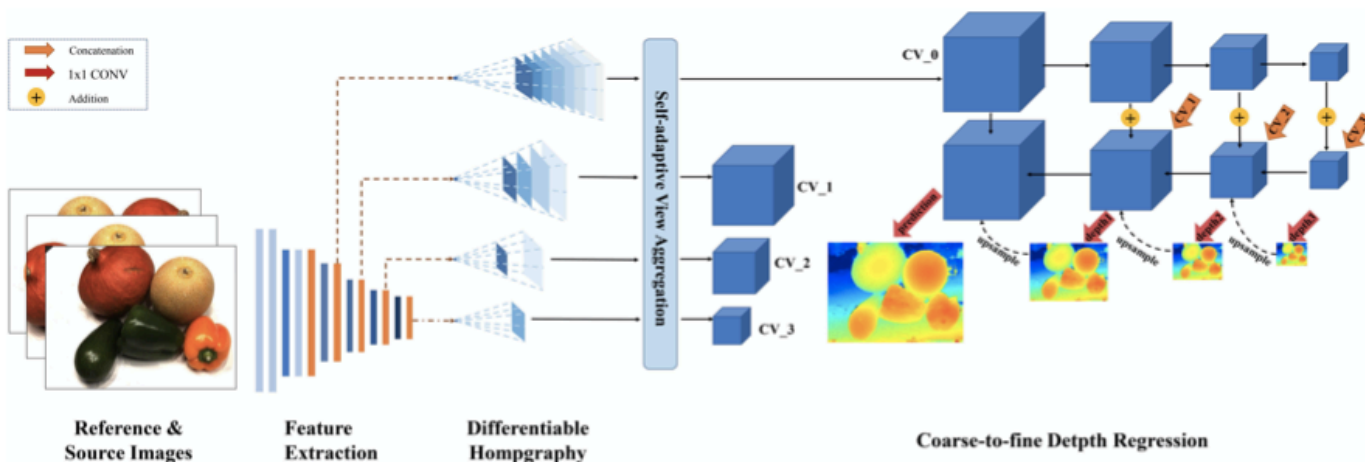
Patch-wise比pixel-wise提高了匹配的精度和鲁棒性



**特点** 借用了mvs中的patch-match的思想，使得depth和重建的点云的准确性和完整性都提高了

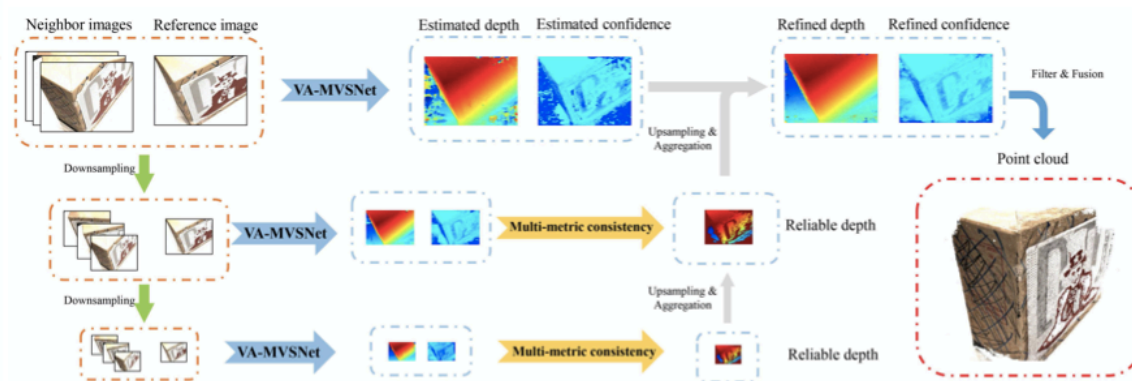
## VA-MVSNet

Coarse to fine  
多尺度特征图构建多cost volume联合估计



## PVA-MVSNet

Coarse to fine  
多尺度输入图分别经过VA-MVSNet结构  
联合估计深度



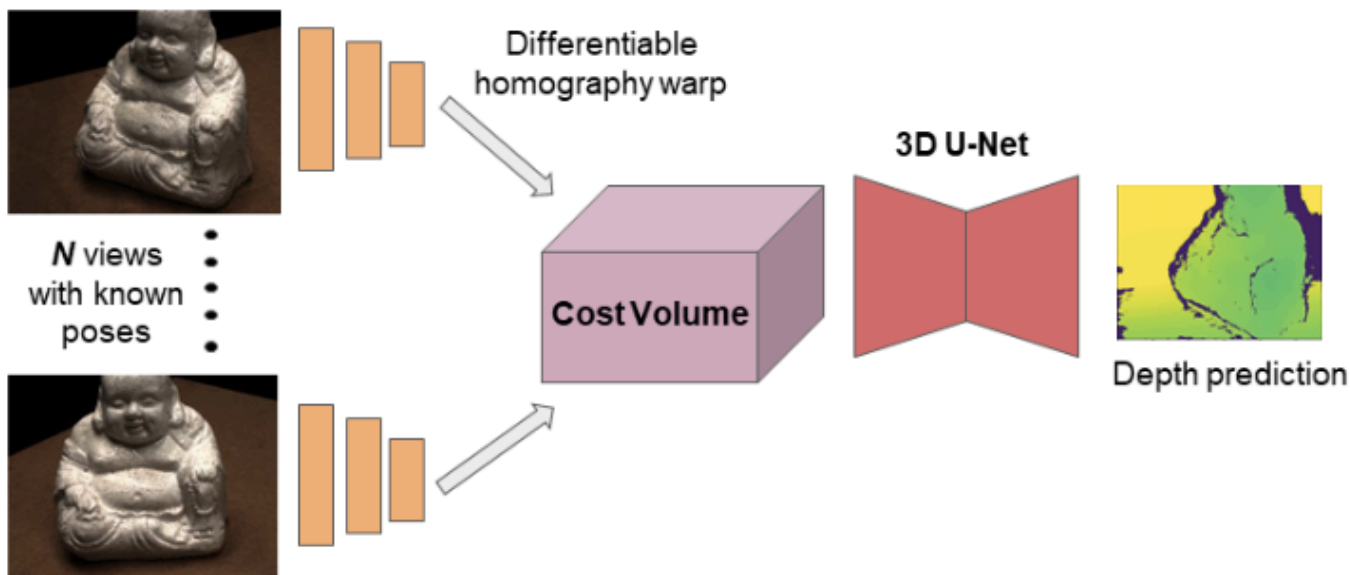
**特点** 特征图金字塔和图片金字塔结构保证了重建的精度和准确度

**缺点** 相对MVSNet需要显存变大，用算力换精度



## 自监督

利用投影光度一致性误差来产生自监督损失  
自监督容易快速拓展到实际的生产应用中



## Loss

$$Loss = \Sigma(\alpha L_{photo} + \beta L_{SSIM} + \gamma L_{smooth})$$

$$I'_r = I_s(KTD_rK^{-1}I_r)$$

$$L_{photo} = \sum_{m=1}^M \|(I_r - I'_r) \cdot V_r^m\| + \|(\nabla I_r - \nabla I'_r) \cdot V_r^m\|$$

$$L_{SSIM} = \Sigma[1 - SSIM(I_r, I'_r)] \cdot V_r^m$$

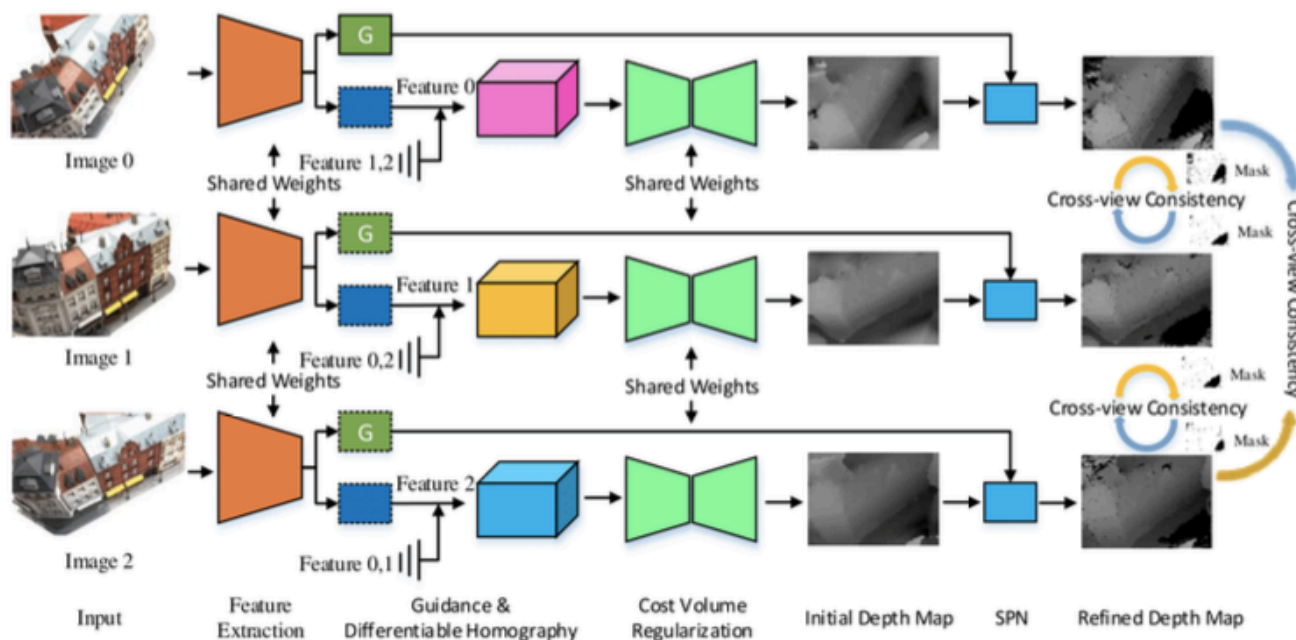
$$L_{smooth} = \Sigma e^{-|\nabla I_r|} \cdot |\nabla D_r|$$

**特点** 利用估计的depth寻找匹配点和图像梯度的融合，克服了多视图下不重叠区域和遮挡，以及不同光照异性问题

## 网络结构

从pixel-wise拓展到patch-wise cost volume

Hybrid 3D U-Net + depth修正高分辨率输出



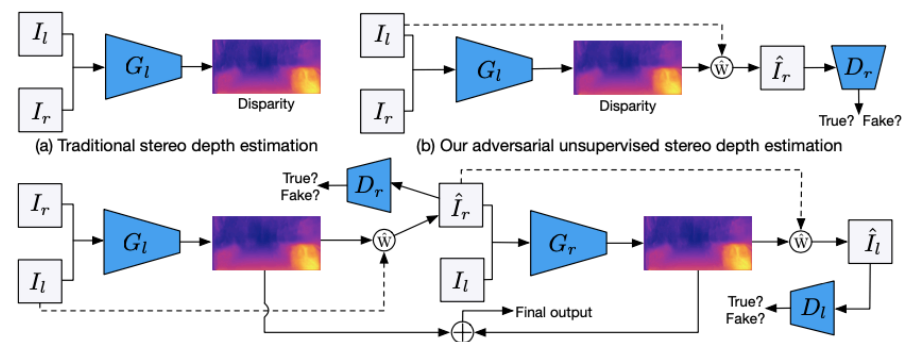
## Loss

在之前Loss的基础上，加入了重投影光度误差，投影几何误差。

$$Loss += \Sigma \delta L_{reprojected}$$

$$L_{reprojected} = L(I_r, I_r'') + L(D_r, D_r') \cdot V_r$$

## 延伸-对抗网络自监督双目深度估计



**特点** 丰富了loss函数，效果更好，但是也更加耗费显存，MVS<sup>2</sup>的显存耗费大约为上一个自监督网络的3倍

谢谢观看！

**MEGVII** 旷视