

开源 | AAAI'21杰出论文-三维重建新探索：解决数据依赖问题，让自监督信号更可靠！

原创 CV开发者都爱看的 极市平台 2021-05-25 22:00:00 手机阅读 𑀓

↑ 点击蓝字 关注极市平台

壹伴图

极市平台
extreme

月发文数目： **

月平均阅读： **

文章工具

已发文

采集图文 合成多

采集样式 查看

作者 | 许鸿斌

来源 | 极市平台

编辑 | 极市平台

极市导读

现有的方法都是假设不同视图之间的对应点具有相同的颜色，这在实践中并不总是正确的。这可能导致不可靠的自监督信号，并损害最终的重建性能。为了解决这个问题，本文提出了一个以语义共切分和数据扩充为指导的更可靠的监控框架。 >>加入极市CV技术交流群，走在计算机视觉的最前沿

Self-supervised Multi-view Stereo via Effective Co-Segmentation and Data-Augmentation

Hongbin Xu^{1,3*}, Zhipeng Zhou^{1*}, Yu Qiao^{1,2†}, Wenxiong Kang³, Qiuxia Wu³

¹ShenZhen Key Lab of Computer Vision and Pattern Recognition,
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

²Shanghai AI Lab, Shanghai, China

³South China University of Technology, Guangzhou, China

导读

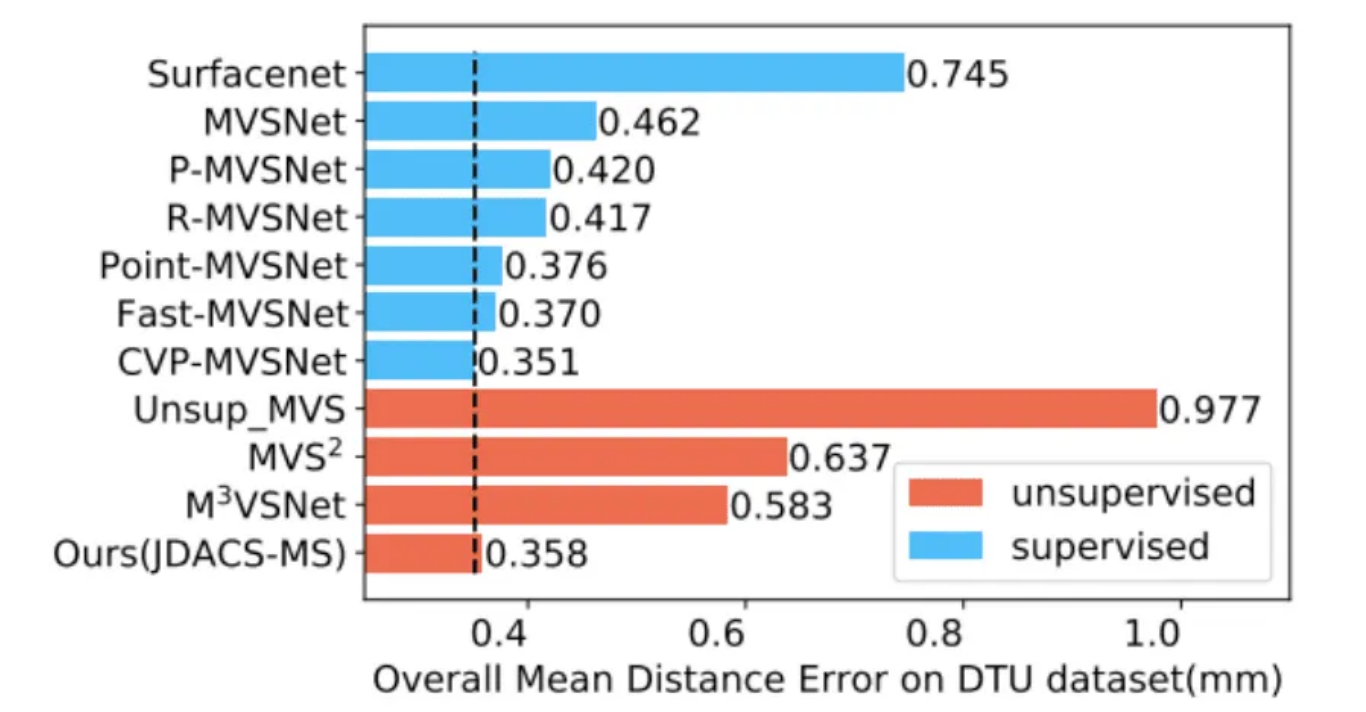
- 目标读者：对多视图立体几何（Multi-view stereo, MVS）感兴趣或有一定了解的人。
- 论文作者：深圳中科院先进技术研究院以及华南理工大学的研究团队。
- 论文：<https://www.aai.org/AAAI21Papers/AAAI-2549.XuH.pdf>
- Arxiv: <https://arxiv.org/abs/2104.05374>
- 代码：<https://github.com/ToughStoneX/Self-Supervised-MVS>

背景

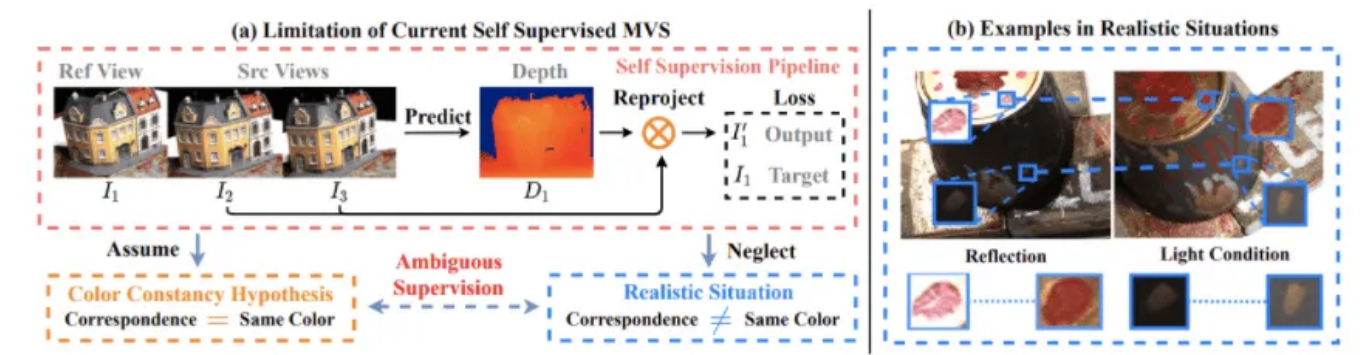
一直以来，基于多视图立体几何（Multi-view stereo, MVS）的三维重建都是一个广受关注的主题。近年来深度学习方法的发展，催生了一系列传统MVS方法与深度学习方法结合的工作，例如：MVSNet^[1]、R-MVSNet^[2]、CascadeMVSNet^[3]等等。以MVSNet为代表这一系列方法，通过可微单应性投影（Differentiable Homography）将立体几何的匹配关系嵌入到代价匹配体（Cost Volume）中，以实现端到端的神经网络。网络的输入是任意数量的多视角图像以及相机的内外参数，输出是某一个参考视

角下的深度图。相比于传统的MVS中的立体匹配方法，这些基于深度学习改进的方法能更好地重建稠密的三维点云信息，并对于弱纹理及噪声干扰的情况鲁棒性更强。在DTU^[4]、Tanks&Temples^[5]等公开数据集上都取得了不错的效果。但是，一个不可忽视的问题就是，这些深度学习方法依赖于大量的3D场景的Ground Truth数据。这会给其在现实场景的应用带来不少的麻烦，因为采集3D Ground Truth数据的成本相对高昂。由此，整个研究社区也开始着眼于无监督/自监督学习在MVS中的应用，以求摆脱对Ground Truth的依赖。

现有的自监督学习方法的基本思路是将需要Ground Truth的深度估计的回归任务转换为一个无监督的图像重建的代理任务。根据网络预测的参考视角下的深度图和其他视角图像通过单应性映射重建为参考视角的图像，只有预测的深度值正确时，重建图像才会尽可能与原图相似。尽管此前的无监督方法进一步地改进自监督方法并取得了不错的结果，例如：Unsup_MVS^[6]、MVS²^[7]、M³VSNet^[8]等，但是如下图所示，此前的无监督/自监督方法与有监督方法依然存在很明显的差异。我们的目标是提升自监督MVS方法的性能，但是在介绍具体方法之前，我们不妨先回顾一下自监督MVS方法本身是否存在一些缺陷导致了有监督与无监督方法的效果差异？



核心问题



由上图(a)可见，以图像重建任务作为代理任务的自监督MVS方法都依赖于一个比较粗糙的假设，即颜色一致性假设（Color Constancy Hypothesis）。该假设认为：多视图之间的匹配点具有相同的颜色。然而，由上图(b)中可见，在实际场景下，多视角图像的颜色值可能被各种外界因素干扰而导致匹配点具有不同的颜色，例如：光照变化、反光，噪声干扰等等。因此，基于颜色一致性假设的自监督信号在这些情况下很有可能引入错误的监督信号，反而干扰模型的效果。我们将这类问题称之为：颜色一致性歧义问题（Color Constancy Ambiguity）。那么，我们该如何解决这类问题呢？

方法

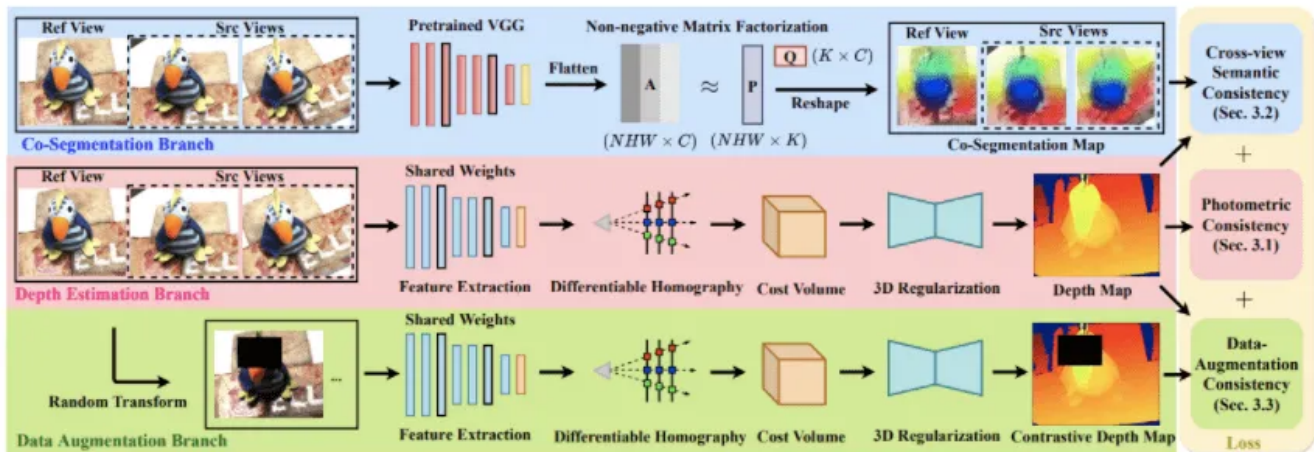
自监督MVS中的颜色一致性歧义问题，其根本原因在于图像重建这个代理任务仅仅考虑了颜色空间上的对应关系（Correspondence）。而这种基于RGB像素值差异的度量指标在表示多视图之间的对应关系时不够可靠，也限制了自监督方法的性能。那么，很自然地我们可以考虑考虑如何引入额外的先验知识，以提供一个更鲁棒的代理任务作为自监督信号。由此可以分为以下两点：

1. **语义一致性**：引入抽象的语义信息来提供鲁棒地对应性关系，将图像重建任务替换为语义分割图的重建任务构建自监督信号。
2. **数据增强一致性**：在自监督训练中引入数据增强，来提升网络针对不同颜色变化的鲁棒性。

但是在构建自监督信号时，依然存在一些不可忽视的问题：

- 对于语义一致性先验来说，获取语义分割图标注的成本是非常高昂的。此外训练集中的场景是动态变化的，我们无法像自动驾驶任务那样明确地定义好所有场景中所有元素的语义类别。这也是此前的自监督方法中不曾使用语义信息构建自监督损失的原因。为此，我们通过对多视角图像进行无监督的协同分割（Co-Segmentation），以挖掘出多视角图像之间的共有语义信息来构建自监督损失。
- 对于数据增强一致性先验来说，数据增强本身就会带来颜色分布的改变，换言之可能反过来引发颜色一致性歧义的问题，干扰自监督信号。为此，我们将单分支的自监督训练框架划分为双分支，使用原始分支的预测结果作为伪标签来监督数据增强分支的预测结果。

基于此，我们提出了一个新的自监督MVS训练框架：JDACS，如下图所示。



整个框架分为三个分支：

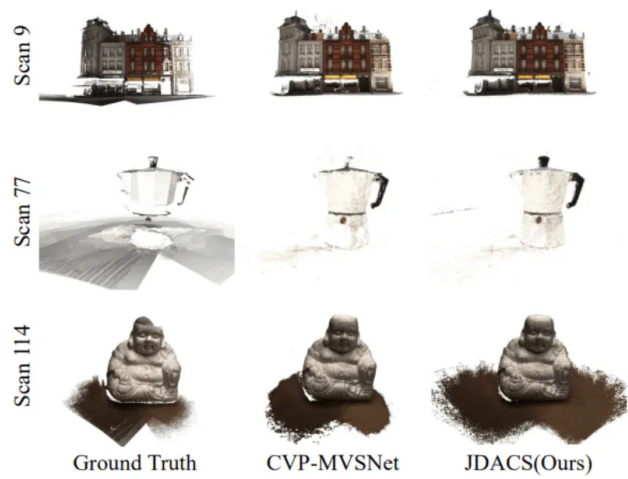
1. **深度估计分支**：输入参考视角（Reference View）以及源视角图像（Source View）到网络中，利用预测的深度图和源视角图像来重建参考视角图像。比较参考视角下重建图像和原图的差异，构建光度立体一致性损失（Photometric Consistency）。
2. **协同分割分支**：将输入多视图送入一个预训练的VGG网络，对其特征图进行非负矩阵分解（NMF）。由于NMF的正交约束，其过程可以看做多多视图之间的共有语义进行聚类，并输出协同分割图。随后通过预测的深度图和多视角的协同分割图构建分割图像重建任务，即语义一致性损失。
3. **数据增强分支**：对原始多视图进行随机的数据增强，并送入到网络中。以深度估计分支预测的深度图作为伪标签来监督数据增强分支的预测结果，构建数据增强一致性损失。

实验结果

DTU数据集上的定量实验结果：

	Method	Acc.	Comp.	Overall
Geo.	Furu	0.613	0.941	0.777
	Tola	0.342	1.190	0.766
	Camp	0.835	0.554	0.694
	Gipuma	0.283	0.873	0.578
Sup.	Surfacenet	0.450	1.040	0.745
	MVSNet	0.396	0.527	0.462
	P-MVSNet	0.406	0.434	0.420
	R-MVSNet	0.383	0.452	0.417
	Point-MVSNet	0.342	0.411	0.376
	Fast-MVSNet	0.336	0.403	0.370
	CVP-MVSNet	0.296	0.406	0.351
UnSup.	Unsup_MVS	0.881	1.073	0.977
	MVS ²	0.760	0.515	0.637
	M ³ VSNet	0.636	0.531	0.583
	JDACS	0.571	0.515	0.543
	JDACS-MS	0.398	0.318	0.358

DTU数据集上的定性实验结果：



Tanks&Temples数据集上的定量实验结果：

Method	Mean	Family	Francis	Horse	Lighthouse	M60	Panther	Playground	Train
MVS ²	37.21	47.74	21.55	19.50	44.54	44.86	46.32	43.38	29.72
M ³ VSNet	37.67	47.74	24.38	18.74	44.42	43.45	44.95	47.39	30.31
Ours	45.48	66.62	38.25	36.11	46.12	46.66	45.25	47.69	37.16

Tanks&Temples数据集上的定性实验结果：



有监督与无监督训练效果对比：

Method	Supervised	Input Size	Depth Map Size	Acc.	Comp.	Overall
MVSNet	✓	1152 × 864	288 × 216	0.456	0.646	0.551
JDACS	×	1152 × 864	288 × 216	0.571	0.515	0.543
CVP-MVSNet	✓	1600 × 1152	1600 × 1152	0.296	0.406	0.351
JDACS-MS	×	1600 × 1152	1600 × 1152	0.398	0.318	0.358

结语

这篇工作是我们对于自监督MVS的一些探索，重新思考了自监督信号的有效性，因为此前的基于图像重建代理任务的自监督方法都是基于颜色一致性假设。然而，现实场景中天然存在的颜色干扰会干扰到自监督训练，并引发颜色一致性歧义问题，导致训练过程中引入了错误的监督信号。我们提出的自监督MVS框架则试图引入额外的Correspondence先验知识来使得自监督信号更加可靠。一方面，通过无监督地挖掘协同分割图中的语义一致性信息来引入抽象的匹配关系；另一方面，通过双分支结构引入数据增强一致性的先验来提升网络应对噪声的鲁棒性。从实验结果可以看出，我们相比于此前的自监督MVS方法有一定的性能提升。

当然，我们提出的方法依然存在一些待解决的问题：首先，在非纹理区域如黑色/白色背景等等，是不存在有效的自监督信号的，因为所有背景像素点的颜色乃至语义都是相同的；其次，我们通过协同分割的方法只挖掘出了相对粗糙的语义信息，这是由于基于ImageNet分类任务预训练的VGG模型并不适用于需要关注到细节语义的分割任务。

最后附上我们论文的引用：

```
1 @inproceedings{xu2021self,
2   title={Self-supervised Multi-view Stereo via Effective Co-Segmentation and Data-Augmentation},
3   author={Xu, Hongbin and Zhou, Zhipeng and Qiao, Yu and Kang, Wenxiong and Wu, Qiuxia},
4   booktitle={Proceedings of the AAAI Conference on Artificial Intelligence},
5   year={2021}
6 }
```

参考文献

- 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.
- Yao Y, Luo Z, Li S, et al. Recurrent mvsnet for high-resolution multi-view stereo depth inference[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 5525-5534.
- Gu X, Fan Z, Zhu S, et al. Cascade cost volume for high-resolution multi-view stereo and stereo matching[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 2495-2504.

本文亮点总结

1.自监督MVS中的颜色一致性歧义问题，其根本原因在于图像重建这个代理任务仅仅考虑了颜色空间上的对应关系（Correspondence）。

如果觉得有用，就请分享到朋友圈吧！



极市平台

专注计算机视觉前沿资讯和技术干货，官网：www.cvmart.net

624篇原创内容

公众号

▲点击卡片关注极市平台，获取最新CV干货

公众号后台回复“目标检测竞赛”获取目标检测竞赛经验资源～

极市干货

- YOLO教程：一文读懂YOLO V5 与 YOLO V4 | 大盘点 | YOLO 系目标检测算法总览 | 全面解析YOLO V4网络结构
- 实操教程：PyTorch vs LibTorch：网络推理速度谁更快？ | 只用两行代码，我让Transformer推理加速了50倍 | PyTorch AutoGrad C++层实现
- 算法技巧（trick）：深度学习训练tricks总结（有实验支撑） | 深度强化学习调参Tricks合集 | 长尾识别中的Tricks汇总（AAAI2021）
- 最新CV竞赛：2021 高通人工智能应用创新大赛 | CVPR 2021 | Short-video Face Parsing Challenge | 3D人体目标检测与行为分析竞赛开赛，奖池7万+，数据集达16671张！



极市原创作者激励计划

极市平台深耕CV开发者领域近5年，拥有一大批优质CV开发者受众，覆盖微信、知乎、B站、微博等多个渠道。通过极市平台，您的文章的观点和看法能分享至更多CV开发者，既能体现文章的价值，又能让文章在视觉圈内得到更大程度上的推广。

对于优质内容开发者，极市可推荐至国内优秀出版社合作出书，同时为开发者引荐行业大牛，组织个人分享交流会，推荐名企就业机会，打造个人品牌 IP。

投稿须知：

- 1.作者保证投稿作品为自己的原创作品。
- 2.极市平台尊重原作者署名权，并支付相应稿费。文章发布后，版权仍属于原作者。
- 3.原作者可以将文章发在其他平台的个人账号，但需要在文章顶部标明首发于极市平台

投稿方式：

添加小编微信Fengcall（微信号：fengcall19），备注：姓名-投稿



△长按添加极市平台小编

觉得有用麻烦给个在看啦~

阅读原文

喜欢此内容的人还喜欢

15个目标检测开源数据集汇总

极市平台