The following publication Yingyu Wang, Yakun Ju, Muwei Jian, Kin-Man Lam, Lin Qia, and Junyu Dong "Self-supervised depth completion with attention-based loss", Proc. SPIE 11515, International Workshop on Advanced Imaging Technology (IWAIT) 2020, 115152T (1 June 2020) is available at https://doi.org/10.1117/12.2566222.

Self-supervised Depth Completion with Attention-Based loss

Yingyu Wang, Yakun Ju
Ocean University of China
Qingdao, China
{wangyingyu, juyakun}@stu.ouc.edu.cn

Muwei Jian
Shandong University of Finance and Economics
Jinan, China
muwei.jian@connect.polyu.hk

Kin-Man Lam
Hong Kong Polytechnic University
Hong Kong
enkmlam@polyu.edu.hk

Lin Qi*, Junyu Dong*
Ocean University of China
Qingdao, China
{ qilin, dongjunyu}@ouc.edu.cn

***Abstract*—Deep completion which predicts dense depth from sparse depth has important applications in the fields of robotics, autonomous driving and virtual reality. It compensates for the shortcomings of low accuracy in monocular depth estimation. However, the previous deep completion works evenly processed each depth pixel and ignored the statistical properties of the depth value distribution. In this paper, we propose a self-supervised framework that can generate accurate dense depth from RGB images and sparse depth without the need for dense depth labels. We propose a novel attention loss that takes into account the statistical properties of the depth value distribution. We evaluate our approach on the KITTI Dataset. The experimental results show that our method achieves state-of-the-art performance. At the same time, ablation study proves that our method can effectively improve the accuracy of the results.**

Keywords—Deep completion; self-supervised; attention loss; statistical properties; monocular depth estimation

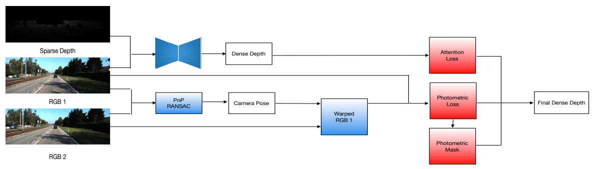
I. Introduction

In our work, we consider the statistical characteristics of the depth data in the deep completion task to pay more attention to pixels in the distance. So far, we are the first to consider this feature in the deep completion task. In addition, we further utilized the photometric loss to solve the problem caused by occlusion and dynamic objects. To summarize, we propose the following contributions:

- We propose an end-to-end CNN framework that can self-supervise learning dense depth from RGB images and sparse depth without the need for additional depth labels. In addition, the framework uses dual branches to separately process the inputs and implement the post-fusion strategy.
- We propose a novel loss function that can focus more on distant pixels to improve the accuracy of depth complementation. In addition, in the loss function we further exploit the photometric loss to solve the occlusion problem.

II. Proposed Method

The pipeline is shown below. The pipeline consists of two parts: the supervised part (top part) and the self-supervised part (lower part). The supervised part inputs the sparse depth and the corresponding RGB images, using the sparse depth as the supervisor, directly estimating the dense depth through the encoder-decoder network. The self-supervised part with adjacent RGB images as input, calculate the camera pose by PnP and RANSAC algorithm and further calculate the photometric loss as self-supervised signal. The self-supervisd path further utilizes photometric loss to generate photometric mask.



Our novel loss function consists of four parts: 1) attention loss, 2) photometric loss, 3) smoothness loss and 4) photometric mask. The attention loss is defined as:

$$\mathcal{L}_{\text{attention}} = \frac{1}{N}\sum_{i=1}^{N} \alpha_D \cdot \| \mathrm{Mask}_{\{d>0\}} \cdot \left(\mathrm{d}_i^{\mathrm{pred}} - \mathrm{d}_i^{\mathrm{GT}} \right) \|_2^2$$

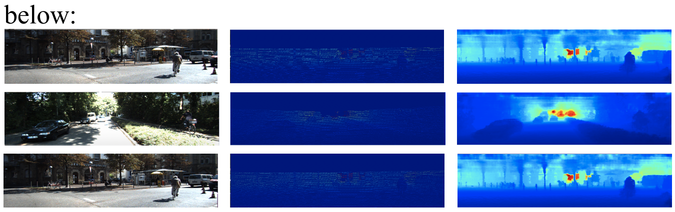
$$\alpha_D = 1 + \frac{d_i^{GT}}{\max\left(d_i^{GT}\right)},$$

the total loss function for the entire self-supervised framework becomes:

$$\mathcal{L}_{\text{total}} = Mask_{\text{photometric}} \odot \mathcal{L}_{\text{attention}} + \alpha \mathcal{L}_{\text{photometric}} + \beta \mathcal{L}_{\text{smoothess}}$$

III. Expeiments And Conclusion

We have conducted many experiments to prove the effectiveness of my method, including comparison with other self-supervised deep completion methods and ablation study. The experiment results show that our method achieves state-of-the-art performance and ablation study prove that our novel loss can effectively reduce the error. The results are shown below:



(a) RGB input (b) Sparse depth input (c) Dense depth ouput

---

* Corresponding author