Stereo Video Depth Estimation in 2025

3D Vision Group 18

Hepeng Fan, Qingrui Deng, Tong Su, Yixin Zhou

Supervisor: Haofei Xu

1 Introduction

Recently, advances in depth models have yielded complementary strengths, but problems still exist (See Fig1).

This project aims to **develop a stereo video depth** model that combines both strengths, delivering **accurate and temporally stable** metric depth for 3D/4D applications.

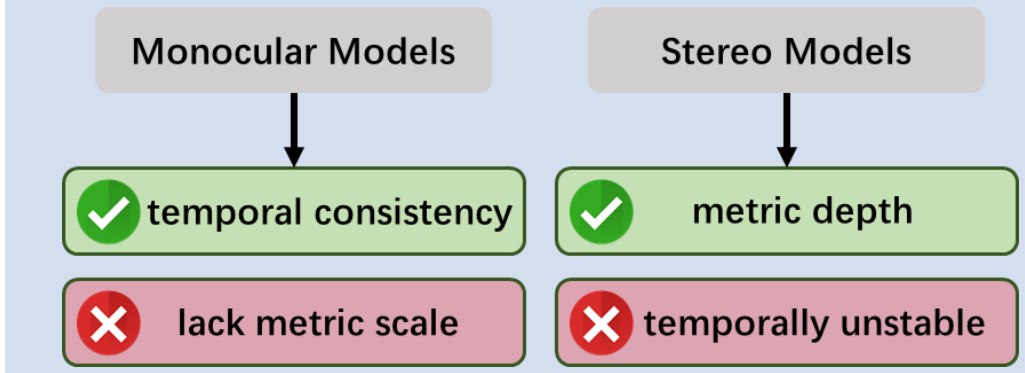


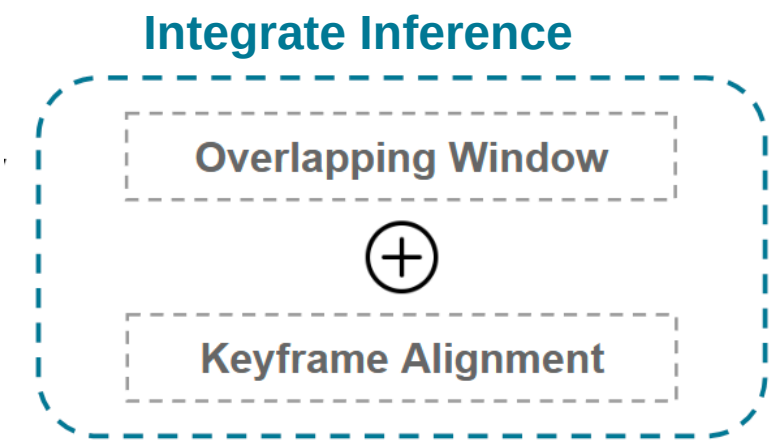
Fig 1. Strength and Weakness of Monocular and Stereo Models

4 Training Strategy

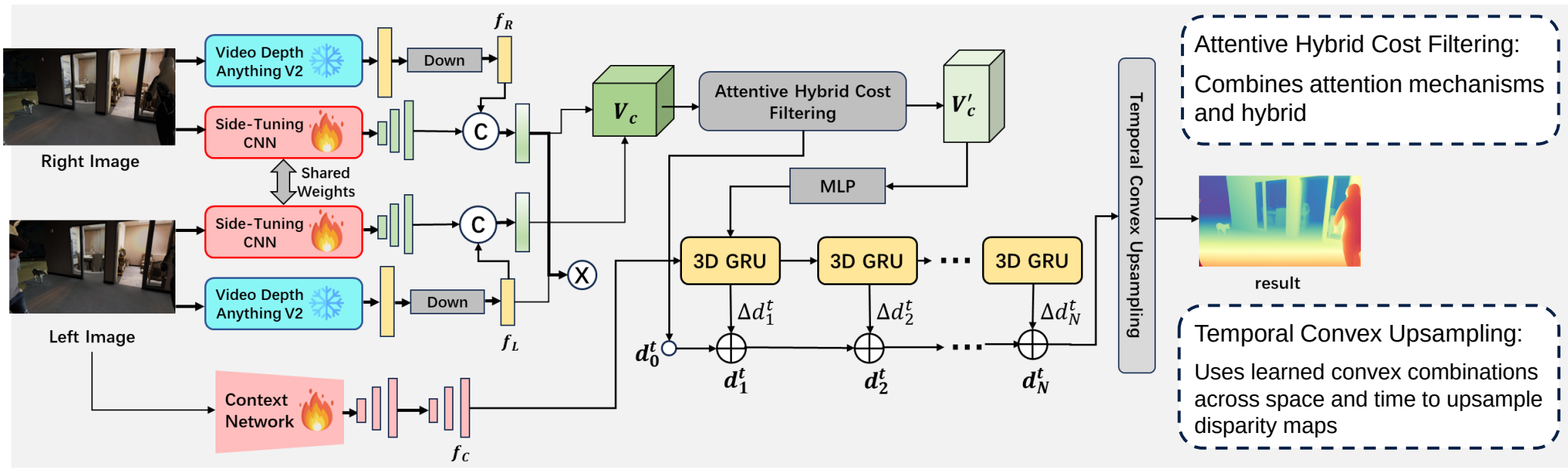
Data Augmentation: Brightness; Contrast; Crops

Optimizer: AdamW with learning rate 5e-6 and weight decay 1e-5

Learning Rate Scheduling: Cosine annealing schedular with gradient clipping at 0.1.



2 Method Overview



4 Dataset and Evaluation Metrics

Loss Function:

$$\mathcal{L} = \sum_{t=1}^{T} \sum_{n=1}^{N} \gamma^{N-n} \left\| D_{\text{gt}}^{t} - D_{n}^{t} \right\|$$

Evaluation Metrics

$$\text{TEPE}(\hat{D}, D) = \sqrt{\sum_{t=1}^{T-1} \left((\hat{D}_t - \hat{D}_{t+1}) - (D_t - D_{t+1}) \right)^2}$$

Dataset: Dynamic Replica and Sintel

Built on the static Replica indoor models and Sintel movies

5 Results

Method	TEPE	EPE
RAFTStereo	2.09	4.89
IGEVStereo	1.74	2.44
DynamicStereo	1.43	4.26
BidaStereo	1.26	1.99
Stereo Any Video	1.07	1.75
Ours	**0.51**	**5.61**

Table1. The comparison of our model and other methods. The result shows that we **significantly increase the TEPE**.

References

1. Wen, B., Trepte, M., Aribido, J., Kautz, J., Gallo, O., & Birchfield, S. (2025). *FoundationStereo: Zero-Shot Stereo Matching*.
2. Jing, J., Luo, W., Mao, Y., & Mikolajczyk, K. (2025). *Stereo Any Video: Temporally Consistent Stereo Matching*.