

# Splat and Distill: Augmenting Teachers with Feed-Forward 3D Reconstruction For 3D-Aware Distillation



David Shavin



Sagie Benaim



האוניברסיטה העברית בירושלים  
THE HEBREW UNIVERSITY OF JERUSALEM



**ICLR**  
International Conference On  
Learning Representations

## Excellent at 2D Tasks

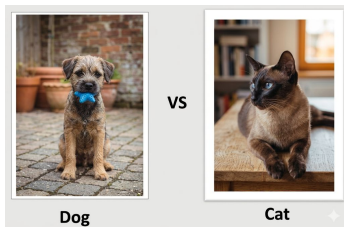


Image  
Classification

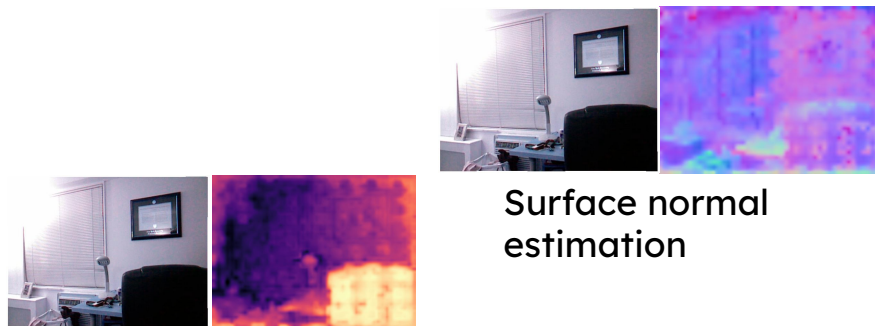


Instance Retrieval



Semantic Segmentation

## Struggle with 3D Awareness



Depth estimation

Surface normal  
estimation



Multiview geometric  
correspondence

Excellent at 2D Tasks

Struggle with 3D Awareness

# How can we make them 3D aware?

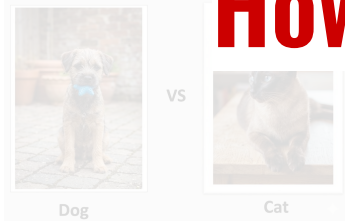
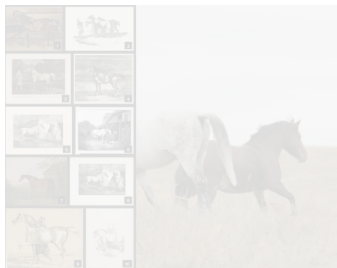
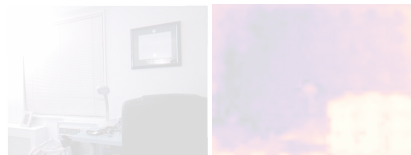


Image Classification



Instance Retrieval

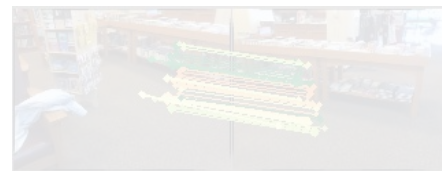
Semantic Segmentation



Depth estimation

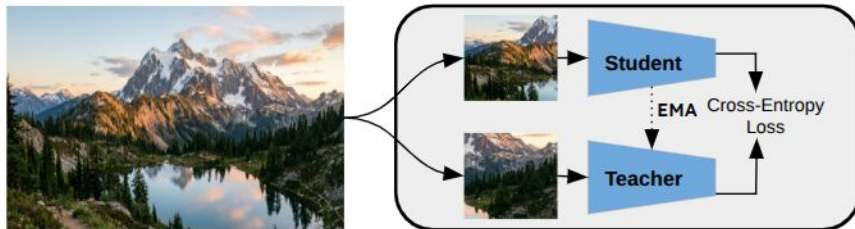


Surface normal estimation



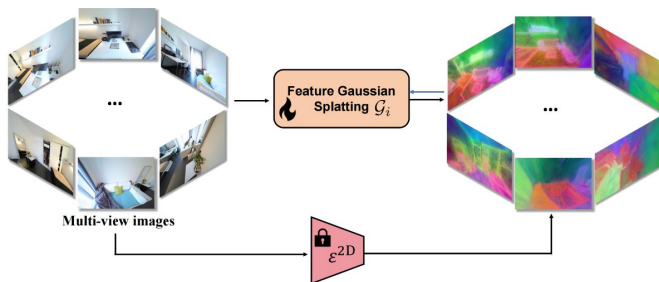
Multiview geometric correspondence

## DinoV2 [Oquab et al.]

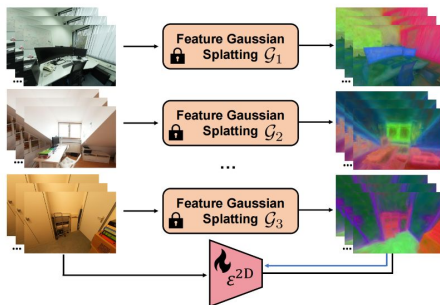


## FiT3D [Yue et al.]

Stage I: Lifting Features to 3D (Per-scene)



Stage II: 3D-Aware Fine-Tuning (Multi-scene)



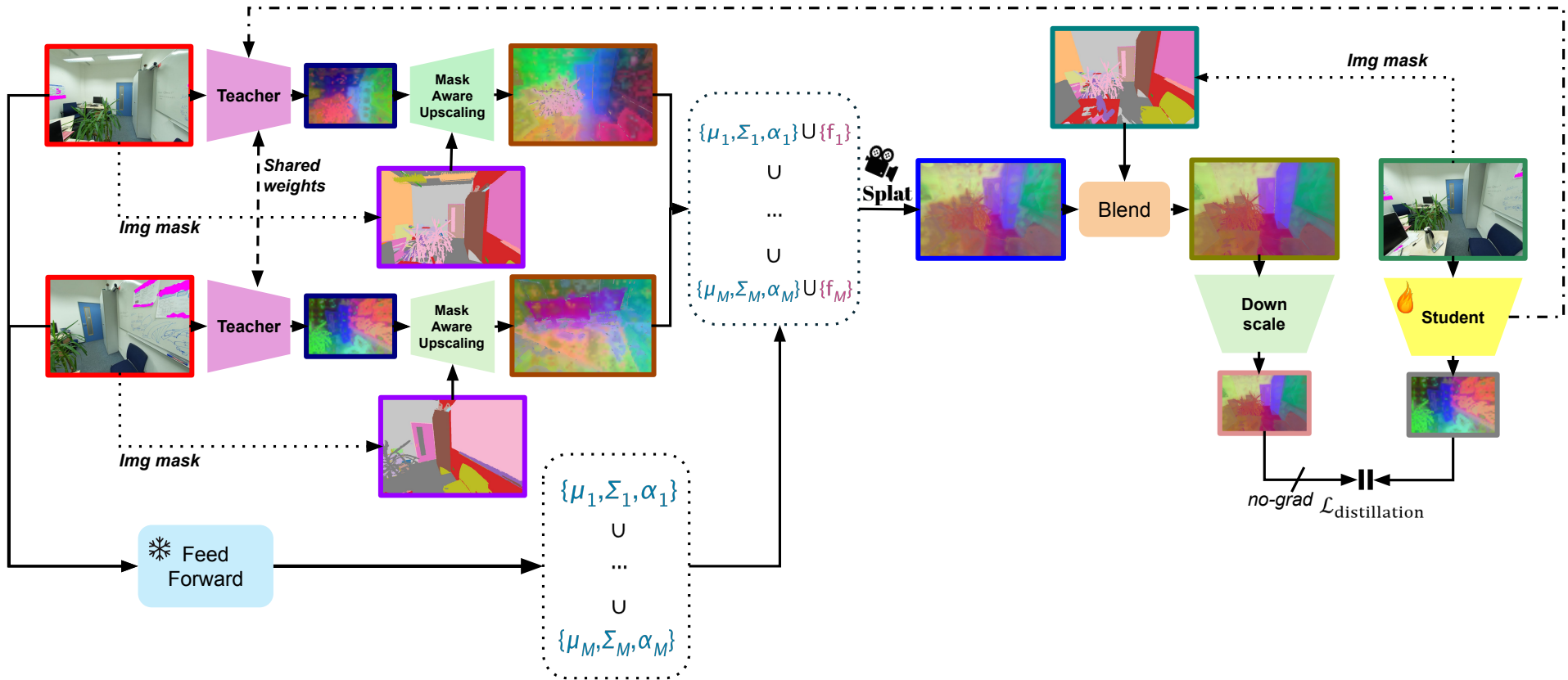
## Ours - SnD

3D aware

Adaptive teacher  
self-distillation

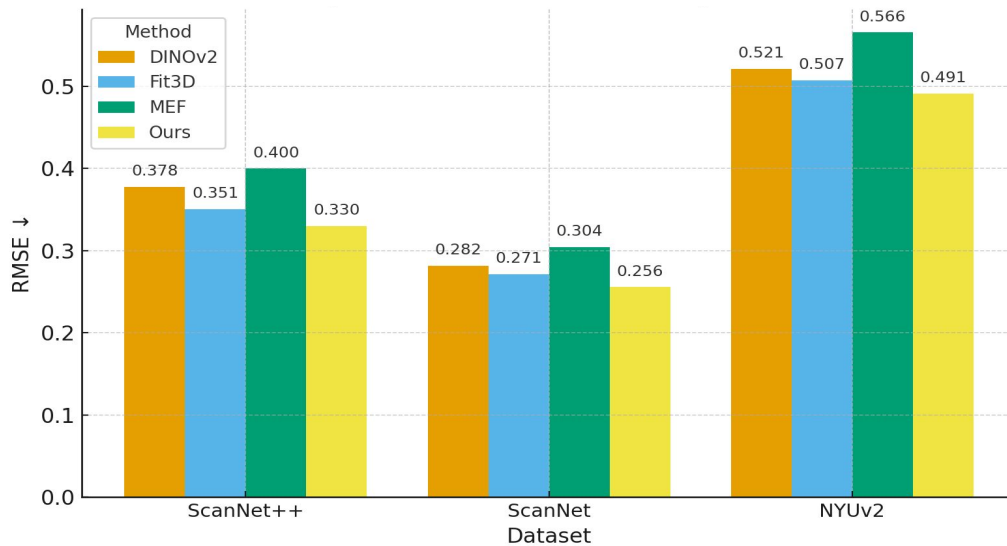


# EMA



# Results

# Depth Estimation



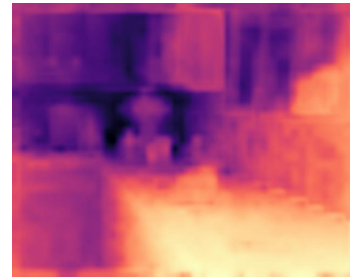
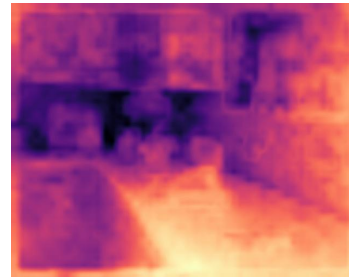
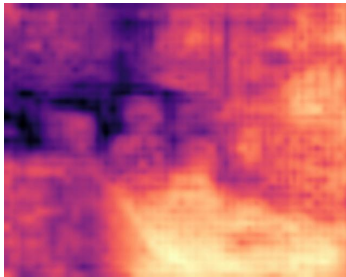
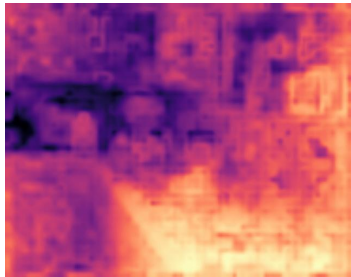
Image

DinoV2

MEF

Fit3D

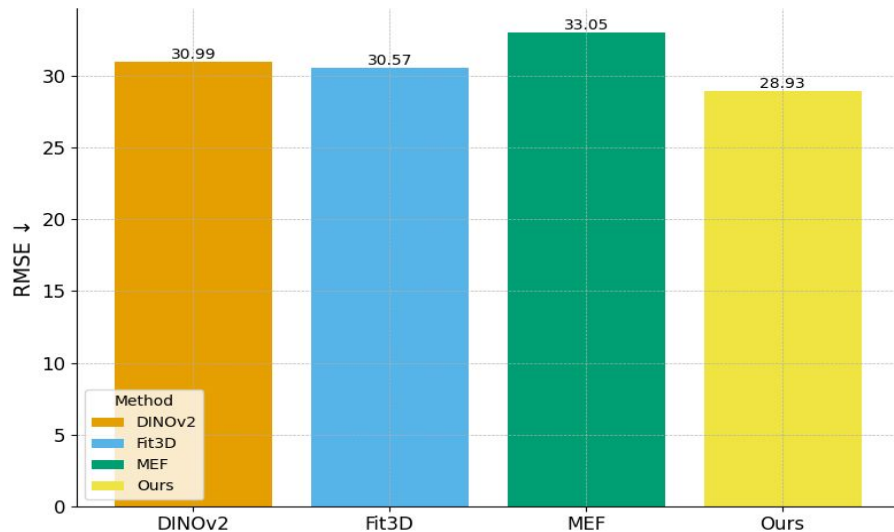
Ours



MEF - Multiview Equivariance Improves 3D Correspondence Understanding with Minimal Feature Finetuning

Fit3D - Improving 2D Feature Representations by 3D-Aware Fine-Tuning

# Surface Normal Estimation



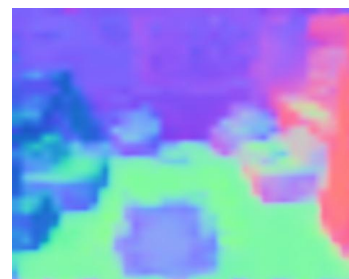
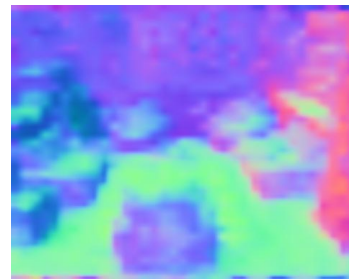
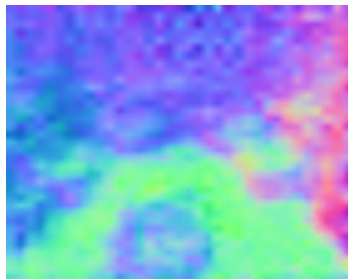
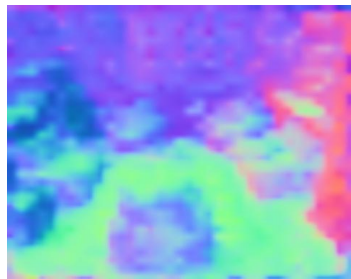
Image

DinoV2

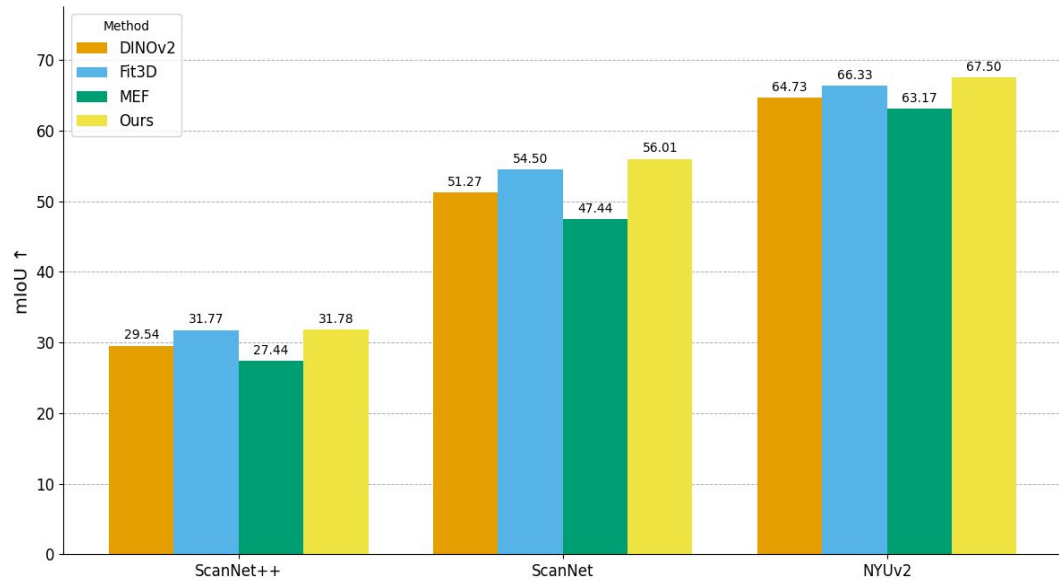
MEF

Fit3D

Ours



# Semantic Segmentation



Image

DinoV2

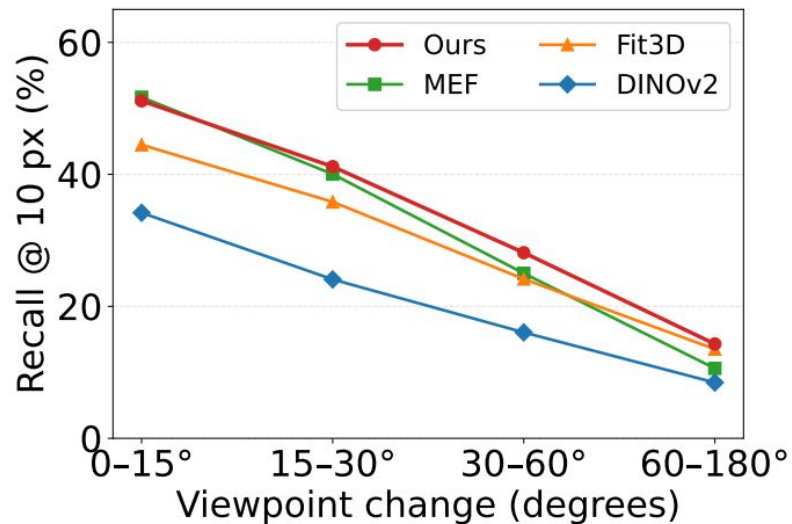
MEF

Fit3D

Ours



# Multiview - Correspondence



DinoV2



MEF



Fit3D



Ours



# Conclusion

- **Boosts 3D awareness** via an adaptive student-teacher distillation framework.
- **Eliminates slow optimization** by leveraging fast feed-forward Gaussian Splatting.
- **Enhances memory and computational efficiency** by requiring significantly fewer Gaussians than optimization-based methods
- **Achieves SOTA results** across diverse 3D-aware evaluations.

# Thank you for listening!

