Semantic Depth Extraction: Updates

Utkarsh Singhal, Stella Yu {s.utkarsh, stellayu}@berkeley.edu
Chun Kai Wang, Michael Lou {ckwang, ylou}@amazon.com

Proposed Method:

Architecture: Since the two main factors in performance of different models are proper ranking and grouping, we propose a semantically-guided affinity-based approach. These approaches have been used successfully for sparse depth completion [16] and they tend to generalize better across multiple datasets. Our approach combines this methodology with the rich geometric features extracted by the depth feature encoder used in SA-net [7]. There are two parts to the process: **Affinity estimation**, and **Affinity Propagation**. We start by estimating the pairwise affinity between neighboring pixels. In order to achieve this, we use the semantic feature map extracted from a pre-trained semantic segmentation module [17] along with the original high-resolution RGB image. These feature maps are used to estimate the affinity, and once affinity is estimated, we use Affinity Propagation to selectively propagate the geometric features. These propagated geometric features are finally used to predict the final output.

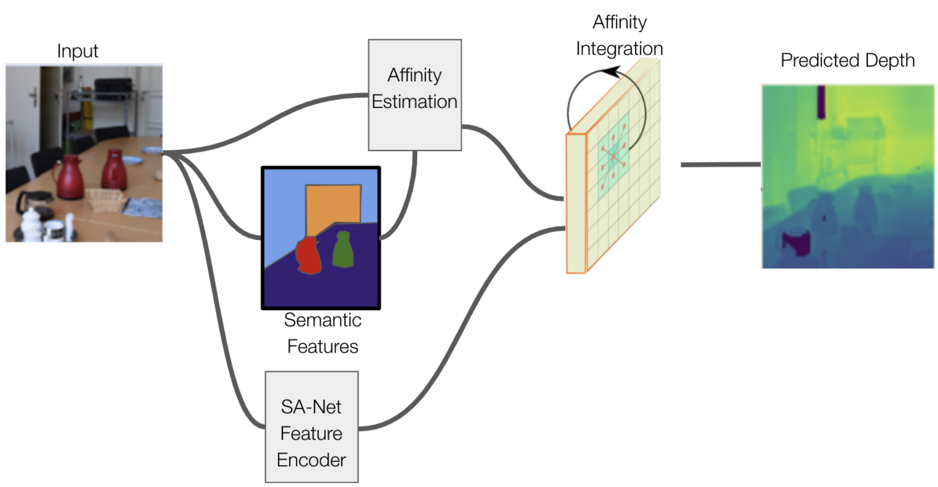
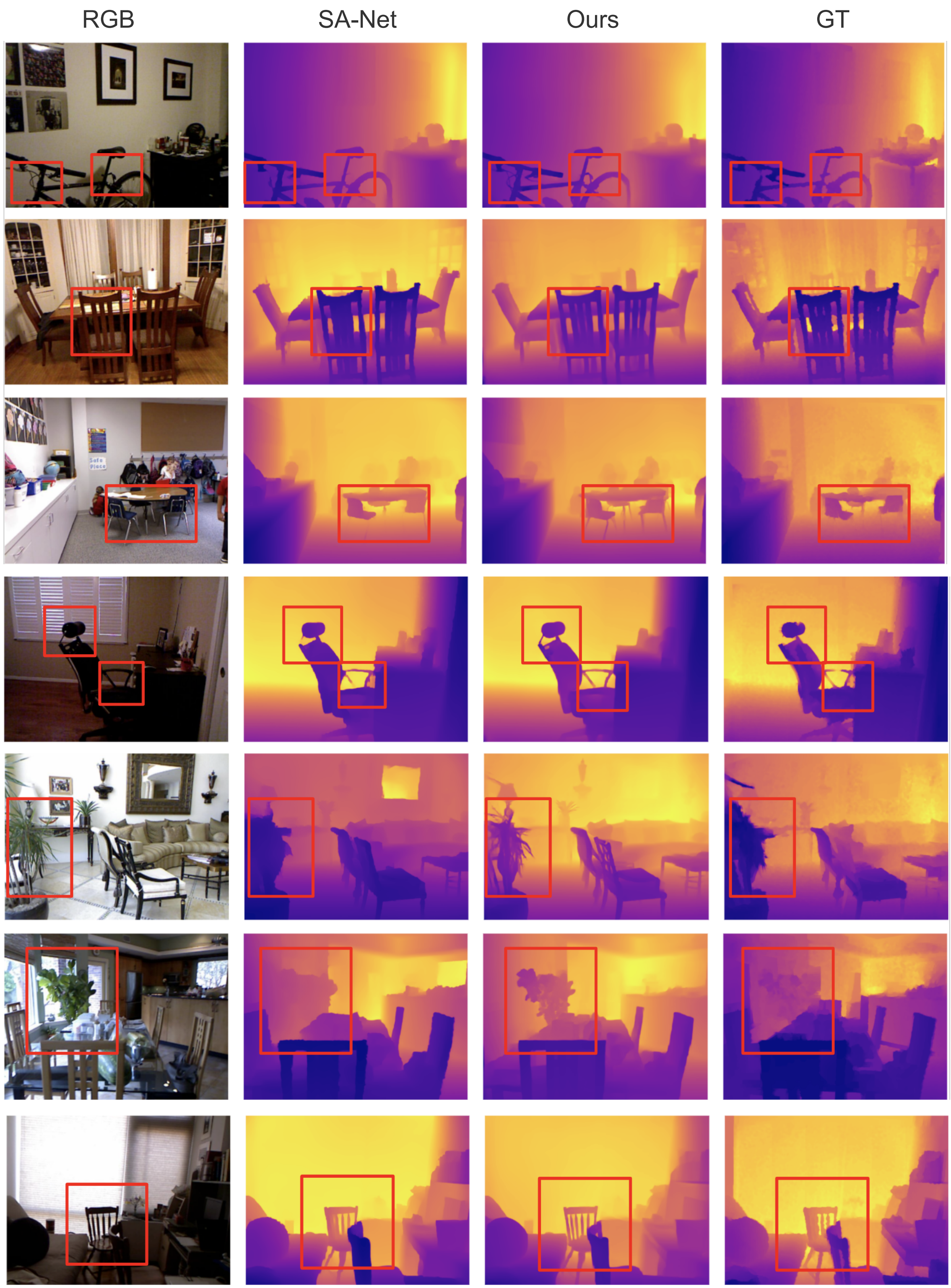


Figure 5: Proposed network architecture; we compute pairwise affinities using semantic features, and integrate the encoded features using this estimated affinity.

Dataset: Choice of dataset is crucial as RGBD datasets like NYUD tend to have poor information close to edges. We experimented with MegaDepth, HR-WSI, Hypersim, and the SA-Net, and in our ablation experiments, we found the combination of NYUD and the custom SA-net dataset to be the best combination, using Hypersim for fine-tuning.

Results: Our method generalizes better to the iBims-100 dataset, achieving **4%** better DBE completeness error, and **11%** better DBE accuracy than previous art even without a refinement stage. The following visualization shows our method is better at preserving thin structures which are parts of the foreground objects. A compilation of visual results can be found here.



RGB
SA-Net
Ours
GT

Citations:

[1] Tsung-Wei Ke, Jyh-Jing Hwang, Ziwei Liu, Stella X. Yu: **Adaptive Affinity Fields for Semantic Segmentation, ECCV 2018**
[2] Jin Han Lee, Myung-Kyu Han, Dong Wook Ko, Il Hong Suh: **From Big to Small: Multi-Scale Local Planar Guidance for Monocular Depth Estimation, arXiv**
[3] Michael Ramamonjisoa, Vincent Lepetit: **SharpNet: Fast and Accurate Recovery of Occluding Contours in Monocular Depth Estimation, ICCV 2019 workshops**
[4] Wei Yin, Yifan Liu, Chunhua Shen, Youliang Yan: **Enforcing geometric constraints of virtual normal for depth prediction, ICCV 2019**
[5] Alireza Fathi, Zbigniew Wojna, Vivek Rathod, Peng Wang, Hyun Oh Song, Sergio Guadarrama, Kevin P. Murphy: **Semantic Instance Segmentation via Deep Metric Learning,**
[6] Jyh-Jing Hwang, Tsung-Wei Ke, Jianbo Shi, Stella X. Yu: **Adversarial Structure Matching for Structured Prediction Tasks, CVPR 2019**
[7] Simon Niklaus, Long Mai, Jimei Yang, Feng Liu: **3D Ken Burns Effect from a Single Image**
[8] Michael Ramamonjisoa, Yuming Du, Vincent Lepetit: **Predicting Sharp and Accurate Occlusion Boundaries in Monocular Depth Estimation Using Displacement Fields**
[9] Pierluigi Zama Ramirez, Matteo Poggi, Fabio Tosi, Stefano Mattoccia, Luigi Di Stefano: **Geometry meets semantics for semi-supervised monocular depth estimation**
[10] Po-Yi Chen, Alexander H. Liu, Yen-Cheng Liu, Yu-Chiang Frank Wang: **Towards Scene Understanding:Unsupervised Monocular Depth Estimation with Semantic-aware Representation**
[11] Peng Wang, Xiaohui Shen, Zhe Lin, Scott Cohen, Brian Price, Alan Yuille: **Towards Unified Depth and Semantic Prediction from a Single Image**
[12] Jianbo Jiao, Ying Cao, Yibing Song, and Rynson Lau: **Look Deeper into Depth: Monocular Depth Estimation with Semantic Booster and Attention-Driven Loss**
[13] Ke Xian, Jianming Zhang, Oliver Wang, Long Mai, Zhe Lin, and Zhiguo Cao: **Structure-Guided Ranking Loss for Single Image Depth Prediction**
[14] Zhenyu Zhang, Zhen Cui, Chunyan Xu, Yan Yan, Nicu Sebe, Jian Yang: **Pattern-Affinitive Propagation across Depth, Surface Normal and Semantic Segmentation**
[15] René Ranftl, Katrin Lasinger, David Hafner, Konrad Schindler, Vladlen Koltun: **Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-shot Cross-dataset Transfer**
[16] Xinjing Cheng, Peng Wang, Ruigang Yang: **Learning Depth with Convolutional Spatial Propagation Network**
[17] Alexander Kirillov, Yuxin Wu, Kaiming He, Ross Girshick: **PointRend: Image Segmentation as Rendering**