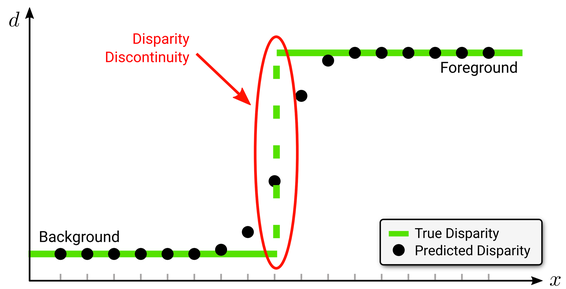
SMD-Nets: Stereo Mixture Density Networks

Fabio Tosi[1], Yiyi Liao[2], Carolin Schmitt[2], Andreas Geiger[2]
[1]University of Bologna, [2]Autonomous Vision Group, MPI-IS/University of Tübingen

Introduction

Problems:

- Stereo matching networks poorly reconstruct object boundaries due to smoothness bias, causing **bleeding artifacts** in 3D point clouds.
- Existing methods are limited to **discrete predictions** at pixel locations of a fixed resolution image grid.
- Absence of large-scale, realistic and high-resolution stereo datasets with pixel-accurate ground truth.



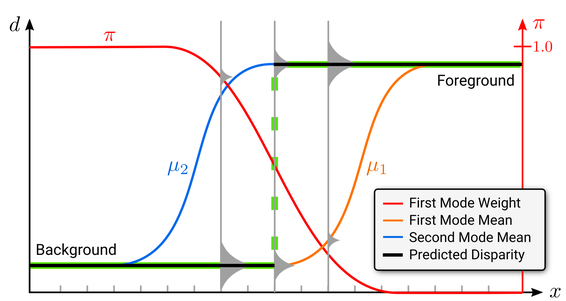
Stereo Regression Network

Goals:

- Predicting accurate and sharp depth boundaries.
- High-resolution outputs with constant memory.

Key Contributions:

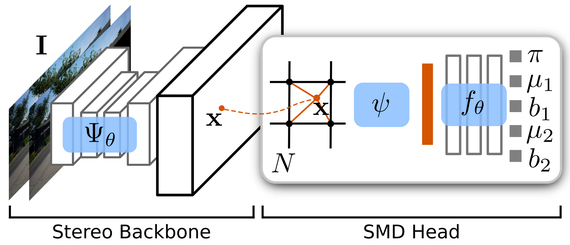
- A **bimodal mixture distribution** as output representation such that sharp discontinuities can be regressed.
- A **continuous function formulation** aimed at estimating disparities at arbitrary spatial resolution with constant memory footprint.
- A new large-scale synthetic binocular stereo dataset with ground truth disparities at **8Mpx resolution**.

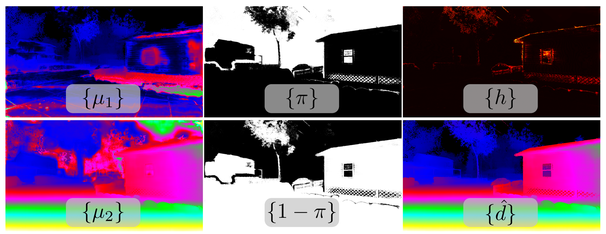


Stereo Mixture Density Network

Our Method

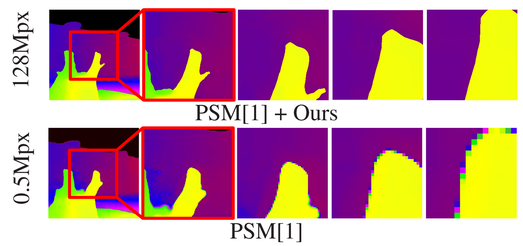
- SMD-Nets predict a **bimodal** (Laplacian) **mixture distribution** which allows to accurately capture **uncertainty** close to depth discontinuities. By doing so, **sharp discontinuities** can be regressed despite the underlying neural networks make smooth predictions.
- Our framework use a (2D or 3D) convolutional **stereo backbone** in combination with a shallow **multi-layer perceptron head** that regresses the five distribution parameters from interpolated features.





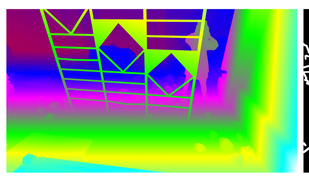
Stereo Super-Resolution

- Our **continuous formulation** allows us to exploit ground truth at higher resolution than the input.

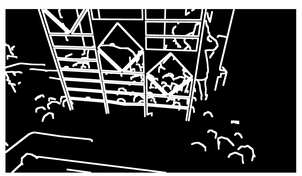


Sampling Strategy

- We adopt a **Depth Discontinuity Aware** sampling approach during training that favors points located near object boundaries.



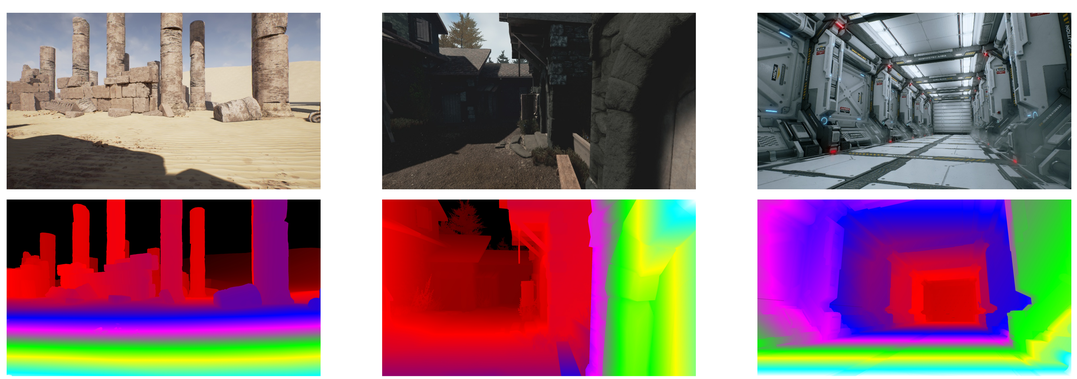
Ground Truth Disparity



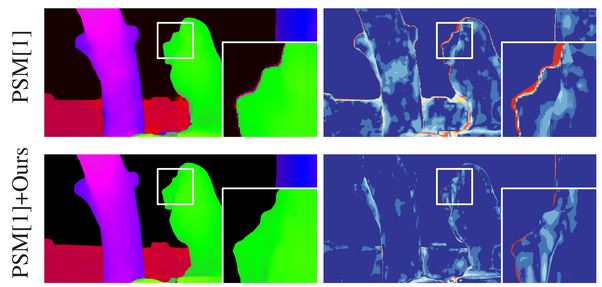
Boundary Mask

UnrealStereo4K Dataset

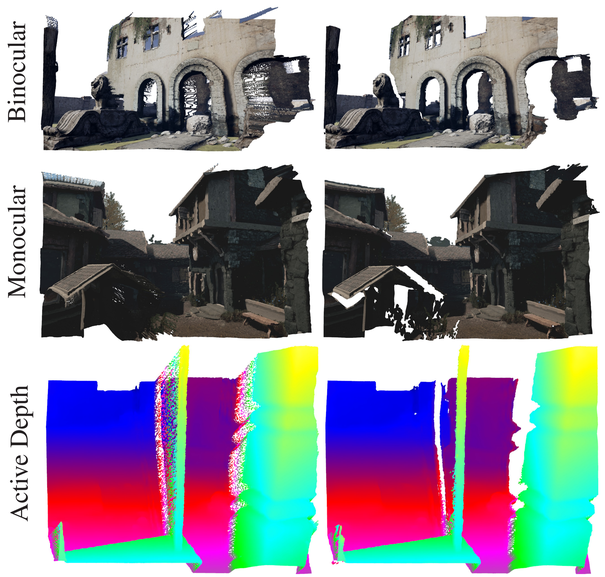
- We introduce a photo-realistic binocular stereo dataset at 3840×2160 resolution with pixel-accurate ground truth.



Qualitative Results



Point Cloud Visualization



Direct Regression | Bimodal

References

[1] Chang and Chen, "Pyramid stereo matching network", CVPR 2018

Links

Paper | Supplement | Code | Video