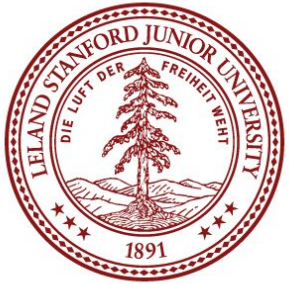


Analysis of Monocular & Stereo Depth Estimation Techniques

CS231A Final Project (Winter 2022)

Varo Ly, Rajeev Patwari

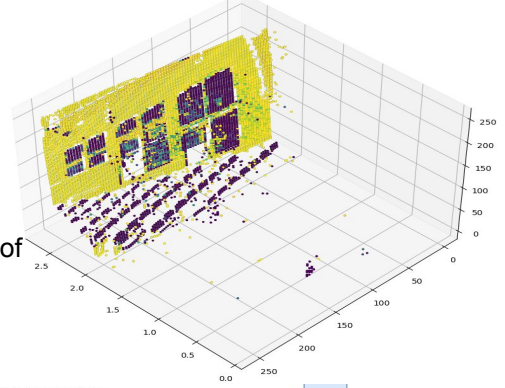
Background, Motivation & Technical Approach

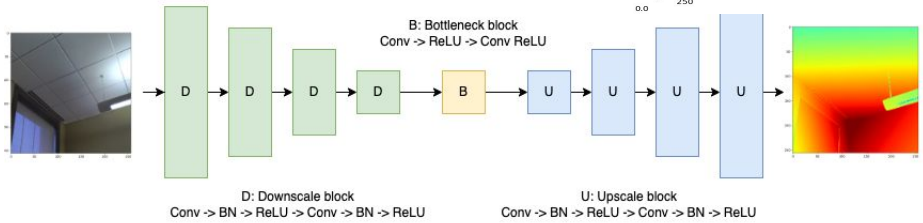
- Depth estimation is critical for autonomous navigation systems
- Fast scene understanding reduces decision making pipeline latency
- Our motivation
  - Analyze learning based Monocular and Stereo depth estimation methods
  - Reduce computational complexity of chosen models
  - Lower compute enables low power, higher frame-rate for battery operated systems
- Technical approach
  - Choose fastest models for both Monocular & Stereo depth estimation
  - Reproduce methods and perform ablation studies
  - Modify models to reduce computational complexity with least accuracy loss

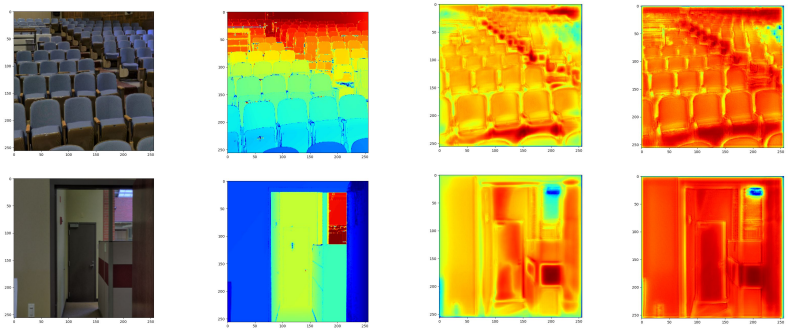
Monocular Depth Estimation

- Smaller version of FastDepth
- Datasets: KITTI, NYU v2, DIODE
  - Used validation dataset of DIODE
  - Train - test split ~ 90%, 10% on 1000 samples of size 256x256
- Metric
  - SSIM, Depth smoothness, L1 Loss
- Training converged within 30 epochs
- Current challenges
  - Used training dataset is small
  - Predicted depth is not exact as ground truth
  - More hyper-parameter search may be required

3D point cloud visualization of input image with input depth







Input | Ground Truth | Predicted (Loss: SSIM + L1 + Depth Smoothness) | Predicted (Loss: SSIM)

FastDepth: https://arxiv.org/pdf/1903.03273.pdf
Monocular Depth Estimation using Keras: https://keras.io/examples/vision/depth_estimation/

Stereo Vision Depth Estimation - AnyNet

- AnyNet is 2nd fastest stereo model
- Dataset used: Monkaa (full)
  - Left/Right RGB and disparity
  - 8664 samples
- Method
  - U-Net + 3D cost volume + SPNet
  - Can truncate at different stages
  - Metric: Smooth L1 loss
- Current challenges
  - Training is compute intensive
  - Analyze outputs with SPNet included

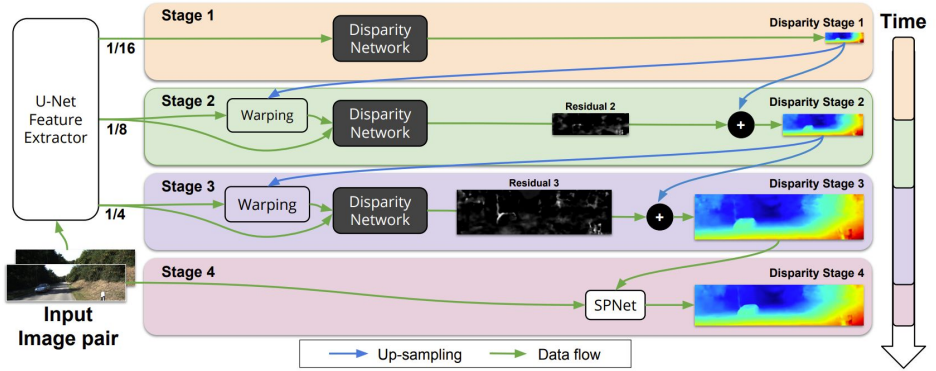
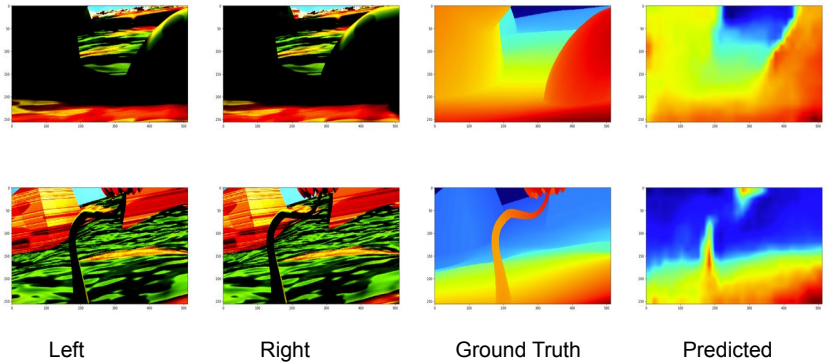


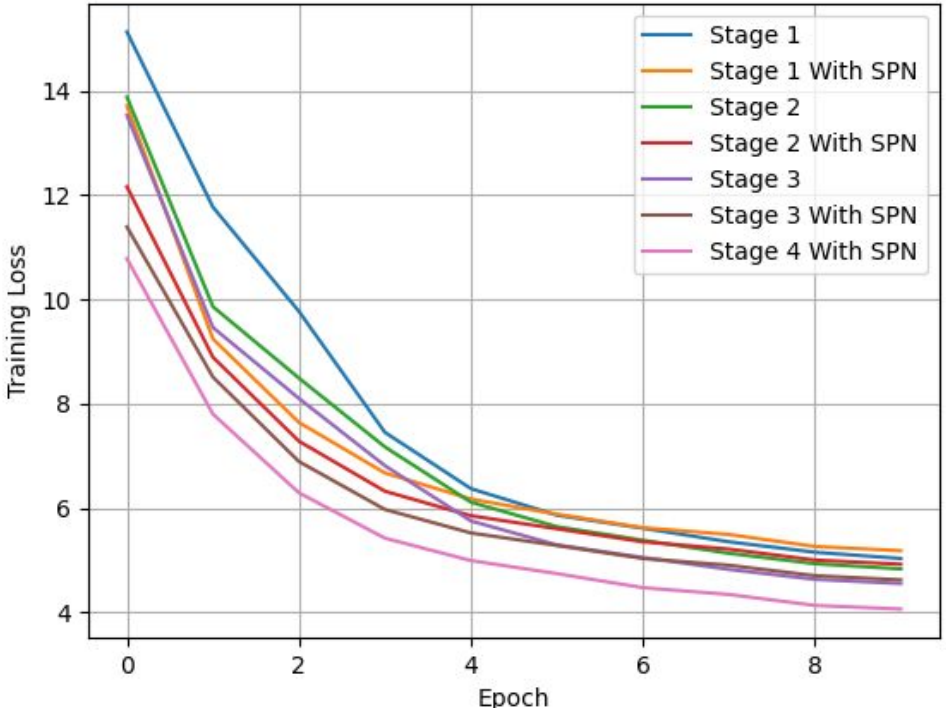
Fig. 2: Network structure of AnyNet.

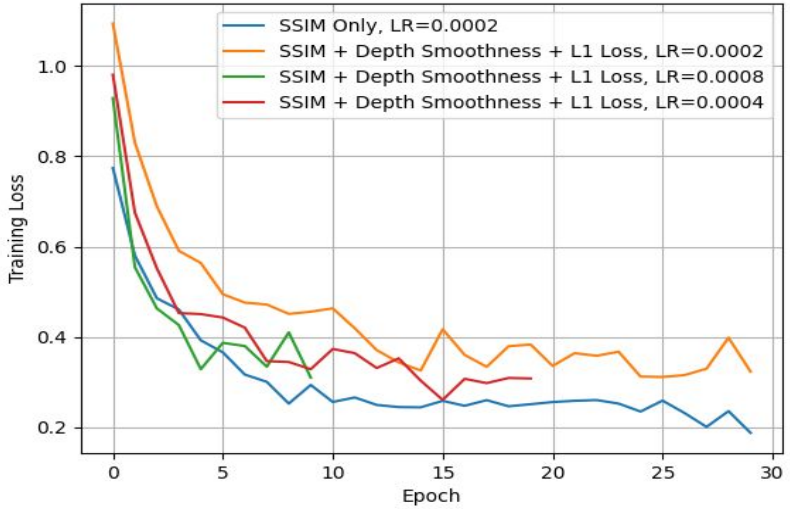


Survey of Stereo Models: https://arxiv.org/pdf/2006.02535.pdf
AnyNet: https://arxiv.org/pdf/1810.11408.pdf

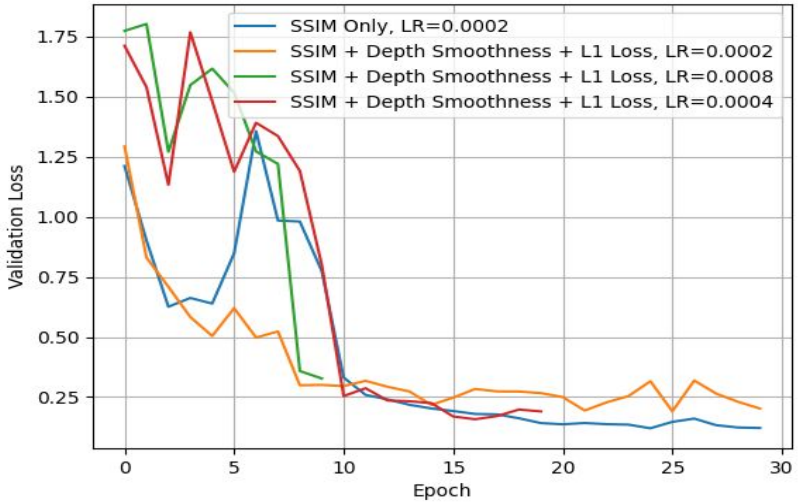
Ablation Studies

Stereo Vision Model





Monocular Model



Experiments in progress

- Monocular:
  - Improve model performance
  - Modify autoencoder structure to reduce computational complexity
  - Use larger dataset for Monocular Depth method
- Stereo:
  - Visualize results of AnyNet with Spatial Propagation Network
  - Modify Spatial Propagation Network configuration to reduce computational complexity
  - Measure test loss
- Use same data on Monocular as used in Stereo method