Residual-Learning-Based Single-Image Depth Estimation

Yokila Arora, Ishan Patil, Thao Nguyen

Stanford University

Objective

- Learning the physical geometry of a scene from a single monocular image without any environmental assumptions is a challenging problem
- Pixelwise depth regression is crucial for automated systems where depth sensing is not available
- Implemented a convolutional residual network architecture in PyTorch, based on the model proposed by Laina et al.

Dataset

- **NYU Depth v2, Make3D Laser+Image datasets**
  - Sample: 1449 scenes from NYU Depth and 534 from Make3D in total, each with corresponding ground-truth depth map
- **Suitability**
  - RGB-D data for both indoor (NYU Depth) and outdoor (Make3D) scenes
- **Preprocessing**
  - Standard resizing to 304x228 pixels, OR
  - Data augmentation: random scaling, rotation, center-crop and horizontal flip as proposed by Eigen et al.

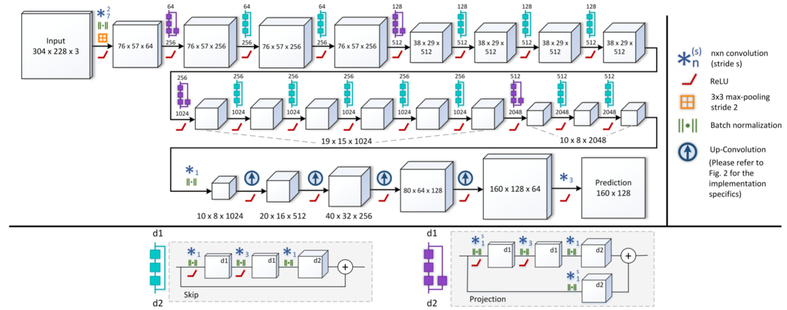
Background

- Eigen et al. were the first to use CNNs for depth estimation: predict a coarse global output and then a finer local network
- Laina et al. explored CNNs pre-trained with AlexNet, VGG16 and ResNet

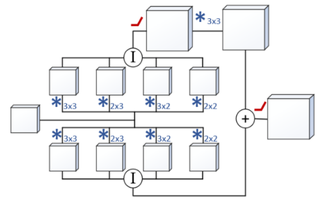
→ **Improvements:** No post-processing and FC layers needed, deeper network without vanishing gradient problem, residual & skip layers, efficient up-sampling of feature maps (to increase accuracy)

Methods & Models

- **Laina et al. Model:**



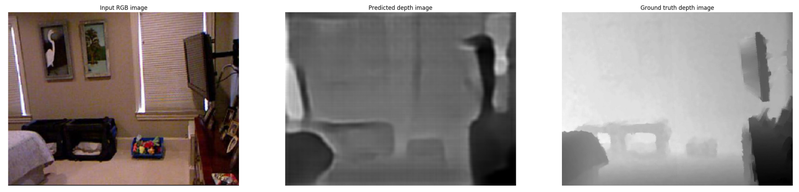
  - Skip and Projection layers
  - Batch normalization
  - Efficient up-projection blocks
  - L2 loss, BerHu loss



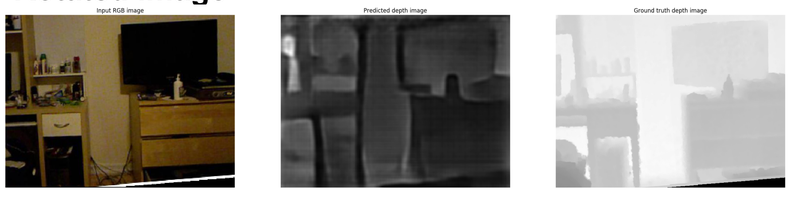
Efficient up-convolutional blocks

Results

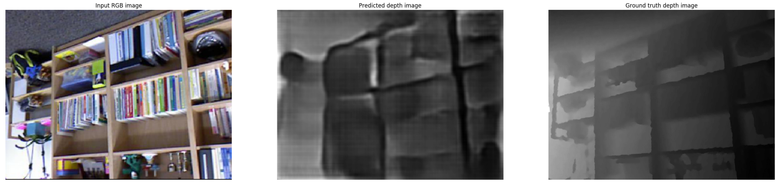
- **Standard image**



- **Rotated image**



- **Flipped image**



Analysis & Observation

- The network is good at deducing the depth of object close to the camera but does poorly when the same object is far away → an useful real-world application may be indoor obstacle detection for automaton

Conclusion

- Fully convolutional network, single-step process
- Fewer parameters and training samples required
- Novel & efficient up-sampling blocks → high-resolution outputs

Future Work

- Fine tuning hyperparameters to get closer results to paper
- Use transfer learning and only tune the last few layers
- Combine depth information with other tasks (object detection, segmentation, etc.)

References

1.D. Eigen, C. Puhrsch, and R. Fergus. Depth map prediction from a single image using a multi-scale deep network.
2. I. Laina, C. Rupprecht, V. Belagiannis, F. Tombari, and N. Navab. Deeper depth prediction with fully convolutional residual networks