

RGB Image Depth Estimation

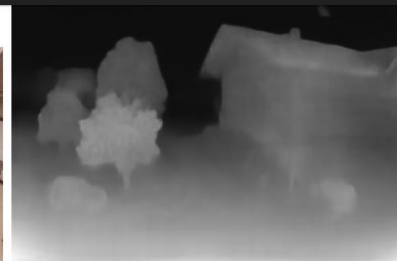
Mentor:
Nikola Velickovic



Done by:
Mihailo Grbic
Branko Grbic

Introduction

- Fundamental task in computer vision
- Mostly based on CNN
- Self-supervised methods appearing



Related work

- Towards Good Practice for CNN-Based Monocular Depth Estimation
 - Different components/factors for monocular depth prediction
 - Analyzing the errors made by different models
 - State-of-the-art results on NYU Depth v2 and competitive results on KITTI dataset
- Estimating Monocular depth using cycle GAN and segmentation
 - (Our starting goal)

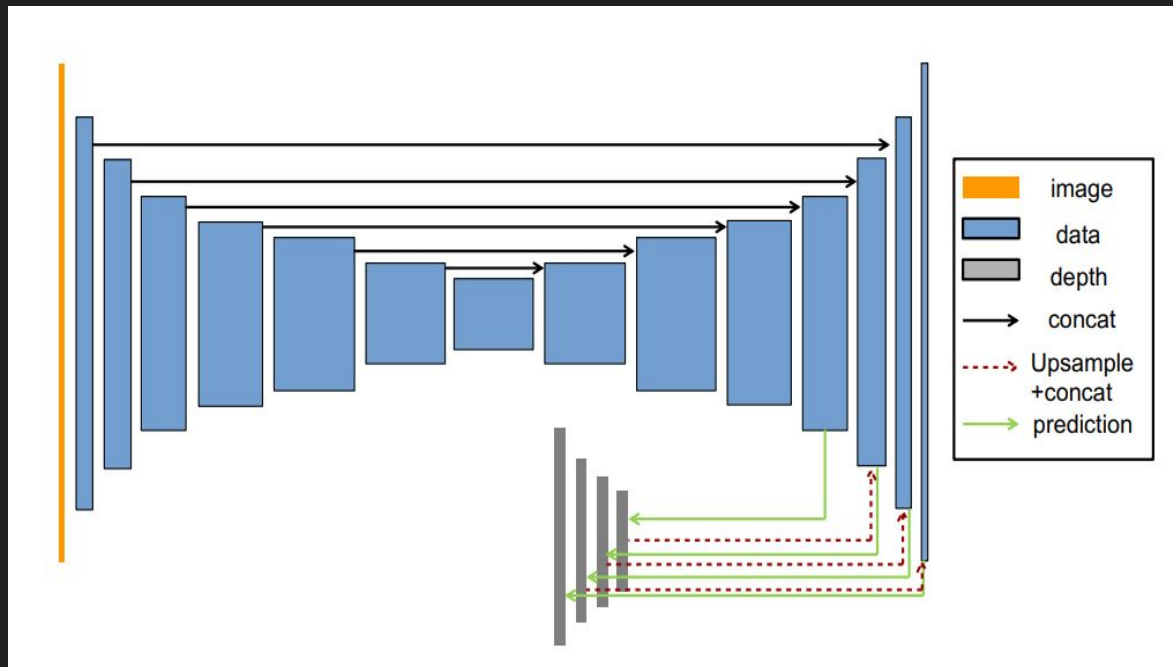
Data Preprocessing

- 1449 images and depths
- Train / Val / Test split (80%/10%/10%)
- Normalization using training dataset statistics
- Rescaling and cropping (both train & test)
- Extending Dataset
 - Random horizontal/vertical flip (train only)
 - Random rotation (train only)
 - Random crop (train only) / Center crop
 - MixUp
 - Blending

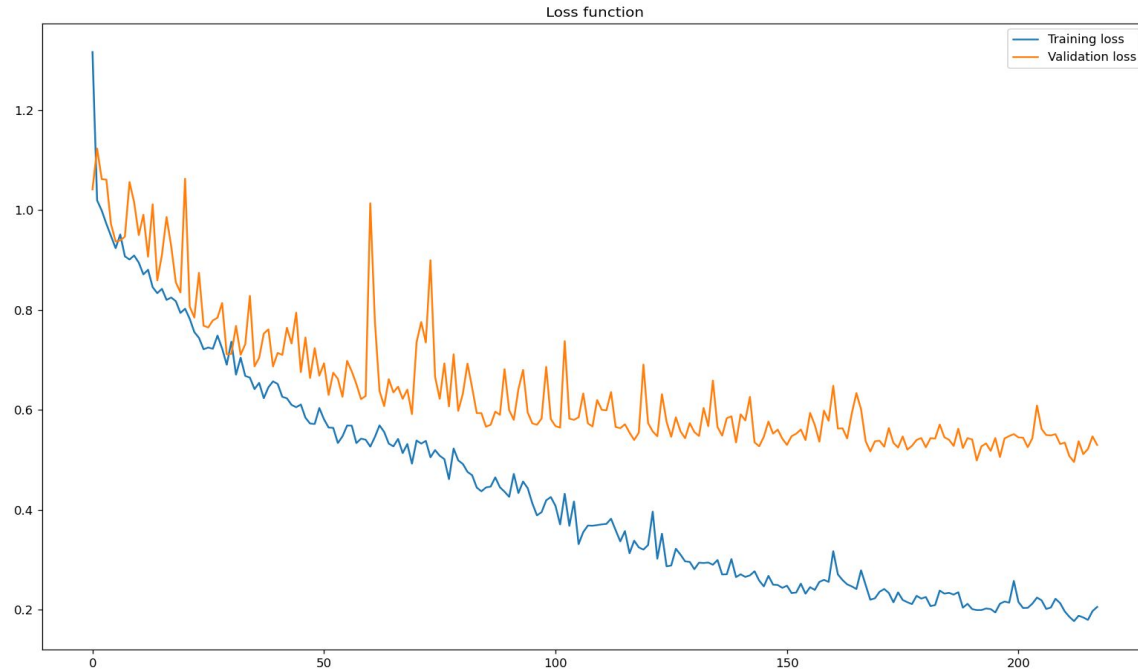


Model

- VGG-16 with BN (Encoder)
- DispNet (Decoder)
- NYUv2 (Dataset)
- Epochs: ~230
- Learning Rate: 0.001
- Batch Size: 32
- Loss: L1Smooth
- Image size: 256x352
- Optimizer: AdamW
- Activation: LReLU

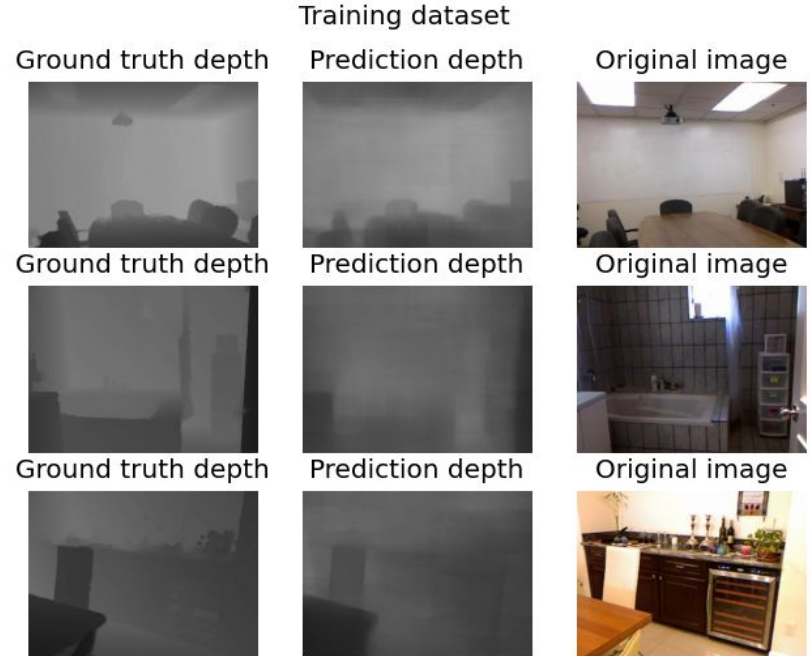


Results (loss function)



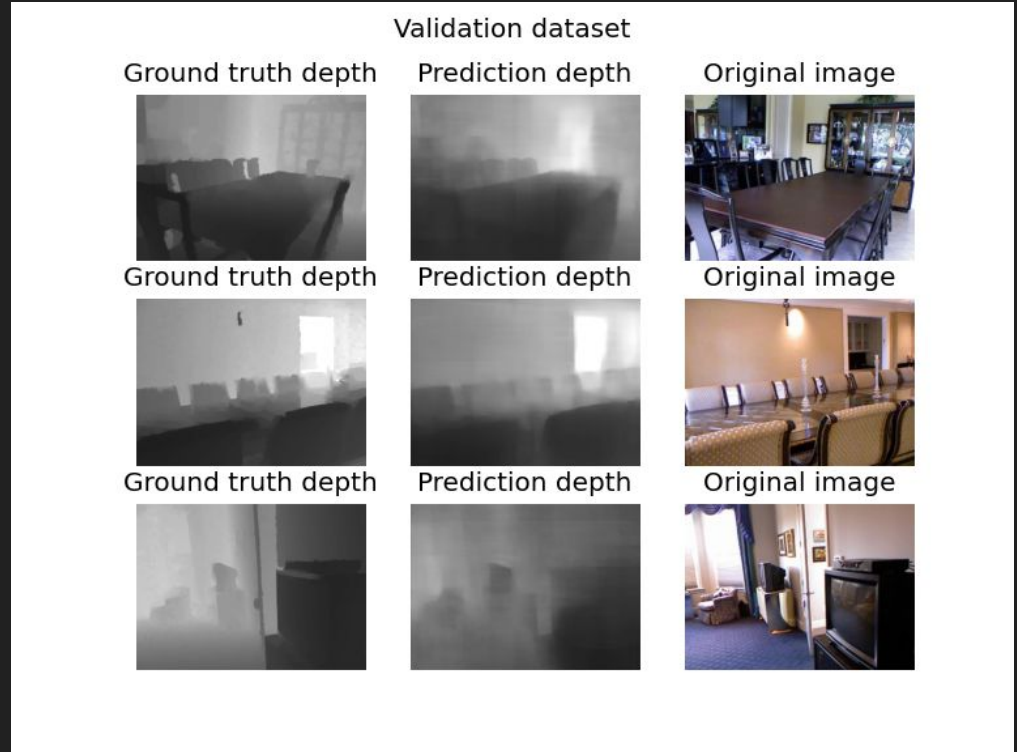
Results (Training)

- Photometric loss: 0.1231
- Smooth loss: 0.0636
- Overall loss: 0.1231
- Mean absolute error: 0.1231
 - Standard Deviation: 0.1072
- Mean Square Error: 0.3263
 - Standard deviation: 0.1305



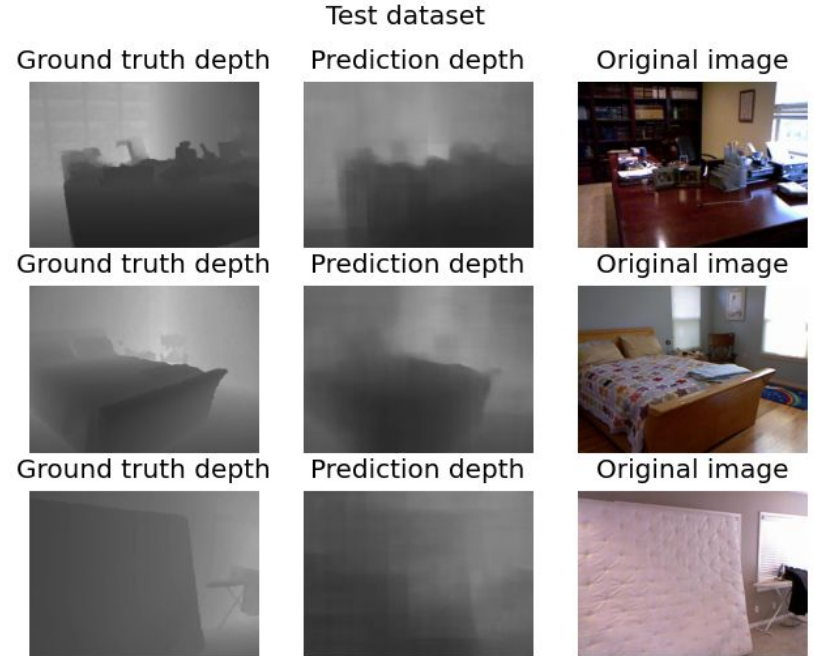
Results (Validation)

- Photometric loss: 0.4389
- Smooth loss: 0.0604
- Overall loss: 0.4389
- Mean absolute error: 0.4389
 - Standard Deviation: 0.3115
- Mean Square Error: 0.6578
 - Standard deviation: 0.3218



Results (Test)

- Photometric loss: 0.4423
- Smooth loss: 0.0603
- Overall loss: 0.4424
- Mean absolute error: 0.4423
 - Standard Deviation: 0.3160
- Mean Square Error: 0.6752
 - Standard deviation: 0.3480



Results (Best training sample)

Best sample in training dataset (loss = 0.0118)

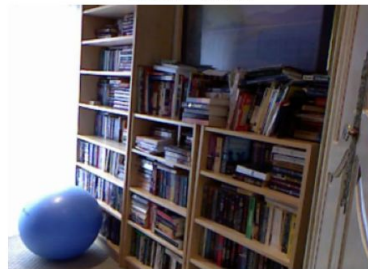
Ground truth depth



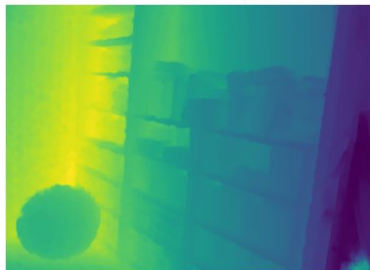
Prediction depth



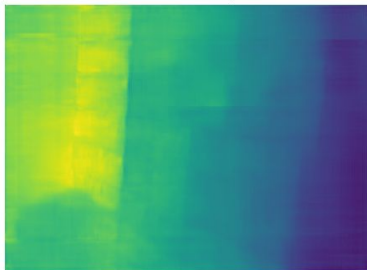
Original image



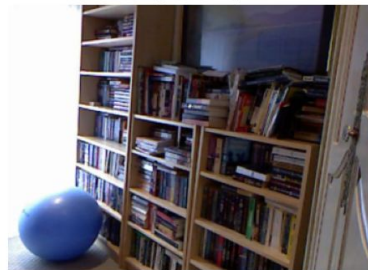
Ground truth depth



Prediction depth



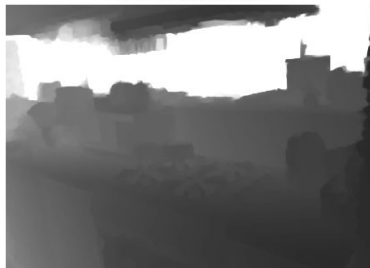
Original image



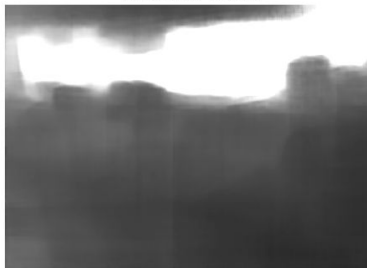
Results (Worst training sample)

Worst sample in training dataset (loss = 0.7317)

Ground truth depth



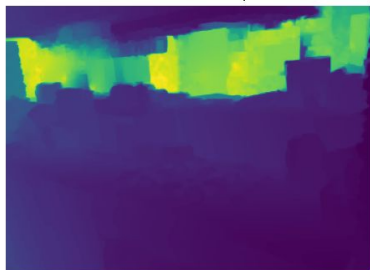
Prediction depth



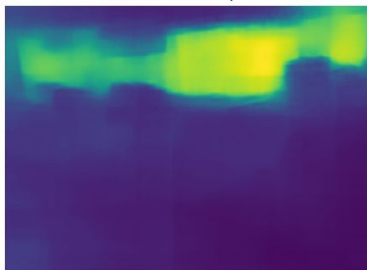
Original image



Ground truth depth



Prediction depth



Original image



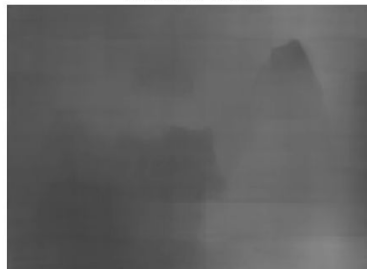
Results (Best validation sample)

Best sample in validation dataset (loss = 0.0419)

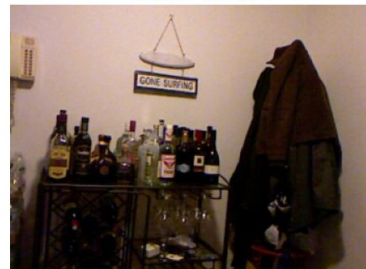
Ground truth depth



Prediction depth



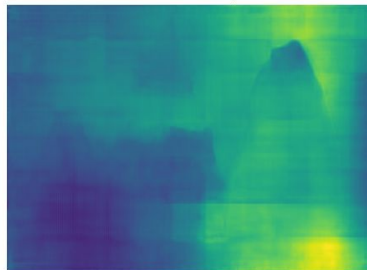
Original image



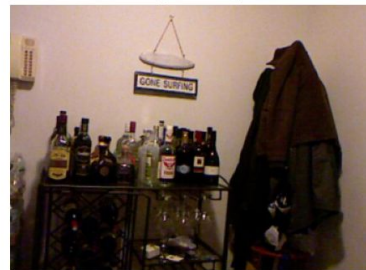
Ground truth depth



Prediction depth



Original image



Results (Worst validation sample)

Worst sample in validation dataset (loss = 1.5702)

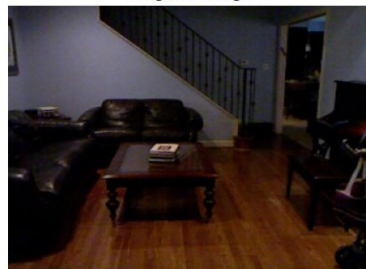
Ground truth depth



Prediction depth



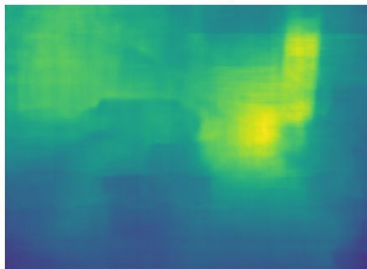
Original image



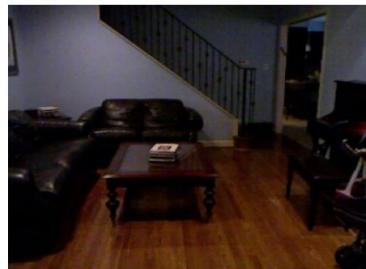
Ground truth depth



Prediction depth



Original image



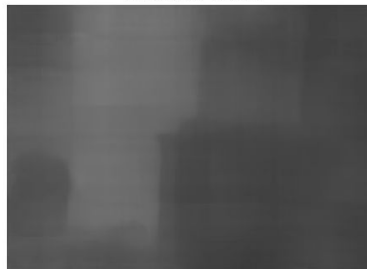
Results (Best test sample)

Best sample in test dataset (loss = 0.0387)

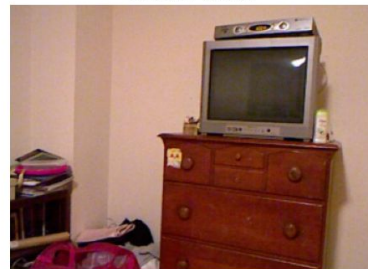
Ground truth depth



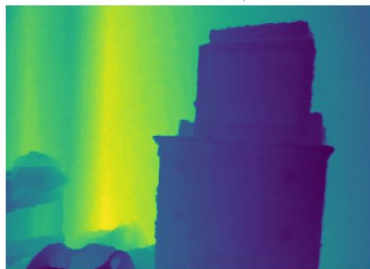
Prediction depth



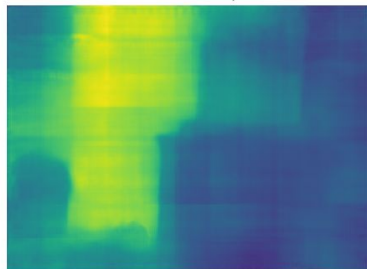
Original image



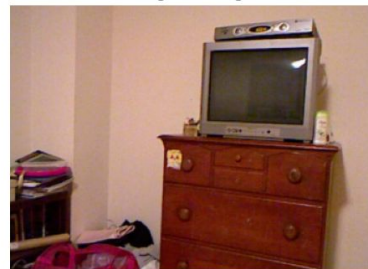
Ground truth depth



Prediction depth



Original image



Results (Worst test sample)

Worst sample in test dataset (loss = 1.7274)

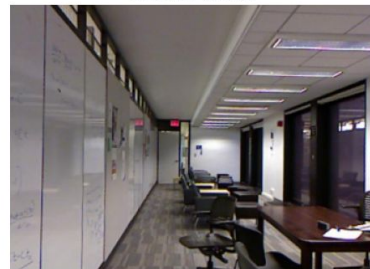
Ground truth depth



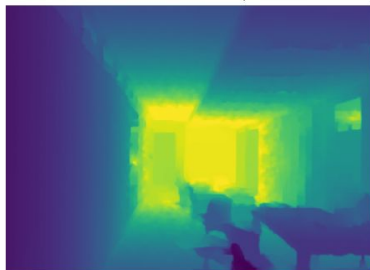
Prediction depth



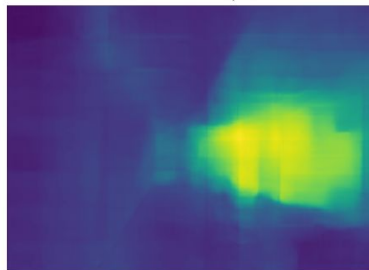
Original image



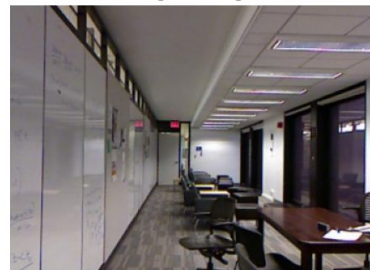
Ground truth depth



Prediction depth



Original image



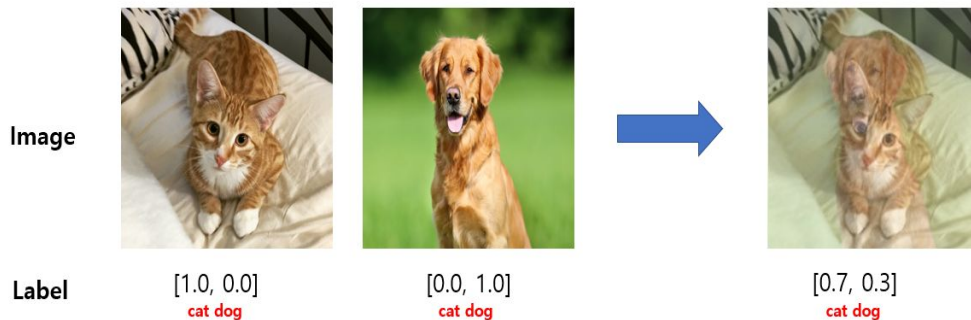
What didn't work?

- Freezed encoder
- LR scheduler
- L2 and Behru loss
- MixUp
- Blending
- Smaller images (?)

MixUp augmentation technique

$$\begin{aligned}\hat{x} &= \lambda x_i + (1 - \lambda)x_j, \\ \hat{y} &= \lambda y_i + (1 - \lambda)y_j,\end{aligned}$$

where $\lambda \in [0, 1]$ is a random number

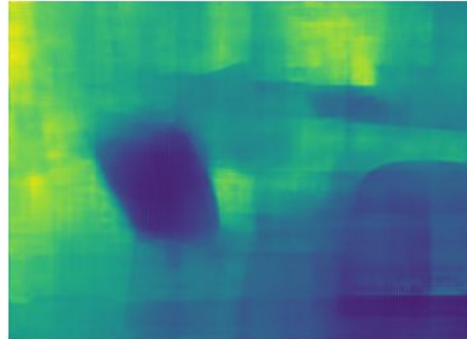


What's next?

- Bigger dataset
- Experimenting with alternative model architectures
- Tweaking parameters

Sample image

Depth prediction



Original image



Thank you!

Useful links:

[Towards Good Practice for CNN-Based Monocular Depth Estimation](https://openaccess.thecvf.com/content_WACV_2020/papers/Fang_Towards_Good_Practice_for_CNN-Based_Monocular_Depth_Estimation_WACV_2020_paper.pdf)

(https://openaccess.thecvf.com/content_WACV_2020/papers/Fang_Towards_Good_Practice_for_CNN-Based_Monocular_Depth_Estimation_WACV_2020_paper.pdf)

[NYUv2 Depth V2 Dataset](https://cs.nyu.edu/~silberman/datasets/nyu_depth_v2.html)

(https://cs.nyu.edu/~silberman/datasets/nyu_depth_v2.html)

[Our github repository with implementation and results](https://github.com/m-grbic/psiml7)

(<https://github.com/m-grbic/psiml7>)

[Depth estimation using cycle GAN](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7249099/)

(<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7249099/>)