Depth Estimation Using Deep Convolution Networks

Endsem Review

NN&DL

Done By:
Saikumar Dande
Chandravaran Kunjeti

1

INTRODUCTION TO DEPTH IMAGES

METHODS TO OBTAIN DEPTH

Deep Learning

Using Deep learning we can either use supervised learning or unsupervised learning to find the depth

STEREO

This works on the principle of reconstruction of an image using 2 images

IR

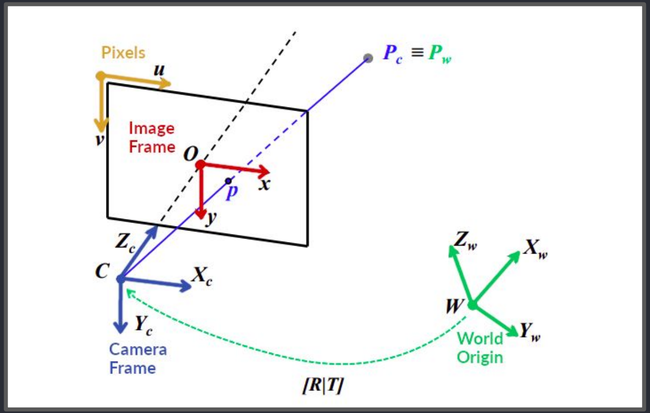
Uses a IR Transmitter and receiver to determine the depth

LIDAR

Uses light transmitted to calculate the distance usually ground truth images are found using this method.

Preprocessing - Camera Calibration

If we want understand the point in an Image we need to know what transforms are involved in the process, there are many types of parameters which are as follows

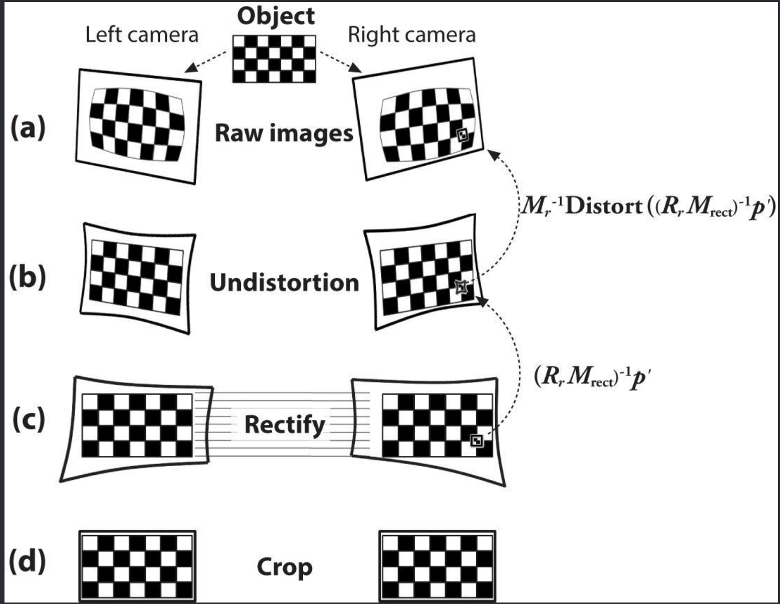


- Extrinsic parameters
  - Rotation matrix
  - Translation matrix

- Intrinsic parameters
  -Intrinsic matrix
  -Focal length
  -Optical centre

Preprocessing - Image Rectification

- After taking the images, they need to be rectified.
- Rectification process
  - Removes lens distortion.
  - Turns the stereo pair into standard form where images are perfectly aligned horizontally.

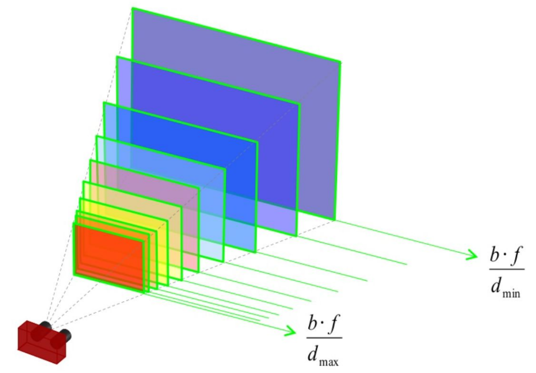


Disparity

$$\frac{b}{Z} = \frac{(b + x_T) - x_R}{Z - f} \rightarrow Z = \frac{b \cdot f}{x_R - x_T} = \frac{b \cdot f}{d}$$

Disparity = X_R-X_T

Depth



Depth = Base_length*Focal_length/Disparity

Base_length = Distance between left and right camera

Dataset

- NYU Dataset
  - The NYU-Depth V2 data set is comprised of video sequences from a variety of indoor scenes as recorded by both the RGB and Depth cameras from the Microsoft Kinect. It features:
  - **1449** densely labeled pairs of aligned RGB and depth images
  - **464** new scenes taken from 3 cities
  - **407,024** new unlabeled frames
  - Dataset is split into 1024 train, 201 test and 224 validation
- KITTI Dataset
  - The depth completion and depth prediction evaluation are related to our work published in Sparsity Invariant CNNs (THREEDV 2017)
  - It contains over 93 thousand depth maps with corresponding raw LiDaR scans and RGB images, aligned with the raw data of the KITTI dataset

2

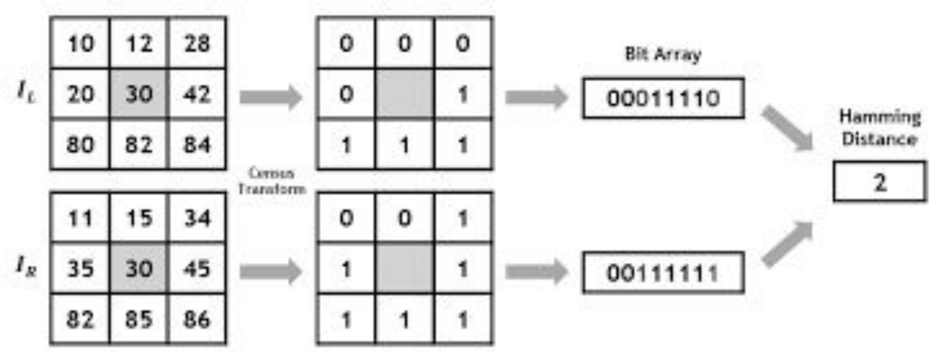
DIFFERENT MODELS

Previous year work

Sum of Absolute Difference(SAD)

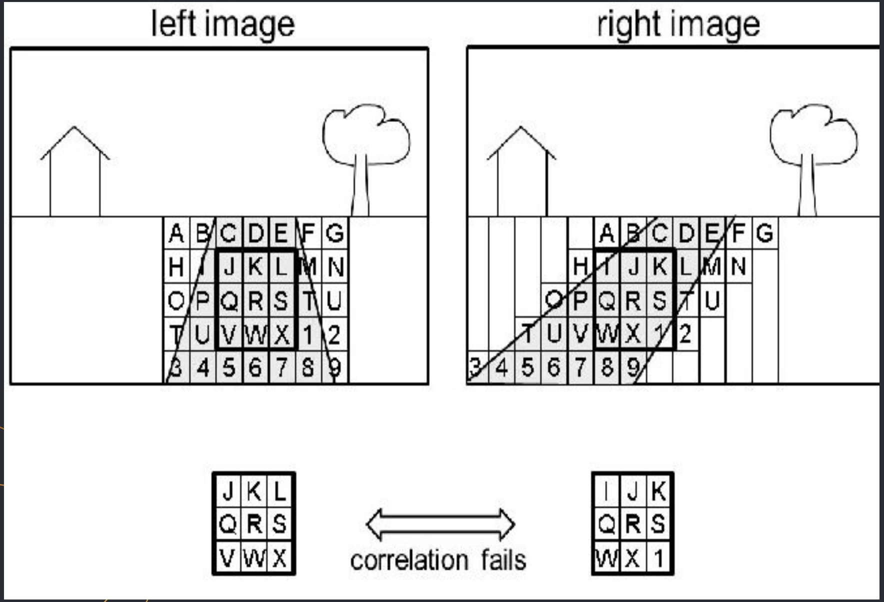
$$s = \sum_{(u,v)\in \mathbf{I}} |\mathbf{I}_1[u,v] - \mathbf{I}_2[u,v]|$$

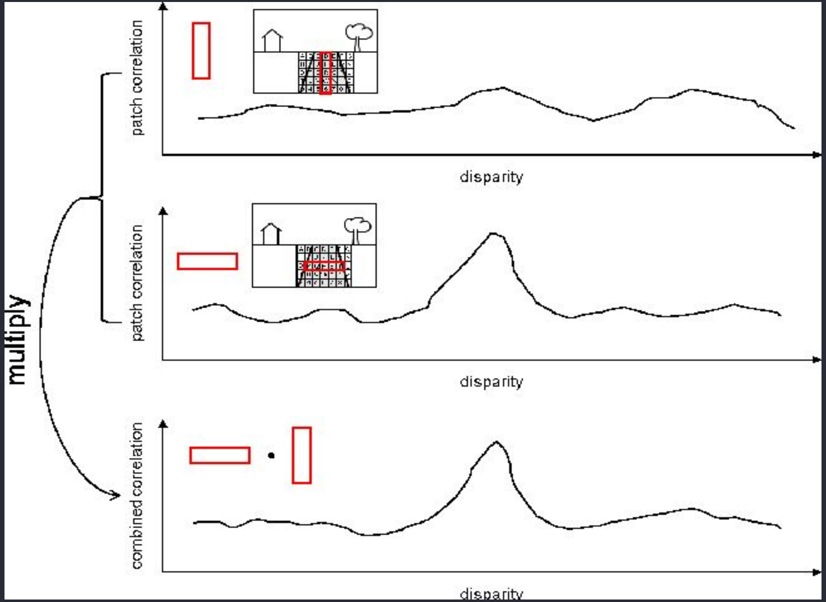
Census Transform(CT)



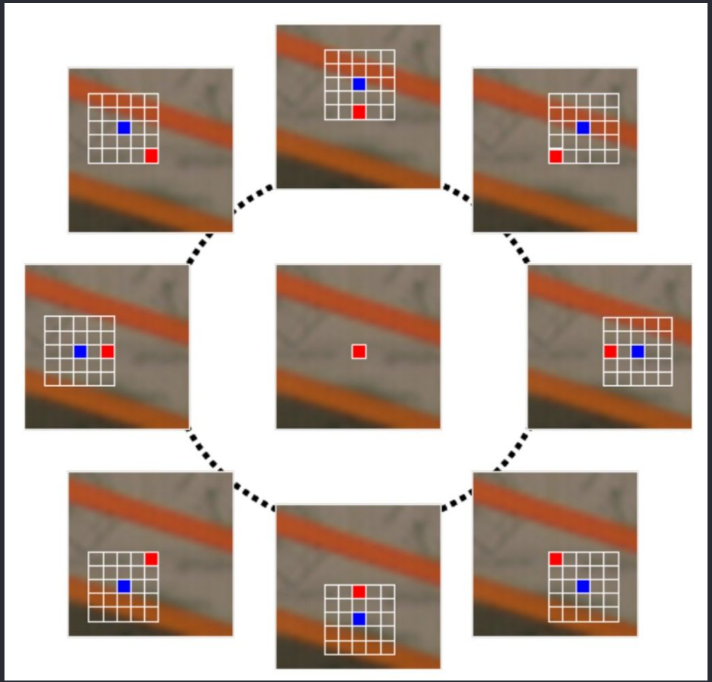
Multi Block Matching (MBM)

Reason to use Multi Block Matching

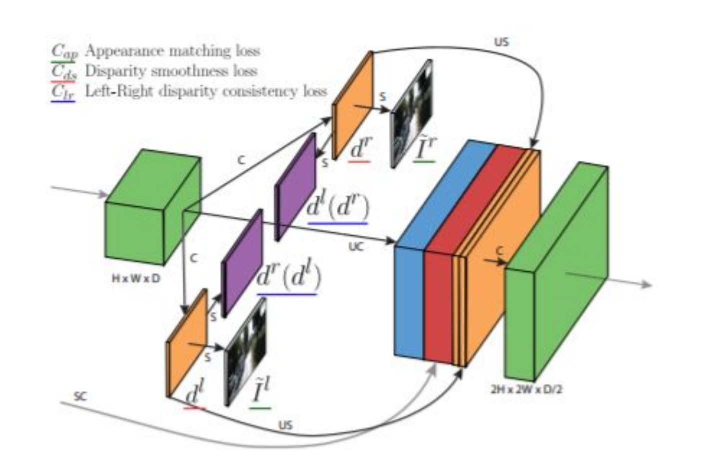


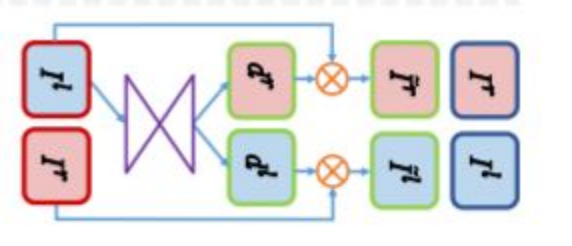


Locally Consistent Disparity Map Generation



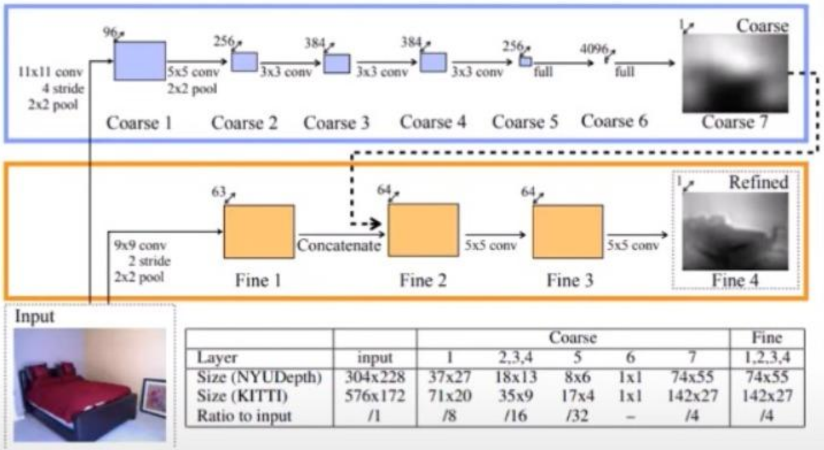
Unsupervised Monocular Depth Estimation with Left-Right Consistency





- ➢ This model is an unsupervised method of finding the depth of an Image. It uses a stereo image as an input
- ➢ The depth is found for a single image, and then the other image is used to evaluate it

Coarse & Fine Net



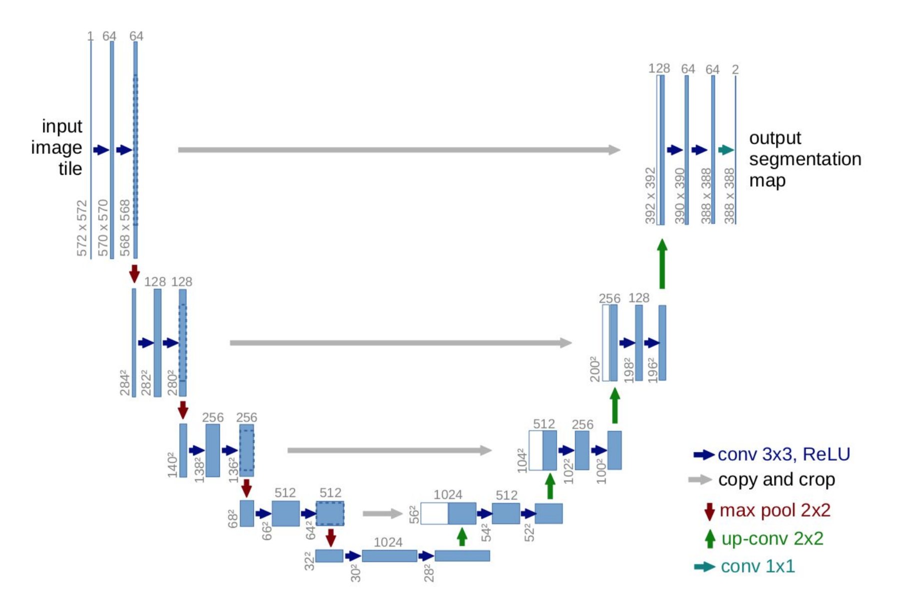
		Coarse					Fine
Layer	input	1	2,3,4	5	6	7	1,2,3,4
Size (NYUDepth)	304x228	37x27	18x13	8x6	1x1	74x55	74x55
Size (KITTI)	576x172	71x20	35x9	17x4	1x1	142x27	142x27
Ratio to input	/1	/8	/16	/32	–	/4	/4

Eigen *et al.* [1]

- ➢ This network has 2 parts: Coarse net and fine net.
- ➢ Coarse net predicts the depth of a scene at a global level.
- ➢ Fine net is used to refine within the local regions.

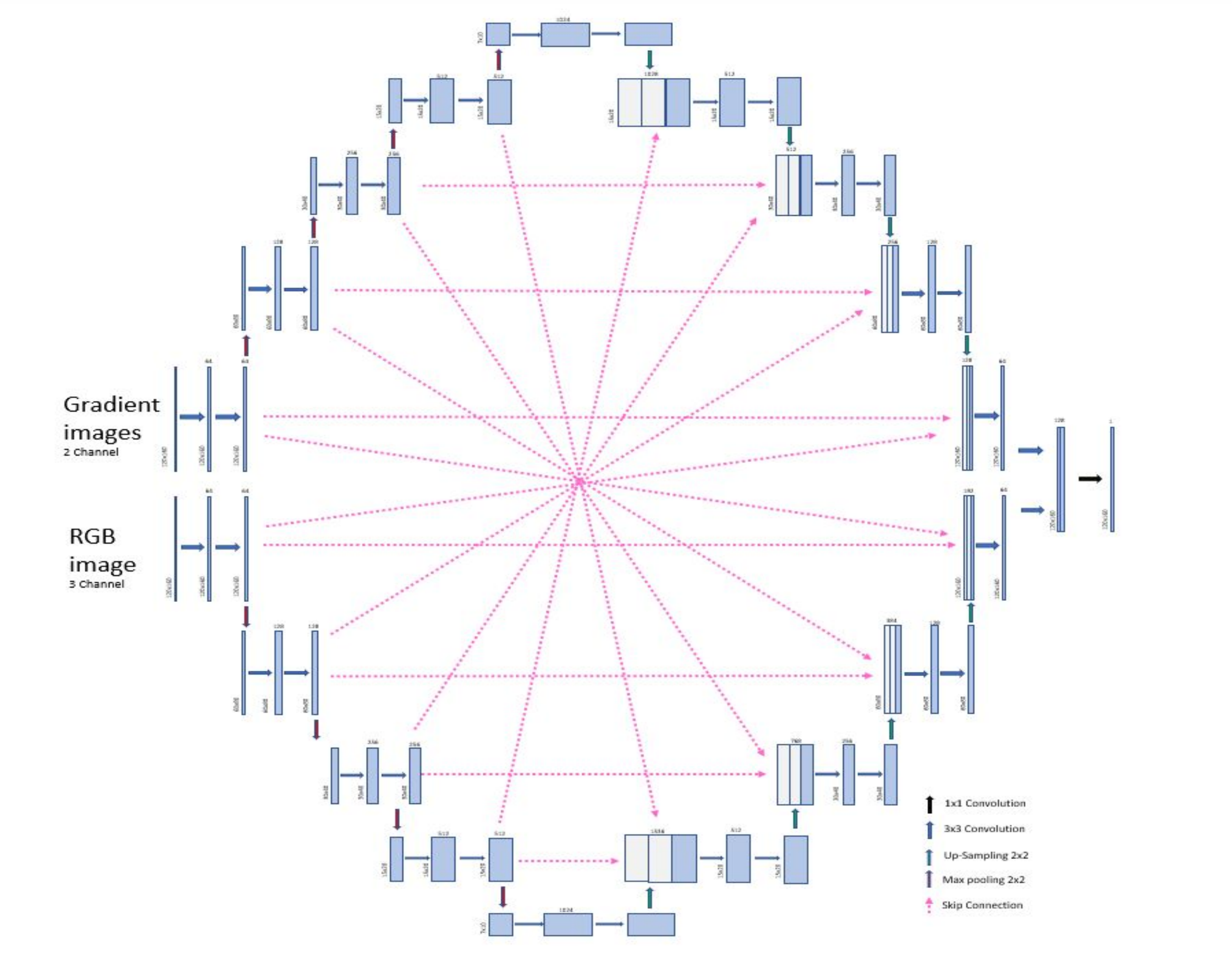
[1] Depth Map Prediction from a Single Image using a Multi-Scale Deep Network

UNet



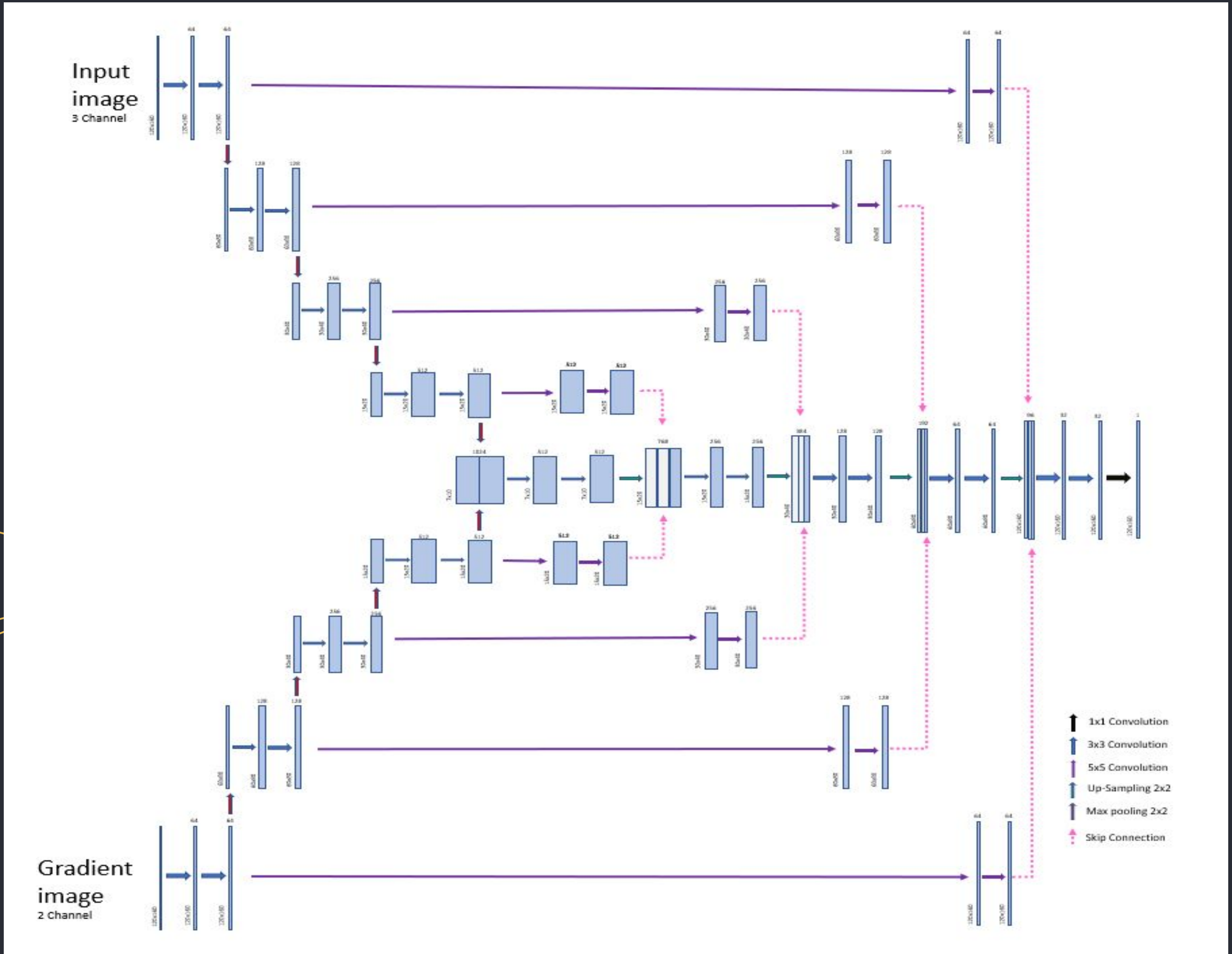
- ➢ Unet is a famous model that can be used for segmentation. It is also used for depth estimation.
- ➢ Unet consists of two parts: Encoder and a Decoder.

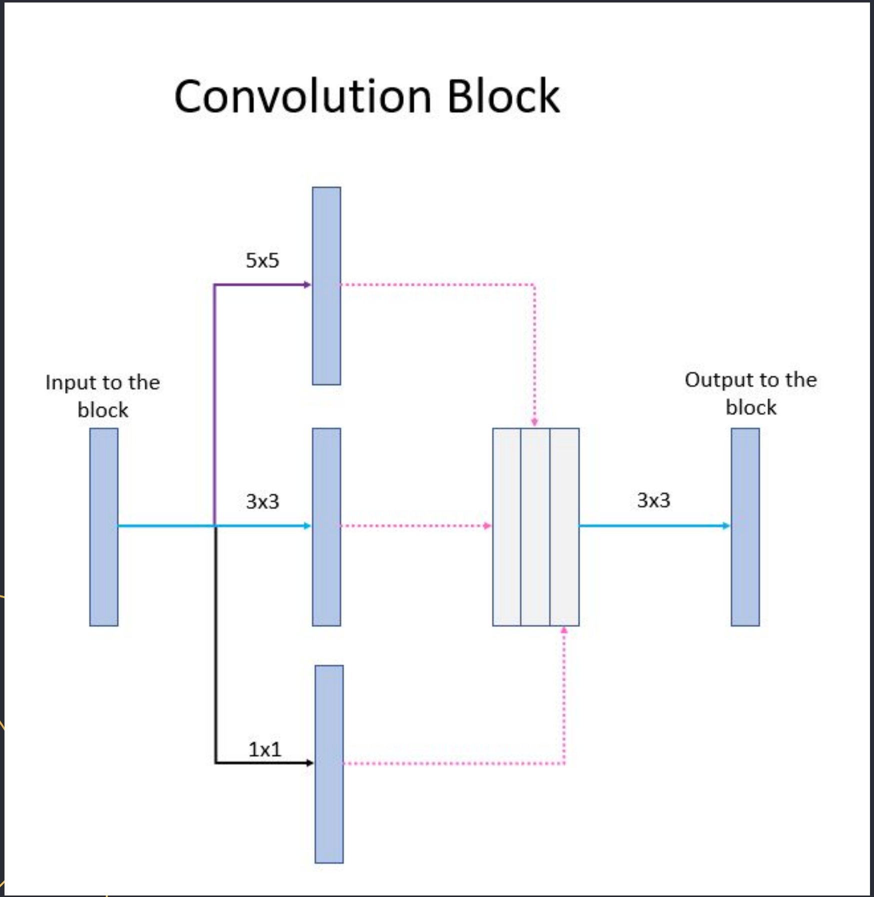
[2] U-Net: Convolutional Networks for Biomedical Image Segmentation



Proposed Model 1: Onet

Proposed Model 2: Ynet



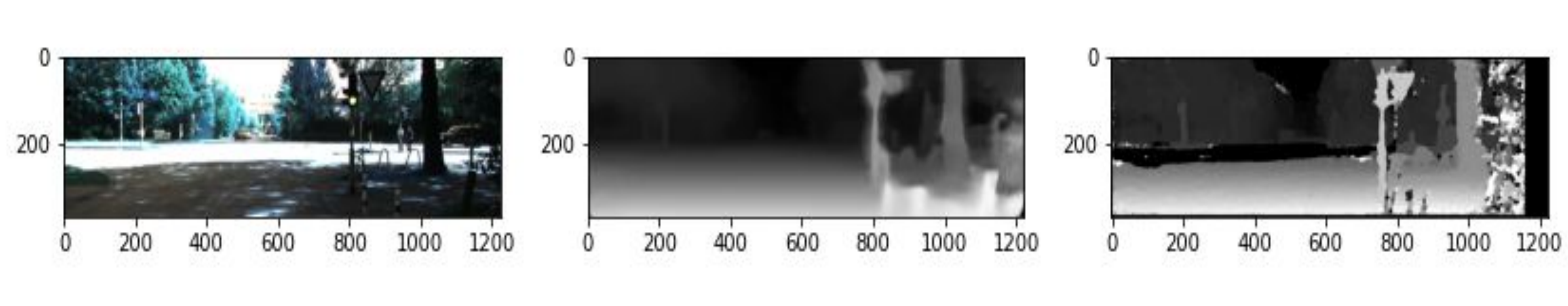


Proposed Convolution block

3

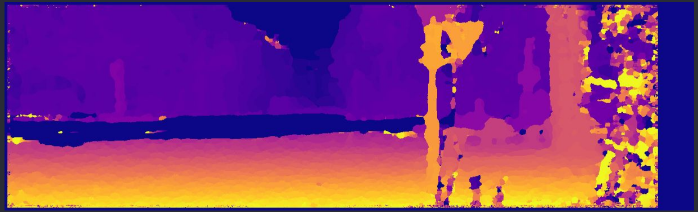
RESULTS OBTAINED

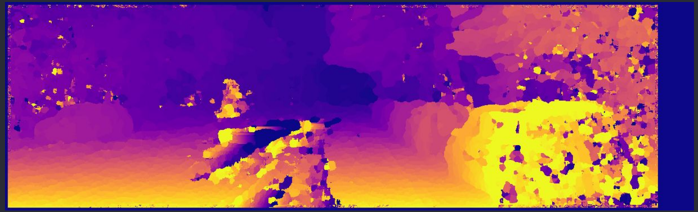
Previous year method results



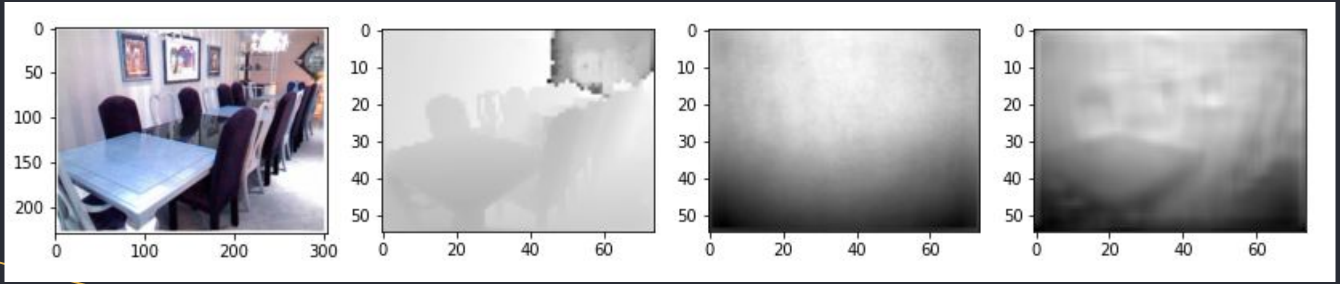
Models	Delta 1 (<1.25)	Delta 2 ($<1.25^2$)	Delta 3 ($<1.25^3$)	RMSE Linear	RMSE Log	ABS rel	Square Relative
IVP	0.462	0.634	0.851	4.393	0.346	0.78	0.835

Previous year method results



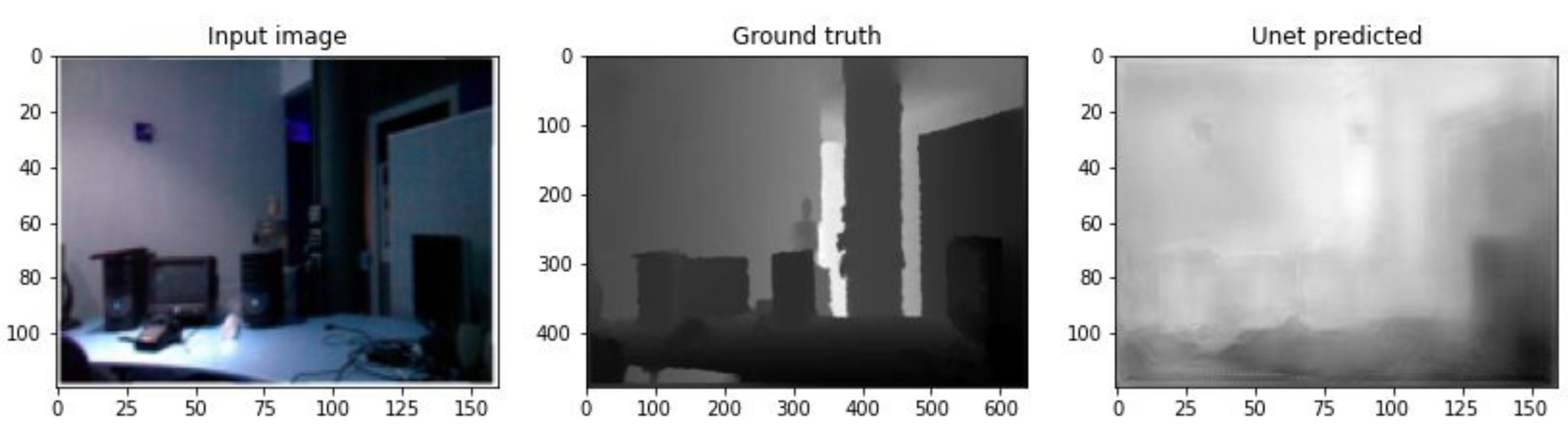


Coarse & Fine Net



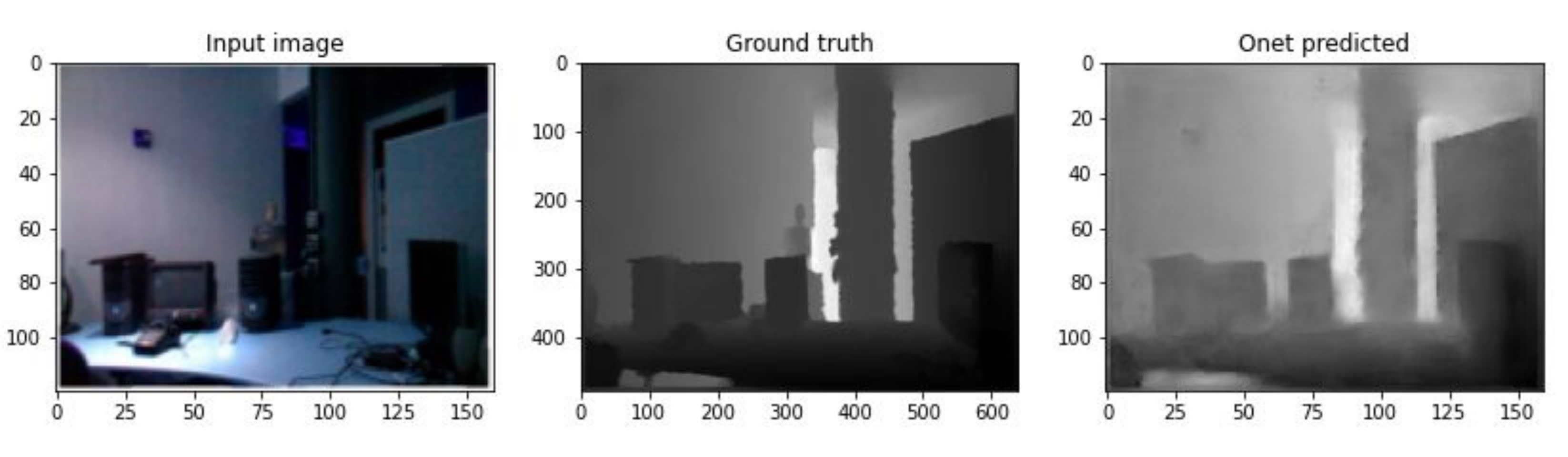
Models	Delta 1 (<1.25)	Delta 2 ($<1.25^2$)	Delta 3 ($<1.25^3$)	RMSE Linear	RMSE Log	ABS rel	Square Relative
Coarse	0.4797	0.8064	0.9449	0.8961	0.1465	0.3709	0.5530
Coarse + Fine	0.5230	0.8268	0.9470	0.8283	0.1415	0.3831	0.5373

UNet



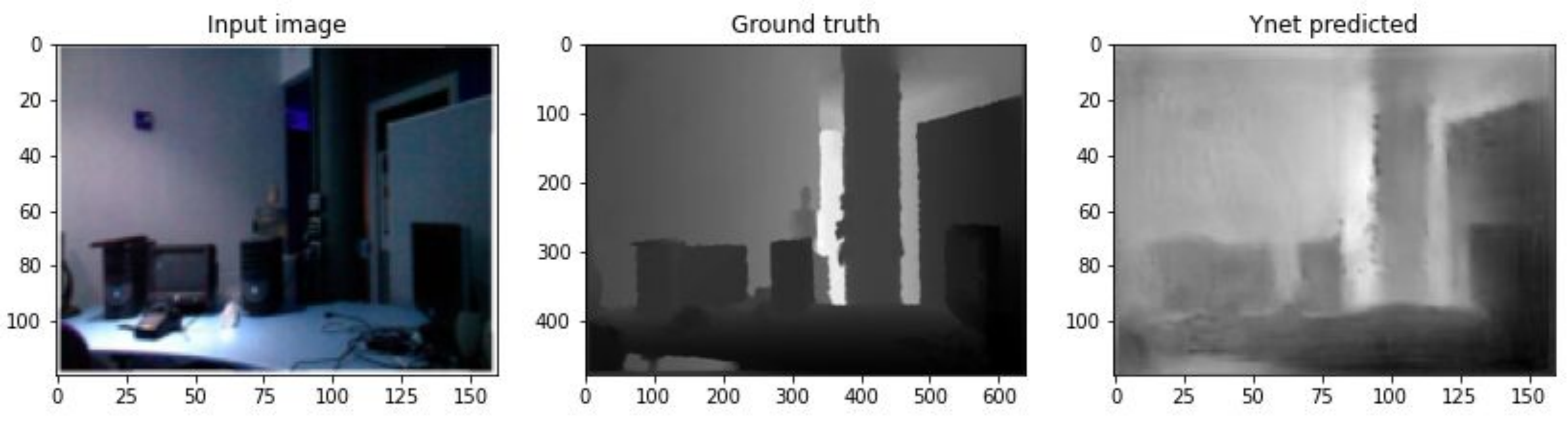
Models	Delta 1 (<1.25)	Delta 2 (<1.25^2)	Delta 3 (<1.25^3)	RMSE Linear	RMSE Log	ABS rel	Square Relative
Unet	0.5625	0.8605	0.9571	0.7873	0.1278	0.3582	0.6016

ONet



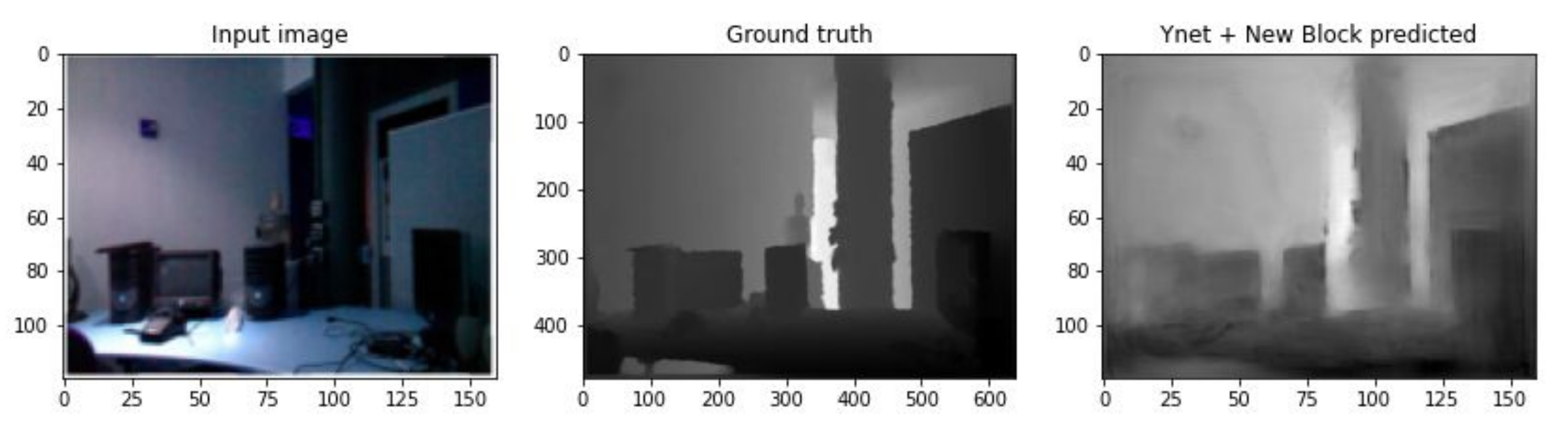
Models	Delta 1 (<1.25)	Delta 2 ($<1.25^2$)	Delta 3 ($<1.25^3$)	RMSE Linear	RMSE Log	ABS rel	Square Relative
Onet	0.6298	0.8984	0.9704	0.6924	0.1057	0.3125	0.5097

YNet

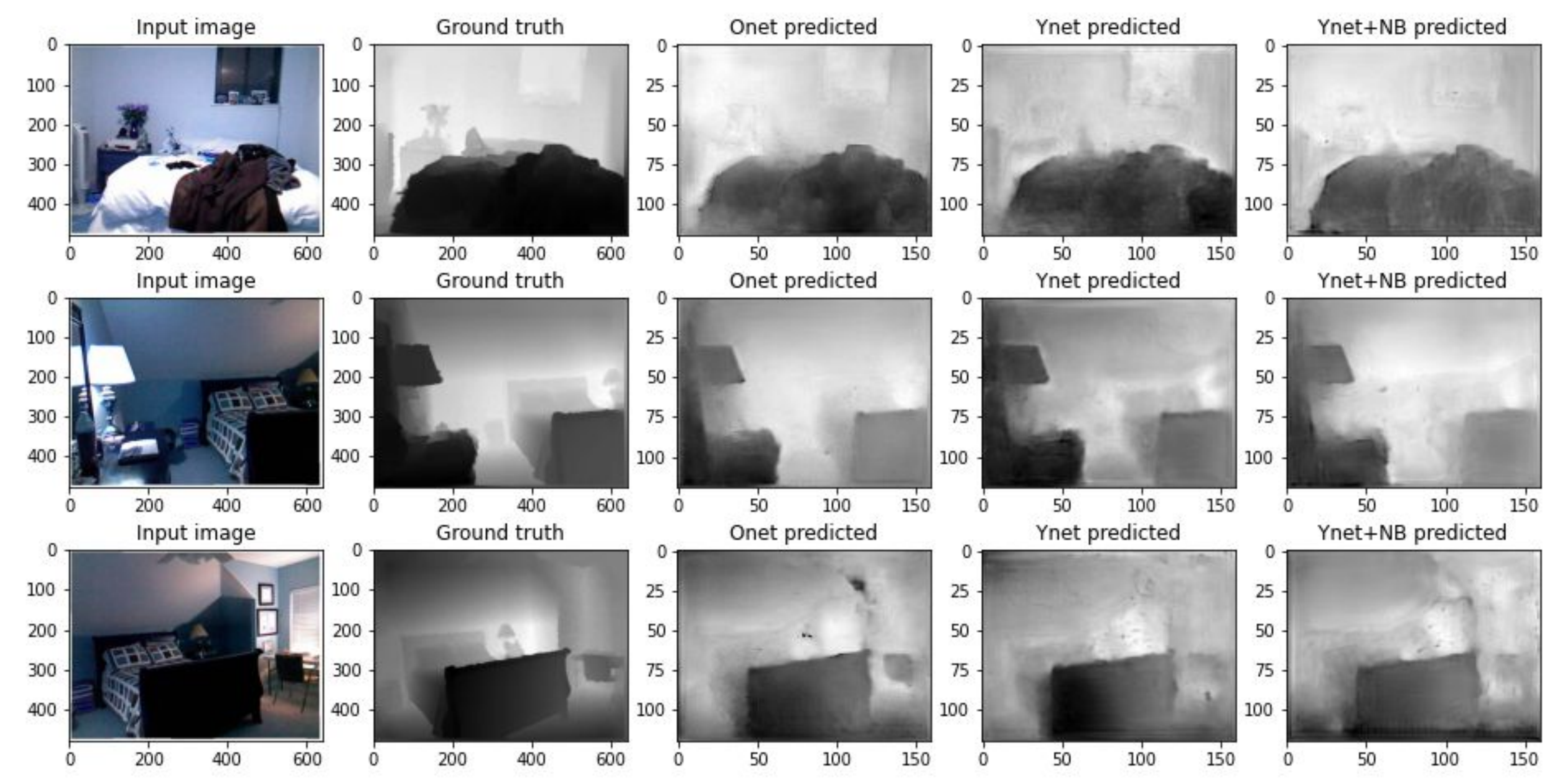


Models	Delta 1 (<1.25)	Delta 2 ($<1.25^2$)	Delta 3 ($<1.25^3$)	RMSE Linear	RMSE Log	ABS rel	Square Relative
Ynet	0.6560	0.9092	0.9725	0.6652	0.1009	0.3074	0.5118

YNet + New Convolution Block



Models	Delta 1 (<1.25)	Delta 2 ($<1.25^2$)	Delta 3 ($<1.25^3$)	RMSE Linear	RMSE Log	ABS rel	Square Relative
Ynet + New block	0.6745	0.9119	0.9772	0.6537	0.0985	0.2971	0.4910

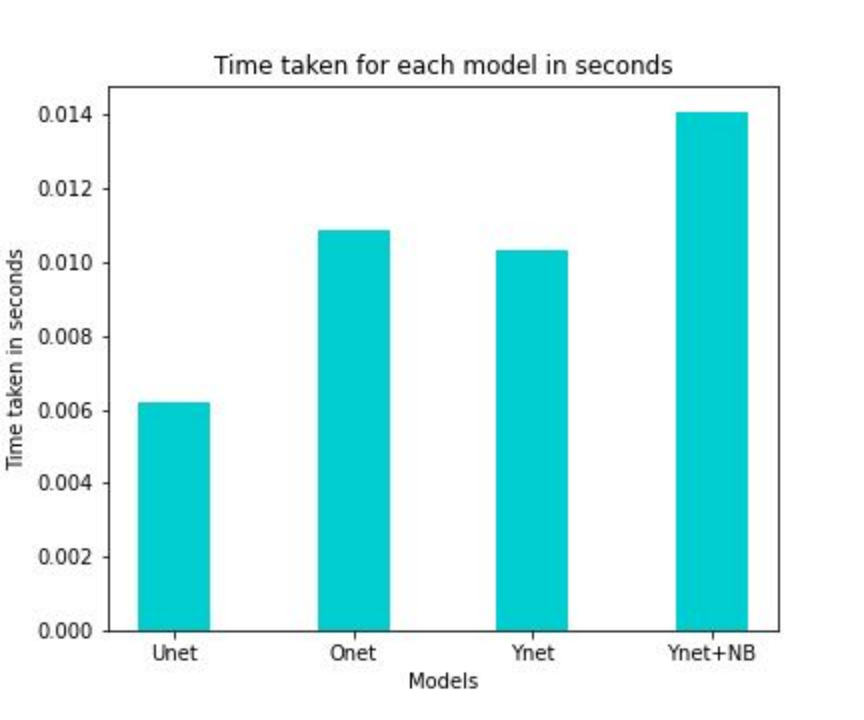


Input image
Ground truth
Onet predicted
Ynet predicted
Ynet+NB predicted
0
100
200
300
400
600
25
50
75
150

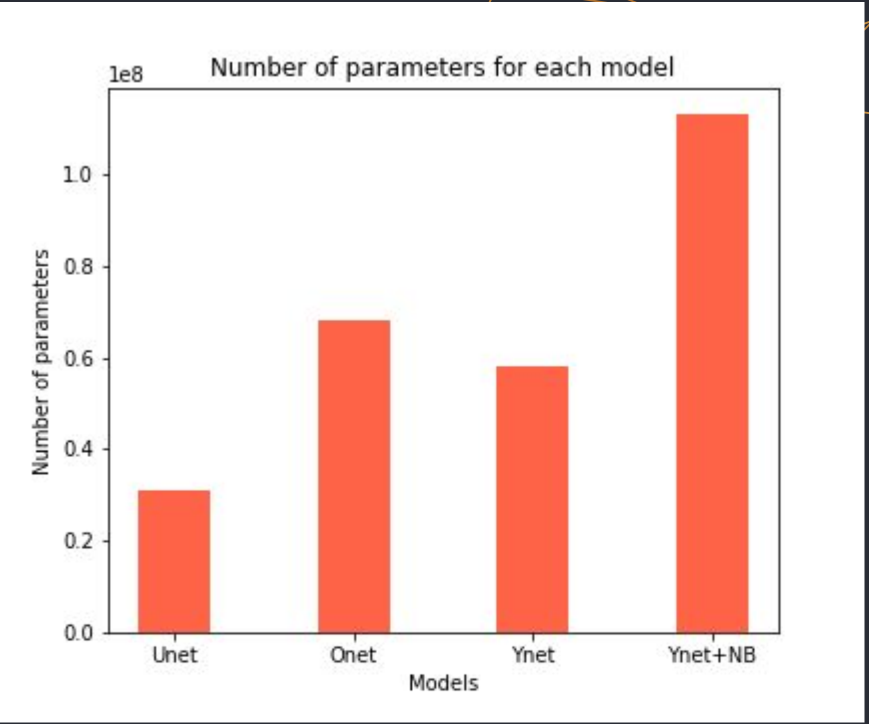
Results

Models	Delta 1 (<1.25)	Delta 2 ($<1.25^2$)	Delta 3 ($<1.25^3$)	RMSE Linear	RMSE Log	ABS relative	Square relative
Coarse + Fine (Paper)	0.611	0.887	0.971	0.907	0.285	0.215	0.212
Coarse + Fine (Trained)	0.5230	0.8268	0.9470	0.8283	0.1415	0.3831	0.5373
Unet	0.5625	0.8605	0.9571	0.7873	0.1278	0.3582	0.6016
Onet	0.6298	0.8984	0.9704	0.6924	0.1057	0.3125	0.5097
Ynet	0.6560	0.9092	0.9725	0.6652	0.1009	0.3074	0.5118
Ynet + New block	0.6745	0.9119	0.9772	0.6537	0.0985	0.2971	0.4910

Results



Time taken for each model in seconds
Time taken in seconds
0.000
0.002
0.004
0.006
0.008
0.010
0.012
0.014
Unet
Onet
Ynet
Ynet+NB
Models



Number of parameters for each model
1e8
Number of parameters
0.0
0.2
0.4
0.6
0.8
1.0
Unet
Onet
Ynet
Ynet+NB
Models

References

- ➢ Dataset: https://cs.nyu.edu/~silberman/datasets/nyu_depth_v2.html
- ➢ David Eigen, Christian Puhrsch, and Rob Fergus. 2014. Depth map prediction from a single image using a multi-scale deep network. In Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2 (NIPS'14). MIT Press, Cambridge, MA, USA, 2366–2374
- ➢ Ronneberger O., Fischer P., Brox T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N., Hornegger J., Wells W., Frangi A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science, vol 9351. Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28
- ➢ Godard, Clement & Aodha, Oisin & Gabriel, Jourdan. (2017). Unsupervised Monocular Depth Estimation with Left-Right Consistency. 10.1109/CVPR.2017.699.

4

FUTURE WORK

FUTURE WORK

- We plan to improve the model by using altros convolution as part of the new-convolution block
- Also we use the full image size if we are able to get enough compute power
- Try to deploy this in a real life scenario using a webcam, also add preprocessing for real world images.

THANKS!

Saikumar Dande
Chandravaran Kunjeti