COMP0122 − Research Project Demo

Obtaining Worldspace Measurements through Virtual Annotation of RGB Images

William Herbosch
Project Supervisor: Niloy J Mitra

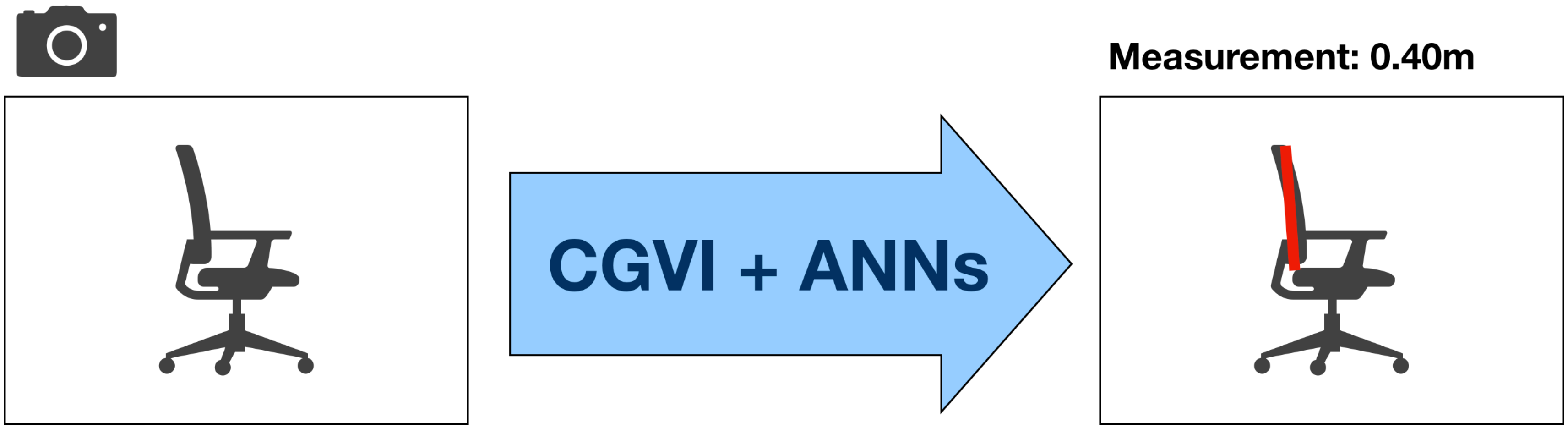
A presentation submitted in part fulfilment of the degree of
MSc (Hons) in Computer Graphics, Vision and Imaging

Department of Computer Science
University College London

2018 − 2019

The Task at Hand…

Take a picture --> Obtain real-world measurements from it!



- Independent Research Project
- (Python/MatLab oriented)
- Unifies Computational Vision and Imaging with Artificial Intelligence
- Architectural Engineering, Spatial Mapping, Medical Imaging, Autonomy, Surveyance…

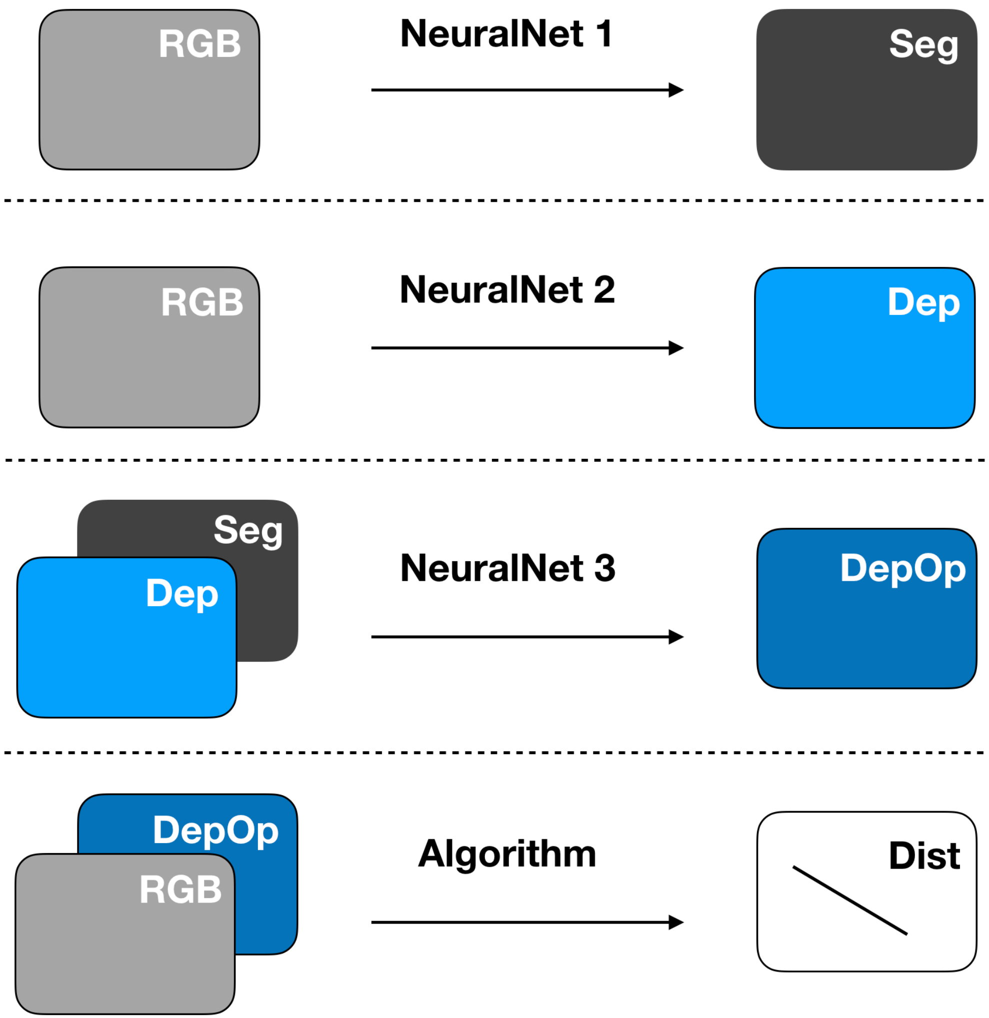
Project Pipeline

Step 1:
Image Segmentation

Step 2:
Depth-Image Estimation

Step 3:
Depth Regression

Step 4:
Measurement Acquisition



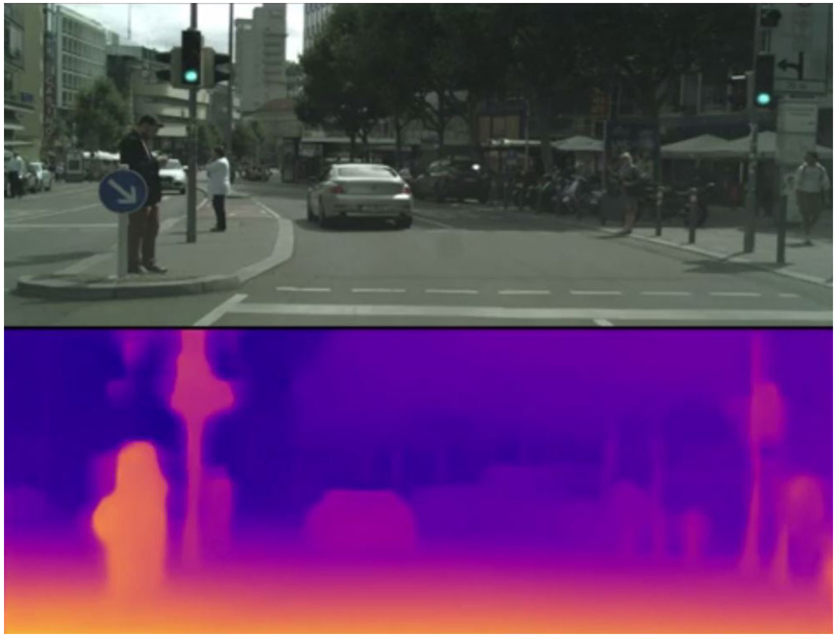
Semantic Segmentation

DeepLabV3+ & PASCAL VOC 2012 Dataset

Aeroplane	Bicycle	Bird	Boat
Bottle	Bus	Car	Cat
Chair	Cow	DiningTable	Dog
Horse	Motorbike	Person	PottedPlant
Sheep	Sofa	Train	TVMonitor
	Background	Void	

- TensorFlow / DeepLab V3+
- A.Verdone, University of Edinburgh, provides a public, readily available implementation
- Identifies key objects within the image (most likely the objects we wish to measure)
- Can be used in our system to produce better depth estimates

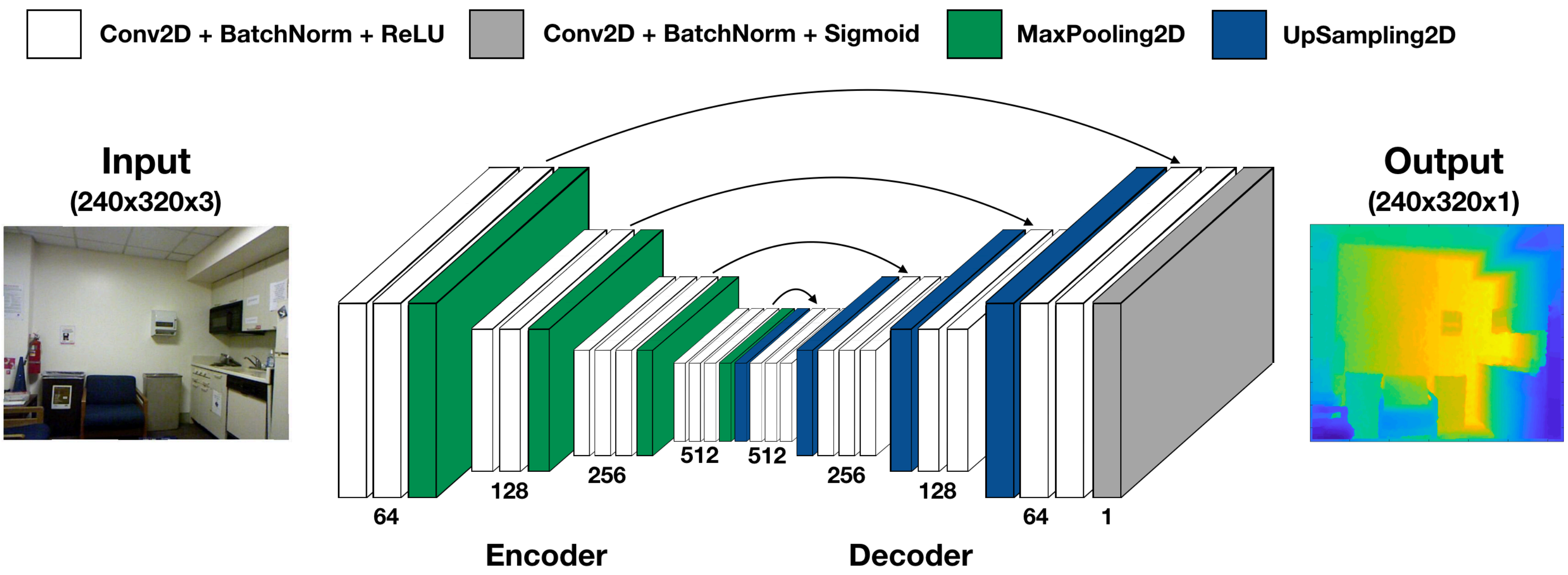
Depth Estimation



- UCL Image to Depth Software
- Spectral Depth Colouring
- Majority Outdoor Scenes

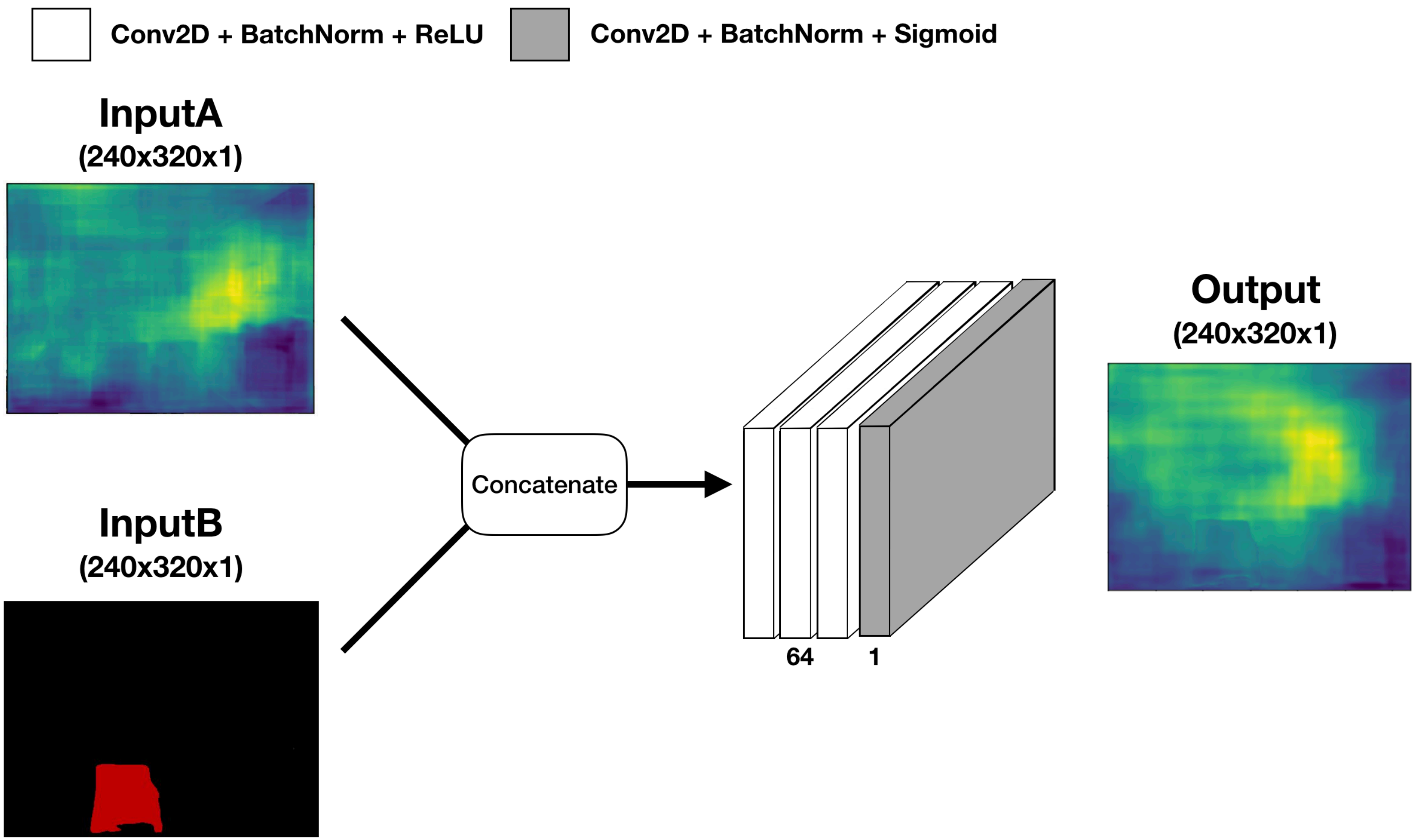
- NYU Depth Dataset V2
- Optimized Colorization
- Majority Indoor Scenes

Depth U-Net Convolution Network



- Series of convolutions (Design originates from ResNet) joined with skip layers
- MaxPooling: Reduces spatial information while increasing feature information
- Upsampling: Increases spatial information (i.e. higher resolution)
- Merging: Enforces prior feature information back onto the up-sampled layer

Optimize: Regress Distance Network



Projection Matrix & Euclidean Distance

Let **I(x, y)** denote a coloured image at pixel location **(x, y)** and let **D(x, y)** denote its corresponding depth map.

Let **p1 = (x1, y2)** and **p2 = (x2, y2)** denote the two query points on the image whose distance is desired to be measured.

Assuming a pinhole camera model, the camera's viewing frustum can be mapped to a projection matrix **K**, with a principle point (i.e. centred pixel) at **(cx, cy)**, a focal length for the camera lens of **(fx, fy)** and a skew of **s**.

$$\mathbf{K} = \begin{bmatrix} fx & s & cx \\ 0 & fy & cy \\ 0 & 0 & 1 \end{bmatrix}$$

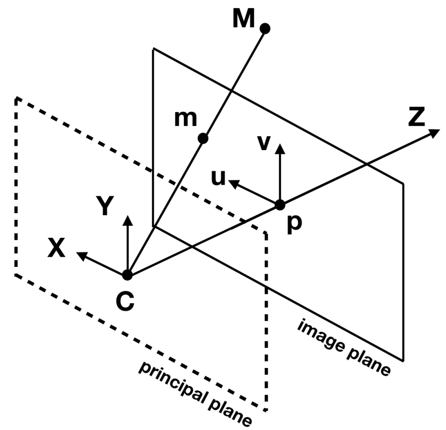
Assume that **fx = fy** and **s = 0** (as do most commercial cameras).

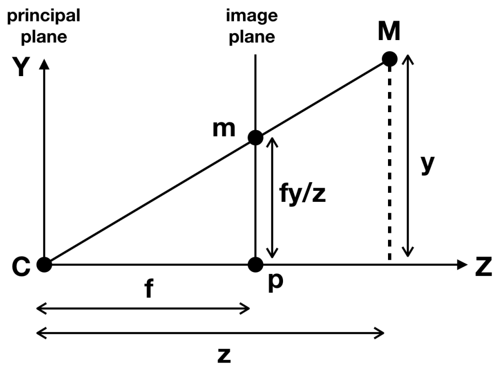
Any point in **I** with associated depth information **Z = D(x, y)** is a projection of a 3D point **P** given by the modified projection matrix:

$$\mathbf{P} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} (x - cx)\, Z / fx \\ (y - cy)\, Z / fy \\ Z \end{bmatrix}$$

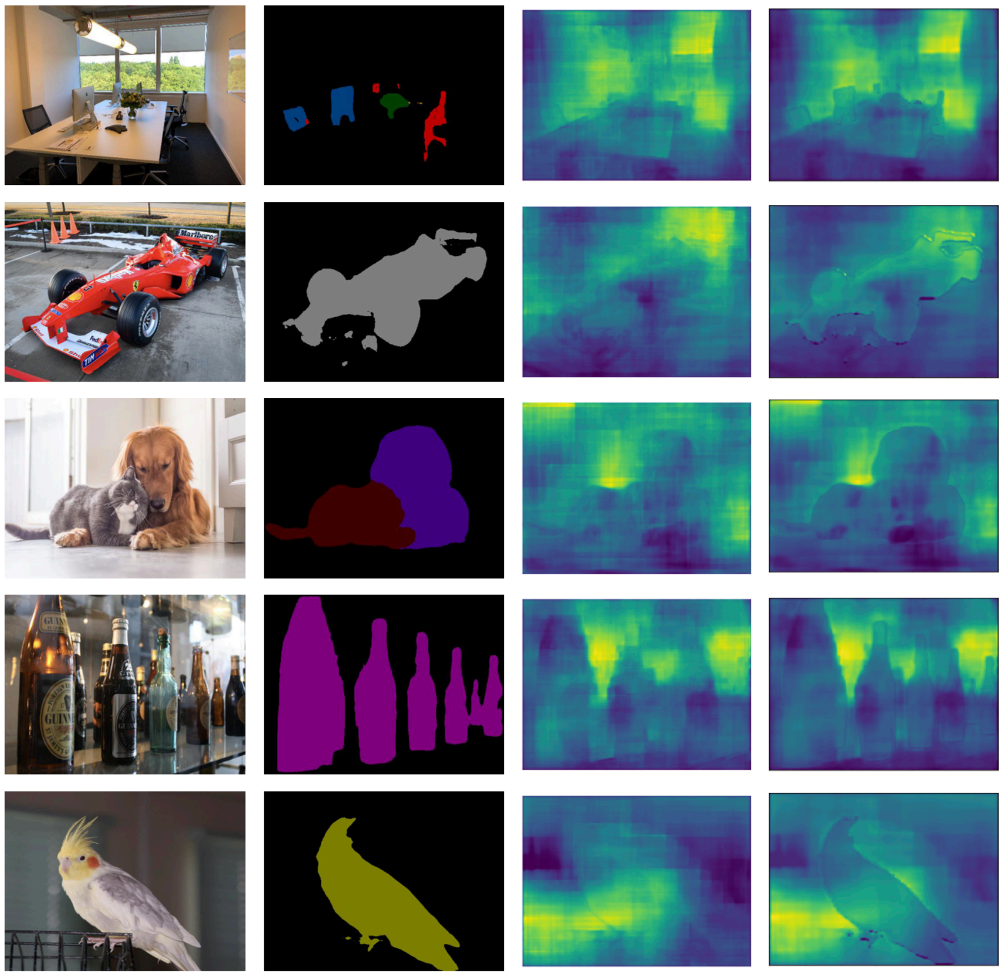
By computing the above for the two query points **p1** and **p2**, their corresponding 3D points **P1** and **P2** can be computed as a Euclidean distance from which the real-world distance **d** can be measured.

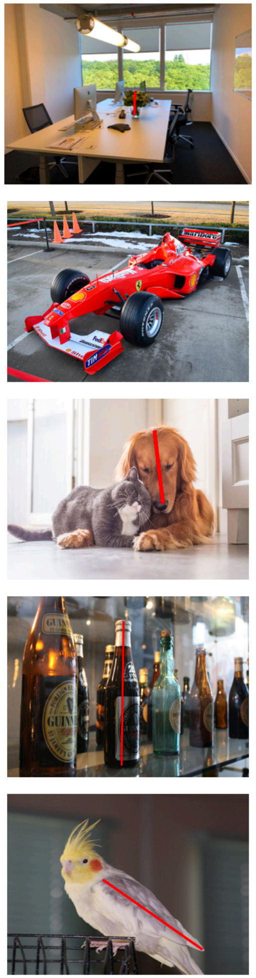
$$\mathbf{d} = \sqrt{(X1 - X2)^2 + (Y1 - Y2)^2 + (Z1 - Z2)^2}$$





Results





Readings	"Looked-up" truths
0.40m 0.40m 0.34m 0.41m 0.38m avg. **0.39m**	**0.37m**
0.47m 0.46m 0.47m 0.45m 0.49m avg. **0.47m**	**0.50m**
0.30m 0.28m 0.31m 0.32m 0.33m avg. **0.31m**	**0.34m**
0.27m 0.28m 0.28m 0.27m 0.25m avg. **0.27m**	**0.22m**
0.18m 0.20m 0.20m 0.21m 0.19m avg. **0.20m**	**0.23m**

…allow me to demonstrate!

References (for personal interest)

- *O.Ronneberger, P.Fischer, & T.Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, University of Freiburg, Germany, 2015*
- *C.Godard, O.M.Aodha & G.J.Brostow, Unsupervised Monocular Depth Estimation with Left-Right Consistency, University College London, 2017*
- *L-C.Chen, Y.Zhu, G.Papandreou, F.Schroff, & H.Adam, Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation, Google Inc., 2018*
- *Visualization of Convolutional Neural Networks for Monocular Depth Estimation, Tohoku University and Center for Advanced Intelligence, Japan, 2019*
- *C.Jaehoon, M.Dongbo, K.Youngjung, & S.Kwanghoon, A Large RGB-D Dataset for Semi-supervised Monocular Depth Estimation, IEEE, 2019*
- *K.Seshadrinathan, O.Nestares & Y.Wu, Accurate measurement of point to point distances in 3D camera images, Intel Corporation, 2017*
- *N.Silberman, D.Hoiem, P.Kohli & R.Fergus, Indoor Segmentation and Support Inference from RGBD Images, Courant Institute at New York University Department of Computer Science, University of Illinois at Urbana-Champaign, Microsoft Research, Cambridge, 2012*