# Depth Estimation Using Deep Neural Networks

Mușat Bogdan-Adrian

December 2017

# Problem Formulation

- Given a pair of stereo images, how to estimate the depth of the existing objects?

# Disparity

- Disparity represents the difference of perspective created by the horizontal or vertical separation of two cameras
- The human brain processes disparity information from both eyes to estimate the depth of real world objects
- Disparity is inverse proportional with depth; given the basline distance $b$ between the cameras and the camera focal lenght $f$, the depth $\hat{d}$ from the predicted disparity $d$ is simply $\hat{d} = bf/d$
- Objects closer to the camera have bigger disparities, whilst objects that are farther have smaller disparities

# Proposed Solution

- Given a pair of rectified [1] stereo images as input, build a model which can accurately predict the disparity per-pixel (i.e. learn the $\Delta$ values with which every pixel from the left image is shifted from the right one on the $x$ axis)
- Such a system can be modeled by a Convolutional Neural Network

[1] Vertically aligned

## Challenges - Textureless Areas

- Depth estimation is an ill-posed problem: a pixel in a textureless area from the left image can belong to multiple pixels in the right image

Figure 1: Example of objects with textureless areas

# Challenges - Object Occlusion

- Parts of an object may be visible in an image, while being absent in the other one because of the difference of perspective

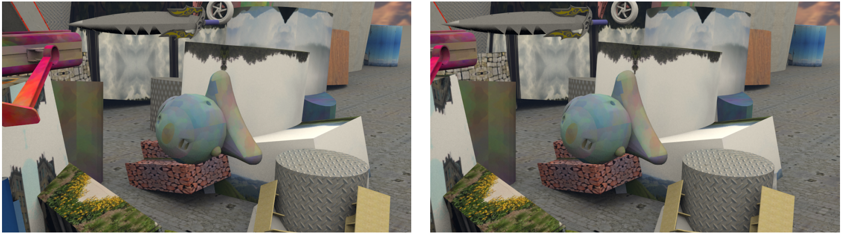

Figure 2: Occlusion in stereo images

# Convolutional Neural Networks Architectures

- Supervised
  - Requires ground-truth disparity, which might be expensive to obtain
  - Based on disparity regression
  - Accurate predictions
- Unsupervised
  - Does not require any form of ground-truth
  - Based on an image reconstruction loss
  - Has artefacts around the edges of objects, caused by occlusion

# DispNet[2]

- Image-to-image autoencoder structure

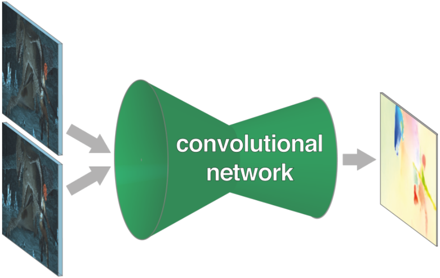


Figure 3: Hourglass structure of DispNet

[2] Philipp Fischer et al. "FlowNet: Learning Optical Flow with Convolutional Networks". In: *CoRR* abs/1504.06852 (2015). arXiv: 1504.06852. URL: http://arxiv.org/abs/1504.06852.

# Unsupervised Monocular Network[3]

- Similar architecture as DispNet

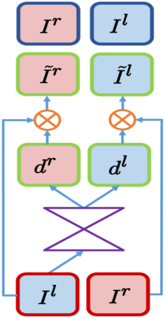


Figure 4: Unsupervised Monocular Network architecture

[3] Clément Godard, Oisin Mac Aodha, and Gabriel J. Brostow. "Unsupervised Monocular Depth Estimation with Left-Right Consistency". In: *CoRR* abs/1609.03677 (2016). arXiv: 1609.03677. URL: http://arxiv.org/abs/1609.03677.