**U-N.O.1T: A U-NET EXPLORATION, IN DEPTH**

John Chuter (jchuter), Geoffrey Boullanger (gbakker), Manuel Nieves (mnieves)

**INTRODUCTION - PREDICTING, MOTIVATION**

**Objective:** Given a video feed, infer a corresponding depth map sequence.

**How:** Analyze spatiotemporally with convLSTM encoding + decoding

**DATASET - FEATURES**

KITTI
- Inputs = RGB video feed
- Groundtruth = LIDAR mapping
- Features are pixel values

**MODEL - MATH**

**convLSTM:**

\[
\begin{align*}
\hat{i} &= \sigma(\text{ReLU}(W_{xi} \cdot X_t + W_{hi} \cdot H_{t-1} + W_{ci} \cdot C_{t-1} + b_i)) \\
\hat{f} &= \sigma(\text{ReLU}(W_{xf} \cdot X_t + W_{hf} \cdot H_{t-1} + W_{cf} \cdot C_{t-1} + b_f)) \\
\hat{g} &= \tanh(\text{ReLU}(W_{xg} \cdot X_t + W_{hg} \cdot H_{t-1} + b_g)) \\
C_t &= \hat{f} \odot C_{t-1} + \hat{i} \odot \hat{g} \\
o_t &= \sigma(\text{ReLU}(W_{xo} \cdot X_t + W_{ho} \cdot H_{t-1} + W_{co} \cdot C_{t} + b_o)) \\
H_t &= o_t \odot \tanh(C_t)
\end{align*}
\]

where \(\odot\) refers to a convolution operation

see 'Model - Specifics'

**MODEL - SPECIFICS**

**FUTURE WORK**

- Bigger images
- Different sequence lengths
- More layers; encoding + decoding
- Layer parameters; # filters, stride, etc.
- Bidirectionality options

**MODEL - IMPLEMENTATION DETAILS**

- Pytorch 0.4.1
- Machine 1: GeForce GTX 1080 Ti
- Machine 2: 2 x GeForce GTX 1080

**ACKNOWLEDGEMENTS - LITERATURE**

- DepthNet
- Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting

**REVIEW - DISCUSSION**

- Complexity - meaningfully, creatively using convolutions and LSTMs for interesting applications; here inside a U-Net encoder-decoder architecture.
- Comparison to baselines (DepthNet) is difficult, because we chose different image size and parameter sizes due to machine constraints.
- LSTM and bidirectionality have not distinguished themselves from convolutions.
- Computation slow.

**MODEL - BIG PICTURE**

U-Net architecture; encoding + decoding

**EXPERIMENTAL RESULTS - PERFORMANCE METRICS**

- Performance metrics
- Complexity - meaningfully, creatively using convolutions and LSTMs for interesting applications; here inside a U-Net encoder-decoder architecture.
- Comparison to baselines (DepthNet) is difficult, because we chose different image size and parameter sizes due to machine constraints.
- LSTM and bidirectionality have not distinguished themselves from convolutions.
- Computation slow.

**ACKNOWLEDGEMENTS**

We would like to thank Professor Ng and Professor Dror, as well as the TAs, for their dedicated instruction.