U-N.o.1T: A U-Net exploration, In Depth

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

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:

$i_t = \sigma(ReLU(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \circ C_{t-1} + b_i))$
$f_t = \sigma(ReLU(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \circ C_{t-1} + b_f))$
$g_t = tanh(ReLU((W_{xg} * X_t + Whg * H_{t-1} + b_g)))$
$C_t = f_t \circ C_{t-1} + it \circ g_t$
$o_t = \sigma(ReLU(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} \circ C_t + b_o))$
$H_t = o_t \circ tanh(C_t)$

where $*$ refers to a convolution operation
see 'Model - Specifics'

Implementation - Model - Big Picture

U-Net architecture; encoding + decoding

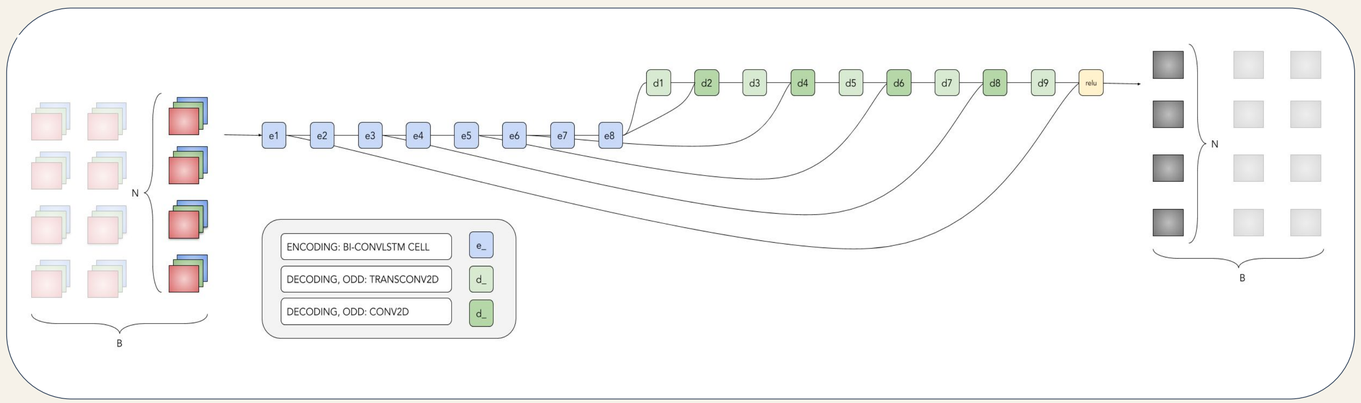


fig.1 Encoder-Decoder U-Net

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 - specifics

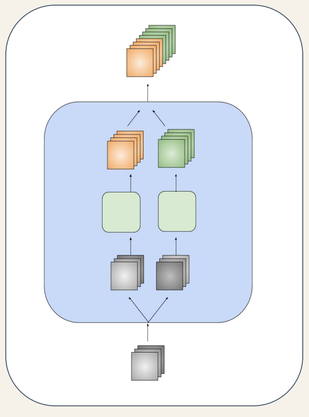


fig.2 bi-ConvLSTM Cell

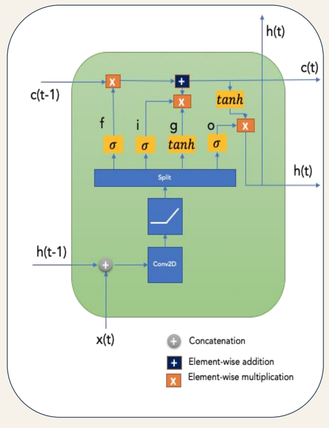


fig.3 ConvLSTM Cell

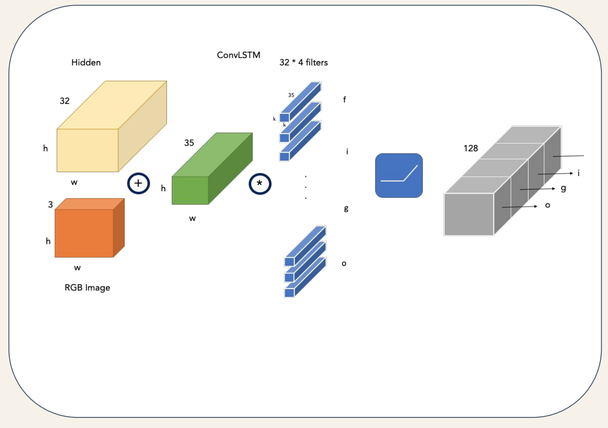
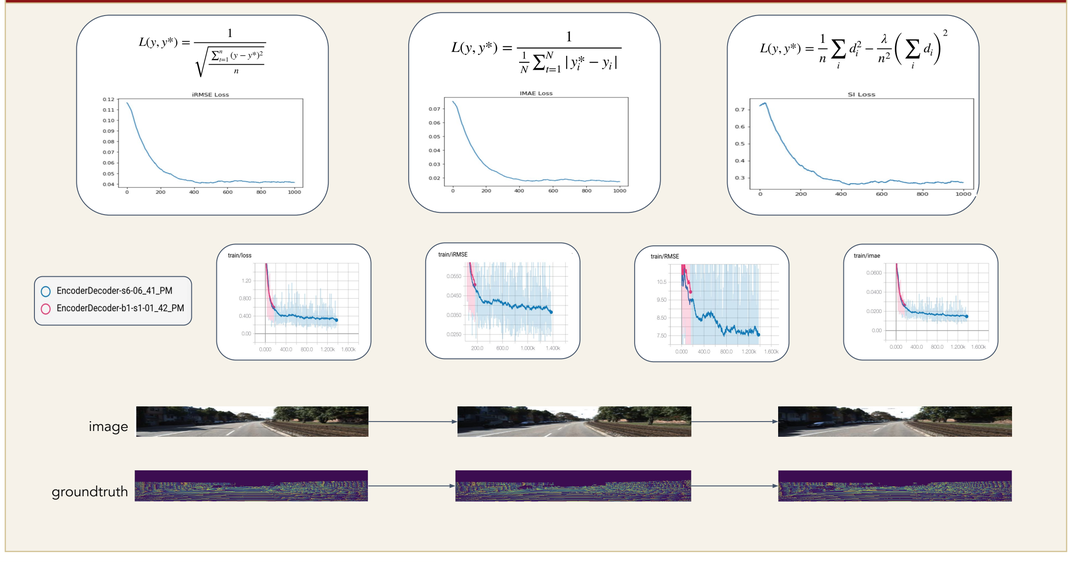


fig.4 ConvLSTM Cell; detail

Experimental Results - Performance Metrics



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*

Acknowledgements

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