Long Short-Term Memory For Autonomous Driving

Michael Treml[1]*, José Arjona-Medina[1]*, Thomas Unterthiner[1]*, Rupesh Durgesh[2], Felix Friedmann[2], Peter Schuberth[2],
Andreas Mayr[1], Martin Heusel[1], Markus Hofmarcher[1], Michael Widrich[1], Ulrich Bodenhofer[1], Bernhard Nessler[1] and Sepp Hochreiter[1]

[1] Institute of Bioinformatics, Johannes Kepler University Linz, Austria
[2] Audi Electronics Venture GmbH, Embedded Software Development / Machine Learning, Gaimersheim, Germany

JOHANNES KEPLER UNIVERSITY LINZ

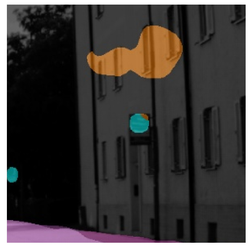
Audi Electronics Venture GmbH

MOTIVATION

Currently much research on autonomous driving is based on inferring driving decisions from single-frame inputs without taking spatio-temporal correlations into account. We propose to exploit these correlations using a recurrent neural network (RNN) architecture known as Long Short-Term Memory (LSTM, [1]) to make more robust, more confident and more timely decisions.

We applied current state of the art convolutional neural network architectures for semantic segmentation like FCN [2], ENet [3] or SqueezeNet [4] to image sequences. We identified three types of undesired artefacts and found that using the output of a segmentation-CNN as an input for a LSTM visibly improves the accuracy.

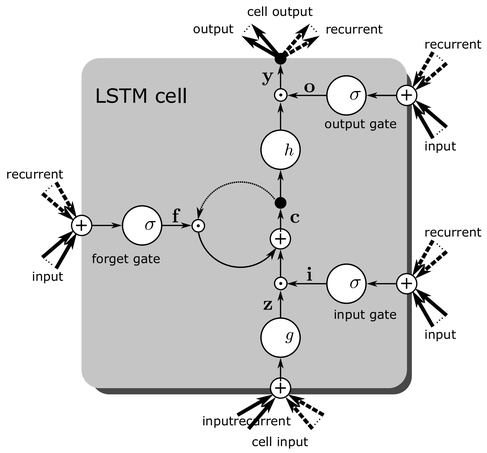
Blurred object edges

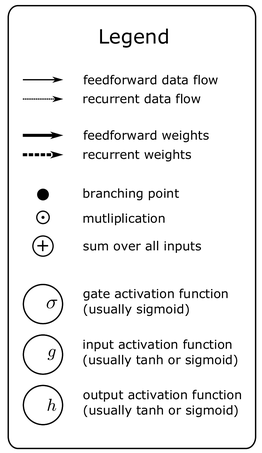


Flickering ghost objects

Glutinous object separation

LSTM





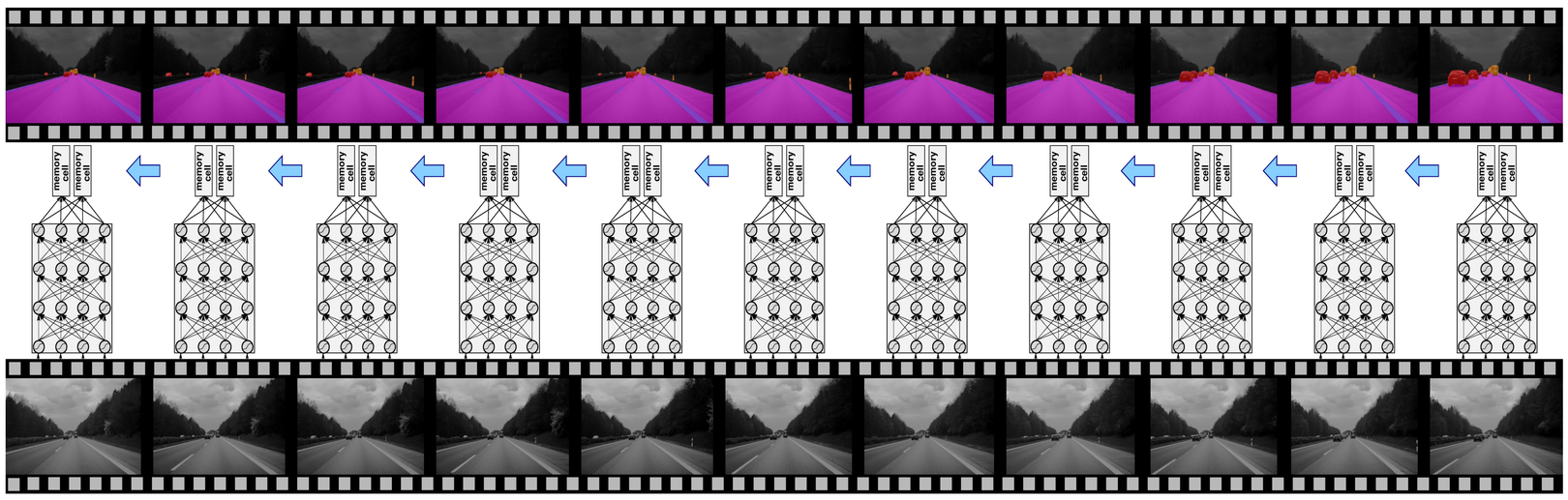
Long Short-Term Memory: LSTM is a special type of RNN that allows for constant error flow during training. It avoids decaying training errors when they are back-propagated through time and enables uniform credit assignment over all inputs over time.

ATTENTION BY LSTM

To make use of the full potential of LSTM we want consider its ability to learn attention. That is especially useful to speed up segmentation by looking at areas that are of higher relevance for autonomous driving first. Furthermore robustness of driving decisions can be improved by paying more attention to critical situations, e.g.

- Abrupt braking maneuvers on highways
- Crossing vehicles at intersections
- Overtaking vehicles in oncoming traffic
- Pedestrians crossing

METHOD



LSTM for video sequences: In this sequence a vehicle in front brakes abruptly. Single frame scene parsing will detect this event at the moment of its occurrence. In contrast LSTM memory cells can predict critical actions already before they occur by looking at past frames of the sequence as well.

SEGMENTATION REFINEMENT

Single frame segmentation

LSTM-refined segmentation

REFERENCES

[1] Hochreiter et al. Long Short-Term Memory. 1997.

[2] Jonathan Long et al. Fully Convolutional Networks for Semantic Segmentation. 2014.

[3] Paszke et al. Enet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation. 2016.

[4] Han et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size. 2016.