HSLU Hochschule Luzern

Technik & Architektur

Master in Engineering

Robust Rail Track Prediction with Temporal Deep Learning

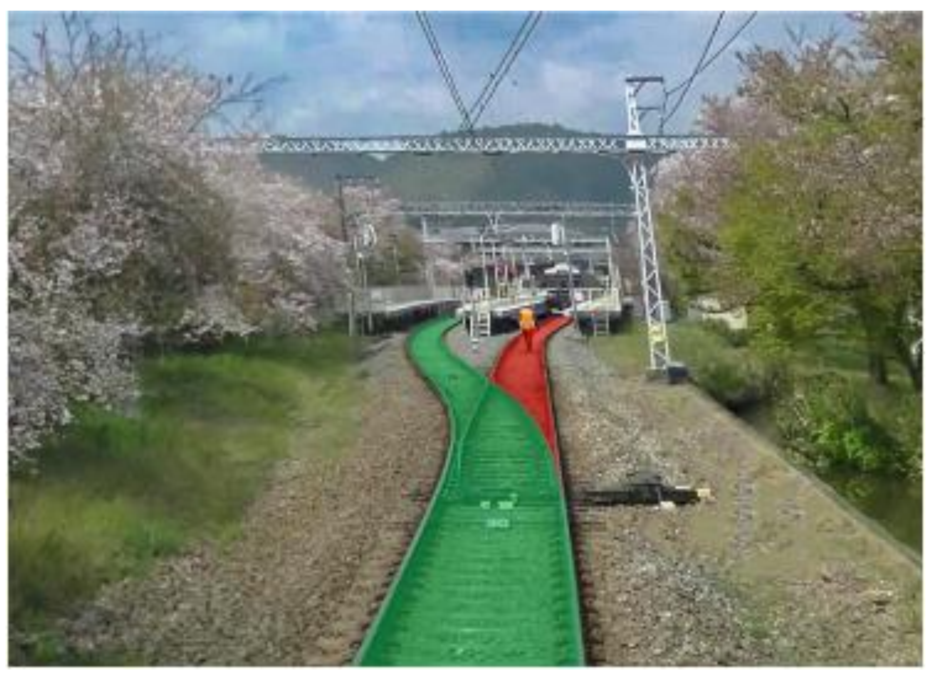

Fig.1: Rail track prediction (green) [1]

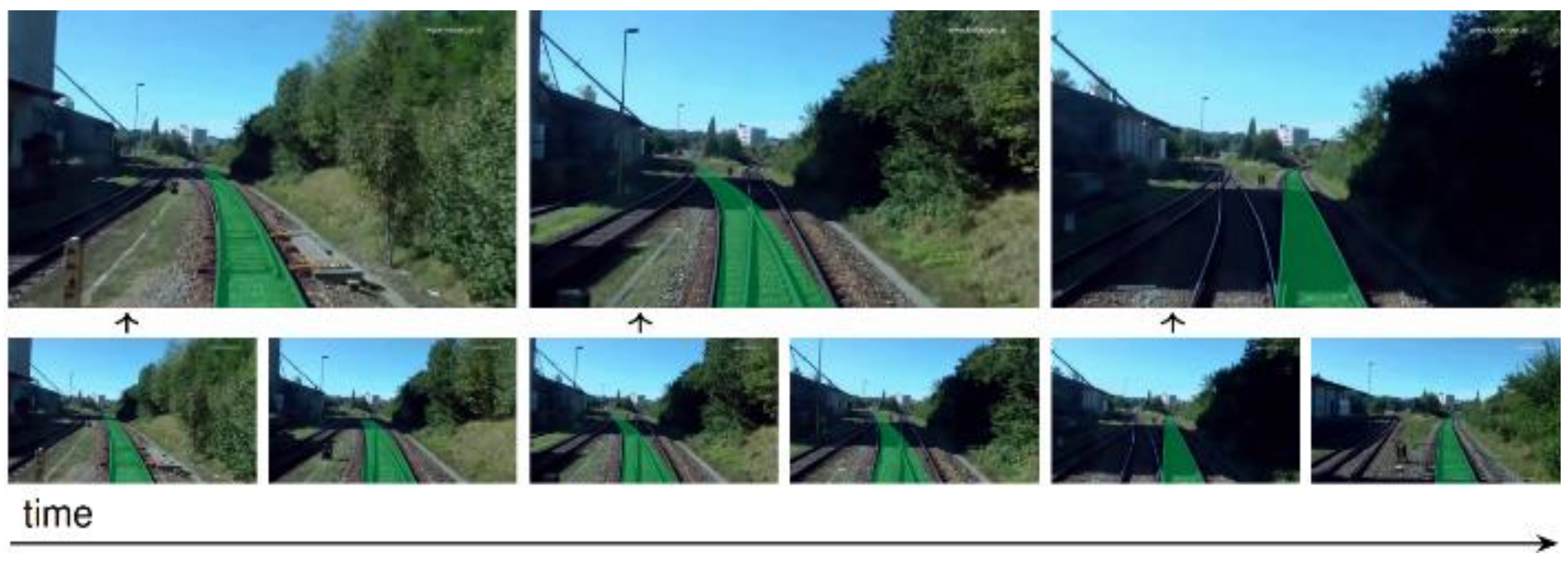

Fig. 2: Limitation of single-frame-based models

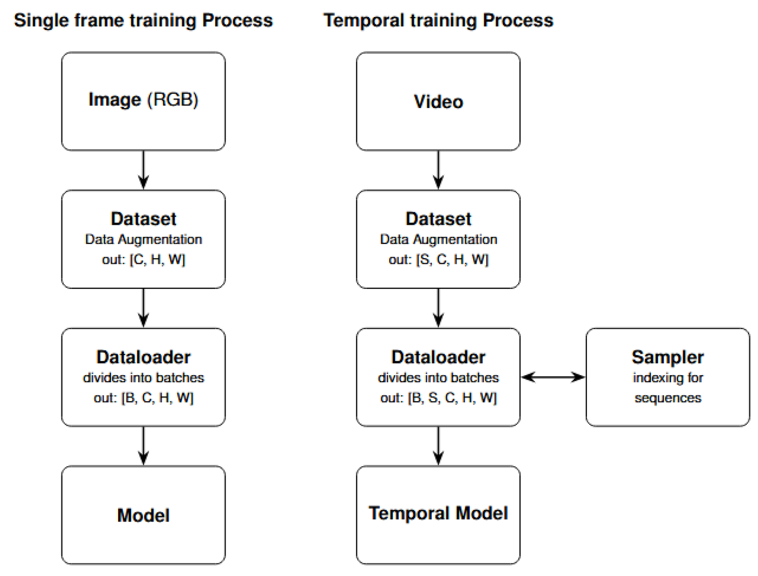

Fig. 3: Comparison of data handling logic

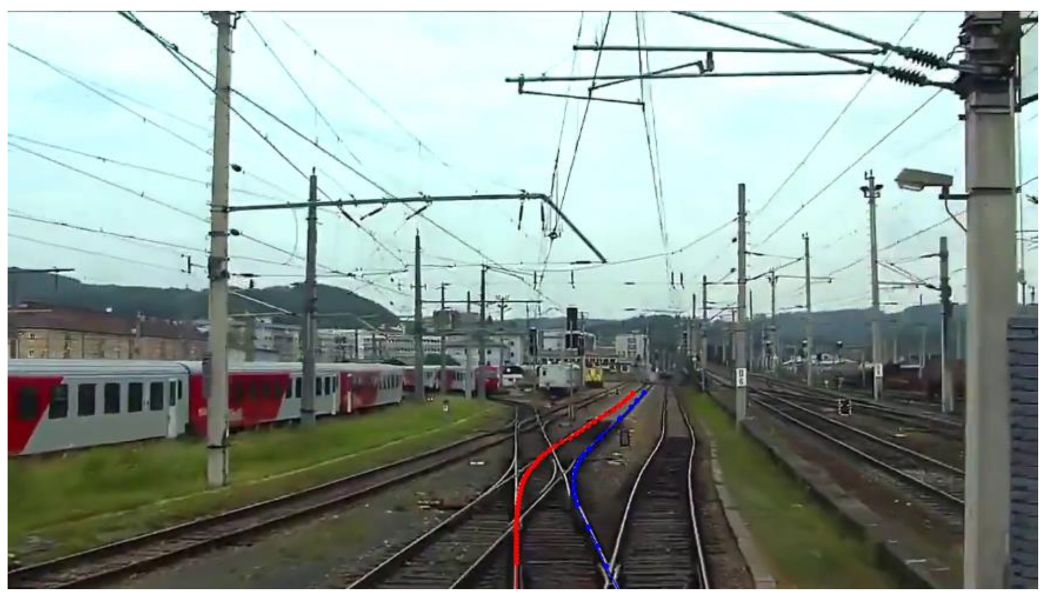

Fig. 4: Annotated test sequence Switch 1

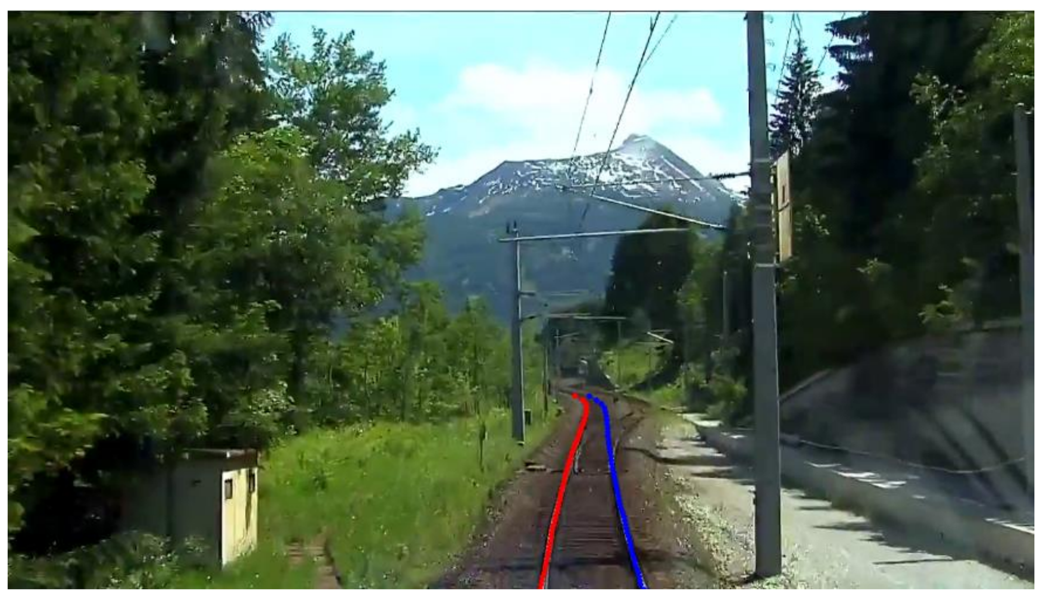

Fig. 5:Annotated test sequence Switch 2

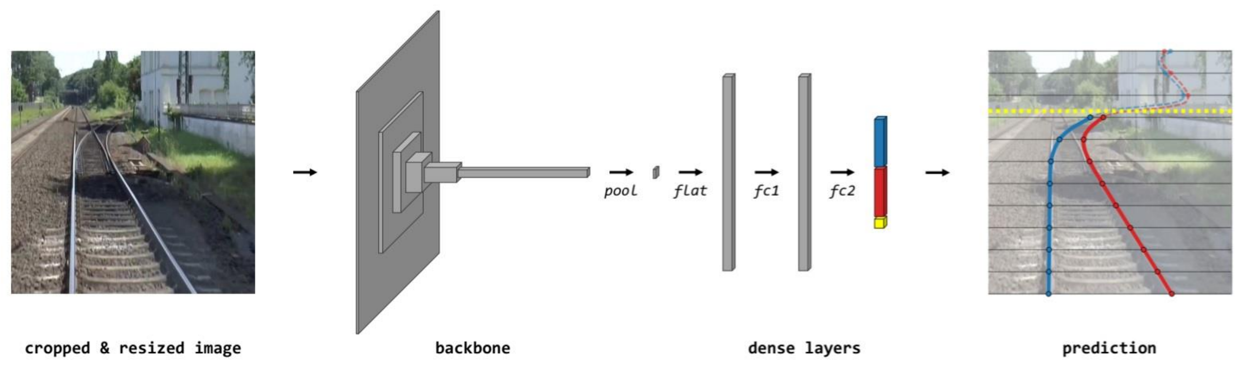

Fig. 8: Chosen Baseline Model (TEP-Net [1])

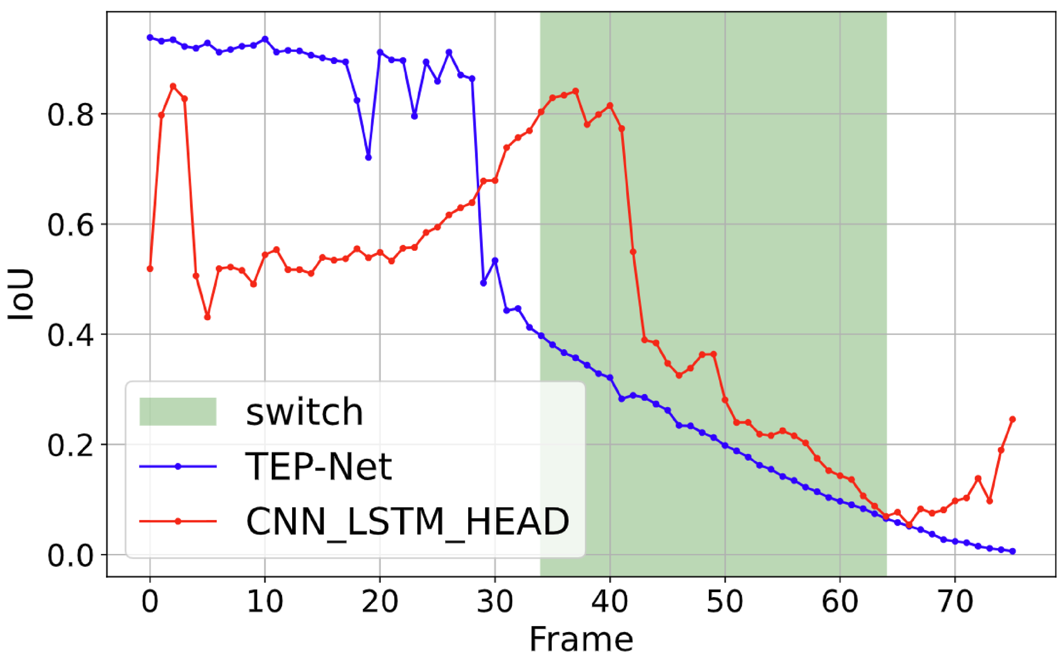

Fig. 6: IoU trend Switch 1

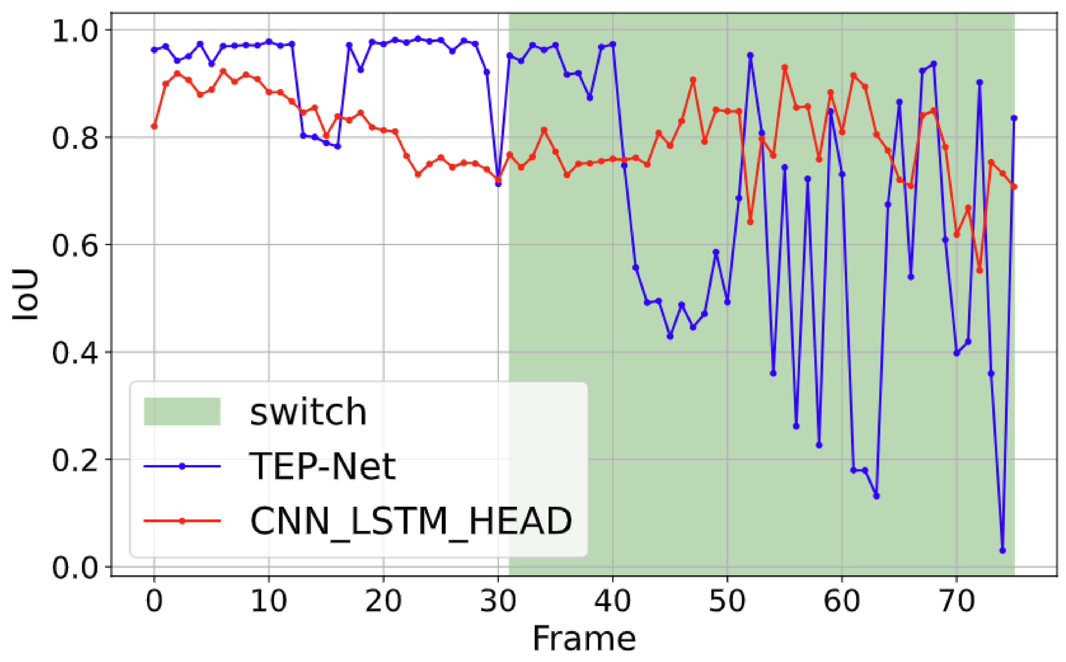

Fig. 7: IoU trend Switch 2

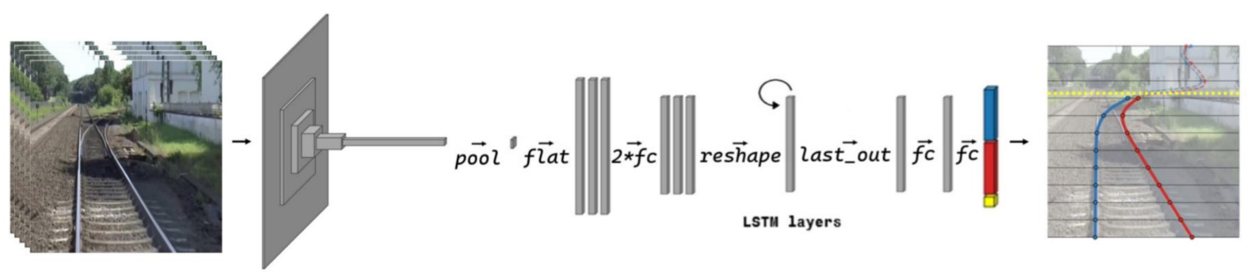

Fig. 9: One the best performing model from 11 temporal models: CNN_LSTM_HEAD

Problem and Goal

In the last few years, rail traffic has increased. Additionally, accidents in rail traffic have risen, often caused by incorrect switch settings or derailments at switches. Autonomous systems have the potential to prevent these accidents and enhance safety. A key challenge is distinguishing the train's track from other tracks in complex rail environments, especially at switches where tracks split (Fig. 1). Therefore, this work focuses on developing a real-time rail track prediction system that accurately identifies the train's path prioritizing scenarios with switches.

TEP-Net [1] is chosen as baseline (Fig. 8), which has a limitation. The model is single-frame-based, and when driving over a switch, the necessary information (start of switchblades) is missing (Fig. 2).

A possible approach to elevate robustness in switch scenarios is to learn from temporal information by incorporating RNNs, bridging the arising information gap.

[1] *Thomas Laurent (2024) "Train Ego-Path Detection on Railway Tracks Using End-to-End Deep Learning"*

Solution concept

The solution concept is a sequence-to-one approach in which the input is a series of images, and the model predicts the rail track of the last image. To learn dependencies between frames, temporal modules (mostly LSTMs) are incorporated into the proposed models (Fig. 9). A new temporal dataset (38 sequences with 76 frames each) is created tailored for the particular switch problem. Data handling and data augmentation are adapted to support video data instead of single images (Fig. 3).

Results

There are two main problematic behaviours of single-frame-based models when driving over switches. Either the prediction fluctuates between the correct and the wrong track (Fig. 5 and 7), or the model chooses the wrong track, and the accuracy drops until the train passes the switch (Fig. 4 and 6). Results show that higher accuracy can be achieved in problematic frames with temporal models.

Switch 1 shows a more stable IoU-trend, and Switch 2 shows a time-delayed accuracy drop when using temporal models. Measurements on the NVIDIA Jetson AGX Xavier reveal that the MobileNetV3 backbone achieves sufficiently low latency for videos with 30 FPS.

Sebastian Goebel

Supervisor:
Prof. Dr. Björn Jensen

2. Supervisor:
Dr. techn. Stefan Thalhammer

Christian Doppler Forschungsgesellschaft