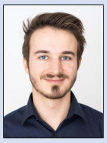
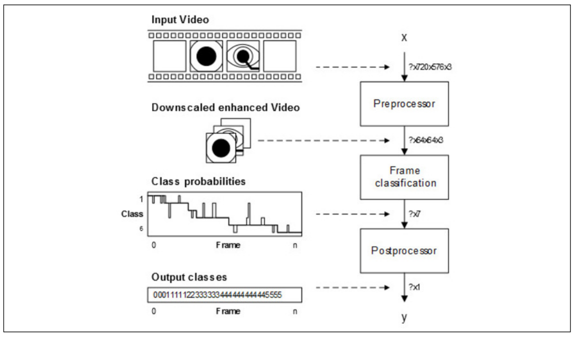

Nicolas
Tobler

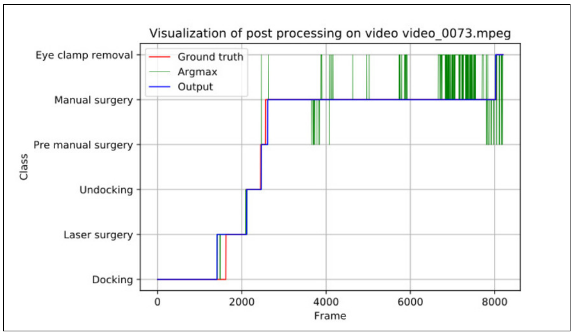
Raphael
Unterer

Graduate Candidates	Nicolas Tobler, Raphael Unterer
Examiner	Prof. Dr. Guido Schuster
Co-Examiner	Gabriel Sidler, Eivycom GmbH, Zürich, ZH
Subject Area	Artificial Intelligence

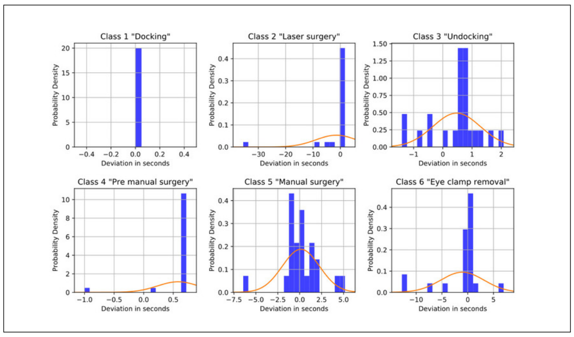
Automatic Shot Transition Detection



Simplified graph of the whole algorithm
Source: Own illustration



Outcome of post-processing shown on an example
Source: Own illustration



Histograms of deviations from ground truth in seconds
Source: Own illustration

Introduction: VisuMAX is a medical laser device for refractive surgery built by Zeiss. Each surgery outputs a video clip that depicts the whole surgical procedure. Every video consists of a consecutive sequence of six different surgical steps. The goal of this work is to automatically segment the eye surgery videos into these classes using a deep learning approach. The segmentation is required to enable further video-based analysis of the surgery. This is task is closely related to the shot boundary detection problem, which finds boundaries between two video shots.

Approach: Two different approaches have been implemented in order to solve this problem.

At first, a 3D convolutional neural network has been trained on artificially generated video sequences taken from a TV data set. This neural network is able to detect shot boundaries accurately on TV clips. However, it has a poor performance on the surgical videos.

Consequently, a second deep convolutional neural network has been designed to perform a content classification on every frame. Six possible classes have been defined, where each represents a part of the surgery. First, each frame is pre-processed and fed through this frame classification network. Then, the most probable class sequence is evaluated by post-processing the class probabilities. The neural network has been trained end-to-end using a provided batch of 30 eye surgery videos. In order to provide ground truth, each video frame has been labeled by a human operator using a dedicated labeling tool.

A Python API using Google Tensorflow has been built. It includes components for training and inference of the neural network, as well as tools for statistics and data set handling.

Conclusion: The deep convolutional neural network in combination with a Viterbi or long short-term memory (LSTM) neural network based post-processing algorithm is able to segment the videos properly. Over 90% of the predicted boundaries have less than two seconds deviation from the labeled data. Less than 10% poorly detected transitions remain. These outliers can be partly attributed to a low amount of training data.