Dense Face Detection and Improving Temporal Convolutional Networks for Automatic Image Captioning

Zozoulenko, Nikita (School: Katedralskolan in Linkoping)

Convolutional neural networks (CNN) are traditionally used in computer vision, but have recently gained traction as sequential models. In this paper two improvements to temporal convolutional networks (TCN) are suggested, together with three different residual blocks, and empirically evaluated to achieve a new state of the art on the task of SequenceMNIST (99.53% versus 99.0%) and Permuted SequenceMNIST (97.86% versus 97.28%). The improved model was then compared to a Gated Recurrent Unit in an area where TCNs have never been applied before, namely automatic image captioning, and is shown to perform similarly. Additionally, since the field lacks a fully derived order 4 example of a convolutional neural network (CNN), we presented a clear derivation of the general case for a CNN with an arbitrary input of a tensor of order 4, including varying zero-padding and strides during forward and backpropagation. The derivation was then used for our convolutional neural network library implementation in Python and C++. Furthermore, we systematically construct a one-shot dense face detector and empirically investigate how factors such as binary cross entropy loss, focal loss, feature pyramid networks, online hard example mining, color jitter data augmentation, and network depth affects its performance. The final model is sufficiently accurate and computationally efficient to be used for real time video, to serve as a baseline model for any system which incorporates surveillance, security or the detection of humans.

Awards Won:

Third Award of $1,000

Association for Computing Machinery: Fourth Award of $500

Association for the Advancement of Artificial Intelligence: First Award of $1,500