Feature analysis of two stream CNNs for action recognition

Will Price (wp13824@bristol.ac.uk), Dima Damen (csxda@bristol.ac.uk)

University of BRISTOL

Abstract

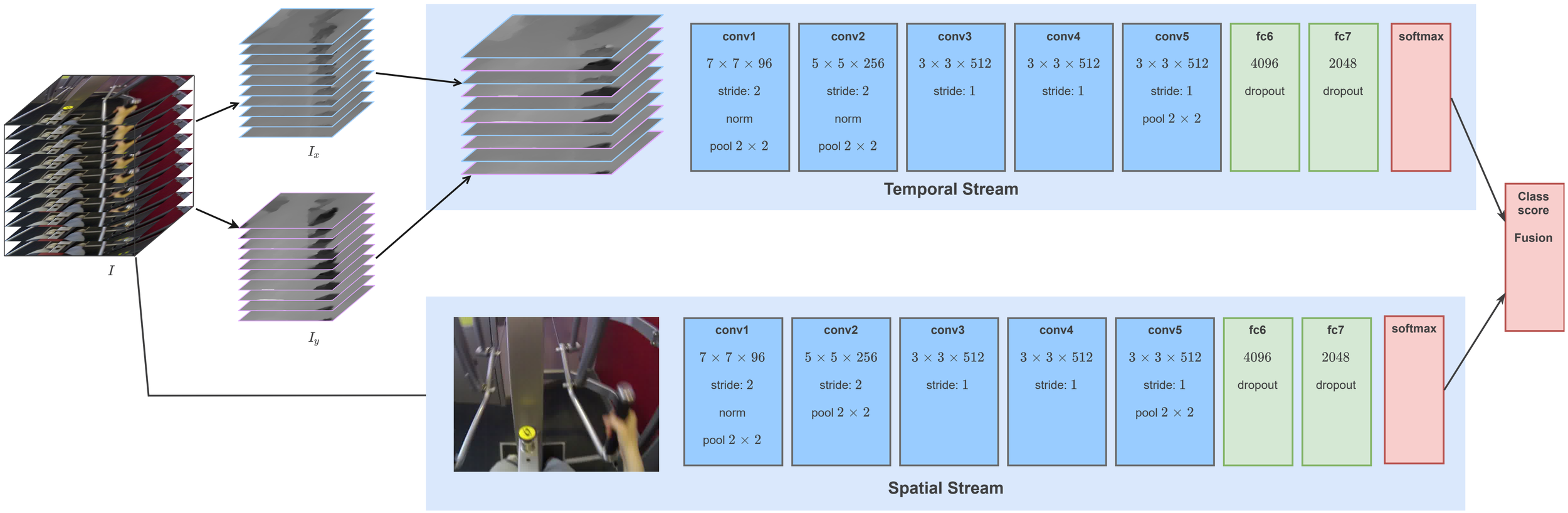
Convolutional neural networks (CNNs) recently achieved state of the art performance in object detection but their lack of transparency makes them difficult to debug and understand. Towards solving this problem, a technique, **excitation backpropagation (EBP)**, has been developed to determine which regions in the input of an image excite arbitrary neurons in the network, this technique can help understand what a network has learnt and whether it has generalised or overfit specific examples.

EBP has been used to inspect object detection networks, but not two stream CNNs, a CNN architecture for action recognition from video sequences.

Can EBP produce interpretable results on two stream CNNs for action recognition?

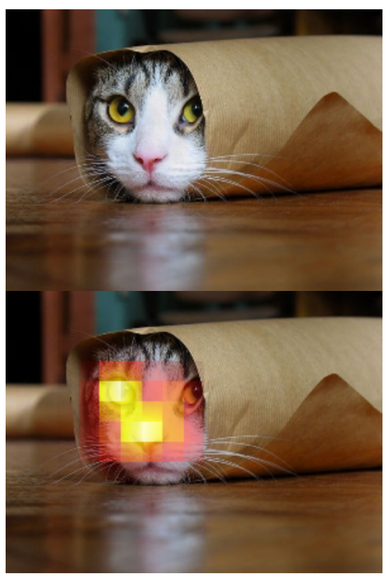
Two stream CNN

Two stream CNNs for action recognition are formed of two separate networks whose results are 'fused' together. One network operates on a single video frame considering *spatial* information. the second receives *temporal* information from a stack of optical flow frames.

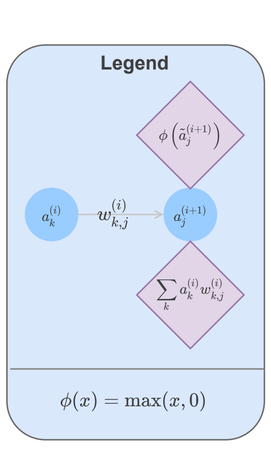


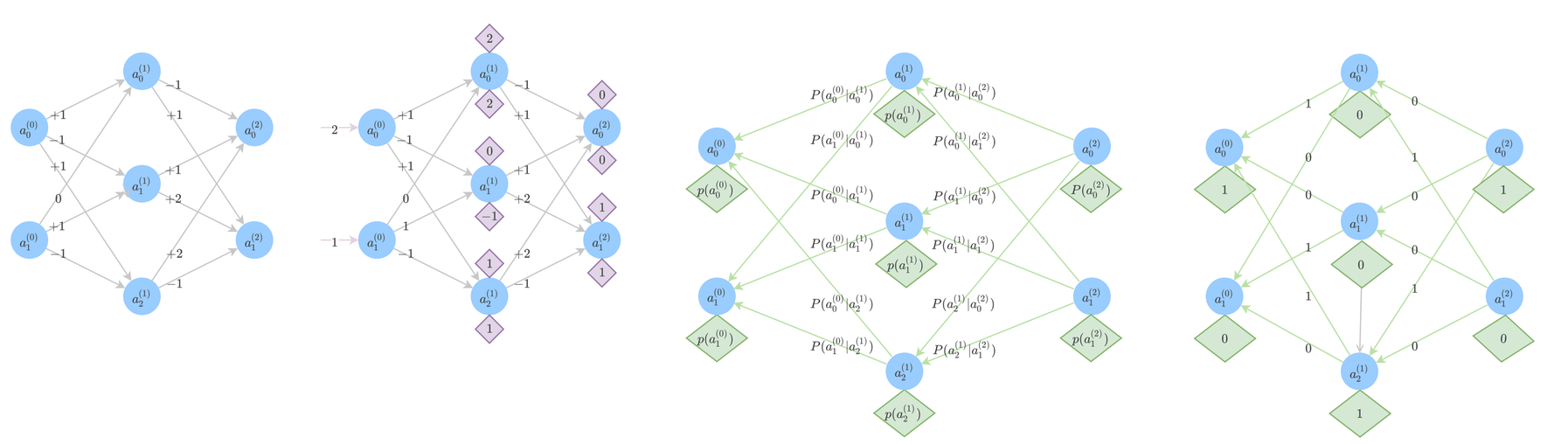
Excitation Back-propagation

Excitation back-propagation (EBP) is a technique to determine the regions in the input of a network that contribute to the excitation of an arbitrary neuron. By understanding the regions exciting a neuron we can infer certain features the network has learnt.



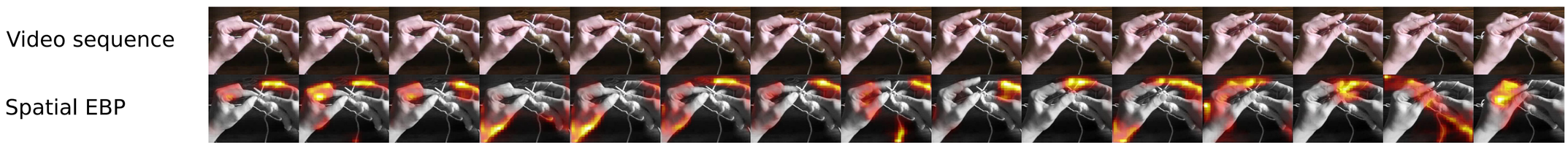
EBP of 'cat' neuron in an object detection CNN

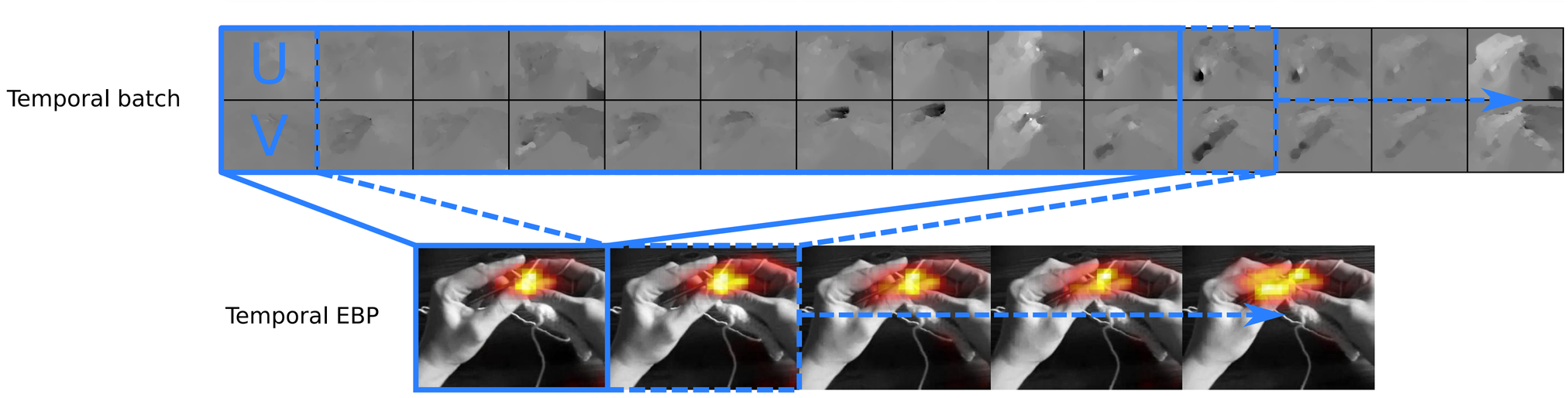




EBP example on a network of fully connected layers

Application of Excitation Back-propagation to two stream CNNs





Results