Go with the Flow: Deep Learning Methods for Autonomous Viscosity Estimations

Michael Walker, Gabriella Pizzuto, Hatem Fakhruldeen, and Andrew I. Cooper

1 Model Architecture

1.1 Viscosity Models

The structure of the 3DCNN consisted of two convolutional layers, a series of batch normalisations, activation functions and three fully connected layers. The first convolutional layer took one input channel and output 32 channels. The initial input size was 256x342 with the time dimension dependant on the liquid (30 for honey, 10 for viscosity standards). These frames were evenly distributed across the entire video. A kernel of size (5, 5, 5) and stride (2, 2, 2) was used for the first convolution. The second convolutional layer had 32 input channels and 48 output channels, with kernel size (3, 3, 3) and stride (2, 2, 2). 3D batch normalisation, a ReLU activation function, and a 3D max pooling layer with kernel size 2, stride 2 and padding 0 was used for each layer. 3 FC layers were used to map the features to the final output of the network. The first two FC layers had 256 hidden units, with the final layer altered to either 1 (regression) or the number of classification categories (in this case, five).

1.2 Solvent Identification Classification Models

The 3DCNN structure consisted of an input layer, which used frames of size 256x342, two convolutional layers, each with a ReLu activation layer and 3-dimensional batch normalisation. This was followed by three fully connected layers, which reduced the features from the convolutional layers from 1499904 to five (number of classification categories). The kernel size and stride for the first convolutional layer was (5, 5, 5) and (2, 2, 2), respectively. The kernel size and stride for the first convolutional layer was (3, 3, 3) and (2, 2, 2), respectively. The kernel size for the first and second convolutional layers was (5, 5, 5) and (3, 3, 3), respectively. Each convolutional layer had a stride of (2, 2, 2) and padding of zero. Dropout was also set to zero. A max pooling layer 18 (3-dimensional) was also used with kernel size 2, stride 2 and padding zero. The time dimension of the network (i.e., the number of frames inputted to the model) was set to 29. The input frames used for the model did not have equal intervals between start and finish. Frames 3-19 and 45-58 were chosen.