Hidden layers	3
Nodes per hidden layer	512
Activation function	ReLu
Output activation function	Softmax
Loss function	categorical crossentropy
Pre-processing	`BatchNormalization`
Dropout rate	20%
Training epochs	100
Batch size	4032/10648/15972
Steps per epoch	52/23/15
Initial learning rate	$5 \cdot 10^{-4}$
Learning rate decay	0.8
Learning rate decay patience	3 epochs