Differences between the models

Baseline	Improved
Conv. 1 (Output_channels = 6)	Conv. 1 (Output_channels = 10)
Conv. 2 (Input_channels = 6, Output_channels = 12, kernel = 4)	Conv. 2 (Input_channels = 10, Output_channels = 20, kernel = 5)
MaxPooling2 (kernel = 3)	MaxPooling2 (kernel = 2)
Epochs = 25	Epochs = 120
Learning rate = 0.001	Learning rate = 0.0009
Optimizer = stochastic gradient descent	Optimizer = ADAM Algorithm