

Logan Griffin

HW and PA 3

8/3/2025

EN.665.645.8VL.SU25

Artificial Intelligence for Robotics

Section 3: Implementing a CNN

Below are two tables that show how varying the hyperparameters such as learning rate, number of epochs, channels, and the optimizing method affect the validation accuracy of a CNN on the CIFAR-10 dataset. The first table shows results for a basic CNN implementation that has 3 convolution layers followed by 1 linear layer. The convolutional layers have a kernel size of 3 with a padding of 1. ReLU activation is applied on each layer. The second table shows results for a CNN implementation that has the same 3 convolutional layers, but with a max pooling layer added after each. The max pooling layers have a kernel size of 2 and a stride of 2. The runs with 512 channels per layer were taking a long time to run on my machine, so there are less trials for that setting. I tried different learning rates and epoch combinations for each combination of channels and optimizer. The validation accuracies are listed as well as if the parameters seemed to lead to overfitting and when that occurred during the test. The highlighted accuracies show which parameters tested led to the best results. It is likely that other combinations may yield a better accuracy, but I was limited by the capabilities of my machine. Overall, the Adam optimizer required less epochs than the SGD optimizer. The max pooling CNN also led to higher validation accuracies than the regular CNN implementation. As the number of channels increased, so did the run time. A higher number of channels, however, did lead to faster convergence among the tests.

Basic CNN Implementation Data:

Channels	Learning Rate	Epochs	Optimizer	Validation Accuracy	Observations
(128, 128, 128)	0.0001	20	SGD	37.91%	
(128, 128, 128)	0.001	5	SGD	41.99%	
(128, 128, 128)	0.001	10	SGD	50.04%	
(128, 128, 128)	0.001	50	SGD	60.65%	
(128, 128, 128)	0.01	10	SGD	63.77%	
(128, 128, 128)	0.01	20	SGD	64.40%	Acc. decrease at 15 th epoch → overfitting
(128, 128, 128)	0.02	10	SGD	64.49%	Acc. decrease at 9 th epoch → overfitting
(128, 128, 128)	0.05	10	SGD	64.73%	Acc. decrease at 7 th epoch → overfitting
(128, 128, 128)	0.1	10	SGD	60.35%	Acc. decrease at 6 th epoch → overfitting
(128, 128, 128)	0.1	5	SGD	63.13%	
(256, 256, 256)	0.001	10	SGD	51.38%	
(256, 256, 256)	0.005	20	SGD	65.36%	
(256, 256, 256)	0.01	10	SGD	64.81%	
(256, 256, 256)	0.01	15	SGD	64.31%	Acc. decrease at 12 th epoch → overfitting
(256, 256, 256)	0.05	10	SGD	64.60%	Acc. Decrease at 6 th epoch → overfitting
(256, 256, 256)	0.1	10	SGD	64.30%	Acc. decrease at 6 th epoch → overfitting

(512, 512, 512)	0.001	10	SGD	54.32%	
(512, 512, 512)	0.01	10	SGD	64.74%	
(512, 512, 512)	0.1	10	SGD	64.89%	
Channels	Learning Rate	Epochs	Optimizer	Validation Accuracy	Observations
(128, 128, 128)	0.00008	5	Adam	67.29%	
(128, 128, 128)	0.00009	5	Adam	66.77%	
(128, 128, 128)	0.0001	5	Adam	66.42%	
(128, 128, 128)	0.0003	4	Adam	66.29%	
(128, 128, 128)	0.0004	4	Adam	66.72%	
(128, 128, 128)	0.0005	5	Adam	64.95%	
(128, 128, 128)	0.001	5	Adam	55.87%	
(128, 128, 128)	0.005	5	Adam	50.53%	Acc. decrease at 4 th epoch → overfitting
(256, 256, 256)	0.00005	5	Adam	67.08%	
(256, 256, 256)	0.00008	5	Adam	66.80%	Acc. decrease at 4 th epoch → overfitting
(256, 256, 256)	0.0001	3	Adam	66.58%	
(256, 256, 256)	0.001	3	Adam	59.83%	
(256, 256, 256)	0.001	5	Adam	50.31%	
(512, 512, 512)	0.00005	3	Adam	66.92%	
(512, 512, 512)	0.00006	3	Adam	66.58%	
(512, 512, 512)	0.0001	5	Adam	62.83%	Acc. decrease at 4 th epoch → overfitting
(512, 512, 512)	0.001	10	Adam	54.76%	Acc. decrease at 4 th epoch → overfitting

Max Pooling CNN Implementation Data:

Channels	Learning Rate	Epochs	Optimizer	Validation Accuracy	Observations
(128, 128, 128)	0.001	50	SGD	56.29%	
(128, 128, 128)	0.001	75	SGD	60.55%	
(128, 128, 128)	0.005	35	SGD	69.92%	
(128, 128, 128)	0.01	25	SGD	70.04%	
(128, 128, 128)	0.01	50	SGD	72.37%	Acc. decrease at 47 th epoch → overfitting
(128, 128, 128)	0.1	25	SGD	73.80%	
(256, 256, 256)	0.001	75	SGD	63.12%	
(256, 256, 256)	0.01	10	SGD	62.77%	
(256, 256, 256)	0.01	20	SGD	69.60%	
(256, 256, 256)	0.05	12	SGD	73.23%	
(256, 256, 256)	0.1	10	SGD	72.04%	Acc. decrease at 8 th epoch → overfitting
(512, 512, 512)	0.01	20	SGD	71.06%	
(512, 512, 512)	0.05	15	SGD	75.07%	
(512, 512, 512)	0.1	12	SGD	74.05%	
Channels	Learning Rate	Epochs	Optimizer	Validation Accuracy	Observations
(128, 128, 128)	0.0001	15	Adam	69.73%	
(128, 128, 128)	0.0005	10	Adam	73.23%	
(128, 128, 128)	0.001	5	Adam	72.96%	

(256, 256, 256)	0.0001	10	Adam	72.65%	
(256, 256, 256)	0.003	8	Adam	74.09%	
(256, 256, 256)	0.005	8	Adam	72.74%	Acc. decrease at 6 th epoch → overfitting
(256, 256, 256)	0.001	10	Adam	69.16%	Acc. decrease at 7 th epoch → overfitting
(512, 512, 512)	0.0001	5	Adam	72.46%	
(512, 512, 512)	0.0005	5	Adam	72.06%	
(512, 512, 512)	0.001	4	Adam	67.22%	