Report: Cifar-10 Classification with MLP and CNN

2018011365 张鹤潇

实验内容

在本次实验中，我用 `pytorch` 实现了 Dropout, Batch Normalizaiton 两个模块，并用其搭建 MLP 和 CNN，在 Cifar-10 数据集上测试模型的性能。

实验设置

MLP 的结构如下，

Layer	Type	In Dim	Out Dim
0	Linear	3072	512
1-3	BN + ReLU + Dropout	512	512
4	Linear	512	10

CNN 的结构如下，

Layer	Type	In Dim	Out Dim
0	Conv	(3, 32, 32)	(128, 32, 32)
1-3	BN + ReLU + Dropout	(128, 32, 32)	(128, 32, 32)
4	MaxPool	(128, 32, 32)	(128, 16, 16)
5	Conv	(128, 16, 16)	(256, 16, 16)
6-8	BN + ReLU + Dropout	(256, 16, 16)	(256, 16, 16)
9	MaxPool	(256, 16, 16)	(256, 8, 8)
10	Flatten	(256, 8, 8)	16384
11	Linear	16384	10

实验结果与思考题

Explain how `self.training` work. Why should training and testing be different?

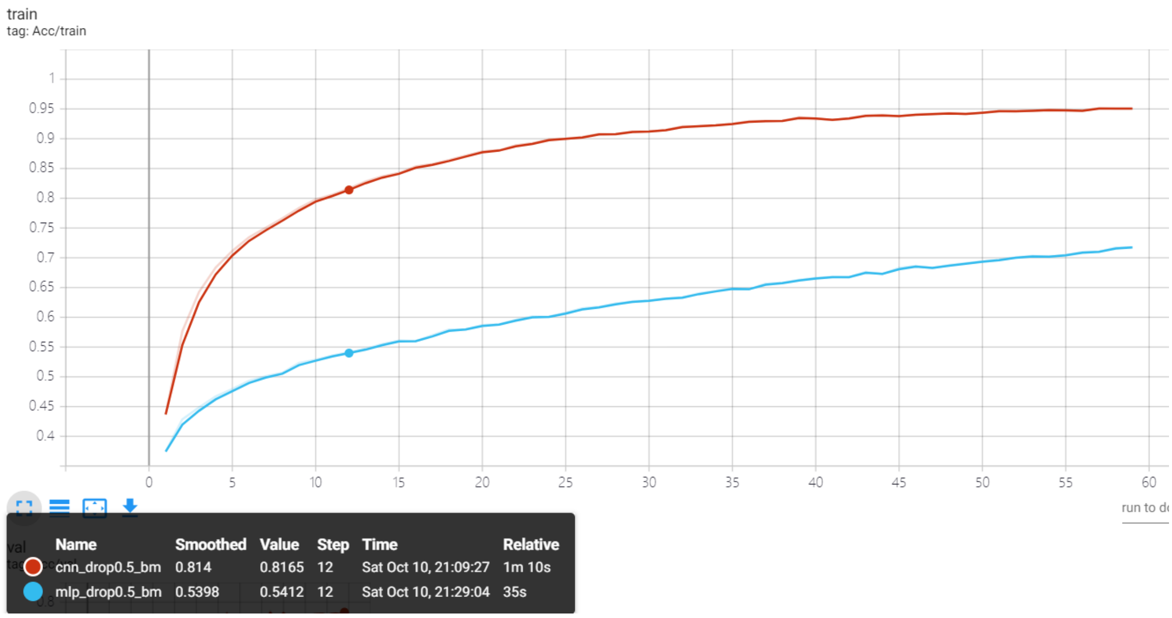
在训练模型 (`model.train()`) 时，`self.training` 设置为 `true`，在测试 (`model.eval()`) 时则为 `false`，借此控制 Dropout 和 Batch Normalization 的行为。

训练时，Dropout 发挥作用，Batch Normalization 使用整个 batch 的均值和方差计算输出，同时更新移动平均值；测试时，Dropout 不再起效，Batch Normalization 用均值和方差移动平均值计算输出，而不再使用该 batch 的统计量。

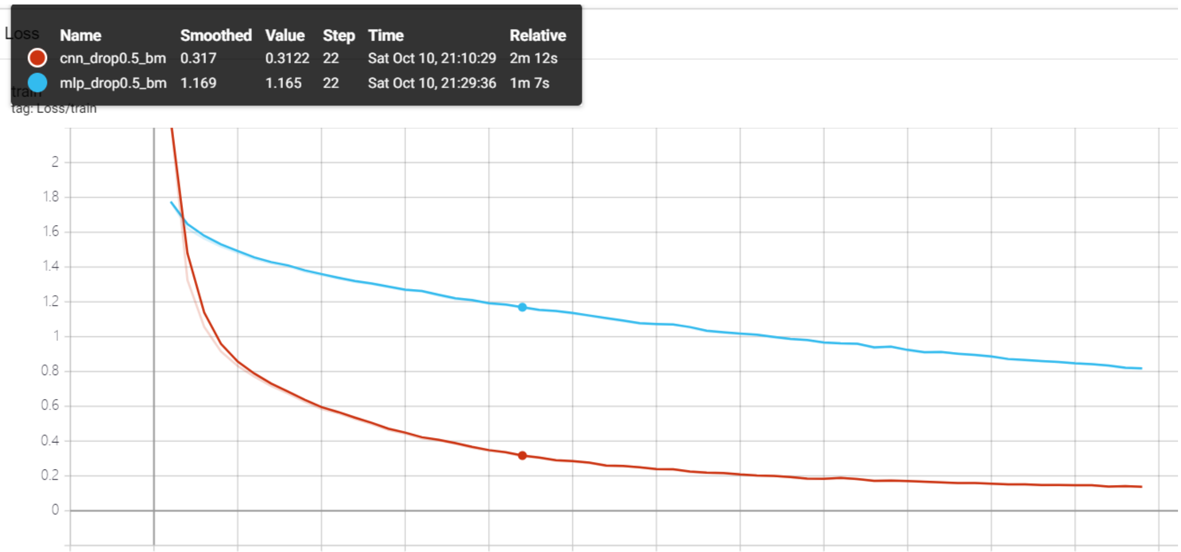
Construct the MLP and CNN with batch normalization and dropout. Write down the hyperparameters that you use to obtain the best performance. Plot the loss value and accuracy (for both training and validation) against to every iteration during training.

`epochs` 设置为60，其它超参数均使用默认值。

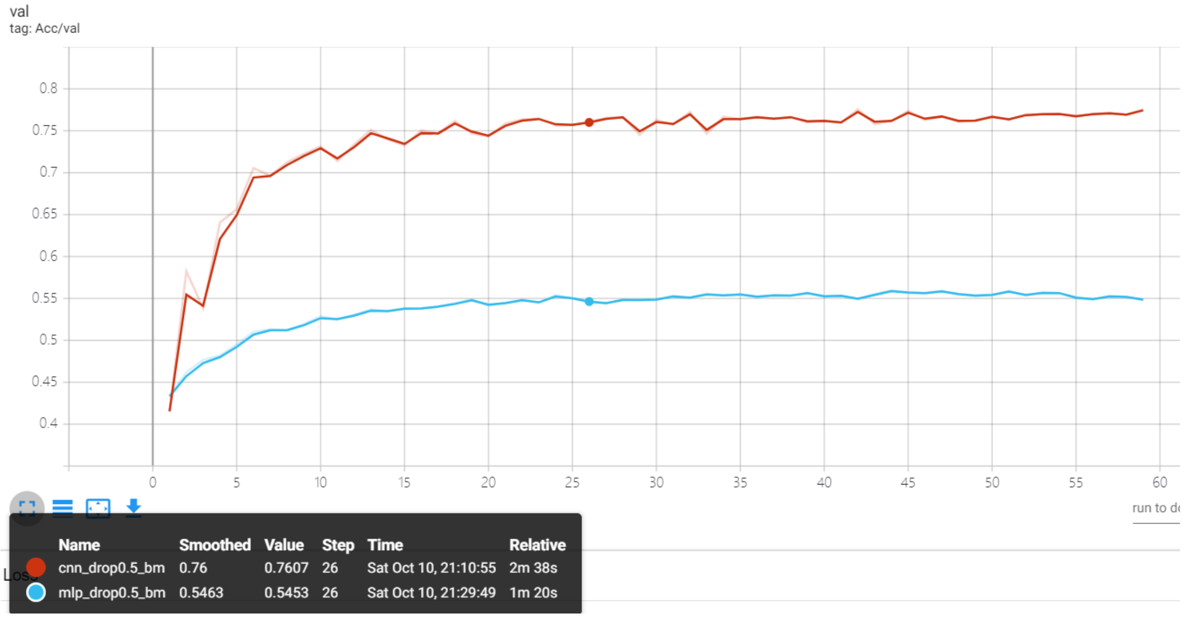
训练集准确率



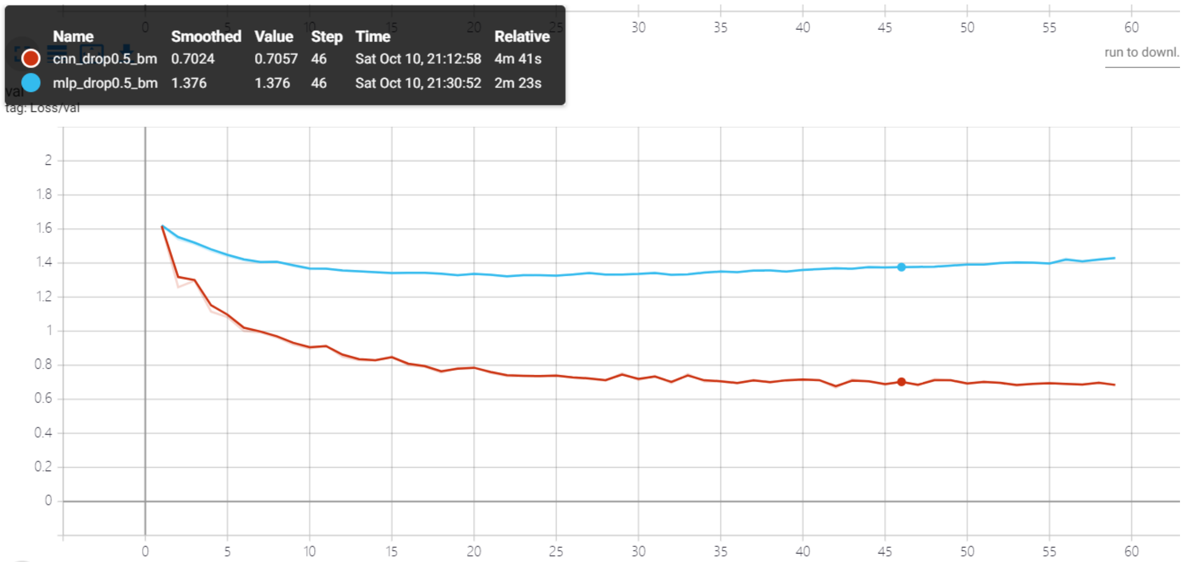
训练集Loss



验证集准确率



验证集Loss



Explain why training loss and validation loss are different. How does the difference help you tuning hyper-parameters?

主要有两方面的原因：训练集与验证集的数据分布不同，模型在两个阶段的计算方式 (Dropout 和 Batch Normalization) 不同。 实验表明，去掉 Batch Normalization 层后，训练集与验证集的 Loss 之差会变小（见下文）。

Loss 的差异可以帮我们诊断模型的问题。若训练集和验证集上的 Loss 都比较高，则模型发生了欠拟合；若训练集上 Loss 较低而验证集 Loss 偏高，则模型的泛化能力可能不足。以此为指标调整超参数，可以优化模型的性能。

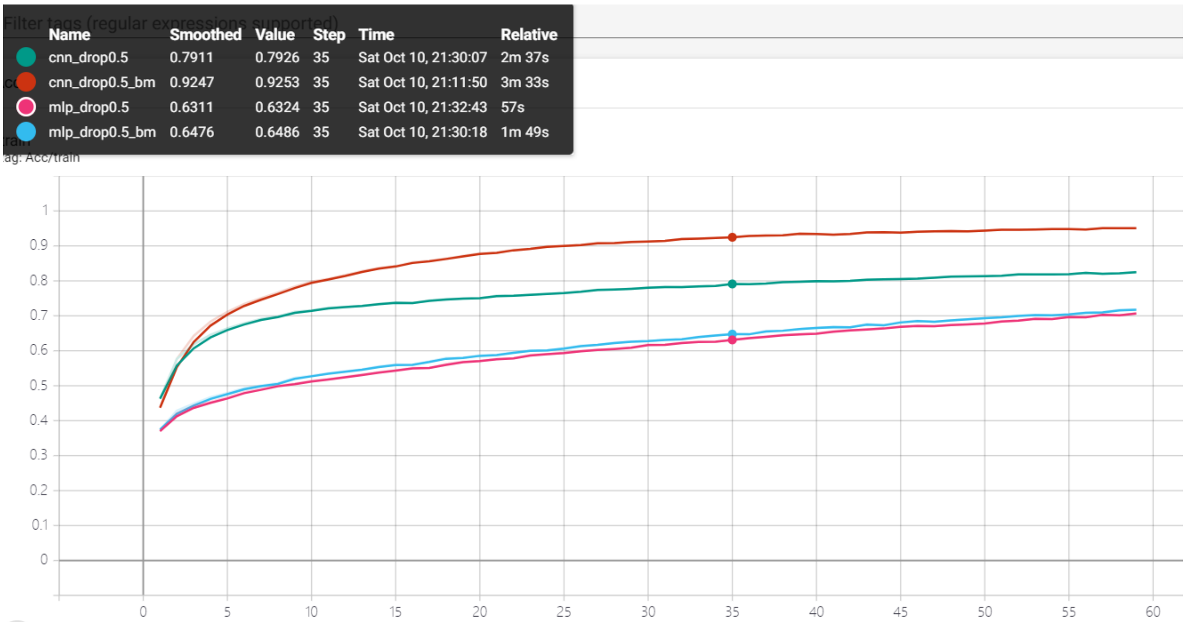
Report the final accuracy for testing. Compare the differences between the results of MLP and CNN.

Model	Train Loss	Train Acc	Val Loss	Val Acc	Test Acc
MLP	0.9442	67.26 %	1.3782	55.97 %	55.16 %
CNN	**0.1366**	**95.06 %**	**0.6818**	**77.57 %**	**77.22 %**

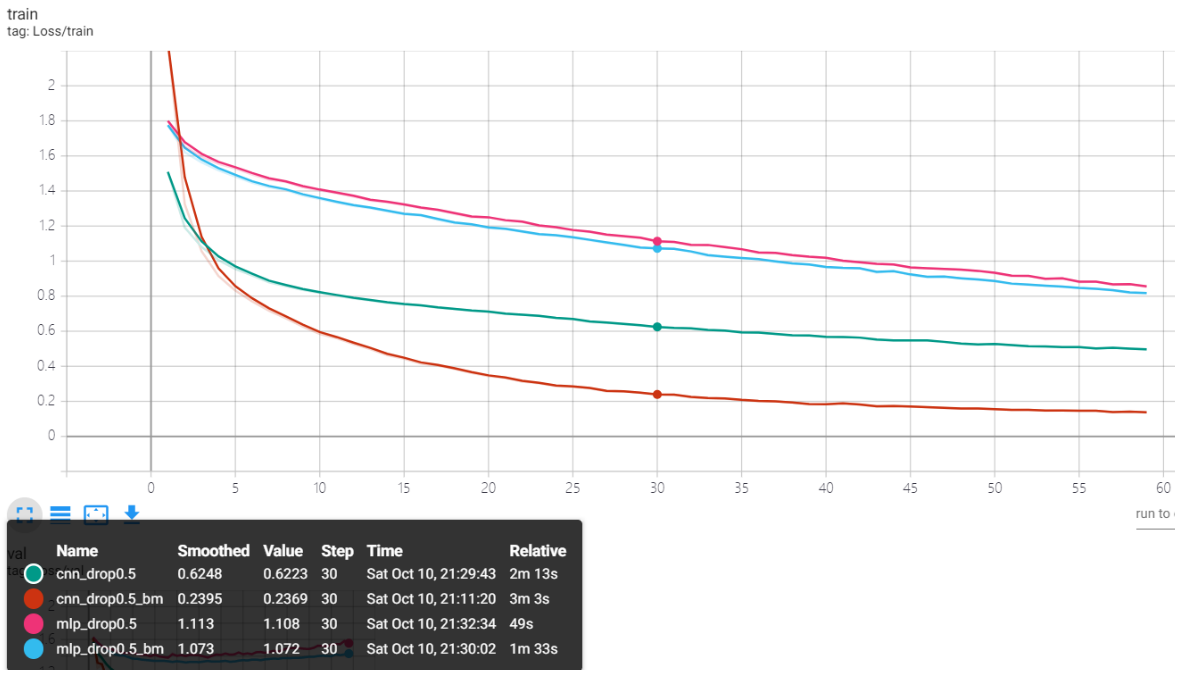
CNN 在各项指标上均优于 MLP。对于图像数据，CNN 有更强的拟合能力和泛化能力，能够达到更优的性能。

Construct MLP and CNN without batch normalization, and discuss the effects of batch normalization.

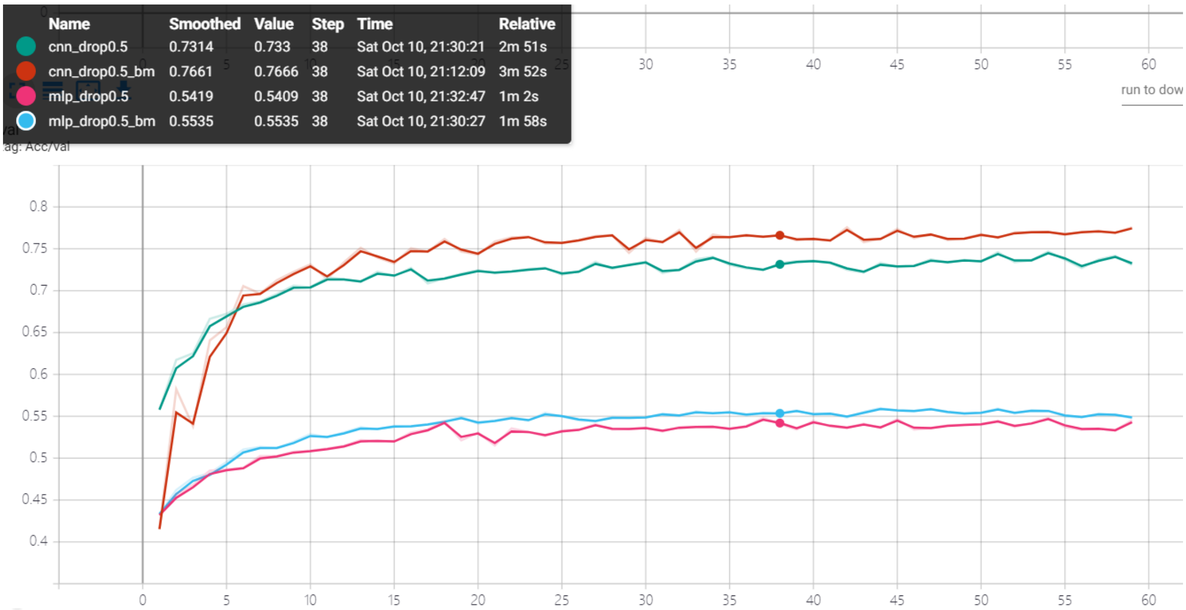
训练集准确率



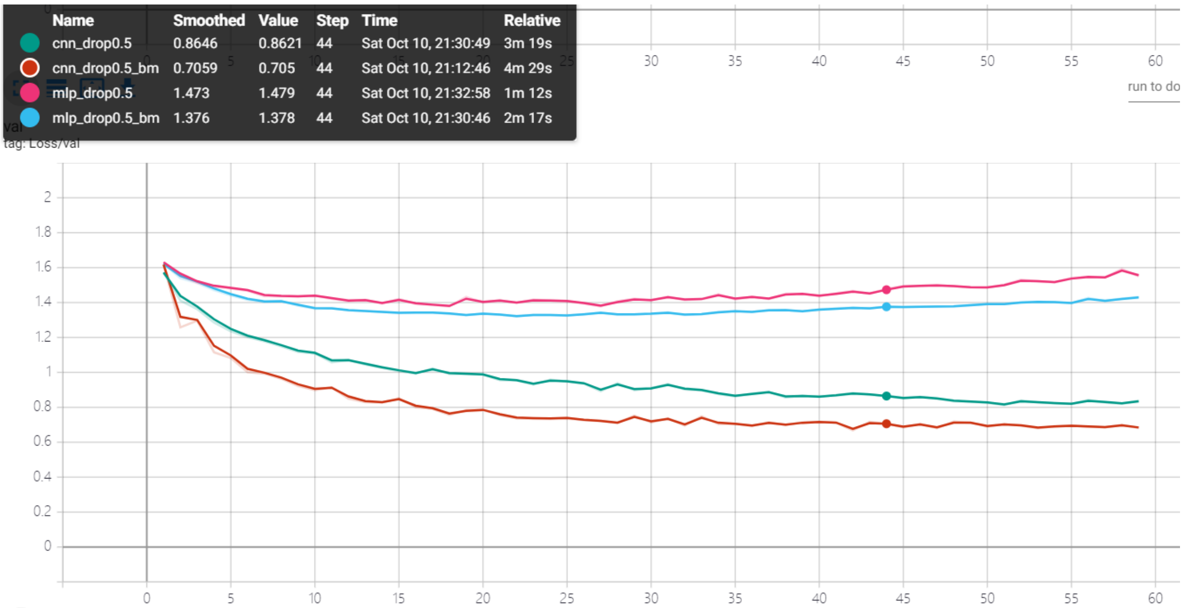
训练集Loss



验证集准确率



验证集Loss

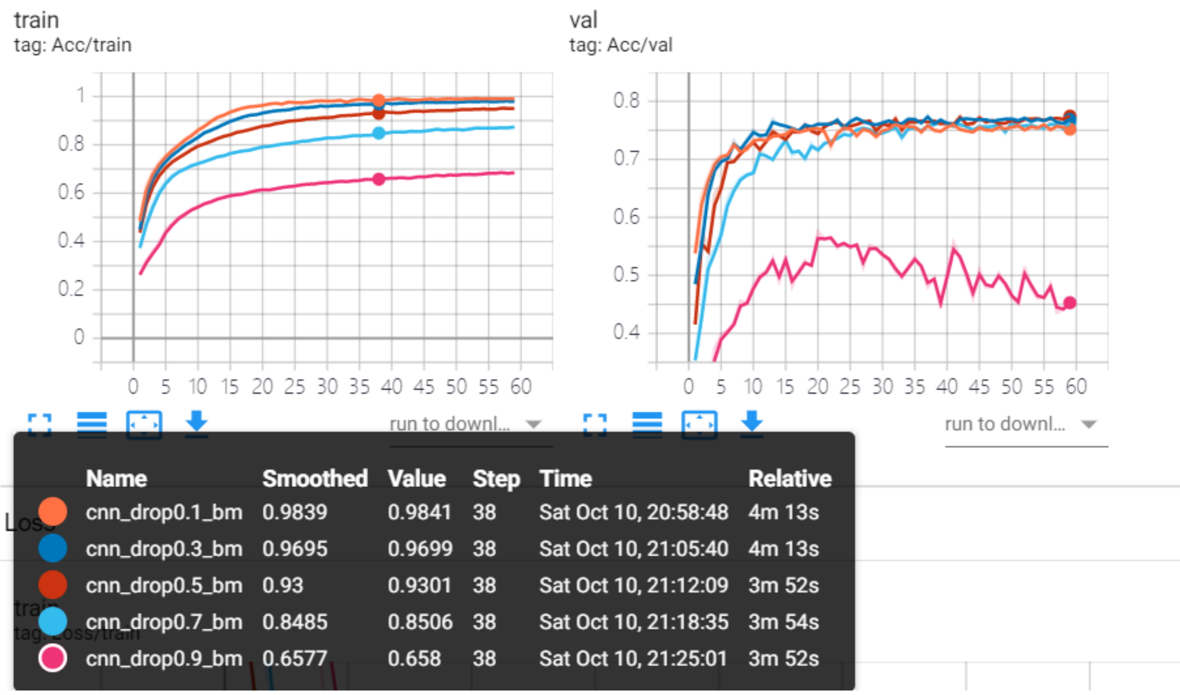


Model	BN	Val Loss	Val Acc	Test Acc
MLP	有	**1.3782**	**55.97 %**	**55.16 %**
MLP	无	1.5157	54.81 %	53.98 %
CNN	有	**0.6818**	**77.57 %**	**77.22 %**
CNN	无	0.8229	74.68 %	74.08 %

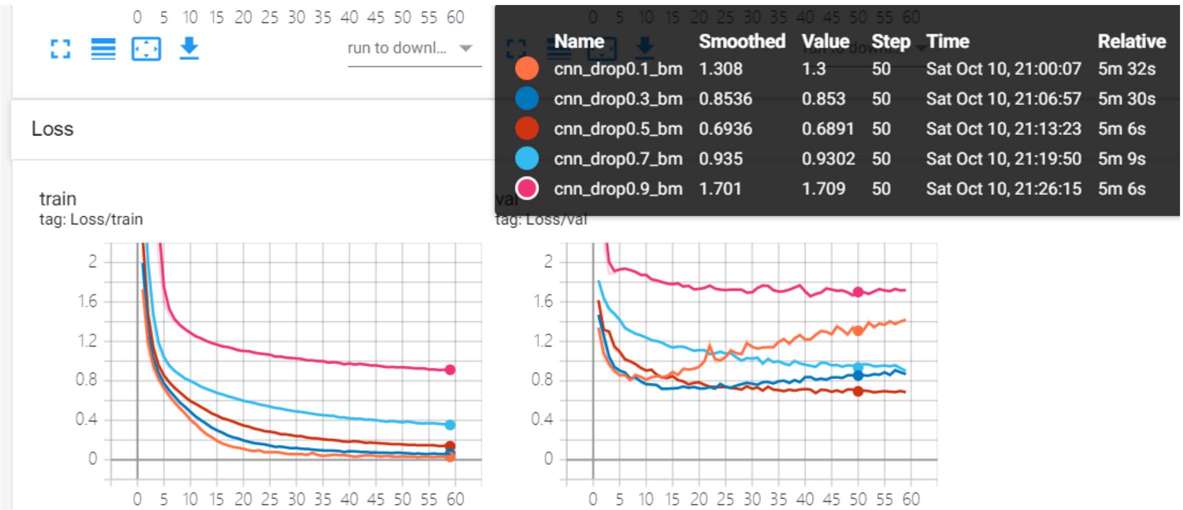
Batch Normalization 能够加速训练过程，帮助模型更快地收敛，获得更高的测试性能。它对 CNN的优化效果高于 MLP。

Tune the drop rate for MLP and CNN, respectively, and discuss the effects of dropout.

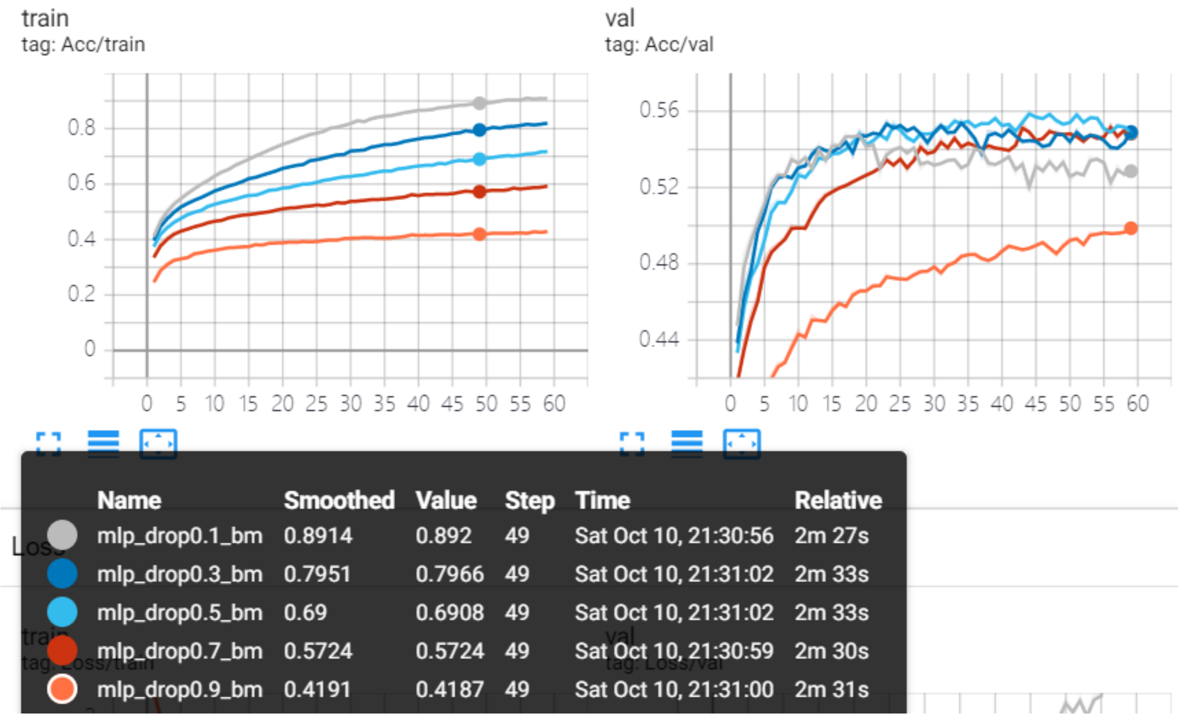
CNN Acc



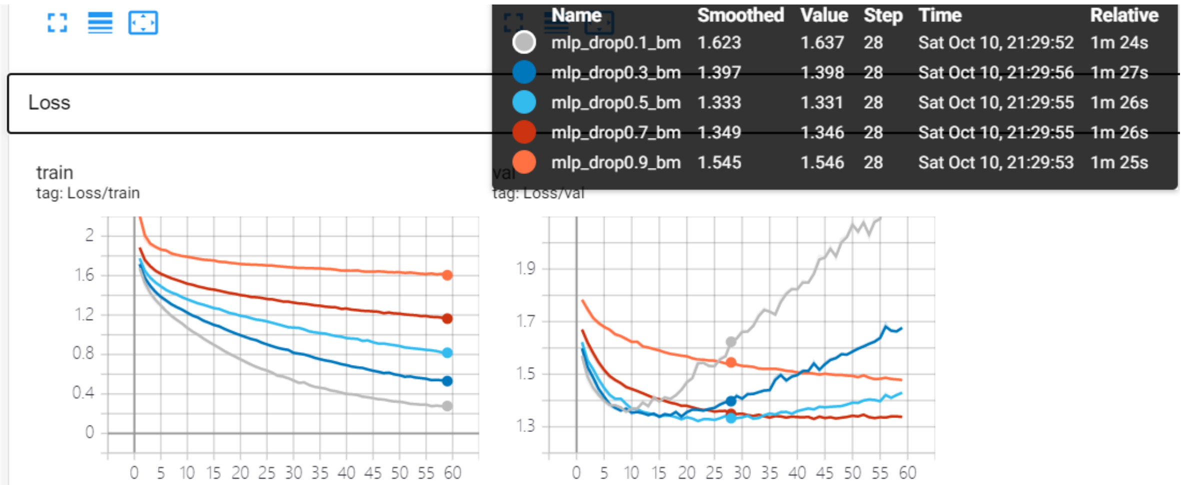
CNN Loss



MLP Acc



MLP Loss



对于我的模型 (MLP/CNN)，其最佳 dropout rate 为 0.5.

在训练阶段，模型的收敛速度随着 dropout rate 的增加而减慢。当 dropout rate = 0.1 时，模型明显发生了过拟合；增大 dropout rate (≥ 0.3) 能有效缓解过拟合问题，但过大的 dropout rate (≥ 0.9) 又会使模型难以收敛。

Dropout 实际上相当于数据增强或者集成学习，具有防止过拟合的效果。