

实验三：卷积神经网络

一、实验任务

本实验旨在设计并实现一个基于卷积神经网络（CNN）的图像分类器，以理解CNN的基本结构、代码实现和训练过程。同时，通过应用dropout和多种normalization方法，分析它们对模型泛化能力的影响，并利用交叉验证找到神经网络的最佳超参数。

二、数据集

本实验使用MNIST手写数字数据集进行训练和测试。MNIST数据集包含60,000个训练图像和10,000个测试图像，每个图像为28x28像素的灰度手写数字。

三、模型结构

模型采用了以下结构：

- **Convolutional Layer 1:** 第一个卷积层使用32个过滤器，每个大小为3x3，步长为1。
- **Convolutional Layer 2:** 第二个卷积层使用64个过滤器，每个大小为3x3，步长为1。
- **Dropout Layer 1:** 在第一个全连接层之前使用dropout层，dropout率为0.25。
- **Fully Connected Layer 1:** 第一个全连接层将特征图转换为128维的向量。
- **Dropout Layer 2:** 在第二个全连接层之前使用dropout层，dropout率为0.5。
- **Fully Connected Layer 2:** 第二个全连接层输出10个类别的概率分布。

fc2

weight $\langle 10 \times 128 \rangle$
bias $\langle 10 \rangle$

fc1

weight $\langle 128 \times 9216 \rangle$
bias $\langle 128 \rangle$

conv2

weight $\langle 64 \times 32 \times 3 \times 3 \rangle$
bias $\langle 64 \rangle$

conv1

weight $\langle 32 \times 1 \times 3 \times 3 \rangle$
bias $\langle 32 \rangle$

四、超参数搜索

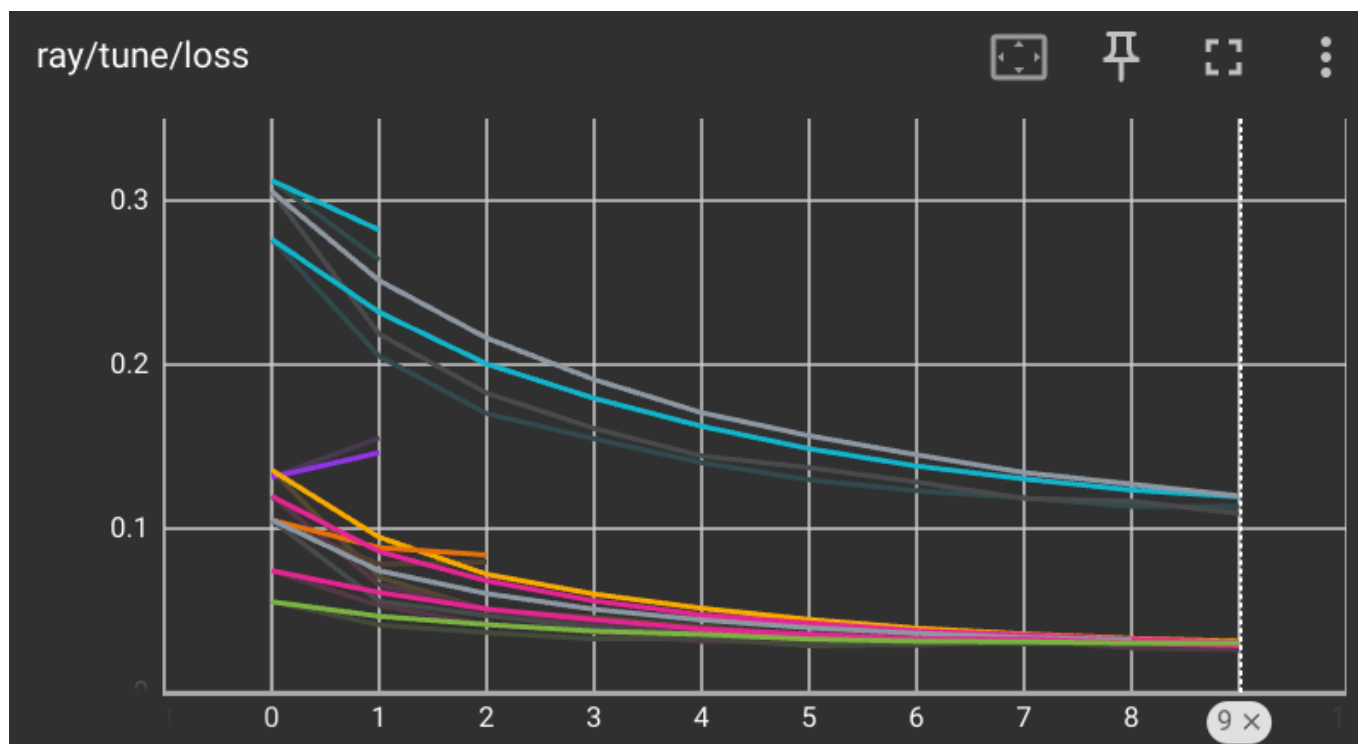
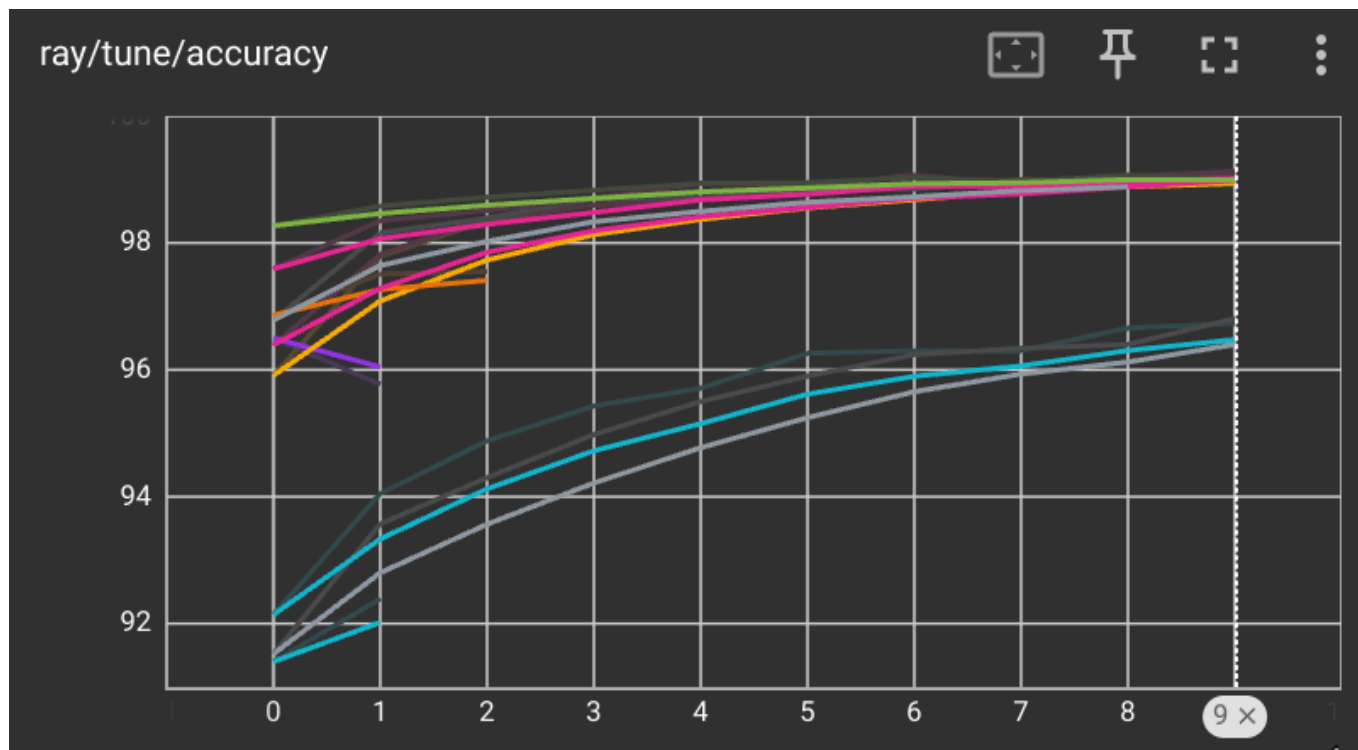
使用Tune的Hyperopt算法完成超参数搜索，最大迭代次数设置为10，每次将训练10个epoch作为测试，找到最佳的超参数组合为（以验证集准确率为标准）：

```
Trial status: 10 TERMINATED
Current time: 2023-11-08 02:04:04. Total running time: 1hr 25min 32s
Logical resource usage: 0/24 CPUs, 1.0/1 GPUs (0.0/1.0 accelerator_type:G)
Current best trial: 16486_00002 with accuracy=99.13 and params={'lr': 0.0014489483362154264, 'batch_size': 32, 'momentum': 0.9035754520022181, 'weight_decay': 0.0001114030365842323}
```

Trial name	status	lr	batch_size	momentum	weight_decay	iter	total time (s)	accuracy	loss
train_mnist_16486_00000	TERMINATED	0.000798787	128	0.88067	0.00799546	9	680.793	96.81	0.109284
train_mnist_16486_00001	TERMINATED	0.00162136	128	0.830231	0.0177367	9	653.442	96.73	0.113431
train_mnist_16486_00002	TERMINATED	0.00144895	32	0.903575	0.000111403	9	656.309	99.13	0.0277521
train_mnist_16486_00003	TERMINATED	0.00129814	32	0.895558	2.01891e-05	9	655.522	99.02	0.0288223
train_mnist_16486_00004	TERMINATED	0.0582407	128	0.880001	0.0232443	1	132.491	95.77	0.155309
train_mnist_16486_00005	TERMINATED	0.00873297	32	0.925621	0.000243345	9	643.942	99	0.0301756
train_mnist_16486_00006	TERMINATED	0.00959941	64	0.852646	0.0147908	2	192.241	97.55	0.0796727
train_mnist_16486_00007	TERMINATED	0.0035915	128	0.961384	0.000924881	8	718.075	98.97	0.0298259
train_mnist_16486_00008	TERMINATED	0.000265855	128	0.981937	0.085343	1	134.007	92.39	0.263929
train_mnist_16486_00009	TERMINATED	0.00439706	32	0.825614	3.02745e-06	9	643.668	99.09	0.0260933

```
{'lr': 0.0014489483362154264, 'batch_size': 32, 'momentum':  
0.9035754520022181, 'weight_decay': 0.0001114030365842323}
```

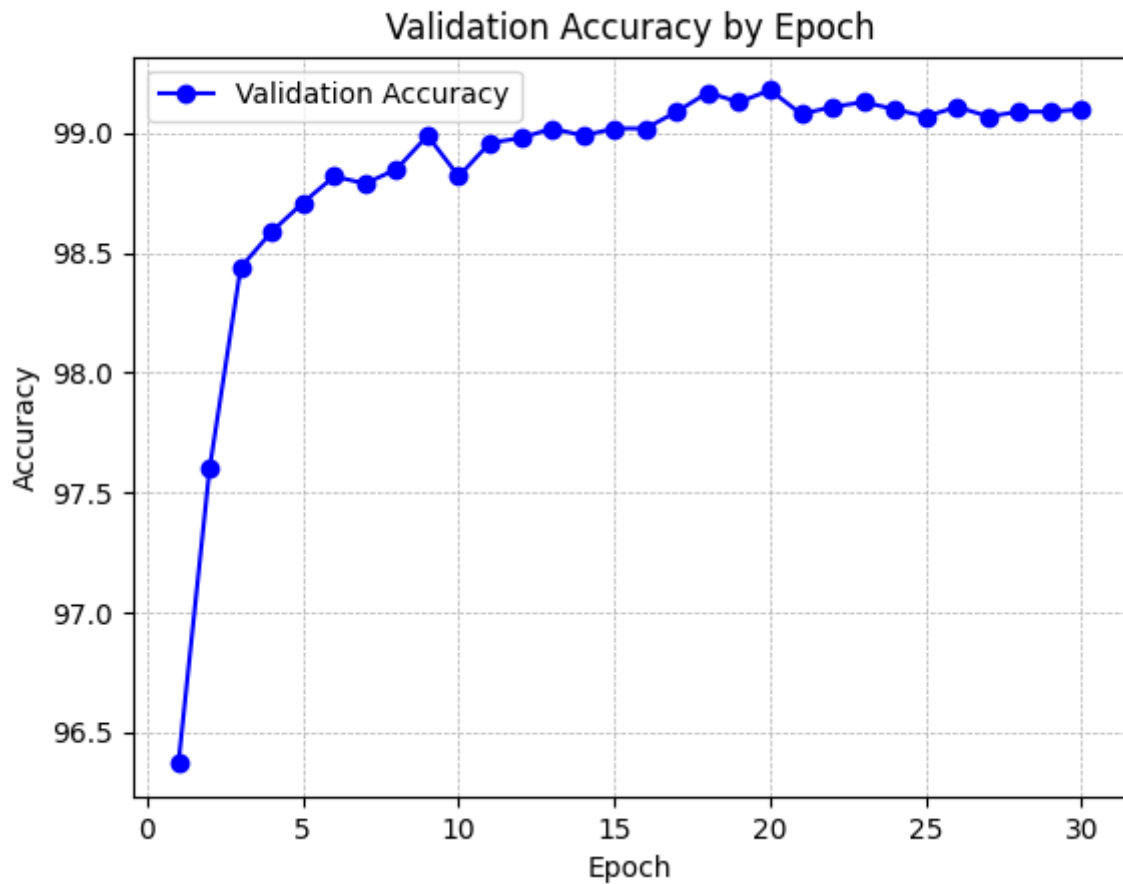
在验证集上的 最佳准确率为99.13%
搜索过程可视化如下：（横轴代表迭代次数）



图中绿线代表最佳的超参数组合

五、训练过程

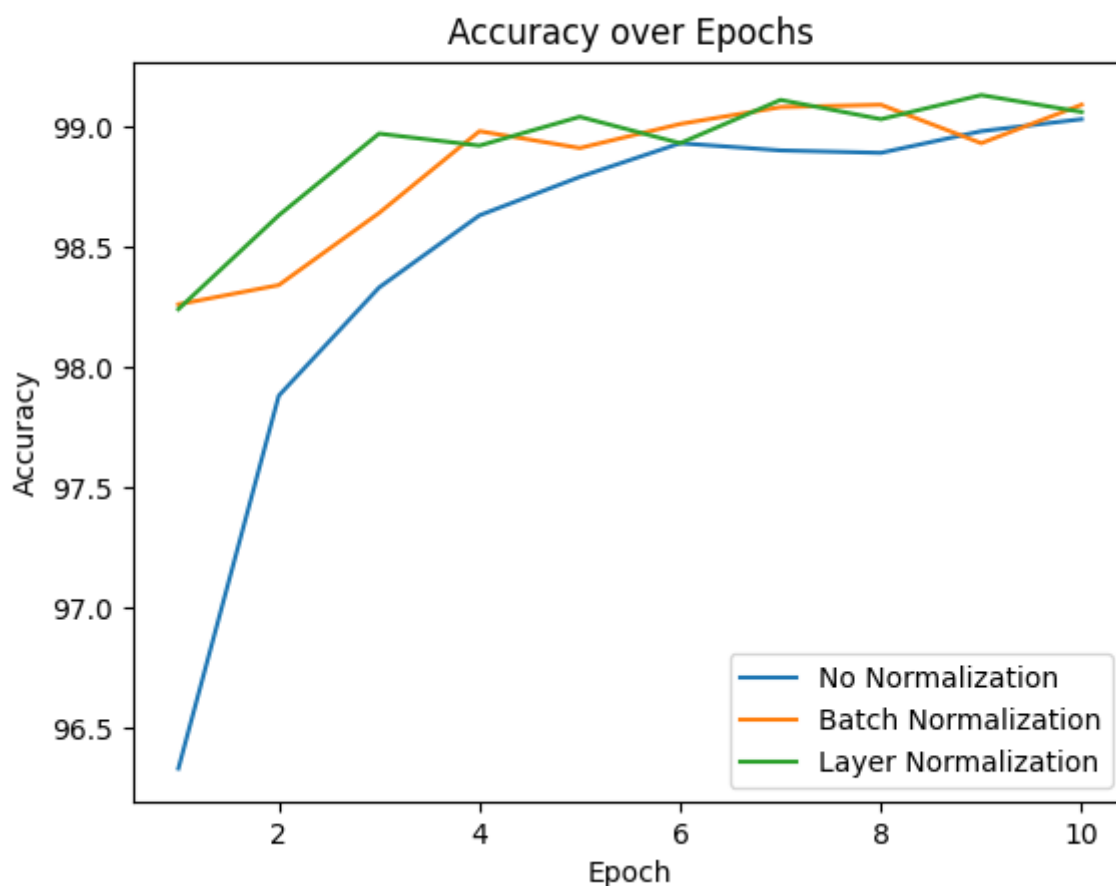
使用搜索得到的最优参数下，使用不含有BatchNorm的模型，训练30个epoch，训练过程如图：



可以看出在第20个epoch处，模型达到最佳的泛化性能。

六、验证Normalization的作用

控制其余变量均相同，搭建三个不同的网络，一个不使用Normalization，一个使用BatchNorm，一个使用LayerNorm，分别进行10个epoch训练，得到的在验证集上的准确率如下图所示。



- 可见使用Normalization可以使模型更快地达到最佳的泛化性能。
实验并没有看出两种规范化方法在效果上有明显的区别。

七、模型集成

为了进一步提高模型的准确率和鲁棒性，实验中尝试了**模型集成**的方法。dropout在训练过程中不仅可以防止过拟合，还可以被视作模型集成的一种形式。

此外，通过训练多个网络并在预测时取它们输出概率的平均或者投票，可以显著提升模型性能。

八、总结

本实验通过实现和训练卷积神经网络，深入理解了CNN的结构和功能。超参数搜索证明了选择合适的学习率、批量大小、动量和权重衰减对提高模型性能至关重要。dropout和模型集成等策略对于增强模型的泛化能力和准确率也起到了积极作用。