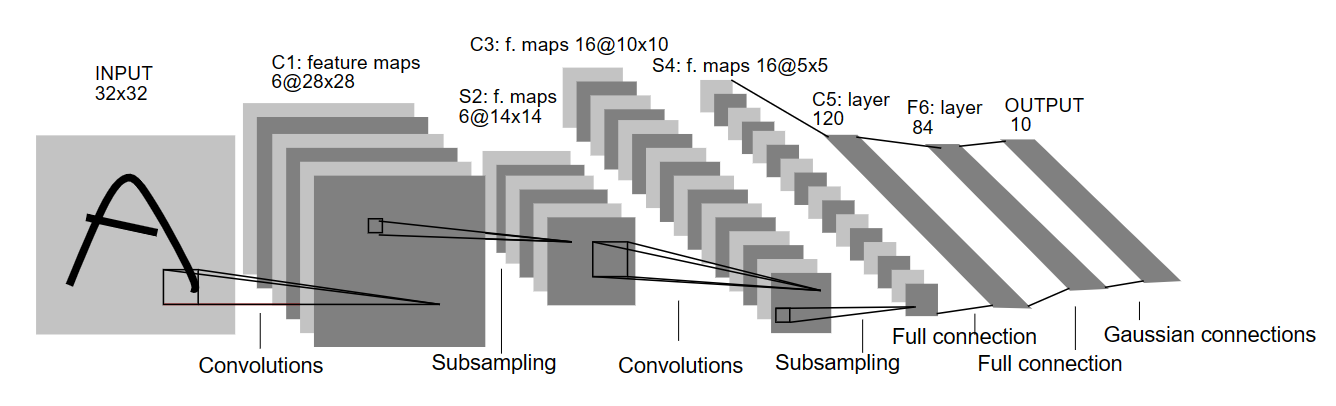
## 手写体数字识别实验

1. 实验原理：

基于神经网络的分类算法（LeNet-5）

LeNet-5模型是Yann LeCun教授于1998年在论文Gradient-based learning applied to document recognition中提出的，它是第一个成功应用于数字识别问题的卷积神经网络。在MNIST数据集上，LeNet-5模型可以达到大约99.2%的正确率。

LeNet-5模型总共有7层，下图展示了LeNet-5模型的架构。



输入为32\*32的手写图片，总共分为0-9十种类别，而输出应为0-9中的一个数字。由此可知，这是一个十分类问题。

二．模型介绍：

各层的具体结构：

C1层：卷积层，采用6个5\*5卷积核，输入为32\*32的手写图片，根据二维卷积原理，32-5+1=28，其输出为6\*28\*28的特征图。

S2层：下采样层（池化层），池化大小选取为2\*2，也就是将6个28\*28特征图分为14\*14个2\*2的小块，分别取每个小块中的最大值作为下采样得到的新像素值，由此便得到6\*14\*14的输出图像。

C3层：卷积层，对于6张14\*14的特征图，依然使用5\*5的卷积核进行卷积运算；对得到的6张10\*10的卷积结果求和，设为P，再由Out=f(P+b)（f为激活函数，b为偏置项），得到最终的16张10\*10的图片。

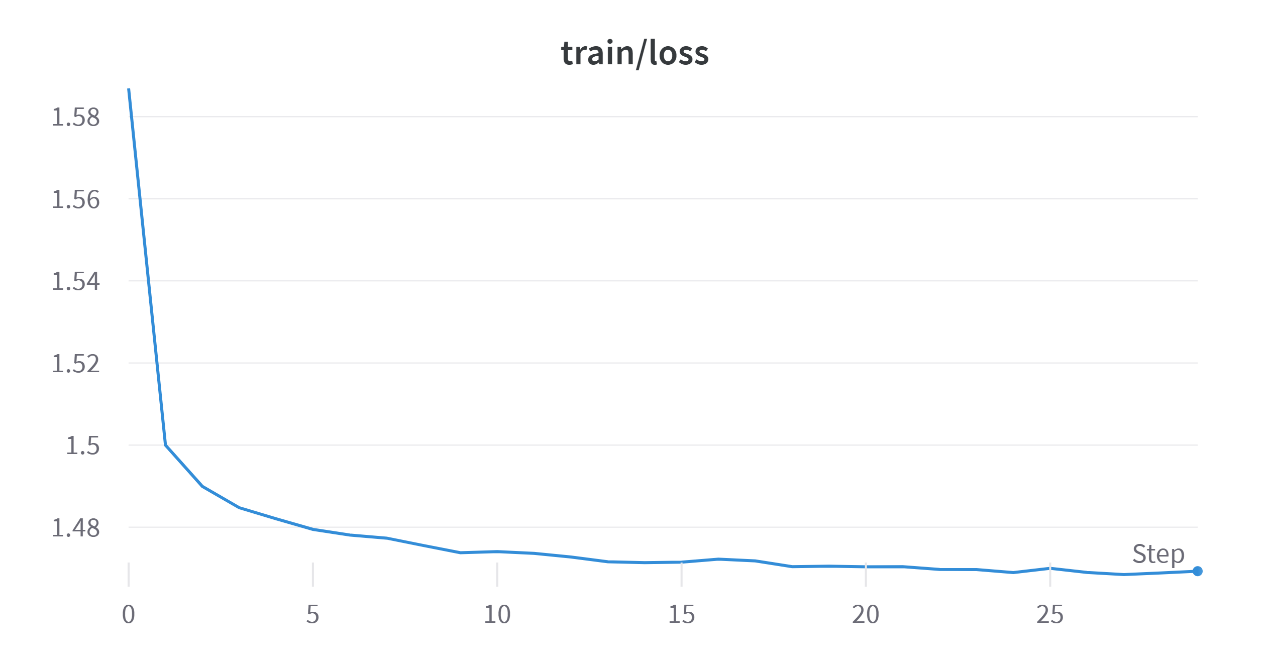
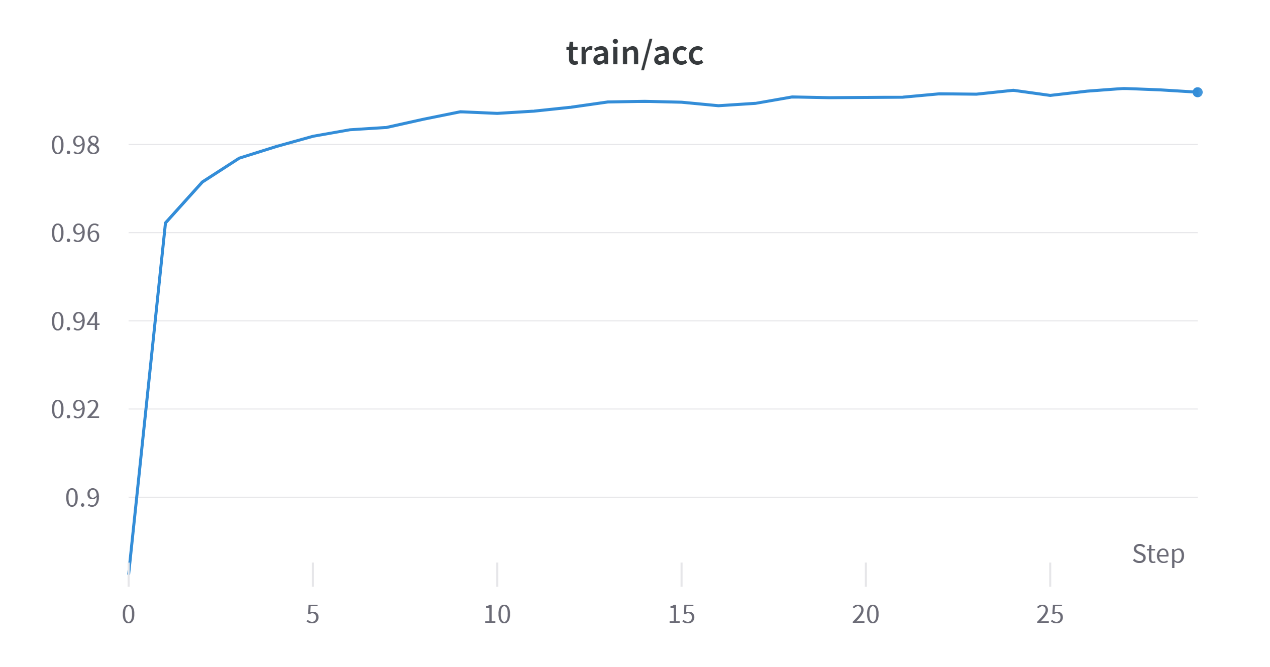
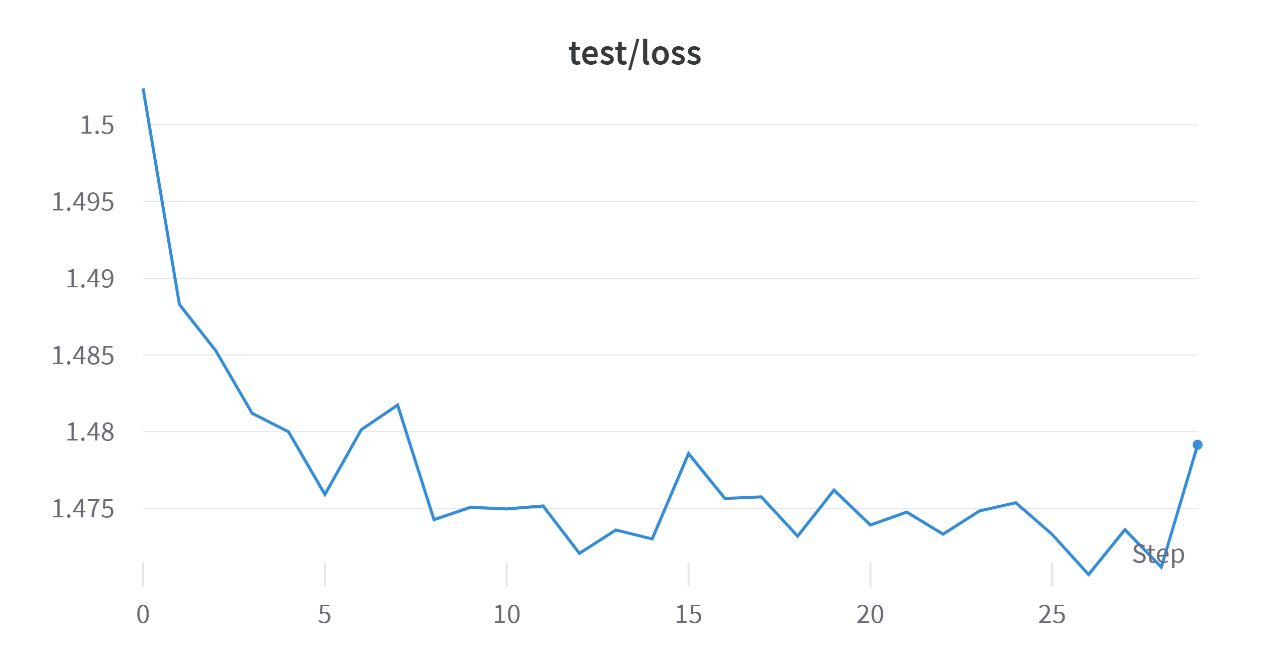
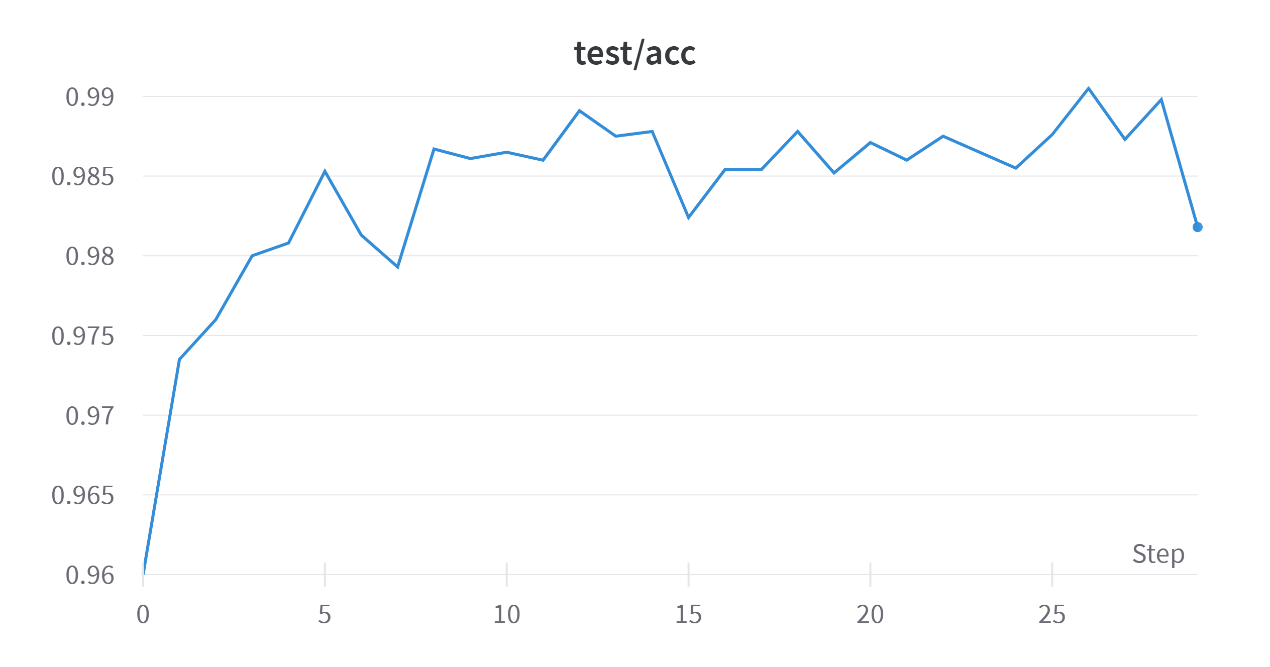
S4层：池化层，池化大小为2\*2。

C5层：卷积层，依然用5\*5的卷积核，得到5-5+1=1，也就是120个单独的像素值。因为输入图片与卷积核大小相同，可以视作全连接层。

F6层：全连接层，输入节点120个，输出节点84个。

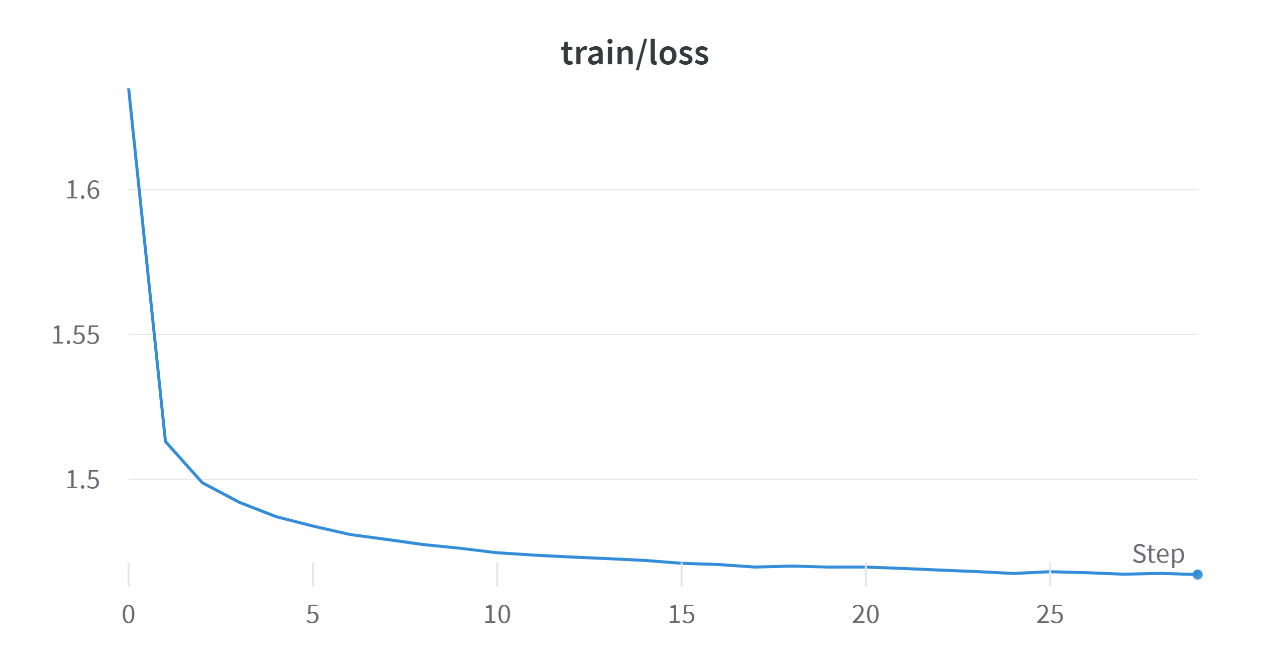
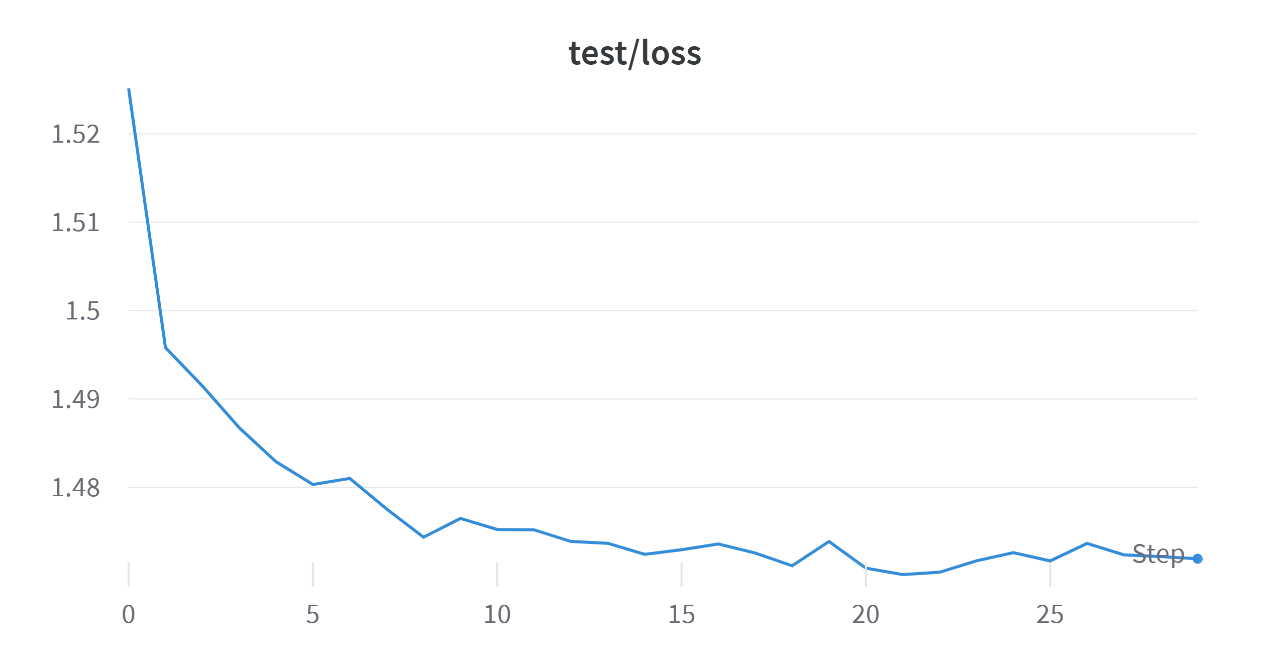
OutPut层：使用softmax函数，输入节点84个，输出节点10个，分别对应0-9类。

1. 实验效果及分析：
2. 学习率设定为0.001，使用adam优化器时的训练集、测试集loss、准确率曲线：

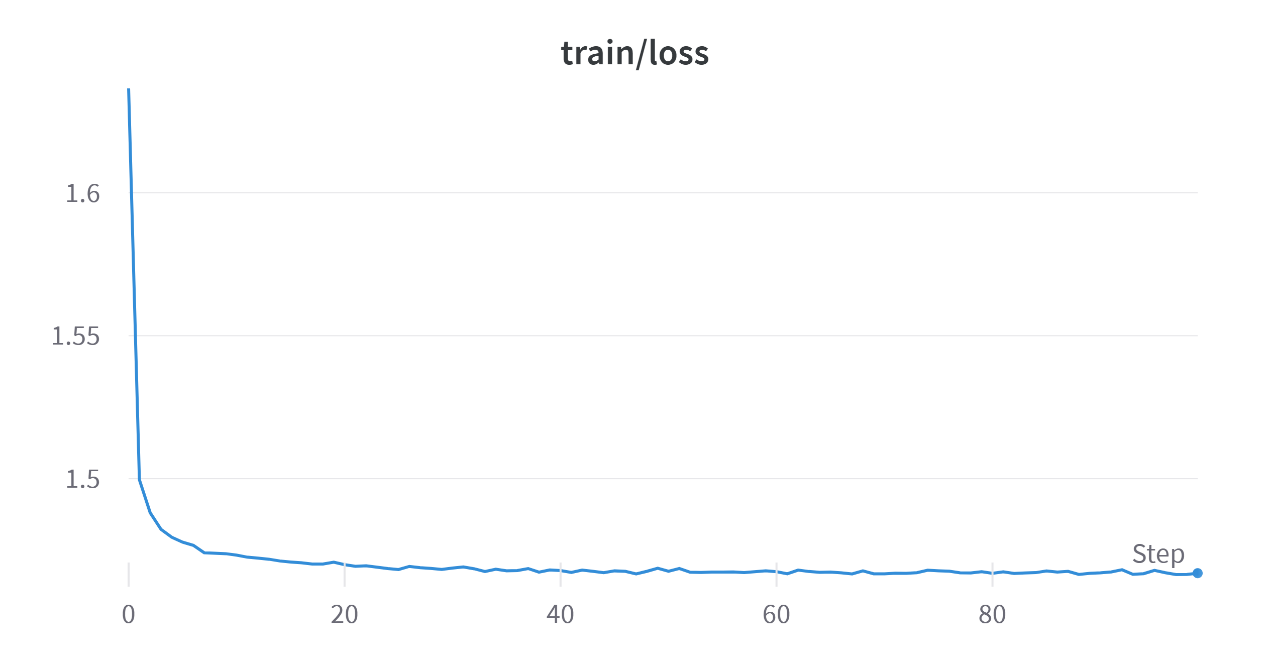
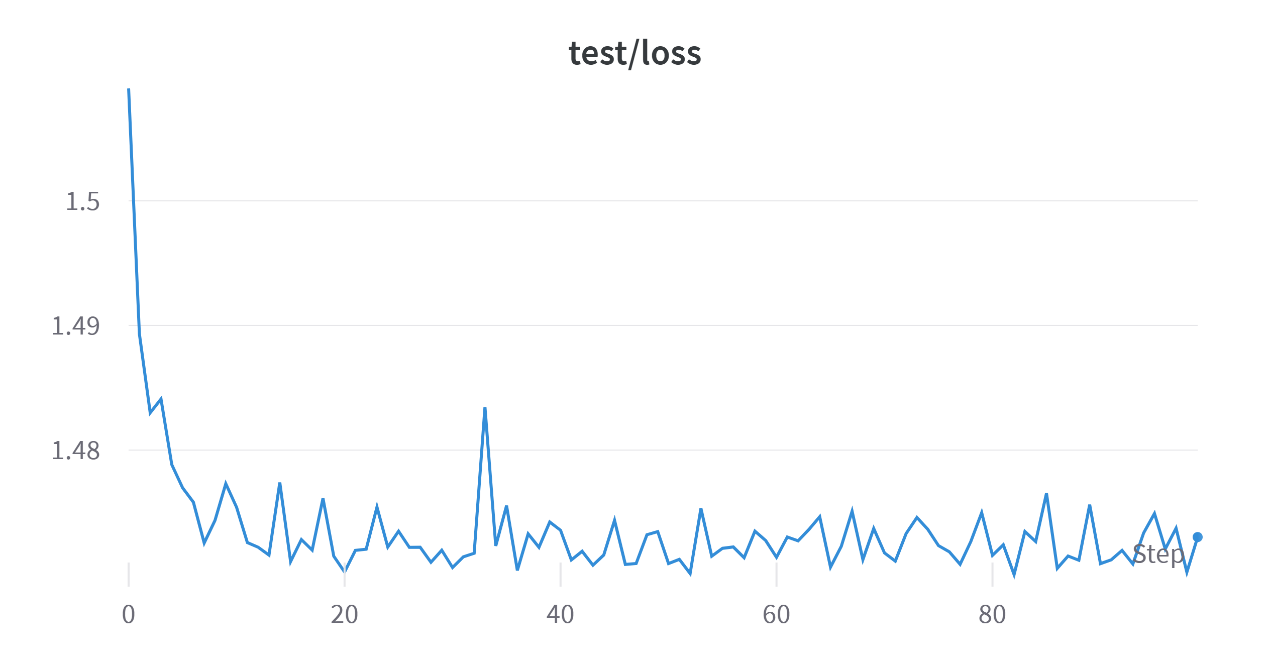


1. 调整学习率对loss曲线的影响：（优化器为adam）：

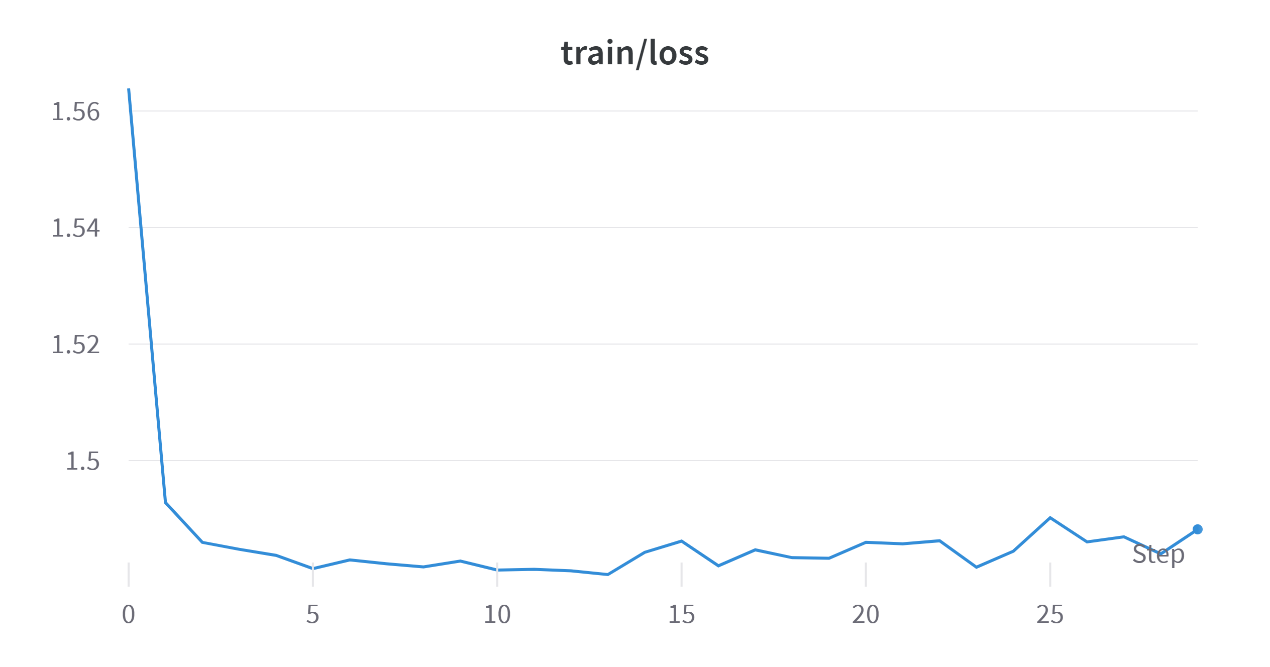
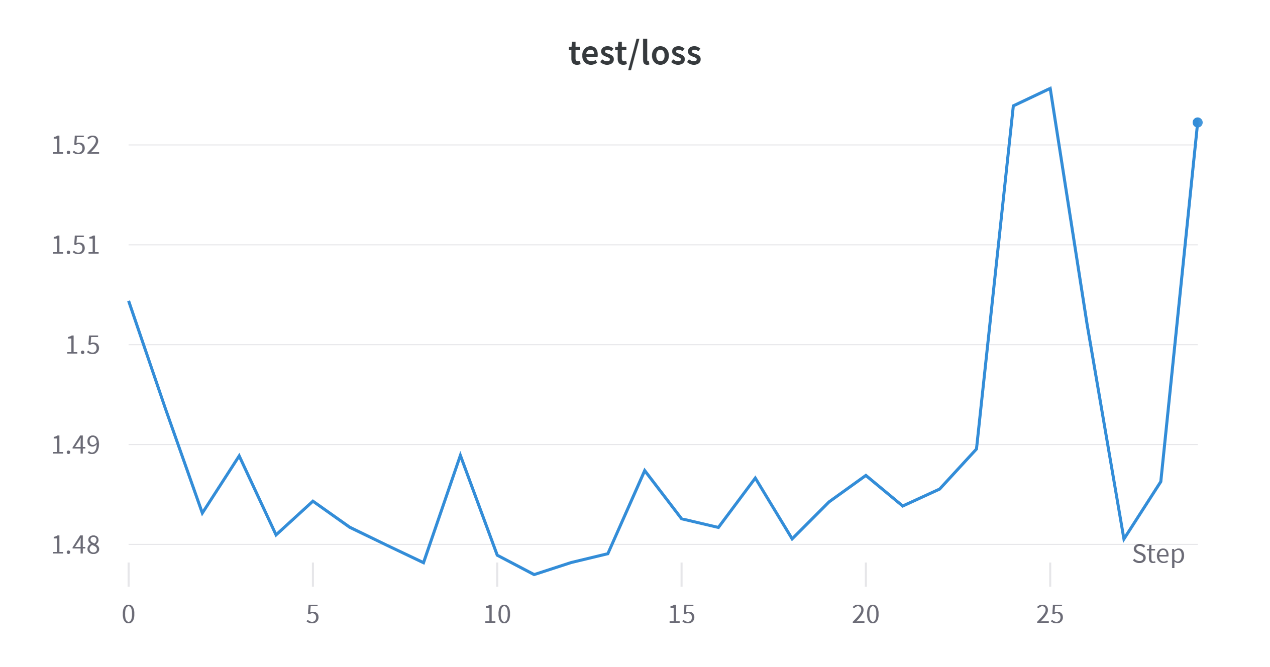
Lr=0.0005：



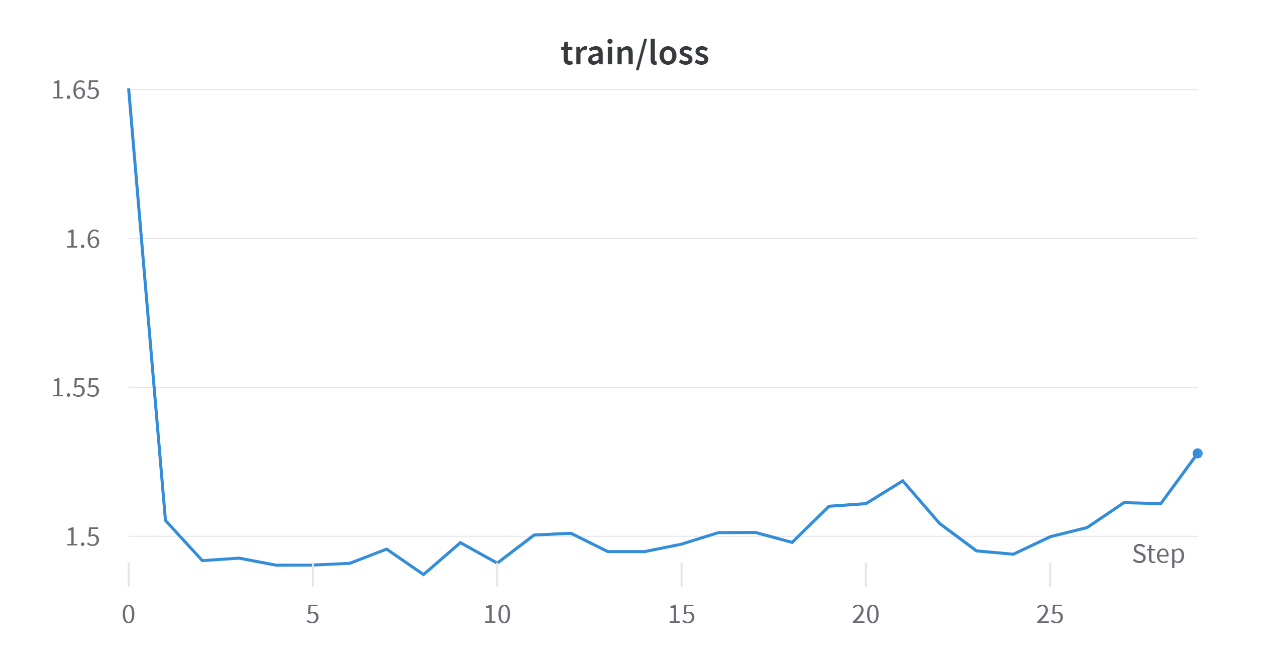
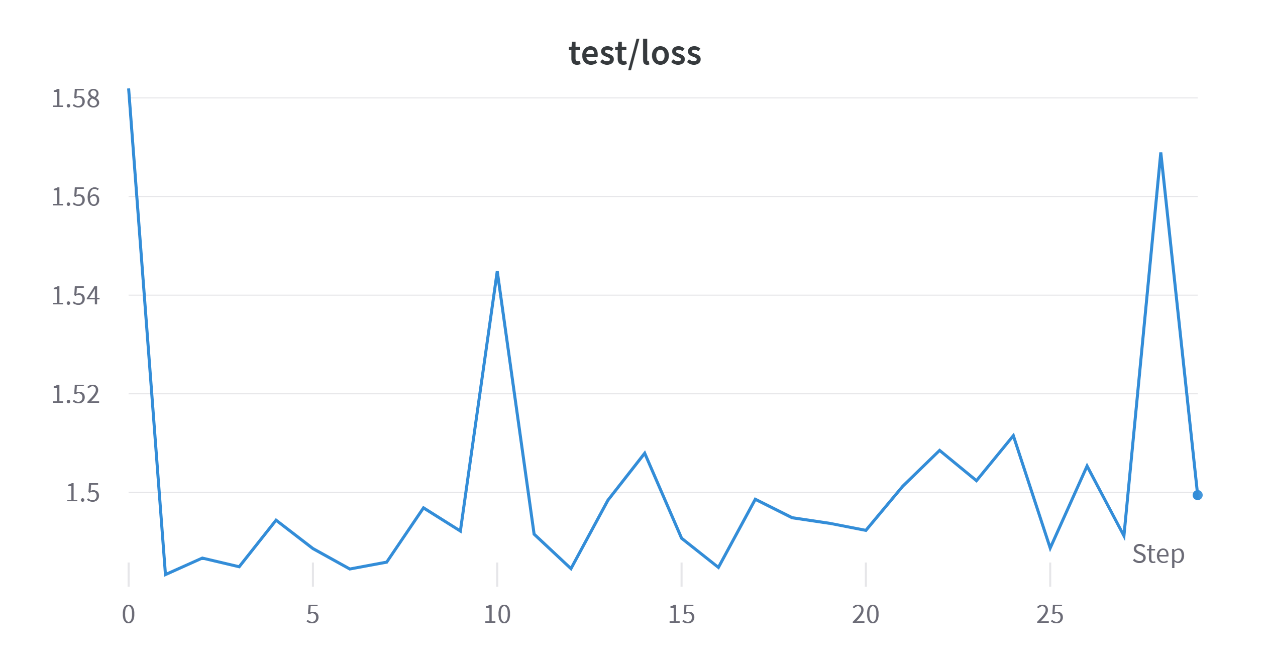
Lr=0.001：



Lr=0.002：



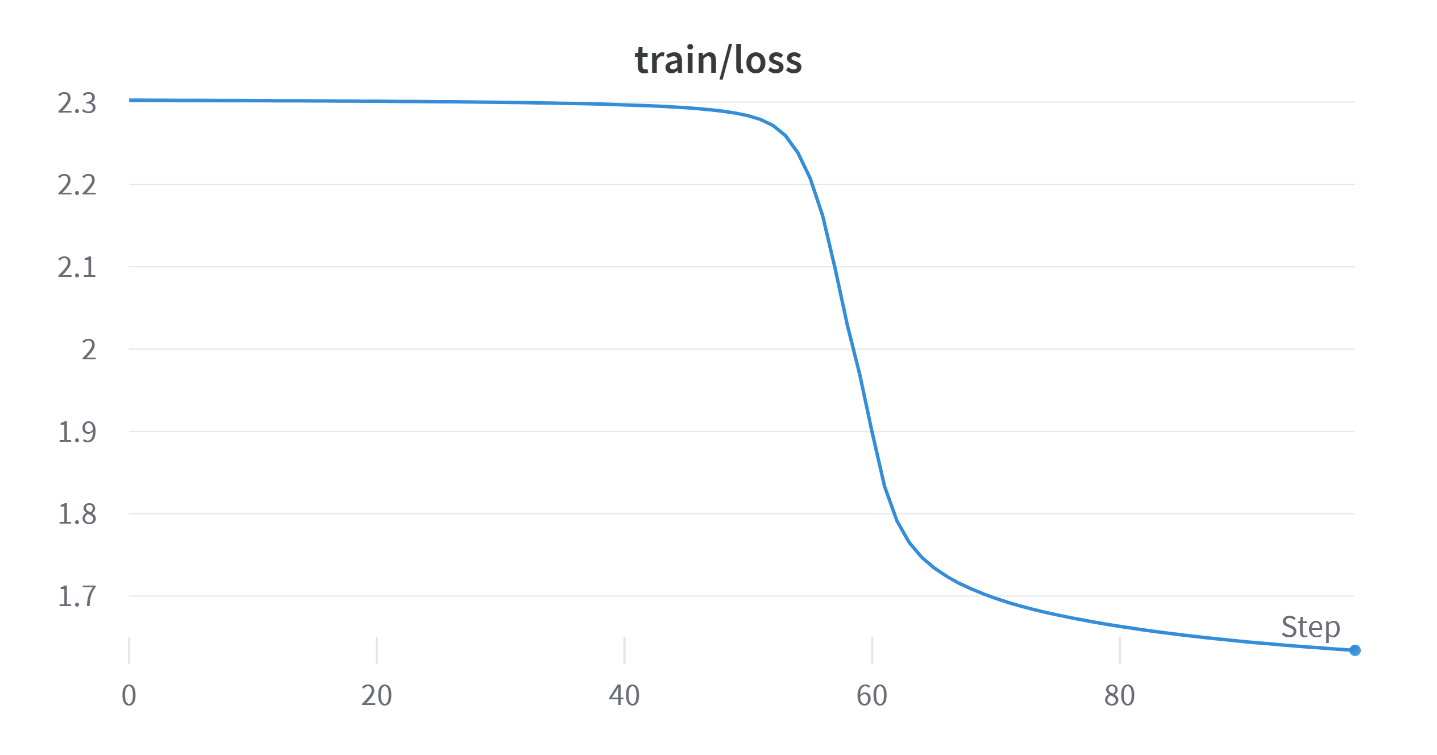
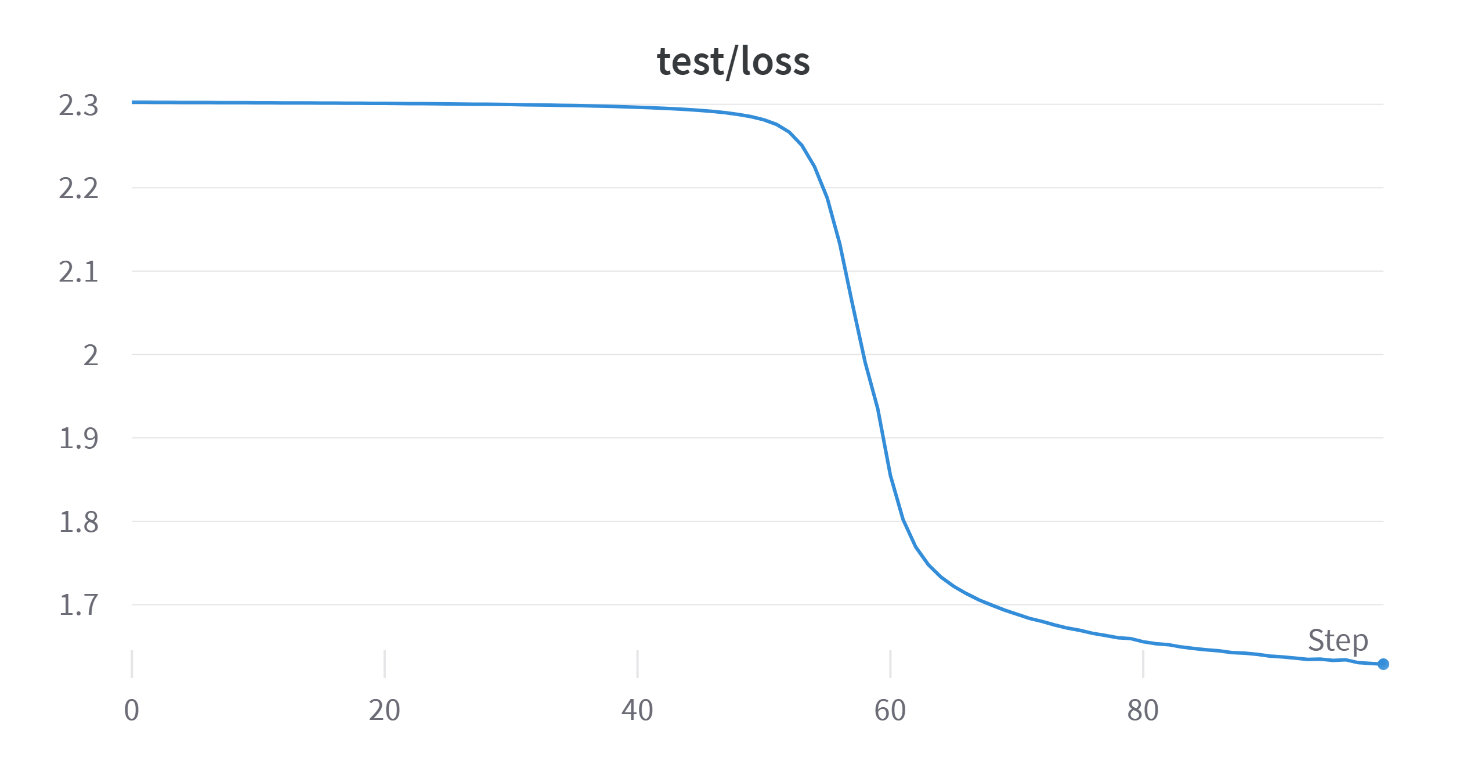
Lr=0.003：



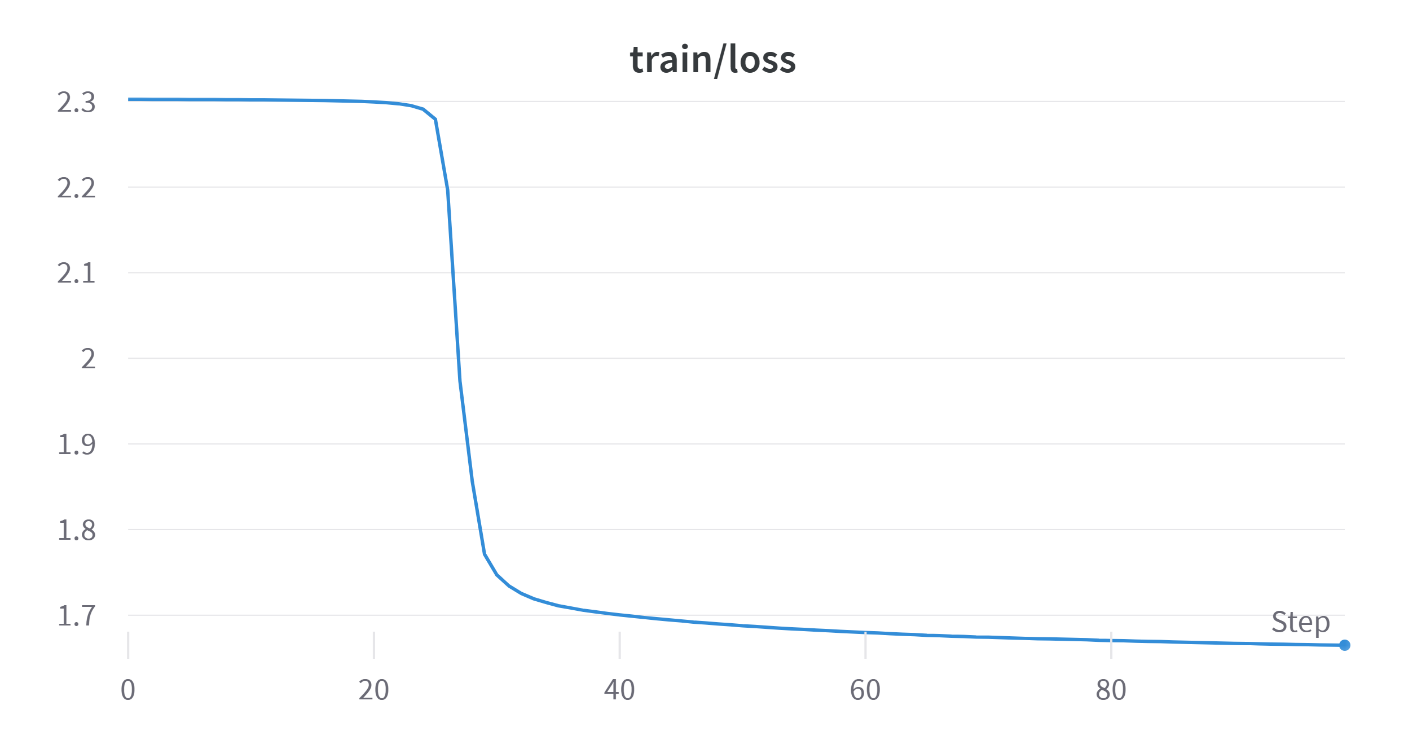
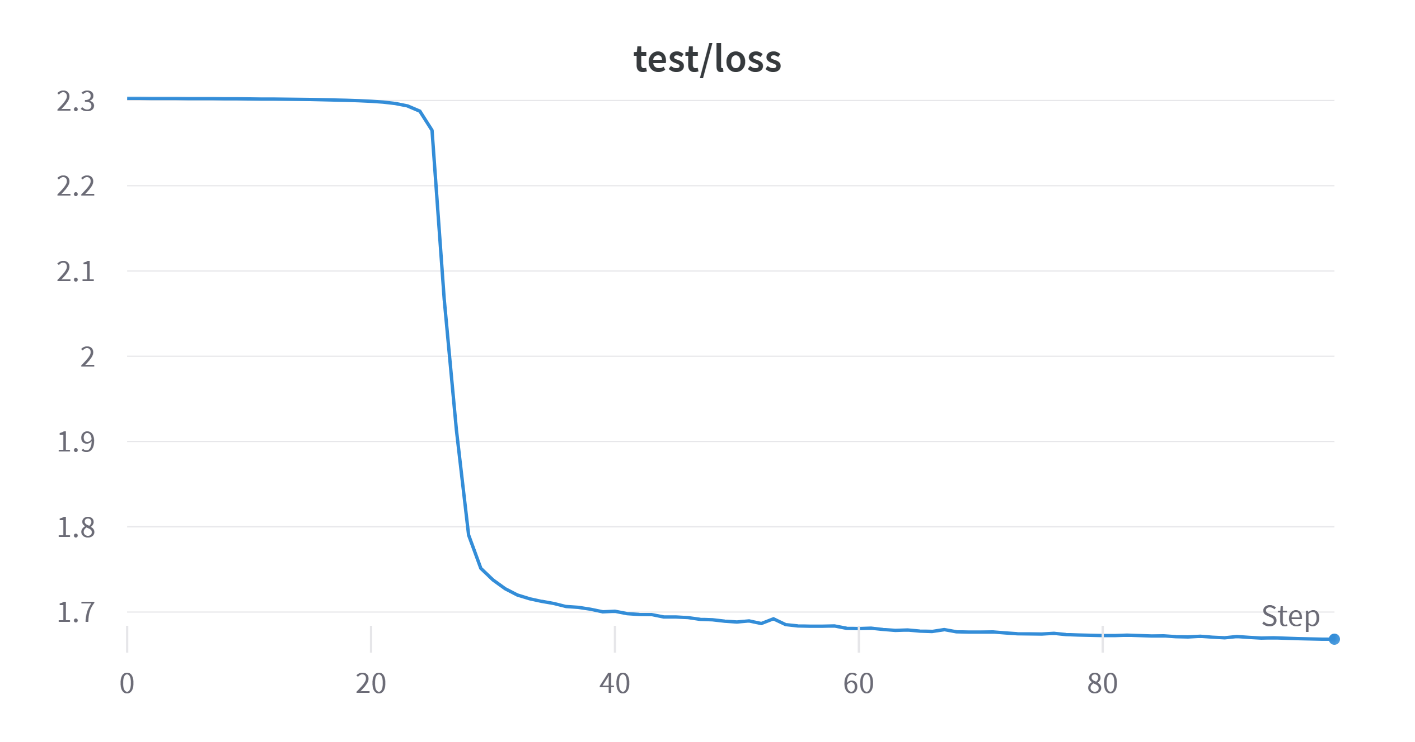
综上，对比可以发现，学习率取适中的值如0.001时，训练的loss值下降速度最快，模型性能最佳。

3.使用不同优化器（sgd、momentum、RMSprop、adam）对loss曲线的影响：

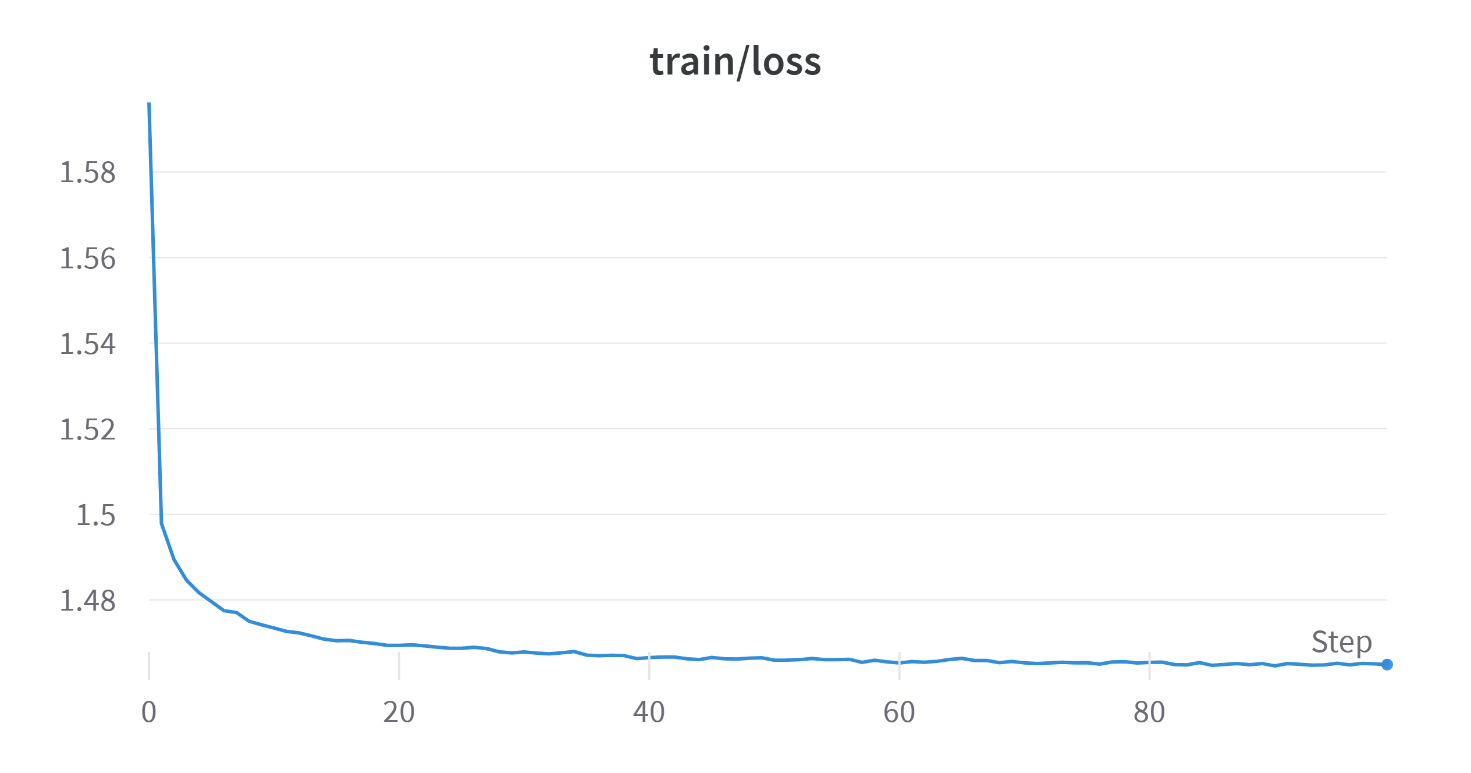
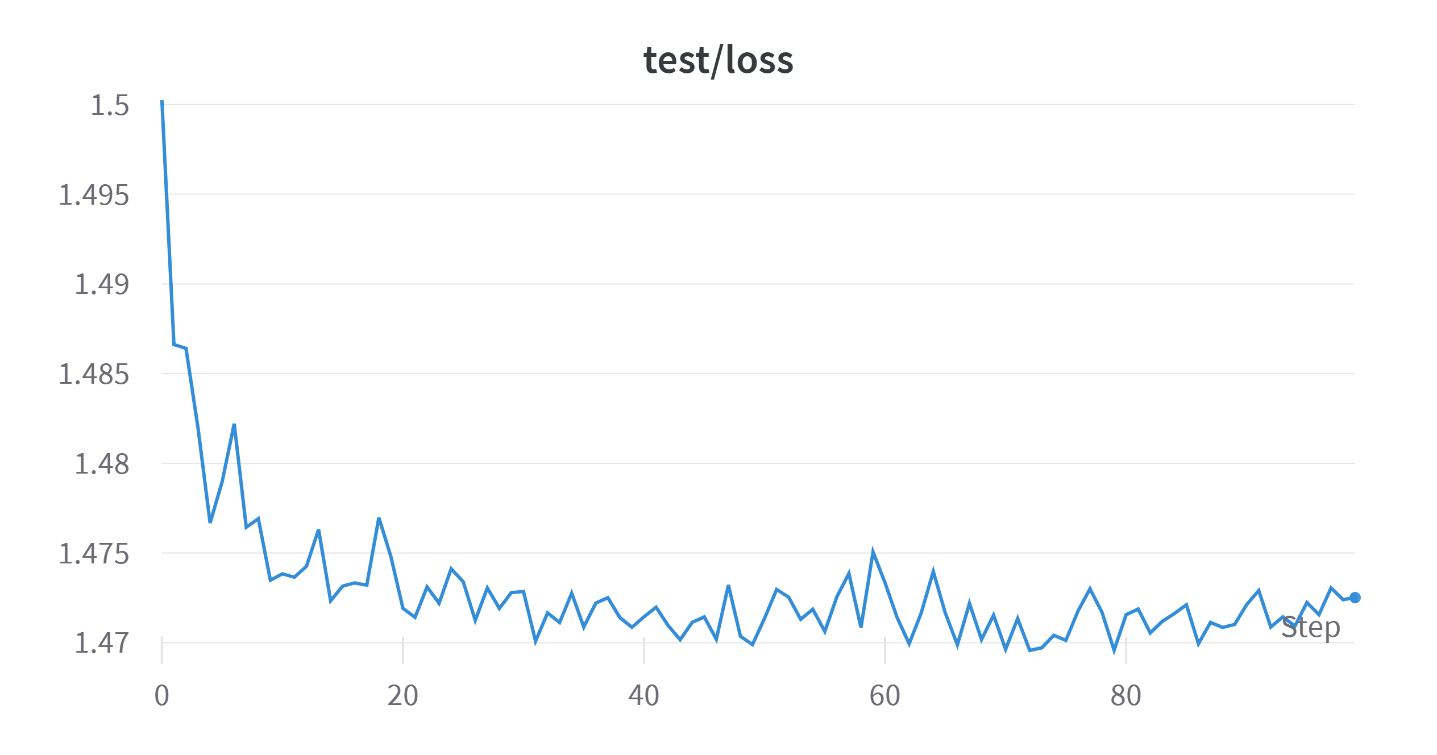
Sgd：



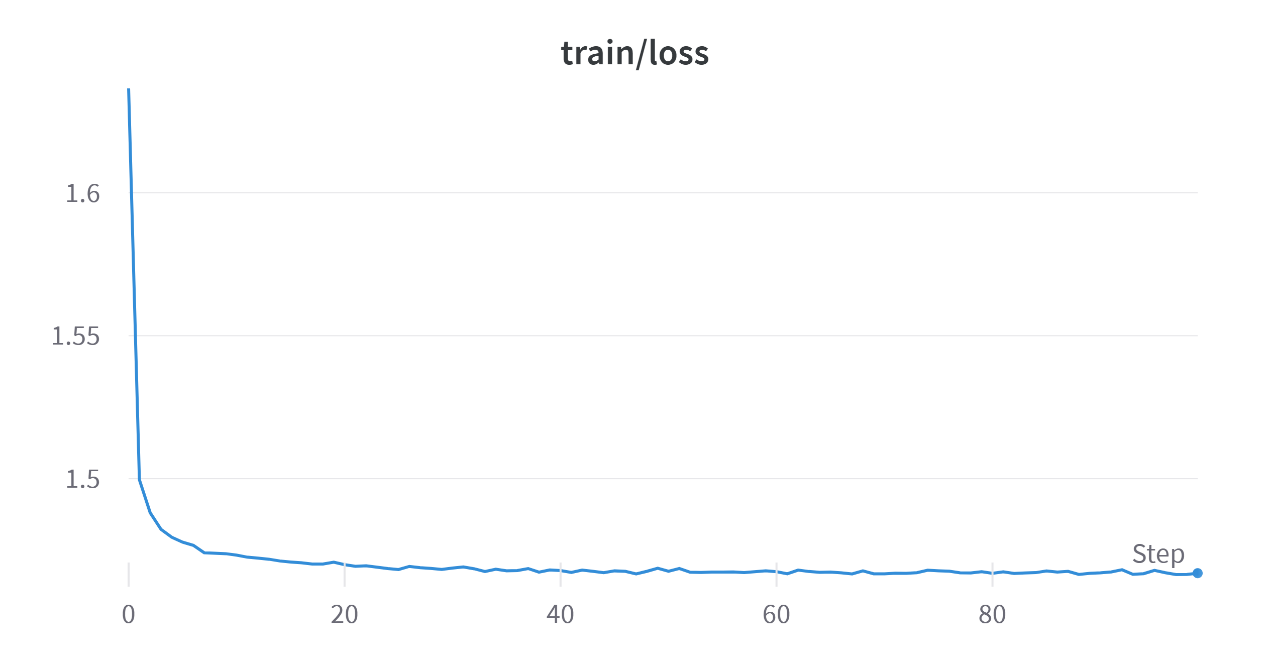
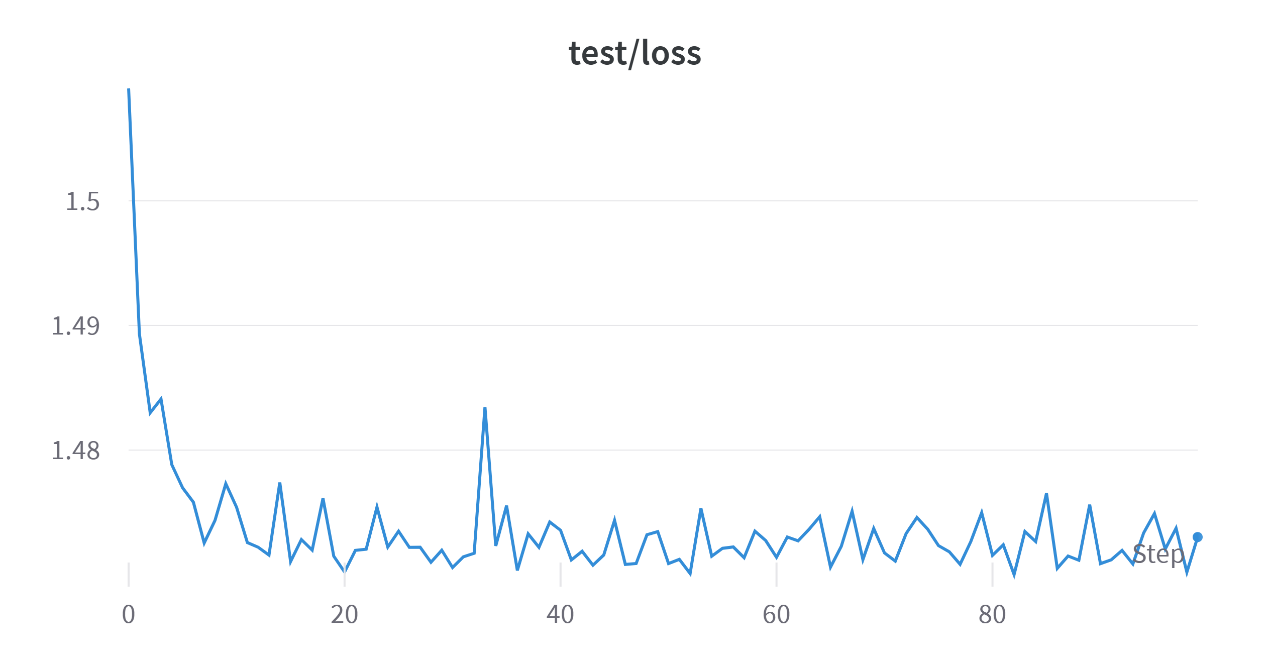
Momentum：



RMSprop：



Adam：



根据几种优化器的原理推断其性能：sgd（随机梯度下降法）：

梯度下降法利用反向传播求损失函数对网络参数的偏导数。然后令参数向偏导数相反的方向更新（参数更新的公式）。随机梯度下降每次仅从数据集中选取一个样本进行迭代更新，可以避免批量过大对显存造成负担。

后面三种为sgd的改进：

momentum通过模拟物理里动量的概念，积累之前的动量来替代真正的梯度，从而实现梯度下降的加速。对比loss曲线，可以看到采用momentum优化器后loss值的下降显著加快；

RMSprop根据自变量在每个维度的梯度值的大小来调整各个维度上的学习率，从而避免统⼀的学习率难以适应所有维度的问题。而且通过对平方项的指数加权移动平均，实现了自变量每个元素的学习率在迭代过程中不会一直降低（或不变），确保可以找到最优解。

Adam在RMSProp基础上对小批量随机梯度也做了指数加权移动平均，进一步修正了偏差。

对比可以发现，RMSprop和Adam的loss曲线下降速度远大于sgd与momentum，且Adam性能进一步优于RMSprop。这与上述原理相符。