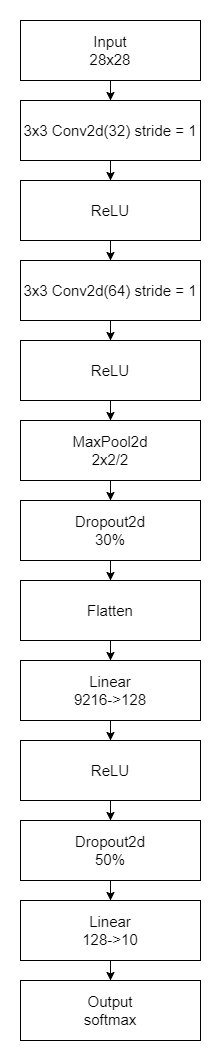
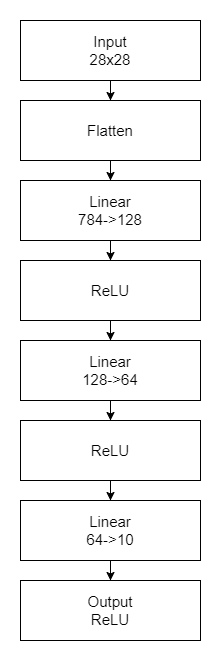
（1）知识蒸馏

我们设计了一个包含两个卷积层、两个dropout层和两个全连接层的卷积神经网络作为教师模型。一个仅由全连接层构成的，不含有任何卷积层的神经网络作为学生模型。采用FasionMINST数据集分别对教师模型和学生模型进行了训练。接着我们用新的损失函数替换了原来学生网络中的交叉熵损失函数，新的损失函数由两部分构成：

1.蒸馏损失：使用 KL 散度（Kullback-Leibler divergence）来衡量学生模型与教师模型输出概率分布的差异，通过调节温度和权重系数来控制其大小；

2.交叉熵损失：用于训练学生模型的常规损失。

通过蒸馏损失和交叉熵损失的加权组合，让学生模型在尽量保持预测准确性的前提下，尽量接近教师模型的输出分布，以获得更好的泛化性能。最终我们得到的实验结果如下：



（2）模型剪枝

在本实验中，我们采用了知识蒸馏中和教师网络相同的卷积神经网络作为原始模型，同样采用了FasionMINST数据集进行训练。接着采用了L1Norm剪枝策略，分别设置了稀疏度为0.6,0.8，0.9对模型进行了剪枝。实验结果如下：