总结：

1. 针对AlexNet这种网络结构相对简单的网络，在epoch相同的情形下，SGD和其他算法的时间没有太多差别，但是SGD和NAG算法比其他算法收敛的要快，精度也要高。建议选择SGD算法。

2. 对于GoogleNet等网络结构较为复杂的网络，SGD也可以收敛到很高的精度，adam算法的收敛速度要快一点而且收敛更稳定；建议选择adam算法

3.建议参数设置（batchsize可以根据情况选择28，64和128）

SGD： 学习率, 动量，学习规则设为step就好;

NAG: 学习率，动量，学习规则设为step就好;

AdaGrad: 学习率=0.001， 学习规则为INV 参数设置power 0.75 gama0.5；

AdaDelta：学习率=1 ， 学习规则fix

RMSprop：学习率=0.001， 学习规则 INV or poly

adam： 学习率=0.001， 学习规则 INV or poly, sgmoid

4.在GoogleNet上当batchsize由28变为56时，收敛的速度要比28时快一点，但是精度没有明显提高。

5在GoogleNet上对adam算法尝试不同的学习规则（INV，poly，exp，sigmoid）,实验结果没有明显差别。原因有可能是训练集较小。

6.在GoogleNet上尝试不同的优化方法和策略时，一般都可以收敛到较高的精度，影响算法

收敛和精度的主要因素还是学习率。

7.在训练网络时，建议先选择参数默认值，和step学习规则进行尝试，然后根据训练结果调整参数。

8.当学习率较大时，网络会震荡收敛，容易不稳定。

9.目前深度网络优化算法使用率较高的为adam算法