（2）第二个阶段开始的标志就是relu, dropout等小技巧，第二波深度学习算法已经基本抛弃了预训练的做法。引入RELU代替sigmoid激活函数。

（3）深度残差学习（deepresidual learning）进一步避免了梯度消失（其实是把高阶特征和低阶做融合）。

（4）归一初始化，各层输入归一化，使得可以收敛的网络的深度提升为原来的十倍。