1. 神经网络因为参数和计算量很多所以需要进行压缩，有人发现只需5%的参数子集就可以完整重构出剩余参数，但是这种冗余在训练阶段是必须的。
2. 最简单的量化算法：标量量化。将权重矩阵向量化，然后进行k均值聚类，将k个聚类中心存储在码本，原权重矩阵记录聚类中心在码本中的索引。
3. 知识蒸馏：监督的信息可以是概率分布、中间层的特征图等。
4. 随机抠取操作一般用较大（约0.8至0.9倍原图大小）的正方形在原图的随机位置处抠取图像块。在此使用正方形的原因是由于卷积神经网络模型的输入一般是方形图像，直接以方形抠取避免了矩形抠取后续的图像拉伸操作带来分辨率失真。尺度变换旋转是为了增加卷积神经网络对物体尺度和方向上的鲁棒性。尺度变换一般是将图像介辨率变为原图的0.8，0.9、1.1，1.2、1.3等倍数。色彩抖动：在RGB颜色空间对原有RGB色彩分布进行轻微扰动。在实践中，往往会将上述种方式叠加使用，如此便可将图像数据扩充至原有数量的数倍甚至数十倍。
5. 网络参数初始化：除以根号n为了维持输入和输出数据分布的方差一致性。将预训练模型上的参数作为新任务上模型的参数初始化。
6. 目标函数：支持向量机有个合页损失函数，它的设计理念是“对错误越大的样本施加越严重的惩罚”，这样造成损失函数对噪声的抵抗能力较差。若某样本标记本身错误或该样本本身是离群点，则由于错分导致该样本分类误差会变得很大，如此便会影响整个分类超平面的学习，从而降低模型泛化能力。非凸损失函数的引入则很好地解决了这个问题。大间隔交叉熵损失函数可以扩大类间距离，中心损失函数减小同类差异。两个可以与交叉熵损失函数搭配使用。
7. 网络正则化：加入l1l2正则化项。随机失活相当于训练多个子网络然后集成，从而提升网络泛化性能。
8. 批规范化操作：两个调节因子，对规范到均值为 0 ，方差为1 的值进行微调。因为经过第一步操作后有可能降低神经网络的非线性表达能力，所以会以此方式来补偿 Normalization 操作后神经网络的表达能力。
9. 样本不平衡：需指出的是下采样并不是直接随机丢弃一部分图像，而是在批处理训练时对每批随机抽取的图像严格控制其样本较多类别的图像数量。