

# 网络压缩论文整理(network compression)

---

## 1.参数修剪和共享

### 1.1量化和二值化

- [使用矢量量化压缩深度卷积网络](#)
- [用于移动设备的量化卷积神经网络](#)
- [提高神经网络在cpus上的速度](#)
- [深度学习的数值精度有限](#)
- [深度压缩：通过修剪，训练量化和霍夫曼编码来压缩深度神经网络](#)
- [接近网络量化的极限](#)
- [Binaryconnect：在传播期间用二进制权重训练深度神经网络](#)
- [二进制网络：训练深度神经网络，权重和激活约束为+1或-1](#)
- [Xnor-net：使用二元卷积神经网络的Imagenet分类](#)
- [深度神经网络对加权二值化和其他非线性失真具有鲁棒性](#)
- [深度网络的丢失感知二值化](#)
- [乘法少的神经网络](#)

### 1.2修剪和共享

- [将最小网络结构的偏差与反向传播进行比较](#)
- [神经信息处理系统的进展2](#)
- [网络修剪的二阶导数：最佳脑外科医生](#)
- [深度神经网络的无数据参数修剪](#)
- [学习有效神经网络的权重和连接](#)
- [使用散列技巧来压缩神经网络](#)
- [用于神经网络压缩的软重量分享](#)
- [使用团体脑损伤快速衔接](#)
- [少即是多：迈向紧凑型cnns](#)
- [在深度神经网络中学习结构化的稀疏性](#)
- [修剪筛选以获得有效的回合](#)

### 1.3设计结构矩阵

- [循环投影的深度网络参数冗余探讨](#)
- [具有循环投影的快速神经网络](#)
- [油炸的小圆点](#)

## 2.低等级分解和稀疏性

- [学习可分离的滤波器](#)
- [利用卷积网络中的线性结构进行有效评估](#)
- [加速低阶扩展的卷积神经网络](#)
- [利用微调cp分解加速卷积神经网络](#)
- [具有低秩正则化的卷积神经网络](#)
- [预测深度学习中的参数](#)
- [具有高维输出目标的深度神经网络训练的低秩矩阵分解](#)

## 3.传递/紧凑卷积滤波器

- [集团等变卷积网络](#)
- [双卷积神经网络](#)

- [通过级联的整型线性单元理解和改进卷积神经网络](#)
- [深度神经网络中的多偏置非线性激活](#)
- [利用卷积神经网络的循环对称性](#)
- [初始-v4：初始阶段和剩余连接对学习的影响](#)
- [Squeezedet：统一的小型低功耗完全卷积神经网络，用于自动驾驶的实时目标检测](#)
- [SQUEEZENET：具有50X较小参数和<0.5MB型号的ALEXNET-LEVEL精度](#)
- [移动网络：用于移动视觉应用的高效卷积神经网络](#)

#### 4.知识蒸馏

- [模型压缩](#)
- [深网真的需要深入吗？](#)
- [在神经网络中提炼知识](#)
- [Fitnet：提示薄的深网](#)
- [贝叶斯黑暗的知识](#)
- [通过从神经元中提取知识来面对模型压缩](#)
- [Net2net：通过知识转移加速学习](#)
- [关注注意力：通过注意力转移提高卷积神经网络的性能](#)

#### 5.其他

- [动态容量网络](#)
- [令人难以置信的大型神经网络：稀疏的门控混合专家层](#)
- [用于多模式手势分割和识别的深度动态神经网络](#)
- [具有随机深度的深度网络](#)
- [具有分离随机深度的深锥体残余网络](#)
- [通过FFT快速训练卷积网络](#)
- [卷积神经网络的快速算法](#)
- [S3pool：随机空间抽样合并](#)

#### 6.调查

- [深度神经网络模型压缩与加速度综述](#)