

关于 Intel caffe 的速度以及相关性测试报告：

测试背景介绍：

在工业界对于深度学习的应用中，网络前馈 SDK 是作为最基础的核心功能模块，承担着非常高比重的计算任务，而核心计算库的选择例如 openblas, mkl, altblas 等，决定了 SDK 计算能力的上下限。

Intel caffe 是 Intel 为 CNN 定制的 caffe，官方说明对于 Xeon 系列机器可以进行有效的优化，但是优化效果按照 cpu 的型号有较大差异。其中性能最好的是 “Intel Xeon Phi™ product family”，其次是 Xeon E5 及以上的系列。

从目前的简单特性层面上看来，intel 具有以下几个特点：

- 1.对于传统的卷积，pooling，bn 层等使用了 openmp 技术对于过程中的 for 循环进行加速（相关的优化技术在 xgboost 中有着广泛的应用，intel 的工作组也给出了较为详细的优化过程）
- 2.使用 MKLDNN 对于卷积等 9 个层进行了优化，封装成新的 engine 层，如 MKLDNNConvolutionLayer 等
- 3.使用 MKL2017 对于卷积等 9 个层进行了优化），封装成新的 engine 层，如 MKLConvolutionLayer 等
- 4.以及一些诸如 multinode 的分布式计算平台优化

实验内容简介：

本次测试主要从以下几点对于 intel/caffe 进行了测试和试验

- 1.BVLC/caffe 中使用 MKL 与 openblas 的加速提升效果
- 2.Intel/caffe 使用 MKL+openmp 和 BVLC/caffe 使用 MKL 的比较
- 3.Intel/caffe 中 MKL+openmp,MKL2017,MKLDNN 三种加速优化方式的比较
- 4.在 intel/caffe 中 resnet50 的各种操作所占比例
- 5.不同型号 cpu 之间的比较,E5-2670 v2 和 E5-2660v3 的比较，（二者主要差异在于后者频率略高，且支持 AVX2 等较新型的指令）

实验环境：

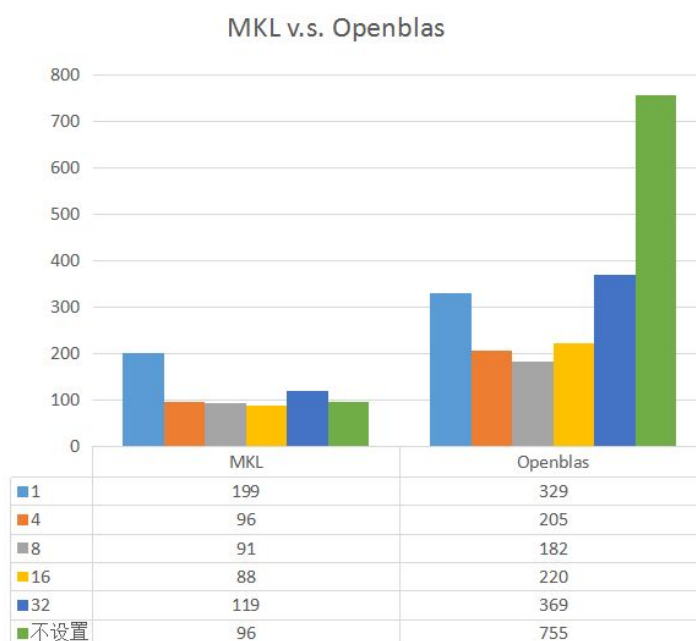
网络：resnet50 1000 类

数据：250*250 人脸图像的 Imdb

Cpu：主要使用 E5-2660v3，在 cpu 对比试验中会使用 E5-2670 v2

方式：可见 reference Intel caffe 时间测试方法以及 openmp 优化说明，其中 `OMP_NUM_THREADS` 是比较 tricky 的环境变量，为了验证他的问题和影响，针对 1,4,8,16,32,不设置 都进行了测试。

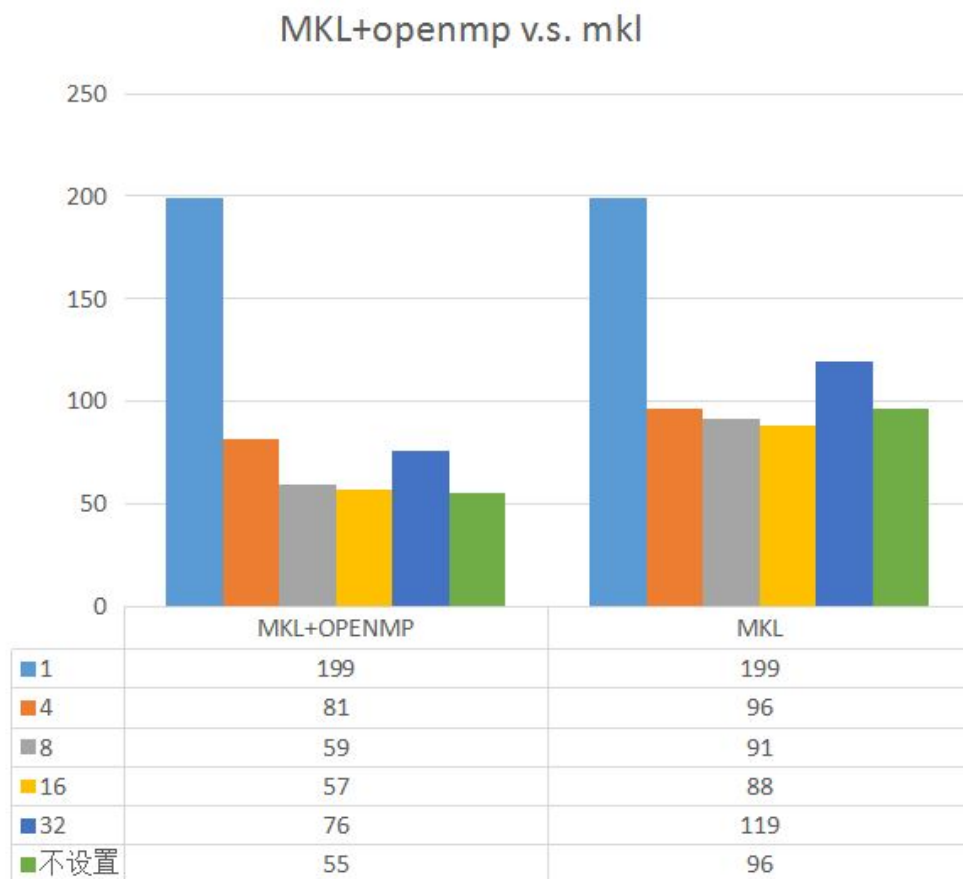
1.BVLC/caffe 中使用 MKL 与 openblas 的加速提升效果：



可以看出 bvlc/caffe 的加速库直接替换成 MKL，就可以有一定的提升

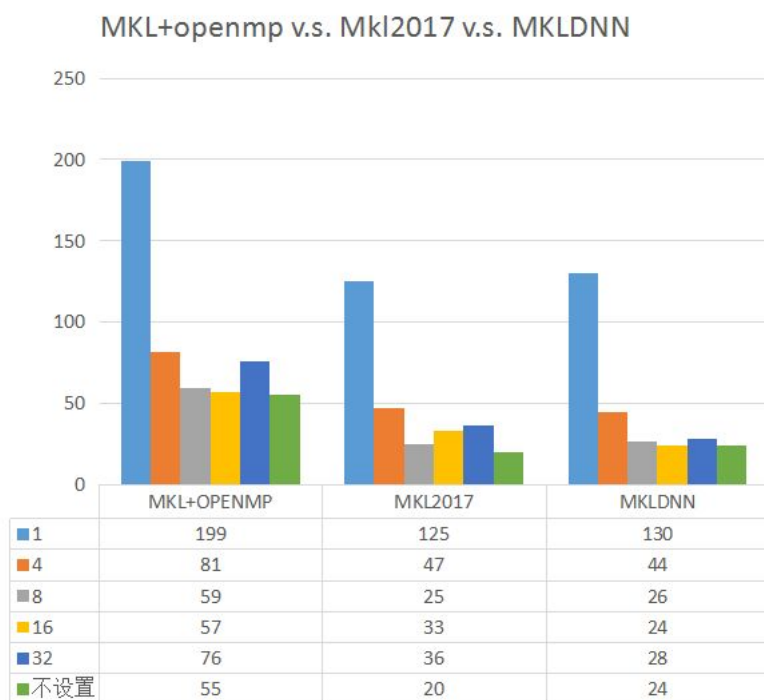
注：图片中所有的时间单位都为 ms，同一簇的不同颜色块表示不同的 openmp 设置

2. Intel/caffe 使用 MKL+openmp 和 BVLC/caffe 使用 MKL 的比较



Intel/caffe 中使用了 MKL+openmp 对于计算层进行了优化，可以看出随着环境中强行设置的 omp_num_threads 的变化，MKL+OPENMP 的性能慢慢从和 MKL 一致，到可以增速 2 倍

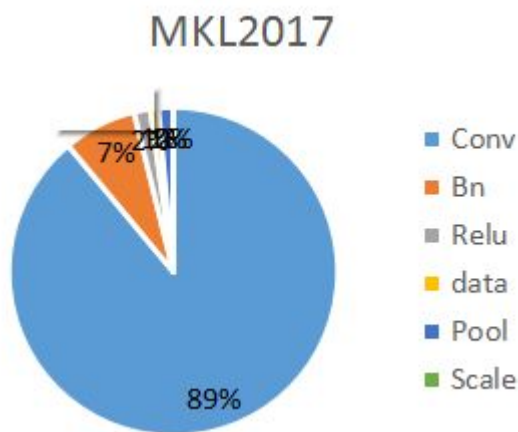
3. Intel/caffe 中 MKL+openmp, MKL2017, MKLDNN 三种加速优化方式的比较



Intel/caffe 内置的加速算法真的是名不虚传，无论是 MKL2017 和 MKLDNN 都在原来的基础上有了较大的提升，在不同 OMP_NUM_THREADS 的情况下都比原来要快 2 倍以上。

注：图片中所有的时间单位都为 ms，同一簇的不同颜色块表示不同的 openmp 设置

4.在 intel/caffe 中 resnet50 的各种操作所占比例

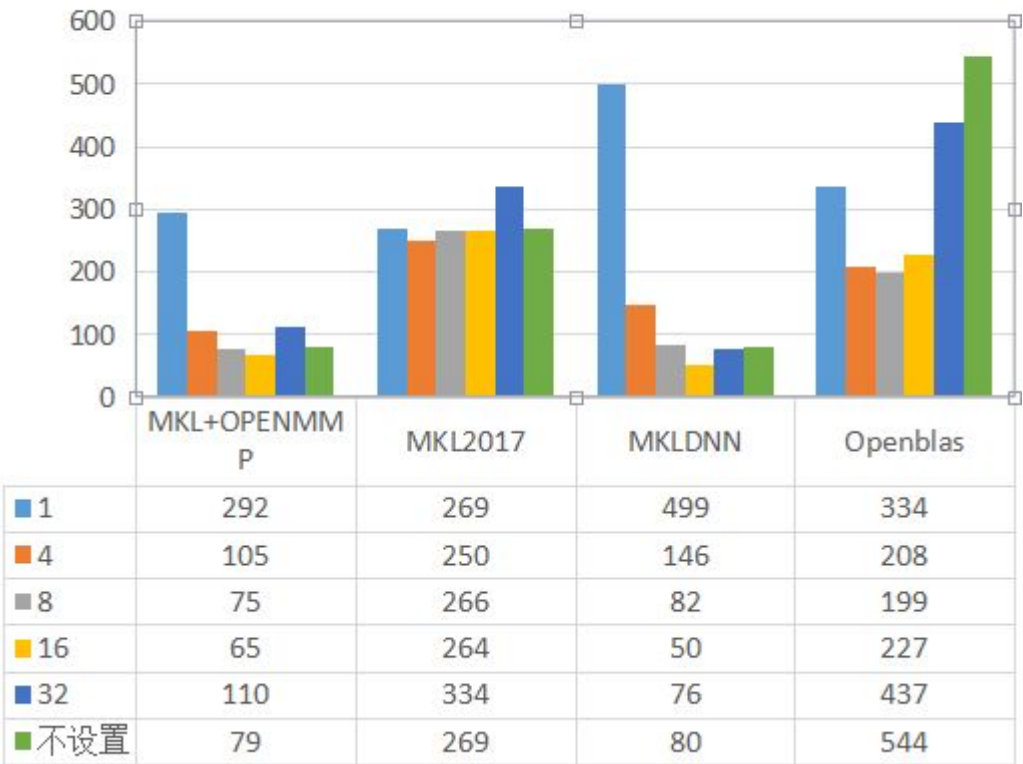


在对算法的每个层计算时间进行统计的情况下，可以看出卷积层的计算比重还是最重的，而且是非常严重，当然这和我们的 resnet50 是全卷积网络德 featuresmap 是有关系的，但值得一提的是 MKL2017 和 MKLDNN 都采用了将 scale 层直接加入 bn 层来计算的方式，可见是非常的机智了

layer	Conv	Bn	Relu	data	Pool	Scale
MKL+OPENMP	142	31	9	1	3	12
MKL2017	111	9	2	1	1.8	0
MKLDNN	114	10	3	1	1.3	0

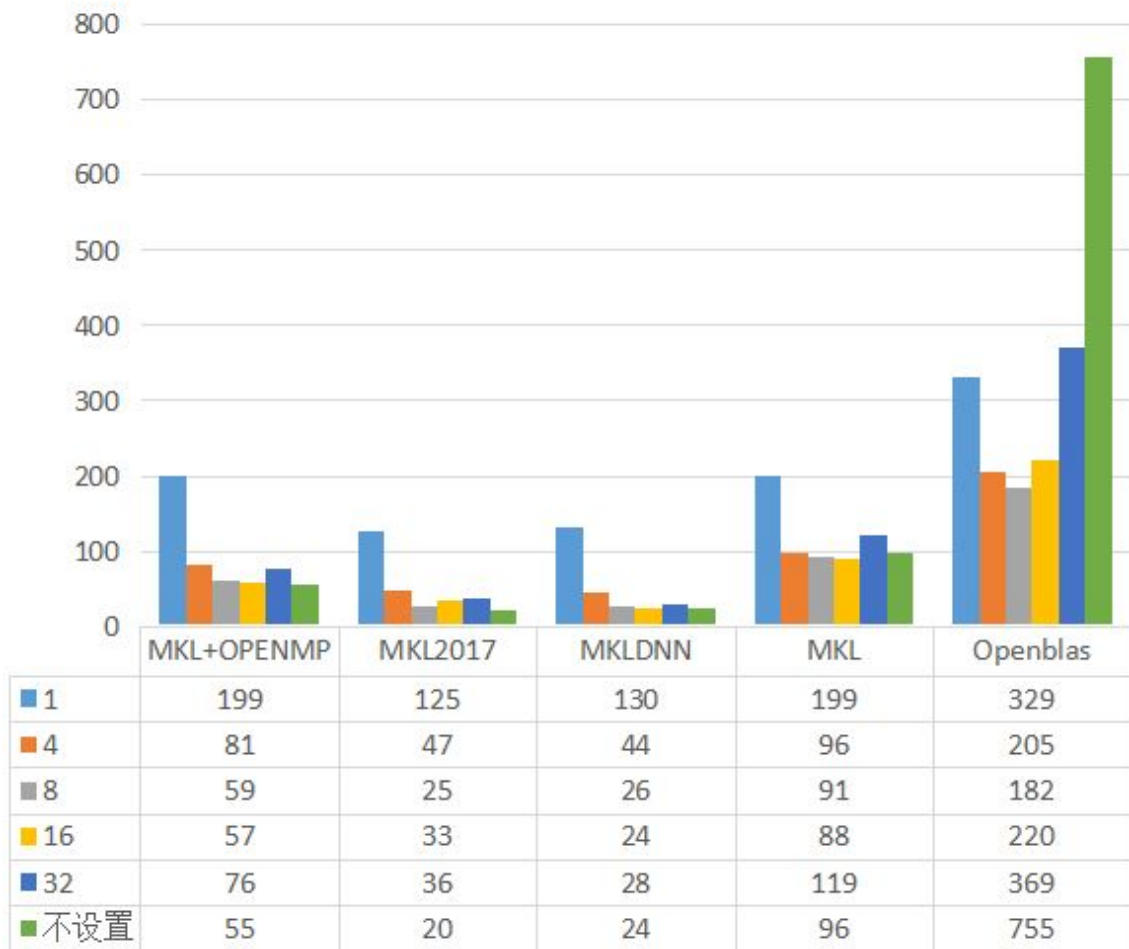
5.不同型号 cpu 之间的比较,E5-2670 v2 和 E5-2660v3 的比较

2670 v2上的性能



注：图片中所有的时间单位都为ms， 同一簇的不同颜色块表示不同的openmp 设置

2660 v3上的性能



这个实验结果可以说尺度达到我伙呆：

2660 v3 比 2670 v2 的速度会快非常大！很明显看出 Intel 是一名心机 boy，为了出货，狂给新技术加持 buff，所以在选择 cpu 的时候要注意了

Reference:

Intel caffe github 地址: <https://github.com/intel/caffe>

Intel caffe 硬件环境优化说明地址: <https://github.com/intel/caffe/wiki/Recommendations-to-achieve-best-performance>

Intel caffe 时间测试方法以及 openmp 优化说明:

<https://software.intel.com/zh-cn/articles/caffe-optimized-for-intel-architecture-applying-modern-code-techniques>

部门: WXG 模式识别中心
于志鹏/zhipengyu 实习生
2017/6/30

注: 图片中所有的时间单位都为 ms, 同一簇的不同颜色块表示不同的 openmp 设置