

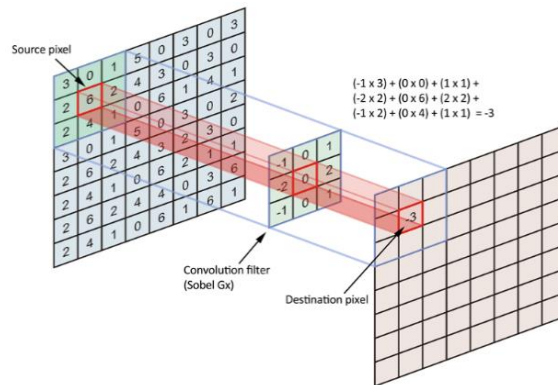
《计算机组成原理》大作业

一、相关知识准备：

1. 学习 MIPS 汇编语言，结合课上所学，熟练掌握相关 MIPS 汇编知识。
2. 通过自己阅读参考文献，理解神经网络推理的过程。
3. 用汇编编写神经网络的推理过程

二、课程设计内容：

1. 使用 MIPS 汇编和 MIPS 仿真器，设计并实现一个普通整数卷积计算算子，要求有完整的输入输出，输入为 $7*7*1$ 格式的张量，对应的卷积核一个，尺寸为 $3*3*1$ ，步长为 1，输出为经过卷积计算后的对应的 $5*5*1$ 张量，要求计算结果正确。参考卷积操作图：



2. 要求：（1）写出上述卷积过程的 C 代码，并验证其正确性；（2）用所学的 MIPS 指令编写汇编代码，并在 MARS 仿真器上运行验证代码正确性；
3. 自己学习 BN, Pooling, Activation 过程，可以尝试在卷积过程后面补充 BN, Pooling, Activation 过程(选择一种就可以,比如 Pooling 可以选择 max pooling 或者 average pooling 都可以),并用汇编实现上述过程，在 MARS 仿真器上运行验证代码正确性；（选做）
4. 阅读论文 1 和论文 2，学习二值神经网络的优化方式，并学习京东二值神经网络在 ARM 上的汇编实现方式，并写一个论文报告阐述论文的优化方法，以及想想我们如何在 MIPS 上实现类似过程；

三、要求

1. 完成课程设计报告，描述汇编代码实现过程，以及展示实验结果
2. 提交基于 MIPS 的代码，并能正确在 MARs 上运行并输出结果(结果输出在一个文件里)
3. 提交时间：2021 年 12 月 15 日

四、参考内容：

1. 论文 1: daBNN: A Super Fast Inference Framework for Binary Neural Networks on ARM devices, <https://arxiv.org/abs/1908.05858>
2. 代码 1: <https://github.com/JDAI-CV/dabnn/blob/master/dabnn>
3. 论文 2: PhoneBit: Efficient GPU-Accelerated Binary Neural Network Inference Engine for Mobile Phones, <https://ieeexplore.ieee.org/document/9116236>（二值神经网络参考）