

## 2022~2023 学年度第二学期《智能芯片设计》期末考查题

要求:

- (1) 课外完成, 2023 年 7 月 15 日 12: 00 前提交到指定邮箱, 逾期不接收。
- (2) 第 1-5 题, 需手写在 A4 打印纸上并转为 pdf 电子版; 第 6 题是报告, 需按照模板要求提交 pdf 电子版。
- (3) 本考查题页面、第 1-5 题手写答题页面、第 6 题页面合并为一个 pdf 文件, 务必按顺序编页码, 文件名按照“学号-姓名”命名。

专业班级: \_\_\_\_\_ 学 号: \_\_\_\_\_ 姓 名: \_\_\_\_\_

题号	1	2	3	4	5	6	总分
分数							

1. (8 分) 对于一个通道的输入特征图  $X$  矩阵 ( $5 \times 5$ , 如左图), 卷积核  $W$  ( $3 \times 3$ , 如右图), 设定卷积步长为 2、边界补 0 扩展长度为 1, 请分别写出直接计算、矩阵向量乘计算卷积的过程, 并对比其计算特性差异。

1	1	1	1	2
1	1	2	1	1
0	0	2	1	2
0	0	0	1	2
1	2	1	1	1

1	0	1
0	0	-1
0	0	-1

2. (10 分) 以尺寸  $H \times L \times N = 64 \times 64 \times 128$  的输入特征图、 $M = 12$  组尺寸  $K \times K \times N = 3 \times 3 \times 128$  的卷积核为例, 请结合示意图和伪代码分别阐述在  $T_m = 3$ 、 $T_n = 4$ 、 $T_h = T_l = 3$ 、卷积步长  $S = 1$  等参数设置情况下, 采用权重数据复用、输入数据复用的卷积计算流程。

3. (12 分)现有一卷积神经网络，其卷积层部分结构如下：16 通道  $64*64$  输入特征图，先通过一常规卷积层，输出 32 通道  $62*62$  特征图，再输入一深度可分离卷积层，最终输出 16 通道  $58*58$  特征图，请回答以下问题：
- (1) 求该神经网络卷积层部分的权重个数。
  - (2) 说明深度卷积与逐点卷积 1) 在网络轻量化方面的作用；2) 在信息流动方面的区别；3) 为何要串联二者构成深度可分离卷积？
4. (8 分)结合实例对比分析 AI 处理器时域架构与空域架构的特点。
5. (12 分)异构 DSA (Domain Specific Architecture) 是当前智能芯片的主流架构，请从计算并行形式、内存层次结构、数据精度以及特定领域语言的角度阐述其在智能计算中相对于冯诺依曼架构效率更高、能耗更低的原因，以及软硬件协同的重要性。
6. (50 分)报告：请围绕智能芯片“高效能、低功耗”这一关键目标，重点从数据复用和高效加速计算的角度来阐述其核心内涵，从数据复用技术、高效并行架构、模型轻量化设计等层面，梳理当前智能芯片的发展脉络，并展望未来的发展方向。(按照提供的模板格式编排，5-6 页纸，标题、内容等自行编排)