

# 团 体 标 准

T/ CESA 1121—2020

## 人工智能芯片 面向端侧的深度学习 芯片测试指标与测试方法

AI chips-Test metrics and test method of deep learning chips for terminal side

2020 - 10 - 30 发布

2020 - 11 - 10 实施



版权保护文件

版权所有归属于该标准的发布机构,除非有其他规定,否则未经许可,此发行物及其章节不得以其他形式或任何手段进行复制、再版或使用,包括电子版,影印件,或发布在互联网及内部网络等。使用许可可于发布机构获取。

## 目 次

前 言.....	III
1 范围.....	1
2 规范性引用文件.....	1
3 术语和定义.....	1
4 缩略语.....	2
5 技术要求.....	2
5.1 环境要求.....	2
5.2 基本技术要求.....	3
5.3 测试框架.....	3
5.4 测试对象.....	3
5.5 测试流程.....	3
5.6 测试场景信息.....	3
5.7 作业到达.....	3
5.8 模型压缩规则.....	3
6 测试方法.....	4
6.1 测试指标.....	4
6.2 测试指标描述.....	4
6.2.1 最大吞吐性能.....	4
6.2.2 平均前向推理速率.....	4
6.2.3 前向推理时延.....	4
6.2.4 功耗.....	4
6.2.5 能效比.....	4
6.2.6 模型推理精度.....	4
6.3 测试类型.....	5
6.3.1 通用测试.....	5
6.3.2 专用测试.....	5
6.4 指标测试方法.....	5
6.4.1 最大吞吐性能.....	5
6.4.2 平均前向推理速率.....	5
6.4.3 前向推理时延.....	5
6.4.4 功耗.....	5
6.4.5 能效比.....	5
6.4.6 模型推理精度.....	5
附录 A （规范性附录）测试框架说明.....	6
附录 B （资料性附录）测试流程说明.....	7
附录 C （资料性附录）端侧深度学习芯片通用测试参数参照表.....	8