



# 登临 Hamming™ V2

## SDK 文档简介

DL-DG/SW-040B-04

2025-01-17

Copyright©苏州登临科技有限公司，2019 - 2025，版权所有。

未经苏州登临科技有限公司事先书面同意，不得以任何形式或方式复制或传播本文件的任何部分。

## 商标和许可



和其它苏州登临科技有限公司的其它登临科技的图标为苏州登临科技有限公司的商标。本手册中提及的所有其他商标均为其各自所有者的财产。

## 通知

所购买的产品、服务和特性由苏州登临科技有限公司与客户签订的合同规定。本文件中描述的所有或部分产品、服务和特性可能不在采购范围或使用范围内。除非合同中另有规定，本文件中的所有声明、信息和建议均按“原样”提供，无任何明示或暗示的保证或陈述。

本手册中的信息如有更改，恕不另行通知。本文件在编制过程中已尽一切努力确保内容的准确性，本文件中的所有声明、信息和建议不构成任何明示或暗示的保证。

苏州登临科技有限公司

苏州工业园区扬富路11号南岸新地一期商务楼5号1101室，江苏，中国

<http://www.denglin.ai>

email : support@denglin.ai

# 更新历史

版本	更新描述
04	修改 表 1 Goldwasser II™ AI 加速卡文档简介； 修改 表 2 Hamming V2 软件栈软件模块文档简介； 新增 表 3 Hamming V2 软件栈工具使用文档简介
03	新增PyCUDA内容； 增加对不同系统下文档内容差异的说明； 更新模块文档简介
02	内容修订
01	第一次发布

# 登临 Hamming™ V2 SDK 文档简介

登临 Hamming™ V2 软件栈是驱动登临 Goldwasser II™ AI 加速卡的软件包。

成功安装 Hamming V2 SDK 以后，Hamming 软件栈的文档存放在 Hamming SDK 安装目录下的 documents 目录中。

表 1 Goldwasser II™ AI 加速卡文档简介

类型	文档名	简述
Hamming™ 简介	登临Hamming V2 软件栈简介.pdf	登临 Hamming V2 软件栈介绍
SDK安装	登临Hamming V2 SDK 安装指南.pdf	登临 Hamming V2 SDK 安装方法
SDK驱动安装	登临Hamming V2 驱动安装指南.pdf	登临 Hamming V2 SDK 驱动程序的安装方法
软件模块	见 表 2 Hamming V2 软件栈软件模块文档	登临 Hamming V2 软件栈各软件模块的相关文档
SDK配套工具使用说明	见 表 3 Hamming V2 软件栈工具文档	登临 Hamming V2 软件栈各工具的使用相关说明文档

说明：

- 《登临Hamming V2 SDK 安装指南》根据适配的不同操作系统区分多个版本，使用时请选择合适的版本。
- 《登临Hamming V2 驱动安装指南》根据适配的操作系统分为Linux版本和Windows版本，使用时请选择合适的版本。

表 2 Hamming V2 软件栈软件模块文档

软件模块	模块简述	文档名称	内容简述
dICU	dICU 模块提供在登临人工智能处理器里的并行计算引擎上通用的并行计算平台和编程模型，为更高效地解决人工智能领域中许多复杂的计算问题。	dICU-Programming-Guide.pdf	dICU 软件模块编程手册
		dICU-Extended-API.pdf	dICU 软件模块扩展 API 手册
		dICU-Extended-Math-API.pdf	dICU 软件模块扩展数学库函数 API 手册
		Denglin Texture Feature Guide.pdf	dICU Texture 特性介绍

		DengLin-Compute-Compiler-User-Guide.pdf	登临 CUDA & OpenCL 编译器使用简介
dINNE	dINNE 模块提供在登临人工智能处理器上执行高性能推理的编程接口和软件库。	dINNE-Developer-Guide.pdf	dINNE 软件模块开发者手册
		dINNE-API.pdf	dINNE 软件模块 API 手册
		dINNE-Build-Modulator.pdf	dINNE 编译构建调制器 API 手册
		dINNE-TVM-Quantization.pdf	dINNE TVM 量化 API 手册
	dINNE-Quant 工具包提供深度神经网络模型量化工具，量化后的模型可直接在登临人工智能处理器上高性能运行。	dINNE-Quant-Introduction-To-Quantization.pdf	网络模型量化方法介绍
		dINNE-Quant-Quantization-Tutorial-With-ResNet-V1-50.pdf	基于 ResNet-V1-50 的网络模型量化教程
		dINNE-Quant-TU-Operator.pdf	TU (Tensor Uint) 算子介绍
dIDNN	dIDNN 模块提供在登临人工智能处理器上加速深度神经网络算子计算的软件库。	dIDNN-Introduction.pdf	dIDNN 软件模块介绍
dIVID	dIVID 模块提供登临人工智能处理器的多路视频编解码引擎的编程接口和软件库。	dIVID-Programming-Guide.pdf	dIVID 软件模块编程手册
		dIVID-API.pdf	dIVID 软件模块 API 手册
dIJPEG	dIJPEG 模块提供登临人工智能处理器的多路 JPEG 数据编解码引擎的编程接口和软件库。	dIJPEG-Programming-Guide.pdf	dIJPEG 软件模块编程手册
		dIJPEG-API.pdf	dIJPEG 软件模块 API 手册
dIML	dIML 模块提供登临人工智能处理器管理的软件库。	dIML-API.pdf	dIML 软件模块 API 手册
diUTIL	diUTIL 提供登临人工智能处理器的命令行工具。		

		DengLin-System-Management-User-Guide.pdf	登临 dlsmi 使用简介
diBLAS	diBLAS 模块提供登临人工智能处理器的线性代数运算算子库。	diBLAS-Introduction.pdf	diBLAS 模块介绍
diBLASLt	diBLASLt是一个轻量级库，用于GEgeneral Matrix-to-Matrix Multiply ( GEMM ) 操作。	diBLASLt-Introduction.pdf	diBLASLt 模块介绍
diSOLVER	diSOLVER 支持CUDA 10.0 cuSOLVER的部分求解功能API。	diSOLVER 使用说明.pdf	diSOLVER 模块介绍
PyCUDA	使用PyCUDA可以通过Python访问NVIDIA的CUDA API。	Denglin Hamming V2 PyCUDA API.pdf	PyCUDA API接口手册
-	MIG ( Multi-instance GPU ) 功能允许GPU被安全地划分为多个实例从而实现GPU的最佳利用率。	Denglin Hamming V2 MIG Feature Guide.pdf	MIG 功能介绍
diSPARSE	CUDA cuSPARSE 是一个用于CUDA加速库的稀疏矩阵计算库，登临提供了diSPARSE兼容CUDA cuSPARSE。	登临 Hamming V2 diSPARSE 使用说明.pdf	介绍登临diSPARSE v1.0 与 CUDA 11.7 cuSPARSE的差异和使用方法
DCCL	DCCL 基于NVIDIA开源的NCCL开发并且基于pci-e进行p2p通信。	登临 Hamming V2 NCCL 使用说明.pdf	介绍登临DCCL与NVIDIA NCCL的差异和使用方法

提示：

Hamming V2 软件栈的模块和文档在不同操作系统中略有区别，详细情况请咨询support@denglin.ai。

表 3 Hamming V2 软件栈工具文档

工具	文档名称	内容简述
Profiler	登临Hamming V2 Profiler 使用说明.pdf	登临Profiler工具用于获取软硬件相关的 Profile 信息，并对信息完成初步分析。本文介绍登临Profiler工具的使用方法。
nnexec	登临nnexec工具使用说明.pdf	nnexec工具是登临提供的命令行包装工具。本文介绍登临nnexec工具的使用方法。
PyTorch	登临 Hamming V2 PyTorch 使用说明.pdf	介绍在Ubuntu环境下搭建和配置登临 AI 加速卡相关的软件和硬件环境的方法和步骤。

工具	文档名称	内容简述
NPI	登临Hamming V2 NPI 工具集使用说明.pdf	NPI工具集是登临在新产品导入中评估硬件环境正确性及硬件关键指标的一组测试工具。本文介绍登临NPI工具集的使用方法。
GPU P2P	登临Hamming V2 GPU P2P使用说明.pdf	登临GPU间P2P既支持DMA操作，也支持cu kernel之间的互操作，应用程序的编程接口与标准CUDA API基本兼容。本文介绍登临 GPU P2P 的使用方法。
GDS	登临Hamming V2 GDS使用说明.pdf	本文介绍了 GDS 在登临平台上的外部安装依赖包、驱动安装和功能验证步骤以及cufile的函数说明。