

NPU Unified Driver v2196 Release Notes

Wednesday 28 February 2024 11:45

Executive Summary

Build Details

Ingredient	Versions and Links
Build Details	24WW08 UD v2196
MCDM VPU driver	32.0.100.2196
VPUFW	MTL - ci tag ud202408 vpu rc 20240221 2101 dKMB - ci tag dkmb vpu rc 20240220 0201
OpenVINO VPUx	ci tag ud202408 vpux rc 20240220 0201
OpenVINO VPUx Tools	ci tag ud202408 vpux rc 20240220 0201
Public NN Models	https://wiki.ith.intel.com/pages/viewpage.action?pageId=3442936943#id-24ww08UD2024.0832.0.100.2196-ModelsVerification
DML Models	https://ubit-artifactory-or.intel.com/artifactory/movidius-vpu-or-local/Archive/Automation/Binaries/DML/Models/
OWR Release Links	NPU driver: OpenVINO:

Release Validation

Test Case Summary

QAC.xml

<https://wiki.ith.intel.com/pages/viewpage.action?pageId=3442936943#id-24ww08UD2024.0832.0.100.2196-QAC>

Full Validation report:

MTL:

http://releases.ti.intel.com/reports/meteorlake/NightlyRegressions/2024/2024.02.22/VPU_FW_REGRESSION/ci_tag_ud202408_vpu_rc_20240221_2101/MTL_UD202408_regression_-_VPU_FW_REGRESSION_-_2024.02.22_-_2101.html

Total results						
date	type	passing rate	time_min	tests	successes	failures
2024.02.22 21:01	DML	85.57% (4.72) ↑	218.80 (142.56) ↑	97 (3) ↑	83 (7) ↑	14 (-4) ↓
	Functional	89.22% (-0.36) ↓	2347.71 (-18.86) ↓	1763 (103) ↑	1573 (86) ↑	31 (17) ↑
	Stability	99.7%	17605.01 (12940.86) ↑	12579 (10153) ↑	12541 (10119) ↑	38 (34) ↑

dKMB:

http://releases.ti.intel.com/reports/DiscreteKeemBay/NightlyRegressions/2024/2024.02.22/VPU_FW_REGRESSION/ci_tag_ud202408_vpu_rc_20240222_2101/dKMB_ud202408_regression_-_VPU_FW_REGRESSION_-_2024.02.22_-_2101.html

Total results						
date	type	passing rate	time_min	tests	successes	failures
2024.02.22 21:01	Functional	99.38% (-0.62) ↓	12464.21 (10390.50) ↑	160 (11) ↑	159 (10) ↑	1 (1) ↑

Config Details

BKC	RPL-P-SV2-CORP-23.25.3.160
IFWI	RPLPFWI1.R00.4221.A00.2305271351

REFERENCE HARDWARE CONFIGURATION:

Host Info (wiki)				
RVP:	RPL P LP5 T4 RVP FAB1 BRD	MTL-H682 & P681	MTL-U282	ARL-S
MCP:	RPL P CPU J02	C0		
BKC:	RPL-P-SV2-CORP-23.25.3.160	MTL-H 682 MTL-H-CORP-RST-24.05.2.332	MTL-U282-9W-CORP-RST-23.51.5.304	ARL-S-MTP-S-SV2-CORP-24.02.6.99
Additional info:	Windows 11 22H2 (OS Build 22621.2134)			
Reworks:	-			
BIOS settings:	RPL_PR01_J0B1-XXXADPP_RPSF_SES5_01F70498_CVF_2023WW23.3.01.bin	MTLPFWI1.R00.3471.D89.2401091901	MTLMFWI1.R00.3471.D83.2312140812	MTLSFWI1.R00.3494.D91.2401112108
VPU hardware :	FAB-C			

KNOWN AND/OR NEW ISSUES:

MTL:

Key	Summary	Priority	Status	Component
EISW-114599	[MTL] Accuracy issue of Wondershare[HumanSegV3.1.0.4][INT8] model	P1- Stopper	OPEN	ip.sw.host.compiler.nn
EISW-114516	[MTL][CONFIG_5] ~25-30% performance drop on MSFT and public models with benchmark app	P1- Stopper	OPEN	ip.sw.host
EISW-114512	[MTL][CONFIG_1] ~25% performance drop on MSFT models with benchmark app	P1- Stopper	OPEN	ip.sw.host
EISW-112858	Review profiling-manager coverity hits for UD2024.08 release	P1- Stopper	IMPLEMENTED	ip.sw.tools.profilin g
EISW-111849	[MTL][NPUW/Report] NPU can't compile models in parallel	P1- Stopper	IN PROGRESS	ip.sw.host.ov.pipelines , ip.sw.host.ov.plugin
EISW-111821	[MTL][OVEP] ONNX OVEP test cases fail on loading onnxruntime_providers_openvino.dll library	P1- Stopper	OPEN	ip.sw.host.ov_validation.system
EISW-108393	IE-MDK smoke_OVClassBasicTestP test fails if run on release pipeline	P1- Stopper	IN PROGRESS	ip.sw.host.ov.plugin
EISW-97997	Memory usage is higher than estimation when run Benchmark app on NPU for stable diffusion model	P1- Stopper	VERIFIED	ip.sw.host.compiler.nn , ip.sw.host.drv.win.umd.L0

dKMB:

--	--	--	--	--

Key	Summary	Priority	Status	Component s
EISW-77455	[dKMB][IRv11] Performance drop for xml "ie" cases	P3- Medium	IN REVIEW	validation.s ystem
EISW-110667	[dKMB] Multitply power cycle fails on Raptor Lake	P3- Medium	IN PROGRESS	validation.c ore
EISW-102752	[dKMB] Report doesn't refresh after secure tests - wrong TAG parameter (EISW-100597 part2)	P2- High	AWAITING DEVELOPMENT	devops.ci_c d.build.1cl ic, validation .reporting
EISW-113544	[dKMB][Secure][Report] Refresh of report not triggered after secure jobs	P2- High	OPEN	validation.r eporting
EISW-113018	[dKMB][Security] Page_fault test fails to trigger Engine Reset	P2- High	OPEN	ip.sw.host, validation.s dl
EISW-111612	[dKMB] Performance IR with regression models on npudriver-31.0.100.2076	P3- Medium	OPEN	validation.c ore, validati on.system