



NPU 算子支持列表

版本号：1.5

发布时间：2025-04-03

版本历史

版本	日期	责任人	版本描述
1.0	2023-10-12	AWA0385	创建文档。
1.1	2024-02-04	AWA1382	1) 新增第 1 章 前言； 2) 第 2 章 a. 更新标题为 V85X/R853/MR527/T527/AI985 算子支持列表； b. 补充运行模块、数据类型和约束规格； c. 补充算子 acosh、atan、atanh、inverse_sigmoid、tan、lpnorm、reducel2、avg_pool3d、globalpool、maxunpool、crop_and_resize、resize_3d、reversesequence； 3) 新增附录 各框架的算子支持列表。
1.2	2024-04-11	AWA1382	新增第 3 章 MR536/T536 算子支持列表。
1.3	2024-06-17	AWA1382	第 2、3 章，补充算子约束规格。
1.4	2024-11-13	AWA1382	第 3 章，补充 A733 说明。
1.5	2025-04-03	AWA1382	第 3 章，补充 T736 说明。

目录

版本历史.....	i
目录.....	ii
表格目录.....	iii
1 前言.....	1
1.1 文档简介.....	1
1.2 目标读者.....	1
1.3 适用范围.....	1
1.4 文档约定.....	1
1.4.1 缩写约定.....	1
2 V85X/R853/MR527/T527/AI985 算子支持列表.....	2
2.1 Basic Operations.....	2
2.2 Activation Operations.....	3
2.3 Elementwise Operations.....	5
2.4 Normalization Operations.....	5
2.5 Reshape Operations.....	6
2.6 RNN Operations.....	7
2.7 Pooling Operations.....	7
2.8 Miscellaneous Operations.....	8
3 MR536/T536/A733/T736 算子支持列表.....	10
3.1 Basic Operations.....	10
3.2 Activation Operations.....	11
3.3 Elementwise Operations.....	13
3.4 Normalization Operations.....	14
3.5 Reshape Operations.....	15
3.6 RNN Operations.....	16
3.7 Pooling Operations.....	17
3.8 Miscellaneous Operations.....	18
附录：各框架的算子支持列表.....	20

表格目录

表 1 TensorFlow 的算子支持列表.....	20
表 2 TFLite 的算子支持列表.....	22
表 3 Keras 的算子支持列表.....	24
表 4 Onnx 的算子支持列表.....	24
表 5 Caffe 的算子支持列表.....	26
表 6 Darknet 的算子支持列表.....	26



1 前言

1.1 文档简介

给出各平台的 NPU 算子支持列表。

1.2 目标读者

本文档主要适用于以下人员：

- 技术支持工程师
- AI 算法开发工程师
- AI 软件开发工程师

1.3 适用范围

硬件平台：V85x、R853、MR527、T527、AI985、MR536、T536、A733、T736

软件平台：Tina 系统、Android13 及以上系统

工具版本：Acuity6.21.x、IDE5.8.2 及以上

1.4 文档约定

1.4.1 缩写约定

本文档在描述数据类型和运行模块时使用的缩写具体如下：

模块	缩写	对应全写
数据类型	i8	int8
	u8	uint8
	i16	int16
	fp32	float32
	fp16	float16
	bf16	bfloat16
运行模块	NN	Neural-Network Engine
	TP	Tensor Processor
	PPU	Parallel Processing Unit

2 V85X/R853/MR527/T527/AI985 算子支持列表

2.1 Basic Operations

算子	运行模块	数据类型	约束规格
conv1d	NN	i8、u8、i16	input/output size: width, height: [1, 8191] channel:[1, 16383] kernel size: x, y: [1, 255] $x*y \leq 255$ stride: x, y:[1, 2] padding: [-15, 16], 仅支持 constant mode
	PPU	fp32、fp16、bf16	
conv2d	NN	i8、u8、i16	input/output size: width, height: [1, 8191] channel:[1, 16383] kernel size: x, y: [1, 15] stride: x, y:[1, 2] padding: [-15, 16], 仅支持 constant mode
	PPU	fp32、fp16、bf16	
conv3d	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
deconvolution	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
deconvolution1d	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
fcl2	TP	i8、u8、i16、fp16	Input size(2dim): x:[1, 65535], y == 1 output size(2dim): x:[1, 512], x != 1, y == 1 kernel size(2dim): x:[1, 65535], y:[1, 512]
	PPU	fp32、bf16	
grouped_conv1d	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	

算子	运行模块	数据类型	约束规格
grouped_conv2d	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	

2.2 Activation Operations

算子	运行模块	数据类型	约束规格
abs	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
acosh	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
atan	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
atanh	PPU	i8、u8、i16、fp32、fp16、bf16	
celu	TP	i8、u8、i16	
	PPU	fp32、fp16、bf16	
clip	TP	i8、u8、i16	
	PPU	fp32、fp16、bf16	
cos	PPU	i8、u8、i16、fp32、fp16、bf16	
elu	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
erf	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
exp	PPU	i8、u8、i16、fp32、fp16、bf16	
gelu	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
hard_sigmoid	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
inverse_sigmoid	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
leaky_relu	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
linear	PPU	i8、u8、i16、fp32、fp16、bf16	
log	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	

算子	运行模块	数据类型	约束规格
log_softmax	PPU	i8、u8、i16、fp32、fp16、bf16	
mish	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
neg	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
prelu	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
rcp	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
relu	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
relun	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
rsqrt	PPU	i8、u8、i16、fp32、fp16、bf16	
selu	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
sigmoid	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
sign	PPU	i8、u8、i16、fp32、fp16、bf16	
sin	PPU	i8、u8、i16、fp32、fp16、bf16	
softmax	PPU	i8、u8、i16、fp32、fp16、bf16	
softrelu	PPU	i8、u8、i16、fp32、fp16、bf16	
softsign	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
sqrt	PPU	i8、u8、i16、fp32、fp16、bf16	
square	PPU	i8、u8、i16、fp32、fp16、bf16	
swish	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
tan	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
tanh	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	

2.3 Elementwise Operations

算子	运行模块	数据类型	约束规格
add	NN	i8、u8、i16	input/output size: x, y:[1, 8191] z:[1, 16383]
	PPU	fp32、fp16、bf16	
addn	PPU	i8、u8、i16、fp32、fp16、bf16	
divide	PPU	i8、u8、i16、fp32、fp16、bf16	
floordiv	PPU	i8、u8、i16、fp32、fp16、bf16	
logical_not	PPU	bool8	
logical_ops	PPU	bool8	
matrixmul	PPU	i8、u8、i16、fp32、fp16、bf16	
maximum	PPU	i8、u8、i16、fp32、fp16、bf16	
minimum	PPU	i8、u8、i16、fp32、fp16、bf16	
mod	PPU	i8、u8、i16、fp32、fp16、bf16	
multiply	PPU	i8、u8、i16、fp32、fp16、bf16	
pow	PPU	i8、u8、i16、fp32、fp16、bf16	
relational_ops	PPU	i8、u8、i16、fp32、fp16、bf16、bool8	输出 bool8
select	PPU	i8、u8、i16、fp32、fp16、bf16	
subtract	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	

2.4 Normalization Operations

算子	运行模块	数据类型	约束规格
batch_norm	PPU	i8、u8、i16、fp32、fp16、bf16	
batchnorm_single	PPU	i8、u8、i16、fp32、fp16、bf16	
group_norm	PPU	i8、u8、i16、fp32、fp16、bf16	
instance_norm	PPU	i8、u8、i16、fp32、fp16、bf16	
l2_normalize	PPU	i8、u8、i16、fp32、fp16、bf16	
layer_norm	PPU	i8、u8、i16、fp32、fp16、bf16	
lpnorm	PPU	i8、u8、i16、fp32、fp16、bf16	
ln2	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
moments	PPU	i8、u8、i16、fp32、fp16、bf16	

2.5 Reshape Operations

算子	运行模块	数据类型	约束规格
argmax	PPU	i8、u8、i16、fp32、fp16、bf16	输出可支持 int32
argmin	PPU	i8、u8、i16、fp32、fp16、bf16	输出可支持 int32
batch2space	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
concat	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
depth2space	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
expand_broadcast	PPU	i8、u8、i16、fp32、fp16、bf16	
pad2	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
permute	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
reduce	PPU	i8、u8、i16、fp32、fp16、bf16	
reducel2	PPU	i8、u8、i16、fp32、fp16、bf16	
reorg	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
reshape2	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
reverse	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
shufflechannel	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
slice	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
space2batch	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
space2depth	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
split	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
squeeze	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	

算子	运行模块	数据类型	约束规格
stack	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
strided_slice	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
unstack	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	

2.6 RNN Operations

算子	运行模块	数据类型	约束规格
conv2d_lstm	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
conv2d_lstm_cell	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
gru	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
grucell	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
lstm_ovxlib	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
lstmunit_ovxlib	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
svdf	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	

2.7 Pooling Operations

算子	运行模块	数据类型	约束规格
avg_pool3d	PPU	i8、u8、i16、fp32、fp16、bf16	
globalppool	PPU	i8、u8、i16、fp32、fp16、bf16	
lppool	PPU	i8、u8、i16、fp32、fp16、bf16	
max_pool3d	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
maxpoolwitharg max	PPU	i8、u8、i16、fp32、fp16、bf16	
maxunpool	PPU	i8、u8、i16、fp32、fp16、bf16	

算子	运行模块	数据类型	约束规格
pool	NN	i8、u8、i16	MAX: input/output size: x, y: [1, 8191] z:[1, 16383] pooling size: x, y:[1, 3], x==y pooling stride: x, y:[1, 2], x==y pooling pad: [0, 3],仅支持 constant mode AVG: input/output size: x, y: [1, 8191] z:[1, 16383] pooling size: x, y:[1, 15] pooling stride: x, y:[1, 2], x==y pooling pad: [0, 3], 不支持 AVG_ANDROID
	PPU	fp32、fp16、bf16	
poolwithargmax	PPU	i8、u8、i16、fp32、fp16、bf16	
roi_pool	TP	i8、u8、i16、fp16	
	PPU	fp32、bf16	
upsample	PPU	i8、u8、i16、fp32、fp16、bf16	

2.8 Miscellaneous Operations

算子	运行模块	数据类型	约束规格
bucketize	PPU	i8、u8、i16、fp32、fp16、bf16	
cast	PPU	i8、u8、i16、bool8、fp32、fp16、bf16、int32	
ceil	PPU	i8、u8、i16、fp32、fp16、bf16	
concatshift	PPU	i8、u8、i16、fp32、fp16、bf16	
crop_and_resize	PPU	i8、u8、i16、fp32、fp16、bf16	
cumsum	PPU	i8、u8、i16、fp32、fp16、bf16	
dataconvert	TP	i8、u8、i16、fp16	

算子	运行模块	数据类型	约束规格
	PPU	fp32、bf16	
dropout	PPU	i8、u8、i16、fp32、fp16、bf16	
embedding_lookup	PPU	i8、u8、i16、fp32、fp16、bf16	
floor	PPU	i8、u8、i16、fp32、fp16、bf16	
gather	PPU	i8、u8、i16、fp32、fp16、bf16	
gather_elements	PPU	i8、u8、i16、fp32、fp16、bf16	
gather_nd	PPU	i8、u8、i16、fp32、fp16、bf16	
grid_sample	PPU	i8、u8、i16、fp32、fp16、bf16	
one_hot	PPU	i8、u8、i16、fp32、fp16、bf16	
proposal	PPU	i8、u8、i16、fp32、fp16、bf16	
repeat	PPU	i8、u8、i16、fp32、fp16、bf16	
resize	PPU	i8、u8、i16、fp32、fp16、bf16	
resize_3d	PPU	i8、u8、i16、fp32、fp16、bf16	
reversesequence	PPU	i8、u8、i16、fp32、fp16、bf16	
round	PPU	i8、u8、i16、fp32、fp16、bf16	
scatter_elements	PPU	i8、u8、i16、fp32、fp16、bf16	
scatter_nd	PPU	i8、u8、i16、fp32、fp16、bf16	
sequence_mask	PPU	i8、u8、i16、fp32、fp16、bf16	
signal_frame	PPU	i8、u8、i16、fp32、fp16、bf16	
tile	PPU	i8、u8、i16、fp32、fp16、bf16	
upsamplescale	PPU	i8、u8、i16、fp32、fp16、bf16	
variable	TP	i8、u8、i16	
	PPU	fp32、fp16、bf16	

3 MR536/T536/A733/T736 算子支持列表

3.1 Basic Operations

算子	运行模块	数据类型	约束规格
conv1d	NN	i8、u8、i16	input/output size: width, height: [1, 8191] channel:[1, 16383] kernel size: x, y: [1, 255] $x*y \leq 255$ stride: x, y:[1, 2] padding: [-15, 16], 仅支持 constant mode
	PPU	fp32、fp16、bf16	
conv2d	NN	i8、u8、i16	input/output size: width, height: [1, 8191] channel:[1, 16383] kernel size: x, y: [1, 15] stride: x, y:[1, 2] padding: [-15, 16], 仅支持 constant mode
	PPU	fp32、fp16、bf16	
conv3d	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
deconvolution	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
deconvolution1d	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
fcl2	NN	i8、u8、i16	input size(2dim) : x:[1, 1048575] y:[1, 8191*8191] output size(2dim): x:[1, 16383] y:[1, 8191*8191] kernel size(2dim): x:[1, 1048575] y:[1, 16383]

算子	运行模块	数据类型	约束规格
grouped_conv1d	PPU	fp32、fp16、bf16	
	NN	i8、u8、i16	
grouped_conv2d	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	

3.2 Activation Operations

算子	运行模块	数据类型	约束规格
abs	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
acosh	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
atan	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
atanh	NN	i8、u8	
	PPU	i16、fp32、fp16、bf16	
celu	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
clip	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
cos	NN	i8、u8	
	PPU	i16、fp32、fp16、bf16	
elu	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
erf	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
exp	NN	i8、u8	
	PPU	i16、fp32、fp16、bf16	
gelu	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
hard_sigmoid	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
inverse_sigmoid	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	

算子	运行模块	数据类型	约束规格
leaky_relu	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
linear	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
log	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
log_softmax	PPU	i8、u8、i16、fp32、fp16、bf16	
mish	NN	i8、u8、i16	input/output size: x, y:[1, 8191] z:[1, 16383]
	PPU	fp32、fp16、bf16	
neg	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
prelu	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
rcp	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
relu	NN	i8、u8、i16	input/output size: x, y:[1, 8191] z:[1, 16383]
	PPU	fp32、fp16、bf16	
relun	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
rsqrt	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
selu	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
sigmoid	NN	i8、u8、i16	input/output size: x, y:[1, 8191] z:[1, 16383]
	PPU	fp32、fp16、bf16	
sign	NN	i8、u8	
	PPU	i16、fp32、fp16、bf16	
sin	NN	i8、u8	
	PPU	i16、fp32、fp16、bf16	

算子	运行模块	数据类型	约束规格
softmax	NN	i8、u8、i16	input/output size: x, y:[1, 8191] z:[1, 16383]
	PPU	fp32、fp16、bf16	
softrelu	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
softsign	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
sqrt	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
square	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
swish	NN	i8、u8、i16	input/output size: x, y:[1, 8191] z:[1, 16383]
	PPU	fp32、fp16、bf16	
tan	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
tanh	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	

3.3 Elementwise Operations

算子	运行模块	数据类型	约束规格
add	NN	i8、u8、i16	input/output size: x, y:[1, 8191] z:[1, 16383]
	PPU	fp32、fp16、bf16	
addn	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
divide	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
floordiv	PPU	i8、u8、i16、fp32、fp16、bf16	
logical_not	NN	bool8	

算子	运行模块	数据类型	约束规格
logical_ops	NN	bool8	
matrixmul	NN	i8、u8、i16	inputA size: x:[1, 1048575], y:[1, 16383] inputB size: x:[1, 8191 * 8191], y:[1, 1048575] transpose mode: transposeA
	PPU	fp32、fp16、bf16	
maximum	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
minimum	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
mod	PPU	i8、u8、i16、fp32、fp16、bf16	
multiply	NN	i8、u8、i16	input/output size: x, y:[1, 8191] z:[1, 16383]
	PPU	fp32、fp16、bf16	
pow	PPU	i8、u8、i16、fp32、fp16、bf16	
relational_ops	NN	i8、u8、i16	输出 bool8
	PPU	fp32、fp16、bf16	
select	NN	i8、u8、i16、bool8	
	PPU	fp32、fp16、bf16	
subtract	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	

3.4 Normalization Operations

算子	运行模块	数据类型	约束规格
batch_norm	NN	i8、u8、i16	input/output size: x, y:[1, 8191] z:[1, 16383]
	PPU	fp32、fp16、bf16	
batchnorm_single	PPU	i8、u8、i16、fp32、fp16、bf16	
group_norm	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	

算子	运行模块	数据类型	约束规格
instance_norm	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
l2_normalize	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
layer_norm	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
lpnorm	PPU	i8、u8、i16、fp32、fp16、bf16	
lnr2	PPU	i8、u8、i16、fp32、fp16、bf16	
moments	PPU	i8、u8、i16、fp32、fp16、bf16	

3.5 Reshape Operations

算子	运行模块	数据类型	约束规格
argmax	PPU	i8、u8、i16、fp32、fp16、bf16	输出可支持 int32
argmin	PPU	i8、u8、i16、fp32、fp16、bf16	输出可支持 int32
batch2space	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
concat	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
depth2space	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
expand_broadcast	PPU	i8、u8、i16、fp32、fp16、bf16	
pad2	NN	i8、u8、i16	input/output size: x,y: [1, 8191] z:[1, 16383] pad size: [-15, 16]，仅支持 constant mode pad axis: x y
	PPU	fp32、fp16、bf16	
permute	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
reduce	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
reducel2	PPU	i8、u8、i16、fp32、fp16、bf16	

算子	运行模块	数据类型	约束规格
reorg	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
reshape2	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
reverse	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
shufflechannel	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
slice	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
space2batch	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
space2depth	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
split	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
squeeze	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
stack	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
strided_slice	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
unstack	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	

3.6 RNN Operations

算子	运行模块	数据类型	约束规格
conv2d_lstm	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
conv2d_lstm_cell	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
gru	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	

算子	运行模块	数据类型	约束规格
grucell	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
lstm_ovxlib	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
lstmunit_ovxlib	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
svdf	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	

3.7 Pooling Operations

算子	运行模块	数据类型	约束规格
avg_pool3d	PPU	i8、u8、i16、fp32、fp16、bf16	
globalppool	PPU	i8、u8、i16、fp32、fp16、bf16	
lppool	PPU	i8、u8、i16、fp32、fp16、bf16	
max_pool3d	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
maxpoolwithargmax	PPU	i8、u8、i16、fp32、fp16、bf16	
maxunpool	PPU	i8、u8、i16、fp32、fp16、bf16	
pool	NN	i8、u8、i16	<p>MAX: input/output size: x, y: [1, 8191] z:[1, 16383] pooling size: x, y:[1, 3], x==y pooling stride: x, y:[1, 2], x==y pooling pad: [0, 3],仅支持 constant mode</p> <p>AVG: input/output size: x, y: [1, 8191] z:[1, 16383] pooling size: x, y:[1, 15] pooling stride:</p>

算子	运行模块	数据类型	约束规格
			x, y:[1, 2], x==y pooling pad : [0, 3], 不支持 AVG_ANDROID
	PPU	fp32、fp16、bf16	
poolwithargmax	PPU	i8、u8、i16、fp32、fp16、bf16	
roi_pool	PPU	i8、u8、i16、fp32、fp16、bf16	
upsample	PPU	i8、u8、i16、fp32、fp16、bf16	

3.8 Miscellaneous Operations

算子	运行模块	数据类型	约束规格
bucketize	PPU	i8、u8、i16、fp32、fp16、bf16	
cast	NN	i8、u8、i16、bool8	
	PPU	fp32、fp16、bf16、int32	
ceil	PPU	i8、u8、i16、fp32、fp16、bf16	
concatshift	PPU	i8、u8、i16、fp32、fp16、bf16	
crop_and_resize	PPU	i8、u8、i16、fp32、fp16、bf16	
cumsum	PPU	i8、u8、i16、fp32、fp16、bf16	
dataconvert	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
dropout	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
embedding_lookup	PPU	i8、u8、i16、fp32、fp16、bf16	
floor	PPU	i8、u8、i16、fp32、fp16、bf16	
gather	PPU	i8、u8、i16、fp32、fp16、bf16	
gather_elements	PPU	i8、u8、i16、fp32、fp16、bf16	
gather_nd	PPU	i8、u8、i16、fp32、fp16、bf16	
grid_sample	PPU	i8、u8、i16、fp32、fp16、bf16	
one_hot	PPU	i8、u8、i16、fp32、fp16、bf16	
proposal	PPU	i8、u8、i16、fp32、fp16、bf16	
repeat	PPU	i8、u8、i16、fp32、fp16、bf16	
resize	PPU	i8、u8、i16、fp32、fp16、bf16	
resize_1d	PPU	i8、u8、i16、fp32、fp16、bf16	
resize_3d	PPU	i8、u8、i16、fp32、fp16、bf16	
reversesequence	PPU	i8、u8、i16、fp32、fp16、bf16	

算子	运行模块	数据类型	约束规格
round	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
scatter_elements	PPU	i8、u8、i16、fp32、fp16、bf16	
scatter_nd	PPU	i8、u8、i16、fp32、fp16、bf16	
scatter_nd_update	PPU	i8、u8、i16、fp32、fp16、bf16	
sequence_mask	PPU	i8、u8、i16、fp32、fp16、bf16	
signal_frame	PPU	i8、u8、i16、fp32、fp16、bf16	
tile	PPU	i8、u8、i16、fp32、fp16、bf16	
upsamplescale	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	
variable	NN	i8、u8、i16	
	PPU	fp32、fp16、bf16	

附录：各框架的算子支持列表

支持的 NN 框架包括如下：

- TensorFlow (TF) 2.15.0
- TensorFlow Lite (TFLite) generated from TF 2.15.0
- Keras generated from TF 2.15.0
- ONNX 1.14.0
- Caffe
- Darknet

各框架的算子支持列表如下（以字母顺序排列）。

表 1 TensorFlow 的算子支持列表

TensorFlow		
tf.abs	tf.expand_dims	tf.minimum
tf.add	tf.floor	tf.multiply
tf.add_n	tf.floordiv	tf.negative
tf.argmax	tf.gather	tf.nn.avg_pool
tf.argmin	tf.gather_nd	tf.nn.batch_normalization
tf.batch_matmul	tf.greater	tf.nn.bias_add
tf.batch_to_space_nd	tf.greater_equal	tf.nn.conv1d
tf.cast	tf.image.crop_and_resize	tf.nn.conv1d
tf.clip_by_value	tf.image.resize_bilinear	tf.nn.conv2d
tf.concat	tf.image.resize_nearest_neighbor	tf.nn.conv2d_transposed
tf.contrib.layers.instance_norm	tf.less	tf.nn.conv3d
tf.depth_to_space	tf.less_equal	tf.nn.depthwise_conv2d
tf.div	tf.logical_add	tf.nn.dynamic_rnn
tf.divide	tf.logical_or	tf.nn.elu
tf.equal	tf.logical_xor	tf.nn.embedding_lookup
tf.erf	tf.matmul	tf.nn.embedding_lookup
tf.exp	tf.maximum	tf.nn.fused_batch_norm

TensorFlow		
tf.nn.fused_batchnorm	tf.nn.dynamic_rnn	tf.scatter_nd
tf.nn.gelu	tf.not_equal	tf.select
tf.nn.l2_normalize	tf.one_hot	tf.signal.frame
tf.nn.leaky_relu	tf.pad	tf.slice
tf.nn.local_response_normalization	tf.pow	tf.space_to_batch_nd
tf.nn.max_pool	tf.realdiv	tf.space_to_depth
tf.nn.max_pool_with_argmax	tf.reduce_any	tf.split
tf.nn.max_pool3d	tf.reduce_max	tf.sqrt
tf.nn.moments	tf.reduce_mean	tf.square
tf.nn.relu	tf.reduce_mean	tf.squeeze
tf.nn.relu6	tf.reduce_sum	tf.stack
tf.nn.rnn_cell_GRUCell	tf.repeat	tf.strided_slice
tf.nn.rnn_cell_GRUCell	tf.reshape	tf.subtract
tf.nn.rnn_cell_LSTMCell	tf.reverse	tf.tile
tf.nn.sigmoid	tf.reverse_sequence	tf.transpose
tf.nn.softmax	tf.rnn_cell.LSTMCell	tf.unstack
tf.nn.swish	tf.round	tf.where
tf.nn.tanh	tf.rsqrt	

表 2 TFLite 的算子支持列表

TFLite		
ABS	FLOOR_MOD	NON_MAX_SUPPRESSION_V5
ADD	FULLY_CONNECTED	NOT_EQUAL
ADD_N	GATHER	ONE_HOT
ARG_MAX	GATHER_ND	PACK
ARG_MIN	GREATER	PAD
AVERAGE_POOL_2D	GREATER_EQUAL	PADV2
BATCH_MATMUL	HARD_SWISH	POW
BATCH_TO_SPACE_ND	L2_NORMALIZATION	PRELU
BROADCAST_ARGS	L2_POOL_2D	RANGE
BROADCASTTO	LEAKY_RELU	RANK
CONCATENATION	LESS	REDUCE_ANY
CONV_2D	LESS_EQUAL	REDUCE_MAX
CONV_3D	LOCAL_RESPONSE_NORMALIZATION	REDUCE_MIN
CUMSUM	LOG_SOFTMAX	REDUCE_PROD
DEPTH_TO_SPACE	LOGICAL_AND	RELU
DEPTHWISE_CONV_2D	LOGICAL_NOT	RELU_N1_TO_1
DEQUANTIZE	LOGICAL_OR	RELU1
DIV	LOGISTIC	RELU6
ELU	LSTM	RESHAPE
EMBEDDING_LOOKUP	MAX_POOL_2D	RESIZE_BILINEAR
EQUAL	MAXIMUM	REVERSE_V2
EXP	MEAN	ROUND
EXPAND_DIMS	MIRROR_PAD	RSQRT
FLOOR	MUL	SCATTER_ND
FLOOR_DIV	NEG	SEGMENT_SUM

TFLite		
SELECT	SQUARE	TOPK_V2
SHAPE	SQUARED_DIFFERENCE	TRANSPOSE
SIN	SQUEEZE	TRANSPOSE_CONV
SLICE	STRIDED_SLICE	UNIDIRECTIONAL_SEQUENCE_LSTM
SOFTMAX	SUB	UNIQUE
SPACE_TO_BATCH_ND	SUM	UNPACK
SPACE_TO_DEPTH	SVDF	UNSTACK
SPARSE_TO_DENSE	TANH	WHERE
SPLIT	TILE	ZEROS_LIKE
SPLIT_V	TOPK	

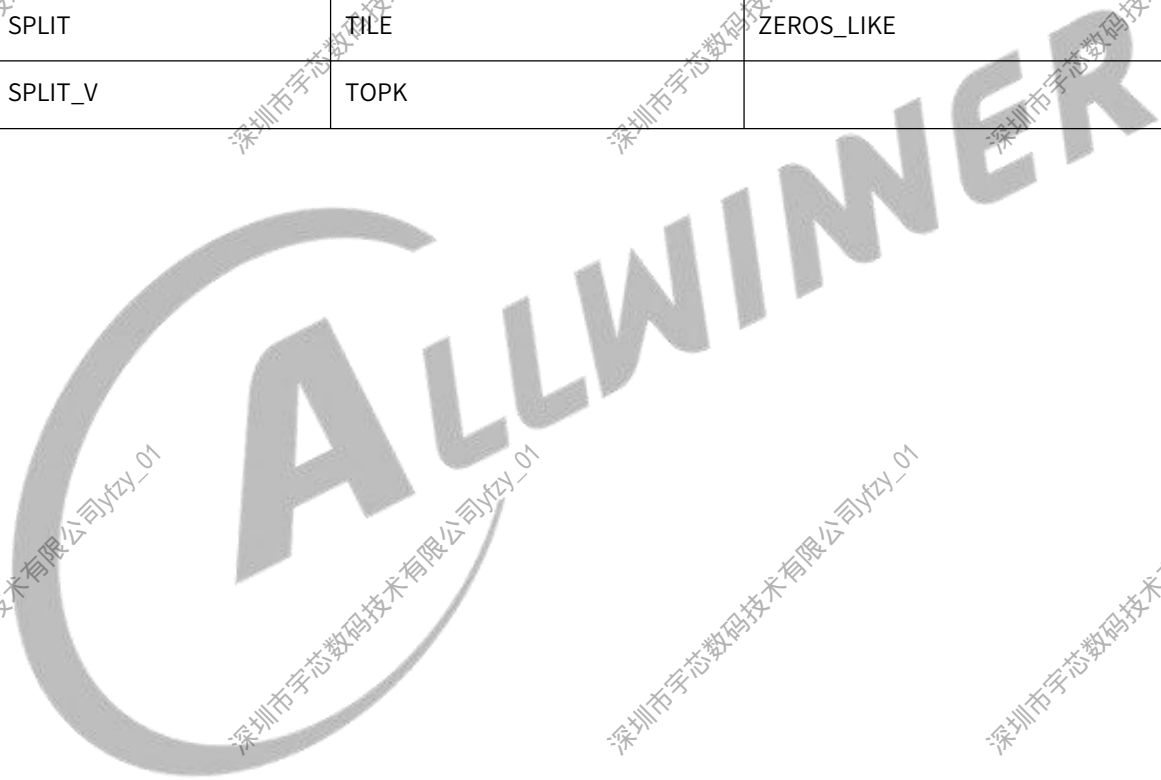


表 3 Keras 的算子支持列表

Keras			
Activation(leaky_relu)	Concatenate	GlobalMaxPooling2D	SeparableConv1D
Activation(relu)	Conv1D	GRU	SeparableConv2D
Activation(sigmoid)	Conv1DTranspose	LayerNormalization	SimpleRNN
Activation(softmax)	Conv2D	LeakyRelu	Softmax
Activation(tanh)	Conv2DTranspose	LSTM	Subtract
ActivityRegularization	Conv3D	Maximum	ThresholdedReLU
Add	ConvLSTM2D	MaxPooling1D	Upsampling1D
AdditiveAttention	Cropping1D	MaxPooling2D	UpSampling2D
AlphaDropout	Cropping2D	MaxPooling3D	UpSampling3D
Attention	Cropping3D	Minimum	ZeroPadding1D
Average	Dense	Multiply	ZeroPadding2D
AveragePooling1D	DepthwiseConv2D	Permute	ZeroPadding3D
AveragePooling2D	ELU	PReLU	
AveragePooling3D	Embedding	RELU	
BatchNormalization	Flatten	Reshape	
BatchNormalizationV1	GlobalAveragePooling2D	RNN	

表 4 Onnx 的算子支持列表

Onnx			
Abs	Atanh	Cast	Col2Im
Add	AveragePool	CastLike	Concat
And	BatchNormalization	Ceil	ConstantOfShape
ArgMax	BitwiseAnd	Celu	Conv
ArgMin	BitwiseOr	CenterCropPad	ConvTranspose
Atan	BitwiseXor	Clip	Cos

Onnx			
Cumsum	LessOrEqual	Reciprocal	Softmax
Elu	Log	ReduceL1	Softplus
Equal	Logsoftmax	ReduceL2	Softsign
Erf	LRN	ReduceLogSum	SpaceToDepth
Exp	LSTM	ReduceLogSumExp	Split
Expand	MatMul	ReduceMax	Sqrt
Flatten	Max	ReduceMean	Squeeze
Floor	MaxPool	ReduceMin	STFT
Gather	MaxRoiPool	ReduceProd	Sub
GatherElements	Mean	ReduceSum	Sum
GatherND	MeanVarianceNormalization	ReduceSumSquare	Tan
Gemm	Min	Relu	Tanh
GlobalAveragePool	Mish	Reshape	Tile
GlobalMaxPool	Mod	Resize	TopK
Greater	Mul	ReverseSequence	Transpose
GreaterOrEqual	Neg	Round	Unsqueeze
GridSample	NonZero	ScatterElements	Upsample
GroupNormalization	OneHot	ScatterND	Where
GRU	Or	Selu	Xor
HammingWindow	Pad	Shape	
HannWindow	Pow	Sigmoid	
HardSigmoid	Prelu	Sign	
HardSwish	QLinearConv	Silu	
InstanceNormalization	QLinearMatMul	Sin	
LeakyRelu	QuantizeLinear	Size	
Less	Range	Slice	

表 5 Caffe 的算子支持列表

Caffe			
absval	eltwise	pooling	roipooling
axpy	elu	poolwithargmax	scale
batchnorm	flatten	prelu	shufflechannel
bn	innerproduct	permute	sigmoid
concat	l2normalizescale	priorbox	slice
convolution	leakyrelu	proposal	softmax
convolutiondepthwise	lrn	relu	swish
deconvolution	lstm	reorg	tanh
depthwiseconvolution	normalize	reshape	
dropout	padchannel	reverse	

表 6 Darknet 的算子支持列表

Darknet			
avgpool	leaky	relu	softmax
batch_normalize	logistic	reorg	swish
connected	maxpool	route	upsample
convolutional	mish	scale_channels	yolo
depthwise_convolutional	region	shortcut	

著作权声明

版权所有©2024 珠海全志科技股份有限公司。保留一切权利。

本文档及内容受著作权法保护，其著作权由珠海全志科技股份有限公司（“全志”）拥有并保留一切权利。

本文档是全志的原创作品和版权财产，未经全志书面许可，任何单位和个人不得擅自摘抄、复制、修改、发表或传播本文档内容的部分或全部，且不得以任何形式传播。

商标声明

、 全志科技、 全志科技、 Allwinner Technology（不完全列举）均为珠海全志科技股份有限公司的商标或者注册商标。在本文档描述的产品中出现的其它商标，产品名称，和服务名称，均由其各自所有人拥有。

免责声明

您购买的产品、服务或特性应受您与珠海全志科技股份有限公司（“全志”）之间签署的商业合同和条款的约束。本文档中描述的全部或部分产品、服务或特性可能不在您所购买或使用的范围内。使用前请认真阅读合同条款和相关说明，并严格遵循本文档的使用说明。您将自行承担任何不当使用行为（包括但不限于如超压，超频，超温使用）造成的不利后果，全志概不负责。

本文档作为使用指导仅供参考。由于产品版本升级或其他原因，本文档内容有可能修改，如有变更，恕不另行通知。全志尽全力在本文档中提供准确的信息，但并不确保内容完全没有错误，因使用本文档而发生损害（包括但不限于间接的、偶然的、特殊的损失）或发生侵犯第三方权利事件，全志概不负责。本文档中的所有陈述、信息和建议并不构成任何明示或暗示的保证或承诺。

本文档未以明示或暗示或其他方式授予全志的任何专利或知识产权。在您实施方案或使用产品的过程中，可能需要获得第三方的权利许可。请您自行向第三方权利人获取相关的许可。全志不承担也不代为支付任何关于获取第三方许可的许可费或版税（专利税）。全志不对您所使用的第三方许可技术做出任何保证、赔偿或承担其他义务。