

# Introduction to sophon-demo

## Introduction

Sophon Demo is developed based on the SophonSDK interface and provides a series of samples for mainstream algorithms. It includes model compilation and quantization based on TPU-NNTC and TPU-MLIR, inference engine porting based on BMRuntime, and pre and post-processing algorithm migration based on BMCV/OpenCV.

SophonSDK is a custom deep learning SDK of Sophon based on its self-developed AI processor, covering model optimization, efficient runtime support, and other capabilities required for the inference phase of neural networks, providing an easy-to-use and efficient full-stack solution for deep learning application development and deployment. It is currently compatible with BM1684/BM1684X/BM1688.

## Directory Structure and Description

contents	category	code	BModel	multi-batch	preprocess
<a href="#">LPRNet</a> (./sample/LPRNet/README.md)	License Plate Recognition	C++/Python	FP32/FP16/INT8	YES	BMCV/OpenCV
<a href="#">ResNet</a> (./sample/ResNet/README.md)	Image classification	C++/Python	FP32/FP16/INT8	YES	BMCV/OpenCV
<a href="#">RetinaFace</a> (./sample/RetinaFace/README.md)	Face detection	C++/Python	FP32	YES	BMCV/OpenCV
<a href="#">segformer</a> (./sample/segformer/README.md)	Semantic segmentation	C++/Python	FP32/FP16	YES	BMCV/OpenCV
<a href="#">yolact</a> (./sample/yolact/README.md)	Instance segmentation	Python	FP32	YES	BMCV/OpenCV
<a href="#">PP-OCR</a> (./sample/PP-OCR/README.md)	OCR	C++/Python	FP32/FP16	YES	BMCV/OpenCV
<a href="#">OpenPose</a> (./sample/OpenPose/README.md)	Keypoint detection	C++/Python	FP32/FP16/INT8	YES	BMCV/OpenCV
<a href="#">C3D</a> (./sample/C3D/README.md)	Video recognition	C++/Python	FP32/FP16/INT8	YES	BMCV/OpenCV
<a href="#">DeepSORT</a> (./sample/DeepSORT/README.md)	Object tracking	C++/Python	FP32/FP16/INT8	YES	BMCV/OpenCV
<a href="#">ByteTrack</a> (./sample/ByteTrack/README.md)	Object tracking	C++/Python	FP32/FP16/INT8	YES	BMCV/OpenCV
<a href="#">CenterNet</a> (./sample/CenterNet/README.md)	Object Detection + pose estimation	C++/Python	FP32/FP16/INT8	YES	BMCV
<a href="#">YOLOv5</a> (./sample/YOLOv5/README.md)	Object Detection	C++/Python	FP32/FP16/INT8	YES	BMCV/OpenCV
<a href="#">YOLOv34</a> (./sample/YOLOv34/README.md)	Object Detection	C++/Python	FP32/INT8	NO	BMCV/OpenCV
<a href="#">YOLOX</a> (./sample/YOLOX/README.md)	Object Detection	C++/Python	FP32/INT8	YES	BMCV/OpenCV
<a href="#">SSD</a> (./sample/SSD/README.md)	Object Detection	C++/Python	FP32/INT8	YES	BMCV/OpenCV
<a href="#">YOLOv7</a> (./sample/YOLOv7/README.md)	Object Detection	C++/Python	FP32/FP16/INT8	YES	BMCV/OpenCV
<a href="#">YOLOv8</a> (./sample/YOLOv8/README.md)	Object Detection	C++/Python	FP32/FP16/INT8	YES	BMCV/OpenCV

contents	category	code	BModel	multi-batch	preprocess
<a href="#">YOLOv5_opt</a> (./sample/YOLOv5_opt/README.md)	Object Detection	C++/Python	FP32/FP16/INT8	YES	BMCV/OpenCV
<a href="#">ppYOLOv3</a> (./sample/ppYOLOv3/README.md)	Object Detection	C++/Python	FP32/FP16/INT8	YES	BMCV/OpenCV
<a href="#">WeNet</a> (./sample/WeNet/README.md)	Speech Recognition	C++/Python	FP32/FP16	YES	-
<a href="#">BERT</a> (./sample/BERT/README.md)	Language	C++/Python	FP32/FP16	YES	-
<a href="#">ChatGLM2</a> (./sample/chatglm2/README.md)	Language	C++/Python	FP16/INT8/INT4	YES	-
<a href="#">Llama2</a> (./sample/Llama2/README.md)	Language	C++	FP16/INT8/INT4	YES	-

## Release Notes

version	description	
0.1.9	Fix documentation and other issues, add segformer/YOLOv7/Llama2, refactor YOLOv34/YOLOv5/ResNet/PP-OCR/DeepSORT/LPRNet/RetinaFace/YOLOv34/WeNet adapt to BM1688, OpenPose postprocess acceleration, chatglm2 support int8/int4 and add compile method in readme.	
0.1.8	Fix documentation and other issues, added BERT/ppYOLOv3/ChatGLM2, refactor YOLOX, added beam search to PP-OCR, added tpu-kernel post-processing acceleration to OpenPose, and updated the SFTP download method.	
0.1.7	Fix documentation and other issues, some demos support BM1684 mlir, refactor PP-OCR/CenterNet, sail support YOLOv5.	
0.1.6	Fix documentation and other issues, add ByteTrack/YOLOv5_opt/WeNet samples.	
0.1.5	Fix documentation and other issues, add DeepSORT sample, refactor ResNet/LPRNet samples.	
0.1.4	Fix documentation and other issues, add C3D and YOLOv8 samples	
0.1.3	Add OpenPose sample, refactor YOLOv5 sample (including adapting arm PCIe, supporting TPU-MLIR to compile BM1684X model, using ffmpeg component to replace opencv decoding, etc.)	
0.1.2	Fix documentation and other issues, refactor SSD related samples, LPRNet/cpp/lprnet_bmcv use ffmpeg component to replace opencv decoding	
0.1.1	Fix documentation and other issues, refactor LPRNet/cpp/lprnet_bmcv with BMNN related classes	Fix documentation and other issues, refactor 0.1.0 LPRNet/cpp/lprnet_bmcv with BMNN related classes.
0.1.0	Provide LPRNet and other 10 samples, adapt BM1684X (x86 PCIe, SoC), BM1684 (x86 PCIe, SoC)	

## Environment dependencies

Sophon Demo mainly depends on tpu-mlir, tpu-nntc, libsophon, sophon-ffmpeg, sophon-opencv, sophon-sail, whose version requirements are as follows:

[sophon-demo|tpu-mlir|tpu-nntc|libsophon|sophon-ffmpeg|sophon-opencv|sophon-sail| Release Date | |-----|-----|-----|-----|-----|

|-----|-----|-----| | 0.1.9 | >=1.2.2 | >=3.1.7 | >=0.4.6 | >=0.6.0 | >=0.6.0 | >=3.7.0 | >=23.10.01| | 0.1.8 | >=1.2.2 | >=3.1.7 | >=0.4.6 | >=0.6.0 | >=0.6.0 | >=3.6.0 | >=23.07.01| | 0.1.7 | >=1.2.2 | >=3.1.7 | >=0.4.6 | >=0.6.0 | >=0.6.0 | >=3.6.0 | >=23.07.01 | | 0.1.6 | >=0.9.9 | >=3.1.7 | >=0.4.6 | >=0.6.0 | >=0.6.0 | >=3.4.0 | >=23.05.01 | | 0.1.5 | >=0.9.9 | >=3.1.7 | >=0.4.6 | >=0.6.0 | >=0.6.0 | >=3.4.0 | >=23.03.01 | | 0.1.4 | >=0.7.1 | >=3.1.5 | >=0.4.4 | >=0.5.1 | >=0.5.1 | >=3.3.0 | >=22.12.01 | | 0.1.3 | >=0.7.1 | >=3.1.5 | >=0.4.4 | >=0.5.1 | >=0.5.1 | >=3.3.0 | - | | 0.1.2 | Not support | >=3.1.4 | >=0.4.3 | >=0.5.0 | >=0.5.0 | >=3.2.0 | - | | 0.1.1 | Not support | >=3.1.3 | >=0.4.2 | >=0.4.0 | >=0.4.0 | >=3.1.0 | - | | 0.1.0 | Not support | >=3.1.3 | >=0.3.0 | >=0.2.4 | >=0.2.4 | >=3.1.0 | - |

**Note:**

1. The version requirements may vary from sample to sample, depending on the README of the routine, and other third-party libraries may need to be installed.
2. BM1688 SDK is not the same as BM1684/BM1684X, it is not released on our official site yet.

## Technical Data

Please get the related documents, materials and video tutorials through [Technical Materials \(https://developer.sophgo.com/site/index.html\)](https://developer.sophgo.com/site/index.html) on the official website of Sophon.

## Community

The Sophon community encourages developers to communicate and learn together. Developers can communicate and learn through the following channels.

Sophon community website: <https://www.sophgo.com/> (<https://www.sophgo.com/>).

Sophon Developer Forum: <https://developer.sophgo.com/forum/index.html> (<https://developer.sophgo.com/forum/index.html>).

## Contribution

Contributions are welcome. For more details, please refer to our [Contributor Wiki \(.CONTRIBUTING\\_EN.md\)](#).

## License

[Apache License 2.0 \(.LICENSE\)](#)