The deep neural network accelerator based on the artificial intelligence processor SPR2801S is used in the field of high-performance edge computing and can be used as visual-based deep learning operation and AI algorithm acceleration. Universal USB interface for more convenient access to a variety of devices.

Features
- Support USB2.0 and 3.0 standard interface communication
- No programming needed, no language barriers
- Open SDK, can be applied to platforms such as X86, ARM, etc.
- Support Android, Linux and other operating systems
- Support VGG, SSD and other neural network models

Specification
- **NPU**
  - Name: Lightspeeur SPR2801S (28nm process, unique MPE and ApM architecture)
  - Energy efficiency: 9.1 TOPs/Watt
  - Peak: 5.6 TOPs@100MHz
  - Low Power: 2.8 TOPs@300mW
  - Hardware interface: SDIO3.0 / eMMC 4.5
  - Package: BGA(7mm*7mm)
  - Manufacturing process: 28nm

- **USB accelerator**
  - Size: 66x19.5x10mm
  - Interface: USB 2.0/3.0 Type-A
  - Transmission Bandwidth: read bandwidth = 68.00 MB/s, write bandwidth = 84.69 MB/s
  - Working Voltage: DC 5V 200mA
  - Operation Temperature: 0°C to 40°C
  - Storage Temperature: -20°C to 80°C
  - Framework: support Pytorch, Caffe framework, follow-up support TensorFlow
  - SDK Provided: ARM, X86 SDK
  - Tools: PLAI model training tool (support for GGG1, GNet18 and GNetfc network models based on VGG-16)
  - Support Ubuntu, Windows operating system

About Lightspeeur® 2801S
Lightspeeur® 2801S is the world’s first commercially available deep learning CNN accelerator chip to run audio and video processing to power AI devices from Edge to Data Center. Lightspeeur® pairs with a host processor to improve AI performance, while significantly reducing energy costs by minimizing host processing and power requirements with no extra memory requirements.

Lightspeeur® 2801S uses 100% proprietary and patented technologies to accelerate CNN processing at extremely high speeds, while consuming very little power.

GTI’s Matrix Processing Engine (MPE™) architecture is a multi-dimensional processing array of physical matrices of digital multiply-and-accumulate (MAC) units that computes the series of matrix operations of a convolutional neural network. The scalable matrix design of the engines allows each engine to directly communicate and interact with adjacent engines, optimizing and
USB Dongle module supports all Firefly development boards, but need the driver module sg.ko (SCSI generic module), please confirm the kernel supported or compile the corresponding kernel driver.