DepthAI: Embedded, Performant Spatial AI and CV

Brandon Gilles
Luxonis
Start with the Why

• Pressing health/safety problem

• Combination 3 available technologies solve the problem:
  • **Spatial sensing** (disparity depth, Lidar)
  • AI (neural inference)
  • CV (feature extraction tracking, platform pose/motion estimation, etc.)

• But only tractable if on an **embedded system** (low size, weight, power, and cost) and performant

• No platform existed that had all 5 of these:
  • Embedded, Performant, Spatial, AI and CV
What is Spatial AI and CV?

- The capability to get neural inference results (e.g. semantic segmentation) in physical space.
  - E.g.: XYZ locations of all the ripe strawberries in real-time.
- Tight fusion of
  - **AI** (object detection, semantic segmentation, etc.)
  - with **CV** (platform motion/pose, object tracking)
  - and **Depth** (disparity)
  - to give **3D position** in world coordinates of detected **objects**, **features**, or **semantic** labels (at the pixel level).
Spatial AI and CV – Basic Example

https://youtu.be/sO1EU5AUq4U
Spatial AI and CV – Fast Motion

OAK-D Neural Inference & Spatial Hard Sync

fps: 30

NO HARD SYNC

https://youtu.be/UQJ6cW7KB34
Traditionally, the capability to handle any of the following necessitated a full-fledged operating system-capable computer:

- High resolution image sensors
- High frame rates
- Multiple image sensors
- Complex AI + CV pipelines
- Spatial sensing

This is now all doable on an embedded system

And this opens up all sorts of applications which were previously intractable
Example: AI-Guided Lossless Zoom

https://youtu.be/H-FjrbWsaKg
Example: AI-Guided Lossless Zoom

https://youtu.be/uyIzG3yLiU
Asymmetric Multi-Processor Embedded System

• Network on Chip (NoC) architecture allows tying together AI/CV/Spatial hardware:
  • 16 Vector processors (think GPU, but architected for computer vision)
  • ~20 fixed-function CV processors (Harris, Canny, warp/dewarp, motion estimation)
  • 2 AI processors
  • 1 semi-global-matching disparity-depth hardware block
• The network on chip is prioritizable, so that it’s selectable which functions drink from the fire-hose of the high-resolution sensor(s)
• Our DepthAI Pipeline Builder allows quick/easy configuration for your application
DepthAI Pipeline Builder (Gen2)

• Node and Graph Based Pipeline

• 3 Node Modalities:
  1. Pre-canned Hardware-accelerated CV/AI/Spatial functionalities
  2. CPython bindings for running scripts directly on DepthAI
  3. OpenCL, ML-Framework-based vectorized math (e.g. [here](#)) for custom hardware-accelerated CV/AI/Spatial functionalities

• The network on chip builds the graph - allowing extremely high data-rate and low-latency connection between the nodes

• The DepthAI resource manager configures the network on chip
Gen2 Pipeline Builder Pre-Canned Nodes

- Neural inference
- 3D object localization
- Object tracking
- Stereo depth
- h.264/h.265 encoding
- Background subtraction
- Feature tracking
- Motion estimation
- Arbitrary crop/rescale/reformat and ROI return (e.g. allowing lossless zoom)
DepthAI Pipeline Builder Example

Input Image

Face detection model
face-detection-retail-0004

Landmark Detection Model
landmarks-regression-retail-0009

Cropped Face

Cropped Left and Right Eye

Cropped Face

Head Pose Estimation Model
head-pose-estimation-adas-0001

Head Pose Angles

Gaze Estimation Model
gaze-estimation-adas-0002

Gaze Coordinates

Camera view

(x=143, y=16) ~ R:71 G:41 B:29

https://github.com/luxonis/depthai-experiments/tree/master/gen2-gaze-estimation#gen2-gaze-estimation
• We built the platform as we, engineers, would want it:
  • Open Source so that it can be autonomously integrated into other codebases
  • Permissively licensed; it can be built into closed-source systems without concern
• DepthAI is Open-Source and MIT-Licensed
  • Hardware
  • Firmware
  • Software
  • ML-Training & Resources
Open-Source Hardware – That You Can Buy

• Can buy all of these directly and us as supported products
• They are also all open-source reference designs, with full Altium Designer source files
Open-Source Hardware
Open-Source Firmware

- Full DepthAI Pipeline Builder Available over SPI Interface, with C++:

```cpp
dai::Pipeline p;

// set up NN node
auto nn1 = p.create<dai::node::NeuralNetwork>();
nn1->setBlobPath(nnPath);

// set up color camera and link to NN node
auto colorCam = p.create<dai::node::ColorCamera>();
colorCam->setPreviewSize(300, 300);
colorCam->setResolution(dai::ColorCameraProperties::SensorResolution::THE_1080_P);
colorCam->setInterleaved(false);
colorCam->setCamId(0);
colorCam->setColorOrder(dai::ColorCameraProperties::ColorOrder::BGR);
colorCam->preview.link(nn1->input);

// set up SPI out node and link to nn1
auto spiOut = p.create<dai::node::SPIOut>();
spiOut->setStreamName("spimetaout");
spiOut->setBusId(0);
nn1->out.link(spiOut->input);

return p;
```

- This means you can no-joke have tinyYOLOv4 running at 30FPS with an ATmega8 “host”.
- ESP32, STM32, MSP430, etc. are commonly used.
- Others are easy to integrate
- microROS (ESP32) example
Open-SOURCE SOFTWARE

• Permissively (MIT-) Licensed so that closed-source products can be built royalty-free.
• We built this how we would want it.

• Python and C++ API parity
• ROS1 and ROS2 Integration
• Unity Plugin
• Wealth of reference pipelines

Usage

```c
#include <dai.h>

int main()
{
    dai::Pipeline pipeline;
    auto mobilenetSpatial = pipeline.create<dai::node::MobileNetSpatialDetectionNetwork>();

    mobilenetSpatial->setBlobPath(nnBlobPath);
    // Will ignore all detections whose confidence is below 50%
    mobilenetSpatial->setConfidenceThreshold(0.5f);
    mobilenetSpatial->input.setBlocking(false);
    // How big the ROI will be (smaller value can provide a more stable reading)
    mobilenetSpatial->setBoundingBoxScaleFactor(0.5f);
    // Min/Max threshold. Values out of range will be set to 0 (invalid)
    mobilenetSpatial->setDepthLowerThreshold(100);
    mobilenetSpatial->setDepthUpperThreshold(5000);

    // Link depth from the StereoDepth node
    stereo->depth.link(mobilenetSpatial->inputDepth);
}
```

© 2021 Luxonis
Open-Source Software
DepthAI Resources

DepthAI Documentation
https://docs.luxonis.com/en/latest/

DepthAI Community Discord
https://discord.gg/EPsZHkg9Nx

Where to buy DepthAI:
• Mouser
• Sparkfun
• OpenCV

2021 Embedded Vision Summit

Luxonis Demos:
• From-Behind Collision Detection for People Who Ride Bikes
• Neural-Inference-Controlled Crop/Zoom and H.265 Encode
• Spatial AI and CV for Human Machine Safety