Summary: RTOS and On-board AI

EEL 4745C: Microprocessor Applications II

Fall 2022

Md Jahidul Islam

Lecture 9

ECE | Florida
Electrical & Computer Engineering

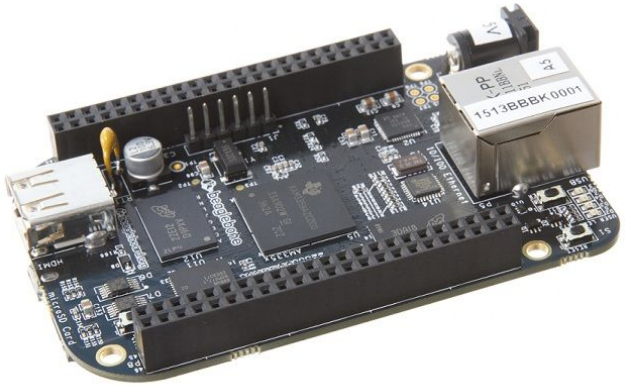
UF | UNIVERSITY of FLORIDA

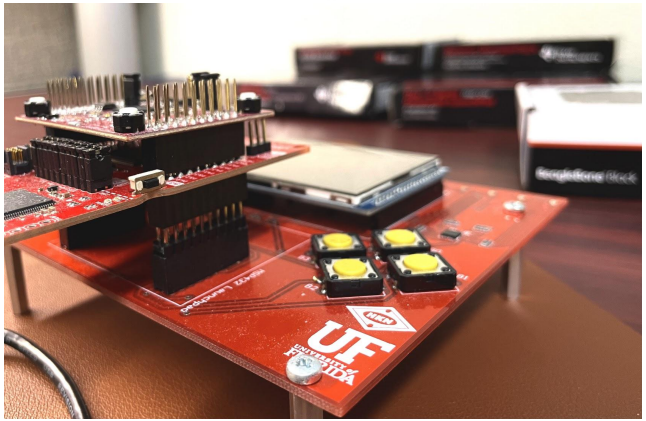
Logistics: Quiz #3 and Project

- **Quiz #3 schedules**
  - Thursday labs: December 1st
  - Monday labs: December 5th
  - Tuesday labs: December 6th
  - Wednesday labs: December 7th

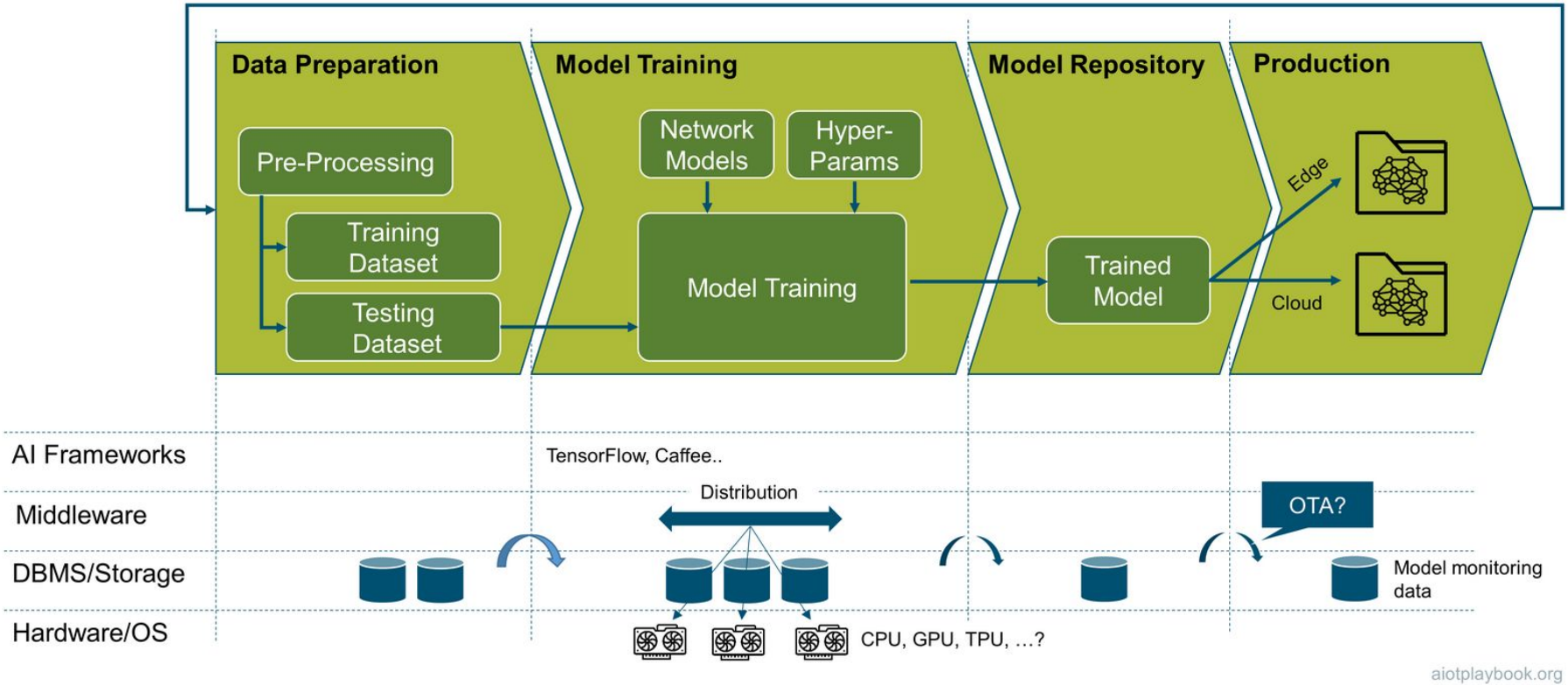
- **Project grading breakdown**
  - RTOS implementation: 70%
  - Completeness: 20%
  - Demo: 10%

- **Project demo**
  - Anytime on/before December 7th
    - In you own labtime or office hours
  - Latest day to demo
    - December 15th 12:30PM - 2:30PM

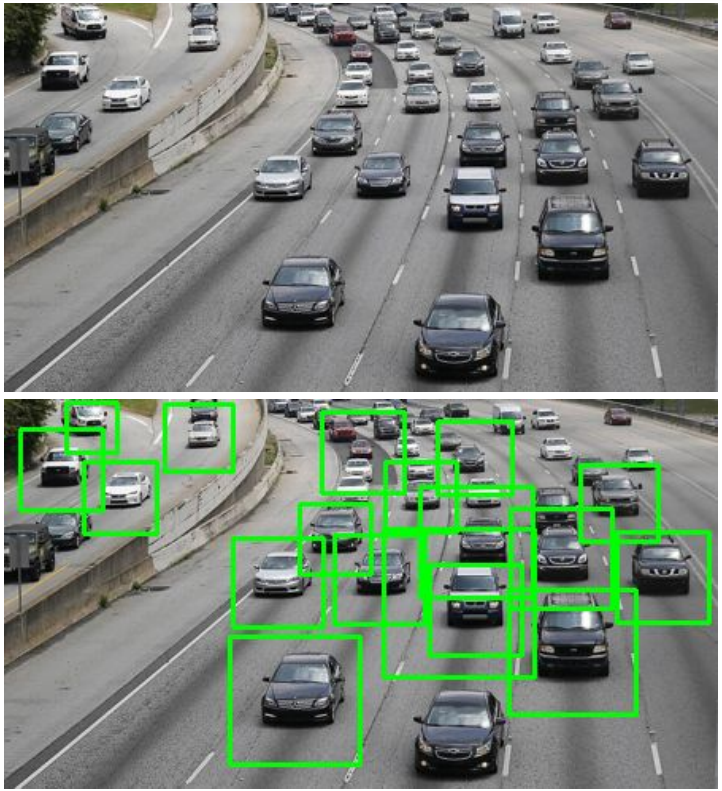




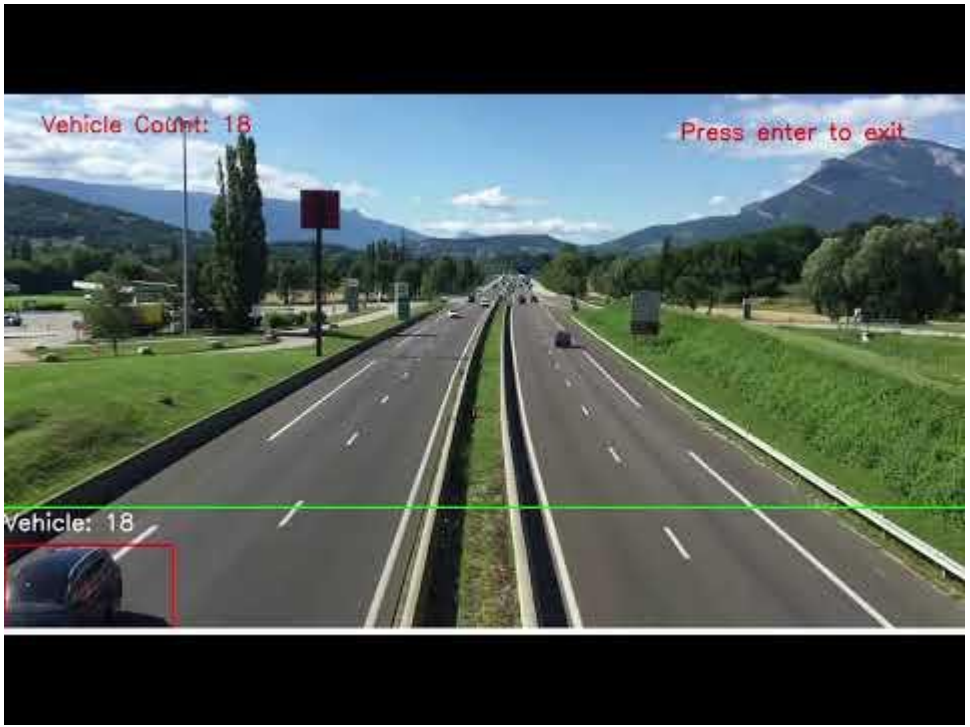
Last Lecture: Edge AIoT Pipeline



Quiz 3 Practice: Vehicle Counting

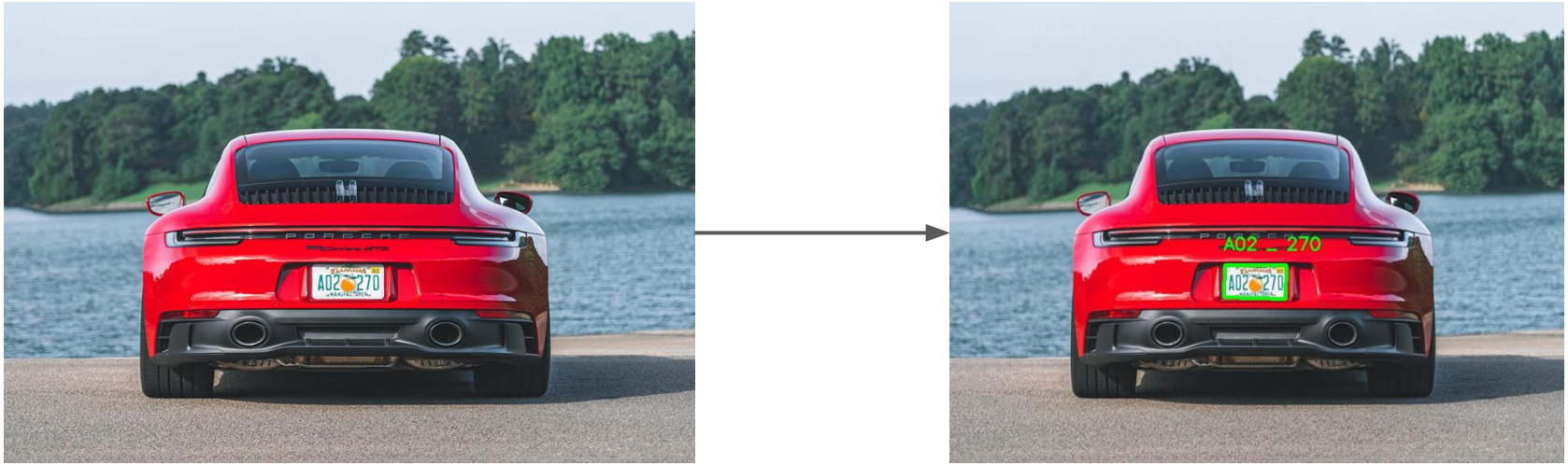


Vehicle Detection by Haar Cascades with OpenCV



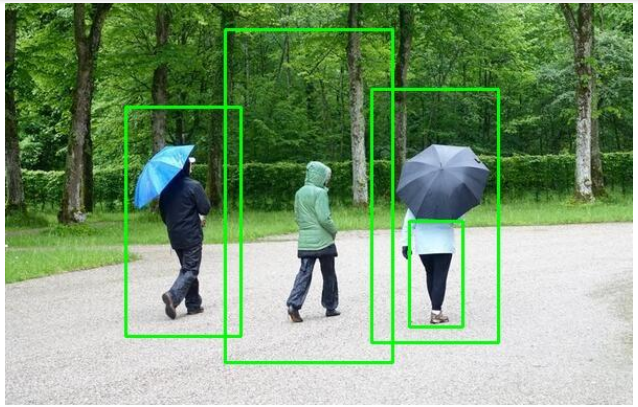
https://youtu.be/ddhWabO6Yvo

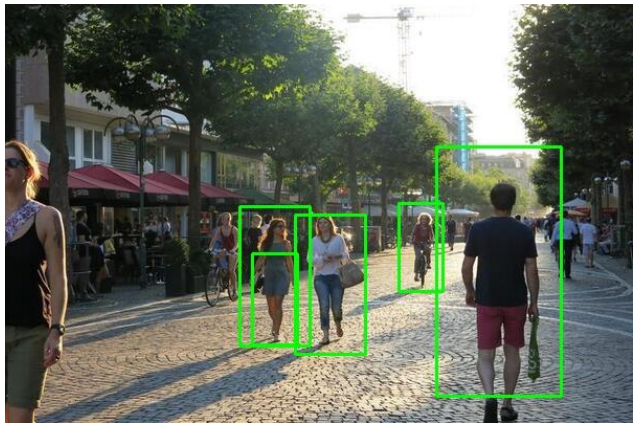
Quiz 3 Practice: License Plate Recognition

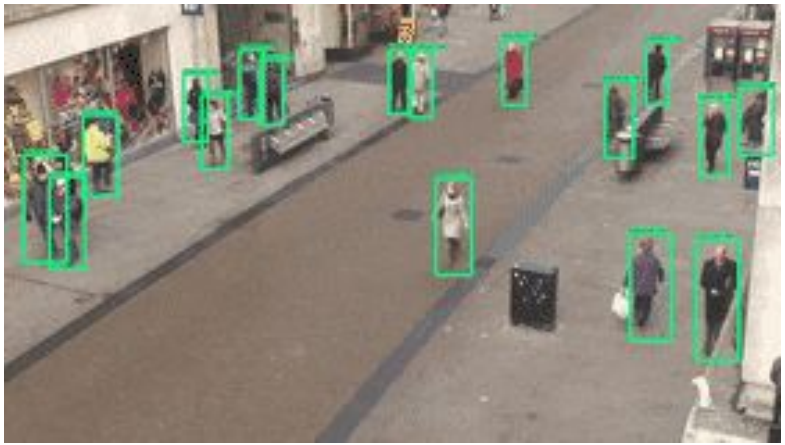


Number Plate Recognition with OpenCV and EasyOCR; see this tutorial:
https://dontrepeatyourself.org/post/number-plate-recognition-with-opencv-and-easyocr/

Quiz 3 Practice: Pedestrian Detection



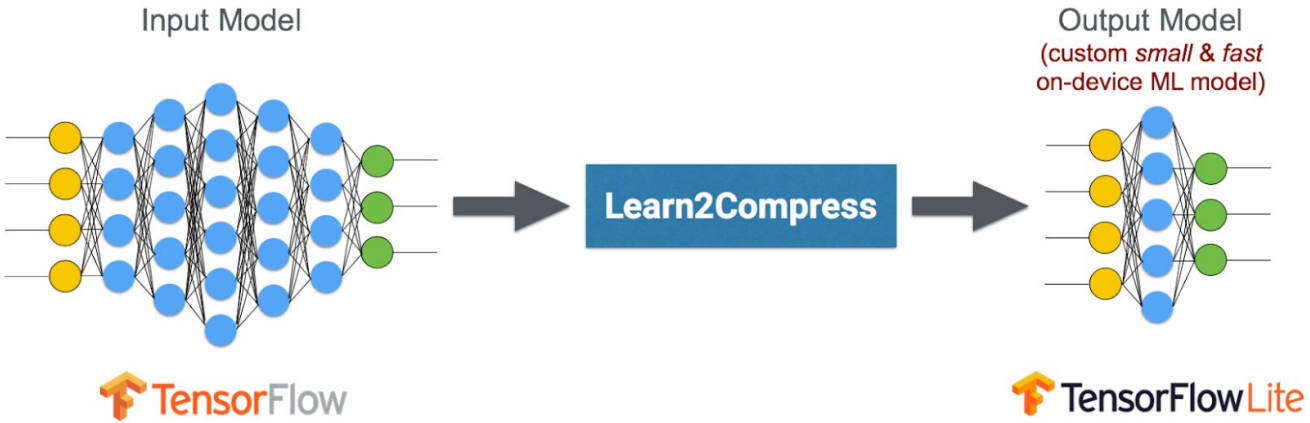




Human (pedestrian) Detection/ counting in OpenCV

- Haar cascade classifiers
- Histogram of Oriented Gradients (HOG)

Running AI Models on RTOS

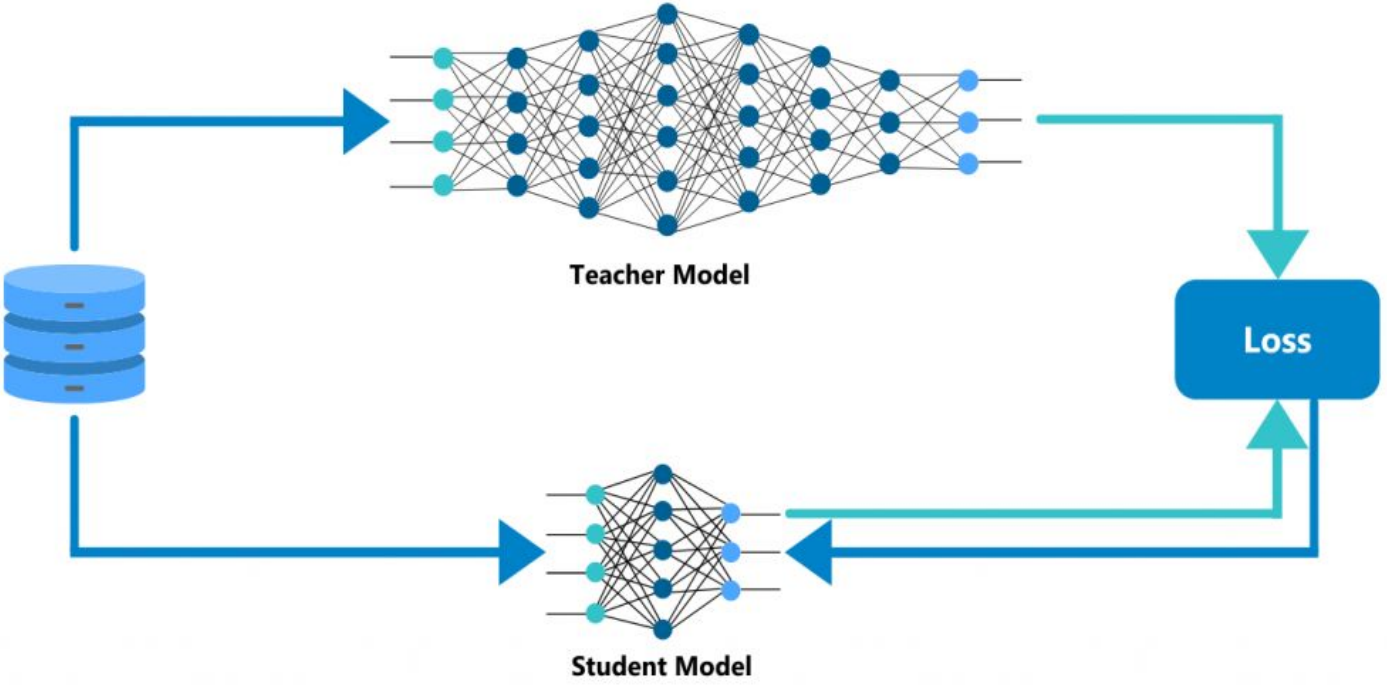


ML/DL model optimization steps

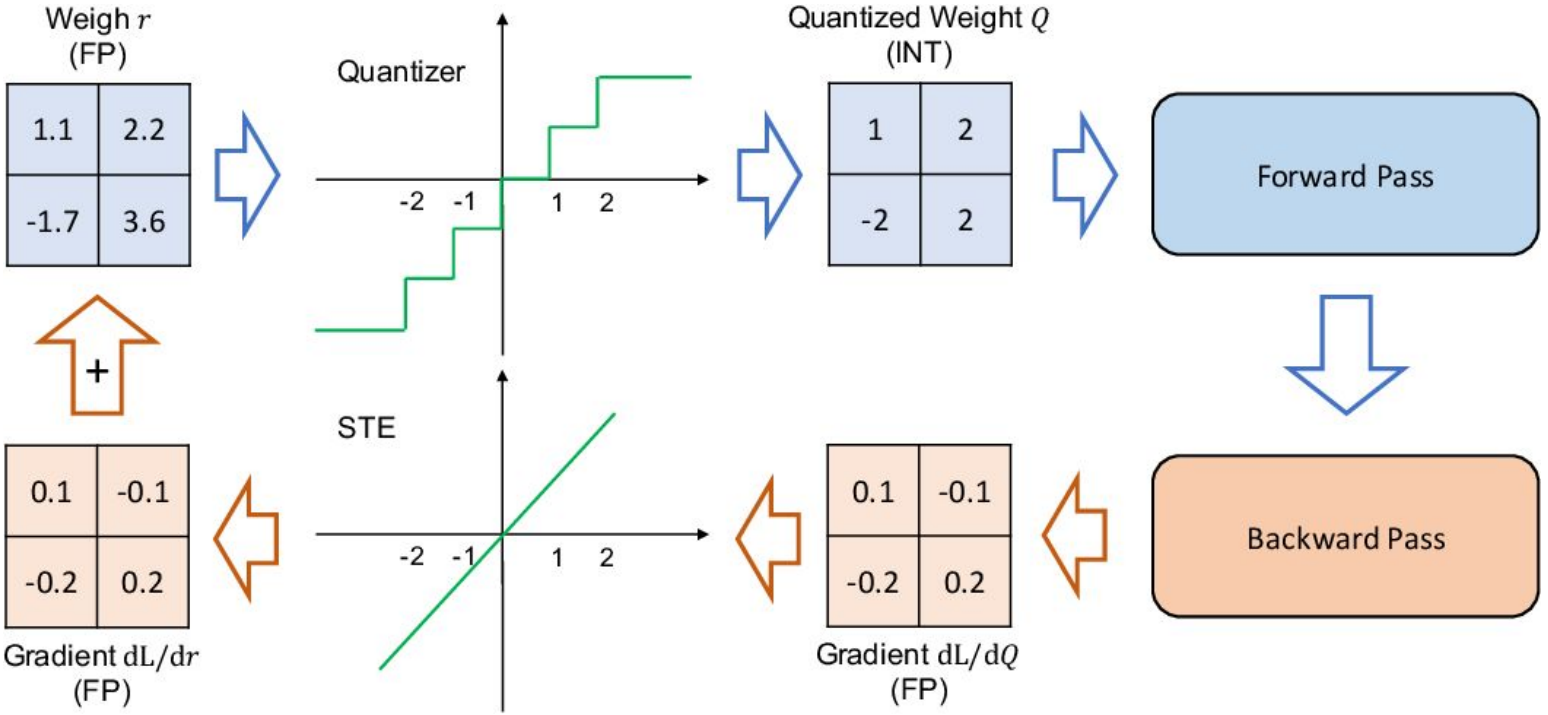
- Model compression and quantization
- Layer fusion, pruning
- Constant folding, matrix factorization

See https://ai.googleblog.com/2018/05/custom-on-device-ml-models.html

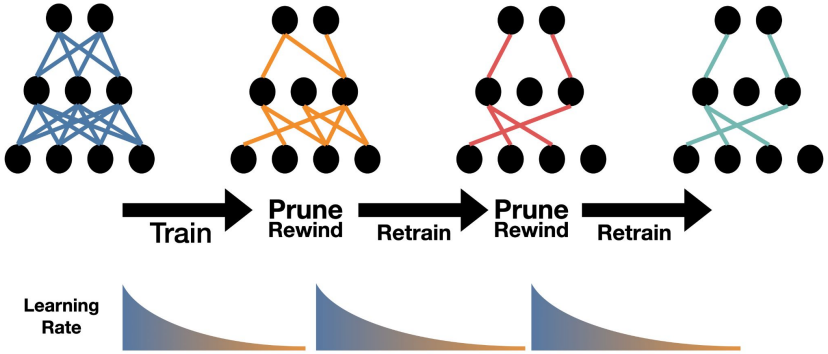
The Teacher-Student Model

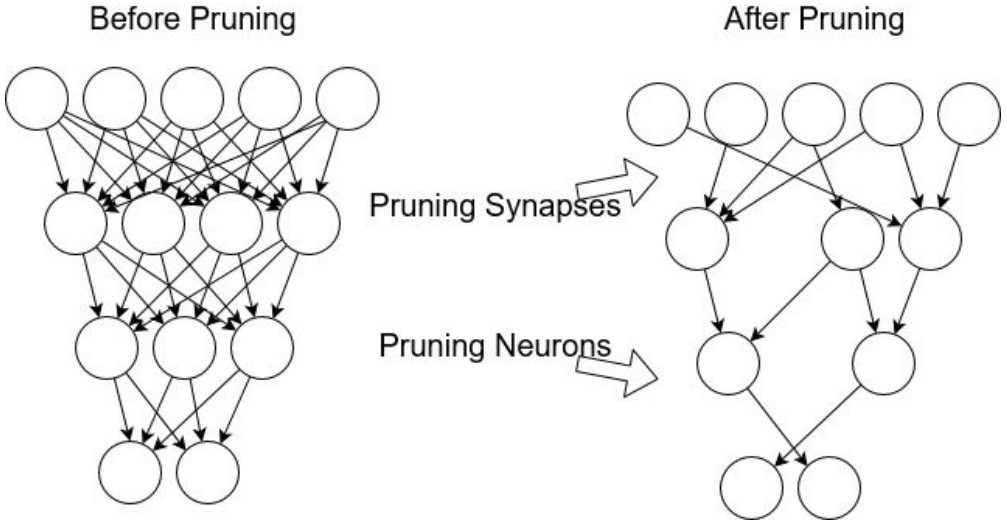


DNN Quantization



Model Compression

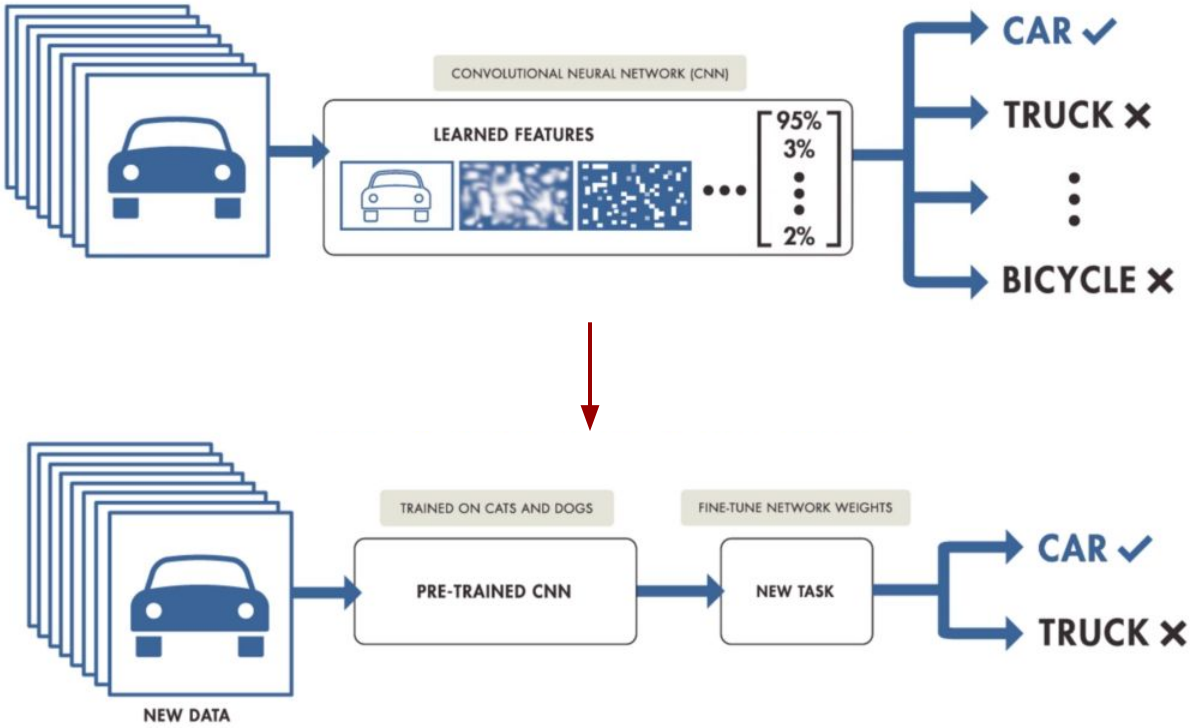




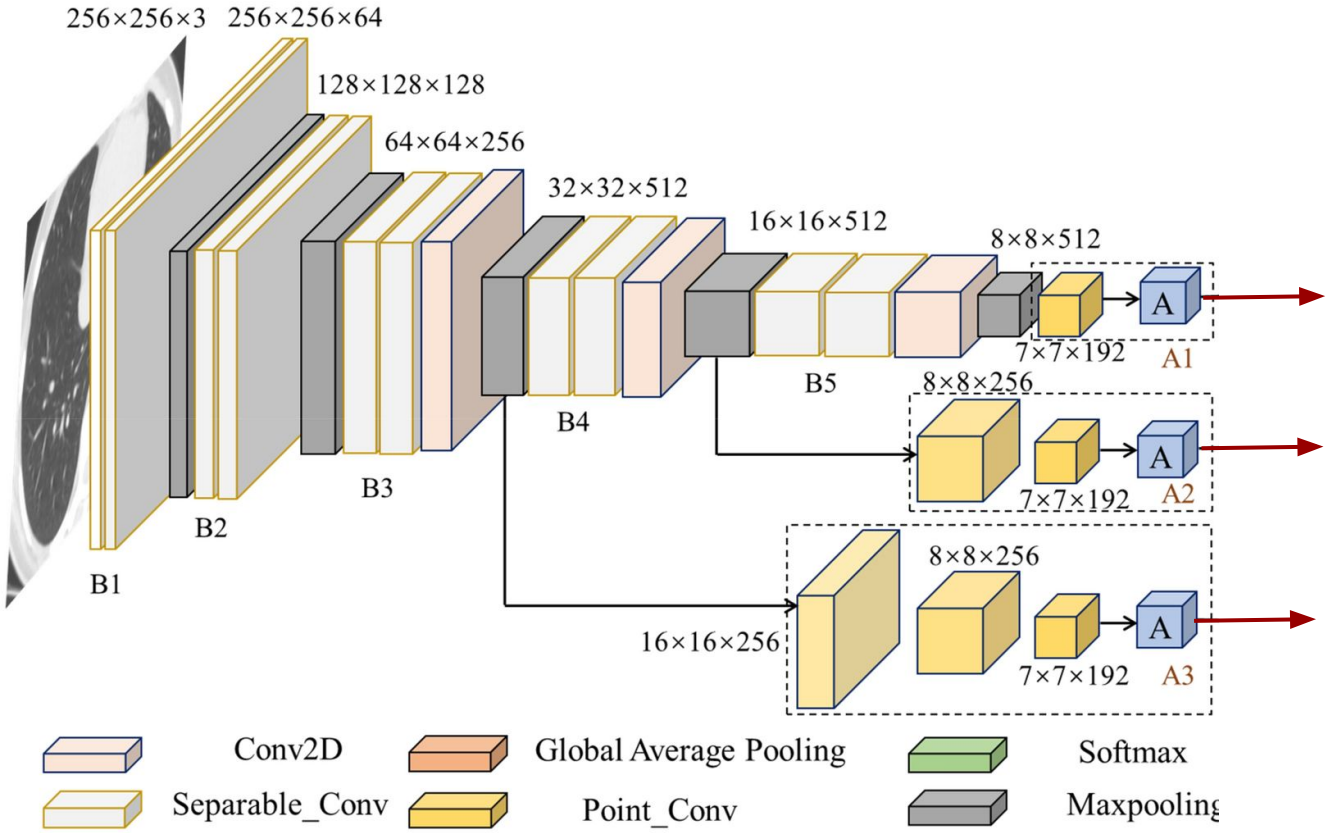
Basic model compression techniques (deep neural nets)

- Prune weights (remove connections)
- Remove nodes (neurons)
- Pruning layers

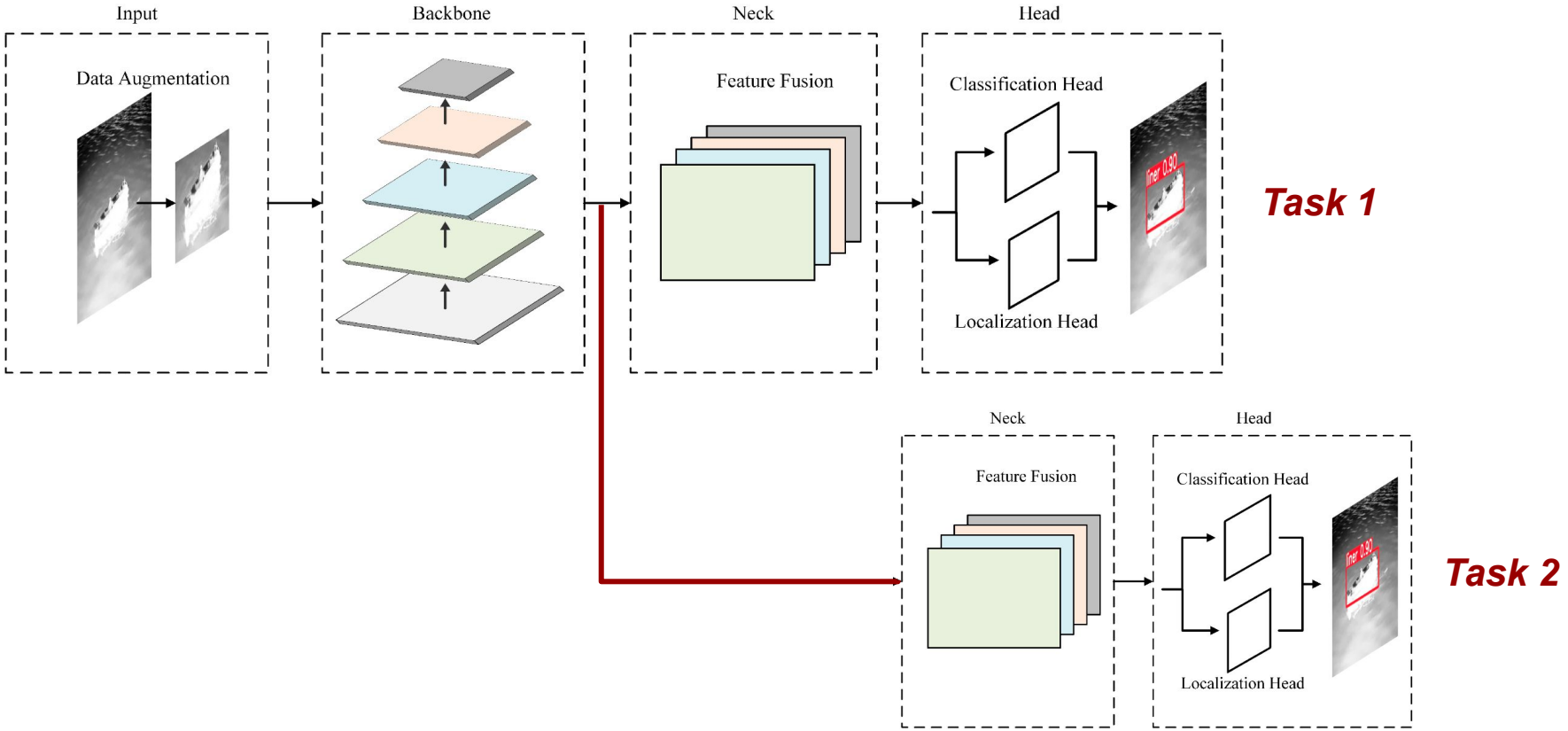
Transfer Learning



Reusing Feature Layers



Using A Single Backbone!

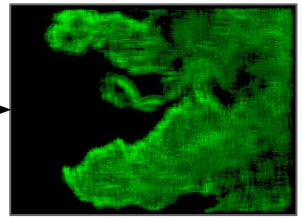


Example #1: SESR

X: Input

Deep SESR
G: $\{X\} \rightarrow \{S, E, Y\}$

S: Saliency



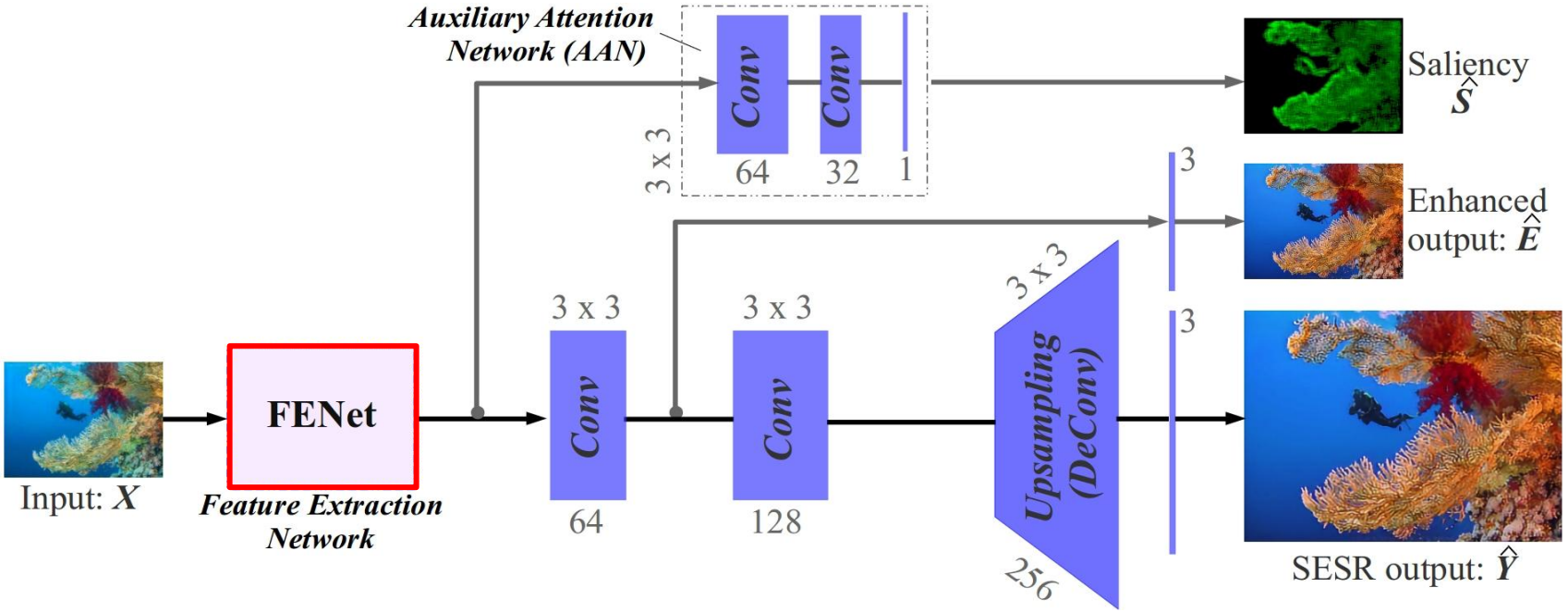
Y: SESR Output (HR)

E: Enhancement

SESR: *Simultaneous Enhancement & Super-Resolution*

See https://github.com/xahidbuffon/Deep_SESR

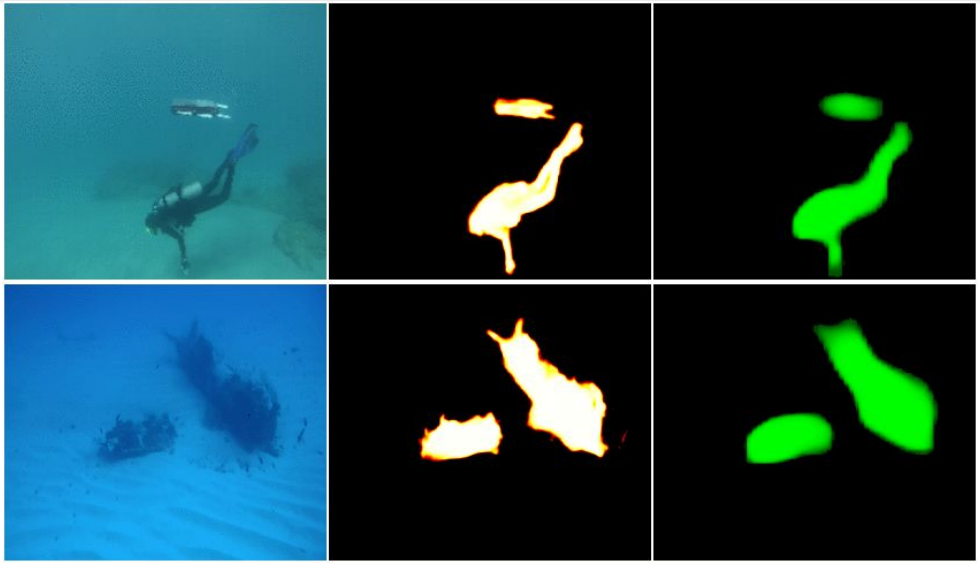
Deep SESR: Network Architecture



SESR: *Simultaneous Enhancement and Super-Resolution*

See https://github.com/xahidbuffon/Deep_SESR

Example #2: SVAM

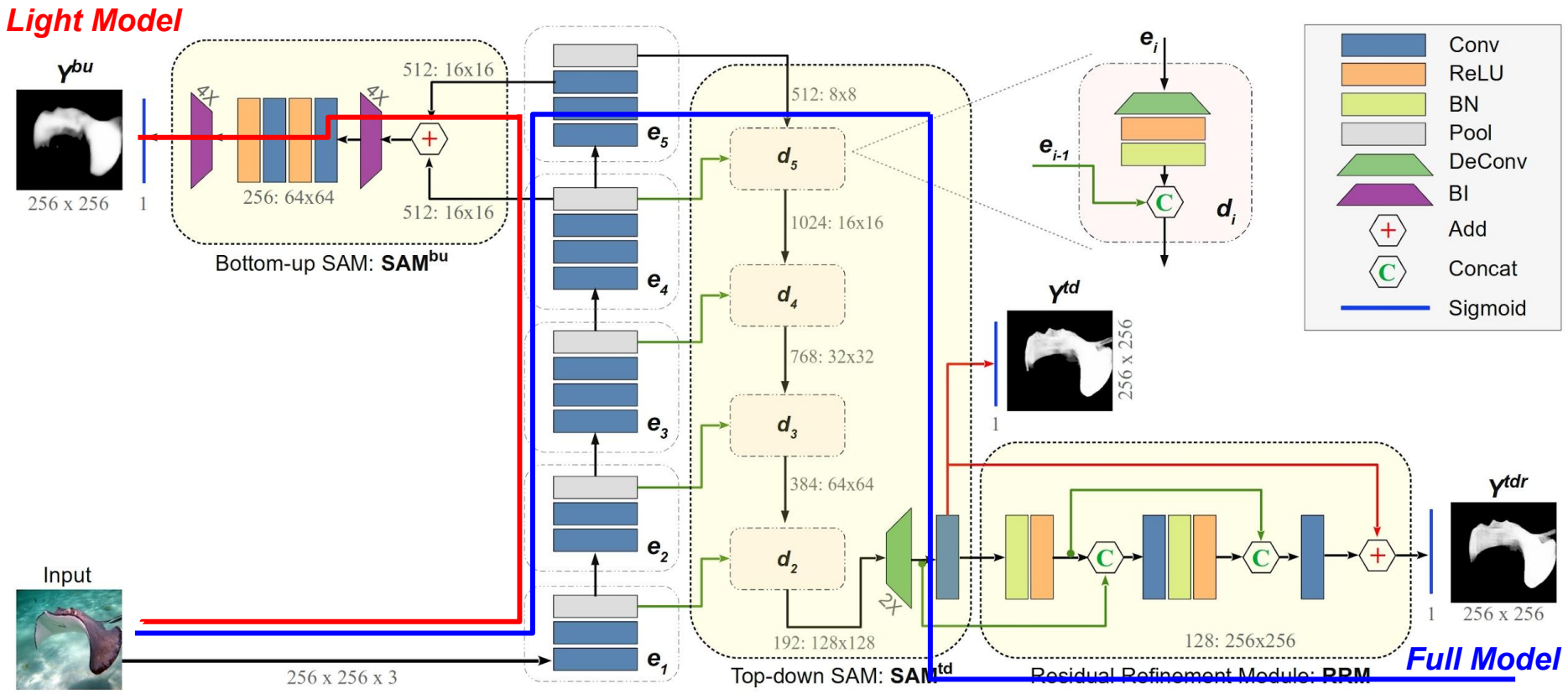


Input | Output Heatmap and Saliency

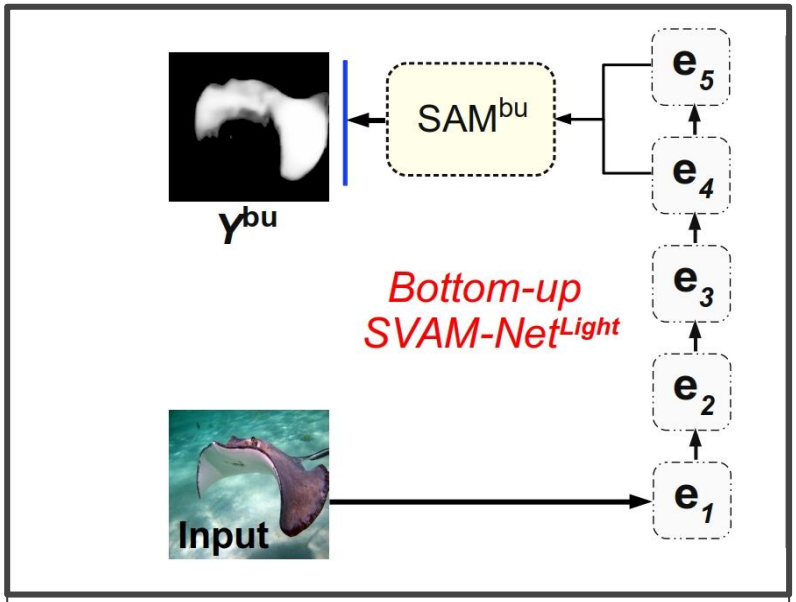
SVAM: *Saliency-guided Visual Attention Modeling By Underwater robots*

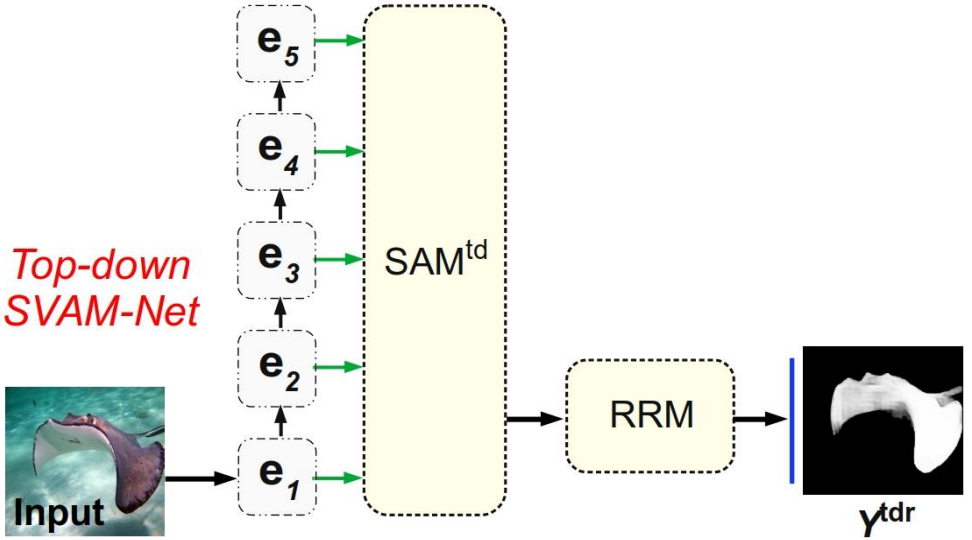
See https://github.com/xahidbuffon/SVAM-Net

Architecture: SVAM-Net



Light Inference: SVAM-Net





⇒ **Bottom-up SVAM-Net^Light**

- Abstract yet accurate estimation
- 86+ FPS on GTX 1080
- 21+ FPS on Jetson Xavier

⇒ **Top-Down SVAM-Net** (full model)

- Fine-grained saliency estimation
- SOTA performance and generalizability
- Fast GPU run-time: 20+ FPS on GTX 1080

uP2 Course Summary (0/6)

⇒ **Basic concepts of RTOS and ARM Cortex M4 processors**

⇒ **Programming RTOS components**: TI Tiva C Series LaunchPad

- Implementing threads and schedulers
- Handling inter-process communication and semaphores
- Interfacing driver libraries for integrated peripherals
  - I2C RGB LEDs drivers, joysticks, LCD touchscreen display
- Mastering a IoT development board (modified daughter-board)
- Interfacing external IoT sensor boards
  - TI sensor booster pack

⇒ **Creating on-device IoT applications**

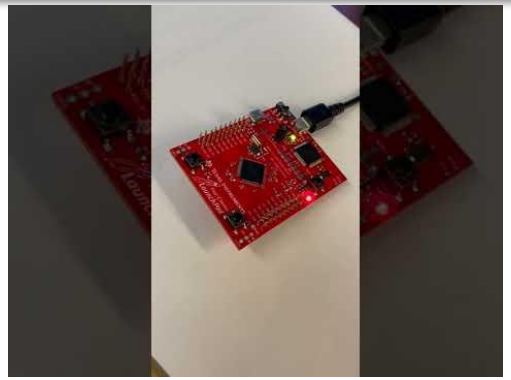
- Interfacing TI BeagleBone Black Board
- Implementing an on-device AI and TinyML applications

uP2 Course Summary (1/6)

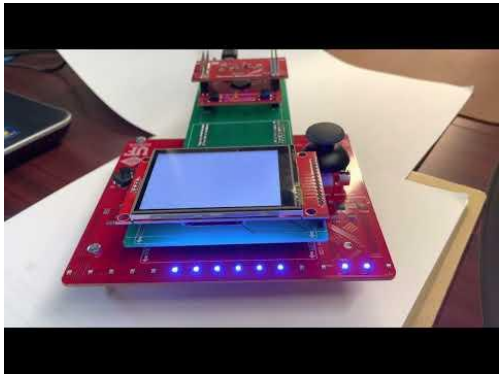
- ARM Architecture and OS overview
  - Instruction set and memory model
  - Board support packages; LED driver interfacing
  - Thumb2 instruction sets; unified assembly programming
  - I2C and UART communication

- **Lab 0:** board setup
  - Part A: Introduction and setup
  - Part B: Blinking TIVA C on-board LEDs

- **Lab1:** Basic Interfacing, Linking, and Communication
  - Part A: Interfacing LED drivers, I2C communication
  - Part B: ARM assembly checksums with the LED driver
  - Part C: Basic UART with LED driver and console I/O



https://youtu.be/dRuUlkuzKnQ

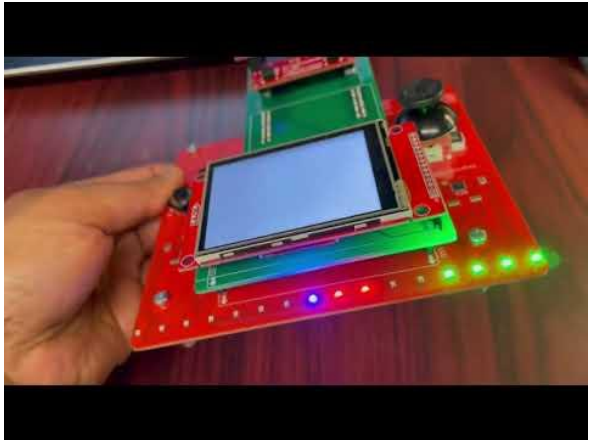


https://youtu.be/2SE7oak54CE

uP2 Course Summary (2/6)

- RTOS Components
  - Threads, interrupts, and schedulers
  - Locks and semaphores
  - Avoiding deadlocks
  - Yielding, blocking, sleeping, etc.
  - Periodic and dynamic threads

- **Lab 2:** G8RTOS Scheduler and Synchronizers
  - Part A: Setting up support packages, drivers, & OS structure
  - Part B: Implementing threads, exception handlers & schedulers
  - Part C: Implementing semaphores & peripheral controls
  - Part D: Adding threads to control LEDs via sensor feedback
  - Part E: Putting it all together!



https://youtu.be/CI1QX1Pv1lg

uP2 Course Summary (3/6)

- Background: OS and RTOS
  - Process and Threads: PCB and TCB
  - Advanced IPC: Inter-Process Communication
    - Various algorithms and process diagrams
  - Scheduling and synchronization
    - Various algorithms their CPU usage
  - Periodic and dynamic threads

- **Lab3:** Periodic Threads and Queueing
  - Part A: Implementing Blocking, yielding, and sleeping
  - Part B: Integrating periodic threads with background threads
  - Part C: Enabling inter-process communication with FIFOs

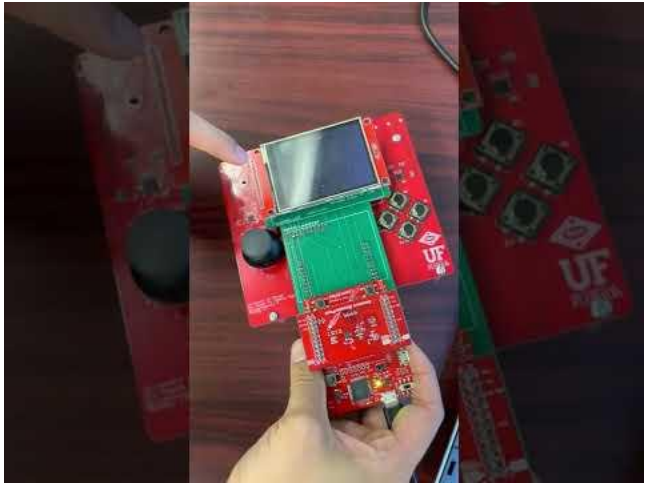
https://youtu.be/kaNBNzkLnko

uP2 Course Summary (4/6)

- Advanced RTOS Concepts
  - More on dynamic and periodic threads
  - Inter-process communication
  - Thread priority: FIFO, round-robin
  - LCD touch display drivers; content of lab 4
  - Aperiodic Event Threads

- Networking Basics
  - OSI model: all 7 layers
  - IPv4/IPv6 and TCP/UDP

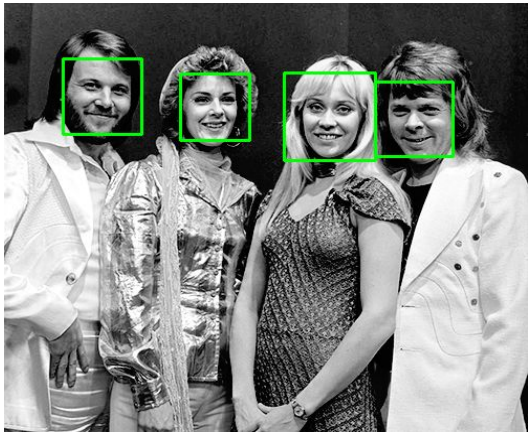
- **Lab 4:** Dynamic Threads and LCD Interfacing
  - Part A: Interfacing a touchscreen color LCD
  - Part B: Incorporating aperiodic/dynamic threads in RTOS



https://youtu.be/umWUxbx3qZc

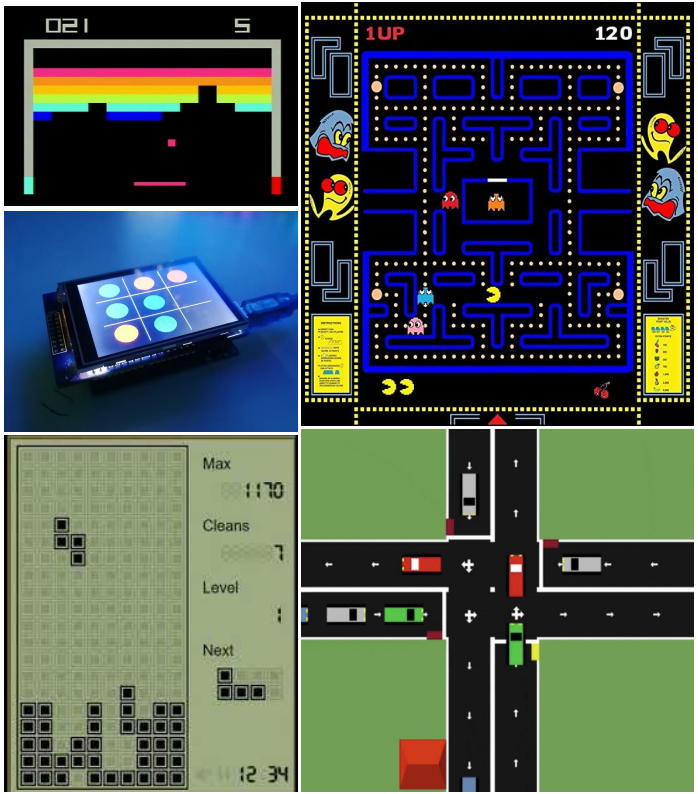
uP2 Course Summary (5/6)

- Real-time on-device AI/AIoT Topics
  - Embedded AI and on-device ML/vision concepts
  - Running AI inference models on Beagle boards
- Sample projects and implementation do/donts
  - System design and RTOS integration
- Integrating RTOS and AIOT
  - ML / DL concepts
  - NN / DNN / CNN concepts
- **Lab 5:** Incorporating AIoT with RTOS
  - Part A: Interfacing and communication with a BeagleBoard
  - Part B: Remote process handling



uP2 Course Summary (6/6)

- Edge AI/AIoT Pipeline
  - Business model and customer viewpoint
  - AIoT sketch and implementation viewpoint
- ML/DL model optimization
  - Model compression and quantization
  - Layer fusion, pruning
  - Constant folding, matrix factorization
- On-device AI / machine vision examples
  - Handwritten zip code recognition
  - Cascade classifiers: face detection
- **Project Options**
  - System design and RTOS integration
  - Full functional game



Thank you!