



MARCH 25—29, 2019 FLORENCE, ITALY FIRENZE FIERA

**DESIGN, AUTOMATION AND TEST IN EUROPE** THE EUROPEAN EVENT FOR
ELECTRONIC SYSTEM DESIGN & TEST



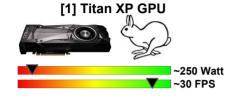


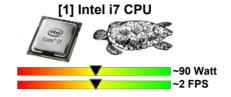
# Energy-Efficient Monocular Depth Estimation on ARM-based Embedded Platforms

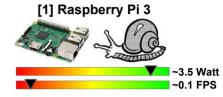
Valentino Peluso, Antonio Cipolletta, Andrea Calimera, Matteo Poggi, Fabio Tosi and Stefano Mattoccia

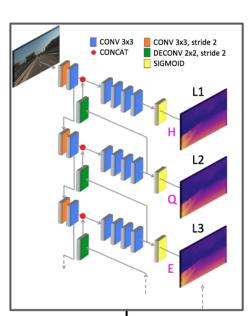
**Monocular depth estimation** is an appealing technique to estimate dense depth maps leveraging unconstrained imaging sensors. State-of-the-art technique [1] deploys energy-hungry deep networks.











## **EQ-Scalable PyD-Net**

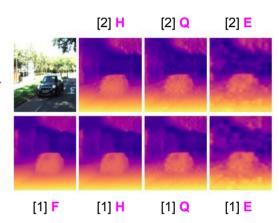
NOW...

#### **Architecture**

**BEFORE** 

Whereas state-of-the-art models [1] count millions of parameters, have large memory footprints and are far from real-time computation on low powered devices, PyD-Net [2] is compact (1.9M vs more than 30M params) and runs at around 1 FPS on Raspberry Pi 3 with comparable accuracy.

Moreover, PyDNet is an energy-scalable architecture with better performance than more complex models like [1].



F: Full resolution
Q: Quarter resolution

H: Half resolution

E: Eighth resolution

#### front-end back-end Q.Neural-Kernels Trained | Neural.Net Quantizer a Training Emulator Compiler Data-Set TensorFlow Computing Library **PvTorch** GPGPU by ARM **PyTorch** FP32 INT16 INT8 Speed [FPS]

A <u>sensing technology</u> [3] with such ability to implement accuracy-energy scaling represents a practical option for adaptive embedded systems: contexts or applications which tolerate lower accuracy might pursue higher energy efficiency by tuning resolution and precision.

### **Energy-Quality Scaling**

- <u>Coarse-Gain Knob:</u> PyD-Net infers disparity maps at different resolutions. Its reconfigurable architecture enables to scale energy with output resolution.

- <u>Fine-Grain Knob:</u> in-house neural kernels enables to scale energy with precision. The optimization flow sketched on the left guarantees marginal loss in output quality at lower bit-widths.



FP32 INT16 INT8



Energy Efficiency [J / Frame]

8.0

7.5

7.5

6.5

6.0

5.5

1 2 3 4 5 6 7

- [2] Poggi et al., "Towards real-time unsupervised monocular depth estimation on CPU", IROS 2018
- [3] Peluso et al., "Energy-Efficient Monocular Depth Estimation on ARM-based Embedded Platforms", DATE 2019

<sup>[1]</sup> Godard et al., "Unsupervised Monocular Depth Estimation with Left-Right Consistency", CVPR 2017