FastDepth: Fast Monocular Depth Estimation on Embedded Systems

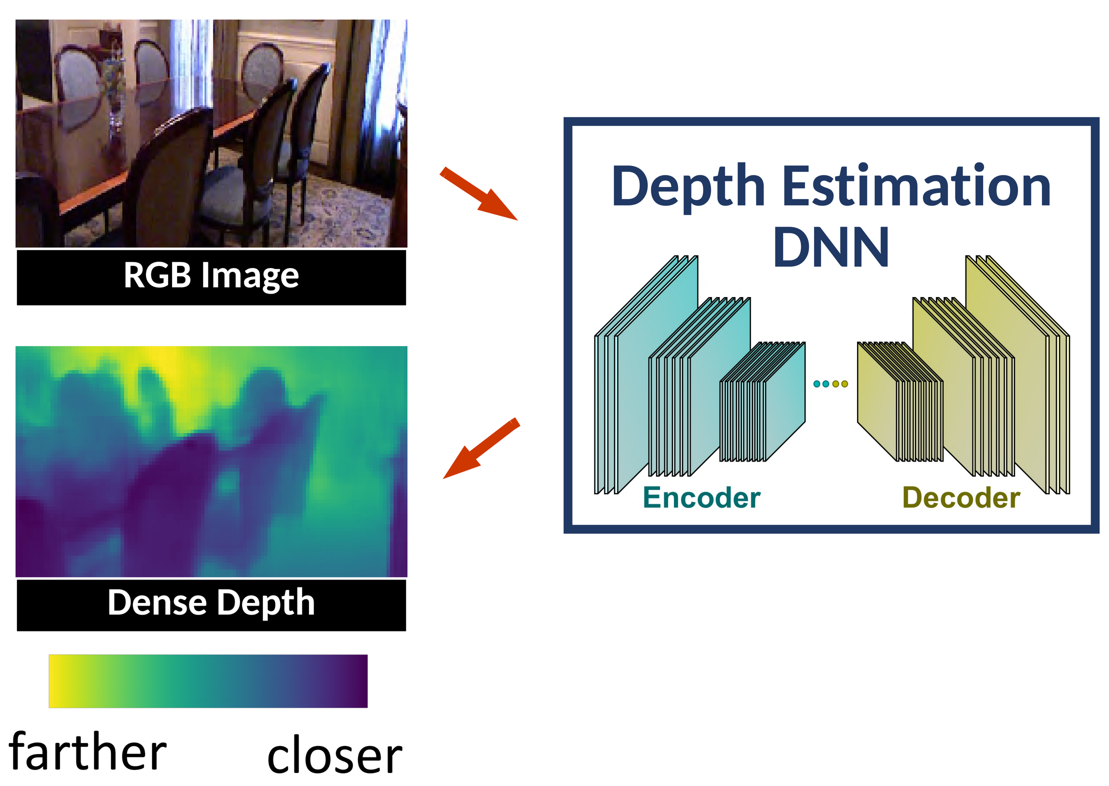
Massachusetts Institute of Technology, USA

Diana Wofk*, Fangchang Ma*, Tien-Ju Yang, Sertac Karaman, Vivienne Sze

Project Webpage

1. Motivation

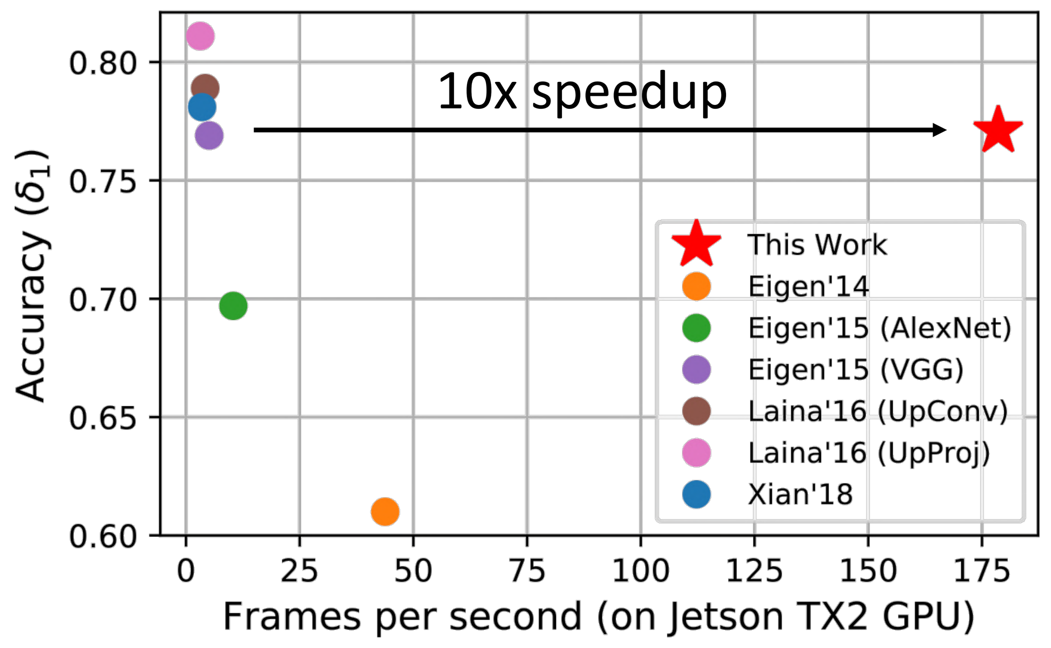
Real-time low-power depth sensing is critical for successful navigation of small robotic vehicles.



Existing Work:
- ✓ High Accuracy
- ✗ High Complexity
- ✗ High Latency
- ✗ High Energy Cost

Our Work:
- ✓ High Accuracy
- ✓ Low Complexity
- ✓ Low Latency
- ✓ Low Energy Cost

2. Contribution



Our work is ***an order of magnitude faster*** than prior work, with ***near state-of-the-art accuracy*** on the NYU Depth v2 dataset.

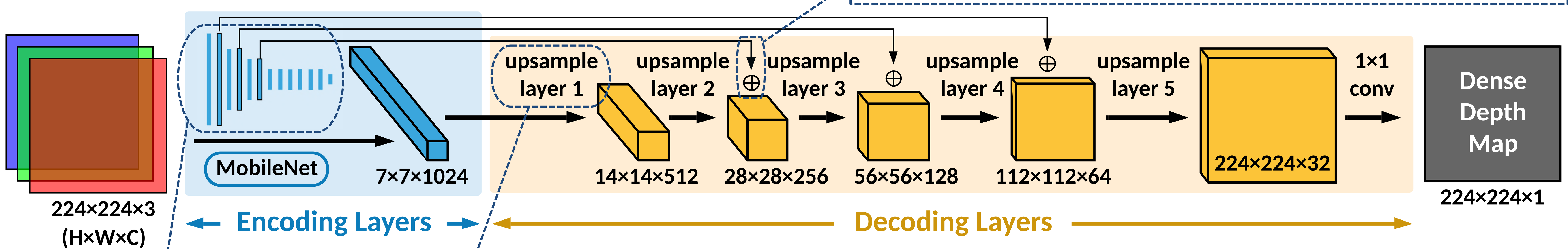
Model Metrics	
# weights [10^6]	1.34
# MACs [10^9]	0.37
Accuracy (δ_1 [%])	77.1
RMSE [cm]	60.4

Jetson	Runtime	Power
TX2 GPU	5.6 ms (178 fps)	12.2 W (3.4 W idle)
TX2 CPU	37 ms (27 fps)	10.5 W (3.4 W idle)

3. Methodology

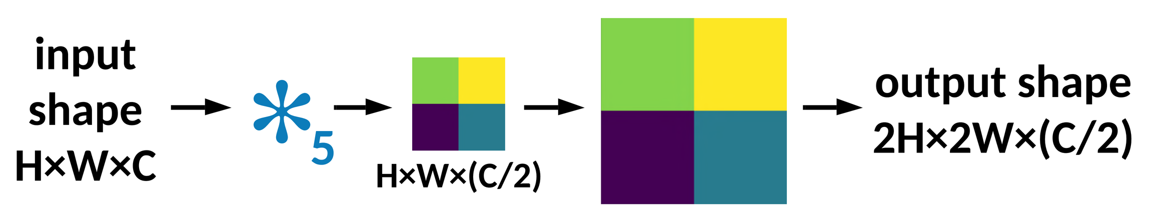
► We use a ***lightweight encoder-decoder architecture:***

Additive skip connections result in improved accuracy.



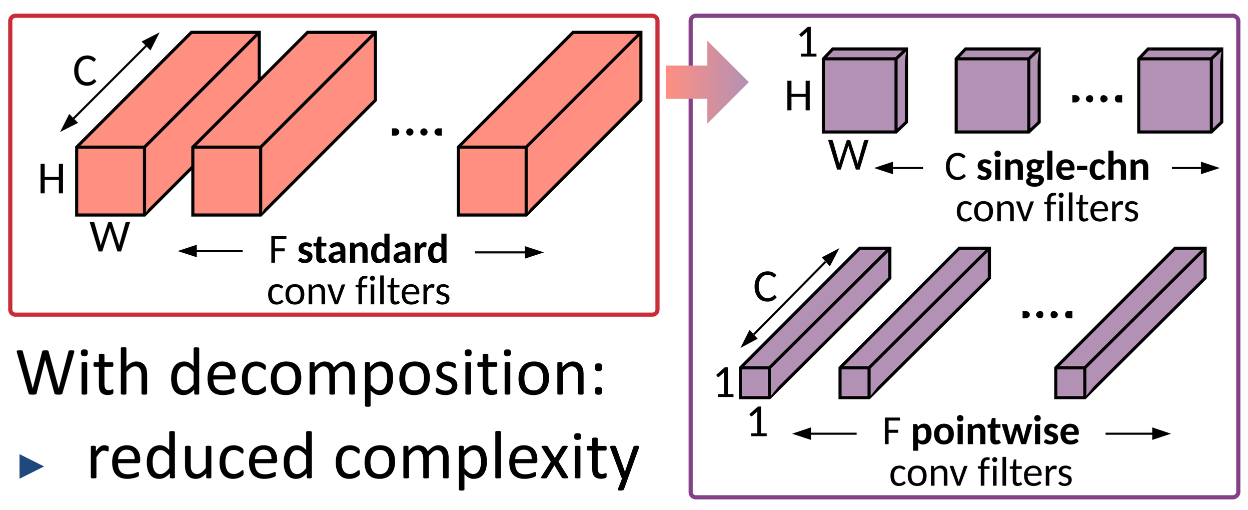
Our encoder is based on a ***computationally-simple and efficient image classification network***.

Each upsample layer performs ***5x5 depthwise separable convolution*** with ***nearest-neighbors interpolation***.



Depthwise Separable Convolution

Used in both the encoder and decoder.



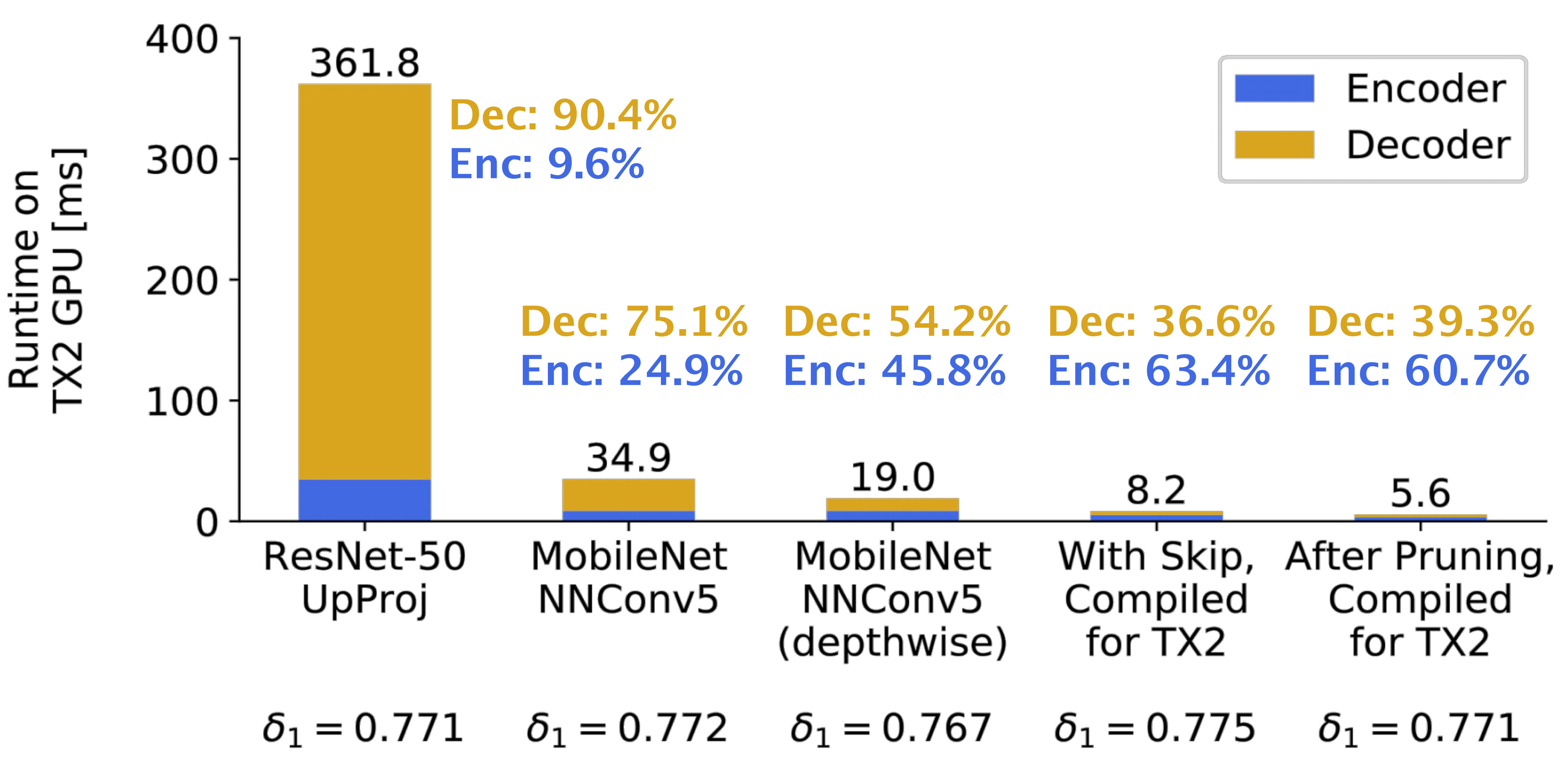
With decomposition:
- reduced complexity
- reduced runtime

► We perform ***hardware-specific compilation*** (TVM) to lower inference runtime on the target embedded platform.

► We apply ***resource-aware network adaptation*** (NetAdapt) to automatically simplify our model to further lower runtime.

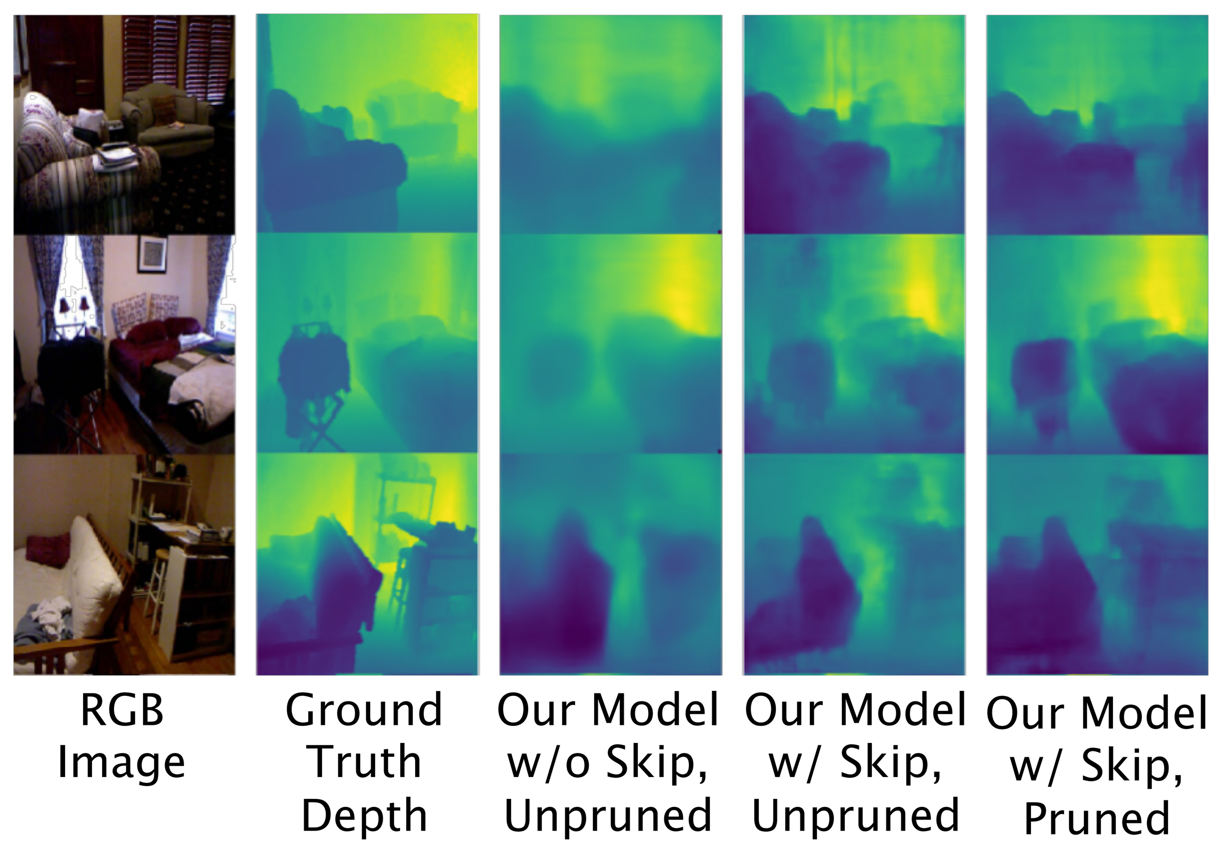
4. Experiments

Runtime reduction achieved with our methodology:

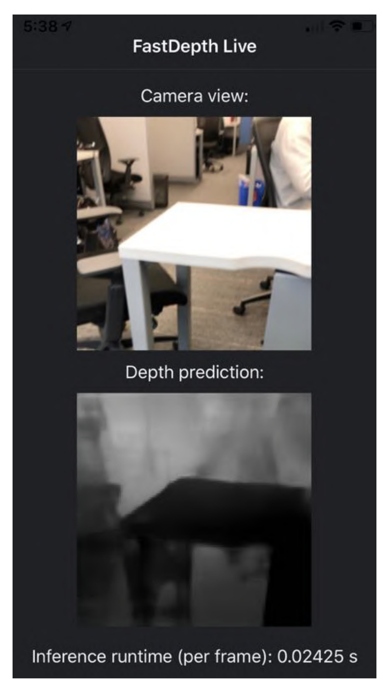


5. Visualization & Phone Demo

Visualized results on NYU Depth v2:



Live Demo:



40 fps on iPhone X