The 40th Annual AAAI Conference on Artificial Intelligence
JANUARY 20 – JANUARY 27, 2026 | SINGAPORE

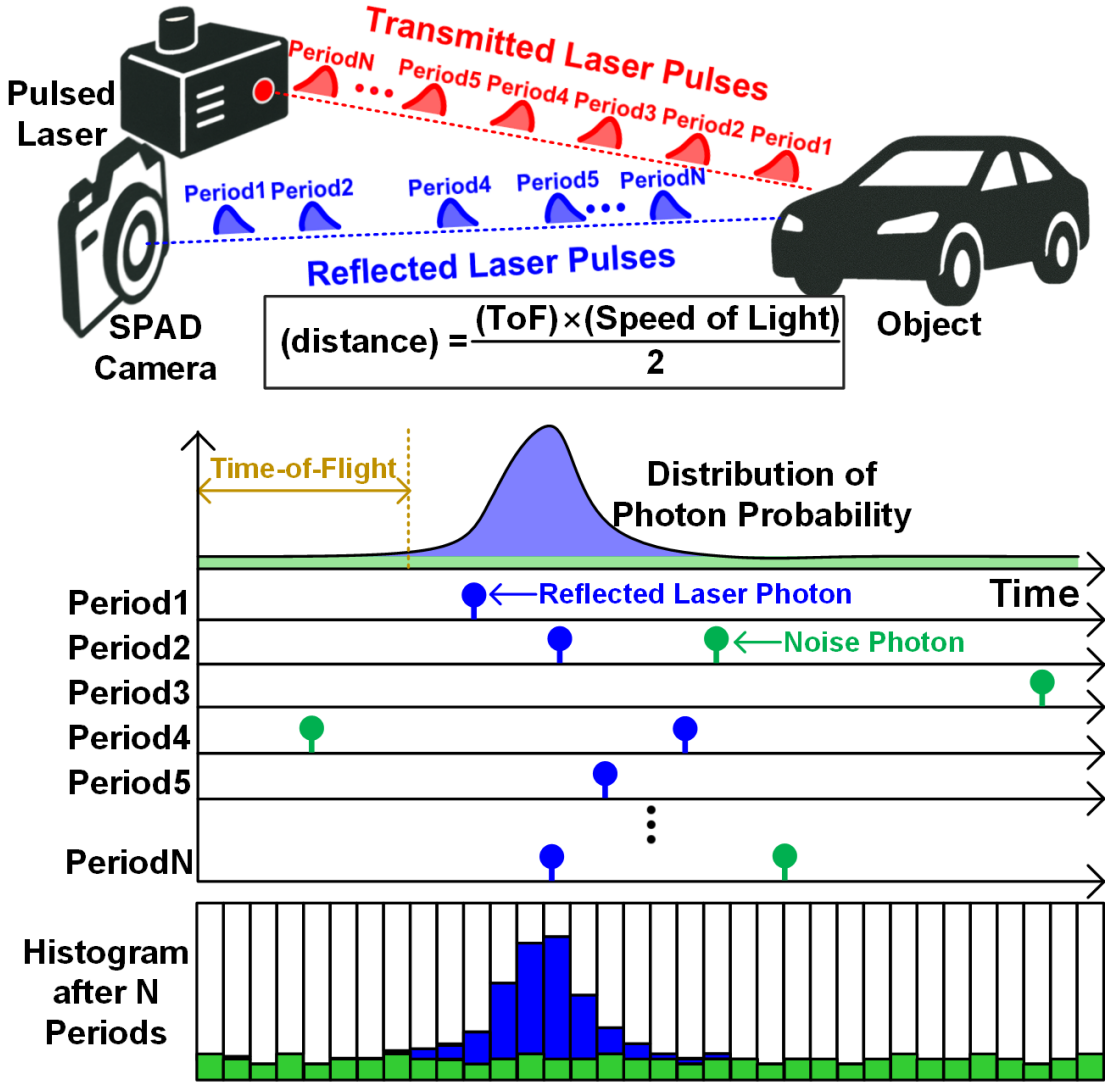
A Paradigm Shift in High-Resolution Depth Estimation:
From Signal Filtering to Lightweight Similarity Learning

Minsung Lee[1*], Seo Hyun Kim[1*], Yeonsu Park[2†], Hyeongseok Seo[3†], Jongmin Lee[1†]
[1]Ajou University, Suwon, Korea, [2]Kangwon National University, Chuncheon, Korea, [3]Pusan National University, Busan, Korea
[*]These authors contributed equally.
[†]Corresponding authors.

What is dToF?

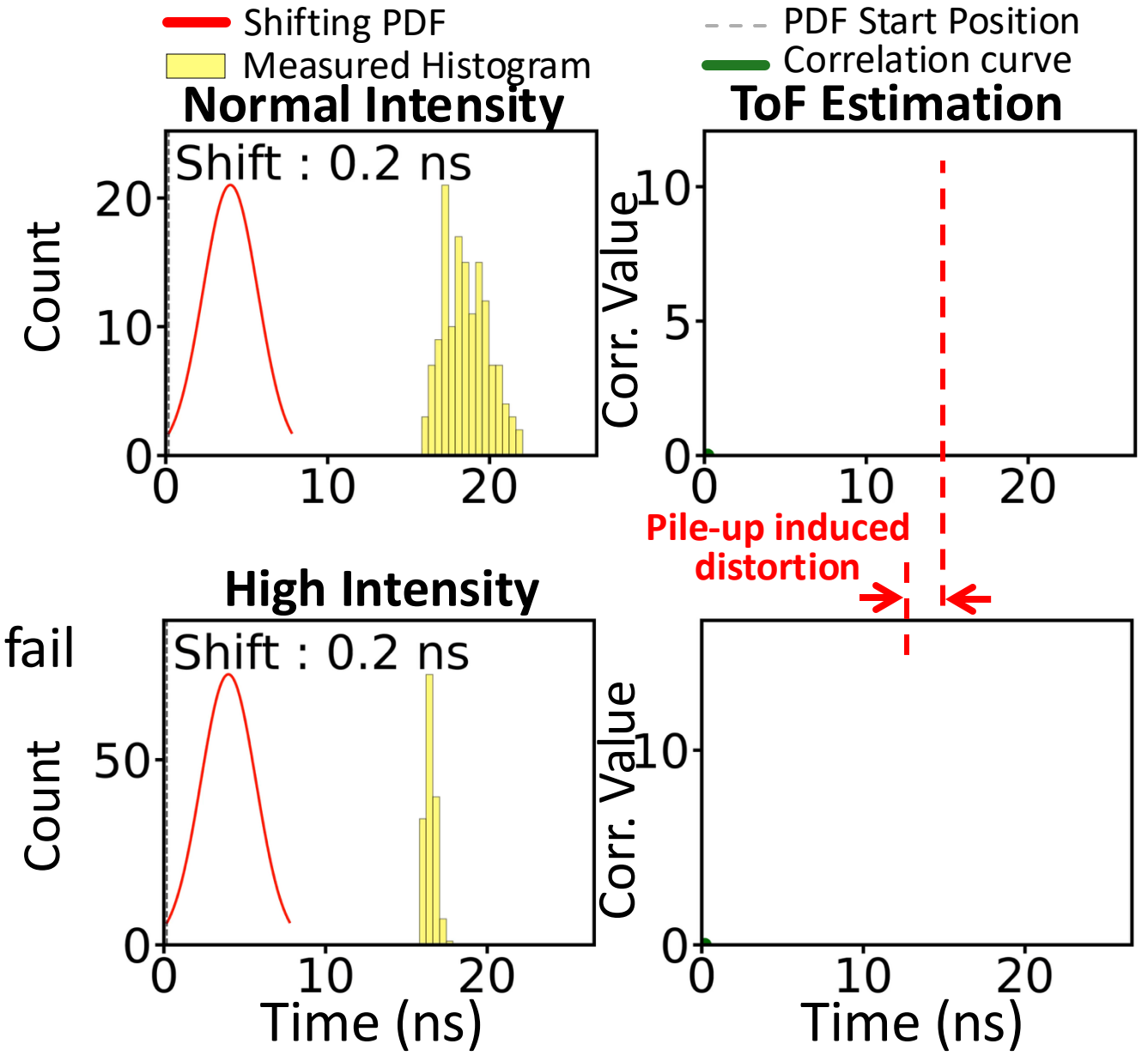
- **Operating principle**
  - Distance measurement via laser round-trip time

- **Histogram construction**
  - Repetitive measurements to handle undesired factors from various noise sources

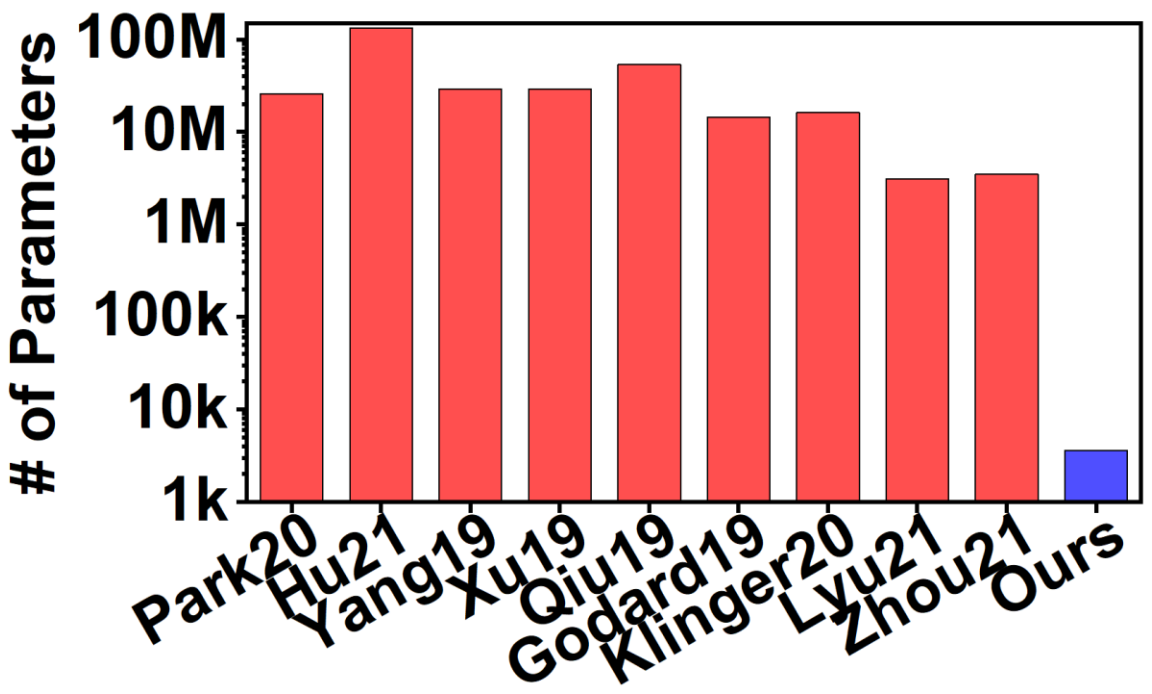


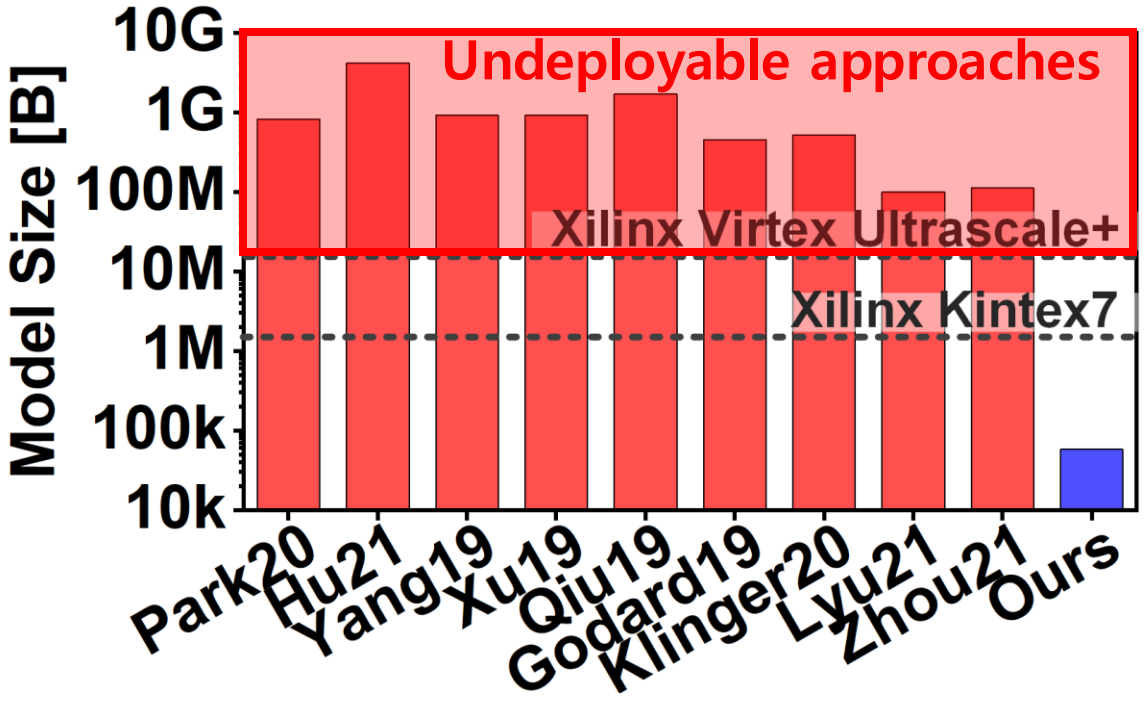
Problem #1: "Pile-Up" Distortion

- **Traditional algorithms**
  - Assume ideal Gaussian pulses
  - High-intensity reflections
    → Signal-induced pile-up
- **Histogram distortion**
  - Pile-up causes classical estimators to fail
    → **Deep learning has emerged as a strong alternative**



Problem #2: Limited Deployability of SOTA AI





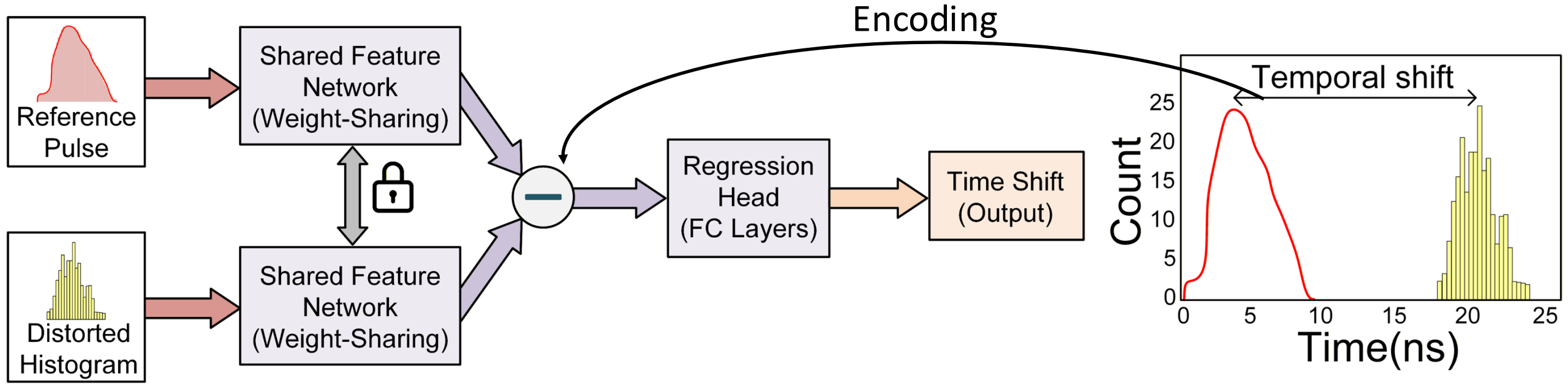
- **Reliance on large models is impractical for edge deployment**
- **Hindrance of real-time inference on resource-limited hardware**

Our Contributions

Enabling distortion-robust and efficient on-device deployment

- **Key contribution #1: A new paradigm**
  - Shift from signal filtering to lightweight similarity learning
- **Key contribution #2: Hardware-validated practicality**
  - Real-time inference at 106.27 fps on FPGA with a 57.61 KB model
- **Key contribution #3: Demonstrated robustness**
  - Consistent accuracy under severe pile-up and reflectance changes

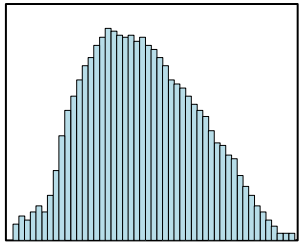
Why Siamese Network?



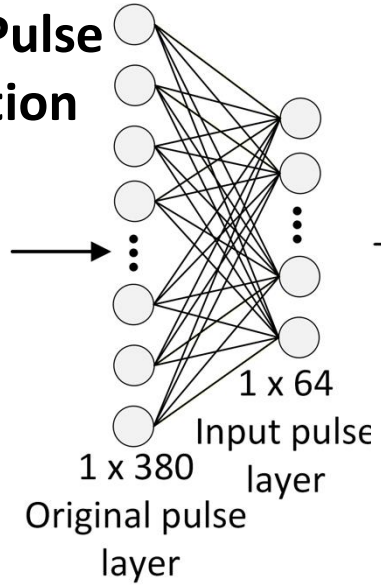
- **Nature of ToF estimation: Essentially a temporal similarity search**
- **Paradigm shift: From signal filtering to similarity learning**
- **Key advantages**
  - Robustness: Feature differences directly encode temporal shifts, bypassing pile-up
  - Efficiency: Weight-sharing enables a compact model (57 KB) for FPGA

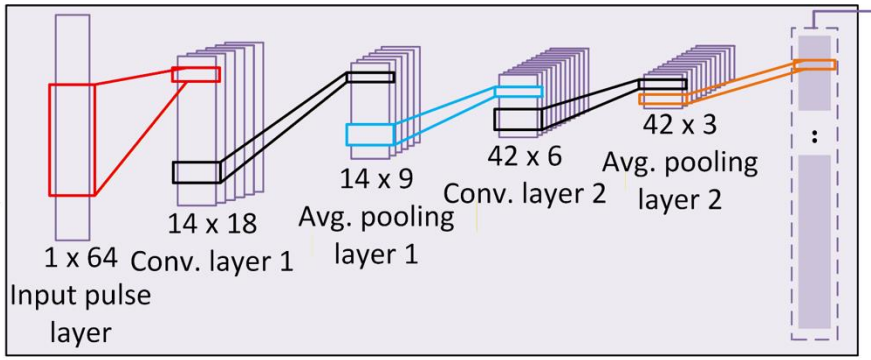
Network Architecture of LiToFNet

① Original Pulse Transformation



Original pulse shape



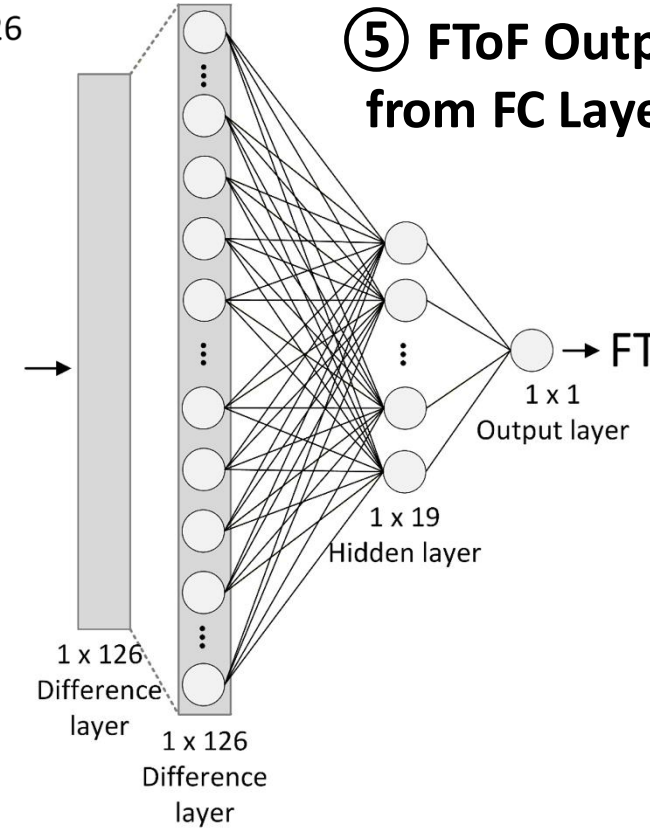


Subnetwork 1

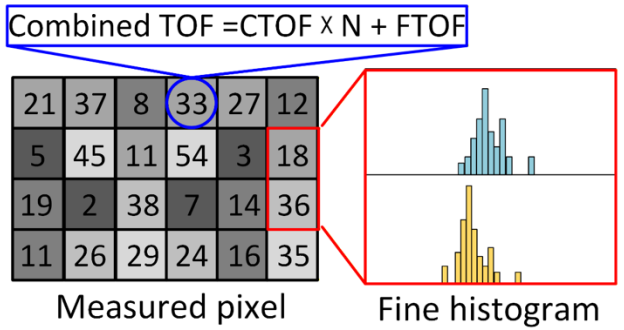
② Flattened feature vector of Subnetwork 1

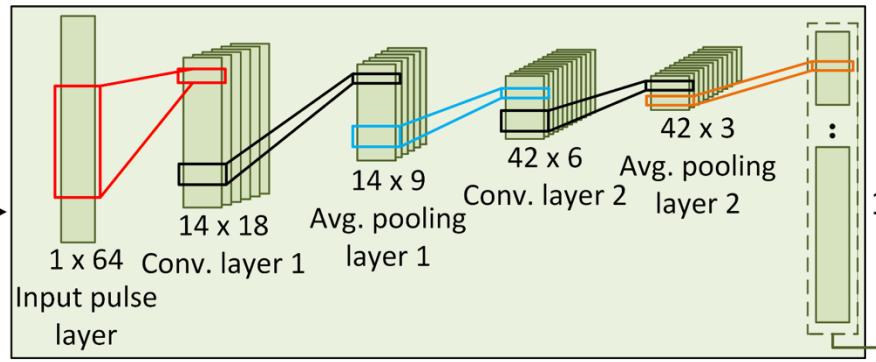
1 x 126

1 x 126



⑤ FToF Output from FC Layer



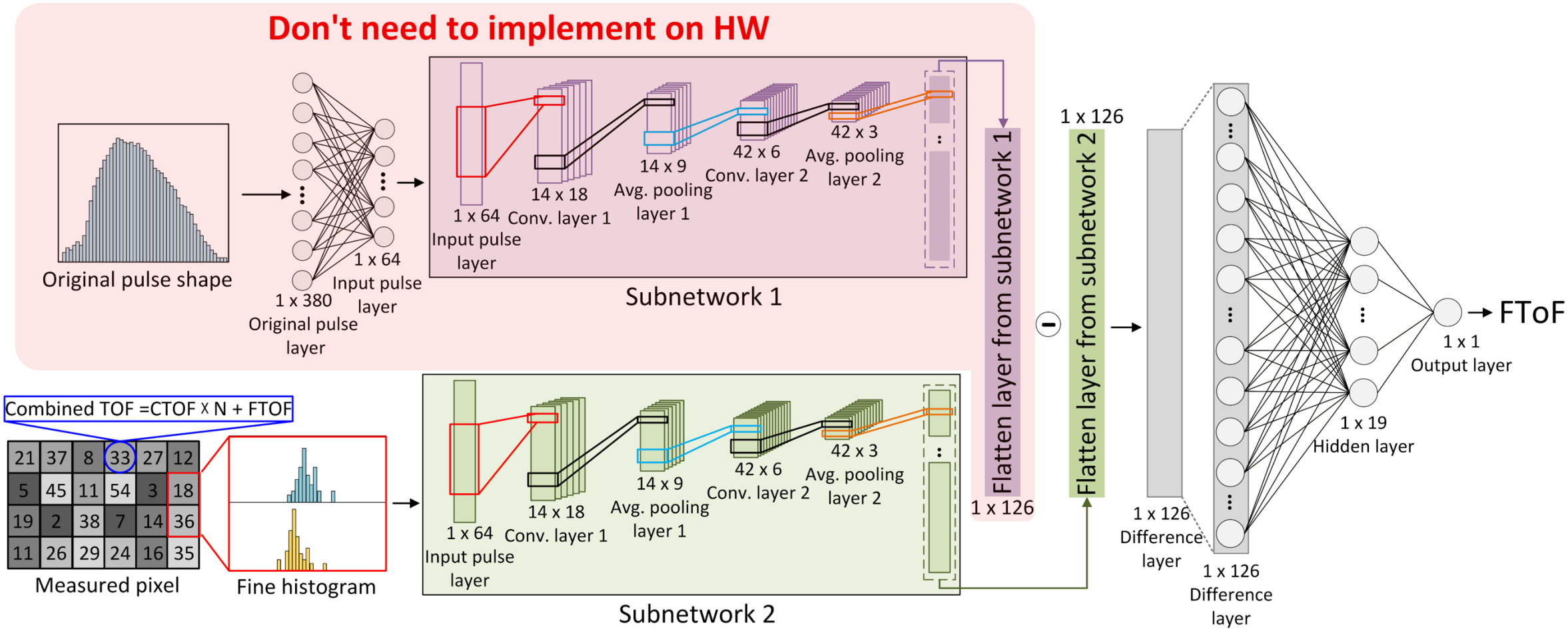


Subnetwork 2

③ Fine Histogram Extraction from Measured Pixel

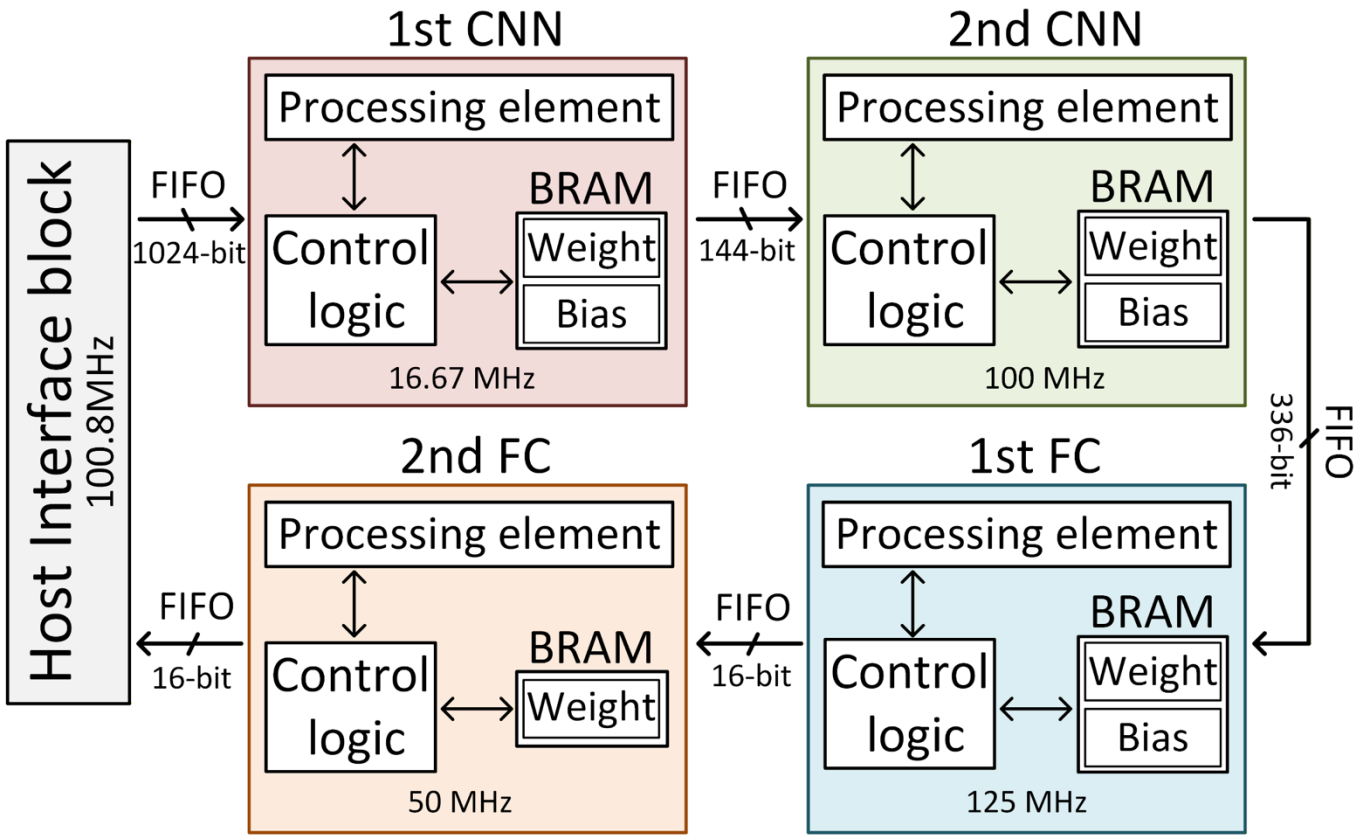
④ Flattened feature vector of Subnetwork 2

Hardware Friendly Implementation Method



- Method #1 : Parameter sharing allows the use of only one instance
- Method #2 : Pre-calculated pulse features used as bias in FC

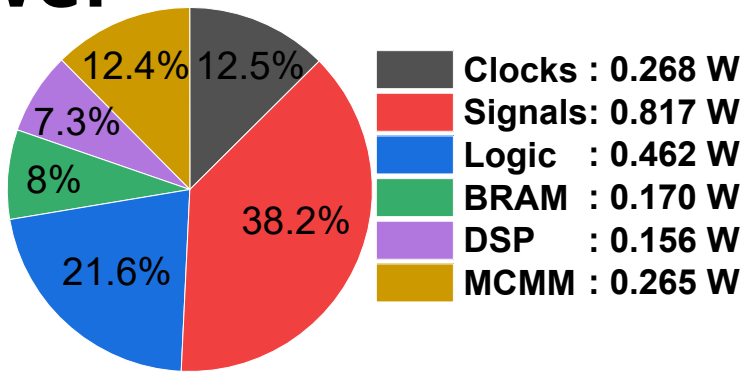
Hardware Implementation



Resource Usage

	Resource	Used
LUT	101400	88161
DSP	600	358
BRAM	325	132.5

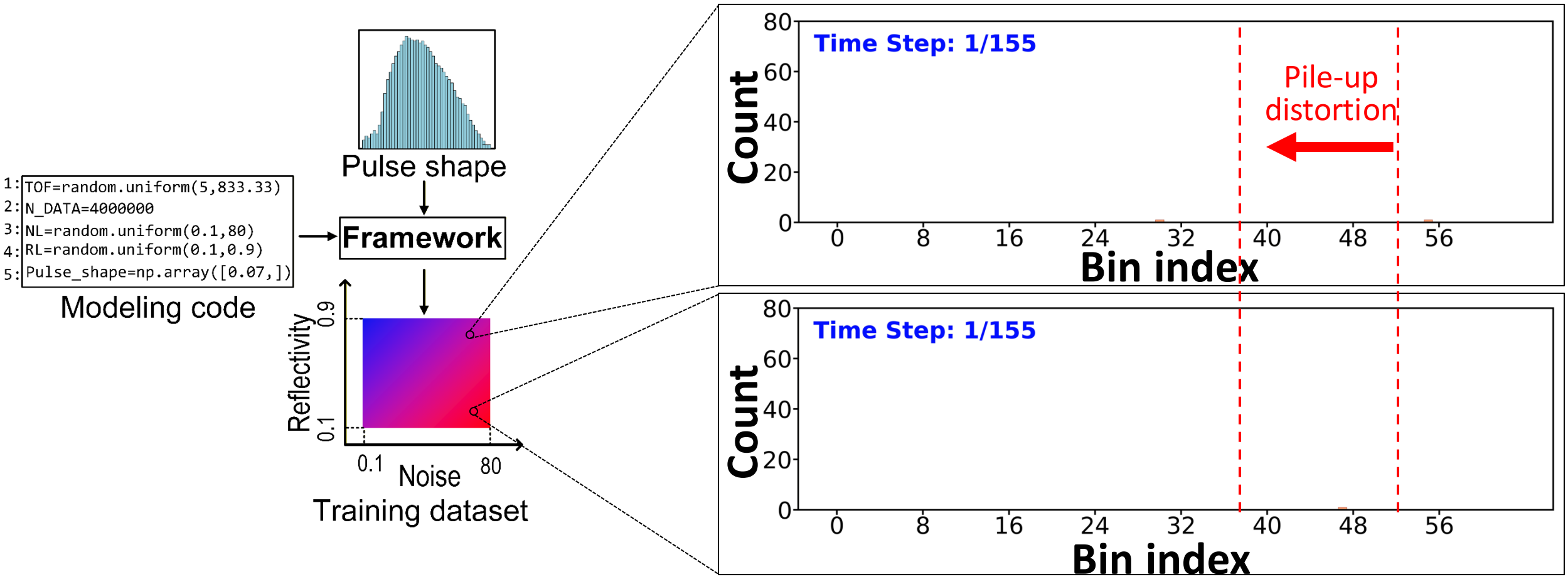
Power



- **Throughput of 840 ns/pixel**
  - Real-time inference at 106 fps* for 192 × 52 resolution

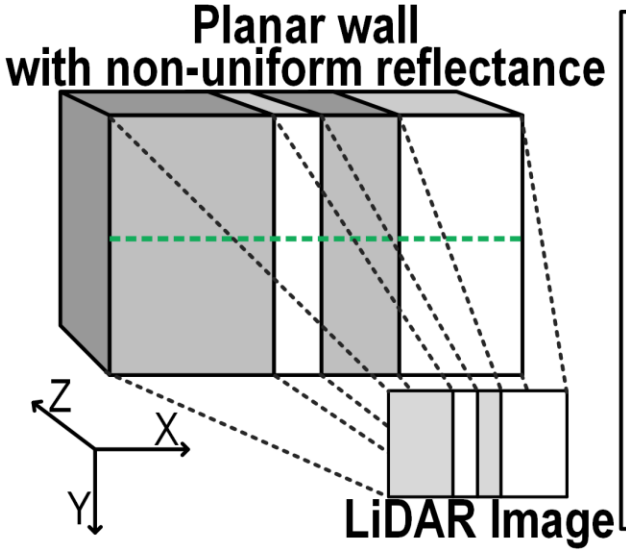
*Latency causes a small mismatch between the ideal and the real throughput.

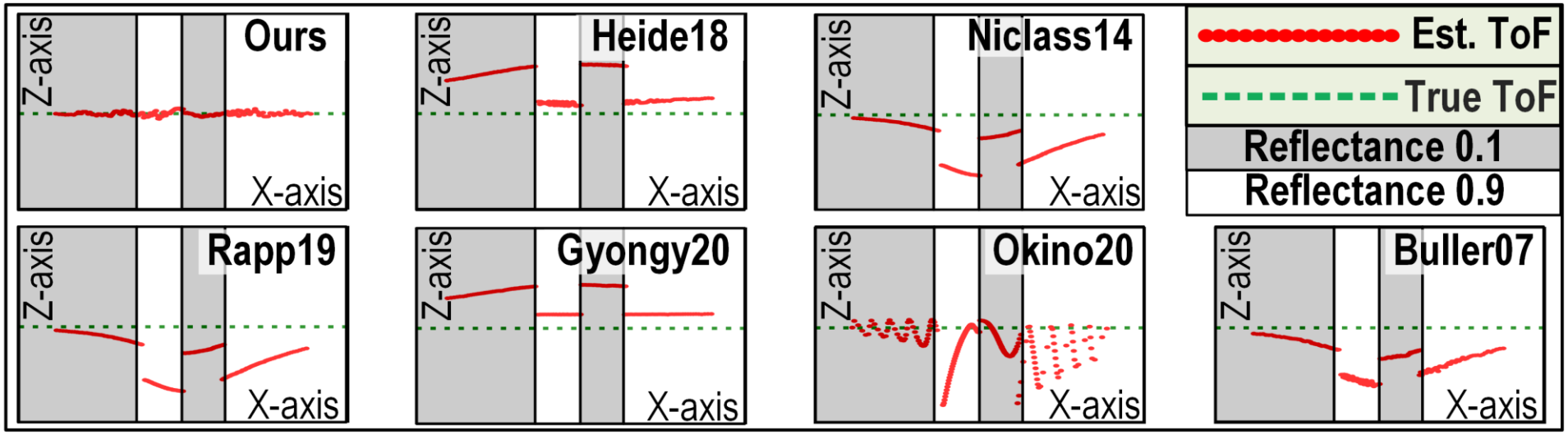
Simulation Framework



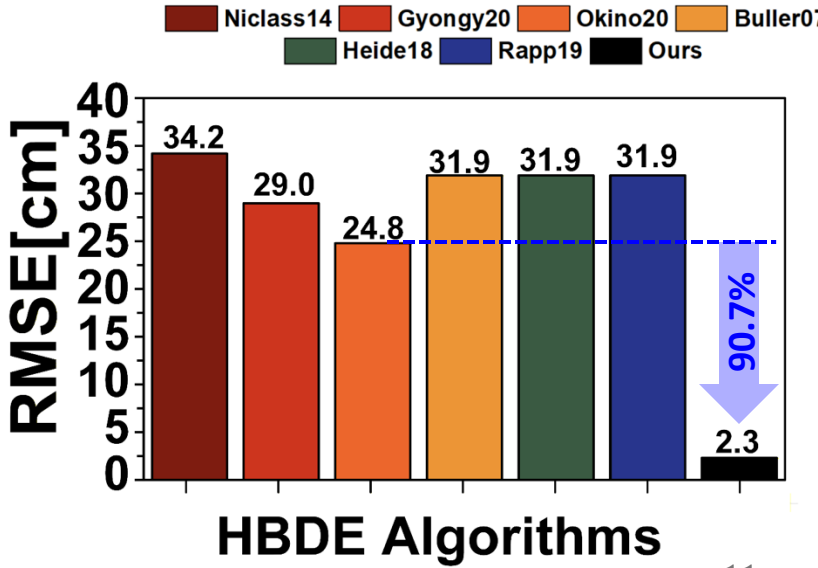
- **Photonic events modeled as an inhomogeneous Poisson process**
  - Time-varying rate reflecting signal & noise
  - Incorporation of SPAD dead time

Experimental Evaluation – Pile-Up Evaluation



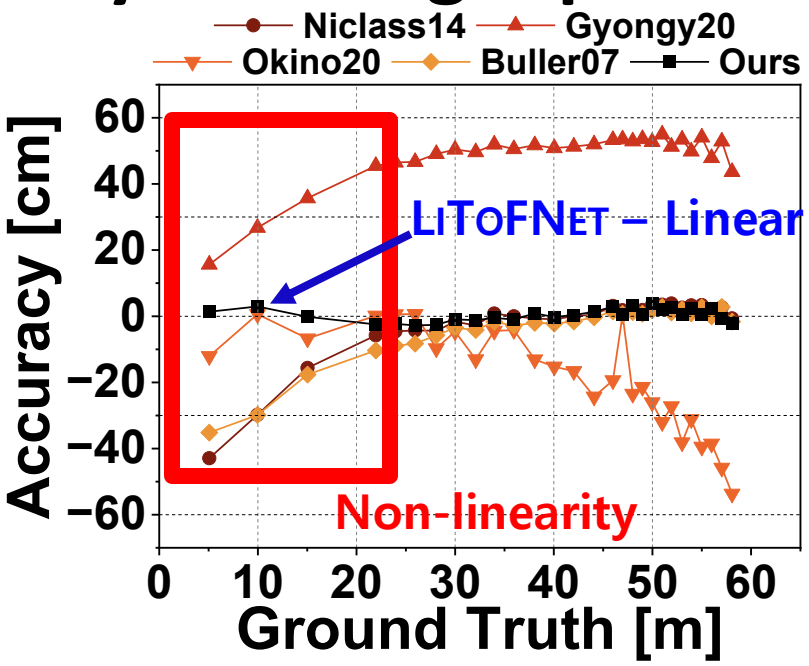


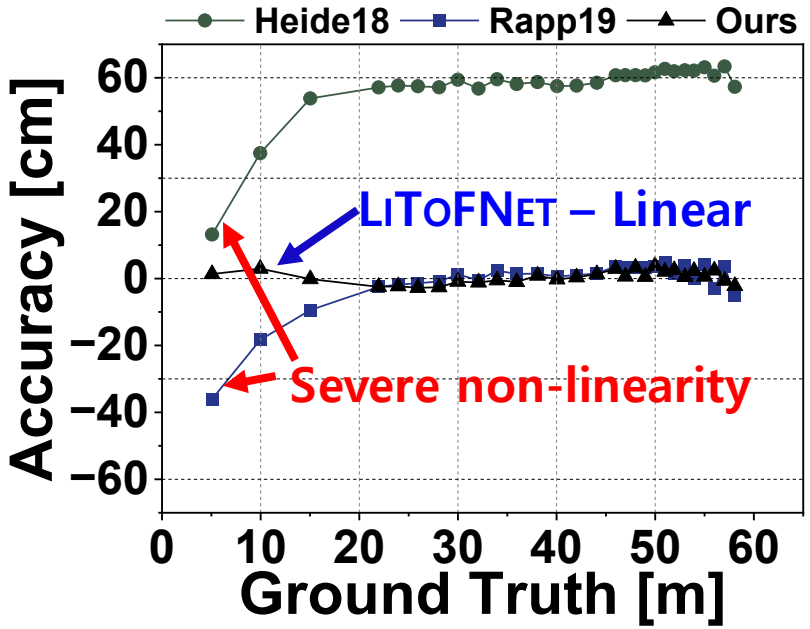
- **Achievement of a minimal RMSE (2.3 cm)**
  - Robustness to abrupt signal intensity changes
- **Elimination of boundary distortion**
  - Maintenance of accuracy under unstable signals

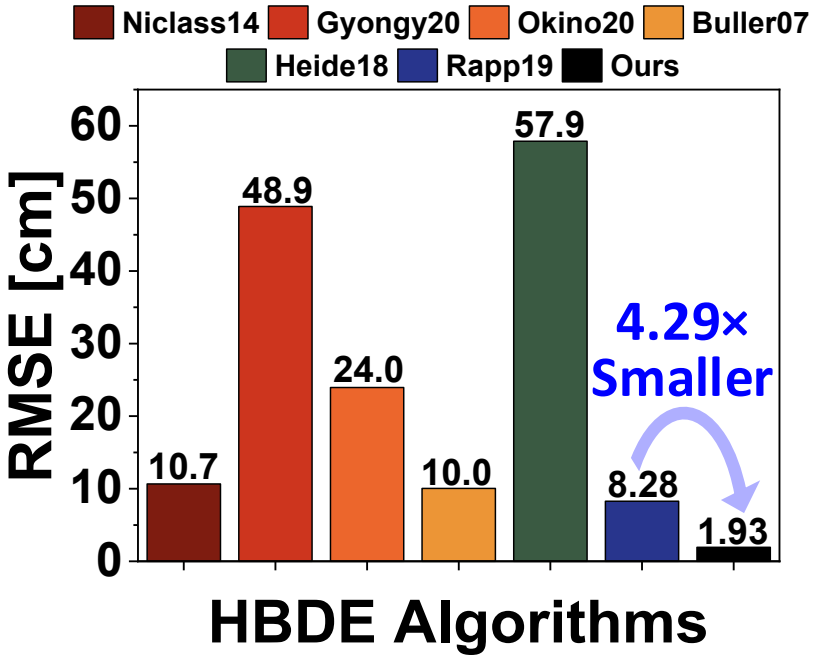


Experimental Evaluation – Real-World Validation (1)

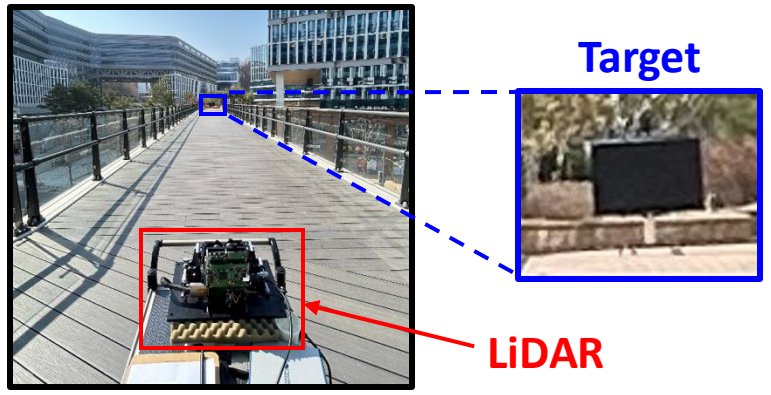
Daytime single-point measurement







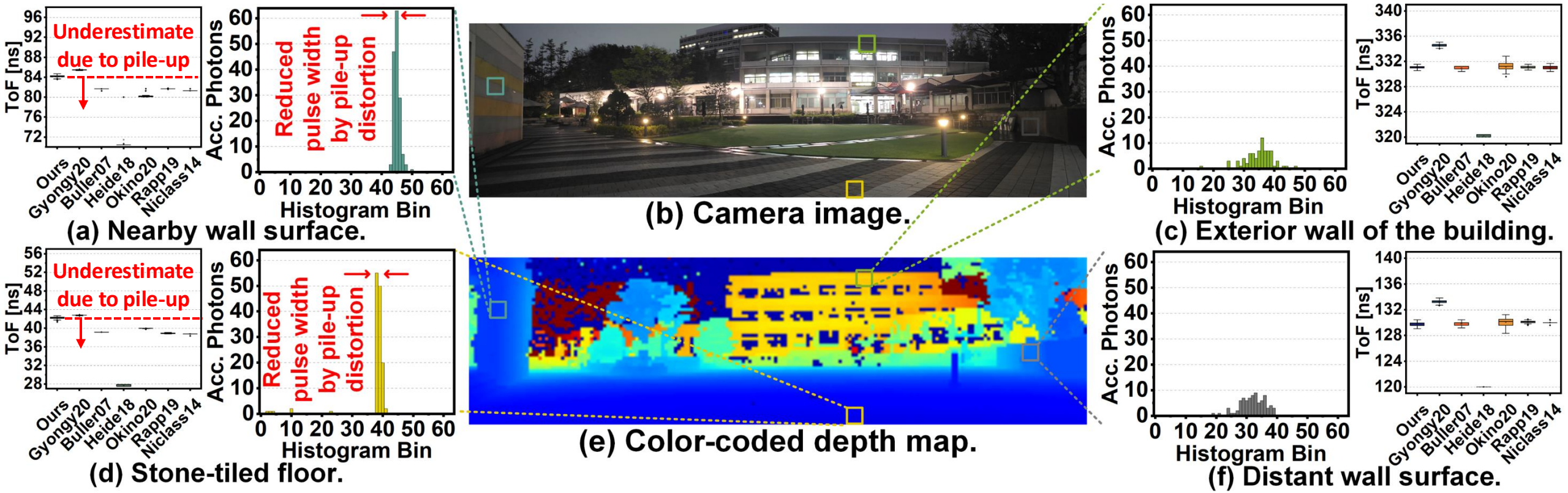
- **Key achievements**
  - Superior accuracy (4.29× lower RMSE)
  - Robust near-field performance (< 15 m)
  - Immunity to signal-induced pile-up



Exp. Setup

Experimental Evaluation – Real-World Validation (2)

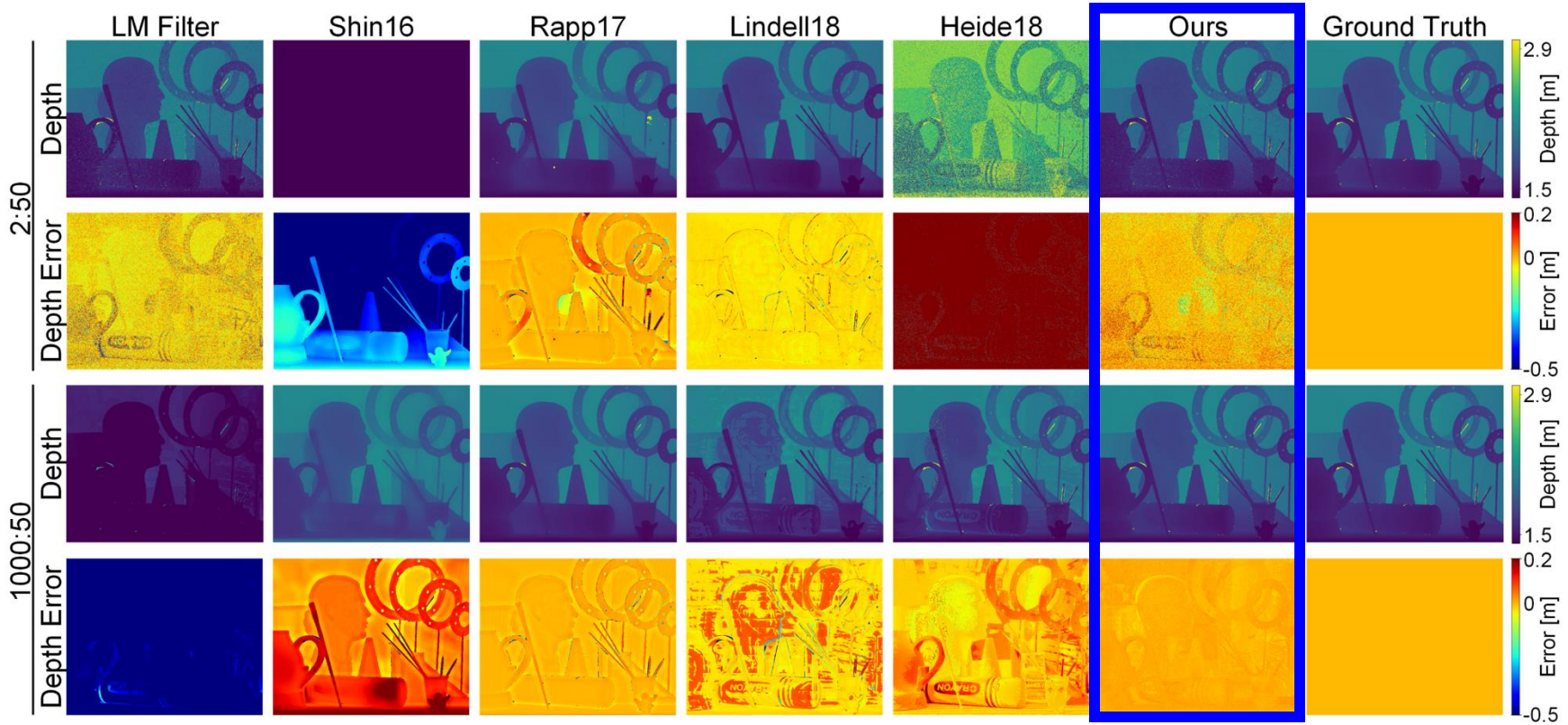
Nighttime outdoor measurement

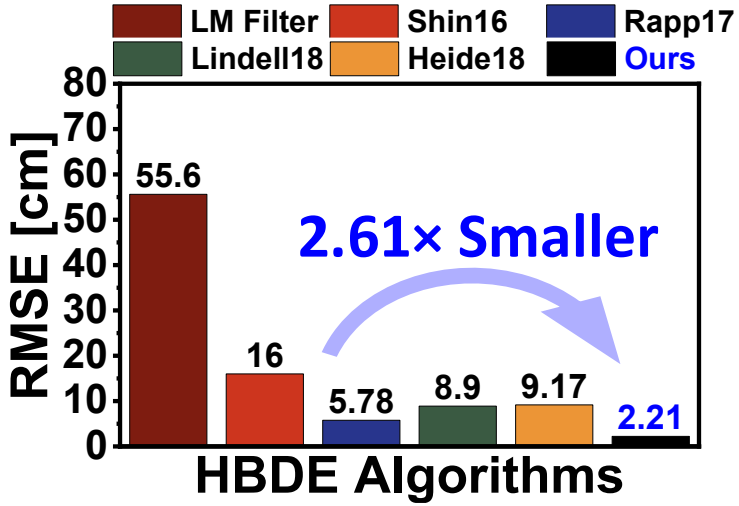


- Occurrence of severe pile-up at nearby surfaces in (a), (d)
- Based on measured data, the hardware implementation of LiToFNet successfully overcomes pile-up distortion

Experimental Evaluation – Middlebury Dataset

Evaluate performance under varying signal and noise conditions





- **Achievement of a minimal RMSE (2.21 cm)**
  - Robustness under strong intensity and severe pile-up
  - Demonstration of superior performance over existing methods

Conclusion

- **Methodology**
  - Paradigm shift from signal filtering to lightweight similarity learning
  - Compact network learns a pile-up–robust similarity metric

- **Hardware-validated efficiency**
  - A **57.61 KB** model, over **215.2× smaller** than prior work
  - Real-time inference at **106.27 fps** on FPGA

- **Robustness & impact**
  - **2.21 cm RMSE** maintained even under severe pile-up conditions
  - This work bridges the gap between high performance and on-device deployability

Thank You!

Please refer to our paper and codes for more details.