PREDICTING HIGH-PRECISION DEPTH ON LOW-PRECISION DEVICES USING 2D HILBERT CURVES

Mykhailo Uss[1,2] Ruslan Yermolenko[1,3] Oleksii Shashko[1,4] Olena Kolodiazhna[5] Ivan Safonov[1]
Volodymyr Savin[1,5] Yoonjae Yeo[6] Seowon Ji[7] Jaeyun Jeong[6]

[1] Samsung R&D Institute Ukraine, Kyiv

[2] Department of Information-Communication Technologies, National Aerospace University, Kharkiv

[3] Faculty of Physics, Taras Shevchenko National University of Kyiv, Kyiv

[4] Department of Artificial Intelligence, Kharkiv National University of Radio Electronics, Kharkiv

[5] Institute of Physics and Technology, NTUU "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv

[6] Samsung Research, Seoul

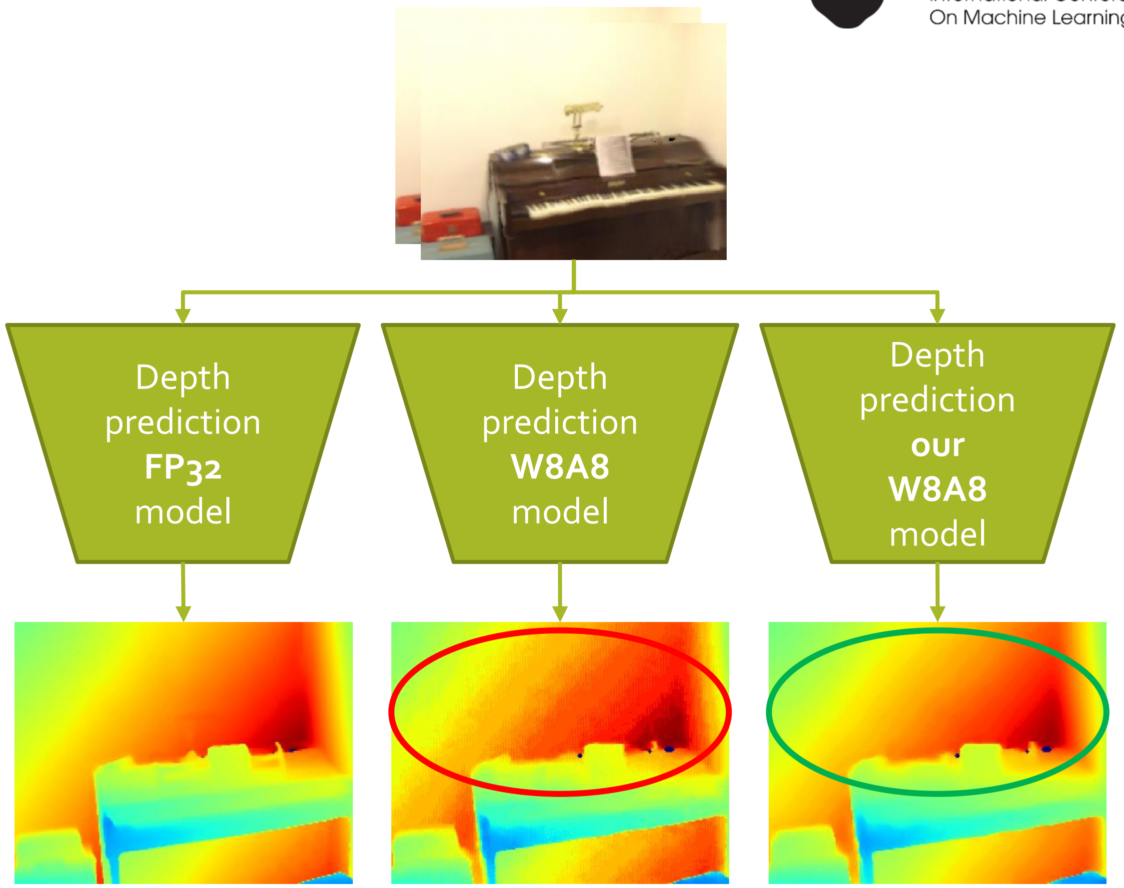
[7] Department of Computer Science and Engineering, Konkuk University, Seoul

ICML
International Conference On Machine Learning

Introduction

Dense depth prediction

- Deep neural networks (DNN) for dense depth prediction are used to predict depth values for each pixel of an input image for both monocular and binocular data.
- Application areas include scene understanding, autonomous driving, robotics, augmented / virtual reality, etc.
- Usually, applications in these areas rely on low-end devices and have strong limitations on hardware capabilities and energy consumption.
- Depth prediction models usually have high computational complexity, restricting their use on low-end devices.



Problem statement

- For inference on low-end devices DNN should be quantized (e.g. to eight-bit weights and activations) using post-training-quantization (PTQ) or quantization-aware-training (QAT).
- Quantization leads to depth prediction model performance degradation that is specific for depth modality.
- **We seek to overcome this problem by representing depth in a form beneficial for quantization**

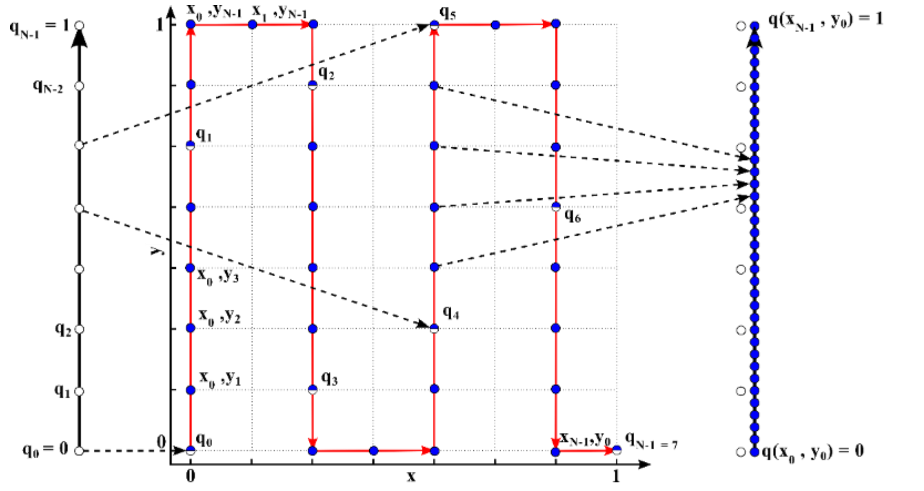
Method

Low-end devices limitation

- Model output is represented in low precision, typically eight bit or less.
- Eight-bit precision is sufficient for RGB images.
- Depth is high dynamic range signal that needs ~10-11 bits for accurate representation. For example, representing depth in the range 0…10 m with 1 cm accuracy requires ten bits

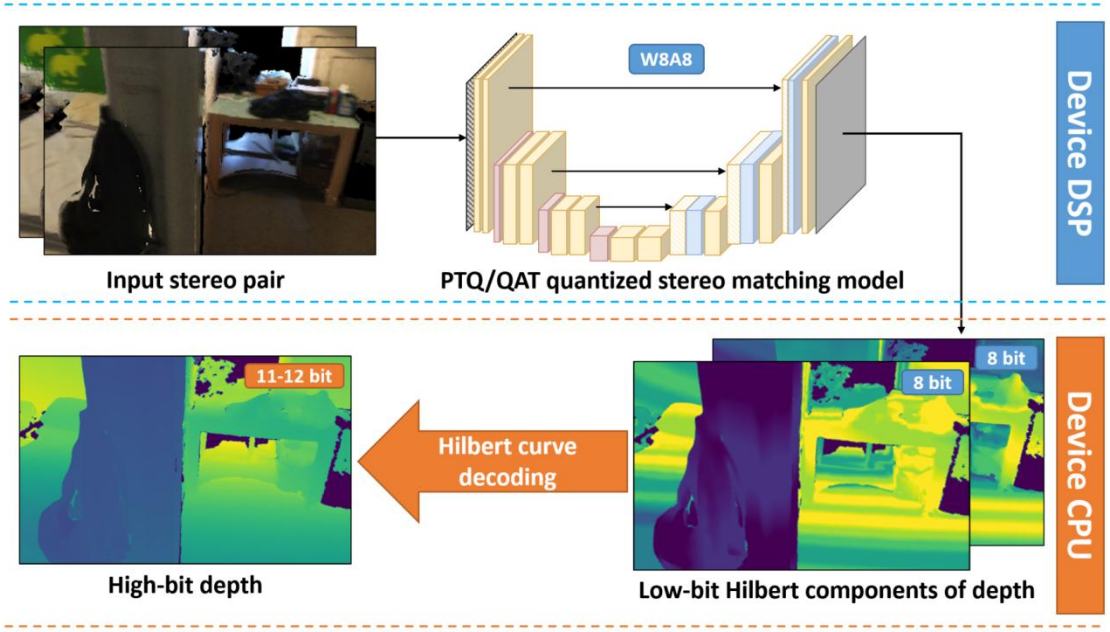
Main idea

- Represent high dynamic range depth as two low dynamic range components -> predict these components on DSP -> reconstruct depth on CPU at the post-processing stage.
- Two components code depth as points on 2D parametric curve. Depth bit-precision can be increased by $\log_2 L$, where L is length of the curve.



Proposed method

- Training the full-precision model to directly predict the Hilbert curve components of depth representation.
- Applying standard quantization methods (either PTQ or QAT).
- Running inference of the modified quantized model on-device and obtaining Hilbert curve components in low-bit precision.
- Applying post-processing to Hilbert curve components and reconstruct depth in higher-bit precision.



Depth representation as points on 2D Hilbert curve

ICML International Conference On Machine Learning

Hilbert curve structure

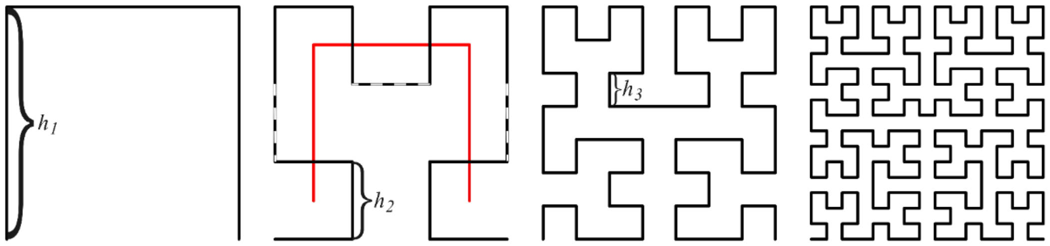
- Continuous fractal space-filling curve that is constructed as a limit of piece-wise linear curves.
- Possesses the following properties: continuity, non-self-intersection, boundedness and self-avoidance.
- Curve length $L_p = 2^p + 1$, where p is the curve order. For example, 5 for p=2, and 9 for p = 3.
- Allows coding signal with dynamic range L_p as two signals with dynamic range 1.

Training process

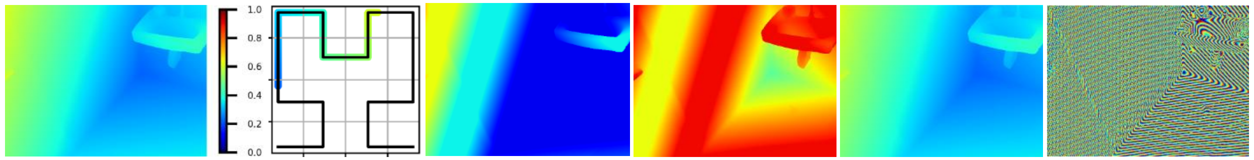
- Depth prediction DNN is modified to predict x and y components of the Hilbert curve.
- Modified loss is applied to learn the modified representation. It includes original depth loss and loss terms applied to Hilbert curve components.

$$\Lambda_{full} = \Lambda(q_{GT}, q_{xy}) + \alpha \cdot \Lambda_H(x_{GT}, y_{GT}, x, y)$$

$$\Lambda_H(x_{GT}, y_{GT}, x, y) = (x_{GT} - x)^2 + (y_{GT} - y)^2 + \beta \cdot r_{xy}^2$$

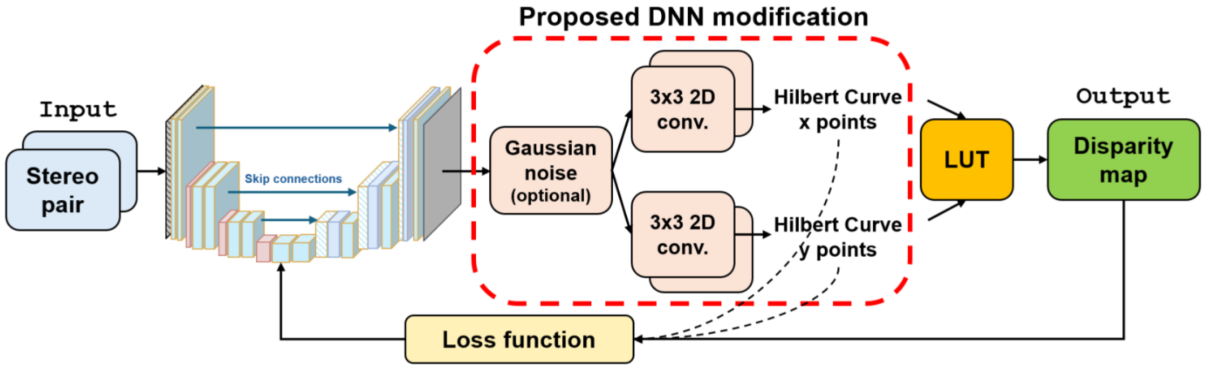


Hilbert curves for orders $p = 1,2,3,4$ (from left to right). Every order is formed by the replacement of every node by an elementary 3-segment sequence.



(a) (b) (c) (d) (e) (f)

Illustration of disparity transforms: (a) disparity map; (b) mapping to 2D with second order Hilbert curve; (c, d) x and y components of the Hilbert curve; (e, f) coarse and fine details of disparity map. Fine details in (f) are the least significant byte of disparity (a) represented in 16-bit format. High-frequency oscillations make it appear different from the original disparity and difficult to predict by a DNN model.



Experimental results

Dataset and models

- Training/validation/test datasets: rendered from ScanNet v2 meshes using PyRender.
- Models:
  - DispNet – U-Net-like architecture;
  - DPT – transformer-based model with MobileViTv3-S encoder.

Evaluation metrics

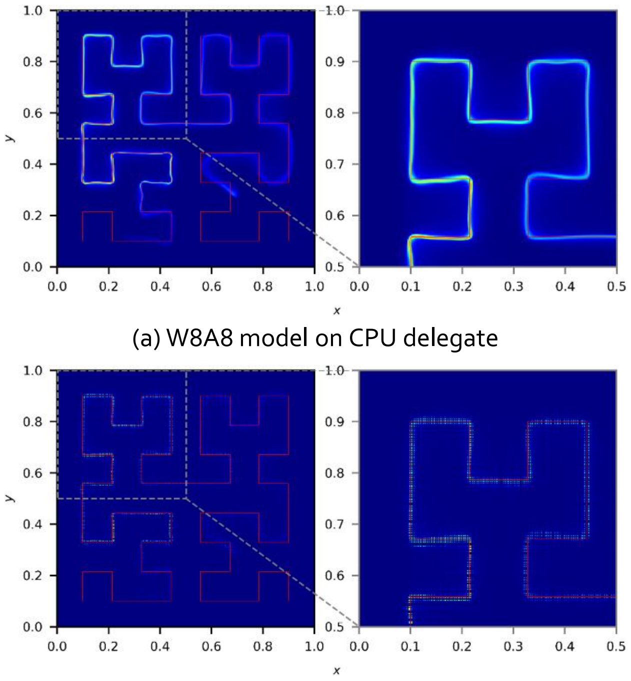
- Basic DFS metrics (mean absolute relative error, EPE, D1);
- Cosine similarity between depth DCT coefficients (S_C) – sensitive to structural errors.
- Inference time, Power consumption.

Results

- The modified DispNet model in W8A8 format shows better quality than the original model in W8A16 format.
- Compared to the original model in W8A16 format, the modified DPT model in W8A8 format shows significantly better Abs Rel metric and only slightly worse Sc.
- For W8A8 model, overhead due to post-processing and Hilbert components prediction is about 14%.
- Compared to original models in W8A16 format, modified models in W8A8 format reduce energy consumption by 35% and latency by 30-54%.
- Quantization error reduced by up to 4.6 times.

Analysis of W8A8 Models

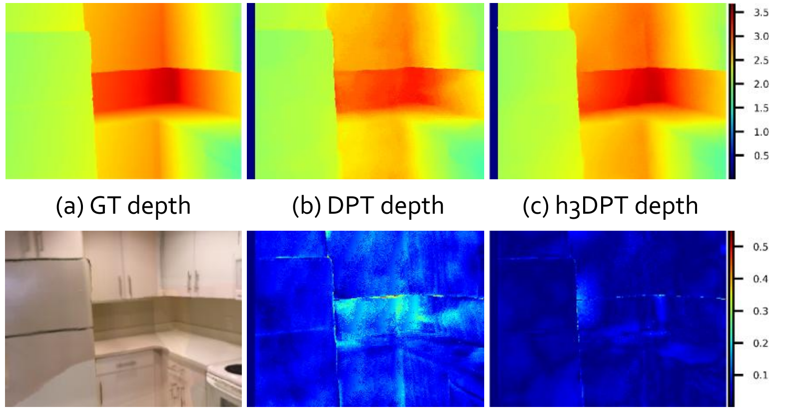
Precision	Abs Rel, %	EPE, px	D1, %	S_C	T, ms	P, mW·s/infr.
DispNet						
FP32	1.01	0.29	1.81	0.858	-	-
FP16	1.50	0.37	1.80	0.855	19.54	19.52
W8A16	1.78	0.63	5.22	0.798	18.7	12.3
W8A8	2.02	0.69	5.34	0.585	10.5	7.1
Ours, W8A8	0.93	0.24	1.26	0.807	12.0	8.7
DPT						
FP32	0.75	0.21	0.95	0.889	-	-
FP16	1.14	0.27	0.97	0.884	54.1	110.5
W8A16	4.03	0.97	5.58	0.825	46.2	64.5
W8A8	4.16	1.03	5.76	0.520	26.7	28.3
Ours, W8A8	1.35	0.32	1.28	0.697	30.4	29.7



(a) W8A8 model on CPU delegate

(d) W8A8 model on DSP delegate

2D histogram of h3DispNet W8A8 model output for CPU and DSP delegates.



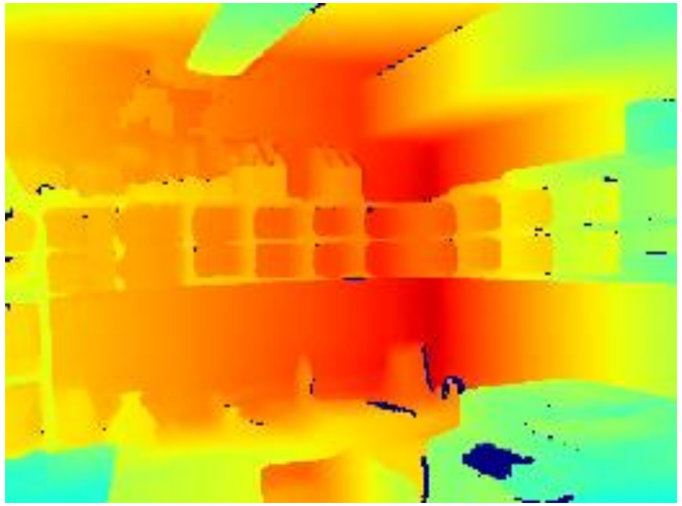
(a) GT depth (b) DPT depth (c) h3DPT depth

(d) Image (e) |GT - DPT| (f) |GT - h3DPT|

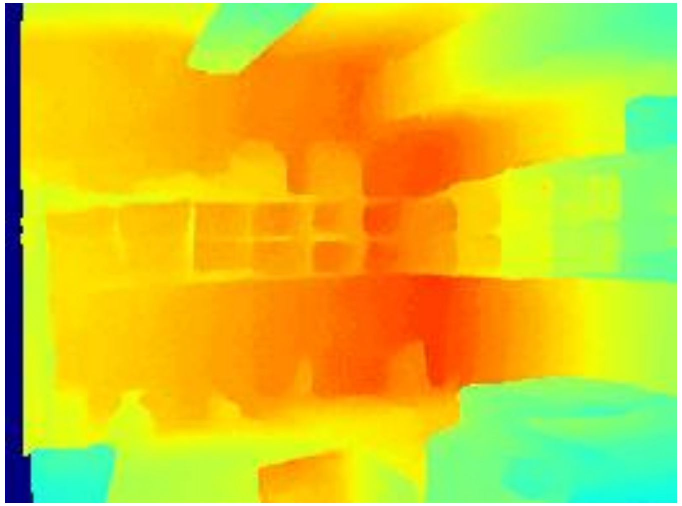
Quantization errors of DPT and h3DPT, W8A8, DSP.

Results

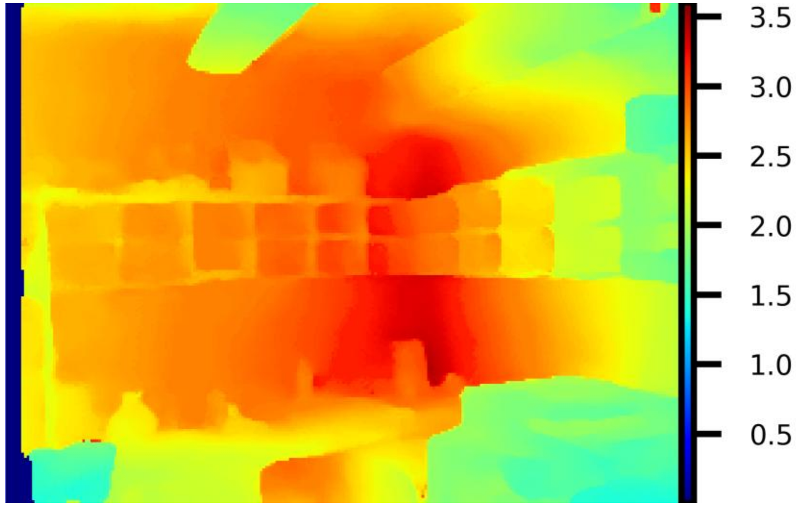
ICML
International Conference
On Machine Learning



(a) GT depth

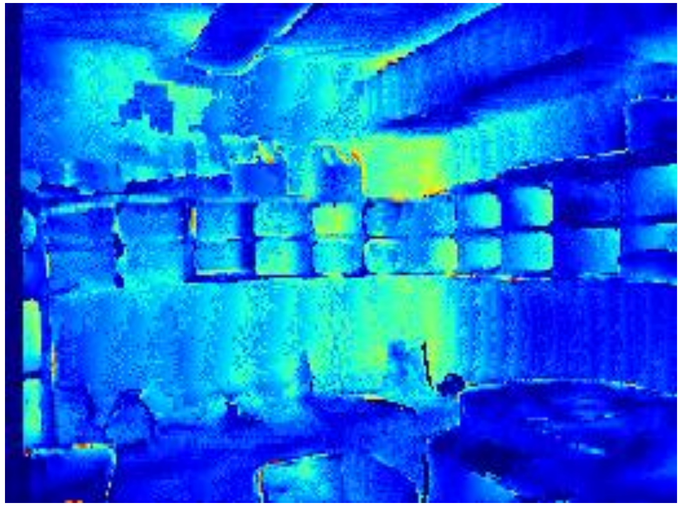


(b) DPT depth



(c) h3DPT depth (third order Hilbert cure)

(d) Image



(e) |GT - DPT|

(f) |GT - h3DPT|

International Conference On Machine Learning

THANK YOU