

Pixel-Level Predictions with Embedded Lookup Tables

Marius Aasan and Adín Ramírez Rivera

University of Oslo, SFI Visual Intelligence



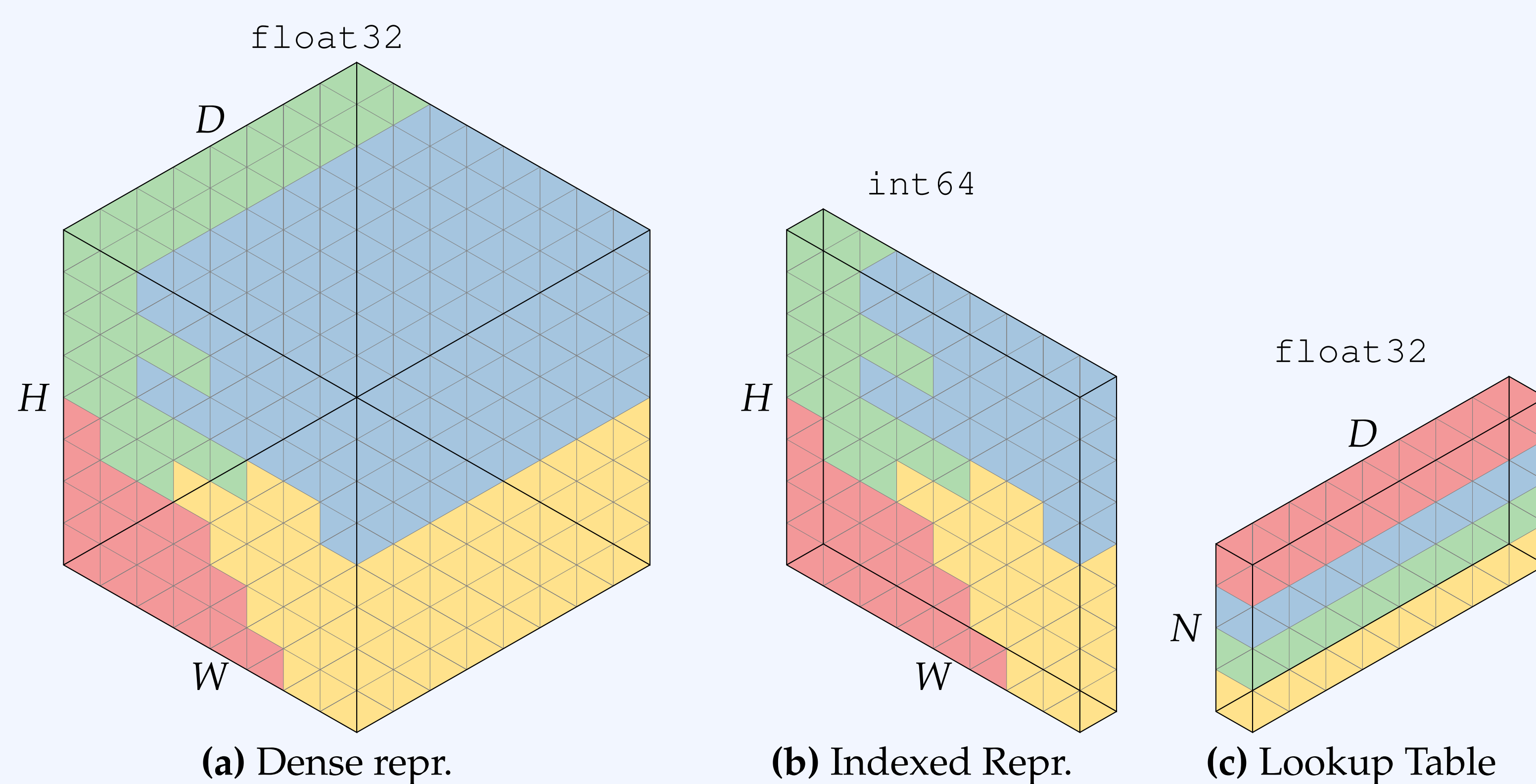
Goal: Improve efficiency in dense tasks

Motivation

Embedded Lookup Tables significantly cut down on dense-tensor bottlenecks.

- **Memory Wall:** Dense representations scale as $\mathcal{O}(DHW)$, increases explode GPU VRAM.
- **Dimensionality:** More classes, bigger images $\rightarrow$ challenging for pixel-level prediction.
- **Redundancy:** Adjacent pixels often share labels $\rightarrow$ *near-identical representations*.
- **Proposal:** Use *indexed representations with lookup tables for embeddings*.

Dense (a) versus ELUT (b + c)



Space Complexity

Let $P = H \times W$ denote the number of pixels in an image, D the feature dimension, and $N \leq P$ the number of regions. We compare the memory footprint of a standard dense feature map

$$M_{\text{dense}} = D \times P \quad (1)$$

against that of an ELUT representation, which is given by

$$M_{\text{ELUT}} = \underbrace{P}_{\text{index map } I \in \{1, \dots, N\}^{H \times W}} + \underbrace{DN}_{\text{lookup table } T \in \mathbb{R}^{N \times D}}. \quad (2)$$

We can express the relative memory usage as

$$\frac{M_{\text{ELUT}}}{M_{\text{dense}}} = \frac{DN + P}{DP} = \frac{N}{P} + \frac{1}{D} = \alpha + \frac{1}{D}, \quad \alpha \equiv \frac{N}{P} \in (0, 1]. \quad (3)$$

Given this, ELUTs provide an overall memory *reduction factor*

$$\rho = \frac{M_{\text{dense}}}{M_{\text{ELUT}}} = \frac{1}{\alpha + 1/D}. \quad (4)$$

Peak GPU memory (GB) and throughput (images/s) for a forward-back-prop pass of cross-entropy (224×224 , float32).

Batch Size	Dense		ELUT	
	Mem. [GiB] ↓	img/ms ↑	Mem. [GiB] ↓	img/ms ↑
32	4.62	5.32	0.03	3.18
64	9.21	5.28	0.06	3.42
128	18.58	5.22	0.13	3.78
256	36.84	5.23	0.25	4.05

Per-image memory footprint over ImageNet (224×224) using SPiT with base capacity ($D = 768$, float32).

Steps t	0	1	2	3	4
No.regions N	50 176	11 940	3 156	795	197
Memory [MiB] M	147.4	35.2	9.4	2.5	0.8
Red.factor ρ	0.99×	4.16×	15.27×	54.21×	158.13×

Compression

Image	Dim.	HPX		PNG	
		bpp	KB	bpp	KB
	1024×1024	3.464	1330.11	3.800	1459.25
	1024×1024	2.539	974.84	5.996	2302.65
	512×512	5.848	561.43	6.255	600.48

Segmentation Example

Preliminary Results – Semantic segmentation on ADE20k.

Left: Original images. Centre: Annotated targets. Right: Predicted labels.

