FOUNDRY: DISTILLING 3D FOUNDATION MODELS FOR THE EDGE

Guillaume Letellier[1], Siddharth Srivastava[2], Frederic Jurie[1], Gaurav Sharma[3]

[1]GREYC, Normandy University, Unicaen, ENSICAEN, UMR CNRS 6072, F-14000 Caen, France
[2]IIT Delhi,[3]IIT Kanpur

{firstname.lastname}@unicaen.fr

CVPR JUNE 3-7, 2026 DENVER COLORADO

1. Motivation & Contributions

- **The Problem :** 3D foundation models are too compute and memory-intensive for edge deployment (e.g., robots, AR/VR headsets).
- **Our Solution :** Foundry, a compress-and-reconstruct distillation paradigm that compresses the teacher's latent space into a highly efficient set of learnable SuperTokens.
- **The Result :** A compact, general-purpose proxy model that retains strong transferability and data efficiency with a fraction of the computational cost.

2. Framework

Foundry distills 3D foundation models in latent space through token-level supervision while compressing inputs into a compact set of supertokens, without requiring downstream-task knowledge. The Gate variant further enables adjustable compression ratios at inference.

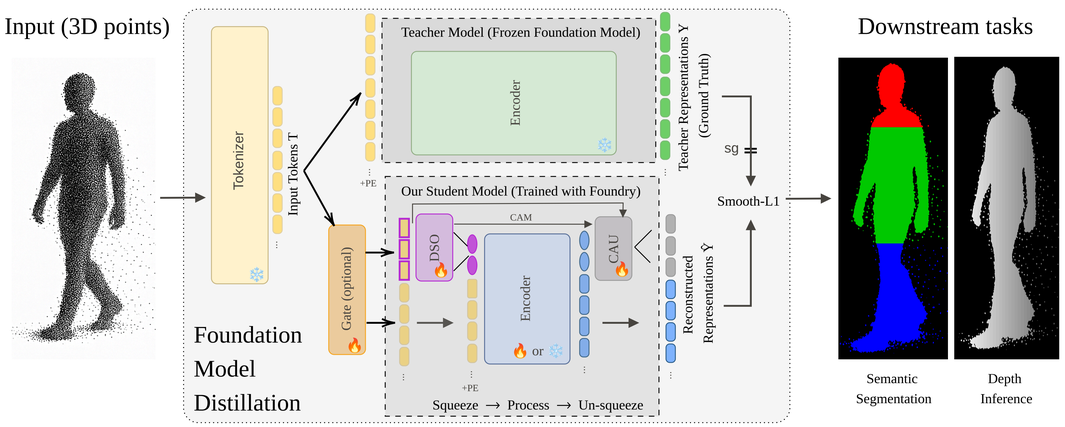

Figure 1: *Overview.*

Method	FLOPs (G)	Throughput (obj/s)
baseline	478.128	747.11
Foundry with $s = 16$ (Ours)	178.394	1160.97
Foundry with $s = 1$ (Ours)	137.274	1365.92

Table 1: *FLOPs analysis of Foundry compared to Transformer architecture (bs=64).*

Takeaway : Foundry is edge-ready, reducing FLOPs by 60%+ and handling large scenes with <4GB VRAM without OOM failures.

3. Quantitative results

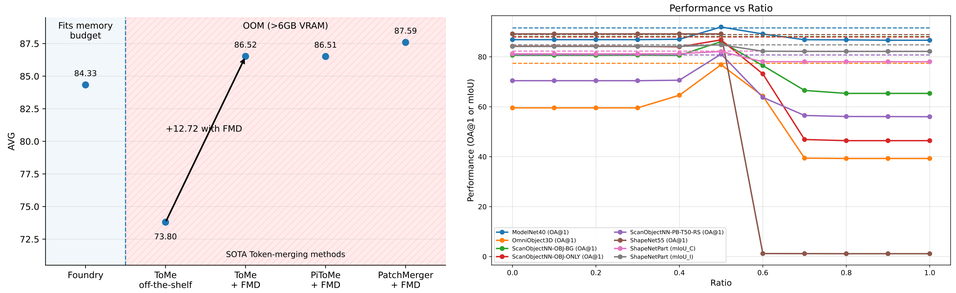

Figure 2: *Comparison with token-merging baselines and gate-based distillation.*

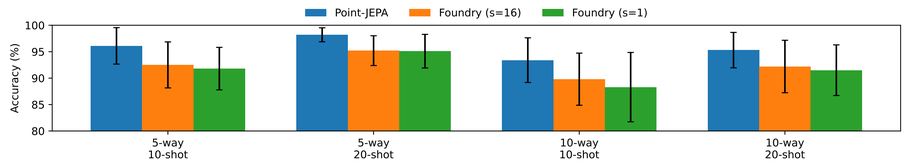

Figure 3: *Few-shot transfer performance on ModelNet40.*

Takeaway : Foundry remains close to token-merging baselines without reaching OOM, stays robust under high compression ratios, and preserves strong generalization capabilities.

4. Qualitative results

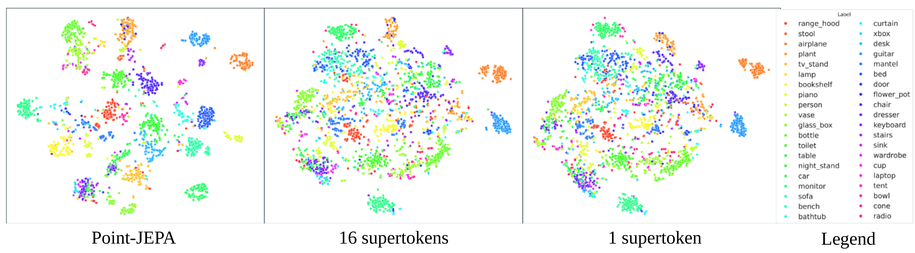

Figure 4: *t-SNE visualization of embedding spaces produced by Point-JEPA and Foundry.*

Takeaway : Despite massive token reduction, Foundry's SuperTokens successfully preserve the semantic structure and class separability of the original Point-JEPA teacher.

The IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 2026, Denver, CO, USA.