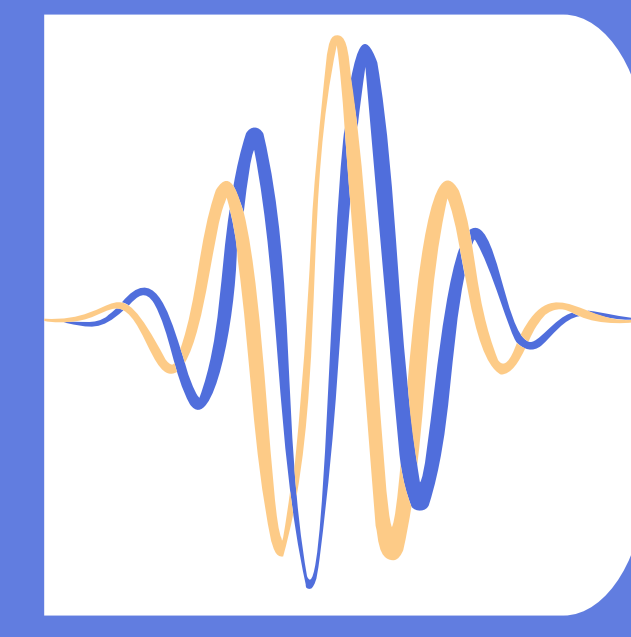
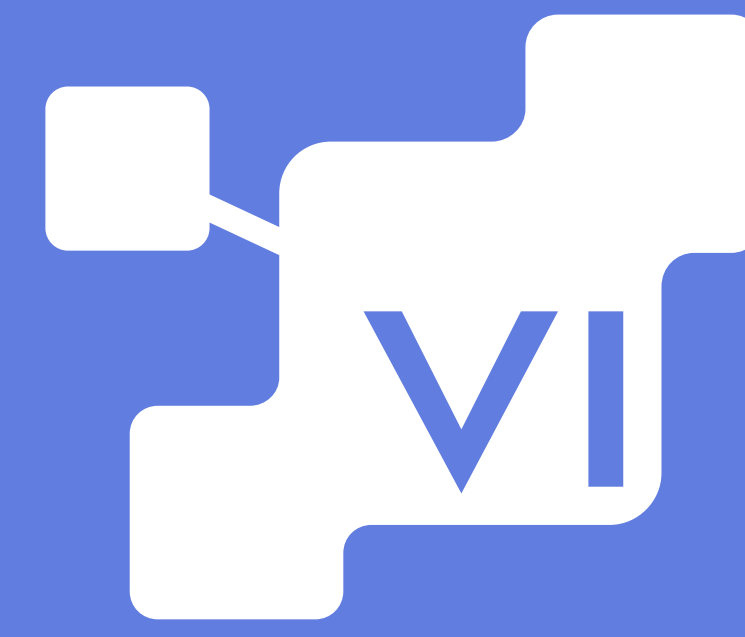


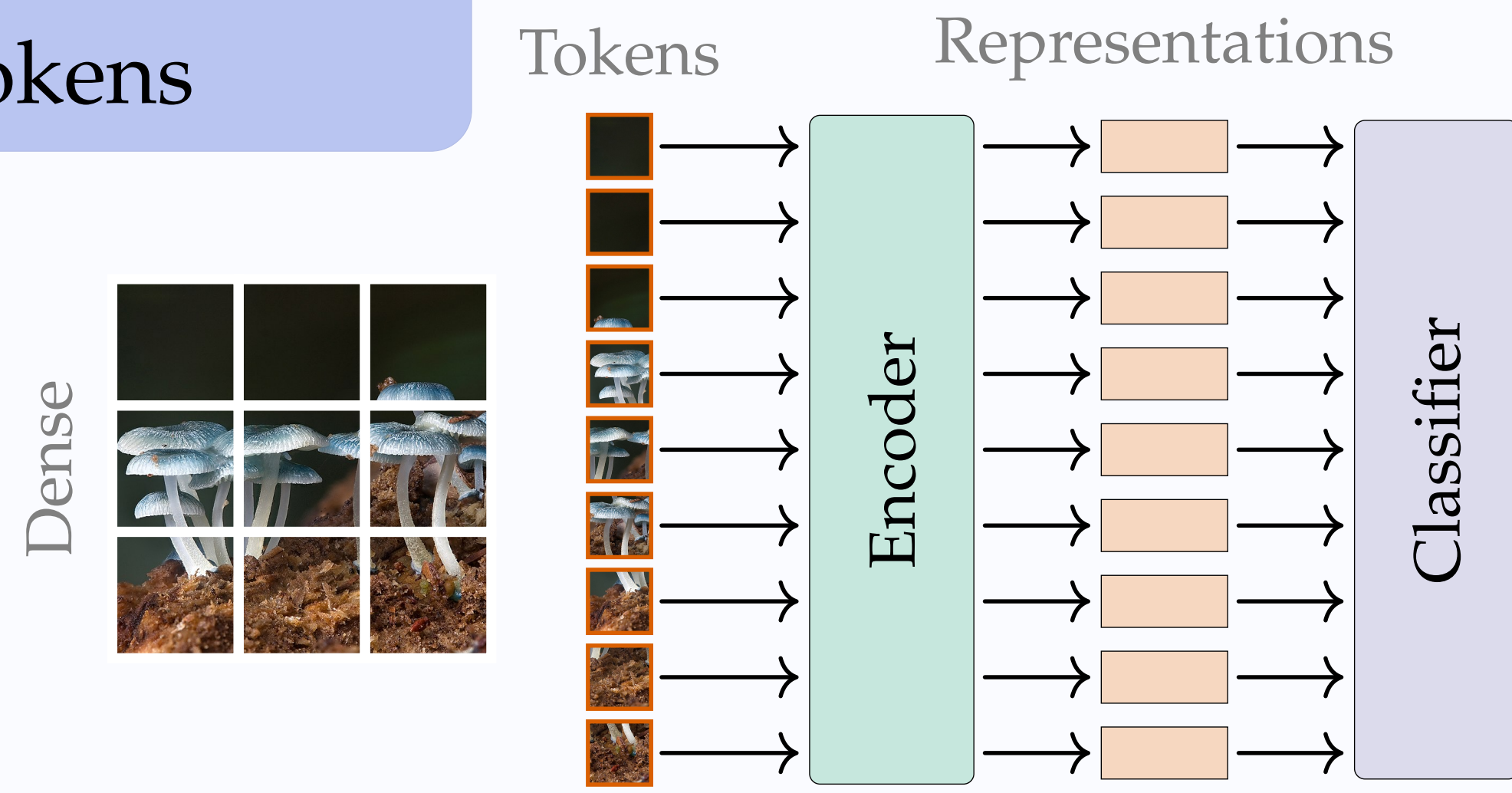
SPoT: Subpixel Placement of Tokens in Vision Transformers

Martine Hjelkrem-Tan, Marius Aasan,
Gabriel Y. Arteaga and Adín Ramírez Rivera
University of Oslo, SFI Visual Intelligence

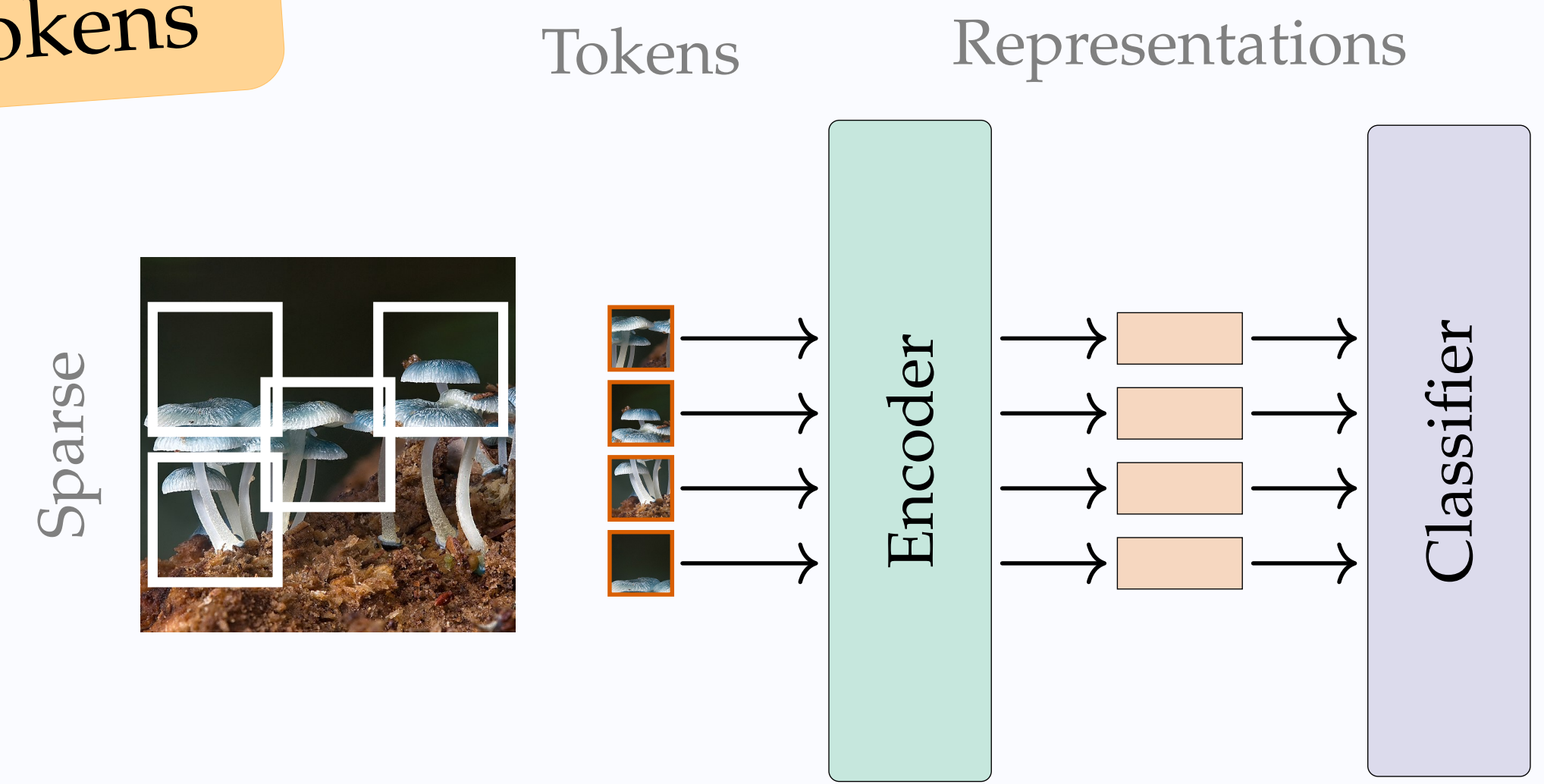


[More info](#)

Current: Dense grid tokens



Goal: An optimal sparse set of tokens

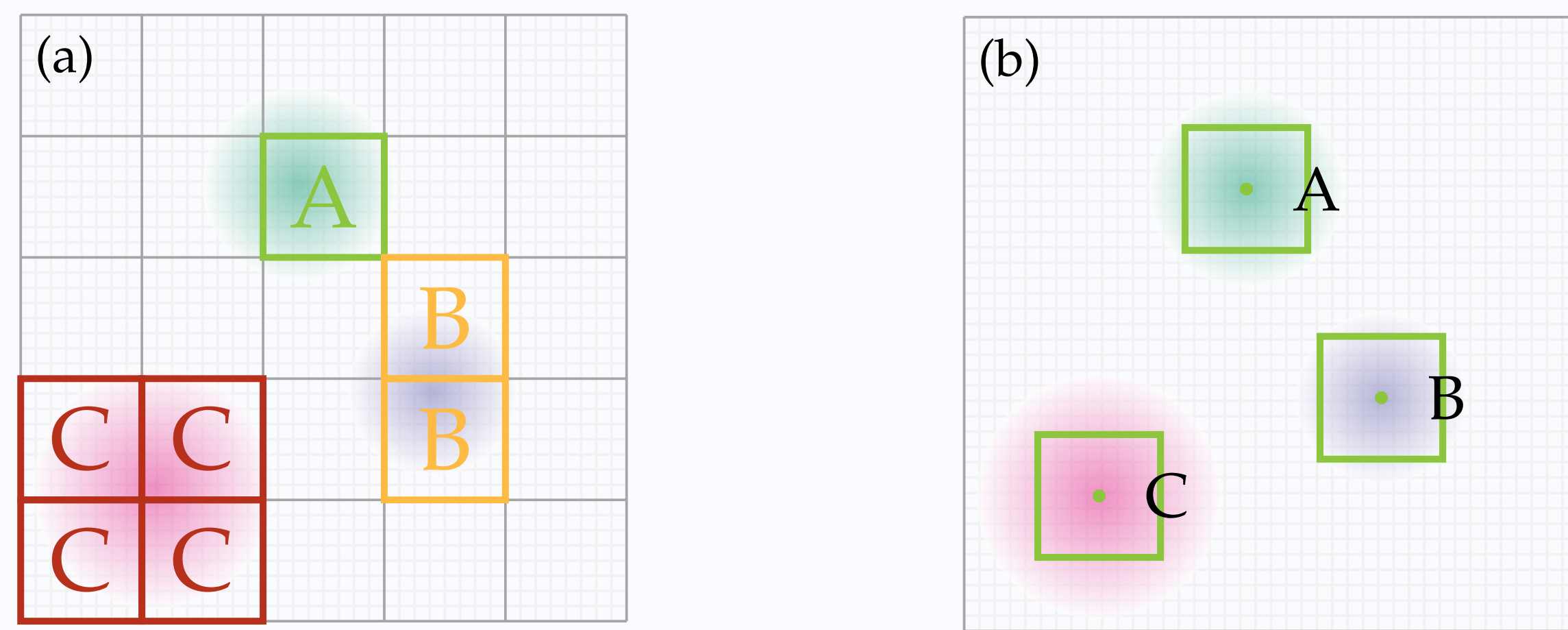


Motivation

Why should tokens be limited to a discrete patch grid?

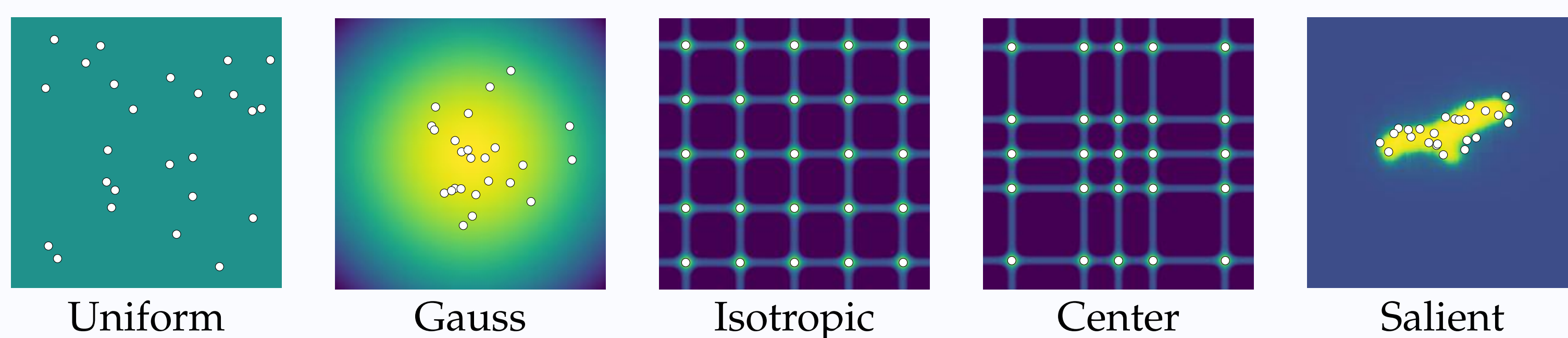
- **Sparse Feature Selection:** Fewer, but better tokens → fast and efficient inference
- **Limitation:** ViTs are constrained by tokens fixed on a discrete grid
- **Our Solution:** Sparse token sampling at *continuous subpixel positions*
- **Benefits:** Reduced inference cost and the ability to learn optimal token sampling location.

Subpixel Placement of Tokens



Fixed grids need more tokens to cover the salient regions due to misalignment.

Sampling placements from a prior



SPoT-ON: Learning placements with oracle neighborhoods

Gradient search for ideal token placements $S = \{s_q, \dots, s_m\}$ for each image I and label y

$$\arg \min_S [\mathcal{L}(g_\theta(I, S), y)] \text{ s.t. } S \subseteq \Omega_{\text{subpixel}}, |S| = m.$$

Reveals ideal locations for classifying each image → *tool for analyzing performance.*

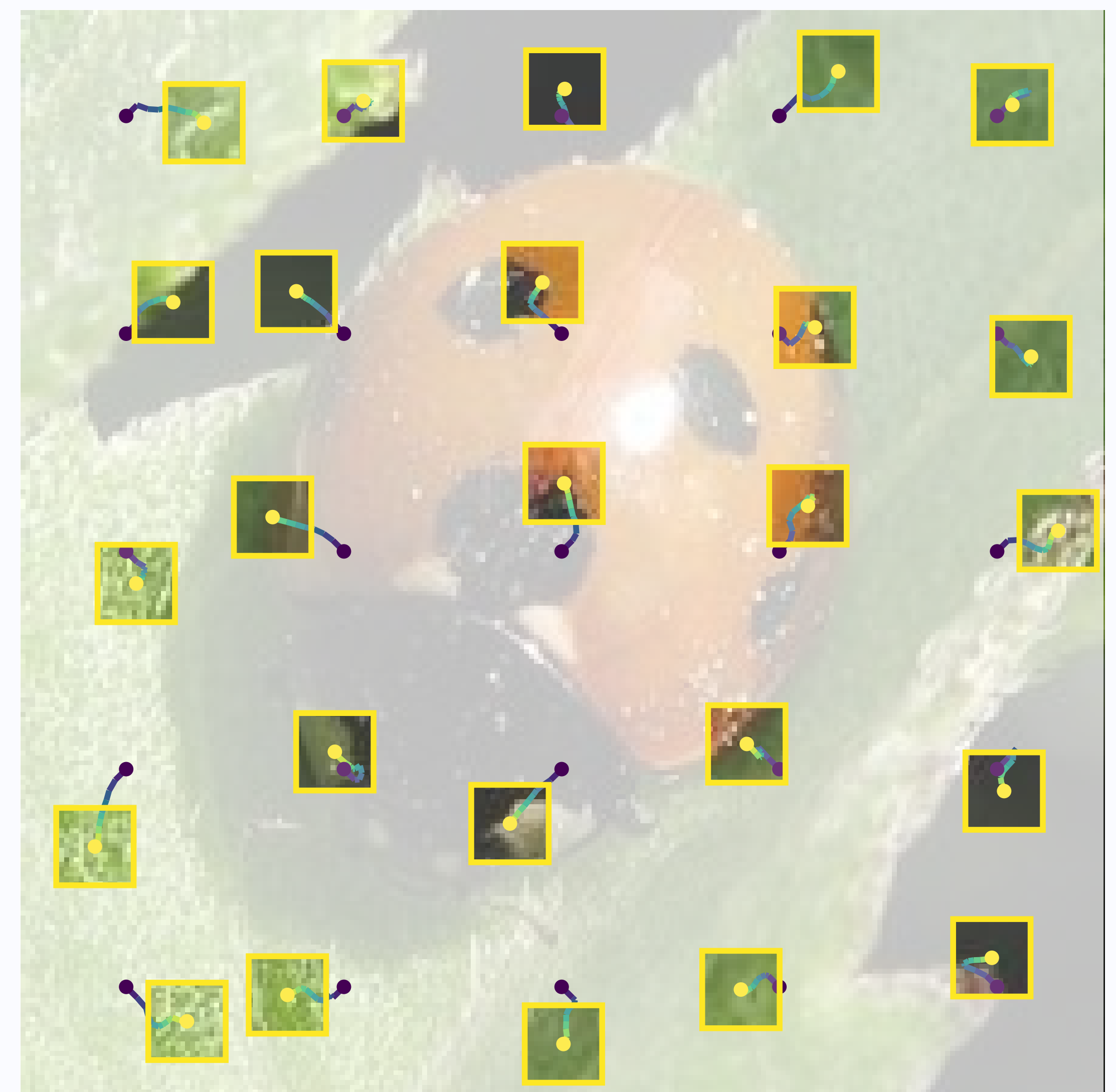
Classification performance for different token priors. Center-biased priors are beneficial in sparse regimes, while coverage becomes more important as token budgets increase.

Model	Prior	Oracle	25 Tokens		49 Tokens		100 Tokens		196 Tokens	
			Acc@1	kNN	Acc@1	kNN	Acc@1	kNN	Acc@1	kNN
Baseline	Patch Grid		24.72	27.86	56.29	57.19	78.75	78.77	85.11	83.96
SPoT	Uniform		44.05	45.23	67.77	66.38	79.64	78.03	83.76	81.85
SPoT	Gaussian		45.22	45.27	68.64	66.96	79.75	77.74	83.45	81.48
SPoT	Sobol		43.67	46.48	69.02	68.60	81.63	79.35	84.66	82.62
SPoT	Isotropic		46.85	48.19	70.61	70.29	82.20	80.73	85.15	83.42
SPoT	Center		52.45	52.18	69.22	68.16	80.84	78.56	84.01	82.23
SPoT	Salient	✓	55.71	56.65	72.89	72.38	79.91	80.56	84.56	82.59
SPoT	Isotropic	✓	81.70	70.65	94.28	88.58	95.97	92.92	96.12	93.52

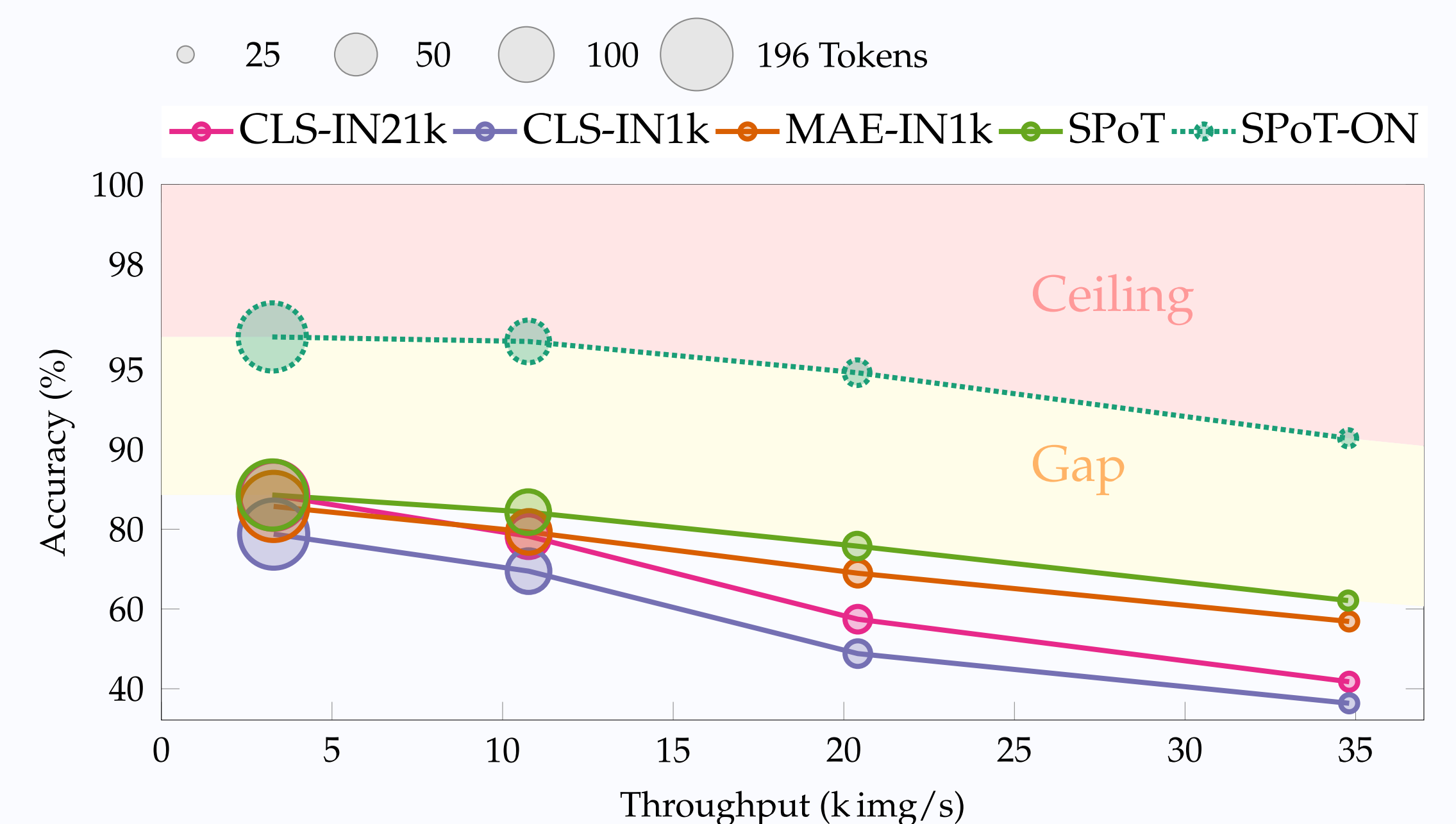
Findings from case studies

- ✓ *Off-grid* > *grid-based* approaches under sparse token settings.
- ✓ Sparse regimes → object-centric placements is better.
Dense regimes → structured coverage is better.
- ✓ Oracle placements do not significantly correlate with saliency.
- ✓ Good placements transfer between models.

Does the oracle prefer salient regions?



Throughput vs. Accuracy



Gap = Possible performance gain through better token placement.

Ceiling = Performance unlikely to be achieved ∴ intrinsic label noise.