Bootstrapping Top-down Information for Self-modulating Slot Attention

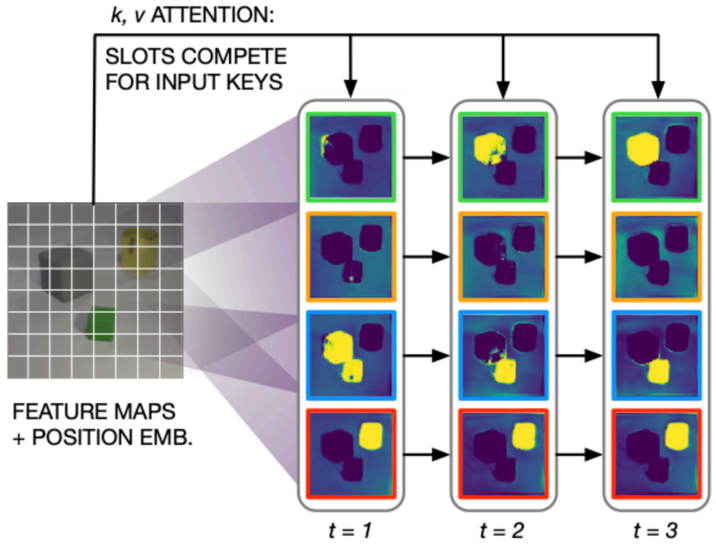
Dongwon Kim, Seoyeon Kim, Suha Kwak
POSTECH

Dongwon Kim, NeurIPS 2024

Task: Object-centric learning

What is object-centric learning?

- Task of learning representations of individual objects from visual scenes without manual labels
- Slot-attention: compute per-object representations (slot) through clustering-like attention

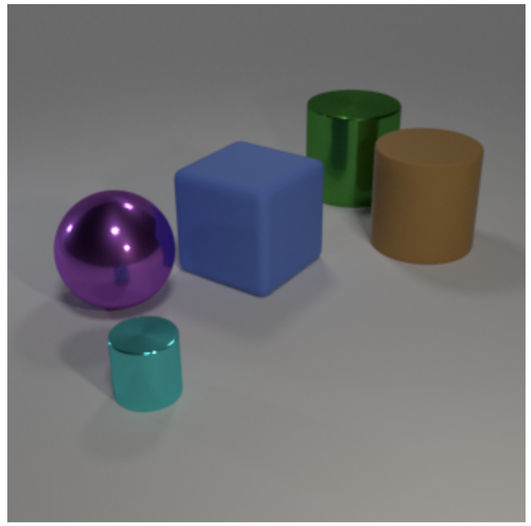


Object-Centric Learning with Slot Attention, NeurIPS 2020

Motivation: Incorporating top-down approaches

Conventional OCLs are bottom-up methods

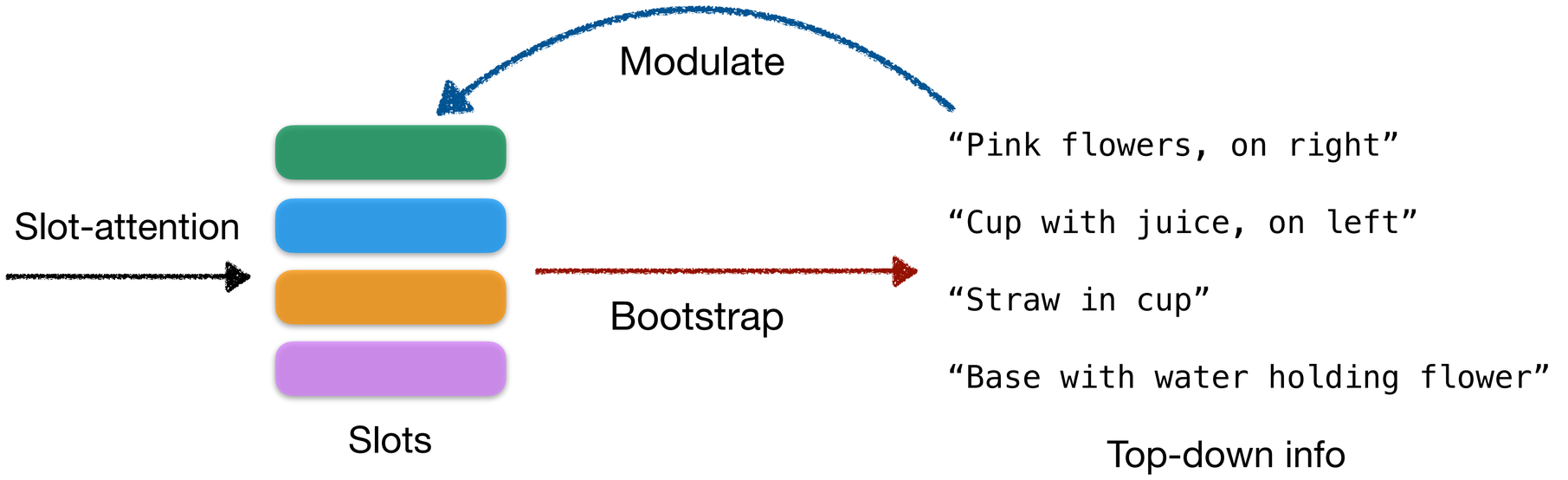
- Slot attention can be considered as a ***bottom-up*** method; it relies on aggregating features without high-level semantic information.
- It assumes that features within an object are homogeneous & can be clustered.
  - But this does not always hold for the real-world images



Motivation: Incorporating top-down approaches

Can we bootstrap top-down info, and then use it to modulate model?

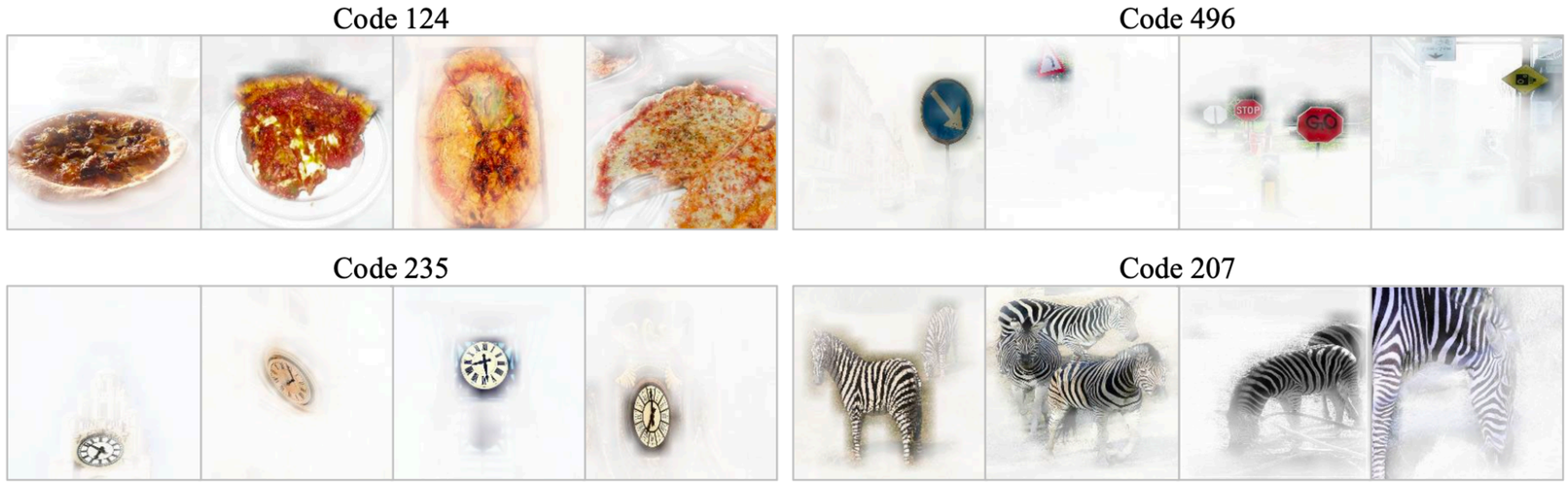
- Incorporating top-down information can fix the issue
  - Problem will be much easier if we can provide model about the semantic information or rough location of each object



Bootstrapping top-down information

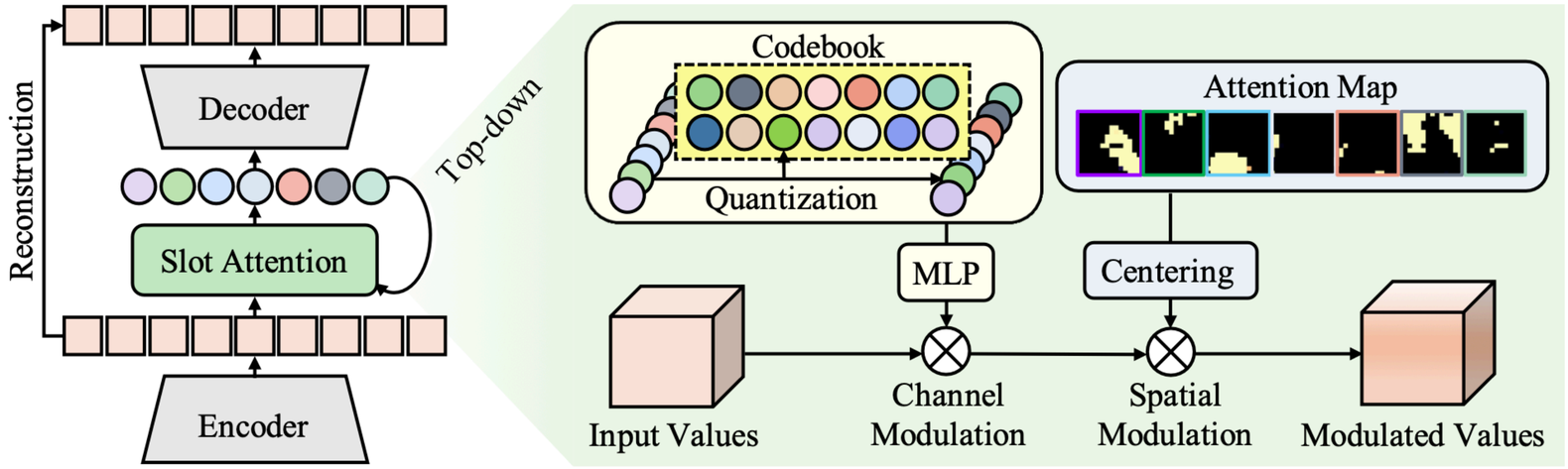
- During training, the codebook of slots is learned together using vector quantization.
  - Codebook learns to store distinct semantic patterns, where each code can act as automatically discovered top-down semantic information.

.



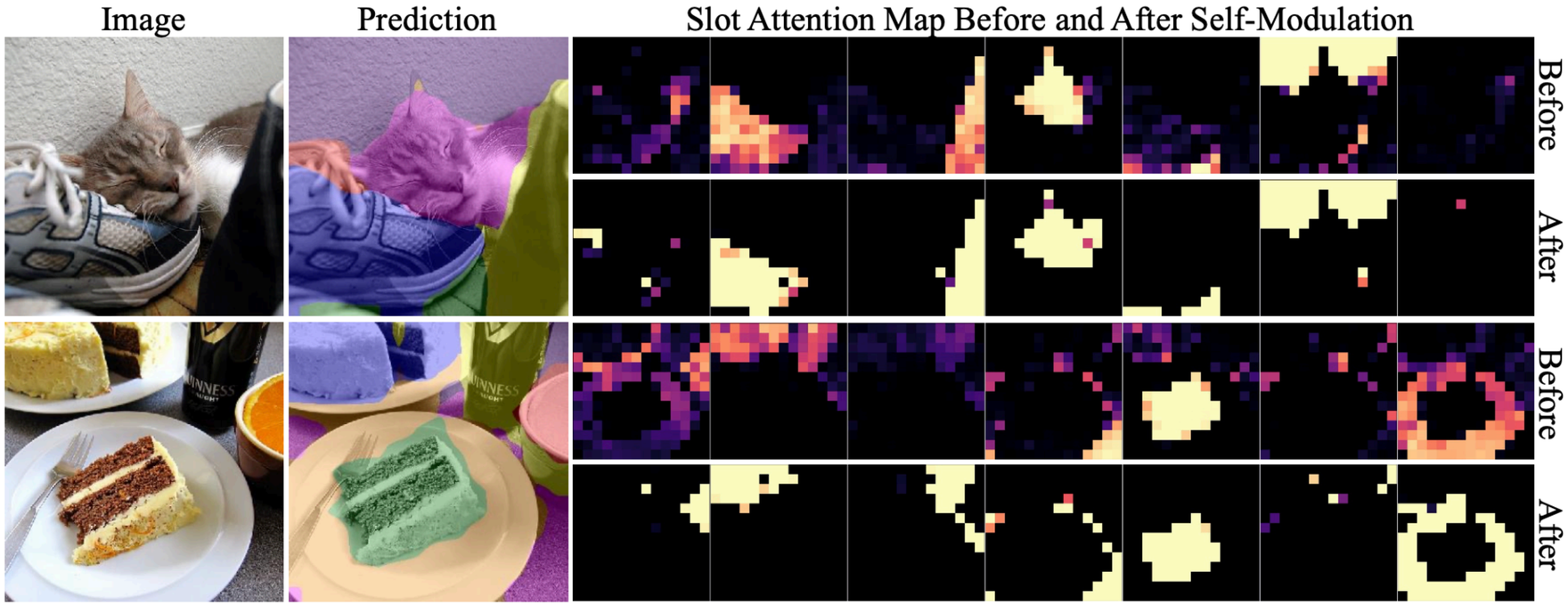
Self-modulating slot attention

- Using bootstrapped top-down information, input value of the slot attention is modulated.
  - The modulation guides the update of the slots, by prioritizing visual features relevant to top-down information



Results

- Our proposed top-down pathway largely improved the quality of object-centric representation in the real-world images.



Thank you!

Visit us: Wed 11 Dec. 11:00 AM — 2:00 PM