Online Object Representations with Contrastive Learning in Videos

Soren Pirk, Mohi Khansari, Yunfei Bai, Corey Lynch, Pierre Sermanet

Objective

- Self-teach to discover and disentangle object attributes from videos without using any labels.
- Use of online adaptation: the longer our online model looks at objects in a video, the lower the object identification error.
- Explore system free of human supervision for robotics applications. A robot collects its own data, trains on it, and then identifies objects.

Datasets

Real Data for Online Training (Complex Scenes, Epic Kitchens)
Automatic Real Data Collection with Robot
Synthetic Data for Evaluation

Approach

- Detect and embed objects to extract their features.
- Use metric loss to contrast similar and dissimilar objects in embedding space.
- Observing objects across different views facilitates learning invariance to scene-specific properties, such as scale, occlusion, lighting, or background.

Experiments

Online Object Identification (red boxes indicate mismatches)
Object Attribute Classification: Comparison to Baselines

Conclusion

Self-supervised online learning of object representations, particularly useful for robotics to increase robustness and adaptability to unseen objects.

Paper and Videos available here: https://online-objects.github.io/