**1. Introduction**

USS

- 초기: image의 two view 간의 mutual information 최대화

- 최근: self-supervised learning으로 patch-level semantic 학습한 per-trained model 사용(STEGO)

- 문제: ViT based method는 patch-level feature 학습 -> 복잡한 feature representation 갖는 object-level understanding 부족함

**2. Related Work**

**2.1. Unsupervised Semantic Segmentation**

- 초기: mutual information 최대화 (IIC, AC)

- (InfoSeg, PiCIE)

- ViT(DINO) top-down feature extraction (STEGO: DINO backbone distillation) (HP: patch-level contrastive learning)

**2.2. Spectral Techniques for Segmentation**

- spectral graph methods: Laplacian matrix의 eigenvector, eigenvalue 이용해 graph property 얻음

- image segmentation -> pixel 간 유사도를 함축하고 있는 affinity matrix를 어떻게 선택할 지가 중요함

- Deep Spectral Methods: feature affinity matrix로부터 얻은 Laplacian eigenvector 사용 / EigenFunction: network로 learnable eigenfunction 학습

**2.3. Object-centric Contrastive Learning**

- patch-level representation learning은 fine detail에 집중해 high-level concept 학습 부족함 -> object-level contrastive learning methods

- unsupervised saliency (MaskContrast, COMUS) / K-means clustering, heuristic masks 이용해 sample 생성 (Odin, DetCon) -> SlotCon: pixel이 slot 학습 / DINOSAUR: decoder에 input 대신 self-supervised features 입력

- 문제: slot에 의존해 high-level image features 놓칠 수 있음