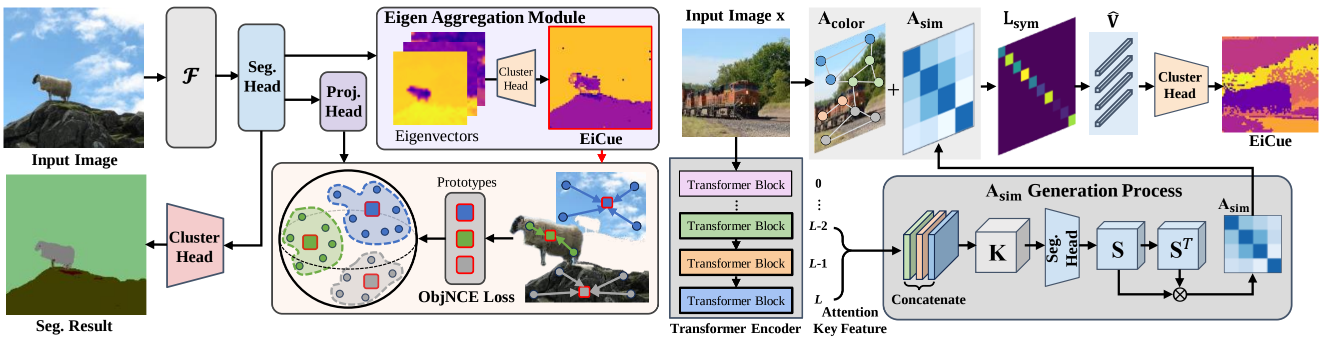
EAGLE: Eigen Aggregation Learning for Object-Centric Unsupervised Semantic Segmentation

Chanyoung Kim* Woojung Han* Dayun Ju Seong Jae Hwang
Yonsei University



(a) Overview of *EAGLE* **(b) Eigen Aggregation Module**

Figure 1. We introduce ***EAGLE***, ***Ei**gen **AG**gregation **LE**arning for object-centric unsupervised semantic segmentation*. (a) We first leverage the aggregated eigenvectors, named EiCue, to obtain the semantic structure knowledge of object segments in an image. Based on both semantic and structural cues from the EiCue, we compute object-centric contrastive loss to learn object-level semantic representation. (b) An illustration of the EiCue generation process. From the input image, both color affinity matrix $\mathbf{A}_{\text{color}}$ and semantic similarity matrix $\mathbf{A}_{\text{seg}}$ are derived, which are combined to form the Laplacian $\mathbf{L}_{\text{sym}}$. An eigenvector subset $\hat{\mathbf{V}}$ of $\mathbf{L}_{\text{sym}}$ are clustered to produce EiCue.

Introduction. Semantic segmentation has innately relied on extensive pixel-level annotated data, leading to the emergence of unsupervised methodologies. Among them, leveraging self-supervised Vision Transformers for unsupervised semantic segmentation (USS) [2, 5] has been making steady progress with expressive deep features. Yet, for semantically segmenting images with complex objects, a predominant challenge remains: the lack of explicit object-level semantic encoding in patch-level features particularly with diverse structures. To address this gap, we present a novel approach, which emphasizes *object-centric representation learning* for unsupervised semantic segmentation, named ***EAGLE***.

Spectral Techniques for Object-centric Perspective. We introduce EiCue, a spectral technique providing semantic and structural cues through an eigenbasis derived from the semantic similarity matrix. We use the Spectral Clustering [1, 4, 6] to obtain unsupervised feature representations that capture the underlying non-linear structures for handling data with complex patterns. This classically operates only in the color space but may easily extend to utilize the similarity matrix constructed from any features. The overall EiCue generation process follows the following procedure: (1) from an adjacency matrix $\mathbf{A}$ (a combination of color space affinity matrix $\mathbf{A}_{\text{color}}$ and deep feature similarity matrix $\mathbf{A}_{\text{sim}}$), (2) construct the graph Laplacian $\mathbf{L}$ from $\mathbf{A}$, and (3) perform the eigendecomposition on $\mathbf{L}$ to derive the eigenbasis $\mathbf{V}$ from which the eigenfeatures are used for the differentiable clustering, resulting in EiCue.

Object-Centric Contrastive Learning. Further, we incorporate ObjNCELoss, a newly designed EiCue-based object-centric contrastive loss method designed to refine feature embeddings through capturing complex inter-object relationships and enhancing feature discriminability for improved semantic segmentation. This loss function uses EiCue to capture complex inter-object relationships, enhancing feature discriminability and establishing object-level prototypes in a projected semantic space. These prototypes act as anchors, attracting similar features while repelling dissimilar ones, thereby promoting semantic clarity among detected object classes in the image. Through a comprehensive learning process, our model effectively captures inherent structures within images, allowing it to precisely identify semantically plausible object representations, the key to advancing feature-based USS.

Experiments. We evaluate *EAGLE* on COCO-Stuff, Cityscapes, and Potsdam-3 datasets to demonstrate the state-of-the-art USS results with accurate and consistent semantic segmentation.

*Equal contribution
Original Paper: Kim and Han *et al.* CVPR 2024 [3]

References

[1] Jeff Cheeger. A lower bound for the smallest eigenvalue of the laplacian. In *Problems in Analysis: A Symposium in Honor of Salomon Bochner (PMS-31)*, pages 195–200. Princeton University Press, 2015. 1

[2] Mark Hamilton, Zhoutong Zhang, Bharath Hariharan, Noah Snavely, and William T. Freeman. Unsupervised semantic segmentation by distilling feature correspondences. In *International Conference on Learning Representations*, 2022. 1

[3] Chanyoung Kim, Woojung Han, Dayun Ju, and Seong Jae Hwang. Eagle: Eigen aggregation learning for object-centric unsupervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. 1

[4] Andrew Ng, Michael Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. *Advances in Neural Information Processing Systems*, 14, 2001. 1

[5] Hyun Seok Seong, WonJun Moon, SuBeen Lee, and Jae-Pil Heo. Leveraging hidden positives for unsupervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19540–19549, 2023. 1

[6] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000. 1