UNIVERSITAS
GADJAH MADA
Unsupervised Panoptic Segmentation dengan Komparasi Backbone Berbasis CNN dan Transformer
William Surya Putra, Ika Candradewi, S.Si., M.Cs
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

INTISARI

Unsupervised Panoptic Segmentation dengan Komparasi Backbone Berbasis CNN dan Transformer

Oleh

William Surya Putra
22/494211/PA/21260

Anotasi manual pada segmentasi panoptik *supervised learning* memakan biaya dan waktu tinggi, sehingga menghambat pengembangan sistem segmentasi pada lingkungan dinamis. Penelitian ini mengimplementasikan *framework Unsupervised Universal Segmentation* (U2Seg) sebagai alternatif *unsupervised panoptic segmentation* untuk mengeliminasi kebutuhan anotasi manual. Metode penelitian dilakukan dengan mengekstraksi fitur menggunakan *backbone* ResNet-50-*DIstillation with NO labels* (DINO) dan Swin Transformer Tiny, yang kemudian diintegrasikan dengan *pseudo-label* MaskCut untuk segmentasi *instance* dan *Self-supervised Transformer with Energy-based Graph Optimization* (STEGO) untuk segmentasi semantik. Model tersebut dievaluasi pada dataset COCO 2017 untuk membandingkan akurasi dan efisiensi komputasi antar *backbone* pada format PyTorch dan *Open Neural Network Exchange* (ONNX). Hasil penelitian menunjukkan ResNet-50 DINO mencapai *Panoptic Quality* (PQ) 15.5, lebih unggul dari Swin Tiny yang memperoleh 14.5. Meski ResNet-50 DINO lebih optimal untuk kategori *things*, Swin Tiny menunjukkan keunggulan pada kategori *stuff* berkat mekanisme *Shifted Window Attention*. Selain itu, format ONNX menyebabkan penurunan efisiensi seragam pada kedua *backbone* akibat keterbatasan perangkat keras lokal. Penelitian ini menunjukkan bahwa pemilihan *backbone* memberikan pengaruh yang berbeda terhadap performa segmentasi panoptik pada kategori *things* dan *stuff* maupun efisiensi komputasi antar format model.

Kata Kunci—Segmentasi Panoptik, *Unsupervised Panoptic Segmentation*, U2Seg, *Backbone, Pseudo-Label*

UNIVERSITAS
GADJAH MADA
Unsupervised Panoptic Segmentation dengan Komparasi Backbone Berbasis CNN dan Transformer
William Surya Putra, Ika Candradewi, S.Si., M.Cs
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

ABSTRACT

Unsupervised Panoptic Segmentation with a Comparison of CNN-Based and Transformer-Based Backbones

by

William Surya Putra
22/494211/PA/21260

Manual annotation in supervised panoptic segmentation is costly and time-consuming, hindering the development of segmentation systems for dynamic environments. This study implements the Unsupervised Universal Segmentation (U2Seg) framework as an alternative approach to eliminate the need for manual labels. The research methodology involved feature extraction using ResNet-50 DIstillation with NO labels (DINO) and Swin Transformer Tiny backbones, which were integrated with MaskCut pseudo-labels for instance segmentation and Self-supervised Transformer with Energy-based Graph Optimization (STEGO) for semantic segmentation. These models were evaluated on the COCO 2017 dataset to compare accuracy and computational efficiency across PyTorch and Open Neural Network Exchange (ONNX) formats. Results indicate that ResNet-50 DINO achieved a Panoptic Quality (PQ) score of 15.5, outperforming Swin Tiny at 14.5. While ResNet-50 DINO proved more optimal for the 'things' category, Swin Tiny demonstrated an advantage in the 'stuff' category, attributed to its Shifted Window Attention mechanism. Furthermore, the ONNX format resulted in a uniform decrease in efficiency for both backbones due to hardware limitations. This research shows that backbone selection has a different influence on panoptic segmentation performance across the things and stuff categories, as well as on computational efficiency across model formats.

Keyword—Panoptic Segmentation, Unsupervised Panoptic Segmentation, U2Seg, Backbone, Pseudo-Label