

BAB II

KAJIAN PUSTAKA

2.1 Dasar Teori

2.1.1 Segmentasi Semantik

Segmentasi semantik merupakan proses pengolahan citra yang bertujuan untuk mengklasifikasikan setiap piksel dalam gambar ke dalam kategori objek yang bermakna (Haq & Huy, 2024). Teknik ini menjadi fundamental dalam berbagai aplikasi komputer vision, khususnya pada sistem kendaraan otonom, pemantauan perkotaan, dan analisis lingkungan. Dalam segmentasi semantik, model memisahkan gambar menjadi beberapa segmen berdasarkan objek yang terkandung di dalamnya, seperti jalan, kendaraan, pejalan kaki, dan rambu lalu lintas (Helnawan et al., 2023). Hal ini memungkinkan sistem untuk memahami struktur lingkungan secara lebih mendalam pada tingkat piksel, yang sangat penting dalam aplikasi yang membutuhkan ketelitian tinggi, seperti deteksi objek kecil seperti marka jalan atau pejalan kaki (Helnawan et al., 2023).

Namun, salah satu tantangan utama dalam segmentasi semantik adalah pergeseran domain (domain shift), yaitu perbedaan distribusi antara data pelatihan dan data dunia nyata. Data pelatihan yang biasa digunakan, seperti dataset sintesis dari GTA5 atau Synthia, seringkali memiliki kondisi visual yang sangat berbeda dengan data dunia nyata, seperti variasi pencahayaan, permukaan jalan yang tidak rata, dan adanya objek yang tidak ada dalam dataset sintesis, seperti pedagang kaki lima dan kendaraan yang parkir sembarangan (Shen et al., 2023). Oleh karena itu, segmentasi semantik yang efektif harus mampu mengatasi pergeseran domain ini agar dapat diimplementasikan pada data dunia nyata dengan akurasi yang tinggi (Nugroho et al., 2023).

2.1.2 Domain Adaptation

Domain adaptation adalah pendekatan dalam pembelajaran mesin yang memungkinkan model untuk beradaptasi dengan data yang berasal dari domain berbeda dengan data pelatihan (Shen et al., 2023). Dalam konteks segmentasi semantik, domain adaptation memungkinkan model yang dilatih dengan dataset sintesis (seperti GTA5 atau Synthia) untuk mengatasi perbedaan besar antara data

pelatihan dan data dunia nyata, misalnya jalanan yang ada di Indonesia (Bai et al., 2024). Teknik ini sangat penting ketika data dunia nyata sulit didapatkan dalam jumlah besar atau tidak representatif terhadap kondisi yang ada di lapangan (Bai et al., 2024).

Metode yang umum digunakan dalam domain adaptation adalah transfer learning, yang memungkinkan model untuk memanfaatkan pengetahuan yang telah diperoleh dari satu domain dan menerapkannya pada domain lain (Pylianidis et al., 2024). Dalam domain adaptation untuk segmentasi semantik, strategi ini dapat meliputi teknik-teknik seperti fine-tuning, di mana model yang telah dilatih pada data sintetis kemudian disesuaikan dengan data dunia nyata untuk meningkatkan kinerjanya (Wang et al., 2024). Salah satu pendekatan yang banyak digunakan adalah penggunaan Optimal Transport (OT) yang bertujuan untuk menyesuaikan distribusi kategori antara domain sumber dan domain target, sehingga memungkinkan model untuk belajar dari data dunia nyata meskipun pelatihan dilakukan pada data sintetis (Guo et al., 2024).

2.1.3 Karakteristik Jalan di Surabaya

Jalan di Indonesia, khususnya di kawasan urban seperti Surabaya, memiliki karakteristik yang sangat berbeda dibandingkan dengan kondisi jalan yang ada di negara maju atau dalam dataset sintetis yang sering digunakan. Jalanan di Indonesia sering kali memiliki permukaan yang tidak rata, dengan variasi material yang sangat berbeda (aspal, beton, kerikil) yang mengarah pada perbedaan dalam tekstur dan ketahanan jalan. Selain itu, lebar jalan juga sangat bervariasi, mulai dari jalan yang sangat sempit hingga jalan yang sangat lebar di area perkotaan dan kawasan pedesaan (Khairunnisa & Buana, 2023).

Keberagaman objek yang ada di jalan juga merupakan tantangan tersendiri. Di beberapa ruas jalan, terdapat pedagang kaki lima yang memanfaatkan trotoar atau bahkan sebagian jalan, serta parkir kendaraan yang menghalangi arus lalu lintas. Hal ini tentu mempengaruhi performa model segmentasi semantik, karena model harus mampu mengidentifikasi objek seperti pejalan kaki, kendaraan yang terparkir, serta elemen-elemen jalan yang tidak terduga (Haq et al., 2022). Variasi dalam kondisi cuaca juga mempengaruhi citra

jalanan, misalnya saat hujan yang menyebabkan permukaan jalan basah atau saat kemarau yang membuat debu menutupi jalan. Pencahayaan yang berubah-ubah, baik di siang hari maupun malam hari, serta adanya perubahan visual yang disebabkan oleh kondisi cuaca dan waktu, memperburuk pergeseran domain ini (Shao et al., 2022).



Gambar 2. 1 Google Street View



Gambar 2. 2 Google Street View



Gambar 2. 3 Google Street View

Oleh karena itu, penting bagi penelitian ini untuk tidak hanya mengandalkan dataset sintetis, tetapi juga mengumpulkan data jalanan lokal di Surabaya yang mewakili kondisi jalan di Indonesia untuk pelatihan model segmentasi semantik.

Data lengkapnya ada di google drive ini :

https://drive.google.com/drive/folders/1CgypWScJKQ6mMmcVgEDPbw0WLGf5X5Xh?usp=drive_link

2.1.4 Metode/ Teori/ Algoritma

Dalam penelitian ini, kami menguji model utama untuk segmentasi semantik domain-adaptif: **SegFormer-R101**.

- **SegFormer-R101** juga berbasis transformer dan menggunakan **encoder hierarkis** yang dapat menangkap konteks global dan detail spasial dengan efisien (Katsamenis et al., 2023). Dengan penggunaan **decoder MLP** yang ringan, SegFormer-R101 tidak hanya mengurangi biaya komputasi tetapi juga mempertahankan akurasi tinggi pada berbagai kategori, seperti trotoar dan langit. Model ini menunjukkan keunggulannya dalam menghadapi variasi tekstur dan permukaan yang kontinu pada citra jalanan.

2.2 Penelitian Terdahulu

Beberapa penelitian telah dilakukan terkait dengan masalah segmentasi semantik domain-adaptif pada gambar resolusi tinggi. Penelitian-penelitian ini fokus pada pengembangan model yang dapat menangani pergeseran domain dan adaptasi pada data dunia nyata.

Tabel 1 Penelitian Terdahulu

No	Previous Research Papers	Author	Metode	Results
1	Domain Adaptation through Photorealistic Enhanced Images for Semantic Segmentation (Katayama et al., 2022)	Takafumi Katayama, Tian Song, Xiantao Jiang, Jenq-Shiou Leu, and Takashi Shimamoto	Image-level Domain Adaptation (Photorealism Enhancement); ADVENT; IntraDA; DeepLab-v2 (ResNet-101)	Kerangka kerja ini menghasilkan akurasi maksimum (mIoU) sebesar 45,0%, yang meningkat menjadi 47,5% ketika IntraDA ditambahkan, mengungguli metode pembandingan sebelumnya sebesar 1,2 poin. Peningkatan akurasi paling signifikan terjadi pada kelas trotoar dan rambu di Cityscapes.

No	Previous Research Papers	Author	Metode	Results
2	Image Segmentation from Sparse Decomposition with a Pretrained Object Detection Network (Wu et al., 2022)	Yulin Wu, Chuandong Lv, Baoqing Ding, Lei Chen, Bin Zhou, Hongchao Zhou	Object Location Segmentation (OLSeg); pretext task sparse decomposition; autoencoder; U-Net; YOLOv4 (pretrained object-detection)	OLSeg mencapai kinerja superior pada set data internet (Pesawat P/J 91,87/61,50, Mobil 90,23/68,45, Kuda 89,56/58,72) dan rata-rata P/J 87,56/65,85 pada MSRC, mengungguli metode benchmark. Secara kualitatif, model ini menghasilkan batas segmentasi yang lebih jelas dan latar belakang yang lebih bersih, yang menunjukkan generalisasi yang baik di seluruh kelas objek.
3	Domain adaptive semantic segmentation by optimal transport (Guo et al., 2024)	Yaqian Guo, Xin Wang, Ce Li, Shihui Ying	Optimal Transport (OT), output-space alignment, attention mechanism (OT coupling), multilevel/multiscale adaptation, DeepLab-v2 (ResNet-101 + ASPP), and Sinkhorn optimization	Pada GTA5→Cityscapes, metode ini mencapai mIoU sebesar 44,3% (multi-OT), melampaui AdaptSegNet yang hanya 42,4%; pada SYNTHIA→Cityscapes, metode ini mencapai mIoU sebesar 48,8%, melampaui tolok ukur lainnya. Model ini juga lebih ringan, dengan 43,55 juta parameter dan 651,190 FLOP dibandingkan dengan AdaptSegNet yang hanya 49,11 dan 705,468 FLOP.

No	Previous Research Papers	Author	Metode	Results
4	Weakly-Supervised Domain Adaptation with Adversarial Entropy for Building Segmentation in Cross-Domain Aerial Imagery (Yao et al., 2021)	Xuedong Yao, Yandong Wang, Yanlan Wu, Zeyu Liang	Adversarial Entropy; Weakly-Supervised Self-Training dengan Pseudolabel; Entropy Alignment (Shannon's Entropy); Adversarial Learning (discriminator); DeepLab-V2 (ResNet-101)	Pada transfer Potsdam→Berlin dan Potsdam→Zurich, metode ini mencapai skor terbaik: Berlin F1/IoU 72,12%/57,04% dan Zurich F1/IoU 72,56%/57,17%, mengungguli nilai dasar masing-masing sebesar +19,18 dan +20,80 poin IoU. Pada transfer Potsdam→Chicago, kinerjanya tetap tertinggi, meningkatkan F1/IoU masing-masing sebesar +1,32/+1,64 poin di atas nilai dasar dan +19,42/+15,92 poin di atas AdaptSegNet.
5	Semantic segmentation for remote sensing images based on an AD-HRNet model (Yang et al., 2022)	Xue Yang, Xiang Fan, Mingjun Peng, Qingfeng Guan, Luliang Tang	HRNet; AD-HRNet; Shuffle-CBAM; Mixed Dilated Convolution (MDC); Dense Upsampling Convolution (DUC); Weighted Cross-Entropy (Median Frequency Balance)	AD-HRNet mencapai mIoU sebesar 75,59% di Potsdam dan 71,58% di Vaihingen. Dibandingkan dengan HRNet_W48, mIoU meningkat masing-masing sebesar +0,74% dan +0,73%, dengan segmentasi tepi yang lebih halus dan pengenalan objek kecil yang lebih baik.
6	HRDA: Context-Aware High Resolution Domain Adaptive Semantic Segmentation (Hoyer et al., 2022)	Lukas Hoyer, Dengxin Dai, Luc Van Gool	multi-resolution training; scale attention; LR context crop & HR detail crop; self-training (DAFormer); DACS; overlapping sliding-window inference; DeepLabV2/DAFormer backbone	HRDA mencapai 73.8 mIoU pada GTA→Cityscapes dan 65,8 mIoU pada Synthia→Cityscapes, melampaui SOTA sebelumnya masing-masing sebesar +5,5 dan +4,9 poin.

No	Previous Research Papers	Author	Metode	Results
7	Bounding Box-Free Instance Segmentation Using Semi-Supervised Iterative Learning for Vehicle Detection (De Carvalho et al., 2022)	Osmar Luiz Ferreira de Carvalho, Osmar Abílio de Carvalho Júnior, Anesmar Olino de Albuquerque, Nickolas Castro Santana, Renato Fontes Guimarães, Roberto Arnaldo Trancoso Gomes, Díbio Leandro Borges	Iterative Learning (GIS integration); Box-Free Instance Segmentation; U-Net (EfficientNet-B7); Interior–Border Separation; Expanded Border Algorithm (1-pixel edge restoration); Mask R-CNN (comparator)	Metode bebas kotak berbasis U-Net mencapai 82% IoU per piksel, mengungguli Mask R-CNN (72%) dan mencatat metrik per objek (presisi/recall) di atas 90%. Pada klasifikasi skala kota pada Dataset Kendaraan BSB, pipeline ini memetakan 122.567 kendaraan dan menunjukkan kinerja rata-rata sekitar 81,88 IoU pada set uji terpisah.
8	Performance Evaluation of Different Object Detection Models for the Segmentation of Optical Cups and Discs (Alfonso-Francia et al., 2022)	Gendry Alfonso-Francia, Jesus Carlos Pedraza-Ortega, Mariana Badillo-Fernández, Manuel Toledano-Ayala, Marco Antonio Aceves-Fernandez, Juvenal Rodriguez-Resendiz, Seok-Bum Ko, Saul Tovar-Arriaga	Mask R-CNN; MS R-CNN; Cascade Mask R-CNN; CARAFE; GCNet; SOLO; PointRend; MMDetection; ResNet-50 + FPN; multiscale training; AdamW; cosine annealing; anchor box configuration; evaluasi COCO (AP, IoU, F1, AUCPR)	Pada REFUGE, beberapa model mencapai AP 1.000 pada IoU = 0,50 dan Cascade Mask R-CNN terendah adalah 0,997; pada set lengkap, skor F1 ≈ 1,00 menunjukkan segmentasi yang hampir sempurna. Pada G1020, PointRend menunjukkan performa terbaik dengan AP 0,956 sementara SOLO mencapai 0,909, yang mengonfirmasi kinerja tinggi di seluruh set data.

No	Previous Research Papers	Author	Metode	Results
9	Comparison of Deep-Learning-Based Segmentation Models: Using Top View Person Images (Ahmed et al., 2020)	Imran Ahmed, Misbah Ahmad, Fakhri Alam Khan, Muhammad Asif	Fully Convolutional Network (FCN); U-Net; DeepLabV3	Model-model ini mencapai IoU sebesar 83%, 84%, dan 86% serta mIoU sebesar 80%, 82%, dan 84% untuk FCN, U-Net, dan DeepLabV3. Secara keseluruhan, DeepLabV3 memberikan akurasi terbaik tetapi sedikit lebih lambat daripada FCN, dan pelatihan pada data top-view meningkatkan kinerja dibandingkan dengan model yang telah dilatih sebelumnya.
10	MFSNet: Enhancing Semantic Segmentation of Urban Scenes with a Multi-Scale Feature Shuffle Network (Qian et al., 2024)	Xiaohong Qian, Chente Shu, Wuyin Jin, Yunxiang Yu, Shengying Yang	MFSNet (Multi-Scale Feature Shuffle Network); Pyramid Shuffle Module (PSM); Efficient Feature Aggregation Module (EFAM); ResNet101; progressive upsampling; pyramid pooling; channel & spatial attention; channel shuffle	Model ini mencapai 80,4% mIoU pada Pascal VOC 2012, 79,4% pada Cityscapes, dan 40,1% pada COCO-Stuff. Dibandingkan dengan DeepLabv3+, model ini berkinerja lebih baik dengan +2,0 poin mIoU pada Pascal VOC dan +0,4 poin mIoU pada Cityscapes, dengan MAcc mencapai hingga 89,3%.

2.3 Analisa Kebaruan Penelitian

Penelitian ini menawarkan kebaruan yang signifikan dengan penggunaan dataset jalan Indonesia yang dikumpulkan khusus dari jalan di Surabaya. Dataset ini memiliki karakteristik visual yang berbeda dengan dataset yang umumnya digunakan dalam penelitian sebelumnya, seperti Cityscapes atau GTA5. Adanya variasi dalam kondisi jalan, keberadaan objek yang tidak terduga (seperti pedagang kaki lima dan kendaraan parkir sembarangan), serta kondisi lingkungan yang berbeda (pencahayaan, cuaca, dan lalu lintas padat) menjadikan dataset ini unik dan penting untuk penelitian adaptasi domain.

Selain itu, penelitian ini juga berfokus pada pengujian dan evaluasi tiga model segmentasi semantik utama (SegFormer-R101) dalam konteks dunia nyata di Indonesia, yang memberikan kontribusi baru dalam pengembangan model yang lebih adaptif terhadap pergeseran domain dalam aplikasi kendaraan otonom dan pemantauan infrastruktur jalan.

