

TESIS

**PENINGKATAN PERFORMA DETEKSI OBJEK DI JALAN
PADA *AUTONOMOUS CAR* DENGAN MODIFIKASI
ALGORITMA U-NET**

*IMPROVING ROAD OBJECT DETECTION PERFORMANCE IN
AUTONOMOUS CAR WITH MODIFIED U-NET ALGORITHM*

**HERLINA ANWAR
D082222026**



**PROGRAM STUDI MAGISTER TEKNIK INFORMATIKA
DEPARTEMEN TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS HASANUDDIN
GOWA
2024**

PENGAJUAN TESIS

PENINGKATAN PERFORMA DETEKSI OBJEK DI JALAN PADA AUTONOMOUS CAR DENGAN MODIFIKASI ALGORITMA U-NET

Tesis

Sebagai Salah Satu Syarat untuk Mencapai Gelar Magister
Program Studi Teknik Informatika

Disusun dan diajukan oleh

HERLINA ANWAR

D082222026

Kepada

**FAKULTAS TEKNIK
UNIVERSITAS HASANUDDIN
GOWA
2024**

TESIS

PENINGKATAN PERFORMA DETEKSI OBJEK DI JALAN PADA *AUTONOMOUS CAR* DENGAN MODIFIKASI ALGORITMA U-NET

**HERLINA ANWAR
D082222026**

Telah dipertahankan di hadapan Panitia Ujian Tesis yang dibentuk dalam rangka penyelesaian studi pada Program Magister Teknik Informatika Fakultas Teknik Universitas Hasanuddin Pada tanggal 5 Agustus 2024 dan dinyatakan telah memenuhi syarat kelulusan

Menyetujui,

Pembimbing Utama



Prof. Dr. Ir. Indrabayu, S.T., M.T., M.Bus.Sys., IPM, ASEAN.Eng.
NIP. 19750716 200212 1 004

Pembimbing Pendamping



Prof. Dr.Eng. Intan Sari Areni, S.T., M.T.
NIP. 19750203 200012 2 002

Dekan Fakultas Teknik
Universitas Hasanuddin



Prof. Dr.Eng. Ir. Muhammad Isran Ramli, M.T. IPM., ASEAN.Eng.
NIP. 19730926 200012 1 002

Ketua Program Studi
S2 Teknik Informatika



Dr. Ir. Zahir Zainuddin, M.Sc.
NIP. 19640427 198910 1 002

PERNYATAAN KEASLIAN TESIS DAN PELIMPAHAN HAK CIPTA

Yang bertanda tangan di bawah ini

Nama : Herlina Anwar

Nomor mahasiswa : D082222026

Program studi : Magister Teknik Informatika

Dengan ini menyatakan bahwa, tesis yang berjudul “PENINGKATAN PERFORMA DETEKSI OBJEK DI JALAN PADA AUTONOMOUS CAR DENGAN MODIFIKASI ALGORITMA U-NET” adalah benar karya saya dengan arahan dari komisi pembimbing Prof. Dr. Ir. Indrabayu, ST., MT., M.Bus.Sys., IPM, ASEAN, Eng. dan Prof. Dr. Eng. Ir. Intan Sari Areni, S.T., M.T., IPU. Karya ilmiah ini belum diajukan dan tidak sedang diajukan dalam bentuk apapun kepada perguruan tinggi manapun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka tesis ini. Sebagian dari isi tesis ini telah dipublikasikan di konferensi International (International Seminar on Intelligent Technology and Its Applications ISITIA 2024). Sebagai artikel dengan judul “A Modification of U-Net Decoder Architecture Improve Performance of Object Detection for Autonomous Vehicles”.

Dengan ini saya melimpahkan hak cipta ini dari karya tulis saya berupa tesis ini kepada Universitas Hasanuddin.

Gowa, 19 Agustus 2024

Yang menyatakan



Herlina Anwar

KATA PENGANTAR

Puji dan syukur penulis panjatkan atas kehadiran Allah SWT karena berkat Rahmat dan Karunia-Nya sehingga Tesis yang berjudul “PENINGKATAN PERFORMA DETEKSI OBJEK DI JALAN PADA *AUTONOMOUS CAR* DENGAN MODIFIKASI ALGORITMA U-NET” ini dapat diselesaikan sebagai salah satu syarat dalam menyelesaikan jenjang Strata-2 pada Departemen Teknik Informatika Fakultas Teknik Universitas Hasanuddin. Penulis menyadari bahwa dalam penyusunan dan penulisan tesis ini tidak lepas dari bantuan, bimbingan serta dukungan dari berbagai pihak, dari masa perkuliahan sampai dengan masa penyusunan tesis sangatlah sulit untuk menyelesaikan tugas akhir ini. Oleh karena itu, penulis dengan senang hati menyampaikan terima kasih kepada:

1. Tuhan Yang Maha Esa atas semua berkat, karunia serta pertolonganNya yangtelah diberikan kepada kami di setiap langkah dalam pembuatan program hingga penulisan laporan skripsi ini.
2. Orang tua penulis, Bapak Muh. Anwar dan Ibu Rahmawati yang tidak pernah lelah dalam memberikan dukungan, doa, dan semangat kepada penulis.
3. Prof. Dr. Ir. Indrabayu, S.T., M.T., M.Bus.Sys., IPM, ASEAN, Eng., selaku pembimbing 1 dan Ketua Departement Teknik Informatika, yang telah mendukung, memotivasi dan membantu selama penulis menempuh pendidikan pascasarjana di Universitas Hasanuddin.
4. Prof. Dr. Eng. Ir. Intan Sari Areni, S.T., M.T., IPU, selaku pembimbing II yang selalu menyediakan waktu, tenaga, pikiran dan perhatian yang luar biasa untuk mengarahkan penulis dalam penyusunan tugas akhir.
5. Bapak Dr. Ir. Zahir Zainuddin, M.Sc., selaku Ketua Program Studi Magister Teknik Informatika atas segala bimbingan dan dukungan selama masa perkuliahan.
6. Ibu Anugryani Bustami S.T., M.T., dan ibu Elly Warni, S.T., M.T., serta Segenap Dosen dan Staf Departemen Teknik Informatika Fakultas Teknik Universitas Hasanuddin yang telah banyak membantu semasa perkuliahan dan dalam menyelesaikan tesis.

7. Teman-teman AIMP research group Unhas yang telah memberikan begitu banyak bantuan selama penelitian, pengambilan data dan diskusi progress penyusunan tesis.
8. Teman-teman Mahasiswa S2 Angkatan 7 atas dukungan, bantuan, semangat yang diberikan selama ini.

Penulis menyadari bahwa tesis ini masih jauh dari kata sempurna, oleh karenanya diharapkan segala bentuk saran serta masukan yang membangun dari berbagai pihak. Semoga tesis ini dapat memberikan manfaat bagi semua pihak yang membaca.

Gowa, 7 Mei 2024

Penulis

ABSTRAK

HERLINA ANWAR. Peningkatan Performa Deteksi Objek Di Jalan Pada *Autonomous Car* Dengan Modifikasi Algoritma U-Net (dibimbing oleh **Indrabayu, Intan Sari Areni**)

Autonomous car adalah teknologi yang semakin populer, dimana ketepatan dalam deteksi objek di jalan sangat penting untuk menjamin keselamatan dan efisiensi dalam berkendara. Penelitian ini mengembangkan sistem segmentasi objek di jalan dikhususkan pada motor dan mobil dengan melakukan modifikasi algoritma U-Net. Modifikasi yang dilakukan pada penelitian ini adalah penambahan *layer input* yang di *resize* pada bagian *decoder*. Pengambilan data dilakukan di Jl. A.P. Pettarani, Jl. Urip Sumoharjo, Jl. Perintis, Jl. Pendidikan, dan Jl. Hertasning kota Makassar, sehingga diperoleh 651 gambar sebagai dataset. *Hyperparameter* yang digunakan yaitu *epoch* 500 dan *batch size* 4 pada tahap *training*. Pada tahap *testing* digunakan 131 data gambar dengan tiga jenis ukuran gambar, yaitu 128x128 piksel, 256x256 piksel, dan 512x512 piksel. Metrik pengujian untuk mengevaluasi performa sistem pada penelitian ini adalah *Intersection over union* (IoU) dan *f1-score*. Hasil penelitian menunjukkan algoritma modifikasi U-Net dapat meningkatkan akurasi segmentasi *background*, *car*, dan *motorcycle*. Pada pengujian dengan ukuran 128x128 piksel, 256x256 piksel, dan 512x512 piksel algoritma modifikasi U-Net mendapatkan akurasi masing masing sebesar 68%, 63% 74% dengan menggunakan IoU dan sebesar 83%,73,3%, 85,3% menggunakan *f1-score*, lebih baik dari pada menggunakan algoritma U-Net dengan akurasi masing masing sebesar 66,3%, 56,3%, 62,3% dengan menggunakan IoU dan 79,6%, 68,6%, 70,6% menggunakan *f1-score*. Hasil ini memperkuat pentingnya modifikasi algoritma dalam meningkatkan ketepatan deteksi objek pada *autonomous car*. Temuan ini memberikan kontribusi penting dalam pengembangan sistem deteksi objek yang lebih canggih dan akurat untuk aplikasi kendaraan otonom.

Kata Kunci—Semantic Segmentation, Autonomous Car, U-Net, Modifikasi U-Net

ABSTRACT

HERLINA ANWAR. Improving Road Object Detection Performance In Autonomous Car With Modified U-Net Algorithm (supervised by **Indrabayu, Intan Sari Areni**)

Autonomous car is an increasingly popular technology, where accuracy in object detection on the road is very important to ensure safety and efficiency in driving. This research develops a road object segmentation system focused on motorcycles and cars by modifying the U-Net algorithm. The modification made in this research is the addition of a *resized input layer* in the *decoder* section. Data collection was carried out on Jl. A.P. Pettarani, Jl. Urip Sumoharjo, Jl. Perintis, Jl. Education, and Jl. Hertasning Makassar city, so that 651 images were obtained as datasets. The hyperparameters used are epoch 500 and batch size 4 in the training stage. In the test stage, 131 image data is used with three types of image sizes, namely 128x128 pixels, 256x256 pixels, and 512x512 pixels. The test metrics for evaluating system performance in this research are *Intersection over union* (IoU) and f1-score. The results show that the U-Net modification algorithm can increase the accuracy of background, car, and motorcycle segmentation. In tests with sizes 128x128 pixels, 256x256 pixels, and 512x512 pixels, the modified U-Net algorithm obtained accuracies of 68%, 63%, 74% using IoU and 83%, 73.3%, 85.3% using f1-score, respectively, better than using the U-Net algorithm with accuracies of 66.3%, 56.3%, 62.3% using IoU and 79.6%, 68.6%, 70.6% using f1-score. These results reinforce algorithm modification's importance in improving object detection accuracy in autonomous cars. These findings contribute to developing more sophisticated and accurate object detection systems for autonomous vehicle applications.

Keyword—Semantic Segmentation, Autonomous Car, U-Net, Modified U-Net

DAFTAR ISI

HALAMAN JUDUL	i
PENGAJUAN TESIS.....	ii
PERSETUJUAN TESIS	iii
PERNYATAAN KEASLIAN TESIS	Error! Bookmark not defined.
KATA PENGANTAR.....	v
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI.....	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xii
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	3
1.3. Tujuan Penelitian	3
1.4. Manfaat Penelitian	4
1.5. Ruang Lingkup	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Autonomous Car.....	5
2.2 Computer Vision	6
2.3 Semantic Segmentation	7
2.4 Fully <i>Convolutional</i> Network for Semantic Segmentation	8
2.4.1 <i>Convolutional Layer</i>	9
2.4.2 Activation Function ReLu.....	10
2.4.3 <i>Pooling Layer</i>	12
2.5 U-Net.....	14
2.6 Penelitian Terkait Arsitektur Modifikasi U-Net.....	15
2.7 <i>Intersection over union</i> (IoU).....	18
2.8 Confussion Matrix	19
2.8.3 <i>Precision</i>	20
2.8.4 <i>Recall</i>	21
2.8.5 <i>F1-Score</i>	21
2.9 Penelitian Terkait.....	22

BAB III METODOLOGI PENELITIAN	26
3.1 Tahapan Penelitian	26
3.2 Waktu dan Lokasi Penelitian	27
3.3 Instrumen Penelitian	27
3.4 Teknik Pengambilan Data	28
3.5 Rancangan Sistem.....	28
3.5.1 Data <i>Input</i>	29
3.5.2 Data Preprocessing.....	30
3.5.3 Modifikasi Arsitektur U-Net.....	32
3.5.4 Model	43
3.6 Analisis Kerja Sistem	44
BAB IV HASIL DAN PEMBAHASAN	46
4.1 Hasil Penelitian.....	46
4.1.1 Pengujian 1	46
4.1.2 Pengujian 2.....	48
4.1.3 Pengujian 3.....	51
4.2 Pembahasan	53
BAB V KESIMPULAN DAN SARAN	66
5.1 Kesimpulan.....	66
5.2 Saran.....	66
DAFTAR PUSTAKA	68
LAMPIRAN.....	71

DAFTAR TABEL

Tabel 1. <i>State of the Art</i>	24
Tabel 2. Fitur warna diubah menjadi bentuk numerik	31
Tabel 3. Performa model U-Net	46
Tabel 4. Performa model modifikasi U-Net	47
Tabel 5. Performa model U-Net	49
Tabel 6. Performa model modifikasi U-Net	49
Tabel 7. Performa model U-Net	51
Tabel 8. Performa model modifikasi U-Net	51
Tabel 9. Waktu penggunaan GPU saat <i>training</i>	58
Tabel 10. Perbandingan IoU dan <i>F1-Score</i> berdasarkan posisi objek.....	61

DAFTAR GAMBAR

Gambar 1. Aplikasi Visi Komputer	7
Gambar 2. Contoh label segmentasi dari inputan gambar	8
Gambar 3. Arsitektur FCN untuk <i>Semantic Segmentation</i>	9
Gambar 4. Proses konvolusi	10
Gambar 5. Grafik aktivasi ReLu	11
Gambar 6. Proses fungsi aktivasi <i>ReLU</i>	11
Gambar 7. Grafik <i>softmax activation function</i>	12
Gambar 8. Proses <i>max-pooling</i> dan <i>average-pooling</i>	13
Gambar 9. Proses <i>max-pooling</i>	13
Gambar 10. Arsitektur U-Net	14
Gambar 11. VGG-16 + Arsitektur U-Net	16
Gambar 12. ResNet-50 + Arsitektur U-Net	16
Gambar 13. Arsitektur MSCU-Net	17
Gambar 14. Arsitektur LETNet	18
Gambar 15. <i>Interest of Union (IoU)</i>	19
Gambar 16. <i>Multi-class Confusion Matrix</i>	20
Gambar 17. Menghitung Nilai TP, FN, FP, dan TN pada <i>Confusion Matrix</i>	20
Gambar 18. Tahapan penelitian	26
Gambar 19. Ilustrasi pengambilan data	28
Gambar 20. Alur perancangan sistem	29
Gambar 21. Contoh anotasi gambar menggunakan Roboflow	30
Gambar 22. Contoh data format Pascal VOC	31
Gambar 23. Arsitektur Modifikasi U-Net	33
Gambar 24. Modifikasi arsitektur yang diusulkan secara sistematis	35
Gambar 25. Contoh Gambar Awal	36
Gambar 26. <i>Input Layer</i>	37
Gambar 27. Blok pertama (<i>encoder</i>)	37
Gambar 28. Blok kedua (<i>encoder</i>)	38
Gambar 29. Blok ketiga (<i>encoder</i>)	38
Gambar 30. Blok keempat (<i>encoder</i>)	39

Gambar 31. <i>Bottom layer</i>	39
Gambar 32. Blok pertama (<i>decoder</i>)	40
Gambar 33. <i>Resized Input Layer 1</i>	40
Gambar 34. Blok kedua (<i>decoder</i>)	41
Gambar 35. <i>Resized Input Layer 2</i>	41
Gambar 36. Blok ketiga (<i>decoder</i>)	42
Gambar 37. <i>Resized Input Layer 3</i>	42
Gambar 38. Blok keempat (<i>decoder</i>)	43
Gambar 39. <i>Output layer</i>	43
Gambar 40. Contoh posisi objek dan kondisi jalan di Kota Makassar.....	45
Gambar 41. IoU antara U-Net dengan modifikasi U-Net.....	48
Gambar 42. <i>F1-score</i> antara U-Net dengan modifikasi U-Net	48
Gambar 43. Perbandingan IoU pada U-Net dan modifikasi U-Net.....	50
Gambar 44. Perbandingan <i>f1-score</i> pada U-Net dan modifikasi U-Net.....	50
Gambar 45. Perbandingan metrik IoU pada U-Net dan modifikasi U-Net	52
Gambar 46. Perbandingan metrik <i>f1-score</i> pada U-Net dan modifikasi U-Net ...	53
Gambar 47. Perbandingan IoU pada ketiga pengujian	54
Gambar 48. Perbandingan <i>f1-score</i> pada ketiga pengujian	55
Gambar 49. Kecepatan deteksi sistem pada <i>input</i> 128x128 piksel.....	59
Gambar 50. Kecepatan deteksi sistem pada <i>input</i> 256x256 piksel	60
Gambar 51. Kecepatan deteksi sistem pada <i>input</i> 512x512 piksel	60
Gambar 52. Contoh hasil dari set data yang dianalisis.....	63

BAB I PENDAHULUAN

1.1. Latar Belakang

Pengembangan kendaraan otonom telah mengalami kemajuan yang signifikan dalam beberapa tahun terakhir, sehingga pentingnya deteksi objek yang aman dan akurat dalam memastikan kinerja mengemudi otonom terutama dalam konteks detektor berbasis pembelajaran mendalam (Balasubramaniam & Pasricha, 2022). Kemajuan terbaru dalam *deep learning* telah meningkatkan akurasi deteksi objek secara signifikan. Salah satu pendekatan yang populer adalah segmentasi semantik, yang mengasosiasikan setiap piksel dalam gambar dengan kelas tertentu (Sohail et al., 2022), telah menjadi salah satu pondasi utama dalam pemrosesan citra dan visi komputer, terutama dalam aplikasi kendaraan otonom. Teknik ini memungkinkan sistem percepatan kendaraan untuk memahami dan membedakan komponen lingkungan secara detail, seperti perbedaan antara jalan, pejalan kaki, kendaraan lain, dan rintangan.

Namun, segmentasi semantik, terutama dalam konteks pengemudian otonom memiliki tantangan yang beragam, terutama pada citra penginderaan jauh resolusi tinggi, dimana perbedaan skala yang kecil, padat, dan intra-kelas menimbulkan kesulitan (Wang et al., 2020). Kombinasi deteksi objek, segmentasi citra, dan pengenalan multi-kelas dalam segmentasi semantik semakin memperumit tugas tersebut (Atif et al., 2019). Pembelajaran mendalam, khususnya melalui penggunaan *Convolutional Neural Network* (CNN), telah mengubah paradigma pemrosesan citra dan visi komputer, memungkinkan pengembangan sistem yang dapat belajar dari data dengan sedikit intervensi manusia. Salah satu algoritma berbasis jaringan saraf konvolusional mendalam, yang digunakan pada *semantic segmentation* adalah U-Net. Algoritma U-Net adalah salah satu jenis algoritma CNN yang biasa digunakan untuk segmentasi citra berjenis semantik (Ronneberger et al., 2015). U-Net awalnya dikembangkan untuk menangani tugas segmentasi gambar medis, dimana tujuannya adalah untuk memisahkan objek dari latar belakang dalam gambar. Arsitektur ini terkenal karena kemampuannya dalam mengatasi masalah segmentasi pada dataset dengan jumlah sampel yang sedikit

dan telah menjadi salah satu pendekatan yang paling populer dan berhasil dalam menangani tugas-tugas segmentasi semantik yang kompleks dan beragam.

Dalam survei yang dilakukan oleh Huang et al., 2022 eksplorasi mendalam tentang aplikasi teknik pembelajaran dalam untuk segmentasi gambar medis menunjukkan potensi luas dari teknologi ini dalam berbagai bidang aplikasi. Survei tersebut menekankan bagaimana inovasi dalam *deep learning*, khususnya melalui modifikasi dan penyesuaian arsitektur U-Net, dapat meningkatkan akurasi dan efisiensi segmentasi semantik.

Berbagai penelitian telah mengeksplorasi penggunaan U-Net untuk segmentasi semantik dalam pengemudian otonom. Indrabayu et al., 2023 lebih lanjut menunjukkan keefektifan U-Net dalam mendeteksi marka jalur jalan Kota Makassar, Sulawesi Selatan, sebuah tugas penting dalam pengemudian otonom dengan performa IoU sebesar 79,8%. Kemudian Sugirtha & Sridevi, 2022 menyajikan sistem *semantic segmentation* pada dataset Cityscapes, yaitu gambar objek di jalan perkotaan dengan menggunakan algoritma U-Net yang telah dimodifikasi pada arsitektur bagian *Encoder* dengan menggantinya dengan arsitektur VGG16 dan ResNet50. Sistem ini meningkatkan akurasi U-Net untuk pengemudian otonom dengan VGG16 menunjukkan kinerja yang lebih baik daripada ResNet50 dengan total perhitungan mIoU tertinggi sebesar 80,5%.

Xu et al., 2023 juga menyajikan sistem segmentasi semantik pada dataset Cityscapes dan dataset Cam Vid dengan algoritma modifikasi LetNet. LetNet menggabungkan arsitektur *Convolutional Neural Network* (CNN) dalam bentuk U-shape dengan Transformer. Sistem ini menghasilkan 72,8% mIoU pada *Cityscapes* dan 70,5% mIoU pada set uji CamVid. Noori et al., 2022 mencapai akurasi piksel sebesar 0,7874 menggunakan model hibrida U-Net dan Xception, dengan 69% mIoU dan 78% akurasi pada dataset *Cityscapes*.

Penelitian yang disebutkan menyoroti pentingnya terus mengembangkan dan menyesuaikan teknik pembelajaran dalam untuk meningkatkan kinerja sistem kendaraan otonom. Melalui adaptasi arsitektur seperti U-Net dan penerapan teknik pembelajaran dalam yang canggih, kemungkinan untuk mencapai tingkat pemahaman lingkungan yang lebih tinggi untuk kendaraan otonom menjadi semakin terjangkau. Penelitian-penelitian ini secara kolektif menunjukkan potensi

U-Net dan modifikasinya dalam mencapai akurasi dan kinerja yang tinggi dalam tugas-tugas segmentasi menjadikannya arsitektur yang sangat berharga dalam pengembangan sistem penglihatan untuk kendaraan otonom. Karya-karya ini membentuk dasar penting dalam literatur segmentasi semantik dan pembelajaran mendalam, menunjukkan jalan menuju inovasi berkelanjutan dalam teknologi kendaraan otonom.

Oleh karena itu, berdasarkan penelitian sebelumnya, penulis mengusulkan sistem dengan melakukan modifikasi pada arsitektur U-Net sehingga dapat menangani tugas dalam segmentasi semantik pada objek di jalan, dengan tujuan akan dapat meningkatkan performa algoritma dalam melakukan segmentasi objek di jalan kota Makassar Sulawesi selatan. Penelitian ini juga difokuskan untuk melihat bagaimana perbedaan performa algoritma dalam melakukan segmentasi semantik dengan posisi objek yang berbeda-beda.

1.2. Rumusan Masalah

Berdasarkan Latar Belakang penelitian adapun rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana memodifikasi arsitektur U-Net dalam melakukan segmentasi objek di jalan?
2. Bagaimana performa modifikasi algoritma U-Net dalam segmentasi objek jalan?

1.3. Tujuan Penelitian

Tujuan pada penelitian ini adalah sebagai berikut:

1. Memodifikasi arsitektur U-Net dalam melakukan segmentasi objek di jalan.
2. Menganalisis performa modifikasi algoritma U-Net dalam segmentasi objek di jalan.

1.4. Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut:

1. Dengan adanya sistem pendeteksian objek di jalan diharapkan dapat memberikan kontribusi terhadap pengembangan *autonomous car* ke depannya dalam memberikan informasi keadaan jalan Kota Makassar.
2. Sebagai referensi dalam pengembangan penelitian topik terkait deteksi objek menggunakan *semantic segmentation*.

1.5. Ruang Lingkup penelitian

Ruang lingkup pada penelitian ini adalah sebagai berikut:

1. Objek yang dideteksi adalah mobil dan motor.
2. Pengambilan data dilakukan pada siang hari pada kondisi cuaca cerah.
3. Menggunakan kamera *dashcam* yang berjumlah satu buah dan diposisikan menghadap depan.

BAB II TINJAUAN PUSTAKA

2.1 Autonomous Car

Autonomous car adalah mobil yang dikendalikan oleh komputer yang dapat memandu dirinya sendiri, mengenali lingkungan sekitar, membuat keputusan dan beroperasi penuh tanpa adanya interaksi manusia. Hal yang menjadi dasar dalam munculnya ide mobil yang dikendalikan oleh komputer tersebut adalah meliputi keselamatan pengemudi, pertumbuhan penduduk seiring dengan peningkatan jumlah kendaraan, pertumbuhan infrastruktur, kebutuhan akan manajemen waktu yang efisien serta pemanfaatan dan optimalisasi sumber daya (Hussain & Zeadally, 2019). Mobil otonom, juga dikenal sebagai mobil tanpa pengemudi, adalah kendaraan yang dapat beroperasi dan bergerak tanpa campur tangan manusia. Mobil ini menggunakan kombinasi teknologi canggih, seperti sensor, kamera, radar, lidar, GPS, dan kecerdasan buatan (AI), untuk mengenali lingkungannya, membuat keputusan, dan mengendalikan pergerakannya (Azani, 2018).

Algoritma AI dan pembelajaran mesin memproses data yang dikumpulkan untuk mengambil keputusan navigasi dalam waktu nyata, menangani fungsi seperti menghindari rintangan, mengganti jalur, dan menghentikan kendaraan secara aman. Kontrol otomatis atas kemudi, akselerasi, dan pengereman, memungkinkan eksekusi keputusan ini, menjadikan perjalanan lebih aman dan efisien. Lebih jauh, kemajuan dalam komunikasi Kendaraan-ke-Kendaraan (V2V) dan Kendaraan-ke-Infrastruktur (V2I) berpotensi meningkatkan keselamatan dan efisiensi lalu lintas dengan memfasilitasi pertukaran informasi antar kendaraan dan dengan infrastruktur jalan (Anderson et al., 2014).

Dengan adanya berbagai tingkatan otonomi, dari 0 (tanpa otomatisasi) hingga 5 (otonomi penuh), seperti diklasifikasikan oleh *Society of Automotive Engineers* (SAE), mobil otonom menjanjikan peningkatan signifikan dalam keselamatan lalu lintas dengan mengurangi kecelakaan yang disebabkan oleh kesalahan manusia. Namun, terdapat tantangan signifikan yang harus diatasi, termasuk pertimbangan hukum, etika, keamanan siber, dan penerimaan sosial.

Kesuksesan implementasi teknologi ini tidak hanya bergantung pada kemajuan teknis tetapi juga pada kemampuan masyarakat untuk menavigasi kompleksitas sosial dan hukum yang terkait (Bonneton et al., 2016).

2.2 Computer Vision

Visi komputer adalah suatu proses transformasi atau perubahan dari data yang berasal dari kamera video maupun foto/gambar ke dalam sebuah hasil keputusan ataupun sebuah presentasi yang baru, dimana hasil dari kegiatan transformasi tersebut memiliki kepentingan untuk mencapai suatu tujuan. Data yang dimasukkan kedalam kegiatan transformasi tersebut memungkinkan untuk memiliki beberapa informasi yang terkontekstual seperti halnya sebuah foto/gambar yang di dalamnya terdapat berbagai objek. Dengan demikian akan didapatkan keputusan-keputusan yang akan diambil pada gambar (Aningtiyas et al., 2020).

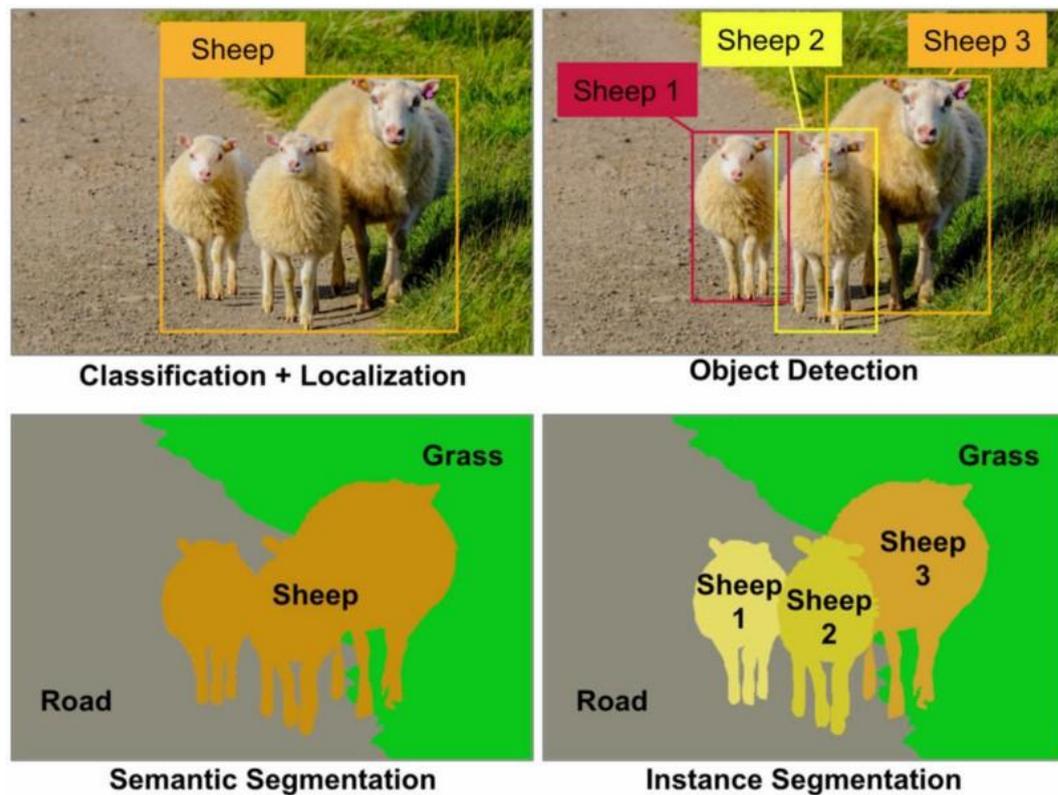
Secara garis besar, visi komputer adalah sebuah teknologi mesin yang mampu mengenali objek yang diamati. Kemampuan untuk mengenali ini merupakan kombinasi dari pengolahan citra dan pengenalan pola (Basuki, 2016). Pengolahan citra adalah proses awal dalam visi komputer untuk menghasilkan citra yang lebih baik atau lebih mudah diinterpretasikan, sedangkan pengenalan pola adalah proses identifikasi objek pada citra. Proses-proses dalam visi komputer secara garis besar dapat dibagi menjadi:

1. Proses mengakuisisi citra digital (*Image Acquisition*)
2. Proses pengolahan citra (*Image Processing*)
3. Proses analisis data citra (*Image Analysis*)
4. Proses pemahaman data citra (*Image Understanding*)

Visi komputer dan deteksi objek adalah dua fitur paling penting dan esensial dari mobil otonom. Untuk meniru perilaku pengemudi manusia, mobil otonom harus mengenali jalan dan mendeteksi setiap rintangan di depan dan di sekitarnya, baik itu mobil lain, pejalan kaki, lubang, atau jenis rintangan lainnya. Kedua fitur utama ini bersama dengan modul lainnya memungkinkan mobil otonom untuk mengemudi di sepanjang jalan dan merespons situasi yang tidak diinginkan dengan cara yang aman, misalnya berhenti di lampu lalu lintas, melambat jika

mobil sebelumnya mengurangi kecepatan, menghindari pejalan kaki, dan sebagainya. Sampai saat ini, banyak hasil penelitian telah dicapai baik dalam visi komputer dan deteksi objek untuk mobil otonom (Hussain & Zeadally, 2019).

Terdapat berbagai aplikasi visi komputer yang membedakan tugas-tugas utama seperti klasifikasi dan lokalisasi objek, deteksi objek, *semantic segmentation*, dan *instance segmentation*. Adapun Perbedaan aplikasi dalam visi komputer dapat dilihat pada Gambar 1.



Gambar 1 Aplikasi Visi Komputer

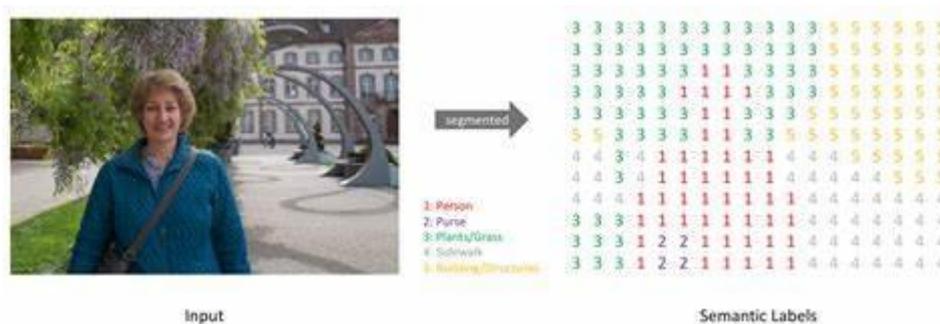
2.3 Semantic Segmentation

Semantic segmentation, juga disebut pelabelan adegan, mengacu pada proses pemberian label semantik (misalnya mobil, orang, dan jalan) ke setiap *pixel* gambar. Ini adalah langkah pemrosesan data penting untuk komputer dan sistem tak berawak lainnya untuk memahami lingkungan di sekitarnya. Terlepas dari upaya selama beberapa dekade, *semantic segmentation* masih merupakan tugas yang sangat menantang karena terlalu banyak variasi dalam pemandangan alam. Dengan pengembangan *deep learning*, penggunaan fitur yang dipelajari dalam

tugas visi komputer, seperti klasifikasi gambar telah mencapai sukses besar dalam beberapa tahun terakhir. Akibatnya, peneliti *semantic segmentation* baru-baru ini menaruh banyak perhatian pada fitur yang dipelajari yang biasanya mengacu pada *Convolutional Neural Network* (CNN) (H. Yu et al., 2018).

Munculnya terminologi *semantic segmentation* dapat dilihat kembali pada tahun 1970-an (Ohta et al., 1978). Pada saat itu, terminologi ini setara dengan segmentasi citra tetapi menekankan bahwa daerah yang tersegmentasi harus “bermakna secara semantik” contohnya pada citra perkotaan, sebuah gedung memiliki jendela dan memiliki ciri-ciri berbentuk persegi panjang sedangkan langit berwarna biru. Lalu penelitian berlanjut menjadi objek segmentasi yang dapat memisahkan antara objek dan latar belakang menggunakan algoritma *Constrained Parametric Min-Cuts* (CPMC) (Carreira & Sminchisescu, 2011).

Sederhananya, tujuannya adalah mengambil gambar warna RGB ($height \times width \times 3$) atau gambar skala abu-abu ($height \times width \times 1$) dan menampilkan peta segmentasi dimana setiap *pixel* berisi label kelas yang direpresentasikan sebagai bilangan bulat, seperti yang diilustrasikan pada Gambar 2.



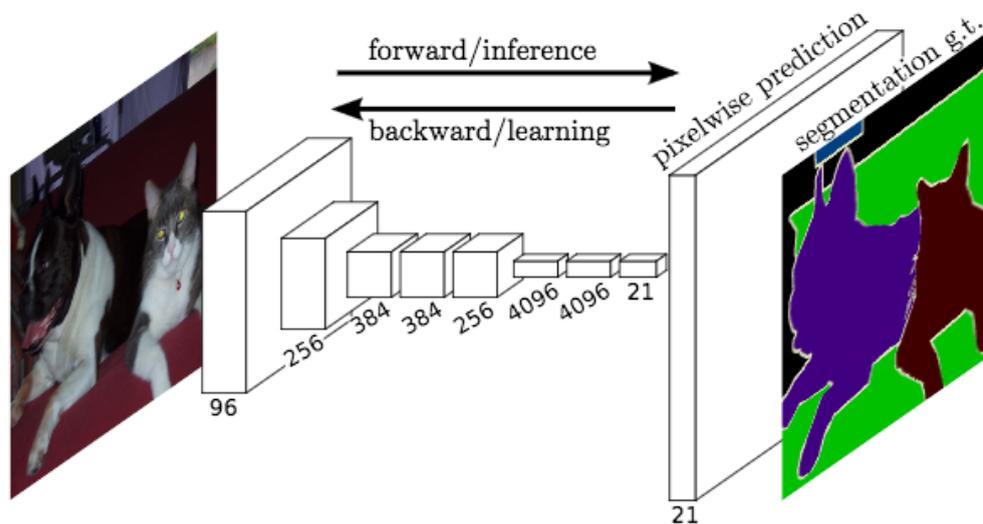
Gambar 2 Contoh label segmentasi dari inputan gambar

Dapat kita lihat ada dua masalah yang dapat diselesaikan oleh *semantic segmentation* yaitu informasi objek dan lokalisasi. Teknik ini dapat mengklasifikasikan berbagai objek seperti *person*, *purse*, *plants*, *sidewalk*, *building*, serta menunjukkan di mana objek-objek tersebut berada.

2.4 Fully Convolutional Network for Semantic Segmentation

Shelhamer dan kawan-kawan menunjukkan bahwa *Fully Convolutional Network* (FCNs) yang dilatih ujung-ke-ujung, *pixel-ke-pixel* pada segmentasi

semantik melebihi hasil terbaik sebelumnya tanpa mesin lebih lanjut. Menurut Shelhamer dkk, ini adalah pekerjaan pertama untuk melatih FCN *end-to-end* untuk prediksi *pixel* dan dari pra-pelatihan yang diawasi. Versi konvolusi penuh dari jaringan yang ada memprediksi keluaran padat dari masukan berukuran arbitrer. Baik pembelajaran dan inferensi dilakukan seluruh gambar pada satu waktu dengan 20 perhitungan *feedforward* yang padat dan *backpropagation*. Lapisan *upsampling* dalam jaringan memungkinkan prediksi dan pembelajaran *pixel* dalam jaring dengan *subsampling* (Long et al., 2015).



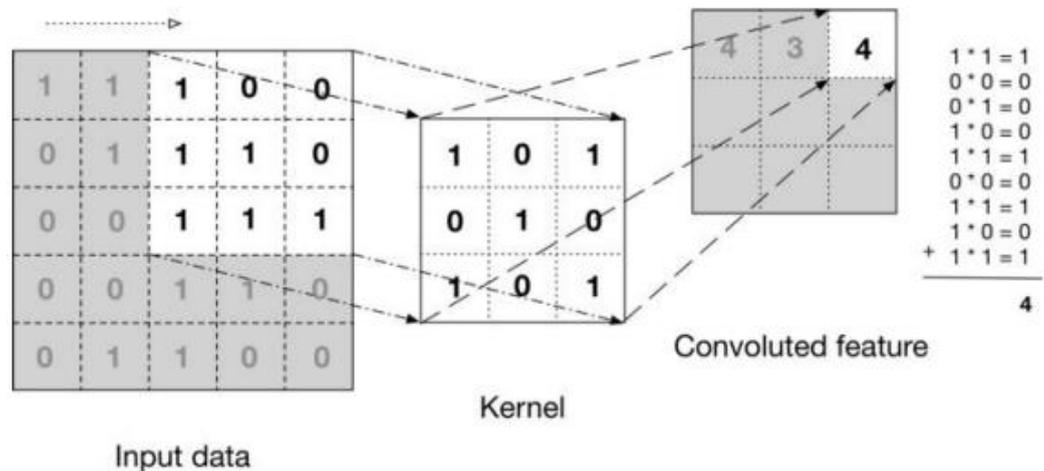
Gambar 3 Arsitektur *Fully Convolutional Network* untuk *Semantic Segmentation*

Setiap keluaran lapisan dalam sebuah *convnet* adalah *array* tiga dimensi dengan ukuran $h \times w \times d$, di mana h dan w adalah dimensi spasial, dan d adalah dimensi fitur atau saluran. Lapisan pertama adalah citra, dengan ukuran *pixel* $h \times w$, dan saluran d . Lokasi di lapisan yang lebih tinggi sesuai dengan lokasi di gambar yang terhubung dengan jalurnya, yang disebut bidang reseptifnya.

2.4.1 *Convolutional Layer*

Convolutional layer adalah komponen pada U-Net Oyang dirancang untuk melakukan pengolahan citra untuk mencari pola tertentu. *Convolutional layer* dapat dikatakan sebagai pencari pola atau fitur yang bertujuan menemukan fitur-fitur penting di dalam gambar, seperti fitur tepi, fitur sudut, dan fitur tekstur. Dalam lapisan konvolusi terdapat beberapa parameter yang dapat diatur, seperti jumlah filter, ukuran kernel, perpindahan (*stride*), dan *padding*. Filter-filter akan

digeser dengan jarak yang telah diatur pada *stride* tertentu dengan ukuran kernel tertentu (3×3 , 5×5 , 7×7) serta *padding same* atau *valid*. Hasil dari lapisan konvolusi adalah *feature map*, yaitu hasil ekstraksi fitur pada gambar dengan nilai filter tertentu (Azmi et al., 2023).



Gambar 4 Proses konvolusi

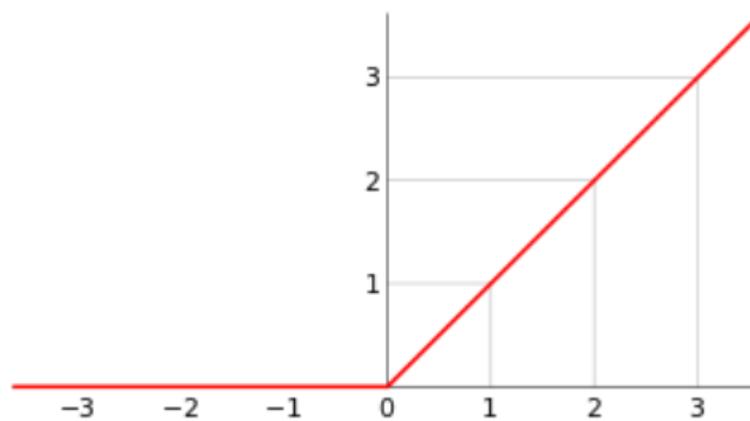
Pada konvolusi, masukan gambar akan dikonvolusi atau di-*filtering* dengan filter yang telah ditetapkan jumlah dan ukuran kernelnya. Proses konvolusi sebenarnya merupakan perkalian dot antara *input* data dengan filter kemudian menjumlahkan hasil perkalian tersebut. Proses konvolusi akan membuat gambar semakin kecil, sehingga hal ini dapat diatasi dengan menggunakan *padding same* yang akan mempertahankan ukuran gambar. Hal tersebut menghasilkan *feature map* atau *activation map* pada citra *output* seperti yang dapat dilihat pada Gambar 4, dan kemudian hasil tersebut diberikan sebagai data *input* untuk *convolution layer* selanjutnya.

2.4.2 Activation Function ReLu

Activation function atau fungsi aktivasi adalah fungsi yang berperan dalam memperkenalkan non-linearitas ke dalam jaringan sehingga variasi data yang terbentuk menjadi lebih detail dan kompleks. Pemilihan fungsi aktivasi dalam arsitektur CNN sangat memengaruhi kinerja model secara signifikan (Wathani & Hidayati, 2023). Terdapat beberapa jenis aktivasi fungsi yang umum digunakan yaitu fungsi aktivasi *ReLU* dan *Softmax*.

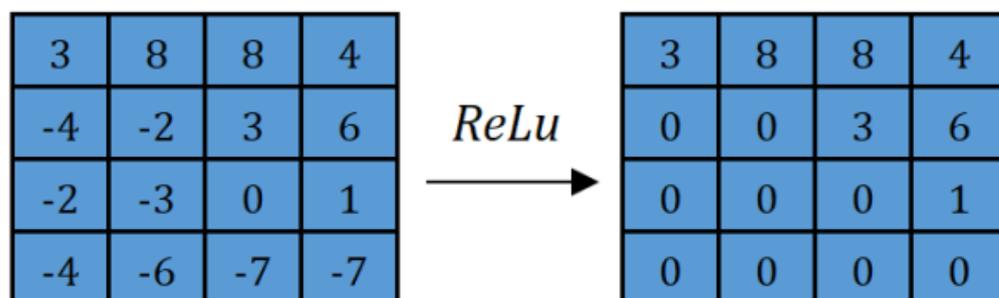
a. Fungsi aktivasi ReLu

Fungsi aktivasi ReLu (*Rectified Linear Unit*) adalah fungsi aktivasi yang mengembalikan nilai *input* jika *input* positif (lebih besar dari 0) atau mengembalikan nilai 0 jika *input* negatif (kurang dari atau sama dengan 0). Fungsi aktivasi ReLu memiliki keuntungan terhadap fungsi aktivasi *sigmoid* dan *tanh* karena tidak pernah menjadi jenuh terhadap nilai x yang besar (Sagala, 2022). Meskipun sederhana, fungsi aktivasi ReLu mampu menerapkan konsep non-linearitas yang dapat memungkinkan jaringan untuk mengabaikan nilai yang *outlier* atau tidak relevan. Grafik fungsi aktivasi ReLu dapat dilihat pada Gambar 5.



Gambar 5 Grafik aktivasi ReLu

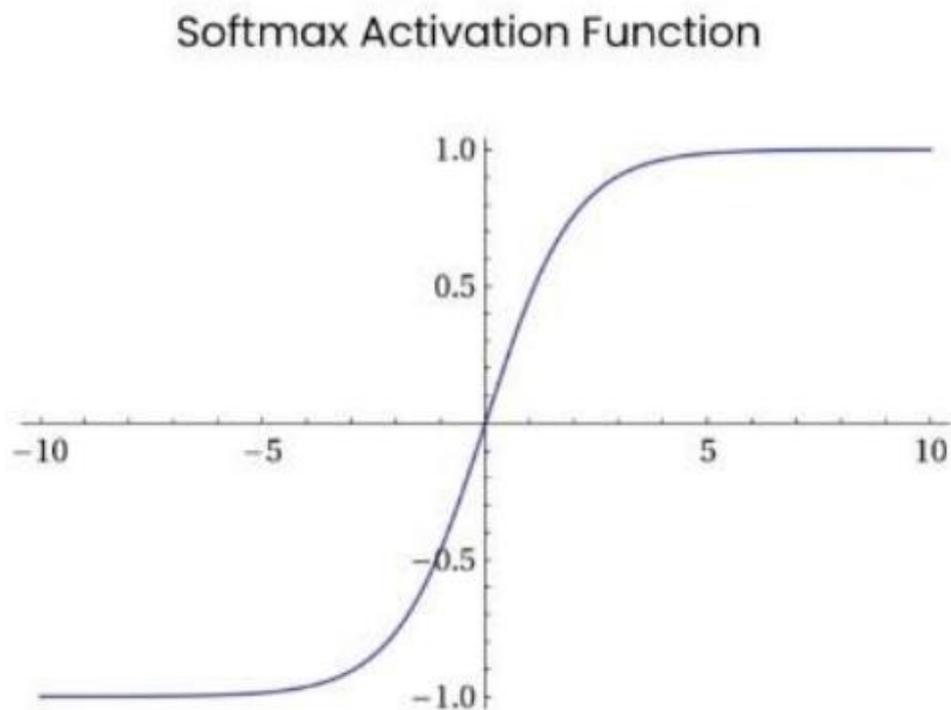
Hasil dari proses konvolusi akan menghasilkan *feature map*. Selanjutnya *feature map* akan dikenakan fungsi aktivasi ReLu yang akan mengubah semua nilai yang lebih kecil atau sama dengan nol menjadi 0 dan mempertahankan nilai yang lebih besar dari 0. Proses dari fungsi aktivasi ReLu dapat dilihat pada Gambar 6.



Gambar 6 Proses fungsi aktivasi *ReLu*

b. Fungsi aktivasi *softmax*

Fungsi aktivasi *softmax* adalah fungsi yang berfokus pada konteks pengenalan pola klasifikasi. Fungsi aktivasi *softmax* adalah cara yang efektif untuk mengubah sejumlah bilangan riil dalam rentang nilai apa pun menjadi distribusi probabilitas yang berjumlah satu. Fungsi aktivasi *softmax* akan mengubah vektor yang berisi nilai-nilai numerik menjadi distribusi probabilitas dengan menjalankan setiap nilai melalui fungsi eksponensial dan akan dinormalisasi. Fungsi aktivasi *softmax* akan menghasilkan masing-masing peluang dari sebuah model multi-kelas (Sagala, 2022). Grafik fungsi aktivasi *softmax* adalah sebagai yang diilustrasikan pada Gambar 7.

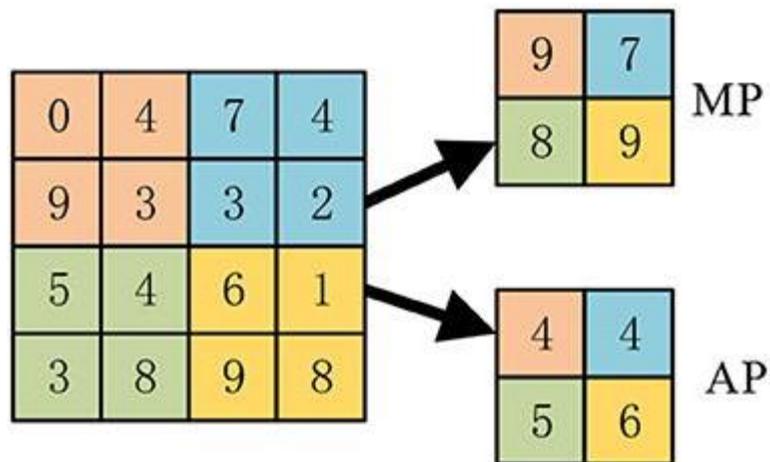


Gambar 7 Grafik *softmax activation function*

2.4.3 Pooling Layer

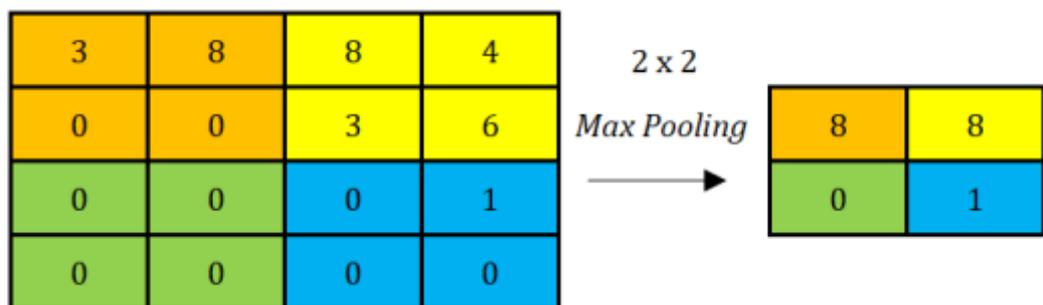
Pooling layer adalah lapisan yang digunakan untuk mereduksi dimensi spasial dari gambar yang dihasilkan setelah proses konvolusi dan aktivasi. Lapisan *pooling* ini bertujuan untuk mengambil fitur utama pada gambar sehingga mengurangi *overfitting* dan mempercepat waktu komputasi. Terdapat dua metode yang dapat dilakukan dalam lapisan *pooling* yaitu *max-pooling* dan *average-*

pooling. Berikut gambaran *max-pooling* dan *average-pooling* yang diilustrasikan pada Gambar 8.



Gambar 8 Proses *max-pooling* dan *average-pooling*

Berdasarkan pada Gambar 8, *max-pooling* mengambil nilai intensitas terbesar dengan ukuran kernel tertentu. Sedangkan *average-pooling* menghitung rata-rata dari nilai intensitas yang terbesar dengan ukuran kernel tertentu. Namun, metode *pooling* yang digunakan pada arsitektur U-Net yaitu metode *max-pooling*. Berikut gambaran proses *max pooling* pada Gambar 9.

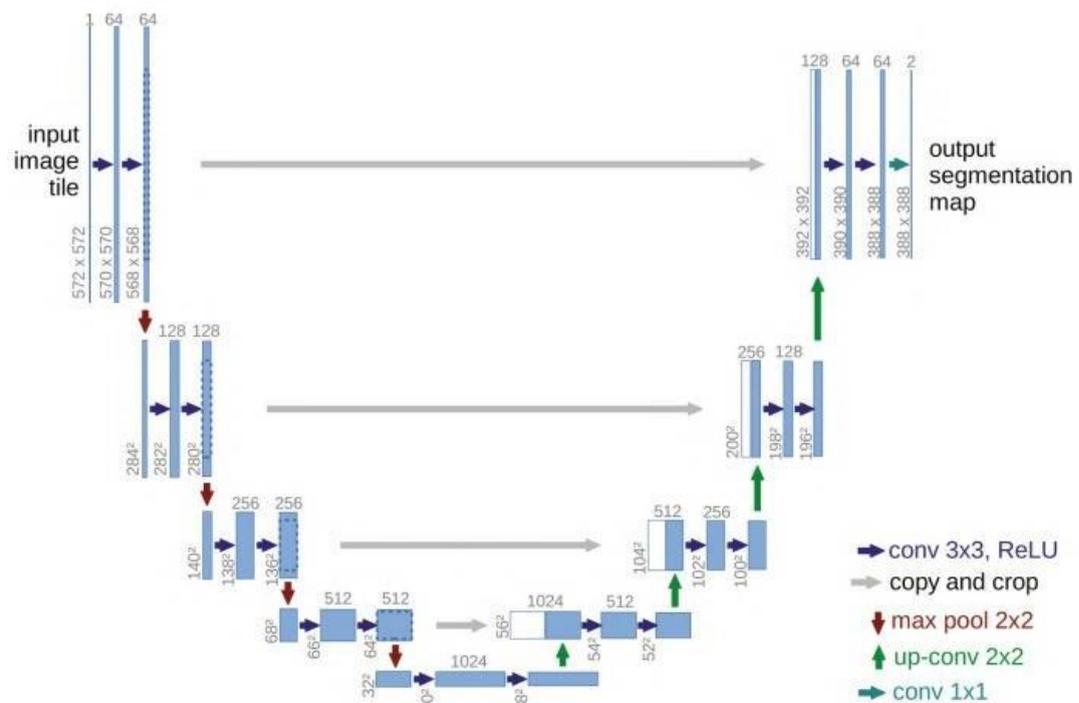


Gambar 9 Proses *max-pooling*

Berdasarkan Gambar 9, *max pooling* bekerja dengan menerapkan kernel berukuran 2×2 dengan *stride* atau perpindahan 2 untuk mengambil nilai maksimum dari matriks *input* 4×4 sehingga menghasilkan hasil *feature map* yang jauh lebih kecil berukuran 2×2 .

2.5 U-Net

Arsitektur U-Net merupakan sebuah arsitektur yang menggunakan *layer concatenate* yang melakukan transmisi informasi spasial pada tiap levelnya dari *encoder* ke *decoder* secara *multilevel*. Fitur ini dapat menghasilkan segmentasi citra yang halus dan akurat. Pada U-Net, *encoder* terdiri atas lapisan konvolusi 2D dengan kernel 3×3 dan *max pooling* dengan kernel 2×2 . Kemudian *decoder* U-Net terdiri atas lapisan konvolusi 2D dengan kernel 3×3 dan *transpose convolution* atau *deconvolution* dengan kernel 2×2 dan *stride 2* (Hadinata et al., 2021). Arsitektur U-Net dapat dilihat pada Gambar 10 berikut.



Gambar 10 Arsitektur U-Net

U-Net (Ronneberger dkk., 2015) menggunakan jaringan *Fully Convolutional Network* untuk melakukan tugas segmentasi semantik. Arsitektur jaringan simetris, memiliki *encoder* yang mengekstrak fitur spasial dari gambar, dan *decoder* yang membangun peta segmentasi dari fitur yang dikodekan. *Encoder* mengikuti formasi khas jaringan konvolusi. Ini melibatkan urutan dua operasi konvolusi 3×3 , diikuti oleh operasi *max-pooling* dengan ukuran *pooling* 2×2 . Urutan ini diulang empat kali, dan setelah setiap *downsampling*, jumlah

filter di lapisan konvolusi menjadi dua kali lipat. Akhirnya, perkembangan dua operasi konvolusi 3×3 menghubungkan *encoder* ke *decoder*.

Di sisi lain, *decoder* pertama-tama mengambil sampel peta fitur menggunakan operasi konvolusi 2×2 yang dialihkan, mengurangi saluran fitur hingga setengahnya. Kemudian urutan dua operasi konvolusi 3×3 dilakukan lagi. Mirip dengan *encoder*, sukseksi *upsampling* dan dua operasi konvolusi ini diulang empat kali, mengurangi separuh jumlah filter pada setiap tahap. Akhirnya, operasi konvolusi 1×1 dilakukan untuk menghasilkan peta segmentasi akhir. Semua lapisan konvolusi dalam arsitektur ini, kecuali yang terakhir, menggunakan fungsi aktivasi ReLU (*Rectified Linear Unit*) lapisan konvolusi akhir menggunakan fungsi aktivasi *sigmoid*. Aspek yang paling cerdas dari arsitektur U-Net adalah *skip connection*. Di keempat level, *output* dari *convolutional layer*, sebelum operasi penyatuan *encoder* dikirim ke *decoder*. Peta Fitur ini kemudian digabungkan dengan *output* dari operasi *upsampling*, dan peta fitur gabungan disebarkan ke lapisan yang berurutan. *Skip connectoin* ini memungkinkan jaringan untuk mengambil informasi spasial yang hilang oleh operasi penyatuan (Carneiro et al., 2016).

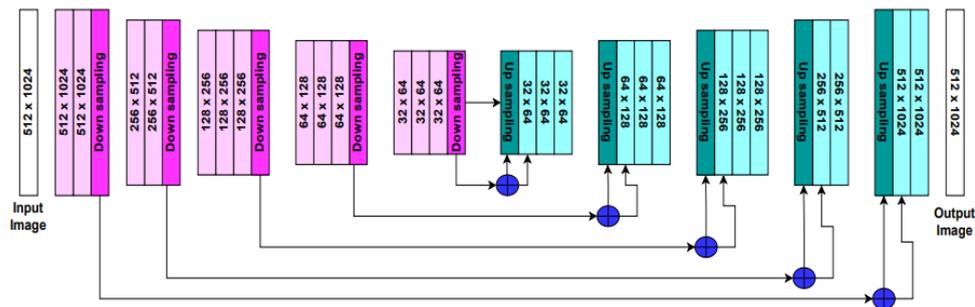
2.6 Penelitian Terkait Arsitektur Modifikasi U-Net

Modifikasi arsitektur U-Net telah banyak diimplementasikan untuk meningkatkan performa segmentasi semantik dalam berbagai aplikasi. Dengan kemajuan teknologi dan kebutuhan yang semakin kompleks dalam berbagai bidang, seperti pengemudian otonom, pemrosesan citra medis, dan pengawasan video, modifikasi pada arsitektur U-Net menjadi sangat penting untuk mencapai hasil yang lebih akurat dan efisien. Penelitian-penelitian terbaru telah menunjukkan berbagai pendekatan inovatif dalam memodifikasi U-Net, baik melalui penambahan lapisan-lapisan baru, penggunaan blok efisien, maupun penggabungan teknik-teknik mutakhir seperti Transformer.

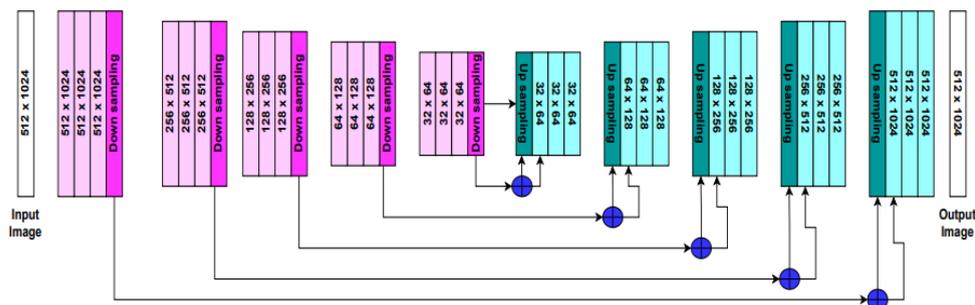
2.6.1. Semantic Segmentation using Modified U-Net for Autonomous Driving (Sugirtha & Sridevi, 2022)

Penelitian ini memfokuskan pada penggunaan U-Net yang dimodifikasi untuk aplikasi mengemudi otonom. Mereka melakukan perubahan pada blok

encoder untuk meningkatkan akurasi dalam mendeteksi batas jalan dan objek sekitar. Modifikasi ini termasuk penambahan lapisan *convolutional* yang lebih dalam serta penggunaan fungsi aktivasi yang lebih kompleks untuk mengatasi variasi pencahayaan dan kondisi jalan yang beragam. Adapun arsitektur modifikasi yang dikembangkan adalah mengganti bagian *Encoder* pada arsitektur U-Net dan menggantinya dengan arsitektur VGG-16 dan Resnet-50. Adapun arsitektur modifikasi dapat dilihat pada Gambar 11 dan Gambar 12.



Gambar 11 VGG-16 + Arsitektur U-Net

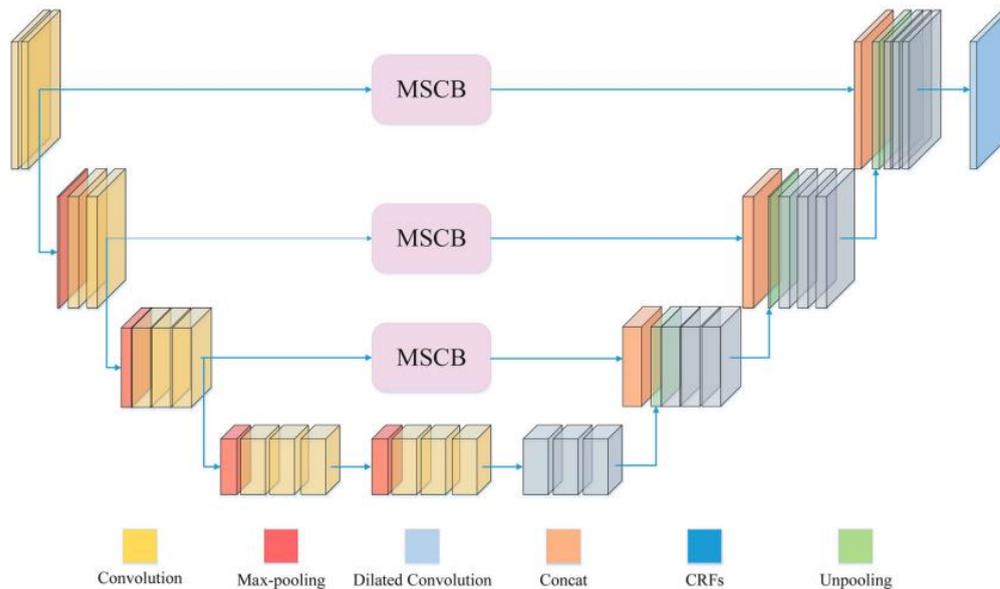


Gambar 12 ResNet-50 + Arsitektur U-Net

2.6.2. Aggregating multi-scale contextual features from multiple stages for semantic image segmentation (D. Jiang et al., 2021)

Dalam penelitian ini, modifikasi utama yang dilakukan terhadap arsitektur U-Net melibatkan penambahan Multi-Scale Context Block (MSCB) dan integrasi Conditional Random Fields (CRF). MSCB menggunakan dilated *convolutions* dengan berbagai tingkat dilasi (6, 12, 18) untuk menangkap informasi kontekstual pada berbagai skala, memperluas receptive field tanpa mengurangi resolusi fitur. Ini memungkinkan jaringan untuk mengenali objek dengan berbagai ukuran lebih akurat. Sementara itu, CRF diterapkan untuk memodelkan hubungan spasial antar

piksel, meningkatkan akurasi segmentasi, terutama pada objek kecil dan batas yang kompleks. Modifikasi ini secara signifikan meningkatkan kinerja segmentasi semantik, menjadikan MSCU-net lebih efektif dibandingkan dengan U-Net klasik.. Adapun arsitektur modifikasi dapat dilihat pada Gambar 13.

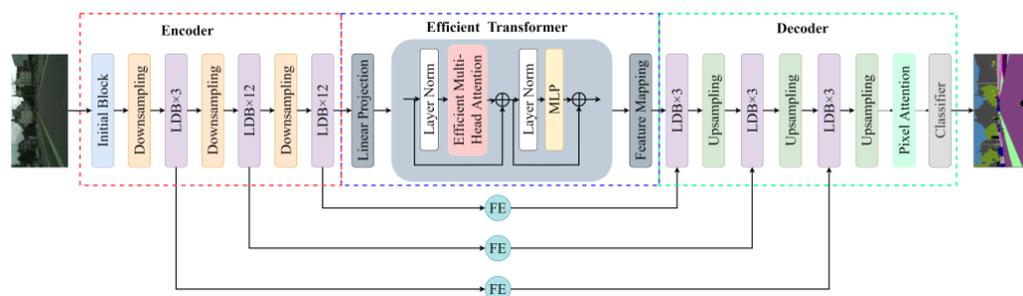


Gambar 13 Arsitektur MSCU-Net

2.6.3 Lightweight Real-time Semantic Segmentation Network with Efficient Transformer and CNN (Xu et al., 2023)

Modifikasi yang dilakukan pada penelitian ini melibatkan pengembangan LETNet, sebuah jaringan segmentasi semantik real-time yang ringan, yang menggabungkan CNN dan Transformer untuk meningkatkan efisiensi dan akurasi. Arsitektur LETNet menggunakan pendekatan *encoder-decoder* yang terinspirasi dari U-Net, di mana CNN berperan sebagai *encoder* untuk mengekstrak fitur lokal, dan *decoder* untuk memetakan kembali fitur tersebut ke resolusi asli. Selain itu, LETNet memperkenalkan Lightweight Dilated Bottleneck (LDB) untuk mengurangi jumlah parameter dan beban komputasi tanpa mengorbankan akurasi, dengan menggunakan konvolusi terpisah dan dilatasi. Untuk menangkap informasi global, LETNet mengintegrasikan Efficient Transformer (ET) di tengah jaringan, yang dirancang untuk mengurangi penggunaan memori GPU. Lebih lanjut, LETNet memperkenalkan modul Feature Enhancement (FE) yang

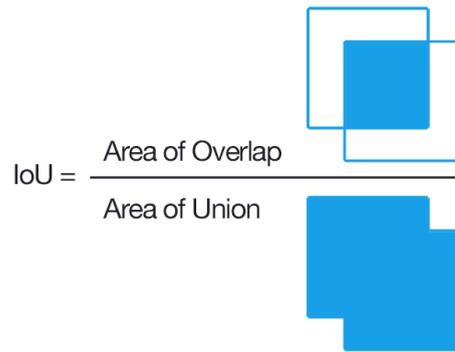
meningkatkan kemampuan ekspresi fitur dengan menggabungkan informasi resolusi rendah dari lapisan bawah dengan informasi resolusi tinggi dari lapisan atas. Akhirnya, mekanisme Pixel Attention (PA) diperkenalkan untuk meningkatkan akurasi segmentasi dengan mempelajari bobot berdasarkan pentingnya fitur pada posisi piksel yang berbeda, sehingga memungkinkan LETNet untuk menghasilkan peta segmentasi yang lebih halus dan detail. Modifikasi-modifikasi ini dirancang untuk mencapai keseimbangan optimal antara ukuran model, akurasi, dan kecepatan inferensi, menjadikannya sangat cocok untuk aplikasi real-time seperti kendaraan otonom. Adapun arsitektur modifikasi dapat dilihat pada Gambar 14.



Gambar 14 Arsitektur LETNet

2.7 Intersection over union (IoU)

Intersection over union (IoU), juga dikenal sebagai *Jaccard Index* dalam statistik, adalah salah satu metode paling mendasar untuk membandingkan kesamaan antara sampel data dalam *machine learning*. Dalam tugas visi komputer, IoU adalah salah satu metode evaluasi inti dalam berbagai tolak ukur termasuk *object detection*, *object tracking*, *semantic segmentation*, dll. Secara umum, IoU didefinisikan untuk menghitung tumpang tindih antara dua set elemen A dan B yang diberikan (J. Yu et al., 2021).



Gambar 15 Interest of Union (IoU)

IoU mengukur kesesuaian antara area yang diprediksi oleh model (misalnya, kotak pembatas objek) dengan area sebenarnya (*growth truth*). Nilai IoU dihitung dengan membagi area irisan (*intersection*) antara prediksi dan *growth truth* dengan area gabungan (*union*) dari kedua area tersebut. Nilai IoU berkisar antara 0 hingga 1, dimana nilai mendekati 1 menunjukkan prediksi yang sangat akurat, sedangkan nilai mendekati 0 menunjukkan prediksi yang tidak akurat.

Secara matematis, IoU didefinisikan sebagai:

$$IoU = \frac{A \cap B}{A \cup B} \quad (1)$$

dimana:

A = Nilai Prediksi

B = Nilai Aktual

2.8 Confussion Matrix

Confusion matrix adalah sebuah matriks yang menunjukkan jumlah kejadian antar dua nilai label, yaitu label aktual dan label prediksi. Sebuah *confusion matrix* memiliki tabel berukuran $l \times l$, dimana l adalah jumlah nilai label atau kelas yang berbeda (Kohavi & Provost, 1998). Pada *Multi-class Confusion Matrix*, kelas-kelas diurutkan dalam urutan yang sama dalam baris dan kolomnya seperti pada Gambar 14. Oleh karena itu, elemen yang diklasifikasi benar (*True Positive*) terletak pada diagonal utama dari kiri atas ke kanan bawah dan merupakan titik temu antara prediksi dan aktual (Grandini et al., 2020). Berikut merupakan contoh *Multi-class Confusion Matrix* yang ditunjukkan pada Gambar 16.

		PREDICTED classification				Total
		Classes	a	b	c	
ACTUAL classification	a	6	0	1	2	9
	b	3	9	1	1	14
	c	1	0	10	2	13
	d	1	2	1	12	16
Total		11	11	13	17	52

Gambar 16 Multi-class Confusion Matrix

2.8.3 Precision

		PREDICTED		Total
		Classes	Positive (1)	
ACTUAL	Positive (1)	TP = 20	FN = 5	25
	Negative (0)	FP = 10	TN = 15	25
Total		30	20	50

Gambar 17 Menghitung Nilai TP, FN, FP, dan TN pada *Confusion Matrix*

Precision (Pr) adalah nilai pecahan dari elemen *True Positive* dibagi dengan jumlah total unit yang diprediksi secara positif (jumlah kolom dari prediksi positif). Berdasarkan pada Gambar 17, persamaan *precision* dapat ditulis sebagai berikut,

$$Pr = \frac{TP}{TP+FP} \quad (2)$$

Dimana:

TP = *True Positive*,

FP = *False Positive*.

TP (*True Positive*) adalah elemen yang telah dilabeli sebagai positif pada model dan nilai sebenarnya positif, sedangkan FP (*False Positive*) adalah elemen yang telah dilabeli sebagai positif pada model, namun nilai sebenarnya negatif (Grandini et al., 2020).

2.8.4 Recall

Recall (Re) adalah nilai pecahan dari elemen *True Positive* (TP) dibagi dengan jumlah total unit yang diklasifikasi positif (jumlah baris dari yang sebenarnya bernilai positif). Persamaannya dapat ditulis sebagai berikut,

$$Re = \frac{TP}{TP+FN} \quad (3)$$

Dimana:

TP = *True Positive*,

FN = *False Negative*.

FN (*False Negative*) adalah elemen yang telah dilabeli sebagai negatif pada model, namun nilai sebenarnya positif (Grandini et al., 2020).

2.8.5 F1-Score

F1-score (F) menilai kinerja model klasifikasi dari *Confusion Matrix*. Rumus *f1-score* dapat diartikan sebagai rata-rata antara *precision* (Pr) dengan *recall* (Re), seperti persamaan berikut,

$$F = 2 \times \frac{Pr \times Re}{Pr + Re} \quad (4)$$

di mana *f1-score* mencapai nilai terbaiknya pada 1 dan skor terburuk pada 0. Kontribusi relatif dari *precision* dan *recall* adalah sama pada *f1-score* dan rata-rata *harmonic* berguna untuk menemukan *trade-off* terbaik antara dua kuantitas (Grandini et al., 2020).

2.9 Penelitian Terkait

Berikut beberapa penelitian terkait terhadap penelitian yang dilakukan:

1. U-Net: *Convolutional Networks for Biomedical Image Segmentation*

(Ronneberger et al., 2015)

Pada penelitian ini, Ronneberger memodifikasi arsitektur *Fully Convolutional Network* menjadi bentuk menyerupai huruf *U*. Dengan menggunakan dataset PhC-U373 yang merupakan kumpulan data gambar sel untuk kebutuhan biomedis. Ronneberger mendapatkan akurasi 92% menggunakan *Intersection over union* (IoU), yang secara signifikan lebih baik daripada algoritma terbaik kedua dengan 83%. Kumpulan data kedua “DICHela”³ adalah sel HeLa pada kaca datar direkam dengan mikroskop kontras interferensi diferensial (DIC). Ini berisi 20 gambar pelatihan beranotasi sebagian. Di sini peneliti mencapai rata-rata IoU 77,5% yang secara signifikan lebih baik daripada algoritma terbaik kedua dengan 46%.

2. Real-time lane departure warning with cascade lane segmentation

(Indrabayu et al., 2023)

Penelitian ini mengembangkan sistem peringatan keluar jalur secara real-time menggunakan teknik segmentasi jalur kaskade, yang dirancang khusus untuk kendaraan otonom. Sistem ini meningkatkan keamanan dengan memantau posisi kendaraan dan memberikan peringatan jika terdeteksi penyimpangan dari jalur. Dua komponen utama sistem ini adalah deteksi dan pengukuran jalur. Untuk mencapai ini, penelitian ini menggunakan segmentasi semantik yang diterapkan melalui arsitektur U-Net untuk mengklasifikasikan setiap piksel gambar, membedakan antara tanda jalur padat dan putus-putus serta jalan. Selain itu, sistem ini mengintegrasikan *Inverse Perspective Mapping* (IPM) untuk menghasilkan memfasilitasi perhitungan jarak yang lebih akurat. Hasil dari penelitian ini menunjukkan bahwa model yang dikembangkan mampu melakukan deteksi jalur dan pengukuran jarak dengan akurasi yang tinggi. Dari pengujian yang dilakukan, sistem mencapai akurasi *Intersection over union* (IoU) sebesar 79,8% dan hasil pengujian jarak menggunakan metode *Root Mean Square*

Error (RMSE) menunjukkan variasi dari RMSE terendah sebesar 0.025254 dan RMSE terbesar 0.134345.

3. Semantic Segmentation using Modified U-Net for Autonomous Driving (Sugirtha & Sridevi, 2022)

Penelitian ini membahas penggunaan modifikasi arsitektur U-Net dalam aplikasi pengemudian otonom melalui teknik segmentasi semantik. Segmentasi semantik sangat penting dalam pemahaman adegan urban untuk mendukung berbagai tugas mengemudi otonom seperti deteksi objek, perencanaan jalur, dan kontrol gerak. Dalam makalah ini, peningkatan akurasi model U-Net dilakukan dengan mengganti bagian *encoder* dengan arsitektur *Convolutional Neural Network* (CNN). Dua arsitektur CNN, yaitu VGG-16 dan ResNet-50, dibandingkan untuk mengevaluasi performanya. Penelitian ini menggunakan dataset *Cityscapes* yang luas untuk analisis, menunjukkan bahwa U-Net dengan *encoder* VGG-16 memberikan performa yang lebih baik dibandingkan dengan ResNet-50. Peningkatan rata-rata pada metrik *Intersection over union* (mIoU) sebesar 2% dicapai, menunjukkan efektivitas modifikasi pada U-Net dalam aplikasi segmentasi semantik untuk pengemudian otonom.

4. ShelfNet for Fast Semantic Segmentation (Zhuang et al., 2019)

Penelitian ini memperkenalkan ShelfNet, arsitektur baru untuk segmentasi semantik cepat dan akurat. Berbeda dari struktur *encoder-decoder* tunggal yang umum, ShelfNet memiliki beberapa pasangan cabang *encoder-decoder* dengan koneksi skip di setiap level spasial, yang menyerupai rak dengan beberapa kolom. Struktur seperti rak ini dapat dilihat sebagai gabungan dari beberapa jalur dalam dan dangkal, sehingga meningkatkan akurasi. Dibandingkan dengan model segmentasi *real-time* seperti BiSeNet, model mencapai akurasi yang lebih tinggi dengan kecepatan yang sebanding pada Dataset *Cityscapes*, memungkinkan aplikasi dalam tugas yang membutuhkan kecepatan seperti pemahaman adegan jalan untuk mengemudi otonom. Selanjutnya, ShelfNet kami mencapai 79,0% mIoU pada Dataset *Cityscapes* dengan *backbone* ResNet34, mengungguli PSPNet dan BiSeNet dengan *backbone* besar seperti ResNet101.

5. Lightweight Real-time Semantic Segmentation Network with Efficient Transformer and CNN (Xu et al., 2023)

Dalam penelitian ini, LETNet diusulkan sebagai jaringan segmentasi semantik yang ringan namun efektif, mengintegrasikan teknologi *Convolutional Neural Network* (CNN) dan *Transformer* dalam sebuah arsitektur yang dioptimalkan untuk aplikasi *real-time*. Model ini menggabungkan komponen utama yang inovatif untuk meningkatkan performa segmentasi. Komponen tersebut meliputi *Lightweight Dilated Bottleneck* (LDB), yang dirancang untuk mengurangi beban parameter sambil tetap mengekstraksi informasi semantik yang penting menggunakan konvolusi yang dilatasi dan konvolusi yang terpisah secara kedalaman. Selain itu, *Efficient Transformer* digunakan untuk mengolah ketergantungan jarak jauh dan meningkatkan representasi global model, yang membantu dalam memahami konteks visual secara keseluruhan dengan lebih baik. *Feature Enhancement Module* diperkenalkan untuk meningkatkan detail batas dan kualitas segmentasi saat resolusi gambar dipulihkan. Hasil dari eksperimen yang dilakukan pada dataset *Cityscapes* dan *CamVid* menunjukkan performa yang mengesankan dari LETNet. Model ini berhasil mencapai mIoU 72.8% pada set tes *Cityscapes* dan mIoU 70.5% pada set tes *CamVid* dengan menggunakan GPU RTX 3090.

Adapun State of The Art penelitian yangyang berhubungan dengan segmentasi jalan perkotaan pada *autonomous car* dipaparkan pada Tabel 1.

Tabel 1 State of the Art

No	Judul	Penulis & Tahun	Objek Permasalahan	Metode Penyelesaian	Hasil
1	U-Net: <i>Convolutional Networks for Biomedical Image Segmentation</i>	Ronneberger et al. (2015)	Objek: Citra sel biomedis Hela Permasalahan: Bagaimana performa segmentasi pada dataset	Modifikasi arsitektur Fully <i>Convolutional Network</i> (CNN) menjadi bentuk menyerupai huruf U	Mencapai rata-rata IoU 77,5% yang secara signifikan lebih baik daripada algoritma terbaik kedua

No	Judul	Penulis & Tahun	Objek Permasalahan	Metode Penyelesaian	Hasil
					dengan 46%.
2	Real-time lane departure warning with cascade lane segmentation		Objek: Jalur jalan kota Makassar Permasalahan:	U-Net, IPM	Sistem mencapai IoU sebesar 79,8% dan hasil pengujian jarak RMSE menunjukkan variasi dari 0.025254 hingga 0.134345
3	Semantic Segmentation using Modified U-Net for Autonomous Driving	T. Sugirtha and M. Sridevi (2022)	Objek: pemandangan jalan Perkotaan (dataset Cityscapes) Permasalahan: Bagaimana meningkatkan akurasi model U-Net dengan mengganti Bagian <i>Encoder</i> dengan arsitektur VGG16 dan Resnet50	U-Net, VGG16, ResNet50.	Meningkatan Performa Algoritma dalam segmentasi image dengan perhitungan kinerja mIoU masing masing 79,4% dan 80,5 %.
4	ShelfNet for Fast Semantic Segmentation	Zhuang et al. (2019)	Objek: Cityscapes Permasalahan: Bagaimana Performa Shelfnet dalam melakukan segmentasi	ShelfNet + Resnet50	Sistem ini mencapai 79,0% mIoU dengan backbone ResNet 34
5	Lightweight Real-time Semantic Segmentation Network with Efficient Transformer and CNN	Xu et al. (2023)	Objek: Dataset Cityscapes dan Cam Vid Dataset Permasalahan: Bagaimana performa LetNet dalam melakukan segmentasi	LetNet(U-Net+Transformer)	Sistem menghasilkan 72,8% mIoU pada Cityscapes dan 70,5% mIoU pada set uji CamVid