



TUGAS AKHIR - EC184801

**DETEKSI PEJALAN KAKI PADA ZEBRA CROSS
UNTUK PERINGATAN DINI PENGENDARA
MOBIL MENGGUNAKAN MASK R-CNN**

**Agung Wicaksono
NRP 0721 17 4000 0002**

**Dosen Pembimbing
Prof. Dr. Ir. Mauridhi Hery Purnomo, M.Eng.
Dr. Eko Mulyanto Yuniarno, S.T., M.T.**

**DEPARTEMEN TEKNIK KOMPUTER
Fakultas Teknologi ELEKTRO DAN INFORMATIKA CERDAS
Institut Teknologi Sepuluh Nopember
Surabaya 2021**



TUGAS AKHIR - EC184801

DETEKSI PEJALAN KAKI PADA ZEBRA CROSS UNTUK PERINGATAN DINI PENGENDARA MOBIL MENGGUNAKAN MASK R-CNN

**Agung Wicaksono
NRP 0721 17 4000 0002**

**Dosen Pembimbing
Prof. Dr. Ir. Mauridhi Hery Purnomo, M.Eng.
Dr. Eko Mulyanto Yuniarno, S.T., M.T.**

**DEPARTEMEN TEKNIK KOMPUTER
Fakultas Teknologi ELEKTRO DAN INFORMATIKA CERDAS
Institut Teknologi Sepuluh Nopember
Surabaya 2021**

PERNYATAAN KEASLIAN TUGAS AKHIR

Dengan ini saya menyatakan bahwa isi sebagian maupun keseluruhan Tugas Akhir sada dengan judul "**Deteksi Pejalan Kaki pada Zebra Cross untuk Peringatan Dini Pengendara Mobil menggunakan Mask R-CNN**" adalah benar-benar hasil karya intelektual mandiri, diselesaikan tanpa menggunakan bahan-bahan yang tidak diijinkan da bukan karya pihak lain yang saya akui sebagai karya sendiri.

Semua referensi yang dikutip maupun dirujuk telah ditulis secara lengkap pada daftar pustaka.

Apabila ternyata pernyataan ini tidak benar, saya bersedia menerima sanksi sesuai peraturan yang berlaku.

Surabaya, 10 Agustus 2021

Agung Wicaksono
0721 17 4000 0002

[Halaman ini sengaja dikosongkan]

LEMBAR PENGESAHAN

DETEKSI PEJALAN KAKI PADA ZEBRA CROSS UNTUK PERINGATAN DINI PENGENDARA MOBIL MENGGUNAKAN MASK R-CNN

Tugas Akhir ini disusun untuk memenuhi salah satu syarat memperoleh gelar Sarjana Teknik di Institut Teknologi Sepuluh Nopember Surabaya

Oleh: Agung Wicaksono (NRP. 0721 17 4000 0002)

Tanggal Ujian : 21 Juli 2021
Periode Wisuda : September 2021

Disetujui Oleh:

Prof. Dr. Ir. Mauridhi Hery Purnomo, M.Eng. (Pembimbing I)
NIP: 19580916 198601 1 001

Dr. Eko Mulyanto Yuniarso, S.T., M.T. (Pembimbing II)
NIP: 19680601 199512 1 009

Dr. I Ketut Eddy Purnama, S.T., M.T. (Penguji I)
NIP: 19690730 199512 1 001

Ahmad Zaini, ST., M.Sc. (Penguji II)
NIP: 19750419 200212 1 003

Diah Puspito Wulandari, ST., M.Sc. (Penguji III)
NIP: 19801219 200501 2 001

Ir. Hany Boedinugroho, M.T. (Penguji IV)
NIP: 19610706 198701 1 001

Mengetahui,
Kepala Departemen Teknik Komputer FTEIC - ITS

Dr. Supeno Mardi Susiki Nugroho, ST., MT.
NIP. 19700313 199512 1 001

ABSTRAK

Nama Mahasiswa : Agung Wicaksono
Judul Tugas Akhir : Deteksi Pejalan Kaki pada *Zebra Cross* untuk Peringatan Dini Pengendara Mobil menggunakan *Mask R-CNN*
Pembimbing : 1. Prof. Dr. Ir. Mauridhi Hery Purnomo, M.Eng.
 2. Dr. Eko Mulyanto Yuniarno, S.T., M.T.

Dewasa ini, fitur keselamatan pada kendaraan roda empat atau mobil sudah sangat berkembang pesat. Hal tersebut terbukti dengan banyaknya produsen mobil yang menerapkan teknologi *seat belt*, *air bag*, *adaptive cruise control*, *electronic stability control*, *autonomous emergency braking*, *blind spot monitoring* dan lain sebagainya. Namun, fitur yang sudah disebutkan diatas dinilai masih kurang ramah bagi pejalan kaki. Terbukti menurut data dari WHO, terdapat 270.000 pejalan kaki meninggal dunia setiap tahun atau sekitar 22% dari seluruh korban meninggal akibat kecelakan di jalan. Berawal dari permasalahan tersebut, penulis akan melakukan penelitian mengenai pendektsian pejalan kaki pada *zebra cross* untuk peringatan dini pengendara mobil sebagai topik penelitian. Pada tugas akhir ini, terdapat 3 objek yang akan dideteksi yaitu pejalan kaki, *zebra cross* dan pengendara motor dengan menggunakan metode *Mask R-CNN*. Hasil terbaik yang didapatkan adalah pada penggunaan *ResNet-101* untuk *backbone Mask R-CNN* dengan skor *mAP* sebesar 76.605%, *mAR* sebesar 85.375% serta *F1-Score* sebesar 80.302%.

Kata Kunci: Pejalan Kaki, *Zebra Cross*, *Mask R-CNN*, Pengolahan Citra.

[Halaman ini sengaja dikosongkan]

ABSTRACT

*Name : Agung Wicaksono
Title : Pedestrian Detection on Zebra Cross for Car Driver Early Warning using Mask R-CNN
Advisors : 1. Prof. Dr. Ir. Mauridhi Hery Purnomo, M.Eng.
2. Dr. Eko Mulyanto Yuniaro, S.T., M.T.*

Today, safety features on four-wheeled vehicles or cars have developed very rapidly. This is evidenced by the number of car manufacturers that apply seat belt technology, air bags, adaptive cruise control, electronic stability control, autonomous emergency braking, blind spot monitoring and so on. However, the features mentioned above are still considered less friendly for pedestrians. It is proven that according to data from the WHO, there are 270,000 pedestrians who die every year or about 22% of all victims die due to road accidents. Starting from these problems, the author will conduct research on the detection of pedestrians at zebra cross for early warning car drivers as a research topic. In this final project, there are 3 objects to be detected, namely pedestrians, zebra cross and motorcyclists using the Mask R-CNN method. The best results obtained are the use of ResNet-101 for backbone Mask R-CNN with a score of mAP of 76,605%, mAR of 85.375% and F1-Score of 80.302% .

Keywords: Pedestrian, Zebra Cross, Mask R-CNN, Image Processing

[Halaman ini sengaja dikosongkan]

KATA PENGANTAR

Puji dan syukur kehadirat Tuhan Yang Maha Esa atas segala karunia-Nya, penulis dapat menyelesaikan penelitian ini dengan judul **Deteksi Pejalan Kaki pada Zebra Cross untuk Peringatan Dini Pengendara Mobil menggunakan Mask R-CNN**.

Penelitian ini disusun dalam rangka pemenuhan bidang riset di Departemen Teknik Komputer ITS, sera digunakan sebagai persyaratan menyelesaikan pendidikan Sarjana. Penelitian ini dapat diselesaikan tidak lepas dari bantuan berbagai pihak. Oleh karena itu, penulis mengucapkan terimakasih kepada:

1. Keluarga, Ibu, Bapak dan Saudara tercinta yang telah memberikan dorongan baik secara spiritual dan material dalam penyelesaian buku penelitian ini.
2. Bapak Dr. Supeno Mardi Susiki Nugroho, ST., MT. selaku Kepala Departemen Teknik Komputer, Fakultas Teknologi Elektro dan Informatika Cerdas, Institut Teknologi Sepuluh Nopember.
3. Bapak Prof. Dr. Ir. Mauridhi Hery Purnomo, M.Eng. selaku dosen pembimbing I dan Bapak Dr. Eko Mulyanto Yuniarno, S.T., M.T. selaku dosen pembimbing II yang selalu memberikan arahan selama mengerjakan penelitian tugas akhir ini.
4. Bapak-ibu dosen pengajar Departemen Teknik Komputer, atas pengajaran dan bimbingan yang diberikan kepada penulis.
5. Seluruh teman-teman dari angkatan e57, Teknik Komputer, Laboratorium B401 dan B201 Teknik Komputer ITS serta Saturasi ITS.

Kesempurnaan hanya milik Allah SWT, untuk itu penulis memohon segenap kritik dan saran yang membangun. Semoga penelitian ini dapat memberikan manfaat bagi kita semua. Amin.

Surabaya, 10 Agustus 2021

Agung Wicaksono

[Halaman ini sengaja dikosongkan]

DAFTAR ISI

ABSTRAK	i
ABSTRACT	iii
DAFTAR ISI	vii
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Permasalahan	2
1.3 Tujuan	2
1.4 Batasan Masalah	3
1.5 Sistematika Penulisan	3
2 TINJAUAN PUSTAKA	5
2.1 <i>State of The Art</i>	5
2.1.1 Real-Time Pedestrian Detection With Deep Network Cascades	5
2.1.2 Pedestrian Detection: The Elephant In The Room	5
2.1.3 Fast Vehicle and Pedestrian Detection Using Improved Mask R-CNN	6
2.1.4 MASK R-CNN for Pedestrian Crosswalk Detection and Instance Segmentation	7

2.2	Teori Penunjang	8
2.2.1	<i>Machine Learning</i>	8
2.2.2	<i>Deep Learning</i>	9
2.2.3	<i>Convolutional Neural Network</i>	10
2.2.4	<i>Regional-Based CNN</i>	11
2.2.5	<i>Fast R-CNN</i>	12
2.2.6	<i>Faster R-CNN</i>	13
2.2.7	<i>Mask R-CNN</i>	14
2.2.8	<i>Regional Proposal Network (RPN)</i>	16
2.2.9	<i>Region of Interest Align</i>	16
2.2.10	Pejalan Kaki	17
2.2.11	<i>Zebra Cross</i>	17
2.2.12	ResNet-50	18
2.2.13	ResNet-101	18
2.2.14	MobileNet-v1	18
3	DESAIN DAN IMPLEMENTASI	21
3.1	Deskripsi Sistem	21
3.2	Pengumpulan <i>Dataset</i> Gambar	22
3.3	Pemisahan Data	23
3.4	<i>Pre-Processing</i>	24
3.5	Membangun Model Mask R-CNN	26
3.6	<i>Training Data</i>	28
3.7	<i>Validating Data</i>	29
3.8	<i>Testing Data</i>	30
4	PENGUJIAN DAN ANALISIS	33
4.1	Pengujian Jenis <i>Backbone</i>	33
4.1.1	Resnet-50	34

4.1.2	Resnet-101	38
4.1.3	MobileNet-V1	42
4.1.4	Perbandingan Hasil Prediksi	48
4.2	Pengujian Perbedaan Waktu	51
4.2.1	Pengujian di Pagi Hari	52
4.2.2	Pengujian di Siang Hari	54
4.2.3	Pengujian di Malam Hari	55
4.2.4	Perbandingan Hasil Evaluasi pada Perbedaan Waktu	57
5	PENUTUP	61
5.1	Kesimpulan	61
5.2	Saran	61
DAFTAR PUSTAKA		63
BIOGRAFI PENULIS		67

[Halaman ini sengaja dikosongkan]

DAFTAR GAMBAR

2.1	Gambaran <i>Supervised Learning</i> [1]	9
2.2	Gambaran <i>Unsupervised Learning</i> [2]	9
2.3	Gambaran <i>Reinforcement Learning</i> [3]	10
2.4	Gambaran Konsep Arsitektur CNN [4]	11
2.5	Arsitektur R-CNN [5]	12
2.6	Arsitektur <i>Fast R-CNN</i> [6]	13
2.7	Gambaran Deteksi Objek pada <i>Faster R-CNN</i> [7] . .	14
2.8	Struktur dari Arsitektur <i>Mask R-CNN</i> [8]	15
2.9	Arsitektur ResNet-50 [9]	18
2.10	Arsitektur ResNet-101 [10]	18
3.1	Blok Diagram Metodologi	21
3.2	Contoh Gambar dari Caltech Pedestrian Database .	22
3.3	Visualisasi Pembagian Data	23
3.4	Diagram Alir <i>Pre Processing</i>	25
3.5	Contoh <i>Image Resizing</i>	26
3.6	Contoh <i>Image Augmentation</i>	27
3.7	Blok Diagram Alur Mask R-CNN	28
3.8	Visualisasi <i>K Fold Cross Validation</i>	30
3.9	Diagram Alir <i>Testing Data</i>	31
4.1	Grafik Perubahan <i>Training Loss</i> pada <i>Resnet-50</i> . .	36
4.2	Grafik Perubahan <i>Validation Loss</i> pada <i>Resnet-50</i> . .	37
4.3	Grafik Perubahan <i>Validation mAP</i> pada <i>Resnet-50</i> . .	38

4.4	Grafik Perubahan <i>Training Loss</i> pada <i>Resnet-101</i>	41
4.5	Grafik Perubahan <i>Validation Loss</i> pada <i>Resnet-101</i>	42
4.6	Grafik Perubahan <i>Validation mAP</i> pada <i>Resnet-101</i>	43
4.7	Grafik Perubahan <i>Training Loss</i> pada <i>MobileNet-V1</i>	45
4.8	Grafik Perubahan <i>Validation Loss</i> pada <i>MobileNet-V1</i>	46
4.9	Grafik Perubahan <i>Validation mAP</i> pada <i>MobileNet-V1</i>	47
4.10	Grafik Perbandingan Model dengan <i>backbone</i> yang berbeda	48
4.11	Hasil Uji dari <i>Backbone Resnet-50</i>	49
4.12	Hasil Uji dari <i>Backbone Resnet-101</i>	50
4.13	Hasil Uji dari <i>Backbone Mobilenet-V1</i>	51
4.14	Perbandingan Hasil pada Pagi Hari	52
4.15	Perbandingan Hasil pada Siang Hari	54
4.16	Perbandingan Hasil pada Malam Hari	56
4.17	<i>Confusion Matrix</i> ResNet-50	58
4.18	<i>Confusion Matrix</i> ResNet-101	59
4.19	<i>Confusion Matrix</i> MobileNet-v1	60

DAFTAR TABEL

4.1	Spesifikasi <i>hardware Google Colaboratory</i>	33
4.2	Spesifikasi <i>hardware Komputer yang Digunakan</i>	33
4.3	Konfigurasi Model menggunakan Resnet-50	34
4.4	Konfigurasi Model menggunakan Resnet-101	38
4.5	Konfigurasi Model menggunakan Mobilenet-V1	43
4.6	Tabel Perbandingan Model dengan <i>backbone</i> yang berbeda	47
4.7	tabel Perbandingan Hasil <i>Testing</i>	51
4.8	Perbandingan Hasil Evaluasi pada Pagi Hari	53
4.9	Waktu Prediksi pada <i>File Input</i> Video dalam <i>MM:SS</i>	53
4.10	Perbandingan Hasil Evaluasi pada Siang Hari	55
4.11	Waktu Prediksi pada <i>File Input</i> Video dalam <i>MM:SS</i>	55
4.12	Perbandingan Hasil Evaluasi pada Malam Hari	56
4.13	Waktu Prediksi pada <i>File Input</i> Video dalam <i>MM:SS</i>	57

[Halaman ini sengaja dikosongkan]

BAB I

PENDAHULUAN

Penelitian ini di latar belakangi oleh berbagai kondisi yang menjadi acuan. Selain itu juga terdapat beberapa permasalahan yang akan dijawab sebagai luaran dari penelitian.

1.1 Latar Belakang

Mobil merupakan salah satu jenis kendaraan bermotor yang banyak terdapat di Indonesia. Pada tahun 2019 Badan Pusat Statistik mencatat terdapat 15.592.419 mobil penumpang yang berada di Indonesia. Dengan bertambahnya jumlah mobil di Indonesia dari tahun ke tahun, meningkatkan juga jumlah kecelakaan mobil. Fitur keselamatan dan keamanan pada mobil sangat penting bagi para pengendara dan penumpang, sehingga para produsen mobil berusaha meningkatkan teknologi keselamatan dan keamanan pada mobil buatannya. Sebagai contoh beberapa fitur keselamatan dan keamanan yang terdapat pada mobil antara lain, *adaptive cruise control*, *hill strat assist*, *blind spot monitoring*, *electronic stability control* dan lain sebagainya.

Menurut data dari WHO, terdapat 270.000 pejalan kaki meninggal dunia setiap tahun atau sekitar 22% dari seluruh korban meninggal akibat kecelakaan di jalan[11]. Sedangkan untuk di Indonesia sendiri, mengutip dari laman *Global Road Safety Facility*, presentase kematian pejalan kaki akibat kecelakaan lalu lintas sebesar 38% dari total 31.282 kematian di jalan raya yang dilaporkan pada tahun 2016 [12]. Melihat kegiatan para pejalan kaki yang jarang berada di badan jalan, angka tersebut tentu cukup tinggi. Para pejalan kaki hanya menggunakan badan jalan ketika hendak menyebrang jalan lewat *zebra cross*. Kelalaian dari pejalan kaki maupun pengendara mobil merupakan faktor utama mengapa angka kematian pejalan kaki cukup tinggi. Salah satu contoh kelalaian pejalan kaki adalah pada saat menyebrang jalan tidak memperhatikan kendaraan yang

akan lewat dan atau mengabaikan rambu serta lampu lalu lintas. Di sisi pengendara mobil, kelelahan, kurangnya fokus saat berkendara dan tidak memperhatikan rambu maupun marka dapat berakibat fatal baik kepada pejalan kaki dan pengendara lain.

Teknologi *artificial intelligent* tentu dapat digunakan untuk mengatasi masalah yang sudah disebutkan pada penjelasan sebelumnya. Melihat penerapan *artificial intelligent* pada teknologi keselamatan pengendara mobil seperti *adaptive cruise control*, *hill start assist* dan lain sebagainya. Bagian dari *artificial intelligent* yang sangat cocok untuk masalah seperti ini adalah *deep learning* dengan memanfaatkan pengolahan citra pada data berbentuk gambar tentu dapat digunakan untuk deteksi pejalan kaki di *zebra cross* guna mengurangi jumlah korban akibat kecelakaan. Deteksi pejalan kaki ini kedepannya dapat digabungkan dengan *buzzer* dan atau *LED* sebagai komponen *output* untuk mengingatkan kepada pengendara bahwa ada pejalan kaki yang sedang menyebrangi jalan serta mengembalikan fokus untuk berkendara.

1.2 Permasalahan

Berdasarkan data yang telah dipaparkan pada latar belakang, cukup tingginya angka kematian pejalan kaki akibat kecelakaan lalu lintas dikarenakan belum adanya deteksi pejalan kaki di *zebra cross* untuk peringatan dini kepada pengendara mobil. Oleh karena itu, diperlukan sebuah sistem yang mampu mendeteksi adanya pejalan kaki yang berada disekitar jalan raya untuk selanjutnya dapat digunakan sebagai peringatan kepada pengendara mobil.

1.3 Tujuan

Berdasarkan rumusan permasalahan di atas, tujuan dari penelitian ini adalah untuk mendeteksi dan mensegmentasikan pejalan kaki dan *zebra cross* di jalan raya untuk peringatan dini pengendara mobil menggunakan metode *Mask R-CNN*.

1.4 Batasan Masalah

Batasan masalah yang timbul dari permasalahan Tugas Akhir ini adalah:

1. Menggunakan *Mask R-CNN* untuk pendekripsi pejalan kaki dan *zebra cross*
2. *File Input* berupa video dengan format *MP4* yang diambil dari sudut pandang pengendara mobil.

1.5 Sistematika Penulisan

Laporan penelitian tugas akhir ini tersusun dalam sistematika dan terstruktur sehingga mudah dipahami dan dipelajari oleh pembaca maupun seseorang yang ingin melanjutkan penelitian ini. Alur sistematika penulisan laporan penelitian ini yaitu :

1. BAB I Pendahuluan

Bab ini berisi uraian tentang latar belakang permasalahan, penegasan dan alasan pemilihan judul, sistematika laporan, tujuan dan metodologi penelitian.

2. BAB II Tinjauan Pustaka

Pada bab ini berisi tentang uraian secara sistematis teori-teori yang berhubungan dengan permasalahan yang dibahas pada penelitian ini. Teori-teori ini digunakan sebagai dasar dalam penelitian, yaitu informasi terkait pejalan kaki, *zebra cross*, algoritma *Mask RCNN*, dan teori-teori penunjang lainnya.

3. BAB III Desain dan Implementasi Sistem

Bab ini berisi tentang penjelasan-penjelasan terkait eksperimen yang akan dilakukan dan langkah-langkah pengambilan data jalan raya serta proses deteksi pejalan kaki pada *zebra cross*. Guna mendukung hal tersebut, digunakanlah blok diagram atau work flow agar sistem yang akan dibuat dapat terlihat dan mudah dibaca untuk implementasi pada pelaksanaan tugas akhir.

4. BAB IV Pengujian dan Analisa

Bab ini menjelaskan tentang pengujian eksperimen yang dilakukan terhadap citra jalan raya, proses klasifikasi pejalan kaki dan *zebra cross*. Serta terkait tingkat akurasi keberhasil-

an pengujian yang dilengkapi dengan analisanya.

5. BAB V Penutup

Bab ini merupakan penutup yang berisi kesimpulan yang diambil dari penelitian dan pengujian yang telah dilakukan. Saran dan kritik yang membangun untuk pengembangan lebih lanjut juga dituliskan pada bab ini.

BAB II

TINJAUAN PUSTAKA

Demi mendukung penelitian ini, dibutuhkan beberapa *state of the art* dan teori penunjang sebagai bahan acuan dan referensi. Dengan demikian penelitian ini menjadi lebih terarah.

2.1 *State of The Art*

2.1.1 Real-Time Pedestrian Detection With Deep Network Cascades

Penelitian ini dilakukan oleh Anelia Angelova dan kawan-kawan pada tahun 2015 yang menyajikan pendekatan *real-time* baru untuk deteksi objek yang mengeksplorasi efisiensi *cascade classifiers* dengan akurasi *deep neural network* [13]. *Deep network* telah terbukti unggul dalam tugas klasifikasi, dan kemampuannya untuk beroperasi pada *raw pixel input* tanpa perlu merancang fitur khusus. Namun, *deep network* terkenal lambat pada waktu inferensi. Dalam *paper* tersebut, Anelia Angelova dan kawan-kawan mengusulkan pendekatan *cascades deep network* dan *fast features*, yang sangat cepat dan sangat akurat. Mereka menerapkannya pada permasalahan pada deteksi pejalan kaki. Algoritma mereka berjalan secara *real-time* pada 15 *frame* per detik. Pendekatan yang dihasilkan mencapai tingkat kesalahan rata-rata 26,2% pada *benchmark* deteksi Caltech Pedestrian.

2.1.2 Pedestrian Detection: The Elephant In The Room

Deteksi pejalan kaki digunakan di banyak aplikasi berbasis citra mulai dari pengawasan video hingga pengemudi otonom. Meskipun mencapai kinerja tinggi, sebagian besar masih belum diketahui seberapa baik detektor yang ada menggeneralisasi data yang tidak terlihat. Hal ini penting karena detektor yang praktis harus siap

digunakan dalam berbagai skenario dalam aplikasi. Untuk tujuan tersebut, Irtiza Hasan dan kawan-kawan melakukan studi komprehensif dalam *paper* ini, menggunakan prinsip umum evaluasi *cross-dataset* langsung [14]. Melalui *paper* ini, ditemukan bahwa detektor pejalan kaki terkini yang ada, meskipun berkinerja cukup baik ketika dilatih dan diuji pada kumpulan data yang sama, namun secara umum memiliki performa yang cukup buruk dalam evaluasi *cross-dataset*.

Dalam *paper* ini ditunjukkan bahwa ada dua alasan untuk permasalahan ini. Pertama, desain yang dibuat (misalnya, *anchor settings*) mungkin bias terhadap tolok ukur dalam *training* dan *test pipeline* data tunggal, tetapi akibatnya sebagian besar membatasi kemampuan generalisasi dari keduanya. Kedua, sumber pelatihan umumnya tidak terlalu padat pada pejalan kaki dan mempunyai beragam dalam skenario. Di dalam evaluasi *cross-dataset* langsung, secara mengejutkan, ditemukan bahwa detektor objek dengan tujuan umum, tanpa adaptasi khusus untuk pejalan kaki dalam desain, digeneralisasi jauh lebih baik dibandingkan dengan detektor pejalan kaki terkini yang ada. Lebih lanjut, diilustrasikan bahwa kumpulan data yang beragam dan padat, yang dikumpulkan dengan *crawling web*, berfungsi sebagai sumber pra-pelatihan yang efisien untuk deteksi pejalan kaki. Oleh karena itu, pada *paper* mengusulkan *training pipeline* progresif dan menemukan bahwa *pipeline* tersebut berfungsi dengan baik untuk deteksi pejalan kaki yang berorientasi pada pengemudian otonom. Akibatnya, studi yang dilakukan dalam makalah ini menunjukkan bahwa lebih banyak penekanan harus diberikan pada evaluasi *cross-dataset* untuk desain masa mendatang pada detektor pejalan kaki.

2.1.3 Fast Vehicle and Pedestrian Detection Using Improved Mask R-CNN

Penelitian ini menyajikan algoritma Mask R-CNN yang sederhana dan efektif untuk deteksi kendaraan dan pejalan kaki yang lebih cepat [15]. Metode ini memiliki nilai praktis untuk sistem peringatan anti-tabrakan dalam mengemudi cerdas. *Deep Neural Network* dengan lebih banyak lapisan memiliki kapasitas yang lebih besar, tetapi juga harus melakukan perhitungan yang lebih rumit.

Untuk mengatasi kelemahan ini, penelitian ini mengadopsi jaringan Resnet-86 sebagai *backbone* yang berbeda dari struktur tulang punggung Resnet-101 dalam algoritma Mask R-CNN. Hasilnya menunjukkan bahwa jaringan Resnet-86 dapat mengurangi waktu operasi dan sangat meningkatkan akurasi. Kendaraan dan pejalan kaki yang terdeteksi juga disaring berdasarkan dataset Microsoft COCO. Dataset baru dibentuk dengan menyaring dan melengkapi COCO dataset, yang membuat pelatihan algoritma lebih efisien. Bagian terpenting dari penelitian ini adalah diusulkannya algoritma baru, *Side Fusion FPN*. Parameter dalam algoritma tidak ada perubahan, jumlah perhitungan meningkat kurang dari 0,000001, dan rata-rata presisi (mAP) meningkat 2,00 poin. Hasilnya menunjukkan bahwa, dibandingkan dengan algoritma Mask R-CNN, algoritma pada penelitian ini menurunkan ukuran memori bobot sebesar 9,43%, meningkatkan kecepatan pelatihan sebesar 26,98%, meningkatkan kecepatan pengujian sebesar 7,94%, menurunkan nilai *error* sebesar 0,26, dan meningkatkan nilai mAP sebesar 17,53 poin.

2.1.4 MASK R-CNN for Pedestrian Crosswalk Detection and Instance Segmentation

Pejalan kaki paling rentan mengalami kecelakaan mengingat mayoritas pengendara mengecualikan mereka sebagai pengguna jalan [16]. Pada penelitian ini, pendekslsian objek menggunakan Mask Region-Based CNN dan segmentasi instance diterapkan pada penyeberangan pejalan kaki. Pelatihan dilakukan menggunakan Mask R-CNN untuk deteksi objek dengan backbone ResNet-101, dengan kecepatan pembelajaran 0,001 dan 2 gambar per GPU selama 30 epoch dari 100 batch. Berdasarkan penelitian ini, 500 gambar penyeberangan pejalan kaki dikumpulkan dan dipilih untuk validasi dan pelatihan. 80% dari gambar adalah untuk set pelatihan dan untuk set validasi adalah 20%. 30 gambar pengujian penyeberangan pejalan kaki lainnya dikumpulkan untuk evaluasi model guna memverifikasi stabilitas dan keandalan model yang terlatih. Semua 30 gambar pengujian telah terdeteksi dan akurasi deteksi lebih besar dari 97%. Jika ada 2 atau lebih penyeberangan pejalan kaki dalam sebuah gambar, maka akan membuat warna MASK berbeda untuk setiap deteksi. Ringkasan hasil tes memverifikasi bahwa semua da-

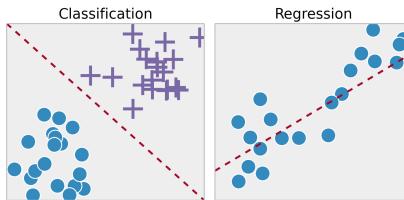
ta yang dikumpulkan lebih tinggi dari 97% untuk dapat mendeteksi pejalan kaki.

2.2 Teori Penunjang

2.2.1 *Machine Learning*

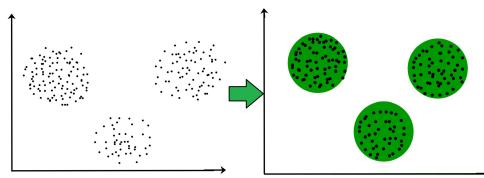
Machine Learning adalah studi tentang algoritma komputer yang memberikan sistem kemampuan untuk belajar secara otomatis dan dapat meningkatkan kemampuan dari pengalaman yang sudah didapatkan [17]. Hal ini umumnya dilihat sebagai sub-bidang kecerdasan buatan. Algoritma pembelajaran mesin memungkinkan sistem membuat keputusan secara mandiri tanpa dukungan eksternal. Keputusan semacam itu dibuat dengan menemukan pola dasar yang berharga dalam data yang kompleks. Berdasarkan pendekatan pembelajaran, jenis data *input* dan *output*, dan jenis masalah yang dipecahkan, ada beberapa kategori utama dari algoritma *machine learning supervised*, *unsupervised* dan *reinforcement learning*. Ada beberapa pendekatan hibrida dan metode umum lainnya yang menawarkan ekstrapolasi alami dari bentuk masalah pembelajaran mesin. Berikut merupakan penjelasan dari beberapa kategori utama dari algoritma *machine learning*:

1. *Supervised Learning* diterapkan ketika data dalam bentuk variabel input dan nilai target output. Algoritma akan mempelajari fungsi pemetaan dari *input* ke *output*. Ketersediaan sampel data berlabel dengan skala besar mempunyai nilai yang tinggi dikarenakan masih terdapat kelangkaan *dataset*. Pendekatan ini secara luas dapat dibagi menjadi dua kategori utama yaitu *classification* dan *regression*. Gambar 2.1 menampilkan visualisasi dari *classification* dan *regression* pada *Supervised Learning*
2. *Unsupervised Learning* diterapkan ketika data hanya tersedia dalam bentuk *input* dan tidak ada variabel *output* yang sesuai. Algoritma semacam itu memodelkan pola yang mendasari data untuk mempelajari lebih lanjut tentang karakteristiknya. Salah satu jenis utama dari algoritma *unsupervised* adalah pengelompokan. Dalam teknik ini, kelompok yang melekat da-



Gambar 2.1: Gambaran *Supervised Learning*[1]

lam data ditemukan dan kemudian digunakan untuk memprediksi *output* untuk *input* yang tidak terlihat. Contoh dari teknik ini adalah untuk memprediksi perilaku pembelian pada pelanggan. Gambar 2.2 merupakan visualisasi dari algoritma *unsupervised learning*.

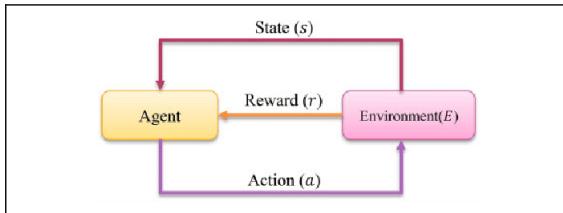


Gambar 2.2: Gambaran *Unsupervised Learning*[2]

3. *Reinforcement learning* diterapkan ketika tugas yang ada adalah membuat urutan keputusan menuju *reward* akhir. Selama proses *learning*, *artificial agent* mendapat *reward* atau *penalties* atas tindakan yang dilakukannya. Tujuannya adalah untuk memaksimalkan total *reward* yang didapatkan. Gambar 2.3 merupakan visualisasi dari algoritma *reinforcement learning*.

2.2.2 Deep Learning

Deep Learning adalah kelas *machine learning* yang berkinerja jauh lebih baik pada data tidak terstruktur[18]. Teknik *deep learning* mengungguli teknik *machine learning* saat ini. Ini memungkinkan model komputasi untuk mempelajari fitur secara progresif



Gambar 2.3: Gambaran *Reinforcement Learning*[3]

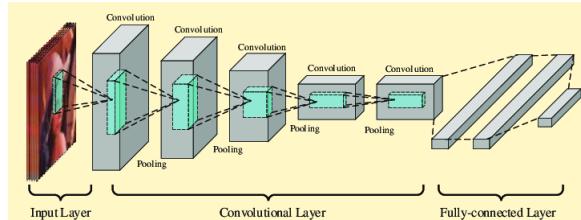
dari data di berbagai level. Popularitas *deep learning* diperkuat karena jumlah data yang tersedia meningkat serta kemajuan perangkat keras yang menyediakan komputer yang kuat.

Arsitektur *deep learning* berkinerja lebih baik daripada jaringan saraf tiruan sederhana, meskipun waktu *learning* dari struktur *deep learning* lebih tinggi dari jaringan saraf tiruan. Namun, waktu *learning* dapat dikurangi dengan menggunakan metode seperti *transfer learning* atau komputasi menggunakan GPU. Salah satu faktor yang menentukan keberhasilan jaringan saraf terletak pada desain arsitektur jaringan yang cermat.

2.2.3 Convolutional Neural Network

Convolutional Neural Network (CNN) adalah jenis khusus dari *multilayer neural network* atau arsitektur *deep learning* yang terinspirasi oleh sistem visual makhluk hidup [19]. CNN sangat cocok untuk berbagai bidang visi komputer dan *natural language processing*. *Convolutional Neural Network* (CNN), juga disebut *ConvNet*, adalah jenis *Artificial Neural Network* (ANN), yang memiliki arsitektur *feed-forward* yang dalam dan memiliki kemampuan generalisasi yang luar biasa dibandingkan dengan jaringan lain dengan lapisan FC (*Fully Connected*), ia dapat mempelajari fitur objek yang sangat abstrak terutama data spasial dan dapat mengidentifikasi-nya dengan lebih efisien. Model CNN yang dalam terdiri dari satu set lapisan pemrosesan yang dapat mempelajari berbagai fitur data *input* (misalnya gambar) dengan beberapa tingkat abstraksi seperti yang ditampilkan pada Gambar 2.4. Lapisan inisiatör mempelajari dan mengekstrak fitur tingkat tinggi (dengan abstraksi yang lebih

rendah), dan lapisan yang lebih dalam mempelajari dan mengekstrak fitur tingkat rendah (dengan abstraksi yang lebih tinggi).



Gambar 2.4: Gambaran Konsep Arsitektur CNN [4]

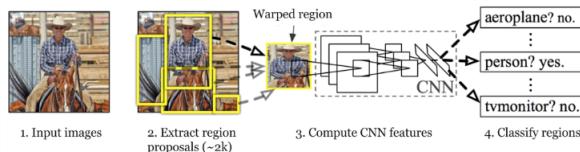
Convolutional Neural Network memiliki beberapa keunggulan dibanding dengan jaringan saraf tiruan lainnya dalam konteks visi komputer, antara lain :

1. Salah satu alasan utama untuk mempertimbangkan CNN dalam kasus tersebut adalah fitur pembagian bobot dari CNN, yang mengurangi jumlah parameter yang dapat dilatih dalam jaringan, yang membantu model untuk menghindari *overfitting* dan juga untuk meningkatkan generalisasi.
2. Pada CNN, lapisan klasifikasi dan lapisan ekstraksi fitur melakukan proses *learning* secara bersama-sama, yang membuat output model lebih terorganisir dan membuat output lebih bergantung pada fitur yang diekstraksi.
3. Implementasi pada jaringan dengan ukuran yang besar akan lebih sulit dilakukan dengan menggunakan jenis jaringan saraf lain daripada menggunakan *Convolutional Neural Network*

2.2.4 *Regional-Based CNN*

Semenjak *Convolution Neural Network* (CNN) dengan *fully connected layer* tidak mampu menangani frekuensi kemunculan dan multi objek. Salah satu cara untuk mengatasinya adalah dengan menggunakan *sliding window brute force search* untuk memilih wilayah dan menerapkan model CNN pada area tersebut, tetapi ma-

salah dari pendekatan ini adalah bahwa objek yang sama dapat di-representasikan dalam gambar dengan ukuran dan aspek rasio yang berbeda. Dengan mempertimbangkan faktor-faktor tersebut, penerapan algoritma *deep learning* (CNN) pada jumlah *region proposal* yang banyak akan menyebabkan proses komputasi yang dijalankan akan menjadi sangat berat dan rumit.



Gambar 2.5: Arsitektur R-CNN [5]

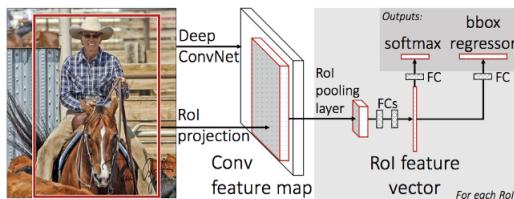
Ross Girshick dan kawan-kawan pada tahun 2013 mengusulkan arsitektur baru yang disebut R-CNN seperti pada Gambar 2.5 (*Regional-based CNN*) untuk menghadapi tantangan deteksi objek ini [20]. Arsitektur R-CNN menggunakan algoritma *selective search* yang menghasilkan sekitar 2000 *region proposal*. *Regional proposal* ini kemudian diterapkan ke arsitektur CNN untuk menghitung jumlah fitur CNN. Fitur-fitur ini kemudian diteruskan dalam model SVM untuk mengklasifikasikan objek yang ada di *region proposal*. Langkah selanjutnya adalah melakukan regresi pada *bounding box* untuk mengetahui lokasi objek yang ada dalam gambar dengan lebih tepat.

2.2.5 Fast R-CNN

Fast R-CNN ditemukan atas metode yang telah ditemukan terlebih dahulu yaitu *R-CNN* untuk mengklasifikasikan proposal objek secara efisien menggunakan *deep convolutional network* [21]. Dibandingkan dengan *R-CNN*, *Fast R-CNN* menggunakan beberapa inovasi untuk meningkatkan kecepatan *learning* dan pengujian sekaligus meningkatkan akurasi deteksi. Fast R-CNN dapat menyelesaikan proses *training* dengan jaringan VGG16 yang sangat dalam membebarkan hasil yang 9 kali lebih cepat dari R-CNN, 213 kali lebih cepat pada waktu pengujian, dan mencapai mAP yang lebih tinggi pada PASCAL VOC 2012. Dibandingkan dengan SPPnet,

Fast R-CNN dapat menyelesaikan proses *training* dengan VGG16 3x lebih cepat, pengujian 10x lebih cepat, dan lebih akurat.

Jika pada R-CNN *region proposal* akan melalui proses konvolusi di CNN, sebaliknya pada *Fast R-CNN* gambar *input*-lah yang akan melalui proses konvolusi di CNN seperti yang ditampilkan pada Gambar 2.6. Hasil dari konvolusi pada *Fast R-CNN* berupa *convolutional feature map* yang akan digunakan untuk identifikasi *region proposal* dan menyatukannya ke dalam bentuk persegi. Selanjutnya dengan menggunakan layer *ROI Pooling* akan dibentuk kembali menjadi ukuran yang tetap sehingga dapat dimasukan ke dalam *fully connected network*. Pada proses deteksi kelas dari *region proposal* dan juga nilai offset dari *bounding box* digunakan *softmax layer* dari *ROI feature vector* yang telah dihasilkan pada proses sebelumnya.



Gambar 2.6: Arsitektur *Fast R-CNN* [6]

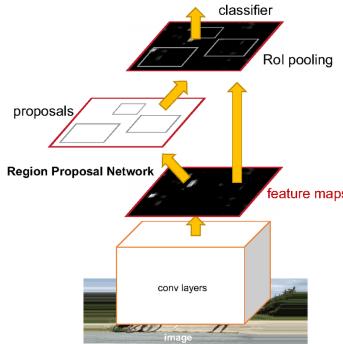
Alasan *Fast R-CNN* lebih cepat daripada R-CNN adalah karena kita tidak perlu memasukkan 2000 *region proposal* ke *convolutional neural network* setiap saat. Sebagai gantinya, operasi konvolusi dilakukan hanya sekali per gambar dan *feature map* dihasilkan operasi tersebut.

2.2.6 Faster R-CNN

Pad kedua algoritma *R-CNN* dan *Fast R-CNN* menggunakan *selective search* untuk mengetahui *region proposal*. *Selective search* sendiri memerlukan waktu yang cukup lama dalam penyelesaian prosesnya sehingga mempengaruhi kinerja dari jaringan [22]. Oleh karena itu, dibuatlah algoritma deteksi objek baru dengan menghilangkan algoritma *selective search* dan memungkinkan jaringan

mempelajari *region proposal*.

Mirip dengan *Fast R-CNN*, gambar dijadikan sebagai input ke jaringan konvolusi yang menyediakan *convolusional feature map*. Disebabkan waktu eksekusi yang lama saat menggunakan algoritma *selective search* pada *feature map* untuk mengidentifikasi *region proposal*, *Faster R-CNN* menggunakan jaringan terpisah untuk memprediksi *region proposal*. Jaringan terpisah ini dinamakan dengan *Region Proposal Network (RPN)*, dimana *RPN* berbagi *full-image convolutional features* dengan jaringan deteksi yaitu *Fast R-CNN*. *Region proposal* yang diprediksi kemudian dibentuk kembali menggunakan *ROI pooling layer* yang kemudian digunakan untuk mengklasifikasikan gambar di dalam *region proposal* dan memprediksi nilai offset untuk *bounding box*. Gambar 2.7 merupakan visualisasi dari proses deteksi objek menggunakan *Faster R-CNN*.



Gambar 2.7: Gambaran Deteksi Objek pada *Faster R-CNN* [7]

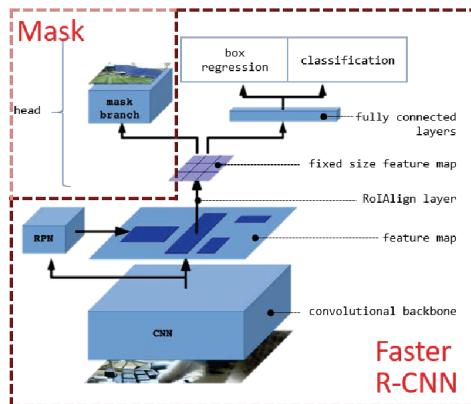
2.2.7 *Mask R-CNN*

Mask R-CNN adalah Convolutional Neural Network (CNN) dan segmentasi citra termutakhir untuk saat ini. Varian Deep Neural Network ini mendeteksi objek dalam gambar dan menghasilkan *mask* segmentasi berkualitas tinggi untuk setiap *instance* [23]. *Mask R-CNN* dibangun menggunakan *Faster R-CNN* dimana *Faster R-CNN* memiliki 2 *output* untuk setiap objek, label kelas dan *bounding box offset*. *Mask R-CNN* adalah penambahan cabang ketiga yang

mengeluarkan *mask* objek seperti yang tertampil pada Gambar 2.8. *output mask* tambahan berbeda dari *output* kelas dan *bounding box*, yang membutuhkan ekstraksi tata letak spasial yang jauh lebih baik dari suatu objek.

Mask R-CNN merupakan perpanjangan dari *Faster R-CNN* dengan menambahkan cabang untuk memprediksi *tmask* objek (*Region of Interest*) secara paralel dengan cabang yang ada untuk pengenalan *bounding box*. Satu keuntungan sederhana dari *Mask R-CNN* dibandingkan *Faster R-CNN* adalah kenyataan bahwa mudah untuk menggeneralisasi tugas lain seperti estimasi pose. Elemen kunci *Mask R-CNN* adalah penyelarasan piksel-ke-piksel, yang merupakan bagian utama dari *Fast/Faster R-CNN* yang hilang. *Mask R-CNN* mengadopsi prosedur dua tahap yang sama dengan tahap pertama yang identik (yaitu *RPN*). Pada tahap kedua, secara paralel untuk memprediksi kelas dan *box offset*, *Mask R-CNN* juga mengeluarkan *mask* biner untuk setiap *RoI*. Ini berbeda dengan sistem terbaru, di mana klasifikasi bergantung pada prediksi *mask*.

Mask R-CNN mudah diterapkan dan dilatih karena *Faster R-CNN framework*, yang memfasilitasi berbagai desain arsitektur yang fleksibel. Selain itu, cabang *mask* hanya menambahkan *overhead* komputasi kecil, memungkinkan sistem dan eksperimen yang cepat.



Gambar 2.8: Struktur dari Arsitektur *Mask R-CNN* [8]

2.2.8 *Regional Proposal Network (RPN)*

Regional Proposal bermanfaat untuk mengetahui kemungkinan lokasi target pada gambar ke tingkat lebih akurat, yang dapat memastikan bahwa tingkat penarikan yang lebih tinggi dipertahankan ketika lebih sedikit jendela yang dipilih. Selanjutnya jendela kandidat yang diperoleh dari proses tersebut, memiliki kualitas yang lebih tinggi dibandingkan dengan algoritma *Sliding Window* pada umumnya. Algoritma *Region Proposal* yang lebih umum digunakan adalah *Selective Search* dan *Bounding Box* [24].

Anchor boxes merupakan satu kumpulan *bounding box* dengan memiliki ukuran dan rasio aspek berbeda yang membentang di seluruh gambar *input* untuk melatih *region proposal network*. Pertama, akan digunakan beberapa lapisan awal dari jaringan yang telah dilatih sebelumnya untuk mengidentifikasi fitur yang menjanjikan dari gambar *input*. Beberapa lapisan awal dari jaringan, akan belajar mendeteksi fitur umum, seperti tepi dan bintik warna pada gambar.

RPN menggunakan klasifikasi dua kelas, yang hanya membedakan latar belakang dari objek, tetapi tidak memprediksi kelas objek. Misalkan setelah gambar dimasukkan dan mengalami serangkaian konvolusi dan *pooling* di *backbone network*, peta fitur ukuran $M \times N$ diperoleh, yang sesuai dengan membagi gambar asli menjadi area $M \times N$. Pusat setiap area gambar asli diwakili oleh koordinat piksel pada peta fitur ini.

RPN digunakan untuk menentukan apakah *anchor boxes* yang sesuai dengan setiap piksel, berisi target objek. Lapisan RPN harus belajar mengklasifikasikan *anchor boxes* sebagai latar belakang atau latar depan, dan menghitung koefisien regresi untuk mengubah posisi, lebar, dan tinggi *anchor boxes* latar depan.

2.2.9 *Region of Interest Align*

Region of Interest Align, atau *RoIAlign*, adalah operasi untuk mengekstraksi peta fitur berukuran kecil dari setiap *RoI* untuk deteksi dan segmentasi. Hal ini menghilangkan kuantisasi pada *RoI Pool* serta menyelaraskan fitur yang diekstraksi dengan *input* [23]. Untuk menghindari kuantisasi batasan pada *RoI*, *RoIAlign* meng-

gunakan interpolasi bilinear untuk menghitung nilai yang tepat dari fitur *input* di empat lokasi sampel reguler di setiap *bin ROI*, dan hasilnya kemudian dikumpulkan berdasarkan nilai maksimum atau rata-rata.

Ide *ROI Align* sangat sederhana yaitu dengan menghapus operasi kuantisasi, dan menggantinya dengan menggunakan metode interpolasi bilinear untuk mendapatkan nilai gambar pada piksel dengan koordinat *floating point*, sehingga mengubah seluruh proses agregasi fitur menjadi operasi berkelanjutan. Perlu dicatat bahwa dalam operasi algoritma tertentu, *ROI Align* tidak hanya melengkapi titik koordinat pada batas area kandidat, dan kemudian menggabungkan titik koordinat ini, tetapi mendesain ulang serangkaian proses yang lebih baik.

2.2.10 Pejalan Kaki

Definisi Pejalan kaki adalah orang yang melakukan aktifitas berjalan kaki dan merupakan salah satu unsur pengguna jalan. (Keputusan Direktur Jendral Perhubungan Darat : SK.43/AJ 007/DR-JD/97). Pejalan kaki harus berjalan pada bagian jalan yang diperuntukan bagi pejalan kaki, atau pada bagian pejalan kaki, atau pada bagian jalan yang paling kiri apabila tidak terdapat bagian jalan yang diperuntukan bagi pejalan kaki (PP No. 43 , 1993)

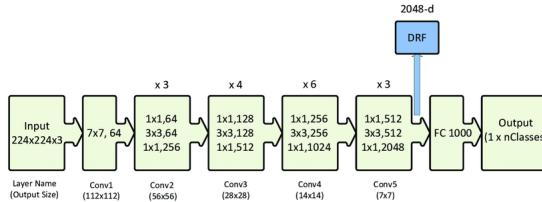
2.2.11 Zebra Cross

Zebra cross merupakan fasilitas penyeberangan bagi pejalan kaki sebidang yang dilengkapi marka untuk memberi ketegasan/batas dalam melakukan lintasan. *Zebra cross* dipasang dengan ketentuan sebagai berikut :

1. *Zebra cross* harus dipasang pada arus lalu lintas, kecepatan lalu lintas, dan arus pejalan kaki yang relatif rendah.
2. Lokasi penempatan *zebra cross* harus memiliki jarak pandang yang cukup, agar tundaan kendaraan yang diakibatkan oleh penggunaan fasilitas penyebrangan masih dalam kondisi aman.

2.2.12 ResNet-50

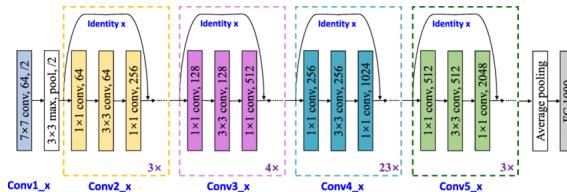
ResNet50 adalah varian dari model ResNet yang memiliki 48 *convolution layer* bersama dengan 1 *MaxPool layer* dan 1 *Average-Pool layer*. ResNet-50 memiliki 3.8×10^9 Operasi *floating point*. Gambar 2.9 merupakan diagram proses dari arsitektur ResNet-50.



Gambar 2.9: Arsitektur ResNet-50 [9]

2.2.13 ResNet-101

ResNet101 adalah varian dari model ResNet yang memiliki 99 *convolution layer* bersama dengan 1 *MaxPool layer* dan 1 *Average-Pool layer*. ResNet-101 memiliki 7.6×10^9 Operasi *floating point*. Gambar 2.10 merupakan diagram proses dari arsitektur ResNet-101.



Gambar 2.10: Arsitektur ResNet-101 [10]

2.2.14 MobileNet-v1

MobileNets, merupakan salah satu arsitektur *convolutional neural network* (CNN) yang dapat digunakan untuk mengatasi kebutuhan akan *computing resource* berlebih. Seperti namanya, *Mobile*, para peneliti dari Google membuat arsitektur CNN yang dapat di-

gunakan untuk *mobile device* atau ponsel [25]. Perbedaan mendasar antara arsitektur MobileNet dan arsitektur CNN pada umumnya adalah penggunaan lapisan atau layer konvolusi dengan ketebalan filter yang sesuai dengan ketebalan dari *input image*. MobileNet membagi konvolusi menjadi *depthwise convolution* dan *pointwise convolution*.

[Halaman ini sengaja dikosongkan]

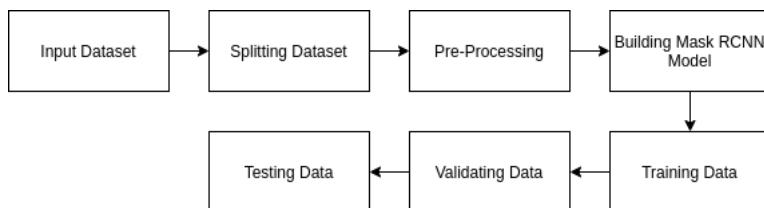
BAB III

DESAIN DAN IMPLEMENTASI

Penelitian ini dilaksanakan sesuai dengan sistem berikut dengan implementasinya. Desain sistem merupakan konsep dari pembuatan dan perancangan infrastruktur dan kemudian diwujudkan dalam bentuk blok-blok alur yang harus dikerjakan. Pada bagian implementasi merupakan pelaksanaan teknis untuk setiap blok pada desain sistem.

3.1 Deskripsi Sistem

Sistem pada tugas akhir ini merupakan implementasi dari salah satu disiplin ilmu *Deep Learning* dan pengolahan citra yang berfungsi untuk mendeteksi adanya pejalan kaki yang berada di pinggir jalan, trotoar dan jalur penyebrangan. Selain pejalan kaki, deteksi juga dilakukan pada jalur penyebrangan atau *zebra cross* dengan tujuan untuk memberi informasi bahwa disekitar area tersebut terdapat banyak aktivitas pejalan kaki yang menyebrang jalan. Blok diagram metodologi sistem yang digunakan pada penelitian ini dapat dilihat pada Gambar 3.1.



Gambar 3.1: Blok Diagram Metodologi

3.2 Pengumpulan *Dataset* Gambar

Pada tugas akhir ini, *dataset* yang digunakan didapatkan dengan beberapa cara, antara lain:

1. *Caltech Pedestrian Database*, merupakan kumpulan gambar yang diambil dari sudut pandang pengendara mobil di California Amerika Serikat dengan ukuran 640 x 480 pixel. Terdapat sekitar 250.000 gambar dengan 350.000 *bounding boxes* dan sekitar 2.300 pejalan kaki dengan kriteria unik diberi tanda. Namun, pada *dataset* ini hanya pejalan kaki saja yang diberi label, sehingga perlu dilakukan proses pelabelan ulang sesuai kelas yang diinginkan. Tidak semua gambar pada *dataset* ini diambil untuk digunakan, gambar yang mempunyai objek berupa pejalan kaki dan *zebra cross* saja yang akan digunakan. Gambar 3.2 merupakan contoh dari gambar yang terdapat pada *Caltech Pedestrian Database*.



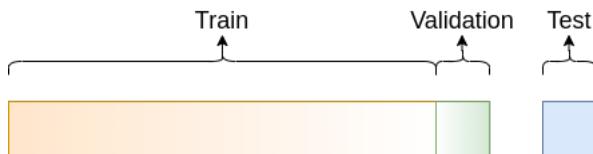
Gambar 3.2: Contoh Gambar dari Caltech Pedestrian Database

2. Tangkapan layar dari beberapa video *online Youtube*. Pada cara ini, penulis mencari video yang berada pada salah satu *website video streaming* yaitu Youtube dengan persyaratan video diambil dari sudut pandang pengendara mobil yang berkendara pada jalan raya dengan ukuran gambar 1360x768 px. Pada *frame-frame* tertentu dilakukan *screenshot* dan disimpan untuk selanjutnya dilakukan proses pemberian label pada objek-objek yang diinginkan.
3. Pengambilan gambar secara mandiri menggunakan kamera *smartphone* yang diambil dari sudut pandang pengendara motor dengan ukuran gambar yang diambil sebesar 1280x720 px.

Pengambilan gambra dilakukan di jalan-jalan Surabaya. Setelah dilakukan pengambilan gambar, proses selanjutnya adalah pemberian label pada objek-objek yang ingin dideteksi.

3.3 Pemisahan Data

Dalam *machine learning* pemisahan data ke beberapa *subset* merupakan suatu hal yang sangat penting. Hal ini dikarenakan setiap *subset* memiliki fungsi masing-masing. Gambar 3.3 merupakan rasio pembagian data ke masing-masing subset.



Gambar 3.3: Visualisasi Pembagian Data

1. *Training Sets*

Training Sets merupakan sampel data yang digunakan untuk melatih model yang sudah kita buat, dalam bidang *Neural Network* bisa disebut juga bobot dan bias. Model yang sudah kita buat mempelajari pola masukan dan keluaran dari data ini.

2. *Validation Sets*

Validation Sets merupakan sampel data yang digunakan untuk mengevaluasi model yang sudah dilatih menggunakan *training sets*. Selain itu, data ini digunakan untuk memperbarui dan menyempurnakan hyperparameter dari model ke tingkat yang lebih tinggi.

3. *Test Sets*

Test Sets merupakan sampel data yang digunakan untuk mengevaluasi model akhir setelah melalui proses *training* dan *validation*. Apabila pengujian model pada data ini sudah sesuai dengan yang diinginkan, maka proses *learning* sudah selesai.

Namun apabila pengujian tidak sesuai dengan yang diharapkan maka diperlukan pengaturan ulang mulai dari proses *training*.

3.4 Pre-Processing

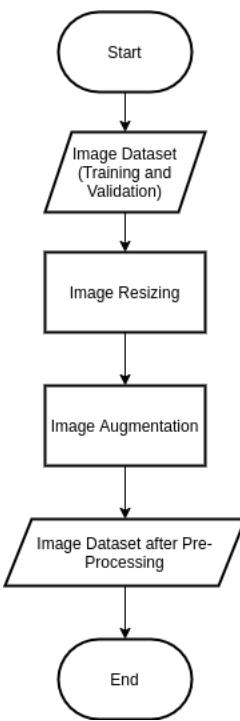
Pada tahap ini, gambar-gambar dari *dataset* akan mengalami proses penyesuaian sebelum masuk ke proses *data training*. Setiap gambar yang akan dijadikan bahan pembelajaran model harus memiliki dimensi dan kedalaman yang sama. Tujuan dari *pre processing* adalah perbaikan data gambar dengan menekan distorsi yang tidak diinginkan atau meningkatkan beberapa fitur gambar yang relevan untuk pemrosesan lebih lanjut. Gambar 3.4 merupakan tahapan dari *pre-processing* gambar *dataset* yang dilakukan.

Berikut merupakan penjelasan mengenai tahapan *pre-processing* yang dilakukan pada tugas akhir kali ini:

1. *Image resizing*

Langkah awal dari proses *pre-processing* adalah memastikan semua gambar dalam *dataset* kita memiliki ukuran yang sama. Selain itu, sama seperti sebagian besar model dari *neural network* lainnya, metode yang dilakukan penulis juga mengasumsikan gambar *input* berbentuk persegi. Jadi diperlukan pemeriksaan gambar di awal, apakah gambar sudah berbentuk persegi atau belum. Berbeda dari metode *image resizing* pada model *neural network* lainnya yang menggunakan teknik *cropping* untuk membuat aspek rasio gambar input menjadi persegi, penulis menggunakan metode yang sudah terdapat pada *Mask R-CNN*.

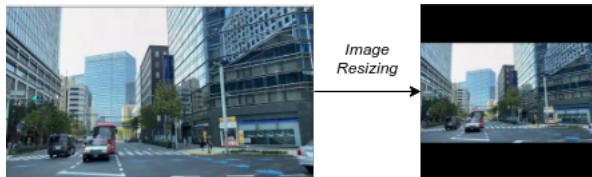
Ukuran gambar yang penulis pilih pada tugas akhir kali ini adalah 512x512 pixel. Pemilihan ukuran gambar ini dilakukan untuk mengurangi beban dan waktu saat *training data*. Apabila terdapat gambar pada *dataset* dengan ukuran baik panjang maupun lebar lebih dari 512 pixel. maka gambar akan di *down scaling* sampai ukuran 512 pixel. Sebaliknya, apabila ada gambar pada *dataset* dengan ukuran lebih kecil dari 512 pixel maka akan dilakukan *up scaling* sampai gambar berukuran 512 pixel. Aspek rasio gambar yang sudah



Gambar 3.4: Diagram Alir *Pre Processing*

melalui proses *scaling* tetap dipertahankan, namun diperlukan penambahan *zero padding* untuk membuat gambar *input* menjadi persegi seperti yang diinginkan.

Gambar 3.5 merupakan salah satu contoh *image resizing* yang dilakukan. Gambar *input* (gambar sebelah kiri) mempunyai ukuran 768x1360 dengan kedalaman 3 atau mempunyai format warna RGB. Setelah mengalami *image resizing* (gambar sebelah kanan) ukuran gambar menjadi 290x512. Namun untuk membuat gambar memiliki aspek rasio 1:1 (berbentuk persegi) maka diperlukan penambahan *zero padding* pada bagian atas gambar sebesar 111 pixel dan pada bagian bawah gambar sebesar 111 pixel. Dengan penambahan *padding* seperti



Gambar 3.5: Contoh *Image Resizing*

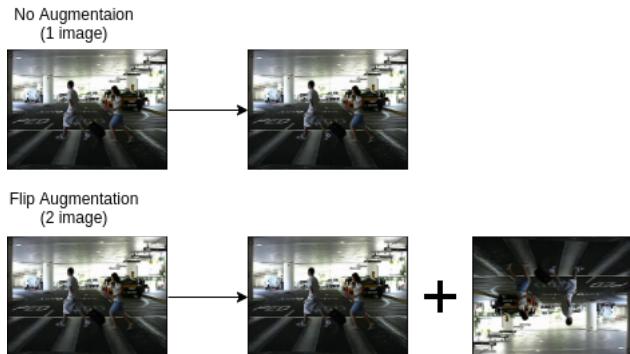
itu membuat gambar *input* berbentuk persegi namun tidak mengurangi informasi gambar.

2. *Image Augmentation*

Langkah selanjutnya pada *pre-processing* adalah *image augmentation*. Proses augmentasi yang dilakukan pada tugas akhir ini adalah rotasi dan transformasi. Tujuan dari penggunaan *image augmentation* adalah untuk mengekspos *neural network* ke berbagai variasi, agar dapat mengenali fitur yang akan dilakukan pada proses *training*. Hal tersebut akan sangat membantu *neural network* untuk mengenali variasi yang tidak terdapat pada dataset. Seperti yang terdapat pada Gambar 3.6, tanpa ada augmentasi maka *neural network* hanya mengenali satu kondisi saja. Jika memakai augmentasi gambar seperti transformasi, maka setidaknya *neural network* akan dapat mengenali 2 kondisi. Semakin banyak augmentasi yang digunakan semakin banyak pula kondisi yang bisa dikenali oleh *neural network*. Namun semakin banyak kondisi yang dikenali, semakin lama dan berat proses *training data* yang dilakukan.

3.5 Membangun Model Mask R-CNN

Mask R-CNN merupakan salah satu metode *deep learning* yang dikembangkan dari Faster R-CNN dengan menambahkan satu cabang di tahap akhir untuk menghasilkan *mask* dari objek yang didekksi. Pengembangan tersebut dilakukan untuk memecahkan masalah *instance segmentation* yang terjadi dalam *machine learning* dan pengolahan citra. Dengan kata lain, *mask r-cnn*, dapat memisahkan

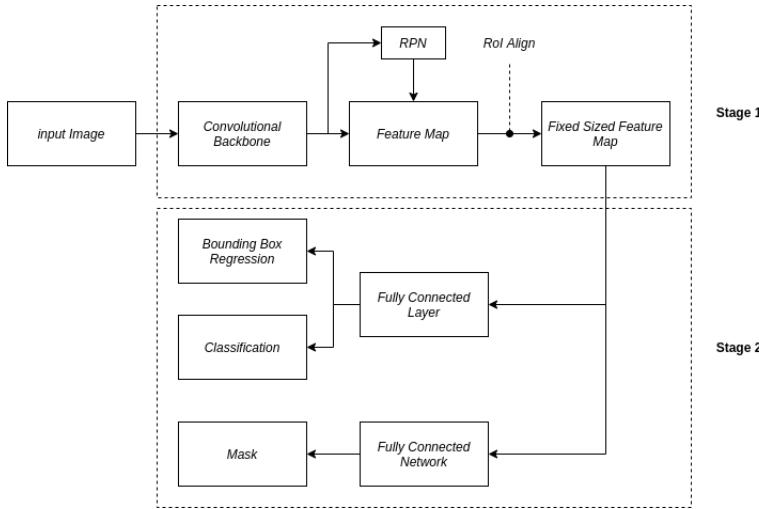


Gambar 3.6: Contoh *Image Augmentation*

an objek yang berbeda walaupun dalam satu kelas yang sama pada gambar atau video. Selain memberikan hasil berupa *bounding box* dan klasifikasi objek seperti kebanyakan algoritma *object detection* lainnya, *mask r-cnn* juga memberikan *mask* dimana hal ini sangat bermanfaat pada segmentasi objek.

Seperti yang ditampilkan pada Gambar 3.7, pada proses *training* menggunakan *mask r-cnn* dibagi menjadi 2 tahapan. Tahap pertama adalah tahap untuk menghasilkan proposal tentang area di mana mungkin ada objek berdasarkan gambar *input*. Lalu tahap kedua adalah tahap untuk memprediksi kelas objek, memperbaiki *bounding box* dan menghasilkan *mask* di tingkat piksel objek berdasarkan proposal tahap pertama. Kedua tahap terhubung ke struktur *backbone*.

Backbone adalah *deep neural network* yang memiliki struktur seperti FPN (*Feature Pyramid Network*). *Backbone* terdiri dari *bottom-up pathway*, *up-bottom pathway* dan *lateral connection*. *Bottom-up pathway* dapat berupa berbagai jenis *Convolutional Network*, biasanya berupa *ResNet* atau *VGG*, yang mengekstrak fitur dari *raw images*. *Up-bottom pathway* menghasilkan *Feature Map Pyramid* yang ukurannya mirip dengan *bottom-up pathway*. *Lateral connection* adalah operasi konvolusi dan penjumlahan antara dua *pathway* dengan tingkat yang sesuai. FPN mempunyai kinerja yang le-



Gambar 3.7: Blok Diagram Alur Mask R-CNN

bih baik dari ConvNet tunggal lainnya terutama karena FPN dapat mempertahankan fitur semantik yang sangat baik pada berbagai skala resolusi.

3.6 *Training Data*

Proses *training data* dilakukan setelah pembuatan model telah selesai dan *dataset* sudah melalui proses *pre-processing*. Pada saat pertama kali menjalankan proses *training*, bobot awal diam-bil dari *pre-trained weight* yang sudah tersedia pada *mask r-cnn*. Hal ini bisa dilakukan dengan menerapkan metode *transfer learning*. *Transfer learning* sendiri adalah teknik yang sangat efisien untuk melakukan proses *training* atau *retrain* pada *neural network*. Penggunaan *transfer learning* mempunyai keuntungan diantara lain proses *training* pada data baru memakan waktu yang lebih cepat daripada memulai dari awal serta masalah dapat dipecahkan dengan menggunakan training data yang lebih sedikit daripada membangun model dari awal.

Ada beberapa hal yang perlu diperhatikan dalam melakukan pengaturan saat akan menjalankan proses *training* antara lain :

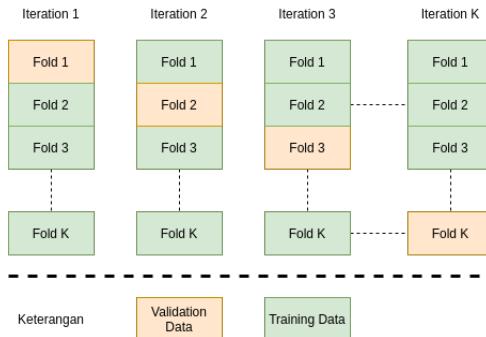
1. *Iteration* adalah banyaknya proses yang dilakukan untuk melakukan *forward* dan *backward pass*. *forward pass* adalah proses dimana *output value* dari *neural network* didapatkan setelah *input value* dari *input-neuron* telah selesai diproses. Sedangkan *backward pass* adalah proses mengkalkulasikan bobot dari *neural network* mulai dari *output neuron* ke *input neuron* untuk mendapatkan *loss* dari setiap *neuron*. Pada *mask r-cnn iteration* bisa disebut juga dengan *step-per-epoch* sesuai dengan yang tercantum pada *file* pengaturan *mask r-cnn*.
2. *Epoch*, ketika seluruh dataset sudah melalui proses *training* pada *neural network* sampai dikembalikan ke awal untuk sekali putaran. Sebagai contoh apabila kita menggunakan *iteration* sebanyak 10 kali, maka satu *epoch* sebanyak 10 *iteration* dan kelipatannya. Pada tugas akhir kali ini penulis menggunakan *epoch* sebanyak 100.

3.7 *Validating Data*

Setelah proses *training* dilakukan, perlu dilakukan apakah model yang dibuat sudah memiliki tingkat akurasi sesuai yang kita inginkan dengan menggunakan teknik validasi. Pada proses inilah, *dataset* yang telah dipisahkan pada proses sebelumnya akan berperan. Evaluasi memungkinkan pengujian model terhadap data yang belum pernah dilihat dan digunakan untuk pelatihan dan dimaksudkan untuk mewakili bagaimana model dapat menyelesaikan permasalahan tersebut. Tahapan pada validasi akan membantu untuk menemukan parameter terbaik untuk model prediktif dan mencegah dari *overfitting*.

Pada tugas akhir kali ini, digunakan salah satu jenis teknik validasi menggunakan *Cross Validation*. Pada *Cross Validation* data akan dibagi menjadi K lipatan, dimana setiap lipatan akan diamati satu data sebagai *validation data* dan sisanya akan digunakan untuk *training data*. Pemilihan *validation data* dilakukan secara menyilang, dengan ketentuan apabila *training* terjadi pada itera-

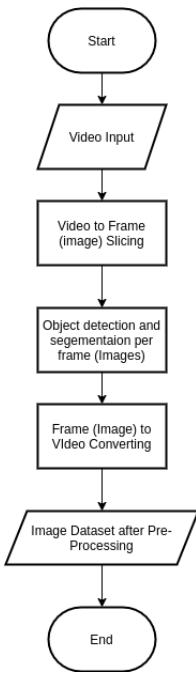
si ke k maka data yang dipilih untuk validasi adalah data k juga. Gambar 3.8 merupakan visualisasi dari K Fold Cross Validation.



Gambar 3.8: Visualisasi K Fold Cross Validation

3.8 Testing Data

Testing data merupakan tahap akhir dalam algoritma *machine learning* secara umum. Pada tahap ini model akan diuji untuk didapatkan keakuratan untuk mendekripsi objek. Data uji yang digunakan pada tugas akhir kali ini adalah data yang berbentuk video. Jadi diperlukan *pre-processing* yang berbeda dengan *training data* dan *validation data*. Data video akan dipecah atau dipotong-potong menjadi format gambar dengan ketentuan 30 *frame per second*. Ketika *data test* sudah dikonversi menjadi bentuk gambar, maka proses deteksi bisa dilakukan. Gambar hasil deteksi berupa gambar asli yang sudah ditambah dengan *bounding box*, *classification*, *mask*. Lalu gambar-gambar tersebut disatukan lagi menjadi format video dengan ketentuan sama seperti saat pemotongan menjadi format gambar, yaitu 30 *frame per second*. Gambar 3.9 merupakan diagram alir dari proses *testing* yang dilakukan pada *input file* berupa video.



Gambar 3.9: Diagram Alir *Testing Data*

[Halaman ini sengaja dikosongkan]

BAB IV

PENGUJIAN DAN ANALISIS

Pada bab ini dipaparkan hasil pengujian serta analisa dari desain sistem dan implementasi. Pengujian dilakukan guna mengetahui tingkat kesalahan dan menarik kesimpulan dari sistem yang telah dibuat.

Pada proses pengujian digunakan salah satu layanan *Google* yaitu *Google Colaboratory* dengan spesifikasi *hardware* seperti pada Tabel 4.1. Sedangkan untuk spesifikasi *hardware* komputer penulis dapat dilihat pada Tabel 4.2.

Tabel 4.1: Spesifikasi *hardware* *Google Colaboratory*

Procesor	Intel Xeon Processor @ 2.3 GHz
Graphic Card	Tesla K80 12 GB GDDR5 VRAM
RAM	16 GB

Tabel 4.2: Spesifikasi *hardware* Komputer yang Digunakan

Procesor	Intel(R) Core(TM) i5-10400F CPU @ 2.90GHz
Graphic Card	Nvidia GeForce GTX 1650 4 GB GDDR6
RAM	8 GB

Pengujian dilakukan dengan membagi model ke beberapa jenis *backbone* yang digunakan, antara lain Resnet-50, Resnet-101, dan Mobilenet-V1.

4.1 Pengujian Jenis *Backbone*

Pengujian pada jenis *backbone* bertujuan untuk mengetahui performa dan akurasi dari setiap model yang dihasilkan dengan

backbone yang berbeda.

4.1.1 Resnet-50

Tabel 4.3 merupakan parameter-parameter yang digunakan untuk membuat model Mask R-CNN dengan menggunakan *backbone* Resnet-50.

Tabel 4.3: Konfigurasi Model menggunakan Resnet-50

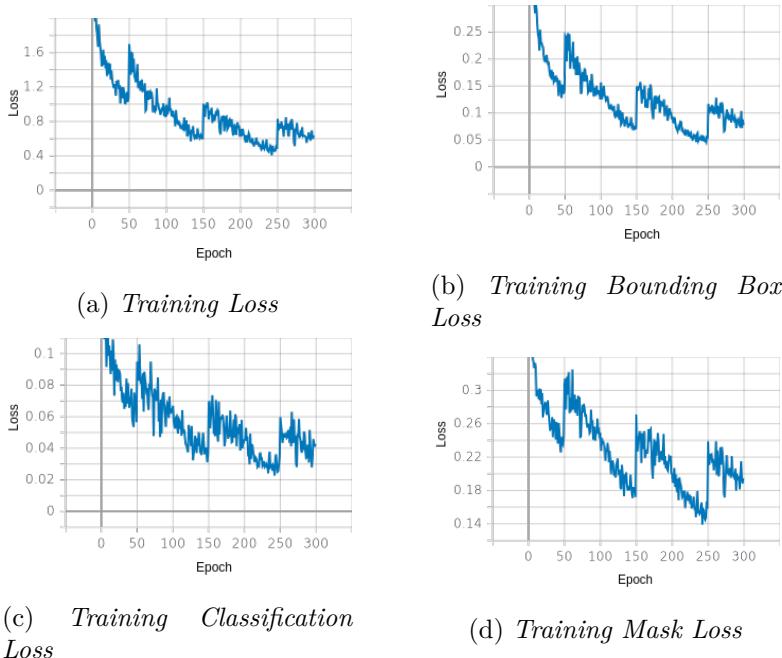
Pengaturan Model Resnet-50	
BACKBONE	resnet50
BACKBONE_STRIDES	[4, 8, 16, 32, 64]
BATCH_SIZE	1
BBOX_STD_DEV	[0.1 0.1 0.2 0.2]
COMPUTE_BACKBONE_SHAPE	None
DETECTION_MAX_INSTANCES	50
DETECTION_MIN_CONFIDENCE	0.9
DETECTION_NMS_THRESHOLD	0.2
FPN_CLASSIF_FC_LAYERS_SIZE	1024
GPU_COUNT	1
GRADIENT_CLIP_NORM	5.0
IMAGES_PER_GPU	1
IMAGE_CHANNEL_COUNT	3
IMAGE_MAX_DIM	512
IMAGE_META_SIZE	16
IMAGE_MIN_DIM	400
IMAGE_MIN_SCALE	0
IMAGE_RESIZE_MODE	square
IMAGE_SHAPE	[512 512 3]
LEARNING_MOMENTUM	0.9
LEARNING_RATE	0.001
LOSS_WEIGHTS	{'rpn_class_loss': 1.0, 'rpn_bbox_loss': 1.0, 'mrcnn_class_loss': 1.0, 'mrcnn_bbox_loss': 1.0, 'mrcnn_mask_loss': 1.0}
<i>Dilanjutkan pada halaman berikutnya</i>	

Tabel 4.3 – Lanjutan dari halaman sebelumnya

MASK_POOL_SIZE	14
MASK_SHAPE	[28, 28]
MAX_GT_INSTANCES	50
MEAN_PIXEL	[123.7 116.8 103.9]
MINI_MASK_SHAPE	(56, 56)
NAME	object
NUM_CLASSES	4
POOL_SIZE	7
POST_NMS_ROIS_INFERENCE	1000
POST_NMS_ROIS_TRAINING	2000
PRE_NMS_LIMIT	6000
ROI_POSITIVE_RATIO	0.33
RPN_ANCHOR RATIOS	[0.5, 1, 2]
RPN_ANCHOR_SCALES	(32, 64, 128, 256, 512)
RPN_ANCHOR_STRIDE	1
RPN_BBOX_STD_DEV	[0.1 0.1 0.2 0.2]
RPN_NMS_THRESHOLD	0.7
RPN_TRAIN_ANCHORS_PER_IMAGE	256
STEPS_PER_EPOCH	100
TOP_DOWN_PYRAMID_SIZE	256
TRAIN_BN	False
TRAIN_ROIS_PER_IMAGE	200
USE_MINI_MASK	True
USE_RPN_ROIS	True
VALIDATION_STEPS	30
WEIGHT_DECAY	0.0001

Setelah dilakukan serangkaian proses training yang memakan waktu sekitar 3 jam 40 menit 24 detik didapatkan *output* berupa *model file* dengan format *h5* yang mempunyai ukuran 170.9 MB. *Training loss* terendah yang berhasil dicapai dengan menggunakan *backbone* Resnet-50 (pada *epoch* ke 242) adalah 0.4061 dengan rincian *training bounding box loss* sebesar 0.04083, *training classification loss* sebesar 0.02268 serta *training mask loss* sebesar 0.139 (dimana $L = L_{bbox} + L_{cls} + L_{mask}$). Gambar 4.1 merupakan grafik

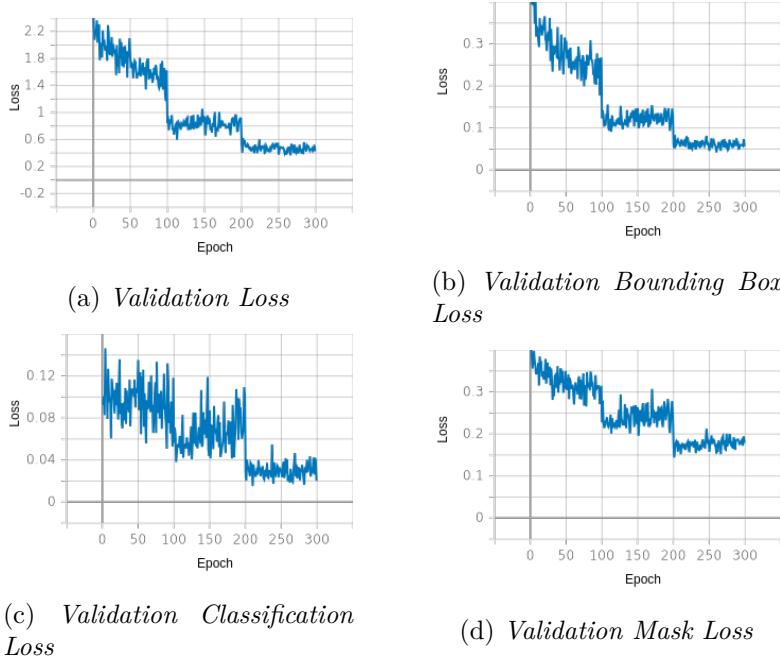
yang menunjukkan perubahan *training loss*, *training bounding box loss*, *training classification loss*, serta *training mask loss* dari epoch 1 sampai 300.



Gambar 4.1: Grafik Perubahan *Training Loss* pada Resnet-50

Sedangkan pada saat proses *validation* sendiri *Loss* terendah yang berhasil dicapai pada epoch ke 266 dengan nilai sebesar 0.3653 dengan rincian *validation bounding box loss* sebesar 0.4556, *validation classification loss* sebesar 0.01912 serta *validation mask loss* sebesar 0.1519. Namun untuk *validation bounding box loss* terendah berada pada epoch ke 260 dengan nilai sebesar 0.4203 sedangkan *validation classification loss* terendah pada epoch ke 210 dengan nilai 0.01525 serta *validation mask loss* terendah pada epoch ke 201 dengan nilai 0.1439. Gambar 4.2 merupakan grafik yang menunjukkan perubahan *validation loss*, *validation bounding box loss*, *validation*

classification loss, serta *validation mask loss* dari epoch 1 sampai 300.

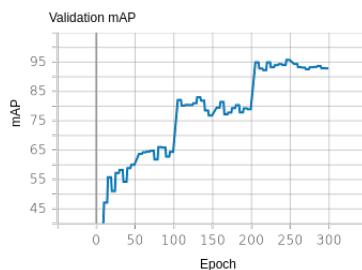


Gambar 4.2: Grafik Perubahan *Validation Loss* pada *Resnet-50*

Selain menggunakan *Loss Function* untuk mengukur peforma hasil *training* yang sudah dilakukan, digunakan juga *mean Average Precision (mAP)*. *Precision* sendiri merupakan fungsi untuk menggambarkan tingkat keakuratan antara data yang diminta dengan hasil prediksi yang diberikan oleh model. Maka, *precision* merupakan rasio prediksi benar positif (TP) dibandingkan dengan keseluruhan hasil yang diprediksi positif (TP dan FP). Rumus untuk mencari *Precision* adalah sebagai berikut :

$$Precision = \frac{TP}{TP + FP} \quad (4.1)$$

Perhitungan *mAP* pada penelitian ini dilakukan setiap 5 *epoch* sekali, karena jika dilakukan setiap *epoch* akan memerlukan *training time* yang lebih lama serta *resource hardware* yang diperlukan lebih besar. Nilai *mAP* tertinggi didapatkan pada *epoch* ke 220 sebesar 94.92. Gambar 4.3 merupakan grafik yang menunjukkan perubahan *validation mean Average Precision* dari *epoch* 1 sampai 300.



Gambar 4.3: Grafik Perubahan *Validation mAP* pada Resnet-50

4.1.2 Resnet-101

Tabel 4.4 merupakan parameter-parameter yang digunakan untuk membuat model Mask R-CNN dengan menggunakan *backbone* Resnet-101.

Tabel 4.4: Konfigurasi Model menggunakan Resnet-101

Pengaturan Model Resnet-101	
BACKBONE	resnet101
BACKBONE_STRIDES	[4, 8, 16, 32, 64]
BATCH_SIZE	1
BBOX_STD_DEV	[0.1 0.1 0.2 0.2]
COMPUTE_BACKBONE_SHAPE	None
DETECTION_MAX_INSTANCES	50
DETECTION_MIN_CONFIDENCE	0.9
DETECTION_NMS_THRESHOLD	0.2
FPN_CLASSIF_FC_LAYERS_SIZE	1024
<i>Dilanjutkan pada halaman berikutnya</i>	

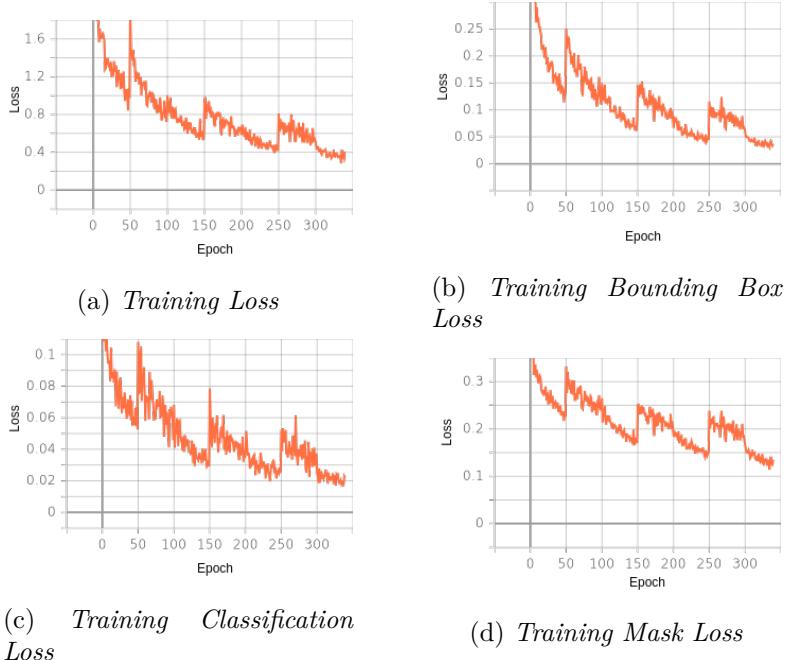
Tabel 4.4 – Lanjutan dari halaman sebelumnya

GPU_COUNT	1
GRADIENT_CLIP_NORM	5.0
IMAGES_PER_GPU	1
IMAGE_CHANNEL_COUNT	3
IMAGE_MAX_DIM	512
IMAGE_META_SIZE	16
IMAGE_MIN_DIM	400
IMAGE_MIN_SCALE	0
IMAGE_RESIZE_MODE	square
IMAGE_SHAPE	[512 512 3]
LEARNING_MOMENTUM	0.9
LEARNING_RATE	0.001
LOSS_WEIGHTS	{'rpn.class_loss': 1.0, 'rpn.bbox_loss': 1.0, 'mrcnn.class_loss': 1.0, 'mrcnn.bbox_loss': 1.0, 'mrcnn.mask_loss': 1.0}
MASK_POOL_SIZE	14
MASK_SHAPE	[28, 28]
MAX_GT_INSTANCES	50
MEAN_PIXEL	[123.7 116.8 103.9]
MINI_MASK_SHAPE	(56, 56)
NAME	object
NUM_CLASSES	4
POOL_SIZE	7
POST_NMS_ROIS_INFERENCE	1000
POST_NMS_ROIS_TRAINING	2000
PRE_NMS_LIMIT	6000
ROI_POSITIVE_RATIO	0.33
RPN_ANCHOR RATIOS	[0.5, 1, 2]
RPN_ANCHOR_SCALES	(32, 64, 128, 256, 512)
RPN_ANCHOR_STRIDE	1
RPN_BBOX_STD_DEV	[0.1 0.1 0.2 0.2]
RPN_NMS_THRESHOLD	0.7
<i>Dilanjutkan pada halaman berikutnya</i>	

Tabel 4.4 – Lanjutan dari halaman sebelumnya

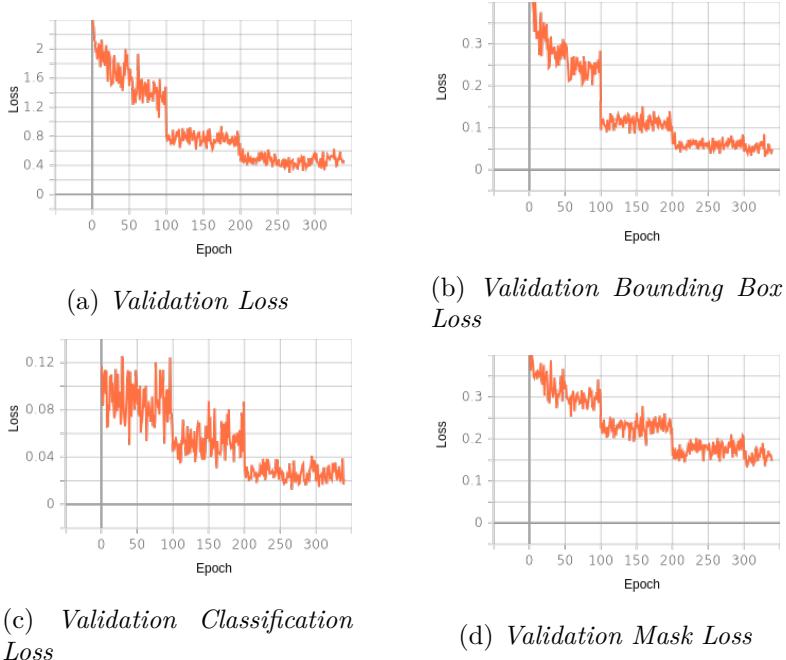
RPN_TRAIN_ANCHORS_PER_IMAGE	256
STEPS_PER_EPOCH	100
TOP_DOWN_PYRAMID_SIZE	256
TRAIN_BN	False
TRAIN_ROIS_PER_IMAGE	200
USE_MINI_MASK	True
USE_RPN_ROIS	True
VALIDATION_STEPS	30
WEIGHT_DECAY	0.0001

Setelah dilakukan serangkaian proses training yang memakan waktu sekitar 4 jam 9 menit 16 detik didapatkan *output* berupa *model file* dengan format *h5* yang mempunyai ukuran 244 MB. *Training loss* terendah yang berhasil dicapai dengan menggunakan *backbone* Resnet-101 (pada *epoch* ke 244) adalah 0.3933 dengan rincian *training bounding box loss* sebesar 0.04164, *training classification loss* sebesar 0.0247 serta *training mask loss* sebesar 0.1403. Namun untuk *training bounding box loss* terendah terdapat pada *epoch* ke 246 dengan nilai sebesar 0.04083, *training classification loss* terendah pada *epoch* ke 234 dengan nilai 0.01957. Gambar 4.4 merupakan grafik yang menunjukkan perubahan *training loss*, *training bounding box loss*, *training classification loss*, serta *training mask loss* dari *epoch* 1 sampai 300.



Gambar 4.4: Grafik Perubahan *Training Loss* pada *Resnet-101*

Sedangkan pada saat proses *validation* sendiri *Loss* terendah yang berhasil dicapai pada *epoch* ke 266 dengan nilai sebesar 0.299 dengan rincian *validation bounding box loss* sebesar 0.03867, *validation classification loss* sebesar 0.01246 serta *validation mask loss* sebesar 0.1461. Gambar 4.5 merupakan grafik yang menunjukkan perubahan *validation loss*, *validation bounding box loss*, *validation classification loss*, serta *validation mask loss* dari *epoch* 1 sampai 300.

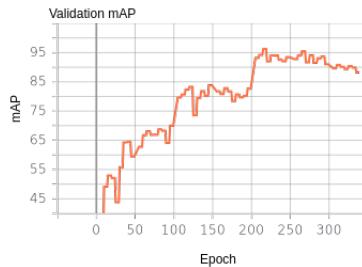


Gambar 4.5: Grafik Perubahan *Validation Loss* pada *Resnet-101*

Nilai *mAP* tertinggi didapatkan pada *epoch* ke 215 sebesar 96.21. Gambar 4.6 merupakan grafik yang menunjukkan perubahan *validation mean Average Precision* dari *epoch* 1 sampai 300.

4.1.3 MobileNet-V1

Tabel 4.5 merupakan parameter-parameter yang digunakan untuk membuat model Mask R-CNN dengan menggunakan *backbone* Mobilenet-V1.



Gambar 4.6: Grafik Perubahan *Validation mAP* pada *Resnet-101*

Tabel 4.5: Konfigurasi Model menggunakan Mobilenet-V1

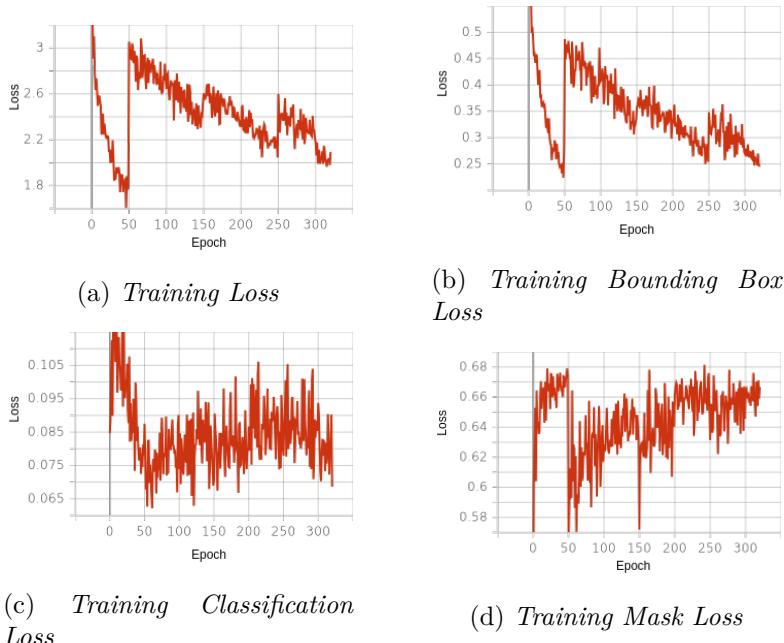
Pengaturan Model Mobilenet-V1	
BACKBONE	mobilenetv1
BACKBONE_STRIDES	[4, 8, 16, 32, 64]
BATCH_SIZE	1
BBOX_STD_DEV	[0.1 0.1 0.2 0.2]
COMPUTE_BACKBONE_SHAPE	None
DETECTION_MAX_INSTANCES	50
DETECTION_MIN_CONFIDENCE	0.9
DETECTION_NMS_THRESHOLD	0.2
FPN_CLASSIF_FC_LAYERS_SIZE	1024
GPU_COUNT	1
GRADIENT_CLIP_NORM	5.0
IMAGES_PER_GPU	1
IMAGE_CHANNEL_COUNT	3
IMAGE_MAX_DIM	512
IMAGE_META_SIZE	16
IMAGE_MIN_DIM	400
IMAGE_MIN_SCALE	0
IMAGE_RESIZE_MODE	square
IMAGE_SHAPE	[512 512 3]
LEARNING_MOMENTUM	0.9
LEARNING_RATE	0.001
<i>Dilanjutkan pada halaman berikutnya</i>	

Tabel 4.5 – Lanjutan dari halaman sebelumnya

LOSS_WEIGHTS	{'rpn_class_loss': 1.0, 'rpn_bbox_loss': 1.0, 'mrcnn_class_loss': 1.0, 'mrcnn_bbox_loss': 1.0, 'mrcnn_mask_loss': 1.0}
MASK_POOL_SIZE	14
MASK_SHAPE	[28, 28]
MAX_GT_INSTANCES	50
MEAN_PIXEL	[123.7 116.8 103.9]
MINI_MASK_SHAPE	(56, 56)
NAME	object
NUM_CLASSES	4
POOL_SIZE	7
POST_NMS_ROIS_INFERENCE	1000
POST_NMS_ROIS_TRAINING	2000
PRE_NMS_LIMIT	6000
ROIPOSITIVE_RATIO	0.33
RPN_ANCHOR RATIOS	[0.5, 1, 2]
RPN_ANCHOR_SCALES	(32, 64, 128, 256, 512)
RPN_ANCHOR_STRIDE	1
RPN_BBOX_STD_DEV	[0.1 0.1 0.2 0.2]
RPN_NMS_THRESHOLD	0.7
RPN_TRAIN_ANCHORS_PER_IMAGE	256
STEPS_PER_EPOCH	100
TOP_DOWN_PYRAMID_SIZE	256
TRAIN_BN	False
TRAIN_ROIS_PER_IMAGE	200
USE_MINI_MASK	True
USE_RPN_ROIS	True
VALIDATION_STEPS	30
WEIGHT_DECAY	0.0001

Setelah dilakukan serangkaian proses training yang memakan waktu sekitar 3 jam 18 menit 44 detik didapatkan *output* berupa *model file* dengan format *h5* yang mempunyai ukuran 83.3 MB.

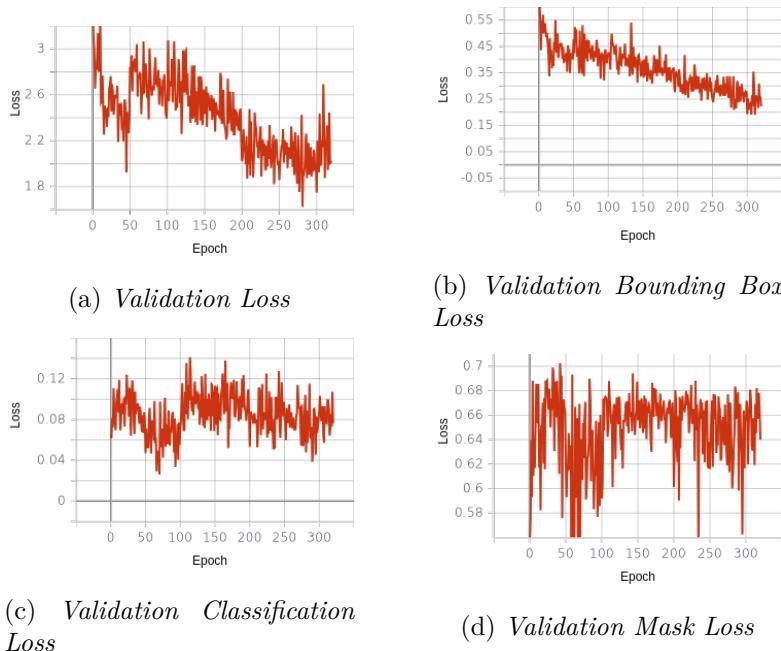
Training loss terendah yang berhasil dicapai dengan menggunakan *backbone* MobileNet-V1 (pada *epoch* ke 46) adalah 1.604 dengan rincian *training bounding box loss* sebesar 0.2322, *training classification loss* sebesar 0.07858 serta *training mask loss* sebesar 0.6729. Namun untuk *training bounding box loss* terendah terdapat pada *epoch* ke 48 dengan nilai sebesar 0.2239, *training classification loss* terendah pada *epoch* ke 61 dengan nilai 0.06219 serta *training mask loss* terendah pada *epoch* ke 61 sebesar 0.5701. Gambar 4.7 merupakan grafik yang menunjukkan perubahan *training loss*, *training bounding box loss*, *training classification loss*, serta *training mask loss* dari *epoch* 1 sampai 300.



Gambar 4.7: Grafik Perubahan *Training Loss* pada MobileNet-V1

Sedangkan pada saat proses *validation* sendiri *Loss* terendah yang berhasil dicapai pada *epoch* ke 281 dengan nilai sebesar 1.624

dengan rincian *validation bounding box loss* sebesar 0.2088, *validation classification loss* sebesar 0.04799 serta *validation mask loss* sebesar 0.6001. Namun untuk *validation bounding box loss* terendah berada pada *epoch* ke 300 dengan nilai sebesar 0.1927 sedangkan *validation classification loss* terendah pada *epoch* ke 70 dengan nilai 0.02607 serta *validation mask loss* terendah pada *epoch* ke 295 dengan nilai 0.5625. Gambar 4.8 merupakan grafik yang menunjukkan perubahan *validation loss*, *validation bounding box loss*, *validation classification loss*, serta *validation mask loss* dari *epoch* 1 sampai 300.



Gambar 4.8: Grafik Perubahan *Validation Loss* pada *MobileNet-V1*

Nilai *mAP* tertinggi didapatkan pada *epoch* ke 210 sebesar 26.04. Gambar 4.9 merupakan grafik yang menunjukkan perubahan *validation mean Average Precision* dari *epoch* 1 sampai 300.

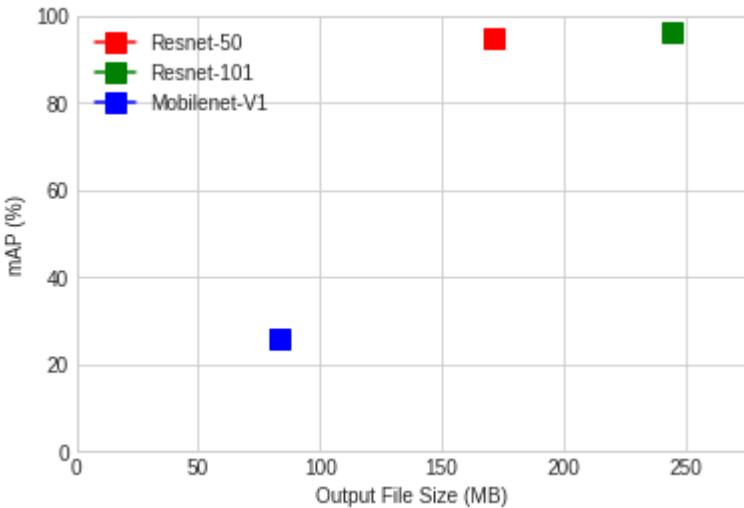


Gambar 4.9: Grafik Perubahan *Validation mAP* pada *MobileNet-V1*

Tabel 4.6 adalah tabel perbandingan dari *output file size* dan *mAP* dari total 3 model dengan *backbone* berbeda yang diuji. Tahap ini masih merupakan hasil tahap *training*. Sedangkan Gambar 4.10 merupakan grafik perbandingan *mAP* dengan *file size*. Proses selanjutnya setelah mendapatkan hasil yang ditunjukkan pada setiap model adalah melakukan proses prediksi atau melakukan *testing data*.

Tabel 4.6: Tabel Perbandingan Model dengan *backbone* yang berbeda

No.	<i>Backbone</i>	<i>Output File Size</i>	<i>mAP</i>
1	Resnet-50	170.9 MB	94.92%
2	Resnet-101	244 MB	96.21%
3	Mobilenet-V1	83.3 MB	26.04%



Gambar 4.10: Grafik Perbandingan Model dengan *backbone* yang berbeda

4.1.4 Perbandingan Hasil Prediksi

Setelah mendapatkan hasil *training* pada Tabel 4.6 terdapat dua parameter yang digunakan yaitu *output file size* dan *mean Average Precision*. Parameter tersebut digunakan untuk melakukan justifikasi dalam pemilihan model yang terbaik. Pada parameter *output file size*, model yang diinginkan adalah model dengan dengan ukuran sekecil mungkin, agar tidak menghabiskan kapasitas penyimpanan terlalu banyak. Sedangkan untuk *mean Average Precision* yang diinginkan adalah model dengan *meand Average Precision* yang tinggi sehingga kemampuan model tersebut untuk melakukan proses prediksi akan semakin tepat dengan kondisi yang sesuai pada dunia nyata.

Setelah membandingkan parameter *output file size* dan *mean Average Precision*, pada tahap selanjutnya akan dilakukan proses membandingkan nilai atau hasil dari prediksi pada setiap model yang telah dibuat. Semakin tinggi *mean Average Precision* dan ke-

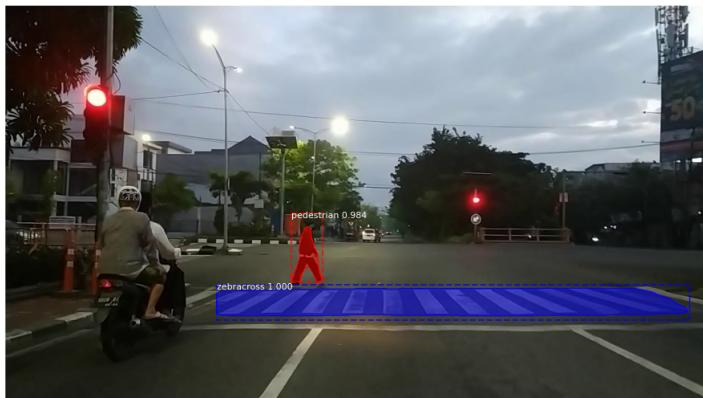
benaran dalam prediksi, maka model tersebut sangat bagus dan seuai dengan permasalahan yang ingin diselesaikan oleh model yang telah dibuat. Tujuan dari tahapan membandingkan ini adalah untuk memilih model yang terbaik dan apakah model yang telah dibuat dapat benar-benar dapat menyelesaikan permasalahan *object detection and segmentation*. Sehingga, dari hasil percobaan ini dapat menentukan model terbaik yang dapat menyelesaikan permasalahan dan juga telah diuji dengan menggunakan *testing set*.

Hasil Pengujian dengan *Backbone Resnet-50*

Pada hasil pengujian dengan menggunakan satu gambar jalan raya pada saat pagi hari yang ditunjukkan pada gambar 4.11, didapatkan hasil skor yaitu :

1. Pejalan kaki (*score* : 0,984)
2. Zebracross (*score* : 0,983)

Dari hasil tersebut, *Backbone Resnet-50* dapat memprediksi gambar dengan benar, serta *evaluation time* dari *backbone* ini sebesar 0,76 detik



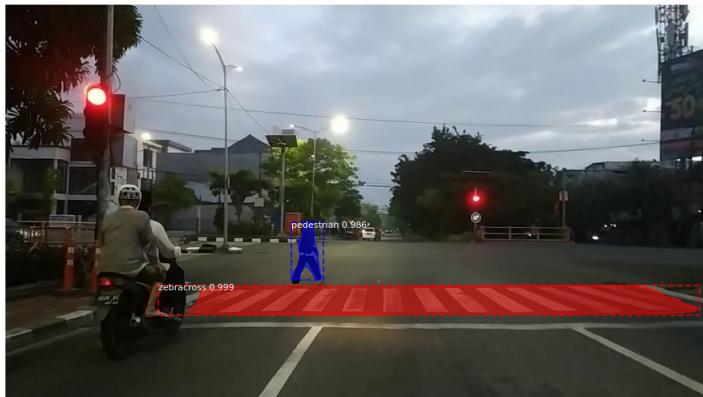
Gambar 4.11: Hasil Uji dari *Backbone Resnet-50*

Hasil Pengujian dengan *Backbone Resnet-101*

Pada hasil pengujian dengan menggunakan satu gambar jalan raya pada saat pagi hari yang ditunjukkan pada gambar 4.12, didapatkan hasil skor yaitu :

1. Pejalan kaki (*score* : 0,986)
2. Zebracross (*score* : 0,995)

Dari hasil tersebut, *Backbone Resnet-101* dapat memprediksi gambar dengan benar, serta *evaluation time* dari *backbone* ini sebesar 0.961 detik



Gambar 4.12: Hasil Uji dari *Backbone Resnet-101*

Hasil Pengujian dengan *Backbone Mobilenet-V1*

Pada hasil pengujian dengan menggunakan satu gambar jalan raya pada saat pagi hari yang ditunjukkan pada gambar 4.13, didapatkan hasil skor yaitu :

1. Pejalan kaki (tidak terdeteksi)
2. Zebracross (terdeteksi 2 dengan *score* : 0,942 dan 0.888)

Dari hasil tersebut, *Backbone Mobilenet-V1* tidak dapat memprediksi gambar dengan benar, serta *evaluation time* dari *backbone* ini sebesar 0.664 detik



Gambar 4.13: Hasil Uji dari *Backbone Mobilenet-V1*

Setelah melakukan *testing* dengan menggunakan gambar pada *testing set* pada keseluruhan model, maka tabel 4.7 menunjukkan ringkasan dari perbandingan hasil *testing* yang telah dilakukan. Parameter yang akan dibandingkan yaitu *evaluation time* dan besar *mean Average Precision*.

Tabel 4.7: tabel Perbandingan Hasil *Testing*

No.	<i>Backbone</i>	<i>Evaluation Time</i>	<i>mAP</i>
1	ResNet-50	0,763 s	100%
2	ResNet-101	0.961 s	100%
3	MobileNet-v1	0.664 s	25%

4.2 Pengujian Perbedaan Waktu

Pengujian pada perbedaan waktu bertujuan untuk mengetahui performa dan akurasi dari setiap model yang dihasilkan dengan waktu yang berbeda-beda (pagi, siang dan malam). *File input* untuk pengujian ini terdapat 2 macam, yaitu gambar dengan format *.jpg* dan video dengan format *.mp4*.

4.2.1 Pengujian di Pagi Hari

Pada pengujian ini *testing data* diambil dengan menggunakan kamera *smartphone* yang menghasilkan video beresolusi 1280×720 px berdurasi 63.43 s. Video tersebut mempunyai banyak *frame* per detik sebanyak 30 fps serta total *frame* yang dihasilkan sejumlah 1903 *frame*. Gambar 4.14 merupakan perbandingan hasil deteksi dan segementasi dari ketiga *backbone* yang digunakan pada waktu pagi hari.



(a) ResNet-50 (b) ResNet-101 (c) MobileNet-v1

Gambar 4.14: Perbandingan Hasil pada Pagi Hari

Untuk megudi performa model yang dihasilkan pada masing-masing *backbone*, digunakan beberapa parameter nilai seperti *precision*, *recall* dan *F1-score* pada masing-masing kelas objek yang didekripsi. Perbandingan hasil evaluasi yang didapatkan masing-masing *backbone* dapat dilihat pada Tabel 4.8.

Pada Resnet-50 *Average Precision* untuk semua kelas yang berhasil dideteksi adalah sebesar 100% dan *Average Recall* 100% dengan *testing time* selama 0,763 s. Sedangkan pada Resnet-101 memberikan hasil yang sama dengan Resnet-50 dengan *AP* 100% dan *AR* 100% dengan *testing time* selama 0,961 s. Namun pada Mobilenet-v1 objek yang berhasil dideteksi hanya *zebracross* (ditambah dengan *background*), sehingga pada objek pejalan kaki mempunyai *Average Precision* bernilai 50% serta *Average Recall* dan *F1-score* bernilai tidak terdefinisi atau *Nan*. Hal ini dikarenakan nilai *True Positif* pada pejalan kaki bernilai 0 (angka 0 dibagi nilai berapun memiliki hasil tidak terdefinisi) dengan *testing time* selama 0.664 s.

Selain menggunakan *input data* berupa gambar, pada pengujian ini juga menggunakan *input data* berupa *file video* dengan format

Tabel 4.8: Perbandingan Hasil Evaluasi pada Pagi Hari

(a) ResNet-50

Kelas	Precision (%)	Recall (%)	F1_Score (%)
Background	100	100	100
Pedestrian	100	100	100
Zebracross	100	100	100

(b) ResNet-101

Kelas	Precision (%)	Recall (%)	F1_Score (%)
Background	100	100	100
Pedestrian	100	100	100
Zebracross	100	100	100

(c) MobileNet-v1

Kelas	Precision (%)	Recall (%)	F1_Score (%)
Background	50	50	50
Pedestrian	0	NaN	NaN
Zebracross	100	50	66,67

.mp4. Proses *testing* dibagi menjadi dua tahapan, tahap pertama memecah *frame* video menjadi gambar dan melakukan prediksi serta tahap kedua menggabungkan kembali setiap *frame* menjadi video. Tabel ?? menunjukkan waktu yang dibutuhkan untuk melakukan semua tahap prediksi pada *file input* berbentuk video pada ketiga *backbone*.

Tabel 4.9: Waktu Prediksi pada *File Input* Video dalam MM:SS

Backbone	Tahap 1	Tahap 2	Total	Output Size
ResNet-50	07:24	00:38	08:03	88.8 MB
ResNet-101	07:54	00:49	08:44	87.7 MB
MobileNet-v1	06:50	00:40	07:31	89.7 MB

4.2.2 Pengujian di Siang Hari

Pada pengujian ini *testing data* diambil dari video streaming *youtube*[26] yang memiliki resolusi 1920×1080 px berdurasi 29.4 s. Video tersebut mempunyai banyak *frame* per detik sebanyak 25 fps serta total *frame* yang dihasilkan sejumlah 735 *frame*. Gambar 4.15 merupakan perbandingan hasil deteksi dan segementasi dari ketiga *backbone* yang digunakan pada waktu siang hari.



(a) ResNet-50 (b) ResNet-101 (c) MobileNet-v1

Gambar 4.15: Perbandingan Hasil pada Siang Hari

Perbandingan hasil evaluasi yang didapatkan masing-masing *backbone* dengan membandingkan nilai *precision*, *recall* dan *F1-score* dapat dilihat pada Tabel 4.10.

Pada Resnet-50 *Average Precision* untuk semua kelas yang berhasil dideteksi adalah sebesar 83.334% dan *Average Recall* 83.334% dengan *testing time* selama 1,253 s. Sedangkan pada Resnet-101 memberikan hasil *AP* sebesar 100% dan *AR* 100% dengan *testing time* selama 1,549 s. Namun pada Mobilenet-v1 objek yang berhasil dideteksi hanya *zebra cross* (ditambah dengan *background*) sama saat pengujian pada pagi hari. Nilai *AP* yang dihasilkan pada evaluasi Mobilenet-v1 sebesar 50% dengan *testing time* selama 1,186 s.

Selain menggunakan *input data* berupa gambar, pada pengujian ini juga menggunakan *input data* berupa *file video* dengan format *.mp4*. Proses *testing* dibagi menjadi dua tahapan, tahap pertama memecah *frame* video menjadi gambar dan melakukan prediksi serta tahap kedua menggabungkan kembali setiap *frame* menjadi video. Tabel ?? menunjukkan waktu total yang dibutuhkan untuk melakukan semua tahap prediksi pada *file input* berbentuk video pada ketiga *backbone*.

Tabel 4.10: Perbandingan Hasil Evaluasi pada Siang Hari

(a) ResNet-50

Kelas	Precision (%)	Recall (%)	F1_Score (%)
Background	50	100	66,67
Pedestrian	100	50	66,67
Zebracross	100	100	100

(b) ResNet-101

Kelas	Precision (%)	Recall (%)	F1_Score (%)
Background	100	100	100
Pedestrian	100	100	100
Zebracross	100	100	100

(c) MobileNet-v1

Kelas	Precision (%)	Recall (%)	F1_Score (%)
Background	50	66.67	57,14
Pedestrian	0	NaN	NaN
Zebracross	100	33,33	50

Tabel 4.11: Waktu Prediksi pada *File Input* Video dalam MM:SS

Backbone	Tahap 1	Tahap 2	Total	Output Size
ResNet-50	05:33	00:30	06:03	52.8 MB
ResNet-101	07:21	00:36	07:58	50.1 MB
MobileNet-v1	04:03	00:30	04:33	42 MB

4.2.3 Pengujian di Malam Hari

Pada pengujian ini *testing data* diambil dari video *streaming youtube* [27] yang memiliki resolusi 1920×1080 px berdurasi 28.04 s. Video tersebut mempunyai banyak *frame* per detik sebanyak 25 fps serta total *frame* yang dihasilkan sejumlah 701 *frame*. Gambar 4.16 merupakan perbandingan hasil deteksi dan segmentasi dari ketiga *backbone* yang digunakan pada waktu siang hari.



(a) ResNet-50

(b) ResNet-101

(c) MobileNet-v1

Gambar 4.16: Perbandingan Hasil pada Malam Hari

Perbandingan hasil evaluasi yang didapatkan masing-masing *backbone* dengan membandingkan nilai *precision*, *recall* dan *F1-score* dapat dilihat pada Tabel 4.12.

Tabel 4.12: Perbandingan Hasil Evaluasi pada Malam Hari

(a) ResNet-50

Kelas	<i>Precision (%)</i>	<i>Recall (%)</i>	<i>F1-Score (%)</i>
Background	100	25	40
Pedestrian	40	100	57.14
Zebracross	100	100	100

(b) ResNet-101

Kelas	<i>Precision (%)</i>	<i>Recall (%)</i>	<i>F1-Score (%)</i>
Background	100	33.33	50
Pedestrian	60	100	74.9
Zebracross	100	100	100

(c) MobileNet-v1

Kelas	<i>Precision (%)</i>	<i>Recall (%)</i>	<i>F1-Score (%)</i>
Background	50	16,64	28,57
Pedestrian	0	Nan	Nan
Zebracross	100	100	100

Pada gambar yang diuji untuk malam hari hanya memiliki 2 kelas saja yaitu *background*, pejalan kaki serta *zebracross*. Resnet-50 mempunyai nilai *Average Precision* untuk semua kelas yang berha-

sil dideteksi adalah sebesar 80% dan *Average Recall* 75% dengan *testing time* selama 1,069 s. Sedangkan pada Resnet-101 memberikan hasil *AP* sebesar 86.67% dan *AR* 77.77% dengan *testing time* selama 1,215 s. Namun pada Mobilenet-v1 objek yang berhasil dideteksi hanya *zebracross* (ditambah dengan *background*) sama saat pengujian pada pagi dan siang hari. Nilai *AP* yang dihasilkan pada evaluasi Mobilenet-v1 sebesar 66,67% dengan *testing time* selama 1,051 s.

Selain menggunakan *input data* berupa gambar, pada pengujian ini juga menggunakan *input data* berupa *file* video dengan format *.mp4*. Proses *testing* dibagi menjadi dua tahapan, tahap pertama memecah *frame* video menjadi gambar dan melakukan prediksi serta tahap kedua menggabungkan kembali setiap *frame* menjadi video. Tabel ?? menunjukkan waktu yang dibutuhkan untuk melakukan semua tahap prediksi pada *file input* berbentuk video pada ketiga *backbone*.

Tabel 4.13: Waktu Prediksi pada *File Input* Video dalam *MM:SS*

<i>Backbone</i>	<i>Tahap 1</i>	<i>Tahap 2</i>	<i>Total</i>	<i>Output Size</i>
<i>ResNet-50</i>	04:23	00:28	04:51	38.7 MB
<i>ResNet-101</i>	04:49	00:34	05:23	39.1 MB
<i>MobileNet-v1</i>	03:39	00:28	04:07	36.1 MB

4.2.4 Perbandingan Hasil Evaluasi pada Perbedaan Waktu

Setelah membandingkan hasil evaluasi model setiap perbedaan waktu, pada bagian ini akan dibandingkan performa setiap model untuk keseluruhan perbedaan waktu (pagi, siang dan malam hari).

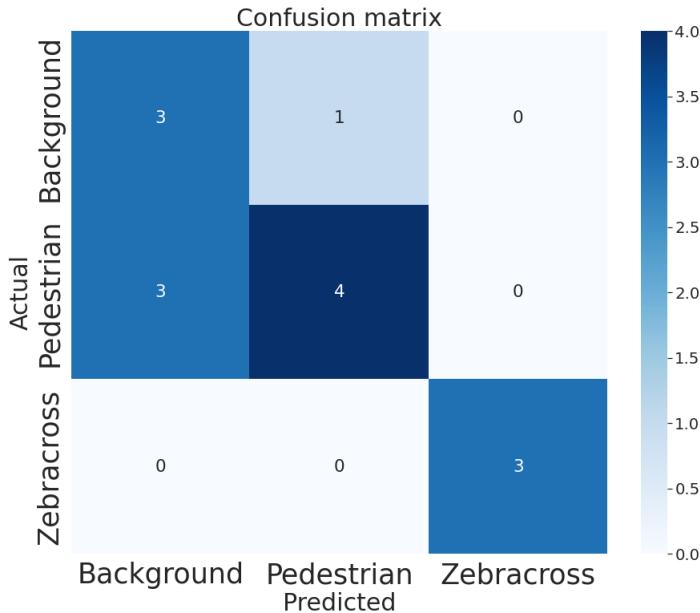
ResNet-50

Pada proses evaluasi model dengan menggunakan seluruh *testing data* (pagi, siang dan malam hari) didapatkan beberapa hasil skor yaitu :

1. *mean Average Precision(mAP)* : 77.381%
2. *mean Average Recall(mAR)* : 76.667%

3. $F1\text{-score} : 75.555\%$

Hasil tersebut didapatkan setelah dilakukan proses pencarian nilai *Trur Positif*, *False Positif* dan *False Negatif* seperti yang ditunjukkan pada *confusion matrix* pada Gambar 4.17.



Gambar 4.17: *Confusion Matrix* ResNet-50

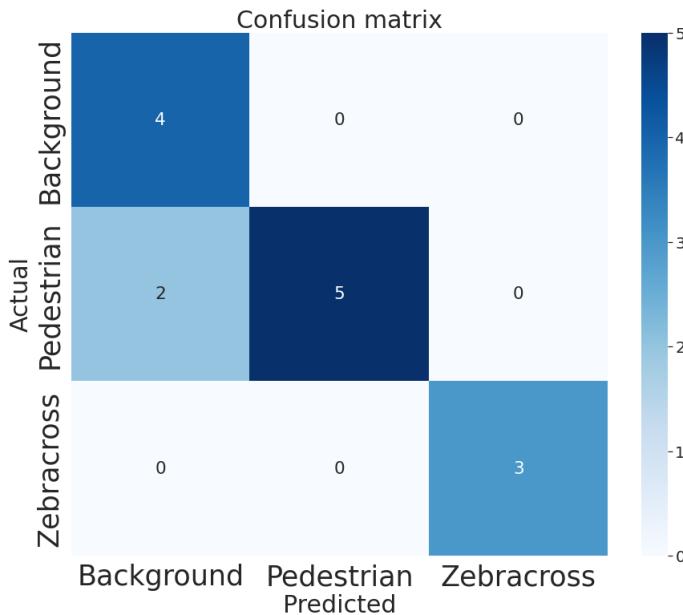
ResNet-101

Pada proses evaluasi model dengan menggunakan seluruh *testing data* (pagi, siang dan malam hari) didapatkan beberapa hasil skor yaitu :

1. $mean\ Average\ Precision(mAP) : 90.476\%$
2. $mean\ Average\ Recall(mAR) : 88.889\%$
3. $F1\text{-score} : 87.777\%$

Hasil tersebut didapatkan setelah dilakukan proses pencarian nilai *Trur Positif*, *False Positif* dan *False Negatif* seperti yang ditunjukkan pada *confusion matrix* pada Gambar 4.18.

ditunjukkan pada *confusion matrix* pada Gambar 4.18.



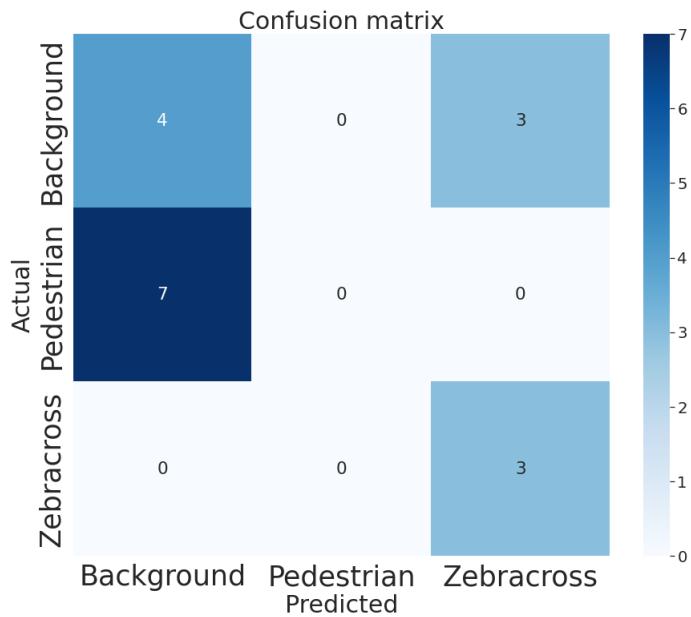
Gambar 4.18: *Confusion Matrix* ResNet-101

MobileNet-v1

Pada proses evaluasi model dengan menggunakan seluruh *testing data* (pagi, siang dan malam hari) didapatkan beberapa hasil skor yaitu :

1. *mean Average Precision(mAP)* : 52.381%
2. *mean Average Recall(mAR)* : *NaN*
3. *F1-score* : *NaN*

Hasil tersebut didapatkan setelah dilakukan proses pencarian nilai *Trur Positif*, *False Positif* dan *False Negatif* seperti yang ditunjukkan pada *confusion matrix* pada Gambar 4.19.



Gambar 4.19: *Confusion Matrix* MobileNet-v1

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil pengujian yang telah dilakukan, penulis dapat menyimpulkan beberapa hal sebagai berikut:

1. Dalam penelitian ini telah diimplementasikan dengan baik proses deteksi dan segmentasi pejalan kaki dan zebra cross dengan menggunakan Mask R-CNN, dengan *mean Average Precision* sebesar 90,476%.
2. *Backbone* ResNet-101 memiliki hasil akurasi yang lebih baik dibanding dengan *backbone* lainnya dengan performa lebih tinggi sebesar 16,92% dibanding ResNet-50 dan 71,73% dibanding MobileNet-v1.
3. Performa pendekripsi dapat berjalan baik dalam beberapa kondisi waktu apabila memiliki cukup pencahayaan..

5.2 Saran

Untuk pengembangan lebih lanjut pada penelitian mendatang, maka penulis memiliki saran sebagai berikut:

1. Menambah jumlah *dataset* yang masih sedikit pada kelas pejalan kaki dan *zebracross*.
2. Pembuatan dataset pejalan kaki di Indonesia karena perilaku pejalan kaki antar negara memiliki perbedaan yang cukup signifikan.

[Halaman ini sengaja dikosongkan]

DAFTAR PUSTAKA

- [1] Supervised vs. unsupervised learning, 2018. URL <https://towardsdatascience.com/supervised-vs-unsupervised-learning-14f68e32ea8d>. (Dikutip pada halaman xi, 9).
- [2] Clustering in machine learning, 2020. URL <https://www.geeksforgeeks.org/clustering-in-machine-learning>. (Dikutip pada halaman xi, 9).
- [3] Hongbo Gao, Guanya Shi, Guotao Xie, and Bo Cheng. Car-following method based on inverse reinforcement learning for autonomous vehicle decision-making. *International Journal of Advanced Robotic Systems*, 15:172988141881716, 11 2018. doi: 10.1177/1729881418817162. (Dikutip pada halaman xi, 10).
- [4] N. Ferracuti, Claudia Norscini, Emanuele Frontoni, P. Gabellini, Marina Paolanti, and Valerio Placidi. A business application of rtls technology in intelligent retail environment: Defining the shopper's preferred path and its segmentation. *Journal of Retailing and Consumer Services*, 47:184–194, 03 2019. doi: 10.1016/j.jretconser.2018.11.005. (Dikutip pada halaman xi, 11).
- [5] R-cnn — region based cnns, 2020. URL <https://www.geeksforgeeks.org/r-cnn-region-based-cnns/>. (Dikutip pada halaman xi, 12).
- [6] Lan Hu. Robot indoor text contents recognition based on visual slam. *Journal of Physics: Conference Series*, 1302:032004, 08 2019. doi: 10.1088/1742-6596/1302/3/032004. (Dikutip pada halaman xi, 13).
- [7] Cuong Nguyen, Giang Son Tran, Thi Nghiem, Nhat Doan, Damien Gratadour, Jean-Christophe Burie, and Chi Luong. Towards real-time smile detection based on faster re-

- gion convolutional neural network. pages 1–6, 04 2018. doi: 10.1109/MAPR.2018.8337524. (Dikutip pada halaman xi, 14).
- [8] Lukasz Bienias, Juanjo n, Line Nielsen, and Tommy Alstrøm. Insights into the behaviour of multi-task deep neural networks for medical image segmentation. pages 1–6, 10 2019. doi: 10.1109/MLSP.2019.8918753. (Dikutip pada halaman xi, 15).
- [9] Ammar Mahmood, Ana Giraldo, Mohammed Bennamoun, Senjian An, Ferdous Sohel, Farid Boussaid, Renae Hovey, Robert Fisher, and Gary Kendrick. Automatic hierarchical classification of kelps using deep residual features. *Sensors*, 20:447, 01 2020. doi: 10.3390/s20020447. (Dikutip pada halaman xi, 18).
- [10] Jiayao Chen, Mingliang Zhou, Dongming Zhang, H. Huang, and Fengshou Zhang. Quantification of water inflow in rock tunnel faces via convolutional neural network approach. *Automation in Construction*, 123:103526, 03 2021. doi: 10.1016/j.autcon.2020.103526. (Dikutip pada halaman xi, 18).
- [11] More than 270 000 pedestrians killed on roads each year, 2013. URL <https://www.who.int/news/item/02-05-2013-more-than-270-000-pedestrians-killed-on-roads-each-year> (Dikutip pada halaman 1).
- [12] Indonesia, indonesia's road safety country profile, 2016. URL <https://www.roadsafetyfacility.org/country/indonesia>. (Dikutip pada halaman 1).
- [13] Anelia Angelova, Alex Krizhevsky, Vincent Vanhoucke, Abhijit Ogale, and Dave Ferguson. Real-time pedestrian detection with deep network cascades. 2015. (Dikutip pada halaman 5).
- [14] Irtiza Hasan, Shengcai Liao, Jinpeng Li, Saad Ullah Akram, and Ling Shao. Pedestrian detection: The elephant in the room, 03 2020. (Dikutip pada halaman 6).
- [15] Chenchen Xu, Guili Wang, Songsong Yan, Jianghua Yu, Baojun Zhang, Shu Dai, Yu Li, and Lin Xu. Fast vehicle

and pedestrian detection using improved mask r-cnn. *Mathematical Problems in Engineering*, 2020:1–15, 05 2020. doi: 10.1155/2020/5761414. (Dikutip pada halaman 6).

- [16] Mon Arjay Malbog. Mask r-cnn for pedestrian crosswalk detection and instance segmentation. In *2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS)*, pages 1–5, 2019. doi: 10.1109/ICETAS48360.2019.9117217. (Dikutip pada halaman 7).
- [17] Shagan Sah. Machine learning: A review of learning types. 07 2020. doi: 10.20944/preprints202007.0230.v1. (Dikutip pada halaman 8).
- [18] Amitha Mathew, Amudha Arul, and S. Sivakumari. *Deep Learning Techniques: An Overview*, pages 599–608. 01 2021. ISBN 978-981-15-3382-2. doi: 10.1007/978-981-15-3383-9_54. (Dikutip pada halaman 9).
- [19] Anirudha Ghosh, A. Sufian, Farhana Sultana, Amlan Chakrabarti, and Debashis De. *Fundamental Concepts of Convolutional Neural Network*, pages 519–567. 01 2020. ISBN 978-3-030-32643-2. doi: 10.1007/978-3-030-32644-9_36. (Dikutip pada halaman 10).
- [20] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Region-based convolutional networks for accurate object detection and segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(1):142–158, 2016. doi: 10.1109/TPAMI.2015.2437384. (Dikutip pada halaman 12).
- [21] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, December 2015. (Dikutip pada halaman 12).
- [22] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39, 06 2015. doi: 10.1109/TPAMI.2016.2577031. (Dikutip pada halaman 13).

- [23] Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017. (Dikutip pada halaman 14, 16).
- [24] Demystifying region proposal network (rpn), 2020. URL <https://medium.com/@nabil.madali/demystifying-region-proposal-network-rpn-faa5a8fb8fce>. (Dikutip pada halaman 16).
- [25] Mobilenet: Deteksi objek pada platform mobile, 2018. URL <https://medium.com/nodelflux/mobilenet-deteksi-objek-pada-platform-mobile-bbbf3806e4b3>. (Dikutip pada halaman 19).
- [26] 4k60 driving around downtown jakarta from kota tua to blok m via thamrin & sudirman street view, 2020. URL https://www.youtube.com/watch?v=_37N1GrVuY&t=519s. (Dikutip pada halaman 54).
- [27] Amazing jakarta indonesia driving downtown - night drive 2021, 2021. URL <https://www.youtube.com/watch?v=JRUasFEC0eI&t=171s>. (Dikutip pada halaman 55).

BIOGRAFI PENULIS



Agung Wicaksono, atau biasa dipanggil Agung, lahir di Madiun Jawa Timur pada tanggal 20 Mei 1999. Merupakan anak kedua dari tiga bersaudara. Penulis lulus dari SMP Negeri 1 Geger dan melanjutkan ke SMA Negeri 1 Geger. Penulis melanjutkan ke jenjang strata satu di Departemen Teknik Komputer Fakultas Teknologi Elektro dan Informatika Cerdas ITS. Dalam masa perkuliahan, penulis tertarik dengan pengembangan *Web Apps* dan *Machine Learning*. Penulis pernah aktif menjadi Staf Departemen Pengembangan Sumber Daya Mahasiswa Badan Eksekutif Mahasiswa Fakultas Teknologi Elektro serta Staf Ahli Kestari Mage 5. Bagi pembaca yang memiliki kritik, saran, atau pertanyaan mengenai tugas akhir ini dapat menghubungi penulis melalui email wicaksonoagun05@gmail.com.

[Halaman ini sengaja dikosongkan]