

KAEDAH AUGMENTASI IMEJ SINTETIK UNTUK
PENGECAMAN BOT DAN KAPAL
MENGUNAKAN RANGKAIAN U-NET

FAKHRUL AZRAN BIN NAWI

UNIVERSITI KEBANGSAAN MALAYSIA

**KAEDAH AUGMENTASI IMEJ SINTETIK UNTUK PENGECAMAN BOT DAN
KAPAL MENGGUNAKAN RANGKAIAN U-NET**

FAKHRUL AZRAN BIN NAWI

**PROJEK YANG DIKEMUKAKAN UNTUK MEMENUHI SEBAHAGIAN
DARIPADA SYARAT MEMPEROLEHI IJAZAH SARJANA SAINS DATA**

**FAKULTI TEKNOLOGI DAN SAINS MAKLUMAT
UNIVERSITI KEBANGSAAN MALAYSIA
BANGI**

2021

PENGAKUAN

Saya akui karya ini adalah hasil kerja saya sendiri kecuali nukilan dan ringkasan yang tiap-tiap satunya telah saya jelaskan sumbernya.

21 Oktober 2021

FAKHRUL AZRAN BIN NAWI
P98822

PENGHARGAAN

Syukur ke hadrat Allah S.W.T. dengan limpah kurnianya diberi kesempatan untuk saya menimba ilmu di Universiti Kebangsaan Malaysia dan diakhiri dengan pelaksanaan kajian dan penyelidikan bagi projek akhir di sini. Ucapan setinggi-tinggi terima kasih kepada penyelia saya iaitu Profesor Madya Dr. Azizi Bin Abdullah yang telah memberi bimbingan, tunjuk ajar dan pandangan berkaitan projek ini.

Ucapan penghargaan juga kepada warga pendidik dan pengurusan Fakulti Teknologi dan Sains Maklumat (FTSM) yang telah memberikan ilmu yang bermanfaat, dan menyediakan kemudahan dan kelengkapan untuk pelajar-pelajar seperti saya. Segala jasa yang telah diberikan ini akan dijadikan panduan untuk saya menjadi seorang yang lebih baik dan dapat memberi bakti kepada agama, bangsa, dan negara.

Tidak ketinggalan juga ucapan terima kasih kepada keluarga saya terutama buat isteri yang tercinta kerana sentiasa memberi sokongan dan dorongan untuk menyiapkan projek akhir ini. Akhir sekali ucapan setinggi-tinggi terima kasih kepada rakan-rakan seperjuangan di fakulti dan di luar fakulti yang tidak jemu-jemu berkongsi pendapat, idea dan memberi komen sama ada secara langsung dan juga tidak langsung

Pusat Sumber
FTSM

ABSTRAK

Pengesanan bot dan kapal dari kamera jarak jauh adalah salah satu tugas yang mencabar kerana saiz objek yang terlalu kecil jika diambil daripada pesawat. Sangat sukar untuk melihat dan mengenal pasti objek bot dan kapal kerana gangguan suasana persekitaran terutamanya daripada gelombang lautan dan pantulan cahaya matahari. Kepentingan untuk menentukan bot dan kapal adalah sangat diperlukan untuk tujuan keselamatan terutamanya jika kawasan adalah berdekatan dengan zon sempadan. Pada masa sekarang, keperluan seorang pakar yang berpengalaman adalah satu kemestian bagi menentukan sama ada objek itu adalah objek bot, kapal ataupun bukan. Cabaran utama dalam mengesan objek bot dan kapal adalah kerana saiz objek yang terlalu kecil jika menggunakan set data yang diambil dari pesawat seperti dron. Set data lain yang ada, yang berkenaan dengan bot dan kapal didapati tidak sesuai untuk kegunaan pemindahan pembelajaran kerana sudut objek yang tidak sama. Set data yang bersesuaian dengan kes penggunaan kajian ini sangat terhad dan telah merumitkan proses untuk mendapatkan model pembelajaran mendalam yang berketepatan tinggi. Teknik-teknik imej augmentasi yang sedia ada juga kurang efektif dengan kes penggunaan kajian ini. Kajian ini menggunakan set data Seagull kerana sudut pandangan dan jarak kamera yang hampir sama dengan kes penggunaan kajian ini. Walau bagaimanapun, set data ini sangat terhad dan ia mempengaruhi prestasi ketepatan model. Objektif kajian ini adalah untuk mengkaji penggunaan pembelajaran mendalam dengan menggunakan teknik imej segmentasi semantik iaitu U-Net bagi mengesan objek bot dan kapal. Selain daripada itu, kajian ini juga berjaya membangunkan dua kaedah imej augmentasi dengan menggunakan teknik imej sintetik yang dapat meningkatkan prestasi model pembelajaran mendalam. Teknik yang pertama memfokuskan kepada proses penempatan semula objek asal ke lokasi lain. Teknik yang kedua memfokuskan kepada proses penggandaan objek asal ke jumlah objek yang lebih banyak. Di akhir kajian ini, satu penilaian dan perbandingan telah dilaksanakan bagi mengukur tahap prestasi ketepatan model hasil daripada pembangunan kaedah imej augmentasi imej sintetik. Kajian ini memfokuskan penyelidikannya dengan menggunakan teknik segmentasi semantik sebagai seni bina utama bagi mengesan objek bot dan kapal. Kajian ini juga melakukan beberapa eksperimen berkenaan dengan penggunaan *hyper parameter* dan fungsi pengoptimum yang berbeza bagi mendapatkan prestasi model pembelajaran yang optimum. Hasil daripada kajian ini menunjukkan bahawa ketepatan *Dice Coefficient* telah meningkat daripada 70% kepada 77% dengan menggunakan imej sintetik penggandaan objek dan meningkat kepada 79% dengan menggunakan imej sintetik penempatan semula. Berdasarkan kepada hasil kajian, ia dapat disimpulkan bahawa terdapat perbezaan yang signifikan dalam menerapkan imej sintetik berbanding dengan hanya menggunakan set data yang asal.

SYNTHETIC IMAGE AUGMENTATION METHOD FOR BOAT AND SHIP DETECTION USING U-NET

ABSTRACT

Boats and ships detection from a range camera is one of the most challenging tasks due to the size of objects that are too small if captured from the aircraft. It is very difficult to see and identify the objects of boats and ships due to the disturbance by the surrounding environment especially from the ocean waves and the reflection of sunlight. The importance of determining boats and ships is very much needed for security purposes especially if the area is close to the border zone. Nowadays, the need for an experienced expert is a must to determine whether the object is a boat, ship, or not. The main challenge in detecting the boat and ship objects is because of the size of the objects that are too small if using the data sets taken from aircraft such as drones. Other available data sets, which apply to boats and ships were found to be unsuitable for transfer learning due to different angles. The appropriate data set that is similar to the use case of this study is very limited and it complicates the process to get high accuracy of deep learning model. Existing data augmentation techniques are less effective for the use case of this study. This study uses the Seagull data set because the viewing angle and camera distance are very similar to the use case. However, this data set is very limited, and it affects the accuracy performance of the model. The objective of this study is to examine the use of deep learning by using semantic image segmentation techniques namely U-Net to detect boat and ship objects. In addition, this study also successfully developed two data augmentation method using synthetic image techniques that can improve the performance of deep learning models. The first technique focuses on the process of relocating the original object to another location. The second technique focuses on the process of multiplying the original object into a higher number of objects. At the end of this study, an evaluation and comparison had been performed to measure the level of model performance accuracy resulting from the development of a synthetic image augmentation data method. This study focuses its research on using semantic segmentation techniques as the main architecture for tracking boat and ship objects. This study also conducted several experiments with respect to the use of hyper parameters and different optimizer functions to obtain the optimal performance of the learning model. The results of this study show that the accuracy of Dice Coefficient has increased from 70% to 77% by using synthetic image of object multiplication and increased to 79% by using synthetic image of relocation. Based on the results of the study, it can be concluded that there is a significant difference in applying synthetic images compared to using only the original data set.

KANDUNGAN

	Halaman
PENGAKUAN	ii
PENGHARGAAN	iii
ABSTRAK	iv
ABSTRACT	v
KANDUNGAN	vi
SENARAI JADUAL	ix
SENARAI ILUSTRASI	x
SENARAI SINGKATAN	xiii
BAB I	PENGENALAN
1.1	Pengenalan 1
1.2	Latar Belakang 3
1.3	Penyataan Masalah 3
1.4	Objektif 5
1.5	Persoalan Kajian 5
1.6	Skop Dan Batasan Kajian 5
1.7	Kepentingan Kajian 5
1.8	Struktur Organisasi Tesis 6
1.9	Kesimpulan 6
BAB II	KAJIAN KESUSASTERAAN
2.1	Pengenalan 7
2.2	Penglihatan Komputer 7
2.3	Rangkaian Neural Konvolusi 8
2.4	Latar Belakang Imej Segmentasi 10
2.5	Seni Bina Rangkaian Neural Konvolusi 13
	2.5.1 <i>Visual Geometry Group (VGG)</i> 13
	2.5.2 <i>Residual Network (ResNet)</i> 15
2.6	Segmentasi Semantik 18
	2.6.1 <i>Fully Convolutional Network (FCN)</i> 20
	2.6.2 <i>U-Net</i> 22

2.7	Kaedah Imej Augmentasi	24
	2.7.1 Teknik Manipulasi Gambar Asas	25
	2.7.2 Teknik Pembelajaran Mendalam	29
2.8	Fungsi Pengaktifan	31
	2.8.1 <i>Rectified Linear Unit</i> (ReLU)	31
	2.8.2 Leaky ReLU	32
	2.8.3 Swish	32
2.9	Sukatan Pengujian Ketepatan	33
	2.9.1 Intersection Of Union (IoU)	34
	2.9.2 Dice Coefficient	35
2.10	Pengesanan Bot, Kapal dan Objek Bersaiz Kecil	35
2.11	Kesimpulan	44
BAB III	METODOLOGI	
3.1	Pengenalan	45
3.2	Kerangka Aliran Proses	45
3.3	Persekitaran Kajian	46
	3.3.1 Perkakasan	46
	3.3.2 Perisian Luar Talian	47
	3.3.3 Perisian Atas Talian	47
	3.3.4 Kerangka Pembangunan	48
3.4	Pengumpulan Data	48
3.5	Pra-Pemprosesan	50
3.6	Penganotasian Data	50
	3.6.1 Topeng Imej	51
	3.6.2 Muat Naik Ke Platform Atas Talian	52
3.7	Pembangunan Model	53
	3.7.1 Penyediaan Sumber Kod Seni Bina U-Net	55
	3.7.2 Penyediaan Imej Asal dan Imej Topeng	55
	3.7.3 Penyediaan Imej Sintetik	55
	3.7.4 Muat Naik Set Data Ke Kerangka <i>Data Loader</i> <i>Pytorch</i>	56
	3.7.5 Muat Naik Seni Bina U-Net Ke Pytorch	56
	3.7.6 Penyediaan Ketetapan Akhir Sebelum Latihan Model	56
	3.7.7 Proses Latihan Model	57
	3.7.8 Proses Penyimpanan Model Untuk Penilaian	58
3.8	Eksperimen Dan Dapatan Kajian	58
3.9	Kesimpulan	58

BAB IV	PEMBANGUNAN PENJANA IMEJ SINTETIK	
4.1	Pengenalan	59
4.2	Penggandaan Objek	60
4.3	Penempatan Semula Objek	63
BAB V	EKSPERIMEN DAN DAPATAN KAJIAN	
5.1	Pengenalan	66
5.2	Penentuan Kadar Pembelajaran Dalam Fasa Latihan	68
5.3	Penentuan Kadar Penggunaan Fungsi Pengaktifan Dalam Fasa Latihan	69
5.4	Ujian Penggunaan Imej Sintetik	70
5.5	Kesimpulan	73
BAB VI	KESIMPULAN	
6.1	Pengenalan	75
6.2	Sumbangan Kajian	75
6.3	Cadangan Kajian Lanjutan	76
6.4	Kesimpulan	76
RUJUKAN		77
Lampiran A	Skrip Penghasilan Topeing Imej	81
Lampiran B	Skrip Penjanaan Imej Sintetik Jenis Penggandaan Objek	84
Lampiran C	Skrip Penjanaan Imej Sintetik Jenis Penempatan Semula	89

SENARAI JADUAL

No. Jadual		Halaman
	Jadual 2.1 Senarai kajian-kajian yang telah dijalankan untuk pengesanan bot dan objek kecil	42
Jadual 3.1	Spesifikasi perkakasan	47
Jadual 3.2	Spesifikasi perisian luar talian	47
Jadual 3.3	Spesifikasi perisian atas talian	47
Jadual 3.4	Senarai kerangka pembangunan	48
Jadual 3.5	Spesifikasi imej set data	50
Jadual 3.6	Spesifikasi topeng imej	52
Jadual 3.7	Parameter yang digunakan di dalam latihan model	57
Jadual 5.1	Jumlah keseluruhan imej yang diguna dalam kajian ini	67
Jadual 5.2	Keputusan peratusan ketepatan bagi ujian kadar pembelajaran	69
Jadual 5.3	Keputusan peratusan ketepatan bagi ujian fungsi pengaktifan	69
Jadual 5.4	Keputusan peratusan ketepatan bagi ujian imej sintetik	70

SENARAI ILUSTRASI

No. Rajah		Halaman
Rajah 1.1	Pemandangan laut yang dirakam oleh agensi penguatkuasaan maritim	1
Rajah 2.1	Biologi algoritma yang mengilhamkan rangkaian neural	8
Rajah 2.2	Hubungan diantara rangkaian neural dengan fungsi di dalam otak	9
Rajah 2.3	Hubungan diantara lapisan-lapisan di dalam rangkaian neural	9
Rajah 2.4	Penggunaan segmentasi imej dalam menentukan objek manusia, basikal dan latar belakang	10
Rajah 2.5	Perbezaan di antara teknik pengesanan objek dan segmentasi objek	11
Rajah 2.6	Perbezaan di antara segmentasi semantik dan segmentasi objek	12
Rajah 2.7	Gambaran segmentasi semantik dalam set data Cityscapes	13
Rajah 2.8	Seni bina VGG-16 yang mengandungi 16 lapisan konvolusi	15
Rajah 2.9	Seni bina ResNet-34	16
Rajah 2.10	Reka bentuk blok konvolusi dalam rangkaian residual	16
Rajah 2.11	Contoh pengeluaran imej selepas proses segmentasi semantik	18
Rajah 2.12	Proses pengeluaran segmentasi daripada imej berwarna RGB kepada imej yang mewakili digit untuk menggambarkan kelas	19
Rajah 2.13	Lapisan <i>one-hot encoder</i> yang mewakili digit kelas masing-masing	19
Rajah 2.14	Penyatuan imej segmentasi ke dalam satu topeng (<i>masking</i>)	20
Rajah 2.15	Seni bina FCN	21
Rajah 2.16	Segmentasi semantik yang diramalkan oleh FCN-32, FCN-16 dan FCN-8 serta imej sebenar yang digunakan untuk pengujian	22
Rajah 2.17	Seni bina U-Net	23
Rajah 2.18	Graf hubungan prestasi pembelajaran mendalam dan jumlah set data	24

Rajah 2.19	Gambar yang dihasilkan melalui proses imej augmentasi	26
Rajah 2.20	Gambar hasil daripada teknik putaran gambar Peralihan Gambar	26
Rajah 2.21	Gambar hasil daripada teknik peralihan gambar	27
Rajah 2.22	Gambar hasil daripada teknik pembalikan gambar	27
Rajah 2.23	Gambar hasil daripada teknik gangguan gambar	28
Rajah 2.24	Gambar hasil daripada teknik pengaburan gambar	29
Rajah 2.25	Langkah-langkah pelaksanaan dalam kajian	30
Rajah 2.26	Graf fungsi pengaktifan ReLU	31
Rajah 2.27	Graf fungsi pengaktifan Leaky ReLU	32
Rajah 2.28	Graf fungsi pengaktifan Swish	33
Rajah 2.29	Intersection of Union	34
Rajah 2.30	Gambar set data daripada ADE20k	36
Rajah 2.31	Contoh set data daripada Kaggle Airbus Ship Detection	37
Rajah 2.32	Contoh gambar bot dan kapal dari sudut pandangan CCTV	38
Rajah 2.33	Gambar sebelah kiri menunjukkan masalah pengesanan berlebihan dan sebelah kanan pengeluaran output menggunakan teknik R-DFPN	39
Rajah 2.34	Gambar set data daripada set data daripada ISPRS Vaihingen 2D dan hasil kajian	40
Rajah 3.1	Kerangka aliran proses	46
Rajah 3.2	Set data Seagull dalam bentuk video	48
Rajah 3.3	Imej statik yang diekstrak daripada video Seagull	49
Rajah 3.4	Penganotasian imej menggunakan perisian VIA	50
Rajah 3.5	Imej asli	51
Rajah 3.6	Imej topeng	52
Rajah 3.7	Platform Kaggle yang digunakan dalam kajian ini	53
Rajah 3.8	Contoh set data Carvana	54
Rajah 3.9	Kerangka aliran proses pembangunan model	54

Rajah 3.10	Proses latihan model yang telah dijalankan	57
Rajah 4.1	Lokasi bot di dalam imej	59
Rajah 4.2	Carta alir penjanaan imej sintetik bagi penggandaan objek	61
Rajah 4.3	Gambar menunjukkan imej asli di sebelah kiri dan imej sintetik di sebelah kanan	62
Rajah 4.4	Gambar menunjukkan imej topeng asli di sebelah kiri dan imej topeng sintetik di sebelah kanan	62
Rajah 4.5	Penggandaan objek	63
Rajah 4.6	Carta alir kaedah penjanaan imej sintetik bagi penempatan semula objek	64
Rajah 4.7	Gambar menunjukkan imej asli di sebelah kiri dan imej sintetik di sebelah kanan selepas proses penempatan semula	65
Rajah 4.8	Gambar menunjukkan imej topeng asli di sebelah kiri dan imej topeng sintetik di sebelah kanan selepas proses penempatan semula	65
Rajah 5.1	Gambar menunjukkan contoh imej bagi setiap jenis set data yang digunakan dalam kajian	68
Rajah 5.2	Imej menunjukkan hasil ujian yang menggunakan model dengan set data imej asli	71
Rajah 5.3	Imej menunjukkan hasil ujian yang menggunakan model dengan set data imej asli	71
Rajah 5.4	Imej menunjukkan hasil ujian yang menggunakan model dengan set data imej asli dan imej sintetik penempatan semula	72
Rajah 5.5	Imej menunjukkan hasil ujian yang menggunakan model dengan set data imej asli dan imej sintetik penempatan semula	72
Rajah 5.6	Imej menunjukkan hasil ujian yang menggunakan model dengan set data imej asli dan imej sintetik penggadaan objek	73
Rajah 5.7	Imej menunjukkan hasil ujian yang menggunakan model dengan set data imej asli dan imej sintetik penggandaan objek	73

SENARAI SINGKATAN

CNN	<i>Convolutional Neural Network</i>
GPU	<i>Graphical Processing Unit</i>
CPU	<i>Central Processing Unit</i>
UAV	<i>Unmanned Aerial Vehicles</i>
IR 4.0	<i>Industrial Revolution 4.0</i>
FCN	<i>Fully Convolutional Network</i>
RGB	<i>Red, Green, Blue</i>
VGG	<i>Visual Geometry Group</i>
ReLU	<i>Rectified Linear Unit</i>
ResNet	<i>Residual Network</i>
SGD	<i>Stochastic Gradient Descent</i>
IoU	<i>Intersection Of Union</i>

BAB I

PENGENALAN

1.1 PENGENALAN

Salah satu tanggungjawab agensi maritim Malaysia adalah untuk mengawal seluruh kawasan perairan Malaysia daripada pencerobohan luar. Selain daripada itu, agensi ini juga bertanggung jawab dalam menjalankan pengawasan udara dan lautan dan menjalankan tugas pencarian dan penyelamatan. Pada hari ini, mereka telah dilengkapi dengan sistem pemantauan pesawat. Namun begitu terdapat banyak lagi kerja-kerja operasi yang dilakukan secara manual yang boleh ditambah baik. Salah satu bidang yang boleh di tambah baik dalam operasi pemantauan adalah dengan mengadakan pengesanan awal informasi berkenaan dengan bot dan kapal. Imej pada Rajah 1.1 menunjukkan bagaimana pemandangan laut diambil oleh agensi tersebut menggunakan pesawat dron.



Rajah 1.1 Pemandangan laut yang dirakam oleh agensi penguatkuasaan maritim

Dengan melihat menggunakan mata kasar, adalah sangat sukar bagi manusia untuk mengenali lokasi objek kapal yang berada di dalam imej. Cara yang paling mudah adalah dengan mengenal pasti ekor ombak yang dihasilkan daripada pergerakan kapal untuk menentukan sama ada objek itu adalah sebuah kapal atau bot. Walau bagaimanapun, terdapat banyak kapal di dalam imej yang tidak berada dalam kedudukan bergerak. Terdapat juga pelbagai jenis objek lain seperti perumahan dan pulau kecil yang boleh menyebabkan salah klasifikasi atau terlepas langsung daripada pengesanan. Cadangan dalam kajian ini adalah untuk memanfaatkan teknologi pembelajaran mendalam dalam pengesanan kapal dan bot. Terdapat banyak seni bina pembelajaran mendalam yang sedia ada dan kajian ini akan membandingkan beberapa teknik dalam menambak baik ketepatan seni bina itu.

Aplikasi pembelajaran mendalam dalam pengesanan objek telah popular sejak beberapa tahun kebelakangan ini bermula dengan terciptanya konvolusi rangkaian neural (CNN) oleh Lecun & Haffner (1999). Kemajuan pesat dalam bidang pengkomputeran membolehkan data diproses dengan lebih pantas. Unit pemprosesan grafik (GPU) bukan hanya memberikan enjin grafik yang kuat tetapi juga pemproses pengkomputeran yang berselari dan dapat mencapai kelajuan yang berterusan berbanding dengan CPU (Boyer & El Baz 2013).

Kerangka pembelajaran mesin seperti Tensorflow (Abadi et al. 2016), PyTorch (Ketkar 2017), Keras ("Keras API reference" t.th.), dan MXNet (Chen et al. 2015) membantu mempercepatkan perkembangan pembelajaran mesin dan juga dapat membantu mempercepatkan proses pembuatan pembelajaran mesin ke pengeluaran dengan lebih mudah. Konsep baru di CNN seperti pemindahan pembelajaran (Hussain et al. 2019) dan model pra-latihan membantu mengurangkan jumlah data dan juga memendekkan masa yang diperlukan untuk latihan model CNN. Pemindahan pembelajaran adalah teknik peningkatan pembelajaran dalam tugas yang baru melalui proses pemindahan pengetahuan yang dihasilkan dari tugas yang telah dihasilkan dari pembelajaran mendalam yang lepas.

Pada hari ini, banyak aplikasi telah dibangunkan bagi membantu dalam menyelesaikan sesuatu masalah. Sebagai contoh dalam industri perubatan, kejuruteraan, dan pembangunan. Walaubagaimanapun, masih terdapat bidang industri yang masih menggunakan cara konvensional untuk mengesan objek sebagai contoh dalam industri maritim. Kajian penyelidikan ini memfokuskan kepada pembelajaran mendalam untuk pengesanan objek kapal dan bot.

Terdapat beberapa seni bina pembelajaran mendalam yang boleh digunakan dalam pengesanan objek kecil. Kajian daripada Fan et al. (2019) telah menilai perbezaan prestasi di antara beberapa seni bina pembelajaran mendalam iaitu CFAR, FRCNN, dan U-Net dan mendapati bahawa seni bina U-Net memperolehi prestasi yang tertinggi dengan ketepatan dan *recall* melebihi 90%.

1.2 LATAR BELAKANG

Pengesanan bot dan kapal merupakan subjek penting dalam industri maritim. Pengawasan menggunakan pesawat dron merupakan prosedur utama dalam menentukan tahap keselamatan di perairan. Dengan adanya teknologi pembelajaran mendalam dapat membantu mempercepatkan proses pengesanan bot dan kapal. Namun begitu, kes penggunaan untuk pengesanan bot mempunyai cabaran yang tertentu. Di antara cabaran utamanya adalah saiz bot yang sangat kecil jika di ambil dari jarak pesawat dron. Oleh yang demikian, set data dengan jumlah yang besar diperlukan untuk meningkatkan prestasi ketepatan pembelajaran mendalam. Malangnya set data dengan jumlah yang besar tidak mungkin dapat diperolehi dalam masa yang singkat. Penggunaan penjana imej augmentasi boleh digunakan untuk meningkatkan jumlah set data. Namun ia bukan merupakan satu-satunya jalan penyelesaian yang boleh digunakan.

Kajian ini telah menjalankan eksperimen untuk menghasilkan imej sintetik berdasarkan set data yang sedia ada. Imej sintetik adalah imej baru yang dijana berdasarkan imej asal dengan menggunakan kaedah manipulasi gambar asas. Terdapat dua jenis sintetik imej telah dibangunkan iaitu jenis penambahan objek dan pemindahan semula. Dengan adanya teknik ini, ia dapat menambah jumlah set data dan meningkatkan prestasi ketepatan pembelajaran mendalam.

1.3 PENYATAAN MASALAH

Salah satu faktor penting bagi mencapai prestasi pembelajaran mendalam yang baik adalah dengan menyediakan set data yang cukup dan berkualiti tinggi. Namun begitu, terdapat banyak situasi yang mana data ini tidak dapat disediakan dengan lengkap. Kemungkinan besar adalah kerana data-data ini tidak di simpan dalam jangka masa yang lama kerana saiz data yang besar. Selain daripada itu, data-data yang rutin yang di rekod namun ia tidak memberi manfaat penting

dalam bahagian operasi akan dibuang dan ini merupakan proses biasa dalam penyelenggaraan teknologi maklumat.

Untuk tujuan kajian ini permasalahan utama yang ditekankan adalah berkisarkan

1. Set data yang terhad dan tidak cukup untuk tujuan latihan pembelajaran mendalam.

Set data yang sangat terhad akan mengurangkan ketepatan pembelajaran mendalam. Berdasarkan kajian yang dijalankan oleh Sun et al. (2017), mereka mendapati bahawa prestasi rangkaian meningkat secara linear dengan catatan jumlah data latihan. Dengan kata lain, semakin banyak data semakin baik prestasi rangkaian.

2. Teknik imej augmentasi yang sedia ada kurang efektif pada kes penggunaan pengesanan bot.

Teknik biasa imej augmentasi yang sedia ada lebih sesuai digunakan untuk objek yang lebih besar manakala kes penggunaan untuk kajian ini mempunyai saiz objek sangat kecil.

3. Sukar mendapatkan kaedah yang sesuai dalam penjanaan imej sintetik.

Proses pembikinan imej sintetik kebiasaannya dilakukan secara manual dengan menggunakan perisian seperti *Adobe Photoshop* atau pun perisian *Blender*. Proses ini merumitkan kerja-kerja penyediaan set data dan memakan masa.

4. Kesukaran untuk menentukan jenis seni-bina yang sesuai dalam pengesanan objek yang bersaiz kecil.

Pengesanan objek yang bersaiz kecil di dalam imej mempunyai cabaran-cabaran yang berbeza berbanding objek-objek yang lebih besar. Kajian ini memfokuskan pada U-Net berdasarkan seni binanya yang mempunyai proses pensampelan ke bawah dan pensampelan ke atas dan diakhiri dengan penentuan klasifikasi objek pada peringkat piksel.

1.4 OBJEKTIF

Berdasarkan pernyataan masalah, berikut adalah objektif utama dalam kajian ini

1. Mengkaji model pembelajaran mendalam untuk tujuan semantik segmentasi dengan menggunakan seni bina U-NET.
2. Membangunkan kaedah penjanaan imej sintetik.
3. Menilai prestasi keputusan dan dapatan kajian.

1.5 PERSOALAN KAJIAN

Bagi memastikan matlamat dan objektif kajian tercapai, berikut adalah soalan-soalan yang akan digunakan:

1. Apakah penalaan halus yang penting diperlukan untuk meningkatkan prestasi ketepatan dalam pembelajaran mendalam?
2. Adakah penggunaan sintetik imej dapat meningkatkan prestasi ketepatan dalam pembelajaran mendalam?
3. Adakah penggunaan U-NET boleh digunakan untuk mengesan objek kecil?

1.6 SKOP DAN BATASAN KAJIAN

Pembelajaran mendalam ini memberi fokus kepada pemprosesan imej untuk kenderaan maritim seperti kapal dan bot. Set data yang digunakan adalah diambil pada waktu siang sahaja.

1.7 KEPENTINGAN KAJIAN

Permasalahan yang di fokuskan pada kajian ini adalah permasalahan yang sering dihadapi oleh pelbagai kes penggunaan. Walaupun kajian ini menumpukan pada pengesanan bot, namun teknik yang dihasilkan boleh juga diguna pakai pada domain lain. Hasil daripada kajian ini dapat memberikan penanda aras sama ada pendekatan menggunakan imej sintetik boleh meningkatkan prestasi ketetapan pembelajaran mendalam ataupun tidak. Sekali gus ia dapat

menyumbang sebagai salah satu teknik untuk meningkatkan ketetapan pembelajaran mendalam.

1.8 STRUKTUR ORGANISASI TESIS

Tesis ini di strukturkan pada lima bahagian. Penerangan untuk setiap bahagian adalah seperti berikut

BAB I Pengenalan, bab ini menerangkan tentang informasi kajian secara umum.

BAB II Kajian Kesusasteraan, bab ini membincangkan tentang kajian-kajian terdahulu yang berkaitan dengan kajian ini.

BAB III Metodologi Kajian, bab ini menerangkan kaedah dan reka bentuk ujian yang dijalankan semasa kajian ini dilaksanakan.

BAB IV Pembangunan Penjanaan Sintetik Imej, bab ini membincangkan tentang penghasilan kaedah penjanaan imej sintetik.

BAB V Eksperimen dan Dapatan Kajian, bab ini membincangkan hasil ukuran ketetapan bagi pengujian yang dijalankan.

BAB VI Kesimpulan, bab ini merumuskan tentang kajian secara menyeluruh dan cadangan lanjutan kajian.

1.9 KESIMPULAN

Bab ini menerangkan secara umum tentang kajian yang dijalankan yang merangkumi latar belakang dan huraian pernyataan permasalahan. Seterusnya di sertakan penerangan mengenai objektif dan hala tuju kajian yang dijalankan. Selain tu, bab ini memaparkan tentang skop kajian dan kepentingan perlunya kajian ini dilaksanakan. Di akhir bab ini, disertakan penerangan tentang struktur organisasi kajian.

BAB II

KAJIAN KESUSASTERAAN

2.1 PENGENALAN

BAB II ini akan menerangkan tentang kajian kesusasteraan yang telah dijalankan sebelum ini yang berkisarkan mengenai topik pengesanan objek. Terdapat beberapa topik dan sub-topik penting yang telah dipetik dari kajian-kajian yang lepas telah digunakan untuk tujuan penambah baikkan kajian ini. Kajian kesusasteraan memberi peranan penting untuk membantu dalam memahami proses dan kaedah bersesuaian yang perlu dijalankan supaya matlamat dan objektif kajian ini dapat dicapai. Selain itu, ia juga dapat membantu dalam memilih kaedah yang terbaik untuk digunakan dalam kajian ini bagi memastikan keputusan yang dihasilkan mempunyai ketepatan yang tinggi.

2.2 PENGLIHATAN KOMPUTER

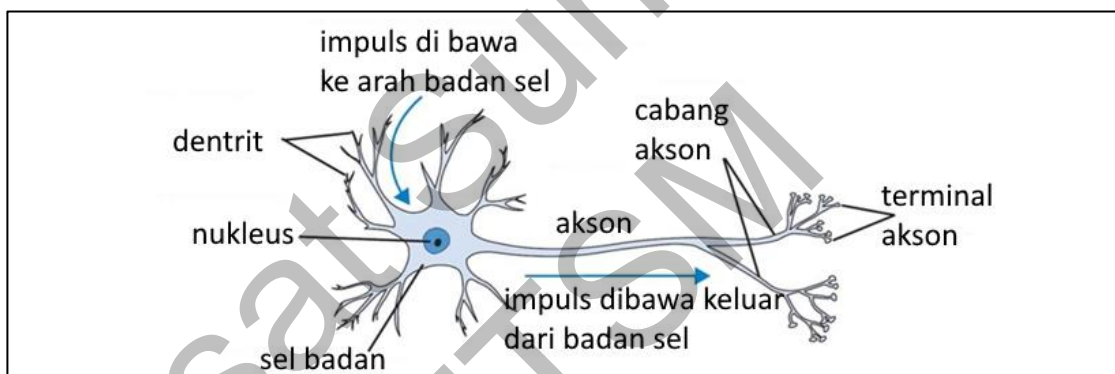
Bidang kecerdasan buatan dalam penglihatan komputer merupakan bidang yang penting dalam era industri revolusi 4.0 (IR 4.0) ini. Terdapat banyak kes penggunaan yang menggunakan pemprosesan imej sebagai cara efektif bagi menyelesaikan sesuatu permasalahan. Salah satu teknik penting dalam topik pemprosesan imej adalah pengesanan objek. Pelbagai seni bina dan kaedah telah dibangunkan sama ada melalui cara pengiraan biasa mahupun menggunakan teknik pembelajaran mendalam. Penglihatan komputer ini mampu memberikan maklumat yang bernilai daripada imej dan video seperti contoh pengesanan objek, pengesanan wajah, dan pengesanan pejalan kaki (Zhao et al. 2019).

Kajian ini menumpukan penggunaan teknik segmentasi imej. Secara amnya, teknik segmentasi imej adalah berbeza dengan teknik pengesanan objek dan klasifikasi objek. Teknik segmentasi imej mengambil kira setiap piksel pada imej bagi tujuan meramalkan dan menentukan jenis objek di dalam imej. Teknik ini telah digunakan

dalam pelbagai industri terutamanya dalam bidang bio perubatan. Sebagai contoh adalah dalam pengesanan tumor (Moon et al. 2002) dan pengesanan papan tanda dan isyarat jalan raya (Maldonado-Bascon et al. 2007), kenderaan tanpa pemandu (Cordts et al. 2016) dan (Fritsch et al. 2013).

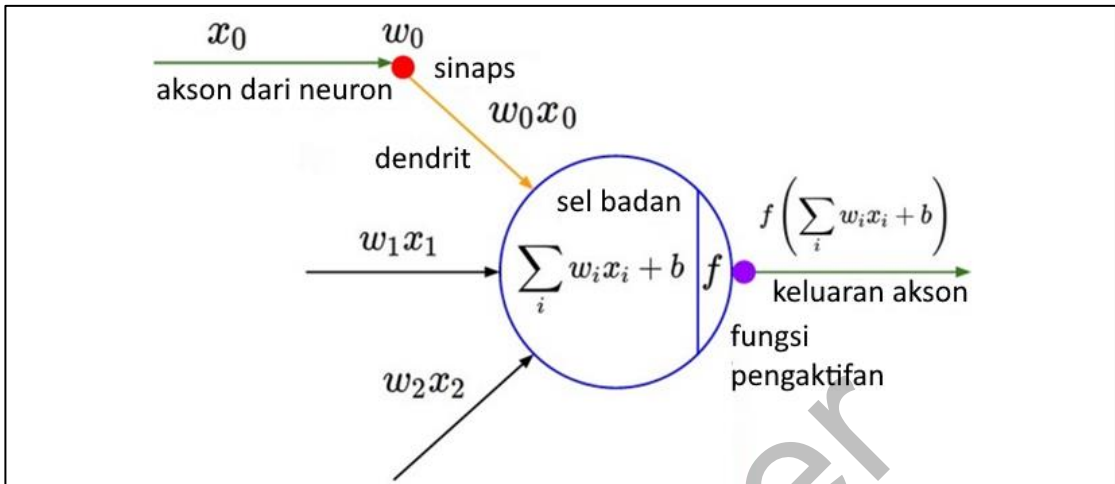
2.3 RANGKAIAN NEURAL KONVOLUSI

Rangkaian neural diilhamkan dari biologi algoritma dengan meniru cara neuron berfungsi di dalam otak. Setiap neuron bertindak untuk membuat pengiraan dengan menerima input dari dentrit dan mengeluarkan isyarat melalui terminal akson. Tindakan ini di aktifkan apabila neuron tertentu telah di aktifkan Rajah 2.1. menunjukkan informasi berkenaan fungsi rangkaian neural di dalam otak.



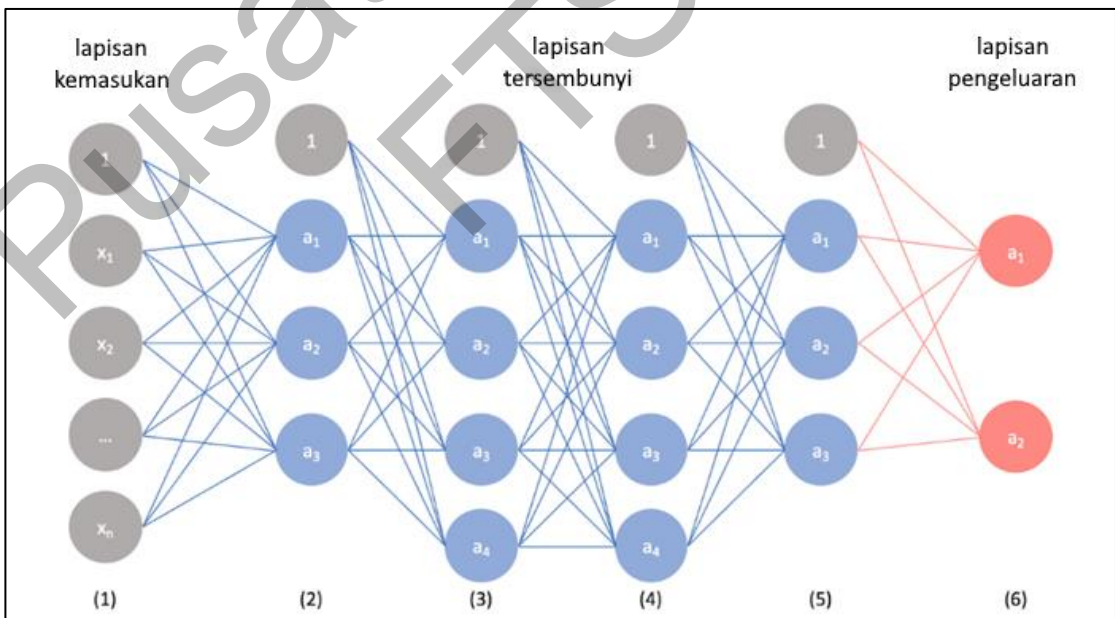
Rajah 2.1 Biologi algoritma yang mengilhamkan rangkaian neural

Secara dasarnya, sel akan bertindak dengan menggunakan input melalui dentrit dan mengeluarkan output melalui terminal akson. Keseluruhan idea di sebalik rangkaian saraf adalah mencari jalan untuk menjalankan fungsi tersebut dan menghubungkan neuron bersama-sama untuk mendapatkan manfaat yang berguna. Rajah 2.2 menunjukkan bagaimana rangkaian neural konvolusi berfungsi yang diilhamkan daripada fungsi otak.



Rajah 2.2 Hubungan diantara rangkaian neural dengan fungsi di dalam otak

Dalam teknologi pembelajaran mendalam, secara asasnya ia merangkumi tiga jenis lapisan iaitu lapisan kemasukan (*input layer*), lapisan tersembunyi (*hidden layer*), dan lapisan pengeluaran (*output layer*). Berdasarkan Rajah 2.3, lapisan pertama adalah merupakan lapisan kemasukan yang mengandungi nod-nod mewakili vektor ciri-ciri kemasukan. Nilai-nilai kemasukan ini kemudian dikira dengan menetapkan pemberat yang sesuai dan ditambahkan dengan nilai berat sebelah.



Rajah 2.3 Hubungan diantara lapisan-lapisan di dalam rangkaian neural

Lapisan tengah pula dikenali sebagai lapisan tersembunyi. Lapisan ini boleh di tambah bilangannya dan juga boleh dikurangkan mengikut kesesuaian. Kekuatan

sebenarnya rangkaian neural adalah berpusat pada lapisan ini. Setiap lapisan tersembunyi ini mesti berada di kedudukan tengah di antara lapisan masukan dan lapisan pengeluaran. Nilai pengeluaran akan ditentukan oleh modul mengoptimimum. Terdapat beberapa jenis lapisan tersembunyi, dan jenis rangkaian konvolusi neural merupakan salah satu lapisan yang sangat sesuai untuk pembelajaran mendalam dalam pemprosesan imej. Lapisan konvolusi adalah lapisan pengekstrakan ciri-ciri (*feature extraction*) di mana penapis (*filter*) diterapkan pada gambar asli. Lapisan konvolusi biasanya terdapat di dalam lapisan tersembunyi. Lapisan yang terakhir adalah lapisan pengeluaran, iaitu lapisan yang menentukan klasifikasi objek dan peratusan keyakinan.

Segmentasi imej pula merupakan proses pelabelan kawasan tertentu di dalam imej bagi menerangkan klasifikasi objek yang dikesan. Tugas utamanya adalah untuk memberikan informasi tentang jenis objek yang berada di dalam imej berserta dengan lokasi. Segmentasi imej secara semantik pula meramalkan dan melabelkan pada setiap piksel di dalam imej. Tugas ini seringkali di kenali sebagai ramalan. Rajah 2.4 menunjukkan bagaimana imej segmentasi dapat meramal jenis objek iaitu manusia, basikal dan latar belakang berserta lokasi objek pada setiap piksel.



Rajah 2.4 Penggunaan segmentasi imej dalam menentukan objek manusia, basikal dan latar belakang

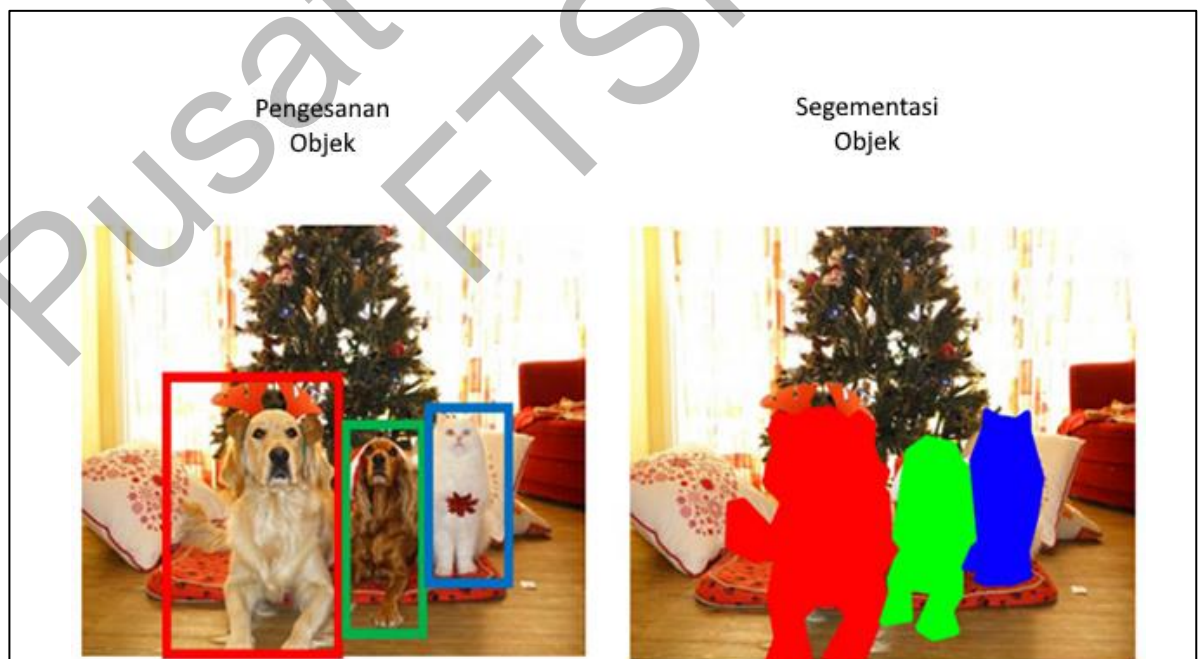
2.4 LATAR BELAKANG IMEJ SEGMENTASI

Segmentasi imej adalah salah satu cabang dalam penglihatan komputer yang mana ia merupakan teknik untuk meramal kawasan tertentu pada imej mengikut klasifikasi yang

ditetapkan. Tugas utama adalah untuk menentukan jenis objek di dalam imej dan di manakah posisi objek itu berada di dalam imej.

Kandungan di dalam imej boleh dibahagikan dalam beberapa bahagian yang di namakan sebagai segmen. Dengan cara pembahagian imej ini, ia dapat mengurangkan pemprosesan pada bahagian imej yang tidak penting dan menumpukan pemprosesan pada bahagian imej yang lebih penting. Imej terdiri daripada koleksi sekumpulan piksel yang berbeza. Dan setiap piksel yang mempunyai atribut yang sama akan diklasifikasi mengikut kelas objek-objek yang tertentu.

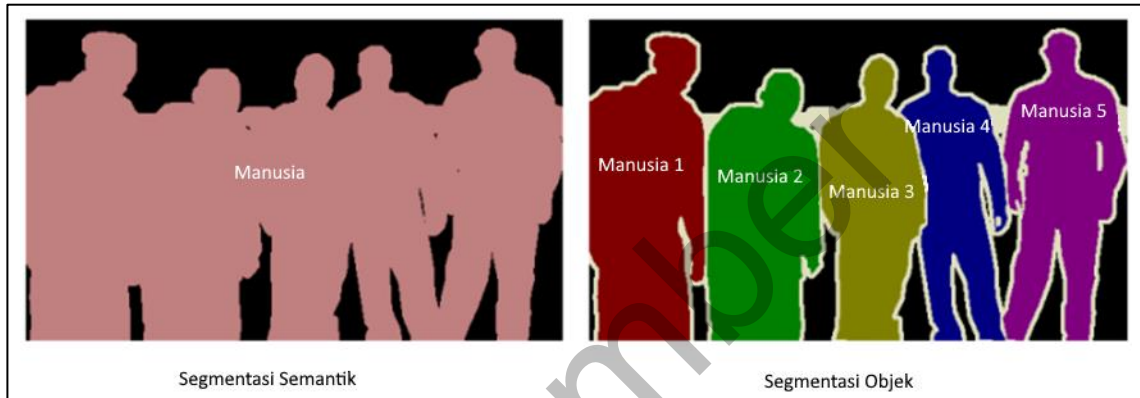
Seni bina dan kaedah pengesanan objek membolehkan pemprosesan imej untuk mendapatkan kotak pembatas yang sesuai mengikut kelas dan kedudukan objek pada imej. Namun begitu, informasi ini tidak cukup untuk memberikan secara tepat bentuk objek yang di berada di dalam imej tersebut. Sebagai contoh imej pada Rajah 2.5 menunjukkan perbezaan diantara informasi dalam bentuk kotak pembatas dan segmentasi.



Rajah 2.5 Perbezaan di antara teknik pengesanan objek dan segmentasi objek

Dengan menggunakan teknik imej segmentasi, ia boleh menghasilkan lapisan piksel untuk setiap objek di dalam imej. Teknik ini dapat memberikan pemahaman yang

jauh lebih tepat dan terperinci tentang bentuk dan kedudukan sebenar objek di dalam imej. Terdapat dua jenis teknik segmentasi imej iaitu segmentasi semantik dan segmentasi objek (*instance segmentation*). Perbezaan antara kedua-dua teknik ini digambarkan oleh Rajah 2.6.



Rajah 2.6 Perbezaan di antara segmentasi semantik dan segmentasi objek

Imej di kiri merupakan jenis segmentasi semantik. Setiap piksel di dalam imej mewakili kelas-kelas tertentu sama ada ia adalah objek manusia ataupun ia adalah sejenis latar belakang. Setiap piksel yang tergolong pada kelas tertentu diwakili dengan warna masing-masing yang kebiasaannya ditetapkan pada setiap kelas dengan setiap lapisan. Imej kiri ini menunjukkan objek kelas manusia dengan warna merah jambu dan objek latar belakang dengan warna hitam.

Imej di sebelah kanan pula merupakan jenis segmentasi objek. Jenis ini juga menetapkan kelas pada setiap piksel. Namun begitu, setiap objek dalam kelas yang sama juga di bezakan antara satu sama lain. Sebagai contoh terdapat lima objek dalam kategori yang sama iaitu manusia yang berwarna-warni dan satu objek berwarna hitam adalah dalam kategori latar belakang.

Kajian ini akan memfokuskan pada segmentasi semantik. Mengikut kajian yang telah dijalankan oleh (Deng et al. 2018), set data yang boleh digunakan dalam segmentasi imej adalah PASCAL VOC, MS COCO, ADE20K, Cityscapes dan KITTI. Rajah 2.7 menunjukkan contoh imej segmentasi menggunakan set data Cityscapes.



Rajah 2.7 Gambaran segmentasi semantik dalam set data Cityscapes

Dalam seni bina segmentasi objek, terdapat beberapa cara yang boleh digunakan untuk tujuan itu. Antaranya adalah segmentasi semantik berdasarkan kawasan, segmentasi semantik berdasarkan *Fully Convolutional Network* (FCN), dan segmentasi semantik berdasarkan *Weakly Supervised*. Kajian daripada Ronneberger et al. (2015) telah menambah baik seni bina FCN dengan memperkenalkan seni bina baru yang dinamakan sebagai seni bina U-Net. Dimotivasikan oleh kajian Ronneberger et al. (2015), kajian ini telah memberi fokus pada penggunaan seni bina U-Net.

2.5 SENI BINA RANGKAIAN NEURAL KONVOLUSI

Beberapa rangkaian neural konvolusi telah dikaji untuk diimplementasikan sebagai lapisan pengekod dalam kaedah segmentasi semantik yang digunakan dalam kajian ini. Jenis seni bina CNN yang sering diguna pakai bersama dengan FCN dan U-Net adalah VGG, AlexNet, GoogleLeNet dan ResNet. Untuk kajian ini, kajian terperinci dilakukan pada jenis VGG dan ResNet.

2.5.1 *Visual Geometry Group* (VGG)

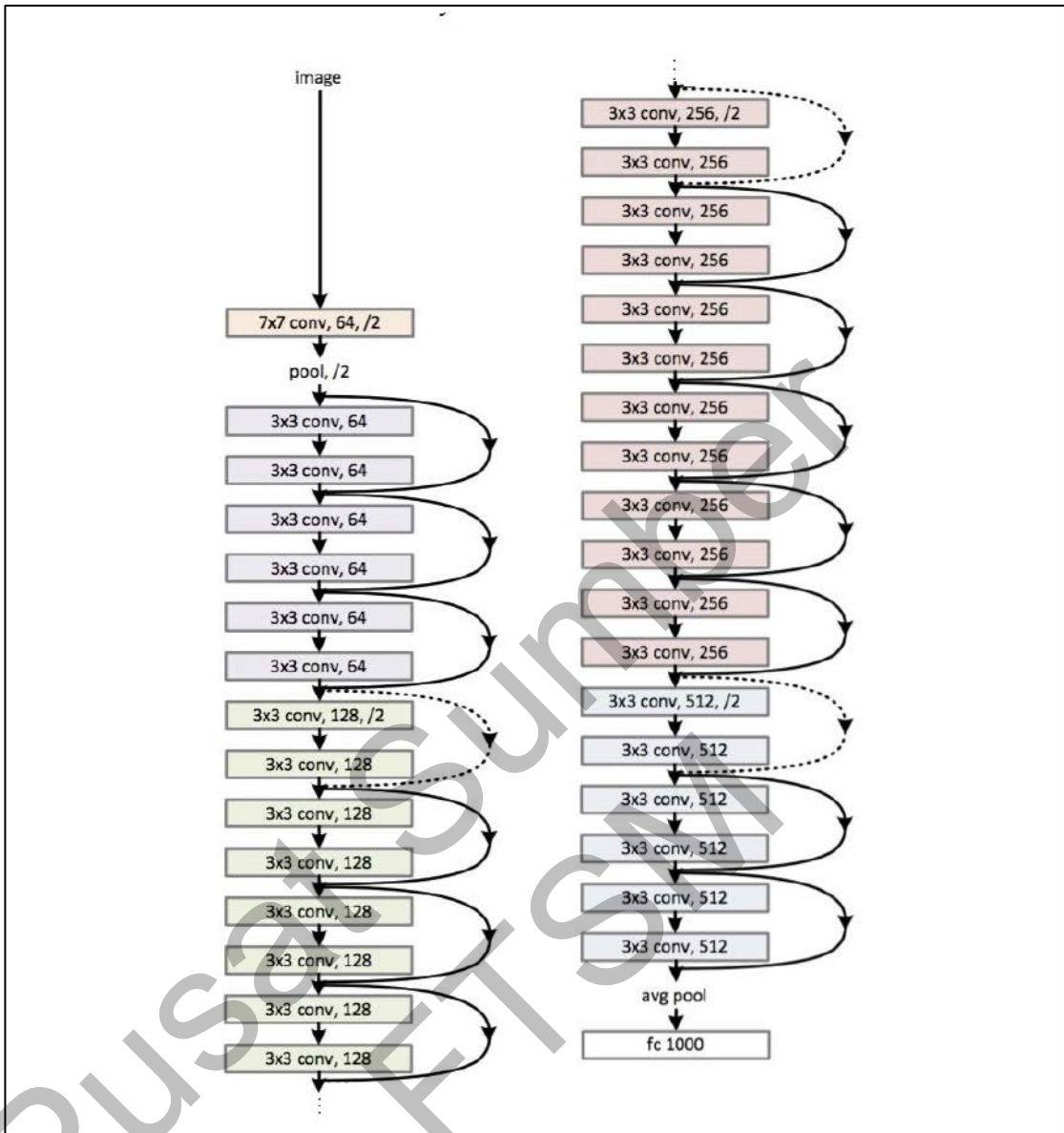
Bahagian ini akan menyentuh tentang rangkaian neural konvolusi VGG. Seni bina ini merupakan hasil daripada kajian Simonyan & Zisserman (2015). Terdapat dua jenis VGG yang telah dihasilkan dalam kertas penyelidikan ini iaitu VGG-16 dan VGG-19. Kedua-dua jenis seni bina ini mempunyai 16 lapisan konvolusi dan 19 lapisan konvolusi

pada lapisan tersembunyi. Latihan model dan pengujian model telah dijalankan dengan menggunakan PASCAL VOC 2007 (Everingham et al. t.th.-a) dan VOC 2012 (Everingham et al. t.th.-b). Dalam kajian ini, ketepatan purata min (mAP) telah digunakan untuk mengukur ketepatan model. Model VGG-16 yang terlatih dengan VOC 2007 mencapai nilai mAP sebanyak 89.3%. Nilai ini adalah sama bagi VGG-19.

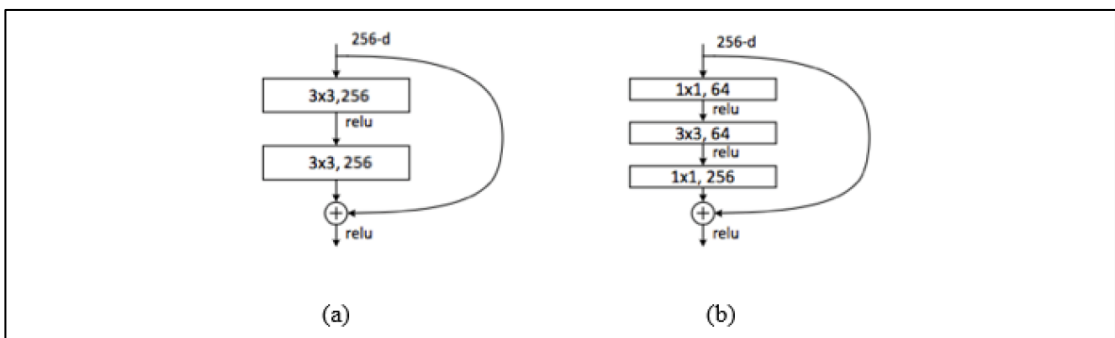
Rajah 2.8 menunjukkan seni bina VGG-16 yang mengandungi 16 lapisan konvolusi. Di dalam VGG-16, setiap blok konvolusi diaktifkan menggunakan ReLU. Seperti yang telah dinyatakan, ReLU digunakan untuk mengelakkan perubahan pada titik pertengahan. Blok konvolusi akan menjalani proses konvolusi dengan menggunakan saiz kernel 3x3. Namun, saiz pembalut (padding) adalah sama dan tidak akan diubah semasa proses konvolusi dijalankan. Selain itu, setiap lapisan pengumpulan maksimum akan menggunakan saiz pengumpulan 2x2 dengan 2x2 langkah. Lapisan pensampelan ke bawah yang pertama mengandungi dua lapisan konvolusi. Selepas itu, imej yang dimasukkan akan menjalani proses pengumpulan maksimum sebelum imej tersebut dihantar ke lapisan konvolusi yang seterusnya.

Lapisan pensampelan ke bawah yang kedua juga mengandungi dua lapisan konvolusi dan jumlah ciri-ciri yang diekstrak daripada imej akan ditingkatkan kepada 128. Imej yang telah menjalani lapisan kedua akan dihantar ke lapisan pengumpulan maksimum.

Dalam lapisan pensampelan ke bawah yang ketiga ini, sebanyak 256 ciri-ciri akan diekstrak. Proses konvolusi akan dijalani selama 3 lapisan. Proses yang mengandungi satu lapisan pensampelan maksimum dan tiga lapisan konvolusi ini akan diulangi dalam lapisan pensampelan ke bawah yang keempat dan kelima. Selepas itu, proses pensampelan maksimum juga akan dijalankan sebelum lapisan persambungan sepenuhnya dijalankan. Lapisan ini mengandungi ReLU sebagai modul pengaktifan. Lapisan persambungan sepenuhnya ini mempunyai 4096 dan akan berkurang kepada 1000 sebelum pengaktifan soft-max dilaksanakan untuk menentu klasifikasi objek.



Rajah 2.9 Seni bina ResNet-34



Rajah 2.10 Reka bentuk blok konvolusi dalam rangkaian residual

Konvolusi ResNet dibina berdasarkan rangkaian Visual Geometry Group (VGG). Antara rangkaian yang paling sering digunakan dalam penyelidikan adalah VGG-16 yang mengandungi 16 lapisan rangkaian neural konvolusi. Dalam blok konvolusi ResNet, proses normalisasi berkumpulan (*batch normalization*) akan diimplementasi selepas setiap lapisan konvolusi dan sebelum fungsi pengaktifan. Berbanding dengan seni bina pembelajaran mesin tradisional yang menggunakan *sigmoid* atau fungsi regresi logistik, ResNet menggunakan ReLU sebagai modul pengaktifan. Hal ini disebabkan modul pengaktifan yang digunakan dalam pembelajaran mesin biasa adalah sensitif kepada perubahan, terutamanya pada titik pertengahan.

Selain itu, menurut Heaton (2018), kecerunan (*gradient*) dan nilai-nilai pada *sigmoid* dan *tanh* sering kali tepu (*saturated*). Apabila nilai besar dan nilai kecil yang dimasukkan ke fungsi pengaktifan yang menggunakan *tanh* atau *sigmoid*, kecerunan dan nilai-nilai dalam fungsi pengaktifan boleh dijejaskan. Sebagai contoh, nilai yang besar akan meningkatkan nilai kepada 1 dan nilai yang kecil akan menurunkan nilai kepada 0 atau -1. Oleh itu, ReLU telah diimplementasi dan digunakan dalam pelbagai seni bina ResNet. Penggunaan ReLU telah mengatasi masalah kecerunan tepu yang dihadapi oleh fungsi pengaktifan yang lain.

Menurut kajian Krizhevsky et al. (2017), penggunaan ReLU telah meningkatkan kadar penumpuan *stochastic gradient descent* (SGD) berbanding dengan *sigmoid* dan *tanh*. Namun, ReLU masih menghadapi masalah seperti *dying ReLU*, di mana neuron ReLU tertentu mati jikalau neuron tersebut mempunyai nilai negatif dan memberi nilai pada kadar 0.

Bagi menyelesaikan masalah ini, penggunaan Leaky ReLU dapat mengatasi masalah ini. Masalah ini dihadapi dengan menambahkan parameter ke algoritma ReLU. Leaky ReLU menggunakan $y = 0.01$ bagi mengelakkan kematian. Namun, Leaky ReLU tidak semestinya adalah pilihan yang lebih baik daripada ReLU, dan ia hanya merupakan alternatif kepada ReLU.

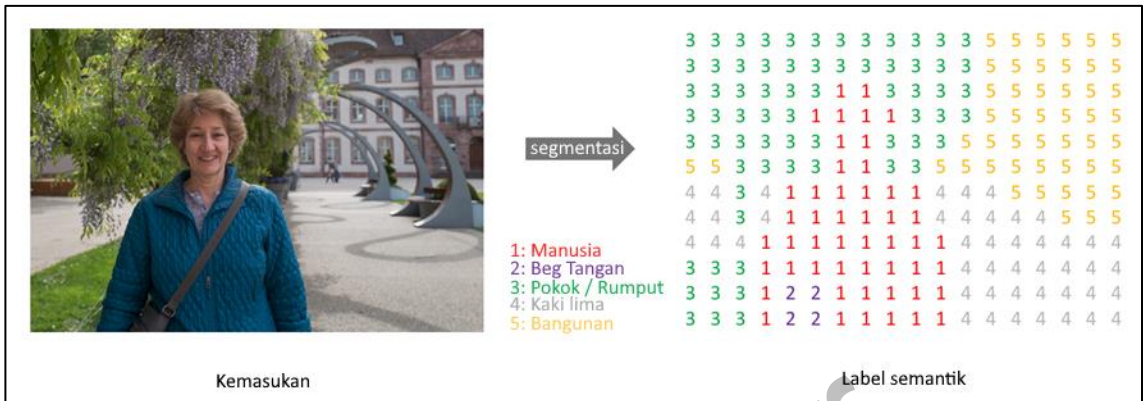
2.6 SEGMENTASI SEMANTIK

Tujuan imej semantik segmentasi adalah untuk menentukan setiap piksel di dalam imej dan mengklasifikasikan kelas yang sesuai dengan apa yang diwakili. Disebabkan teknik ramalan yang digunakan melibatkan setiap piksel di dalam imej, teknik ini juga dikenali sebagai ramalan kepadatan (*dense prediction*). Rajah 2.11 menunjukkan pengeluaran imej selepas diproses dengan menggunakan teknik semantik segmentasi.



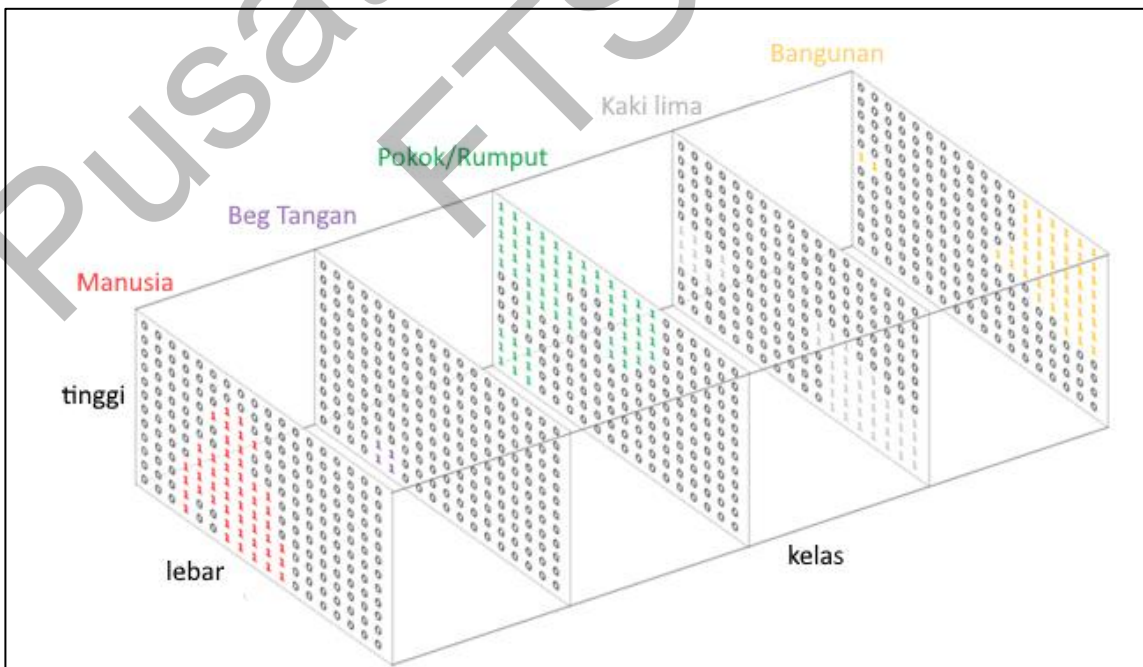
Rajah 2.11 Contoh pengeluaran imej selepas proses segmentasi semantik

Untuk penerangan yang lebih terperinci, imej yang berasal dari imej yang berwarna RGB iaitu yang bersaiz (tinggi x lebar x 3) ataupun imej yang berskala kelabu iaitu yang bersaiz (tinggi x lebar x 1) akan diproses dan akan mengeluarkan peta segmentasi yang mana setiap piksel mengandungi kelas yang diwakili sebagai nombor digit iaitu (tinggi x lebar x 1). Proses ini boleh digambarkan seperti Rajah 2.12.



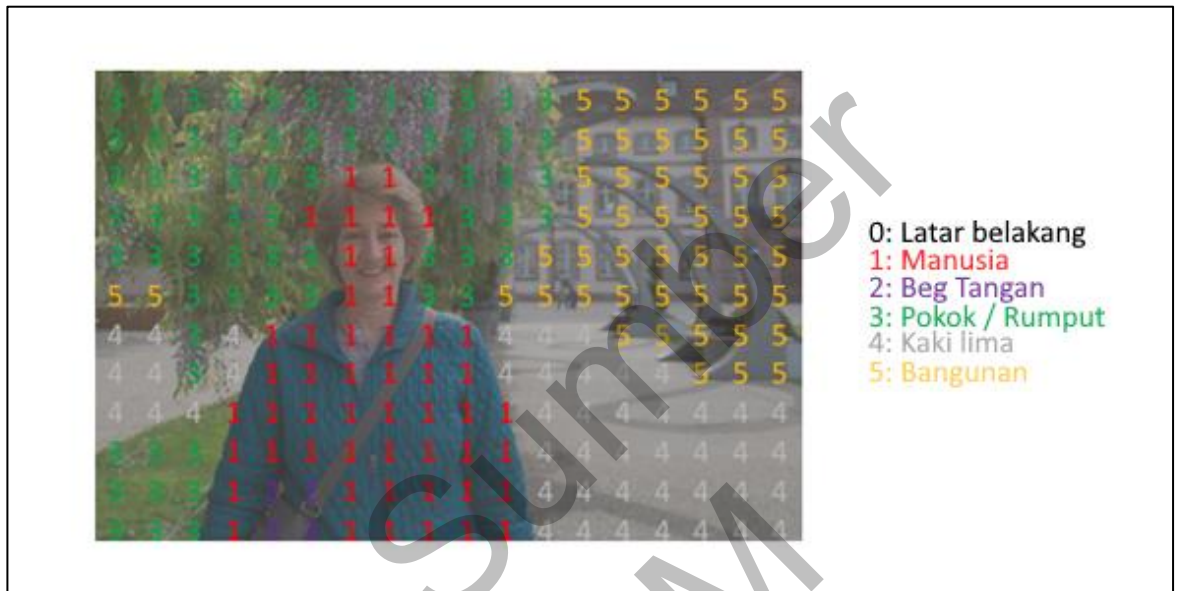
Rajah 2.12 Proses pengeluaian segmentasi daripada imej berwarna RGB kepada imej yang mewakili digit untuk menggambarkan kelas

Konsep yang sama digunakan oleh proses kategori klasifikasi biasa, iaitu teknik sasaran di terjemahkan kepada *one-hot encoder* label kelas. Pada dasarnya setiap kelas akan membuatkan satu saluran pengeluaian seperti yang digambarkan dalam Rajah 2.13. Ramalan ini kemudian dapat disusun ke dalam peta segmentasi seperti yang ditunjukkan pada Rajah 2.13 dengan mengambil kira nilai tertinggi untuk setiap vektor piksel di dalam pengeluaian. Pemeriksaan sasaran adalah lebih mudah dengan meletakkannya pemerhatian pada setiap lapisan.



Rajah 2.13 Lapisan *one-hot encoder* yang mewakili digit kelas masing-masing

Setiap lapisan itu kemudiannya di satukan ke dalam satu lapisan pengeluaran yang lebih dikenali sebagai topeng (masking) yang mana setiap piksel mewakili kelas masing-masing. Rajah 2.14 menunjukkan contoh topeng imej yang melalui proses segmentasi semantik.



Rajah 2.14 Penyatuan imej segmentasi ke dalam satu topeng (*masking*)

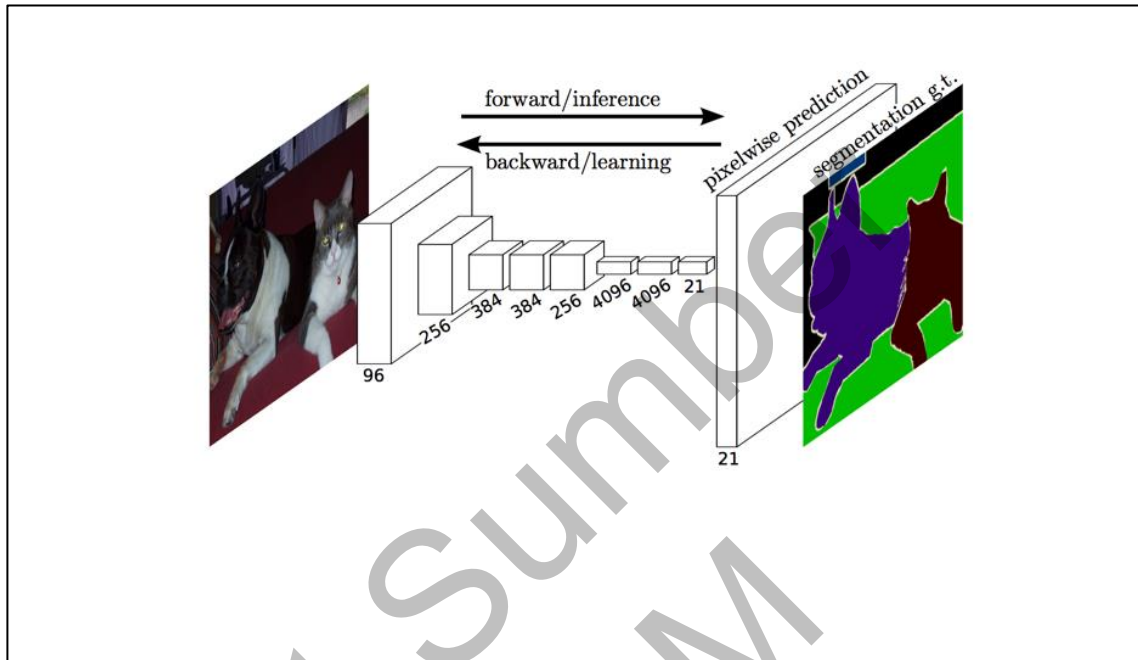
Terdapat beberapa seni bina yang sedia ada yang biasa digunakan untuk pelaksanaan segmentasi semantik. Kajian ini akan U-Net berdasarkan kajian yang telah dijalankan oleh Ronneberger et al. (2015).

2.6.1 Fully Convolutional Network (FCN)

Pendekatan menggunakan rangkaian konvolusi penuh untuk setiap piksel bagi tugas segmentasi imej telah diperkenalkan oleh Long et al. (2014). Penyelidik mencadangkan untuk mengadaptasi rangkaian klasifikasi imej seperti AlexNet (Krizhevsky et al. 2017) yang berfungsi sebagai modul pengkod (*encoder*) dan ditambah dengan modul penyahkod (*decoder*) dengan lapisan konvolusi transposisi untuk memaparkan peta ciri-ciri kasar kepada resolusi penuh segmentasi.

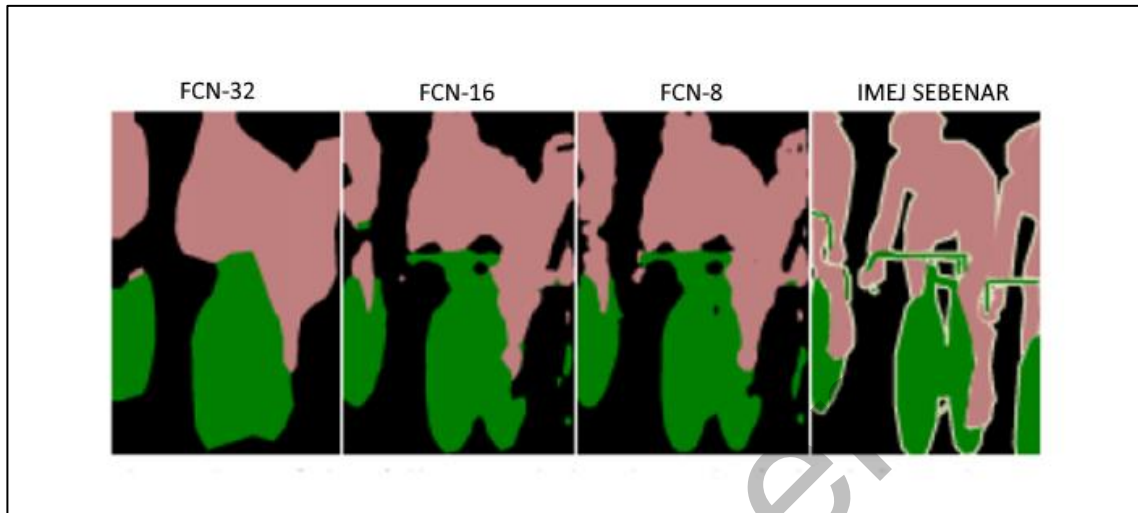
Pengekod merupakan proses permulaan dalam FCN yang mana ia berfungsi untuk menjalankan proses pensampelan ke bawah (*down sampling*) bagi tujuan untuk mengekstrak ciri-ciri dalam imej. Terdapat beberapa lapisan yang digunakan pada

proses pensampelan ke bawah dan ianya bergantung kepada kesesuaian domain. Proses terakhir dalam FCN pula merupakan pensampelan ke atas (*up sampling*) yang mana ia hanya dijalankan sekali sahaja. Rajah 2.15 menunjukkan bagaimana seni bina FCN.



Rajah 2.15 Seni bina FCN

FCN mempunyai tiga jenis lapisan ramalan terakhir iaitu FCN-32, FCN-16, dan FCN-8. FCN-32 mempunyai 32 piksel langkah (*stride*) pada lapisan ramalan terakhir manakala FCN-16 mempunyai 16 piksel langkah pada lapisan ramalan terakhir dan FCN-8 mempunyai 8 piksel langkah pada lapisan ramalan terakhir. Pada bahagian pengkod, CNN boleh ditukar mengikut keperluan sebagai contoh penggunaan VGG-16 (Simonyan & Zisserman 2015), ResNet (He et al. 2015) dan AlexNet (Krizhevsky et al. 2017). Rajah 2.16 menunjukkan segmentasi semantik yang diramalkan oleh FCN-32, FCN-16 dan FCN-8 serta imej sebenar yang digunakan untuk pengujian. Menurut rajah tersebut, FCN-8 menghasilkan segmentasi yang paling tepat manakala FCN-32 menghasilkan segmentasi yang paling kurang tepat.

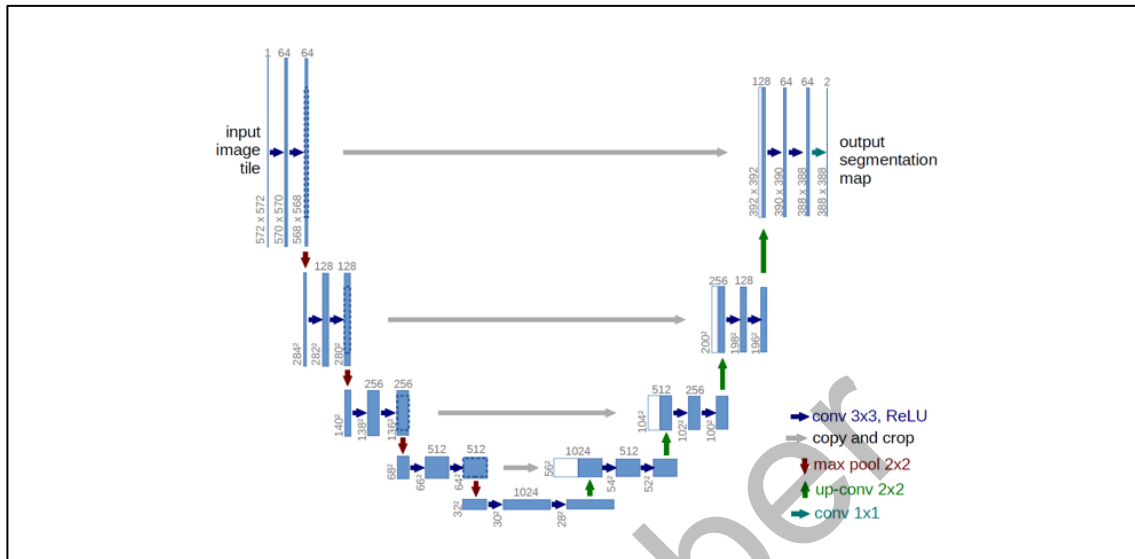


Rajah 2.16 Segmentasi semantik yang diramalkan oleh FCN-32, FCN-16 dan FCN-8 serta imej sebenar yang digunakan untuk pengujian

Menurut kajian Long et al. (2014), penyelidik menyatakan bahawa proses penalaan halus model klasifikasi objek dalam imej kepada segmentasi semantik telah memberikan keputusan dan ketepatan yang berpatutan.

2.6.2 U-Net

Ronneberger et al. (2015) telah menambahbaik seni bina FCN terutamanya dengan mengembangkan kapasiti rangkaian modul pada bahagian penyahkod. Mereka memperkenalkan seni bina baru dengan nama U-Net yang terdiri daripada lapisan simetrik pada bahagian pengekod dan penyahkod. Seni bina ini lebih mudah dan telah diguna pakai di dalam banyak kes penggunaan dan telah menyelesaikan pelbagai masalah segmentasi. Rajah 2.17 menunjukkan seni bina U-Net.



Rajah 2.17 Seni bina U-Net

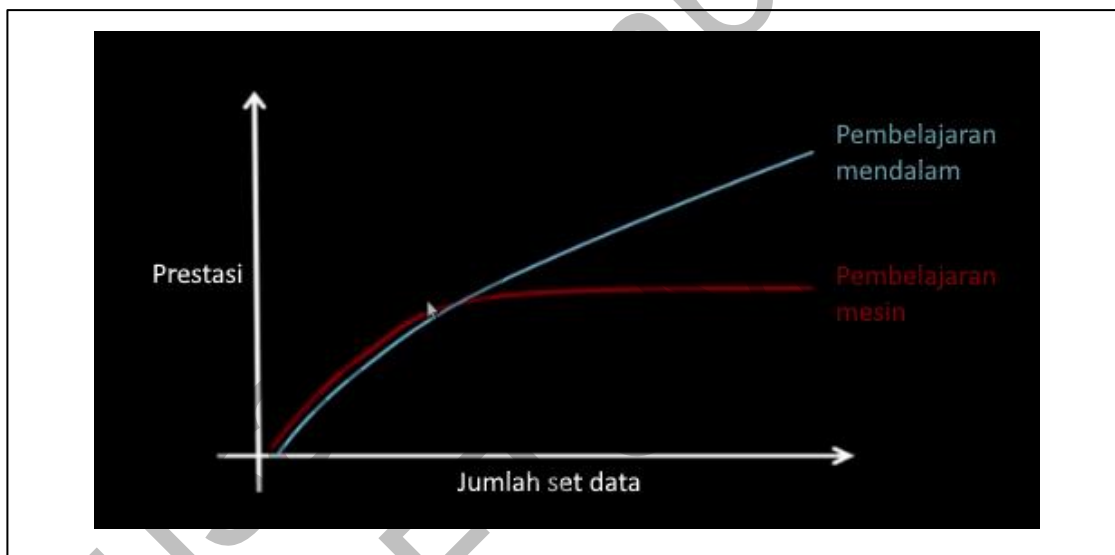
U-Net dibina berdasarkan rangkaian FCN. Seperti FCN, U-Net juga mengandungi proses pensampelan ke bawah dan pensampelan ke atas. Antara rangkaian yang sering diguna pakai dalam U-Net adalah VGG-16 (Simonyan & Zisserman 2015) dan ResNet-50 (He et al. 2015). Seni bina asal U-Net terdiri daripada 23 lapisan konvolusi rangkaian neural yang mengandungi proses pensampelan ke bawah dan pensampelan ke atas. Proses-proses ini bertujuan untuk meningkatkan jumlah pengekstrakan ciri-ciri imej pada setiap lapisan. U-Net bermula dengan menggunakan dua konvolusi 3x3 tanpa pembalut (*padding*), dan setiap lapisan konvolusi menggunakan pengaktifan ReLU. Setiap lapisan konvolusi juga menggunakan 2x2 penggabungan maksimum (*max pooling*) dengan 2 langkah (*stride*) dalam proses pensampelan ke bawah. Selepas proses pensampelan ke bawah selesai, U-Net meneruskan proses pensampelan ke atas dengan menggunakan 2x2 konvolusi dan diikuti dua lapisan 3x3 konvolusi dengan pengaktifan ReLU. Pada lapisan terakhir, 1x1 konvolusi digunakan untuk menyambungkan setiap 64 vektor komponen kepada jumlah kelas objek dalam imej.

Kajian FCN dari Long et al. (2014) melaporkan bahawa peningkatan augmentasi data tidak menghasilkan peningkatan prestasi yang ketara. Manakala di dalam laporan kajian U-Net (Ronneberger et al. 2015) pula menerangkan sebaliknya, yang mana peningkatan augmentasi data menyumbang dan menganggap sebagai konsep penting dalam latihan pembelajaran dalaman. Ini menunjukkan bahawa

penggunaan peningkatan data boleh meningkatkan prestasi namun ia bergantung kepada permasalahan pada setiap domain dan kes penggunaan masing-masing.

2.7 KAEDAH IMEJ AUGMENTASI

Pembelajaran mendalam kebiasaannya akan berfungsi dengan baik jika ianya di bina dengan menggunakan sejumlah besar set data. Secara umumnya, semakin banyak set data yang digunakan maka semakin tinggi prestasi ketetapan model pembelajaran mendalam. Rajah 2.18 menunjukkan graf yang menggambarkan hubungan diantara prestasi pembelajaran mendalam dan jumlah set data yang digunakan.



Rajah 2.18 Graf hubungan prestasi pembelajaran mendalam dan jumlah set data

Untuk memperolehi set data yang mempunyai jumlah yang besar merupakan satu cabaran yang besar. Cabaran ini adalah cabaran yang biasa dihadapi oleh saintis data. Oleh kerana itu teknik imej augmentasi diperkenalkan.

Imej augmentasi adalah teknik mengubah set data yang sedia ada bagi menjana lebih banyak set data baru untuk tujuan proses latihan model. Dengan erti kata lain, ia adalah proses memperluaskan set data yang ada untuk digunakan dalam pembelajaran mendalam.

Kesemua gambar yang dihasilkan adalah berasal dari set data latihan itu sendiri, maka pengumpulan set data secara manual tidak diperlukan lagi. Kaedah ini dapat

meningkatkan jumlah sampel set data tanpa perlu mengumpulkan data-data lain. Label untuk semua gambar yang dihasilkan oleh teknik imej augmentasi adalah label yang sama dengan label gambar asal.

Berdasarkan kajian daripada Song et al. (2020), hasil eksperimen menunjukkan bahawa imej augmentasi memainkan peranan penting dalam meningkatkan ketepatan model pengecaman, namun begitu teknik-teknik berbeza akan menyumbang keputusan yang berbeza. Sebagai contoh dalam kajian itu, teknik pengaburan, penskalaan, dan putaran menyumbang perbezaan yang ketara.

Secara amnya, terdapat dua jenis teknik utama di dalam imej augmentasi iaitu teknik manipulasi gambar asas, dan teknik yang menggunakan pembelajaran mendalam (Shorten & Khoshgoftaar 2019). Di peringkat awal penggunaan imej augmentasi, teknik manipulasi gambar asas lebih banyak digunakan kerana teknik ini lebih mudah untuk di aplikasikan. Teknik transformasi imej asas yang biasa digunakan adalah seperti putaran gambar, peralihan gambar dan peningkatan ruang warna.

2.7.1 Teknik Manipulasi Gambar Asas

Terdapat beberapa jenis teknik manipulasi gambar asas yang sering digunakan untuk tujuan imej augmentasi iaitu putaran gambar, peralihan gambar, pembalikan gambar, gangguan gambar dan pengaburan gambar. Rajah 2.19 menunjukkan gambar di sebelah kiri merupakan gambar asal, dan gambar selebihnya dihasilkan dari teknik imej augmentasi.



Rajah 2.19 Gambar yang dihasilkan melalui proses imej augmentasi

a. Putaran Gambar

Teknik putaran gambar adalah teknik yang selalu digunakan dalam imej augmentasi. Walaupun teknik ini akan memutarakan imej asal kepada imej yang baru, maklumat pada gambar adalah tetap sama. Perbezaannya adalah pada sudut pandangan yang berbeza. Rajah 2.20 menunjukkan contoh teknik putaran gambar.



Rajah 2.20 Gambar hasil daripada teknik putaran gambar Peralihan Gambar

Teknik peralihan gambar adalah teknik pengubahan kedudukan objek di dalam gambar asal ke posisi yang lain. Teknik ini dapat menjana lebih banyak variasi set data. Ia merupakan transformasi geometri yang menetapkan kedudukan setiap objek di dalam gambar ke lokasi baru yang lain. Sebagai contoh objek asal yang berada pada

kedudukan x dan y akan beralih kepada posisi baru iaitu X dan Y dengan perubahan sebanyak dx dan dy .

$$(x,y) = (X,Y) \quad \dots(2.1)$$

di mana $X=x+dx$ dan $Y=y+dy$

Rajah 2.21 menunjukkan gambar hasil daripada teknik peralihan gambar.



Rajah 2.21 Gambar hasil daripada teknik peralihan gambar

b. Pembalikan Gambar

Pembalikan gambar ialah teknik membalikkan gambar berdasarkan paksi sama ada dari arah kiri ke kanan ataupun dari atas ke bawah dan sebaliknya. Rajah 2.22 menunjukkan contoh pembalikan gambar. Gambar di sebelah kiri adalah gambar asal dan dua gambar lain adalah gambar yang telah menggunakan teknik pembalikan gambar.



Rajah 2.22 Gambar hasil daripada teknik pembalikan gambar

c. Gangguan Gambar

Teknik imej augmentasi lain yang popular adalah teknik gangguan gambar yang mana gambar asal di tambah dengan gangguan. Teknik ini juga dapat menjadikan model kita lebih kuat terhadap perubahan dalam gambar. Rajah 2.23 menunjukkan teknik gangguan gambar



Rajah 2.23 Gambar hasil daripada teknik gangguan gambar

d. Pengaburan Gambar

Gambar sebenar yang akan digunakan semasa proses penggunaan akan berasal daripada pelbagai sumber dan pelbagai jenis keadaan. Oleh itu kualiti gambar yang diperolehi tidak akan sama seperti set data yang digunakan dalam proses latihan model. Beberapa gambar sebenar berkemungkinan dalam keadaan yang berkualiti tinggi dan semestinya terdapat juga yang berada dalam kualiti yang sangat rendah. Dalam senario seperti ini, teknik pengaburan gambar dapat digunakan untuk mengaburkan gambar asal. Kaedah ini akan menjadikan model pembelajaran mendalam lebih tepat dalam mengklasifikasikan objek berdasarkan kualiti gambar yang rendah. Rajah 2.24 menunjukkan gambar hasil daripada penggunaan teknik pengaburan gambar.



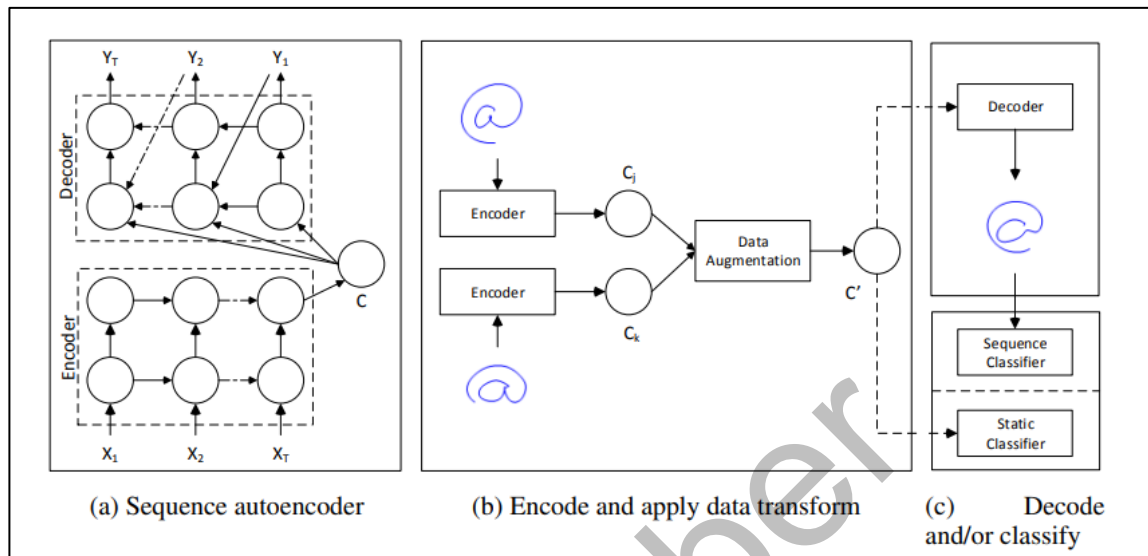
Rajah 2.24 Gambar hasil daripada teknik pengaburan gambar

2.7.2 Teknik Pembelajaran Mendalam

Teknik imej augmentasi yang menggunakan pembelajaran mendalam merupakan teknik yang menggunakan rangkaian konvolusi neural dalam menghasilkan imej augmentasi. Perbezaan yang ketara diantara teknik ini dengan teknik manipulasi gambar asas adalah tahap janaan imej yang boleh dilakukan pada lapisan tersembunyi sedangkan teknik manipulasi gambar hanya dilakukan pada layar asas imej. Terdapat banyak teknik pembelajaran mendalam yang telah diselidik dalam kajian-kajian sebelum ini. Kajian ini akan mengupas beberapa teknik yang telah digunakan. Walaubagaimanapun, teknik ini memerlukan proses pembinaan yang rumit berbanding dengan teknik manipulasi gambar asas. Di antaranya adalah proses latihan model yang perlu dijalankan terlebih dahulu sebelum imej augmentasi itu dapat disediakan.

a. Augmentasi Ciri-Ciri Ruang

DeVries & Taylor (2017) telah memperkenalkan imej augmentasi menggunakan teknik augmentasi ciri-ciri ruang. Teknik ini berfungsi dengan cara mempelajari data asas terlebih dahulu dan kemudian mengaplikasikannya dengan transformasi data sampel dan padanan pada data asas. Rajah 2.25 menunjukkan langkah-langkah pelaksanaan dalam kajian itu.



Rajah 2.25 Langkah-langkah pelaksanaan dalam kajian

Sumber: (DeVries & Taylor 2017)

Rajah 2.25 (a) merupakan langkah pertama dalam seni bina kajian ini. Ia merupakan pengkod secara automatik yang berurutan yang digunakan untuk mempelajari ciri-ciri ruang pada data yang belum dilabelkan, setiap urutan itu akan diwakili dengan konteks vektor C . Rajah 2.25 (b) pula menggabungkan data-data yang telah di enkod dan ditambahkan dengan data transformasi seperti gangguan imej. Rajah 2.25 (c) pula merupakan langkah yang terakhir di mana hasil daripada konteks vektor yang boleh digunakan secara langsung ataupun dinyahkod terlebih dahulu dan di klasifikasikan secara urutan. Kajian ini telah menunjukkan teknik ini boleh digunakan untuk meningkatkan prestasi pembelajaran mendalam.

b. Latihan Permusuhan (Adversarial Training)

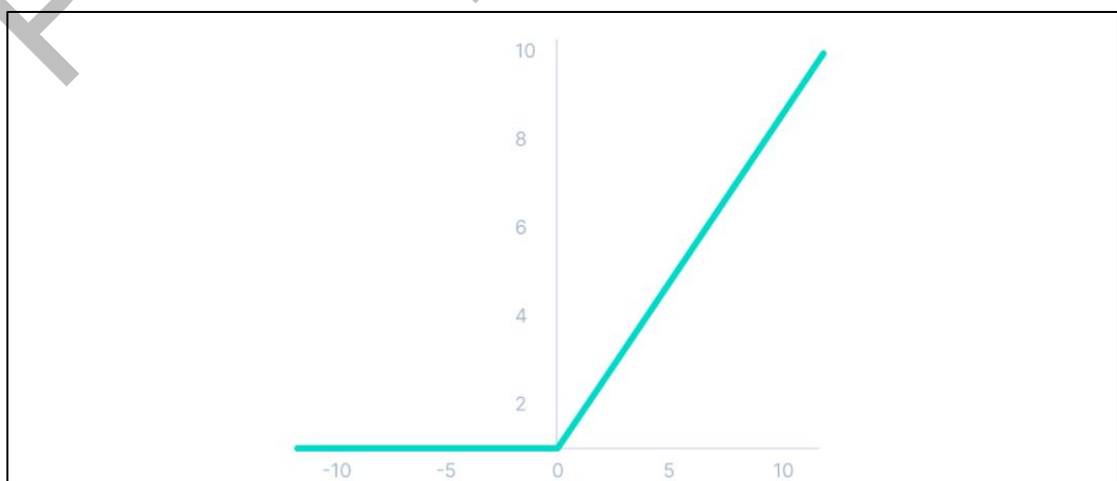
Latihan permusuhan merupakan seni bina yang menggunakan dua atau lebih rangkaian neural yang objektifnya merunglai perbezaan diantara rangkaian neural itu dengan cara menghitung fungsi kehilangan (*loss function*). Teknik permusuhan yang terdiri daripada beberapa rangkaian neural ini akan bersaing diantara satu sama lain dan dalam masa yang sama ia mempelajari ciri-ciri imej dalam tujuan meningkatkan kualiti imej augmentasi berdasarkan kesalahan klasifikasi dalam rangkaian neural saingannya. Dalam kajian Goodfellow et al. (2014), penyelidik telah menunjukkan beberapa contoh penggunaan latihan permusuhan yang dapat meningkatkan prestasi dalam tugas klasifikasi set data MNIST.

2.8 FUNGSI PENGAKTIFAN

Fungsi pengaktifan adalah untuk menentukan bagaimana jumlah keseluruhan pemberat kemasukan akan menjadikan jumlah pengeluaran di antara nod-nod di dalam rangkaian neural. Terdapat beberapa jenis fungsi pengaktifan dalam pemrosesan imej sebagai contoh yang terkenal adalah Rectified Linear Unit (ReLU) dan Leaky ReLU. Berdasarkan kajian yang dilakukan oleh Szandala (2020), penyelidik telah membuat perbandingan diantara fungsi pengaktifan yang sering digunakan di dalam pembelajaran mendalam. Szandala (2020) telah menjalankan ujian dengan menggunakan set data CIFAR-10 yang terdiri daripada 60000 gambar berwarna dan bersaiz 32 x 32 piksel yang dibahagikan kepada 10 jenis kelas. Hasil daripada ujian menunjukkan bahawa Leaky ReLU memperolehi hasil ketetapan yang paling tinggi iaitu sebanyak 73% dan diikuti dengan ReLU sebanyak 72% dan Swish sebanyak 70%. Kajian ini juga membuat perbandingan diantara ketiga-tiga jenis fungsi pengaktifan dengan menggunakan set data bot dan kapal. Hasil daripada perbandingan ini di terangkan pada Bab IV.

2.8.1 *Rectified Linear Unit (ReLU)*

Fungsi pengaktifan ReLU memberikan fungsi linear bagi output yang lebih daripada kosong manakala akan menyahaktifkan jika output adalah kurang daripada kosong. Ini digambarkan di dalam Rajah 2.26.



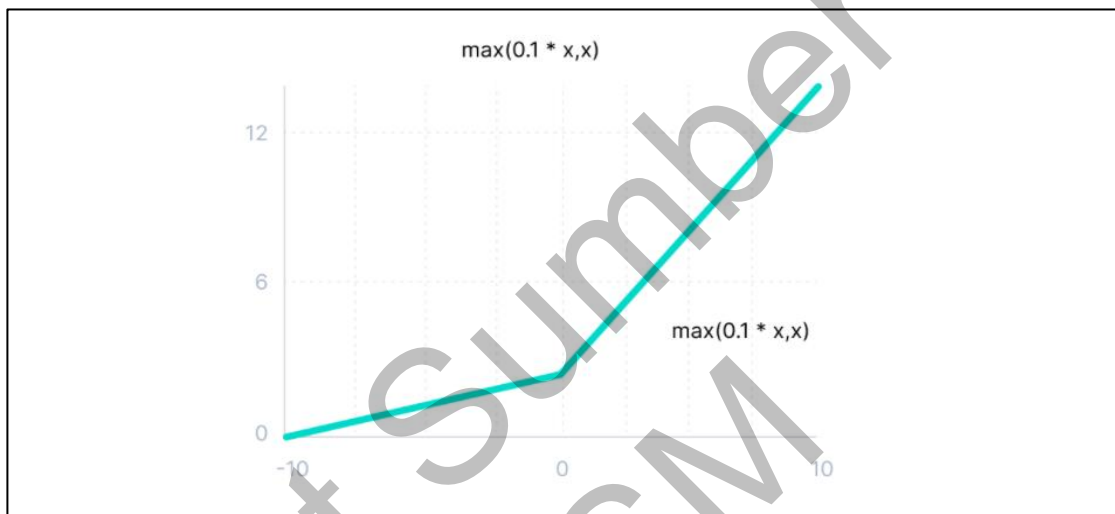
Rajah 2.26 Graf fungsi pengaktifan ReLU

Secara matematik ia dapat diistilahkan sebagai formula

$$f(x) = \max(0, x) \quad \dots(2.2)$$

2.8.2 Leaky ReLU

Leaky ReLU merupakan versi penambah baikkan daripada versi ReLU yang mana ia mempunyai cerun positif kecil di kawasan negatif. Rajah 2.27 menunjukkan graf fungsi pengaktifan Leaky ReLU.



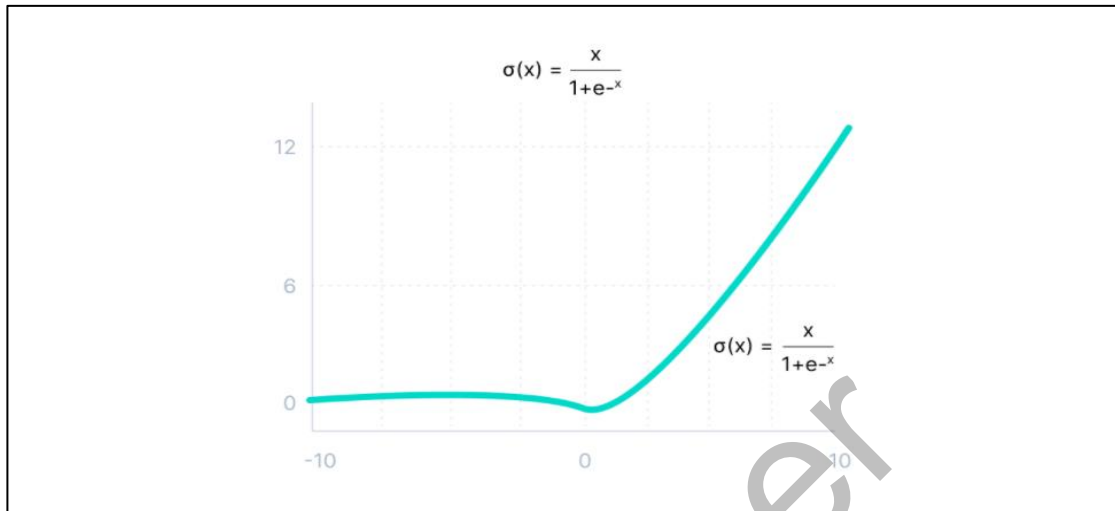
Rajah 2.27 Graf fungsi pengaktifan Leaky ReLU

Secara matematik ia dapat diistilahkan sebagai formula

$$f(x) = \max(0.1x, x) \quad \dots(2.3)$$

2.8.3 Swish

Swish mempunyai lengkungan sedikit ke bawah pada nilai kosong dan tidak terikat pada nilai positif. Rajah 2.28 menunjukkan graf fungsi pengaktifan Swish.



Rajah 2.28 Graf fungsi pengaktifan Swish

Secara matematik ia dapat diistilahkan sebagai formula

$$f(x) = x * \text{sigmoid}(x) \quad \dots(2.4)$$

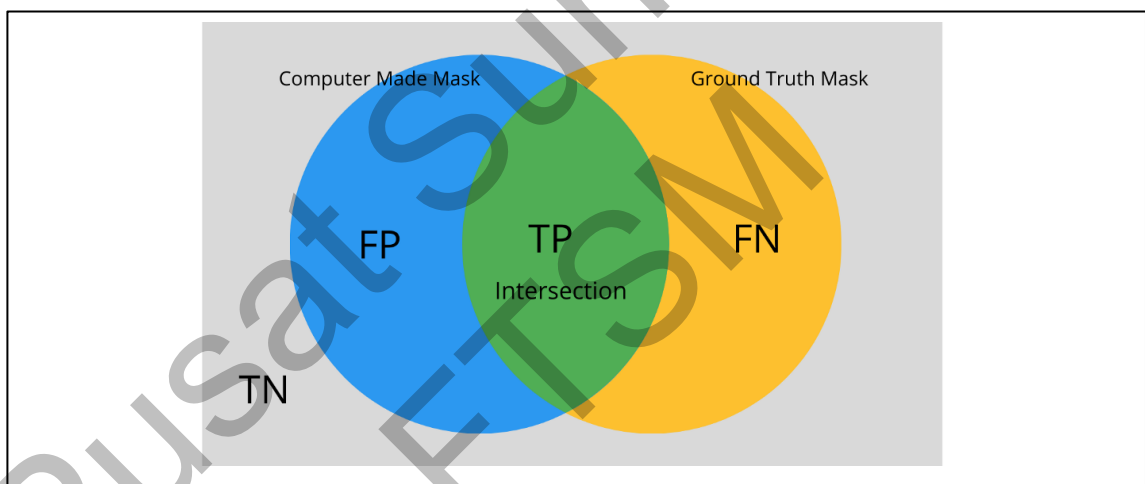
2.9 SUKATAN PENGUJIAN KETEPATAN

Proses latihan dalam pembelajaran mendalam akan menghasilkan sebuah model. Setiap model yang terbina perlu diukur ketepatannya bagi menentukan tanda aras prestasi ramalan. Terdapat beberapa kaedah untuk menentukan ketepatan model yang telah dilatih. Di antara kaedah penilaian yang sering digunakan dalam segmentasi imej adalah jenis *Jaccard's Index (Intersection of Union, IoU)*, dan *Dice Coefficient*. Sumber set data dari Cityscapes (Cordts et al. 2016) menyediakan gambar-gambar yang berasaskan objek umum di kawasan bandar seperti jalan, bangunan, dan lampu isyarat. Sebanyak 25000 set data yang telah di anotasi untuk 30 jenis kelas disediakan. Set data ini banyak digunakan dalam kajian segmentasi pembelajaran mendalam terutamanya untuk kes penggunaan kereta autonomi. Di dalam laman sesawang (“Benchmark Suite – Cityscapes Dataset” t.th.) Cityscapes itu juga memaparkan senarai kedudukan peserta pertandingan yang membina seni bina pembelajaran mendalam imej segmentasi. Penanda aras IoU telah digunakan sebagai unit pengukur tahap ketetapan model yang dibina oleh peserta. Oleh kerana itu, penyelidik telah memilih kaedah penilaian IoU untuk mengukur penanda aras ketetapan model untuk kajian ini. Selain daripada IoU, kaedah *Dice Coefficient* iaitu *F1 Score* juga selalu digunakan untuk tujuan pengukuran ketetapan model imej segmentasi. Dalam kajian Fan et al. (2019), penyelidik

menggunakan *Dice Coefficient* bagi membandingkan prestasi ketepatan diantara beberapa model pembelajaran mendalam. Secara amnya kedua-dua penanda aras akan memberi keputusan yang hampir sama dan kedua-duanya akan memberikan keputusan 0 apabila tiada titik pertindihan dan 1 apabila terdapat titik pertindihan yang sempurna bagi imej sebenar dan imej ramalan. Perbezaan hanyalah pada cara pengiraan ketepatan yang akan di bincangkan pada sub topik seterusnya. Kajian ini menggunakan kedua-dua jenis kaedah bagi menentukan prestasi ketepatan latihan model yang dibangunkan.

2.9.1 Intersection Of Union (IoU)

IoU membandingkan persamaan *union* di antara imej asal (*ground truth*) mask dan imej ramalan (*prediction mask*). Teknik ini dapat digambarkan di dalam Rajah 2.29.



Rajah 2.29 Intersection of Union

Berdasarkan Rajah 2.29, kawasan berwarna hijau merupakan ramalan yang tepat iaitu *True Positive*. Kawasan yang berwarna biru merupakan gambaran *False Positive*. Kawasan yang berwarna kuning merupakan ramalan *False Negative* dan kawasan yang berwarna kelabu merupakan ramalan *True Negative*.

$$Jaccard's\ Index = \frac{Intersection}{Union} = \frac{TP}{TP+FN+FP} \quad \dots(2.5)$$

2.9.2 Dice Coefficient

Dice Coefficient mempunyai persamaan yang dengan Indeks *Jaccard*. Perbezaan hanya pada pengiraan *True Positif* yang di kali dua.

$$Dice\ Coefficient = \frac{2 \times Intersection}{Union + Intersection} = \frac{2TP}{2TP + FN + FP} \quad \dots(2.6)$$

2.10 PENGESANAN BOT, KAPAL DAN OBJEK BERSAIZ KECIL

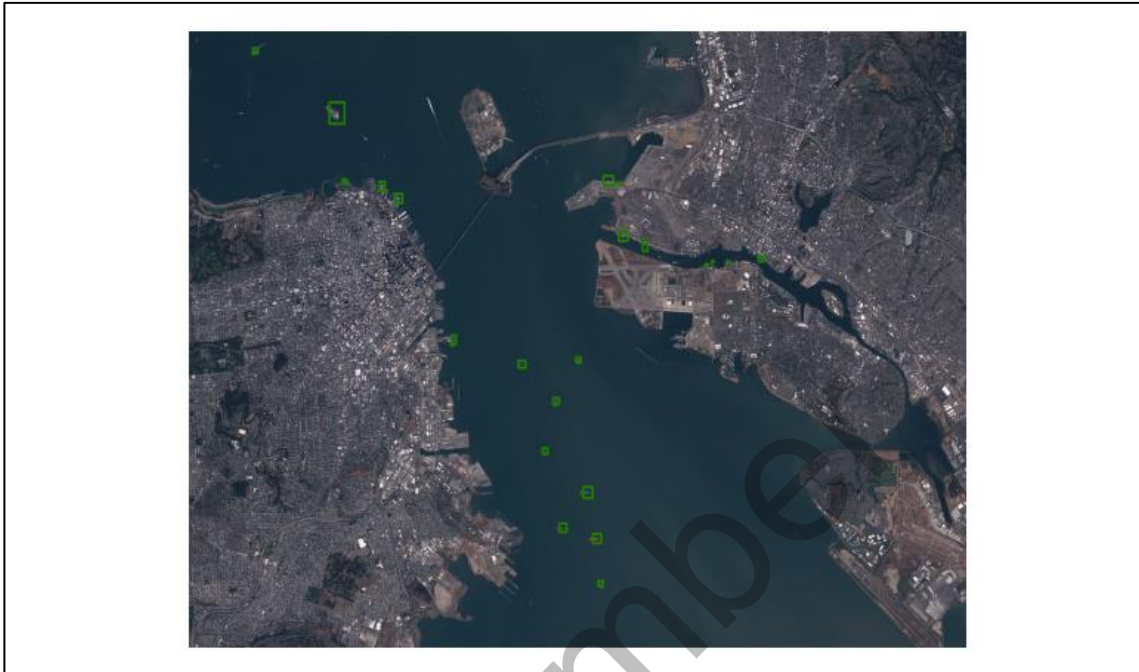
Berdasarkan bacaan kajian kesusasteraan yang dijalankan terhadap kajian-kajian terdahulu, keperluan untuk menentukan objek bot dan kapal adalah sangat penting dalam operasi pengawasan maritim (Cane & Ferryman 2018). Kajian Cane & Ferryman (2018) menyatakan terdapat beberapa cabaran dalam mengadaptasikan pengesanan objek yang diantaranya adalah kerana sifat persekitaran yang tidak terkawal dan keperluan beroperasi dalam masa nyata. Kajian itu telah menggunakan teknik segmentasi semantik iaitu SegNet, ENet, dan ESPNet, dan menggunakan set data daripada ADE20k yang merangkumi gambar-gambar dari pelbagai sudut pandangan mata. Hasil kajian menunjukkan bahawa teknik segmentasi semantik mampu untuk mengesan objek dari pelbagai sudut. Seni bina SeNet dan ENet telah mencapai prestasi ketetapan pengesanan yang lebih tinggi, manakala seni bina ESPNet lebih baik dalam mengklasifikasikan objek dengan betul. Rajah 2.30 menunjukkan sampel set data yang digunakan dalam kajian itu.



Rajah 2.30 Gambar set data daripada ADE20k

Sumber: Cane & Ferryman (2018)

Kajian Štepec et al. (2019) pula memfokuskan pada set data daripada imej satelit. Kajian itu menggunakan set data daripada *Kaggle Airbus Ship Detection* dan menggunakan seni bina Mask R-CNN sebagai seni bina utama dalam pembelajaran mendalam. Penyelidik membangunkan model pembelajaran mendalam dengan menggunakan informasi daripada data *Automatic Identification System (AIS)* sebagai sumber asas untuk pelabelan objek. Hasil daripada kajian mendapati bahawa penggunaan data AIS sebagai sumber anotasi dapat meningkatkan hasil ketetapan berbanding set data yang di anotasi oleh manusia. Rajah 2.31 menunjukkan contoh set data yang digunakan dalam kajian Štepec et al. (2019).



Rajah 2.31 Contoh set data daripada Kaggle Airbus Ship Detection

Sumber: (Štepec et al. 2019)

Kajian daripada Nalamati et al. (2020) pula menggunakan televisyen litar tertutup (CCTV) untuk pengesanan objek bot dan kapal. Kajian ini telah menggunakan beberapa teknik CNN seperti CenterNet dan EfficientDet sebagai seni bina pembelajaran mendalam. Penyelidik menggunakan beberapa set data daripada VAIS, MARVEL, MarDCT, SMD, SeaShips, SeaGull, Boats Re-ID untuk digunakan dalam latihan model. Hasil kajian menunjukkan model CenterNet-DLA dan ResDCN18 adalah pilihan yang terbaik untuk digunakan dalam sistem pengesanan objek masa nyata. Rajah 2.32 menunjukkan contoh gambar bot dan kapal dari sudut pandangan CCTV.



Rajah 2.32 Contoh gambar bot dan kapal dari sudut pandangan CCTV

Sumber: Nalamati et al. (2020)

Di antara cabaran lain dalam pengesanan objek bot dan kapal ialah kerana kerumitan senario aplikasi dalam objek yang intensif dan objek yang dikesan secara berlebihan (Yang et al. 2018). Yang et al. (2018) mencadangkan dalam kajian mereka satu seni bina yang dinamakan sebagai *Rotation Dense Feature Pyramid Networks* (R-DFPN) yang dapat mengesan bot dan kapal dengan berkesan dari pemandangan yang berbeza termasuk lautan dan pelabuhan. Penyelidik membina peta ciri-ciri semantik peringkat tinggi untuk semua skala pada sambungan kepadatan. Selain daripada itu, penyelidik juga memperkenalkan ramalan segi empat tepat untuk setiap objek bot dan kapal bagi mengurangkan kawasan yang dikesan secara berlebihan. Rajah 2.32 menunjukkan imej pengeluaran daripada teknik R-DFPN yang dijalankan.



Rajah 2.33 Gambar sebelah kiri menunjukkan masalah pengesanan berlebihan dan sebelah kanan pengeluaran output menggunakan teknik R-DFPN

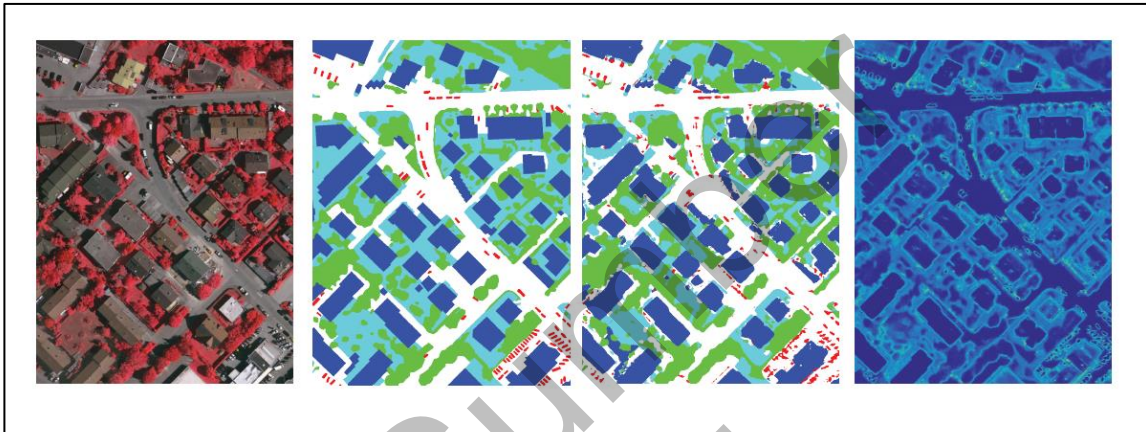
Sumber: Yang et al. (2018)

Di dalam kajian Tang et al. (2021) pula, mereka memperkenalkan novel berasaskan rangkaian gambar beresolusi tinggi bagi mengesan kawasan yang berkepentingan. Kajian ini mengambil pendekatan menggunakan imej berasaskan *hue, saturation, value* (HSV) untuk meningkatkan kecekapan dalam membezakan antara kapal dan latar belakang. Kajian ini menggunakan set data kapal yang didapati daripada Google Earth dan HRSC2016. Eksperimen ini dijalankan dengan menggunakan seni bina YOLO dan penyelidik mendapati bahawa penggunaan saluran HSV telah meningkatkan tahap pengesanan sebanyak 19.01% dan peningkatan ketepatan sebanyak 16.19%.

Terdapat banyak kes penggera palsu (*false alarm*) yang diterima disebabkan oleh keadaan sekeliling lautan dan kekurangan informasi tentang kapal seperti yang dilaporkan dari sistem pemerhatian *compact polarimetric synthetic aperture radar* (CP SAR) (Fan et al. 2019). Dalam kajian Fan et al. (2019), penyelidik menggunakan teknik segmentasi yang berasaskan piksel iaitu U-Net sebagai seni bina untuk mengesan kapal. Sebanyak tiga jenis seni bina di gunakan untuk tujuan perbandingan iaitu Standard CFAR, Faster RCNN dan U-Net. Seni bina U-Net memperolehi prestasi yang tertinggi dengan ketepatan dan *recall* melebihi 90%.

Dalam kajian oleh Kampffmeyer et al. (2016) yang memfokuskan pada pengesanan objek kecil menggunakan imej *remote sensing* mendapati bahawa prestasi

ketepatan klasifikasi secara menyeluruh adalah sebanyak 86% dengan menggunakan teknik segmentasi semantik. Kajian itu menggunakan set data daripada ISPRS Vaihingen 2D yang terdiri daripada 33 imej beresolusi tinggi dan pelbagai saiz. Setiap imej mempunyai diantara 3 juta hingga 10 juta piksel setiap satu. Rajah 2.34 menunjukkan gambar set data yang digunakan di dalam kajian itu.



Rajah 2.34 Gambar set data daripada set data daripada ISPRS Vaihingen 2D dan hasil kajian

Sumber: Kampffmeyer et al. (2016)

Kampffmeyer et al. (2016) menggunakan pendekatan menggabungkan beberapa model untuk meningkatkan ketepatan keseluruhan. Penyelidik menggabungkan kebarangkalian *softmax* daripada kaedah yang berbeza dengan melatih SVM linear. Kajian itu menyimpulkan bahawa penggabungan model memberikan prestasi ketepatan keseluruhan yang bagus untuk objek bersaiz kecil.

Emek & Demir (2020) menjalankan kajian berkenaan pengesanan objek bangunan daripada imej *remote sensing* SAR menggunakan seni bina pembelajaran mendalam U-Net. Penyelidik mengkaji keberkesanan penggunaan U-Net untuk mengesan objek bangunan daripada SAR dan gabungan imej optik. Kajian itu menggunakan set data daripada Sentinel 1 SAR dan Sentinel-2 imej pelbagai sudut. Berdasarkan kajian itu, prestasi ketepatan model yang dibina mencapai sehingga 81%.

Smith (2018) menggunakan set data satelit dalam kajiannya untuk mengesan objek kapal. Set data daripada satelit mempunyai masalah tersendiri seperti terdapatnya gangguan daripada awan, keadaan daratan yang tidak menentu dan gangguan daripada

objek buatan manusia yang lain (Smith 2018). Untuk menangani masalah itu, penyelidik membina seni-bina U-Net yang diubahsuai untuk meningkatkan ketepatan pengesanan objek Smith (2018) mengubahsuai dengan melatih model menggunakan dua jenis seni bina yang berbeza iaitu seni bina nod tunggal dan seni bina berasaskan pangkalan awan yang parameternya yang bertindak sebagai mekanisme peningkatan. Untuk seni bina nod tunggal, kajian ini berjaya mendapatkan 92% ketepatan ke atas set pengesanan dan untuk seni bina berasaskan pangkalan awan mendapatkan 86% ketepatan ke atas set pengesanan.

Pemprosesan objek bersaiz kecil yang berada di dalam imej yang beresolusi tinggi akan memberi masalah kepada pengesanan objek kerana objek yang kecil tidak mempunyai informasi yang terperinci di dalam pembelajaran mendalam (Liu et al. 2021). Kajian Liu et al. (2021) memperkenalkan HRD-Net yang menggunakan beberapa resolusi kemasukan ke dalam *multi-depth backbones*. Penyelidik mencadangkan penggunaan *Multi-Depth Image Pyramid Network* (MD-IPN) and *Multi-Scale Feature Pyramid Network* (MS-FPN) di dalam HRD-Net. Dalam kajian itu, mereka menggunakan set data MS COCO2017, Pascal VOC2007 dan 2012, dan VisDrone 2019 sebagai penanda aras untuk mengukur tahap prestasi. Kajian membuktikan bahawa dengan mengekalkan imej resolusi yang tinggi dapat meningkatkan prestasi pengesanan objek bersaiz kecil.

Berdasarkan kajian yang telah dijalankan sebelum ini, terdapat banyak kajian yang telah dijalankan untuk mengesan objek bot, kapal dan objek-objek lain yang bersaiz kecil. Ini menunjukkan kepentingan untuk mengesan objek kecil ini sangat penting. Terdapat pelbagai set data berkenaan dengan objek bot dan kapal yang berbeza dari sudut pandangan dan jarak. Setiap jenis set data mempunyai cabaran-cabaran yang tersendiri seperti gangguan awan, pantulan cahaya, putaran dan lain-lain. Teknik utama yang sering diguna pakai adalah teknik pengesanan objek, dan segmentasi imej. **Error! Reference source not found.** menunjukkan senarai kajian-kajian yang telah dijalankan untuk pengesanan bot dan objek kecil

Jadual 2.1 Senarai kajian-kajian yang telah dijalankan untuk pengesanan bot dan objek kecil

Bil	Penyelidik	Tajuk	Kaedah	Set Data	Hasil
1.	Cane & Ferryman (2018)	<i>Evaluating deep semantic segmentation networks for object detection in maritime surveillance</i>	Latihan dijalankan menggunakan set data ADE20k yang mengandungi 4 kelas utama iaitu, laut, langit, objek maritim, dan lain-lain. Dimensi imej akhir diselaraskan pada saiz 640x480 piksel. Augmentasi imej di gunakan secara rawak dengan menggunakan teknik potongan, putaran mendatar, putaran, dan gangguan kecerahan dan warna.	ADE20k (merangkumi pelbagai sudut pandangan)	SegNet dan ENet mencapai tahap prestasi ketepatan yang tertinggi untuk pengesanan objek, tetapi ESPNet adalah lebih lebih baik dalam mengklasifikasikan objek dengan betul.
2.	Štepec et al. (2019)	<i>Automated System for Ship Detection from Medium Resolution Satellite Optical Imagery</i>	Kajian ini menguji penggunaan proses anotasi gambar secara automatik yang dilakukan berdasarkan input daripada sistem <i>Automatic Identification System</i> (AIS) sebagai sumber anotasi objek. Penyelidik menggunakan teknik peningkatan skala dalam augmentasi imej yang mana setiap imej ditambah dengan satu lagi imej augmentasi yang dihasilkan dengan menggunakan teknik pembesaran sebanyak julat diantara 50% ke 70%. Penyelidik juga menggunakan augmentasi putaran di antara julat -45 hingga 45 dan <i>Gaussian blur</i> di antara julat 0 hingga 0.5.	Kaggle Airbus, ESA Sentinel-2, Planet Labs Dove	Penggunaan data AIS sebagai sumber anotasi dapat meningkatkan hasil ketetapan berbanding set data yang di anotasi oleh manusia. Secara keseluruhan, hasil daripada kajian menunjukkan prestasi pengesanan objek telah meningkat daripada julat diantara 41% - 54% kepada julat diantara 61% - 84%.
3.	Nalamati et al. (2020)	<i>Automated Monitoring in Maritime Video Surveillance System</i>	Kajian ini menjalankan eksperimen untuk menilai beberapa jenis seni bina pengesanan objek sedia ada seperti CenterNet dan EfficientDet. Sebanyak 27,368 imej telah di gunakan hasil daripada gabungan beberapa set data. Jumlah imej itu kemudian di pecahkan secara rawak kepada 21,906 untuk latihan, 2731 untuk pengesanan dan 2731 untuk ujian. Kajian ini tidak menjalankan augmentasi imej pada set data asal dan berpendapat dengan menggunakan set data campuran daripada sumber yang berbeza dapat mempelbagaikan lagi set data seperti contoh kepelbagaian dari sudut pandangan, cuaca dan jarak kamera.	VAIS, MARVEL, MarDCT, SMD, SeaShips, SeaGull, Boats Re-ID	Hasil ujian menunjukkan model CenterNet-DLA dan ResDCN18 adalah pilihan yang terbaik untuk digunakan dalam sistem pengesanan objek masa nyata. Seni bina CenterNet-DLA memberi keputusan 0.893 untuk prestasi dan kecepatan sebanyak 41.037 <i>frame per second</i> (FPS). Manakala ResDCN18 menunjukkan prestasi sebanyak 0.868 dan kecepatan 49.525 FPS.

bersambung...

...sambungan

- | | | | | | |
|----|---------------------------|---|--|--------------------|--|
| 4. | Yang et al. (2018) | <i>Automatic Ship Detection of Remote Sensing Images from Google Earth in Complex Scenes Based on Multi-Scale Rotation Dense Feature Pyramid Networks</i> | Kajian ini mencadangkan seni bina yang dikenali sebagai <i>Rotation Dense Feature Pyramid Networks (R-DFPN)</i> bagi mengesan kapal dengan lebih berkesan pada set penggunaan pemandangan yang berbeza seperti lautan dan pelabuhan. Set data untuk kajian ini diperolehi daripada <i>Google Earth</i> yang mana selepas pra-pemprosesan, jumlah imej akhir yang digunakan dalam latihan adalah sebanyak 8000 keping. Teknik pembalikan secara rawak, dan mengurangkan nilai min digunakan dalam imej augmentasi. | Google Earth | Kaedah yang dicadangkan iaitu R-DFPN menunjukkan ia dapat mengesan putaran objek daripada pelbagai skala dengan pemandangan yang berbeza. Hasil daripada kajian mendapati prestasi untuk R-DFPN memperolehi 88.2% untuk <i>recall</i> , 91% untuk ketepatan dan 89.6% untuk <i>F-Measure</i> . |
| 5. | Fan et al. (2019) | <i>Ship Detection Using a Fully Convolutional Network with Compact Polarimetric SAR Images</i> | Kajian ini menilai prestasi di antara seni bina CFAR, FRCNN, dan U-Net untuk kegunaan pengesanan bot. Set data CP SAR digunakan di dalam kajian ini yang mana setiap imej dilabelkan secara manual. Sejumlah 1956 imej digunakan untuk latihan, 260 keping untuk pengesanan dan 560 keping untuk ujian yang ditentukan secara rawak. | Imej SAR | Seni bina U-Net memperolehi prestasi yang tertinggi dengan ketepatan dan <i>recall</i> melebihi 90%. |
| 6. | Kampffmeyer et al. (2016) | <i>Semantic Segmentation of Small Objects and Modeling of Uncertainty in Urban Remote Sensing Images Using Deep Convolutional Neural Networks</i> | Kajian ini menjalankan eksperimen untuk mengesan objek kecil dengan menggunakan kaedah gabungan beberapa model bagi meningkatkan prestasi ketepatan. Penyelidik menggunakan set data daripada pertandingan pelabelan semantik ISPRS yang mempunyai 33 keping imej berbeza saiz yang mempunyai bilangan piksel di antara 3 juta hingga 10 juta. Kaedah imej augmentasi yang digunakan di dalam kajian ini adalah dengan mengeluarkan objek seperti kereta yang mempunyai pertindihan sebanyak 50%, di terbalikkan dari kiri ke kanan, dan ke atas dan di putarkan sebanyak 90 darjah yang menghasilkan penambahan imej sebanyak 8 keping. | ISPRS Vaihingen 2D | Dengan menggabungkan beberapa model, ia dapat menaikkan prestasi ketepatan keseluruhan pada pengesanan objek bersaiz kecil. Hasil daripada kajian menunjukkan bahawa ketepatan klasifikasi keseluruhan adalah sebanyak 87%. |
-

2.11 KESIMPULAN

Berdasarkan daripada kajian kesusasteraan yang dijalankan, satu rumusan berkaitan kaedah yang akan digunakan dalam projek ini. Kajian ini secara amnya akan memfokuskan pada seni bina U-Net sebagai seni bina pembelajaran mendalam yang diguna dalam pengujian. Kajian ini juga akan mencuba beberapa jenis fungsi pengaktifan untuk melihat perbezaan prestasi ketepatan diantara satu sama lain. Sukatan pengujian yang akan digunakan dalam kajian ini adalah jenis *IoU* dan *Dice Coefficient*.

Pusat Sumber
FTSM

BAB III

METODOLOGI

3.1 PENGENALAN

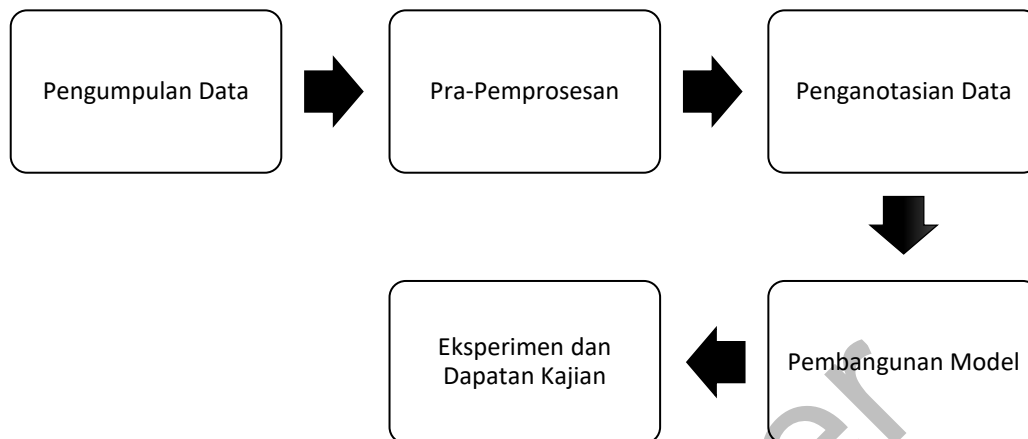
Bab ini menerangkan secara terperinci mengenai metodologi yang digunakan dalam pelaksanaan kajian ini. Setiap proses yang dilakukan adalah mengikut aturan-aturan asas yang sering digunakan dalam pelaksanaan kajian data saintis.

Secara umumnya, kaedah-kaedah yang terlibat dipecahkan kepada beberapa proses penting. Proses-proses tersebut bermula dengan pengumpulan data, pra pemrosesan, pengelasan, latihan, dan analisis.

Bagi memastikan kajian ini berjalan dengan lancar, penyediaan peralatan perkakasan dan perisian juga sangat penting. Kajian yang memfokuskan kepada pemrosesan imej memerlukan komputer berkuasa tinggi kerana setiap bit piksel adalah merupakan ciri-ciri yang akan diguna pakai dalam pembelajaran mendalam. Unit pemrosesan grafik (GPU) adalah merupakan satu keperluan dalam pemrosesan imej kerana keupayaannya untuk memproses setiap piksel secara serentak dalam satu masa yang sama. Ini membolehkan kajian dapat dilakukan dengan lebih pantas.

3.2 KERANGKA ALIRAN PROSES

Kerangka aliran proses menerangkan secara terperinci langkah-langkah yang berlaku semasa pelaksanaan kajian ini. Langkah-langkah yang ditunjukkan bermula dengan pengumpulan data dan berakhir dengan proses kajian dan dapatan. Setiap aliran proses yang utama yang terlibat dalam kajian ini ditunjukkan pada Rajah 3.1.



Rajah 3.1 Kerangka aliran proses

Perjalanan kajian ini bermula dengan pengumpulan data. Set data yang digunakan di dalam kajian ini adalah dari sumber Seagull (“Seagull Dataset – VisLab – Computer and Robot Vision Laboratory” t.th.) dalam bentuk video. Langkah seterusnya ialah pra-pemrosesan yang mana set data dalam bentuk video diekstrak ke bentuk imej. Imej-imej yang diekstrak kemudian di semak dan imej yang tidak sesuai akan diasingkan. Kemudian, langkah penganotasi dijalankan untuk menganotasi objek bot dan kapal. Proses ini adalah proses pelabelan objek di dalam gambar untuk menentukan lokasi dan bentuk objek. Set data akhir hasil daripada proses anotasi kemudian digunakan pada langkah seterusnya iaitu langkah pembangunan model. Setelah selesai pembangunan model, model yang terhasil akan di nilai dan kajian dapatan di lakukan pada langkah yang terakhir.

3.3 PERSEKITARAN KAJIAN

Persekitaran kajian menerangkan tentang peralatan yang digunakan semasa proses kajian dijalankan. Peralatan yang digunakan ini termasuklah perkakasan, perisian komputer, dan perisian atas talian.

3.3.1 Perkakasan

Peralatan komputer merupakan peralatan yang perlu dalam melaksanakan kajian ini. Penggunaan komputer banyak membantu dalam proses-proses penting terutama dalam proses pengumpulan data, pra-pemrosesan dan kajian dapatan iaitu pada langkah yang terakhir. Spesifikasi komputer yang digunakan di dalam kajian ini disenaraikan seperti Jadual 3.1.

Jadual 3.1 Spesifikasi perkakasan

Bil	Perkara	Spesifikasi
1.	CPU Chipset	Intel(R) Xeon(R) W-10855M CPU @ 2.80GHz 2.81 GHz
2.	Memory	32 GB
3.	GPU Chipset	NVIDIA Quadro T2000
4.	GPU Memory	4 GB
5.	Kapasiti simpanan	1 TB

3.3.2 Perisian Luar Talian

Tugas-tugas dalam proses pra-pemprosesan dan pembangunan model memerlukan perisian-perisian yang sesuai. Jadual 3.2 adalah senarai spesifikasi perisian yang digunakan dalam pelaksanaan kajian.

Jadual 3.2 Spesifikasi perisian luar talian

Bil	Perkara	Spesifikasi
1.	Operating System	Windows 10 Pro for Workstations
2.	GPU Driver	Nvidia CUDA dan CuDNN
3.	Development Environment	Anaconda
4.	Programming Compiler	Python 3.8
5.	Programming IDE	Jupyter Notebook
5.	Image Extraction	FFMPEG.exe
5.	Image Annotation Tools	VIA VGG Image Annotator

3.3.3 Perisian Atas Talian

Selain daripada perisian luar talian, kajian ini juga menggunakan beberapa perkhidmatan perisian atas talian. Perkhidmatan atas talian memberikan kemudahan yang cepat terutama pada proses pembangunan model dan penilaian model. Di antara perisian atas talian yang digunakan dalam kajian ini adalah seperti Jadual 3.3.

Jadual 3.3 Spesifikasi perisian atas talian

Bil	Perkara	Spesifikasi
1.	Source Code Storage	GitHub.com
2.	Training and Testing Platform	Kaggle
3.	Coding test	Google Collabs
4.	File Storage	Google Drive
5.	Programming IDE	Jupyter Notebook

3.3.4 Kerangka Pembangunan

Dalam menjalankan kajian ini, beberapa kerangka sedia ada digunakan untuk mempercepatkan proses kajian. Di antara kerangka pembangunan yang digunakan adalah seperti yang disenaraikan dalam Jadual 3.4.

Jadual 3.4 Senarai kerangka pembangunan

Bil	Perkara	Spesifikasi
1.	<i>OpenCV</i>	Kerangka pemprosesan imej
2.	<i>Pytorch</i>	Kerangka pembelajaran mendalam
3.	<i>NumPy</i>	Kerangka pemproses data pelbagai dimensi
4.	<i>Matplotlib</i>	Kerangka pembinaan graf

3.4 PENGUMPULAN DATA

Pengumpulan data adalah proses mendapatkan set data untuk digunakan dalam pembangunan pembelajaran mendalam. Bagi kajian ini, set data yang digunakan diambil daripada sumber set data Seagull (“Seagull Dataset – VisLab – Computer and Robot Vision Laboratory” t.th.) iaitu dalam format video. Sejumlah sepuluh video yang berdurasi di antara 10 saat sehingga 4 minit digunakan di dalam kajian ini. Set data ini merupakan kumpulan data multi spektrum dan multi-kamera yang disimpan bertujuan untuk penyelidikan pemantauan dan pengawasan laut. Set data ini dirakam dari sudut pandangan kenderaan udara tanpa pemandu (UAV) yang terbang di atas laut dan di atas kapal. Set data ini juga memberi beberapa jenis imej yang dirakam oleh kamera *hyper spektrum* namun untuk kajian ini hanya imej jenis RGB yang digunakan. Gambar di Rajah 3.2 menunjukkan contoh video yang didapati daripada set data Seagull



Rajah 3.2 Set data Seagull dalam bentuk video

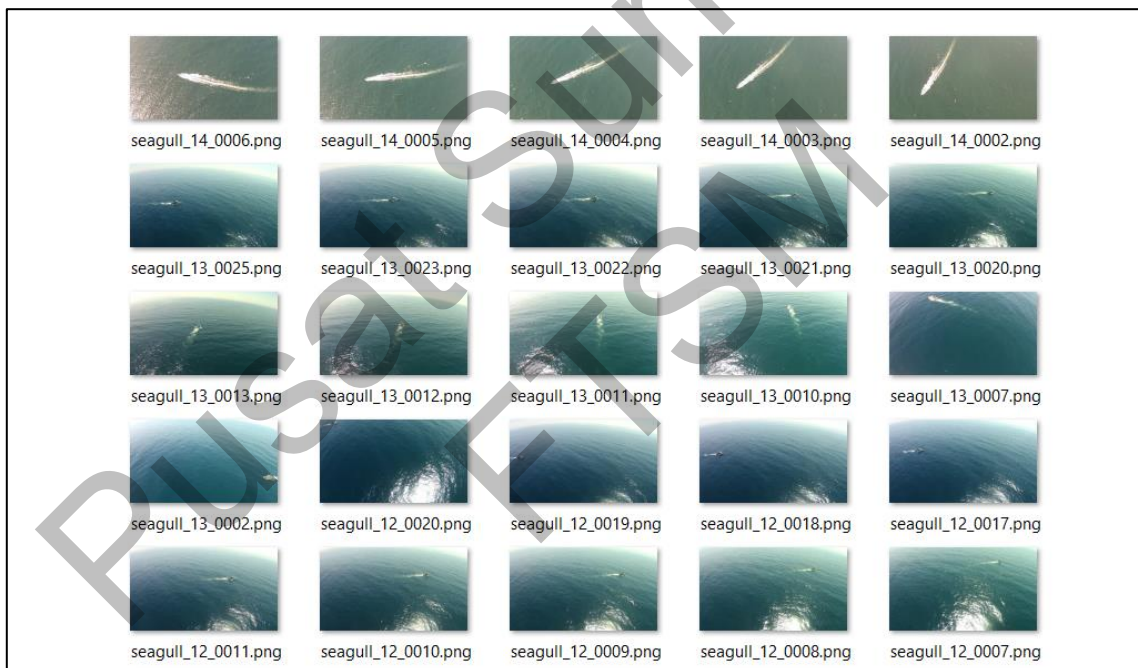
Sumber: Seagull (“Seagull Dataset – VisLab – Computer and Robot Vision Laboratory” t.th.)

Imej dalam video itu kemudian diekstrak setiap bingkai (*frame*) yang kesepuluh bermula daripada awal video hingga akhir video. Sebagai contoh, bingkai yang di ekstrak adalah bingkai yang ke sepuluh, ke dua puluh dan seterusnya sehingga akhir.

Untuk membolehkan proses ini berjalan dengan lancar, kajian ini menggunakan perisian sumber terbuka FFMPEG (FFmpeg 2021) untuk kegunaan pengekstrakan bingkai imej. Perintah kod yang diguna pakai untuk tujuan ini adalah

```
ffmpeg -i video_filename.mp4 -vf fps = 1/10 out_%04d.png
```

Hasil daripada pengekstrakan imej menggunakan FFMPEG ialah pengeluaran imej statik seperti dalam Rajah 3.3.



Rajah 3.3 Imej statik yang diekstrak daripada video Seagull

Jumlah imej yang berjaya diekstrak daripada video ini adalah sebanyak 321 keping. Daripada jumlah itu, sebanyak 300 keping akan digunakan untuk tujuan latihan model dalam pembelajaran mendalam dan sebanyak 21 keping pula akan digunakan untuk tujuan ujian di luar konteks pembelajaran mendalam. Spesifikasi set data yang digunakan dalam proses seterusnya adalah seperti Jadual 3.5.

Jadual 3.5 Spesifikasi imej set data

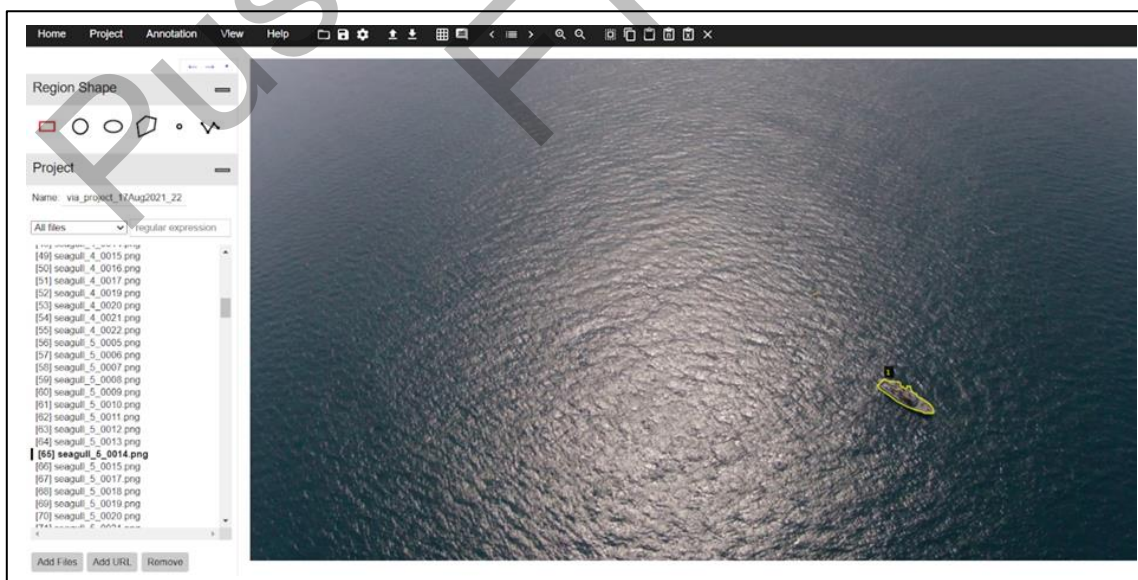
Bil	Perkara	Spesifikasi
1.	Dimensi	1920 x 1080
2.	Lebar	1920 piksel
3.	Tinggi	1080 piksel
4.	Kedalaman Bit	24 bit

3.5 PRA-PEMROSESAN

Pra-pemprosesan merupakan proses awal yang penting dalam menjalankan kajian berkaitan pembelajaran mendalam. Proses ini adalah penting kerana pembelajaran mendalam akan menggunakan sepenuhnya setiap piksel di dalam gambar ini sebagai ciri-ciri kemasukan dalam pembelajaran. Set data yang kurang sempurna seperti tiada objek bot dan kapal, imej kabur, dan refleksi cahaya yang tinggi akan dikeluarkan daripada set data untuk latihan model.

3.6 PENGANOTASIAN DATA

Penganotasian data merupakan proses manual yang dilakukan oleh penyelidik untuk melabel atau menanda bentuk dan lokasi bot dan kapal yang berada di dalam imej. Untuk tujuan ini, kajian menggunakan perisian sumber terbuka iaitu *VGG Image Annotator (VIA)* (Dutta & Zisserman 2019) untuk menganotaskan imej. Proses penganotasian imej digambarkan dalam Rajah 3.4.



Rajah 3.4 Penganotasian imej menggunakan perisian VIA

3.6.1 Topeng Imej

Segmentasi semantik memerlukan dua pasang imej iaitu imej asal dan imej topeng (*masking*) bagi menjalankan proses pembelajaran mendalam. Imej topeng merupakan imej dalam bentuk 1 bit yang berwarna hitam dan putih yang mana ia menggambarkan kedudukan dan bentuk objek dalam imej asal. Kebiasaannya imej yang berwarna putih merupakan lokasi objek dan yang berwarna hitam adalah latar belakang.

Kajian ini menggunakan skrip *Python* untuk menghasilkan topeng imej yang diperlukan dalam pembelajaran mendalam segmentasi semantik. Rajah 3.5 dan Rajah 3.6 menunjukkan hasil untuk beberapa imej asli dan imej topeng yang diguna pakai dalam kajian ini.



Rajah 3.5 Imej asli



Rajah 3.6 Imej topeng

Topeng imej ini telah dijana dengan menggunakan skrip Python yang sumber kodnya boleh dirujuk pada Lampiran A. Secara amnya, skrip ini menggunakan fail anotasi dalam format COCO yang telah dihasilkan daripada proses pelabelan dengan menggunakan perisian VIA. Fail anotasi yang dihasilkan melalui proses itu memberi informasi tentang titik-titik lokasi bagi setiap objek. Skrip ini kemudiannya menggunakan informasi itu untuk membina imej baru iaitu topeng imej. Topeng imej yang dihasilkan adalah dalam bentuk 1 bit iaitu warna hitam untuk bit 0 dan warna putih untuk bit 1. Spesifikasi topeng imej yang telah dihasilkan adalah seperti Jadual 3.6.

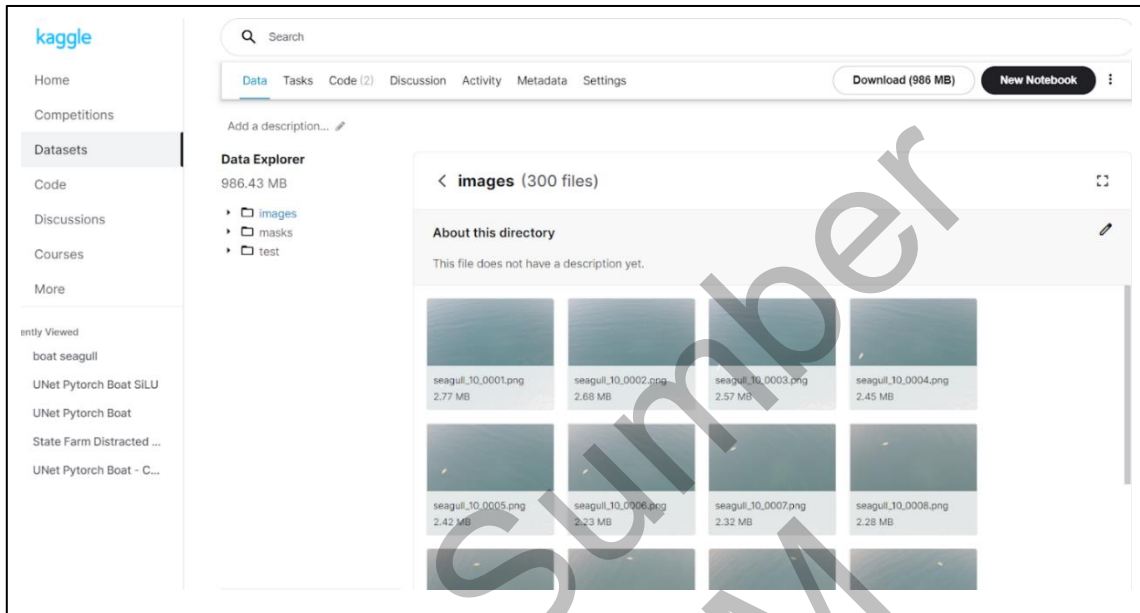
Jadual 3.6 Spesifikasi topeng imej

Bil	Perkara	Spesifikasi
1.	Dimensi	1920 x 1080
2.	Lebar	1920 piksel
3.	Tinggi	1080 piksel
4.	Kedalaman Bit	1 bit

3.6.2 Muat Naik Ke Platform Atas Talian

Set data yang telah disiapkan kemudiannya di muat naik ke platform atas talian iaitu Kaggle (“Kaggle: Your Home for Data Science” t.th.) merupakan sebuah pelantar awan yang menyediakan perkhidmatan pemprosesan imej dengan menggunakan sumber pengkomputeran berdaya tinggi. Sejumlah 16 GB memori diberikan secara percuma dalam suatu masa yang

terhad membolehkan kajian ini dijalankan dengan cepat. Selain itu juga, Kaggle menyediakan tempat untuk penyimpanan set data. Set data Seagull yang telah diproses yang digunakan dalam kajian ini boleh didapati di pautan <https://www.kaggle.com/fakhrul/boat-seagull>. Rajah 3.7 menunjukkan contoh set data yang telah di muat naik ke platform Kaggle.



Rajah 3.7 Platform Kaggle yang digunakan dalam kajian ini

3.7 PEMBANGUNAN MODEL

Pembangunan seni bina U-Net dalam kajian ini menggunakan kerangka pembangunan PyTorch. Sumber asal pengaturcaraan U-Net yang digunakan adalah berasal dari laman sesawang Smith, (“Creating and training a U-Net model with PyTorch for 2D & 3D semantic segmentation: Dataset building [1/4] | by Johannes Schmidt | Towards Data Science” t.th.). Penyelidik telah membangunkan seni bina U-Net ini untuk tujuan penyelidikannya dalam kajian *electron tomograms* iaitu dalam domain industri perubatan.

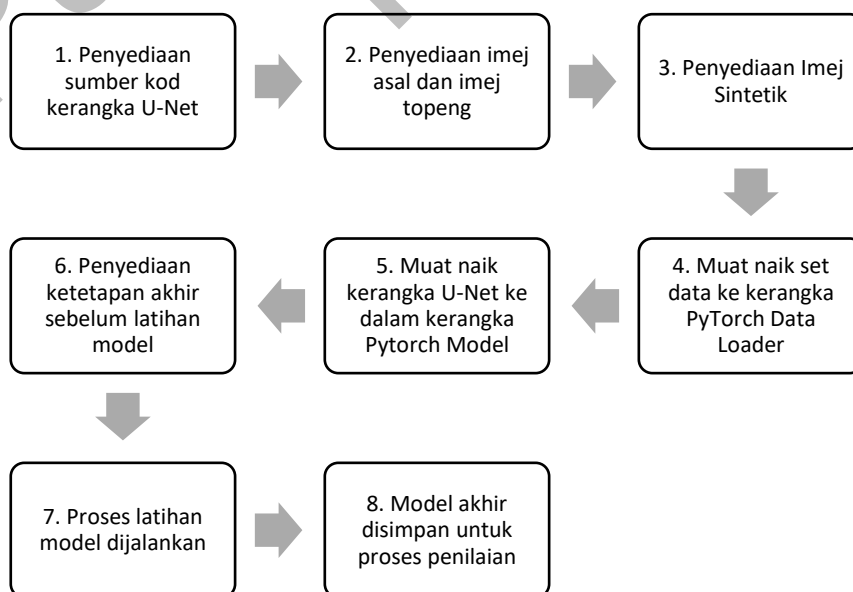
Untuk pembangunan pengaturcaraan seni bina U-Net yang dikongsi oleh Smith menggunakan set data dari Carvana (“Carvana Image Masking Challenge | Kaggle” t.th.) yang merupakan set data terkenal untuk imej segmentasi objek kereta. Rajah 3.8 menunjukkan contoh set data yang digunakan oleh kajian itu



Rajah 3.8 Contoh set data Carvana

Sumber: (“Carvana Image Masking Challenge | Kaggle” t.th.)

Kerangka PyTorch telah menyediakan modul-modul yang memudahkan penyelidik dalam pembangunan pembelajaran mendalam. Modul penting yang mesti digunakan dalam penyediaan latihan model adalah modul *Data Loader*. Modul *Data Loader* membolehkan set data di muat naik dalam kerangka PyTorch sebelum proses latihan di mulakan. Rajah 3.9 menunjukkan proses-proses penting yang digunakan dalam kajian ini bagi tujuan pembangunan model pembelajaran untuk pengesanan objek bot dan kapal.



Rajah 3.9 Kerangka aliran proses pembangunan model

3.7.1 Penyediaan Sumber Kod Seni Bina U-Net

Untuk kajian ini, sumber kod yang digunakan adalah berasal daripada kajian asal oleh Smith di laman sesawang (“johschmidt42/PyTorch-2D-3D-UNet-Tutorial” t.th.). Penyelidik telah menyediakan seni bina ini yang di kongsi di platform perkongsian GitHub. Namun begitu kod asal yang disediakan tidak mempunyai fungsi untuk menilai prestasi ketepatan model. Oleh yang demikian, fungsi penilaian ketetapan menggunakan pengiraan IoU dan *Dice Coefficient* ditambah ke dalam kod asal. Fungsi ini penting untuk membolehkan kajian ini menentukan aras prestasi sesuatu model.

Proses ini kemudiannya akan memuat turun kod akhir ke pelantar Kaggle daripada pelantar GitHub. Pelantar Kaggle merupakan pelantar yang akan digunakan untuk proses pembangunan model

3.7.2 Penyediaan Imej Asal dan Imej Topeng

Set data yang telah disiapkan pada proses penganotasian data digunakan sebagai sumber asas untuk pembangunan model pengesanan bot dan kapal. Sejumlah 300 keping gambar kesemuanya kemudian di muat naik ke pelantar Kaggle. Sejumlah 21 lagi set data untuk ujian juga di muat naik ke pelantar yang sama untuk tujuan penilaian pada proses seterusnya. Setelah dimuat naik ke pelantar, set data ini perlu disediakan di dalam sesi pembelajaran model.

3.7.3 Penyediaan Imej Sintetik

Terdapat dua ujian yang memerlukan penyediaan imej sintetik dalam kajian ini. Secara amnya terdapat dua jenis penjanaan sintetik imej yang telah di hasilkan dalam kajian ini iaitu jenis penggandaan objek dan penempatan semula objek. Penerangan secara terperinci tentang penjanaan jenis sintetik imej ini akan diterangkan dalam Bab IV. Untuk ujian yang memerlukan imej sintetik, proses penyediaan imej sintetik ini merupakan satu keperluan. Proses ini dapat dijalankan dengan menggunakan skrip Python yang telah disediakan. Imej sintetik yang baru akan dijana, dihasilkan dan disimpan pada lokasi yang telah ditetapkan di dalam skrip Python itu. Imej yang dijana melalui proses akan digunakan pada proses yang seterusnya bergantung kepada jenis ujian kajian. Sebanyak 300 imej sintetik baru akan dijana untuk tujuan kajian ini.

3.7.4 Muat Naik Set Data Ke Kerangka *Data Loader Pytorch*

Set data yang telah tersedia akan di muat naik ke dalam kerangka *Data Loader* yang disediakan di dalam platform PyTorch. Set data asal sebanyak 300 keping imej akan di gunakan untuk tujuan latihan model. Untuk ujian yang menggunakan sintetik imej, penambahan 300 imej lagi akan digunakan dalam latihan model. Ini menjumlahkan keseluruhan set data berjumlah sebanyak 600 imej. Senyak 80 peratus daripada keseluruhan jumlah imej akan digunakan sebagai set data latihan dan selebihnya akan digunakan sebagai set data pengesahan.

3.7.5 Muat Naik Seni Bina U-Net Ke Pytorch

Proses seterusnya adalah memuat naik seni bina U-Net ke dalam model PyTorch. Seni bina ini akan digunakan oleh PyTorch untuk proses latihan model. Parameter asas yang diperlukan untuk memuat naik seni bina U-Net adalah bilangan kemasukan saluran dan bilangan pengeluaran saluran. Untuk kajian ini, ia menggunakan kemasukan saluran sebanyak dua saluran yang mewakili jumlah saluran RGB imej dan pengeluaran sebanyak dua saluran yang mewakili objek bot dan latar belakang. Selain daripada itu, seni bina ini juga perlu dinyatakan jenis fungsi pengaktifan yang akan digunakan. Terdapat beberapa fungsi pengaktifan yang digunakan dalam kajian ini seperti contoh ReLU.

3.7.6 Penyediaan Ketetapan Akhir Sebelum Latihan Model

Proses ini adalah proses yang terakhir sebelum proses latihan model dijalankan. Proses ini adalah proses yang menentukan *hyper parameter* yang akan digunakan untuk ujian. Parameter yang digunakan adalah berbeza mengikut jenis ujian yang ingin dilakukan. Saiz yang digunakan di dalam latihan ini dikecilkan kepada dimensi 960 piksel x 536 piksel kerana keupayaan kuasa pemprosesan yang terhad. Jika saiz asal digunakan iaitu 1920 piksel x 1080 piksel akan menyebabkan unit pemprosesan kehilangan kesemua memori. Dalam proses ini juga akan ditentukan jumlah kadar pembelajaran yang akan digunakan sebagai contoh 0.001 ataupun 0.0001. Disini juga ditentukan jenis pengoptimum yang akan digunakan seperti Adam dan juga jenis pengiraan *loss*. Untuk kajian ini kesemua ujian menggunakan teknik pengiraan *Cross Entropy Loss* untuk pengiraan *loss*. Jumlah EPOCHS dalam setiap ujian juga ditetapkan pada kadar yang sama iaitu 50 EPOCHS untuk setiap ujian yang dijalankan. Jadual 3.7 menunjukkan senarai parameter akhir imej untuk tujuan latihan.

Jadual 3.7 Parameter yang digunakan di dalam latihan model

Bil	Perkara	Spesifikasi
1.	Dimensi imej	960 x 536 piksel
2.	Bit kedalaman imej	24 bit
3.	Saluran imej	3 Saluran (RGB)
4.	Dimensi topeng	960 x 536 piksel
5.	Bit kedalaman topeng	1 bit
6.	Saluran topeng	1 Saluran (Hitam dan putih)
7.	Saiz kumpulan (<i>batch size</i>)	2
8.	Kadar pembelajaran	0.001 atau 0.0001
9.	Bilangan EPOCHS	50
10.	Metod pengiraan <i>loss</i>	<i>Cross Entropy Loss</i>

3.7.7 Proses Latihan Model

Proses latihan model adalah proses utama yang mana pembelajaran model secara mendalam dilaksanakan. Proses ini adalah proses di mana rangkaian konvolusi neural akan menggunakan input yang dibekalkan, dan membandingkan dengan kelas yang sebenar, dan mengirakan pemberat dan nilai berat sebelah yang bersesuaian pada setiap nod. Proses ini akan berulang untuk setiap set data sebanyak 50 EPOCHS selesai. Setiap EPOCHS akan dikira nilai ketetapan model yang telah dibina. Rajah 3.10 menunjukkan proses latihan model yang telah dijalankan. Proses ini memakan masa satu jam hingga tiga jam.

```

EPOCH: 45, Validation Loss: 0.0009398345813679043, IOU: 0.6079994257474663, DICE: 0.7562765835784375
EPOCH: 45, TEST Loss: 0.00039116888232952493, IOU: 0.6253276673164309, DICE: 0.7923245267434553
EPOCH: 46, Training Loss: 0.00025662661385770965
EPOCH: 46, Validation Loss: 0.0008665497322605612, IOU: 0.6346573111495325, DICE: 0.7749218911320592
EPOCH: 46, TEST Loss: 0.00044866010880055415, IOU: 0.5696110553038272, DICE: 0.728498556397178
EPOCH: 47, Training Loss: 0.00024512135120554903
EPOCH: 47, Validation Loss: 0.0009172410905496993, IOU: 0.6429463318357008, DICE: 0.7880596862485011
EPOCH: 47, TEST Loss: 0.0004164718543920158, IOU: 0.575202829867386, DICE: 0.736791426485235
EPOCH: 48, Training Loss: 0.00024780436691192637
EPOCH: 48, Validation Loss: 0.0008992757866508327, IOU: 0.6502449928255118, DICE: 0.7800773352384567
EPOCH: 48, TEST Loss: 0.0003847160233735022, IOU: 0.6186127212295836, DICE: 0.7714051766829058
EPOCH: 49, Training Loss: 0.0002284131504590429
EPOCH: 49, Validation Loss: 0.0010520019710384076, IOU: 0.610314712105835, DICE: 0.7515538655066242
EPOCH: 49, TEST Loss: 0.0009558776007245549, IOU: 0.5249490315908042, DICE: 0.6690507422793995
EPOCH: 50, Training Loss: 0.0002647598052438601
EPOCH: 50, Validation Loss: 0.0009371684917520421, IOU: 0.6268546885291386, DICE: 0.7724171803255255
EPOCH: 50, TEST Loss: 0.0004438196708427065, IOU: 0.6115133704004709, DICE: 0.7893685698509216

```

Rajah 3.10 Proses latihan model yang telah dijalankan

3.7.8 Proses Penyimpanan Model Untuk Penilaian

Proses terakhir dalam pembangunan model adalah proses penyimpanan model. Setelah selesai proses latihan model, fail model akan disimpan ke dalam komputer untuk rekod dan tujuan penilaian dalam proses seterusnya

3.8 EKSPERIMEN DAN DAPATAN KAJIAN

Hasil daripada ujian yang dijalankan kemudian di analisis pada proses ini. Terdapat tujuh siri ujian telah dilakukan kesemuanya yang mengambil kira penggunaan tetapan yang berbeza. Di antara tetapan yang di uji adalah penggunaan kadar pembelajaran (*learning rate*) yang berbeza dan penggunaan fungsi pengaktifan yang berbeza. Hasil tetapan yang optimum daripada ujian ini digunakan dalam ujian penggunaan imej sintetik. Penerangan lanjut akan di terangkan pada BAB V, dapatan kajian.

3.9 KESIMPULAN

Bab ini secara keseluruhannya menerangkan proses-proses yang dijalankan di dalam kajian ini bermula dengan pengumpulan data sehingga kajian dapatan. Secara umumnya terdapat enam proses penting yang terlibat iaitu proses pengumpulan data, pra pemprosesan, penganotasian data, pembangunan model, penilaian model dan kajian dapatan.

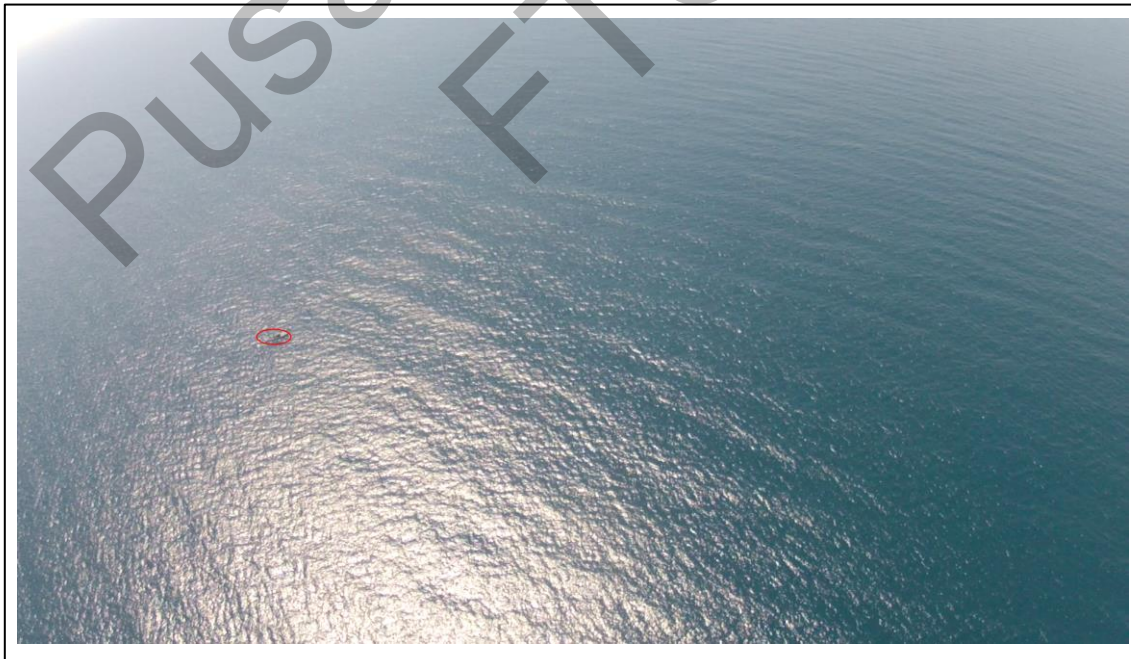
BAB IV

PEMBANGUNAN PENJANA IMEJ SINTETIK

4.1 PENGENALAN

Untuk membolehkan kajian ini menggunakan imej sintetik secara teratur, penyelidik telah membina dua kaedah untuk menjana dua jenis sintetik imej berdasarkan imej asal. Dua jenis sintetik imej yang dijana adalah jenis penggandaan objek, dan penempatan semula objek. Kaedah ini dibangunkan dengan menggunakan pengaturcaraan Python dan kerangka OpenCV.

Imej augmentasi yang sedia ada secara amnya sesuai digunakan untuk kegunaan objek yang besar dan selalunya memenuhi hampir keseluruhan ruang gambar. Untuk kes penggunaan kajian ini, objek di dalam imej yang digunakan adalah sangat kecil yang merangkumi sebahagian peratusan sahaja daripada keseluruhan gambar. Rajah 4.1 menunjukkan lokasi bot di dalam set data yang di guna di dalam kajian ini.

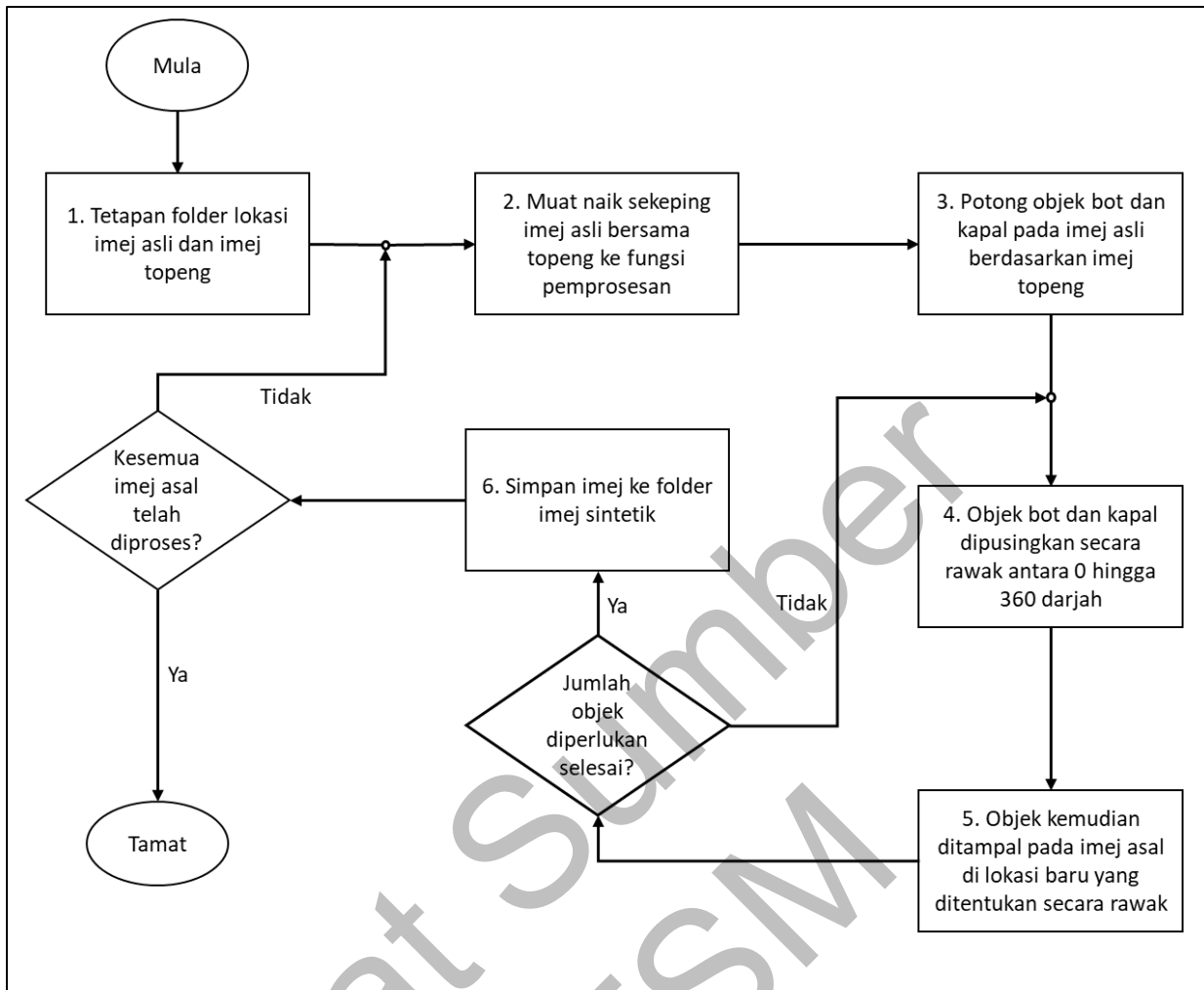


Rajah 4.1 Lokasi bot di dalam imej

Imej ini akan diproses semasa latihan model dalam bentuk pemprosesan secara berkumpulan. Terdapat banyak segmen di dalam gambar yang tidak mempunyai objek bot namun ia juga diperlukan untuk mengenal pasti jenis kelas latar belakang. Daripada permasalahan ini, penyelidik telah membangunkan dua jenis kaedah untuk penjanaan sintetik imej berdasarkan imej asal ini.

4.2 PENGANDAAN OBJEK

Kaedah penggandaan objek merupakan satu fungsi pengaturcaraan yang bertujuan menggandakan jumlah objek di dalam imej. Jumlah objek asal pada imej asli hanya mempunyai satu objek sahaja. Teknik ini membolehkan objek asli itu di ulang mengikut jumlah tetapan yang diberikan. Dengan secara rawak fungsi ini meniru dan menampal imej ke lokasi baru manakala sudut orientasinya juga di pilih secara rawak. Proses-proses yang terdapat di dalam kaedah ini digambarkan pada Rajah 4.2.



Rajah 4.2 Carta alir penjanaan imej sintetik bagi penggantian objek

Berdasarkan carta aliran pada Rajah 4.2, kaedah penggantian objek dijalankan mengikut langkah-langkah seperti berikut

1. Langkah pertama dimulai dengan ketetapan lokasi folder di mana imej asli dan imej topeng disimpan.
2. Kemudian setiap pasang imej asli dan imej topeng akan di muat naik ke dalam fungsi pemrosesan. Untuk kajian ini, terdapat sebanyak 300 keping imej asli dan 300 keping imej topeng yang digunakan.
3. Objek pada imej asli itu kemudiannya di potong berdasarkan nilai yang terdapat pada imej topeng.