

PENKELASAN GAYA RENANG BAGI JUJUKAN
VIDEO MENGGUNAKAN PEMBELAJARAN
MENDALAM

AHMAD HAFRIZALMAN BIN ABDUL HAMID

UNIVERSITI KEBANGSAAN MALAYSIA

PENKELASAN GAYA RENANG BAGI JUJUKAN VIDEO MENGGUNAKAN
PEMBELAJARAN MENDALAM

AHMAD HAFRIZALMAN BIN ABDUL HAMID

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

FAKULTI TEKNOLOGI DAN SAINS MAKLUMAT
UNIVERSITI KEBANGSAAN MALAYSIA
BANGI

2023

PENKELASAN GAYA RENANG BAGI JUJUKAN
VIDEO MENGGUNAKAN PEMBELAJARAN
MENDALAM

PUSAT SUMBER FTSM

UNIVERSITI KEBANGSAAN MALAYSIA

PENAKUAN

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

27 Mac 2023

AHMAD HAFRIZALMAN
BIN ABDUL HAMID
P106877

PUSAT SUMBER FTSM

PENGHARGAAN

Dengan nama Allah yang Maha Pemurah, lagi Maha Penyayang, segala puji bagi Tuhan sekalian alam, saya merafa'kan kesyukuran ke hadrat Ilahi dengan rahmat serta limpah kurnianya dapat disempurnakan kajian ini dengan jayanya.

Pertamanya, saya merakamkan sebanyak penghargaan kepada Prof. Madya Dr. Zalinda Othman, selaku penyelia projek akhir saya di atas segala bimbingan, nasihat serta tunjuk ajar dalam banyak segi yang saya kekurangan, dan dengan ini saya merakamkan ucapan ribuan terima kasih kepada beliau.

Dirakamkan juga sebanyak penghargaan kepada semua tenaga pengajar yang telah mencurahkan ilmu yang bermanfaat, serta kepada staf sokongan, yang terlibat secara langsung ataupun tidak sepanjang pengajian saya di Fakulti Teknologi dan Sains Maklumat, Universiti Kebangsaan Malaysia.

Kesyukuran ke hadrat yang Esa, dikurniakan kasih sayang yang tidak bersyarat dari kedua bonda Jamilah Binti Ketik serta ayahanda Abdul Hamid Bin Harun dengan asbab kasih sayang tersebut saya berusaha untuk memperbaiki diri dalam segala segi. Juga kepada isteri tercinta, Seri Nur Fasa Binti Abu Hasan yang memberi sokongan dalam setiap keadaan jatuh dan bangunnya saya pada setiap masa, juga anak-anak, Ahmad Uwais dan Aaishah Nawwaarah, memberi semangat kepada saya agar dapat pula menjadi inspirasi kejayaan mereka sendiri di kemudian hari.

Ribuan terima kasih diucapkan kepada pembiaya serta penaja saya meneruskan pelajaran iaitu Bahagian Biasiswa, Kementerian Pengajian Tinggi, yang membiaya saya sepanjang pengajian, turut serta majikan Jabatan Pengajian Politeknik dan Kolej Komuniti yang memberi ruang kepada kakitangan untuk meningkatkan ilmu diri dengan memberikan pelepasan cuti untuk tujuan ini.

Akhir kata, terima kasih juga diucapkan kepada semua pihak yang telah terlibat secara langsung atau tidak langsung, membantu saya dengan tidak mengharapkan apa-apa bentuk balasan. Keikhlasan semua tidak dapat dibalas melainkan dengan ganjaran yang dijanjikan oleh Allah SWT bagi semua kebaikan sehinggakan sebesar sebiji zarah.

ABSTRAK

Pengecaman pergerakan manusia mempunyai potensi yang memberangsangkan dalam pelbagai bidang. Penggunaan pembelajaran mesin serta pembelajaran mendalam pada masa kini membantu dalam pembangunan sistem automasi penglihatan komputer. Ia membantu tugas pemantauan yang berterusan selama ini dilakukan oleh manusia. Rangkaian Neural Konvolusi (CNN) setakat ini adalah model rangkaian neural yang paling berjaya dengan kejayaan dalam tugas-tugas pemrosesan isyarat, imej, dan video. Dalam kajian ini, satu set eksperimen telah direkabentuk supaya dapat menghasilkan satu model CNN yang berupaya dalam mengelas jujukan imej dari video-video berdasarkan tiga jenis renangan, iaitu gaya bebas, kuak dada, dan kuak kupu-kupu. Set data yang digunakan dalam kajian ini adalah hasil kompilasi video yang dimuat turun dari laman perkongsian video YouTube. Kriteria utama video ditetapkan supaya model ini dapat digunakan untuk suapan video pandangan dalam air dengan memantau perenang dari pandangan sisi kolam. Melalui kerja-kerja penyuntingan video serta teknik *augmentation*, dua set data *swimClass0* dan *swimClass* telah dihasilkan di mana set data mengandungi video-video kompilasi renangan berdasarkan ketiga-tiga kelas renangan tersebut. Memandangkan faktor seperti sudut pandangan kamera dan hingar mempunyai pengaruh kepada prestasi model, maka kriteria ditetapkan supaya rangkaian konvolusi dapat memberikan prestasi terbaik menggunakan sejumlah data video yang terhad beserta set data yang ditambah. Satu senibina CNN dua dimensi, iaitu 2DCNN telah dipilih untuk membina model pengelasan. Dimulai dengan seni bina asas iaitu seni bina 2DCNN dengan dua lapisan konvolusi, 3 set eksperimen dijalankan dengan gabungan berbeza terhadap setiap set data *swimClass0* dan *swimClass* serta mengubah saiz masukan. Selanjutnya, kedalaman rangkaian ditingkatkan dengan menambah satu lapisan konvolusi kepada 3 lapisan konvolusi berbanding seni bina yang asas. Model 2DCNN ini mengekstrak pergerakan dalam bentuk jujukan imej dari sesebuah video lalu dirambat melalui rangkaian konvolusi 2DCNN. Untuk ini juga, penilaian prestasi model 2DCNN dijalankan dalam enam eksperimen dengan tetapan yang berlainan terhadap dua set data *swimClass0* sebagai set data awalan dan *swimClass* bagi set data yang ditambah dan menilai tetapan yang terbaik. Melalui eksperimen ini, didapati model mencapai ketepatan terbaik sebanyak 99.29%, 99.38% kepersisan dan 99.37% ingatan semula menggunakan gabungan seni bina 2DCNN dengan 3 lapisan konvolusi, saiz masukan 64x64 terhadap set data *swimClass0*. Selepas itu, model yang memberikan keputusan yang terbaik dipilih untuk menentusah model secara lebih lanjut terhadap set data penanda aras UCF15. Dalam kajian ini didapati model 2DCNN yang terbaik menghasilkan nilai ketepatan sebanyak 97.11%, kepersisan sebanyak 97.11% dan ingatan semula sebanyak 97.08% ke atas set data diubah UCF15. Kesemua nilai ini menunjukkan model yang dicadangkan dalam kajian ini mencapai prestasi yang lebih baik berbanding hasil dari model bandingan.

SWIMMING STYLE RECOGNITION FOR VIDEO SEQUENCES USING DEEP LEARNING

ABSTRACT

Human movement recognition has exciting potential in various fields. The use of machine learning as well as deep learning nowadays helps in the development of computer vision automation system. It helps the continuous monitoring tasks that has been done by humans. Convolutional Neural Networks (CNN) are by far the most successful form of neural network architecture with success in signal, image and video processing tasks. In this study, a set of experiments was designed to produce a CNN model capable of classifying image sequences from video based on three types of swimming, namely freestyle, breaststroke, and butterfly. The dataset used in this study is the result of compilation of videos downloaded from the video sharing site, YouTube. The main video criteria are set so that the model can be used for underwater view video feeds by monitoring swimmers from a side view of the pool. Through video editing and augmentation techniques, two datasets *swimClass0* and *swimClass* have been produced where the datasets contain swimming compilation videos based on the three swimming classes. Since factors such as camera viewpoint and noise have an influence on the performance on the model, the criteria are set so that the convolutional network can provide the best performance using a limited amount of video data and augmented datasets. A two-dimensional CNN architecture, 2DCNN was chosen to build the classification model. Starting with the basic architecture which is a 2DCNN architecture with two conditional layers, 3 sets of experiments are run with different combinations of each *swimClass0* and *swimClass* dataset and varying the input size. Furthermore, the depth of the network is increased by adding one convolutional layer to 3 convolutional layers compared to the basic architecture. This 2DCNN model extracts movement in the form of a sequence of images from a video and propagates through the network. The performance evaluation of the 2DCNN was carried out to find the best settings. Through this experiment, it was found that the model achieved the best accuracy of 99.29%, 99.38% precision and 99.37% recall using a combination of 2DCNN architecture with 3 convolutional layers, input size 64x64 for the *swimClass0* dataset. After that, the model that gave the best results was selected to further validate the model against the UCF15 benchmark dataset. In this study, it was found that the chosen 2DCNN model produced an accuracy value of 97.11%, a precision of 97.11% and a recall of 97.08% on the UCF15 dataset. All these values show that the proposed model in this study achieves better performance compared to the results of the comparational study.

KANDUNGAN

		Halaman
PENGAKUAN		iii
PENGHARGAAN		iv
ABSTRAK		v
ABSTRACT		vi
KANDUNGAN		vii
SENARAI JADUAL		x
SENARAI ILUSTRASI		xi
SENARAI SINGKATAN		xiii
BAB 1	Pengenalan	
1.1	Pendahuluan	1
1.2	Motivasi	2
1.3	Permasalahan Kajian	3
1.4	Objektif Kajian	4
1.5	Skop Kajian	4
1.6	Kepentingan Kajian	6
1.7	Organisasi Kajian	6
BAB II	KAJIAN LITERATUR	
2.1	Pengenalan	8
2.2	Latar Belakang Perkembangan Kepintaran Buatan	8
2.3	Penemuan Pengetahuan Data Raya	9
2.4	Kategori Algoritma Pembelajaran Mesin	10
	2.4.1 Pembelajaran diselia	11
	2.4.2 Pembelajaran tak diselia	11
	2.4.3 Pembelajaran perkukuhan	12
2.5	Rangkaian Neural Buatan	12
	2.5.1 Pembelajaran mendalam	15
2.6	Latarbelakang Komputer	15

2.6.1	Algoritma Pengesanan Ciri Penglihatan Komputer	17
2.6.2	Pengesanan Ciri Melalui Penapis yang Dipelajari	18
2.7	Rangkaian Saraf Konvolusi	19
2.7.1	Asas model rangkaian saraf buatan	19
2.7.2	Latar belakang serta seni bina rangkaian CNN	20
2.7.3	Lapisan konvolusi	22
2.7.4	Rangkaian 2DCNN & 3DCNN	24
2.8	Penggunaan CNN Dalam Pemprosesan Imej Dan Video	26
2.9	Pengecaman Pergerakan Manusia	27
2.10	Gaya Renangan	34
2.10.1	Gaya Renangan: Gaya bebas	36
2.10.2	Gaya Renangan: Kuak kupu-kupu	37
2.10.3	Gaya Renangan: Kuak dada	38
2.11	Analisis Pergerakan Renangan Secara Pembelajaran Mesin	39
BAB III	METODOLOGI KAJIAN	
3.1	Pengenalan	41
3.2	Rekabentuk Kajian	41
3.3	Fasa Pengumpulan Dan Pemahaman Data	42
3.4	Fasa Penyediaan Data	45
3.4.1	Fasa Penyediaan Data: Penyuntingan Secara <i>Extraction</i>	48
3.4.2	Fasa Penyediaan Data: Penyuntingan Secara <i>Cropping</i>	50
3.4.3	Fasa Penyediaan Data: Penyuntingan Secara <i>Mirroring</i>	50
3.4.4	Set Data	51
3.5	Fasa Pembangunan Model	52
3.5.1	Model 2DCNN Dengan Dua Lapisan Konvolusi	53
3.5.2	Model 2DCNN Dengan Tiga Lapisan Konvolusi	60
3.6	Tetapan Eksperimen	63
3.7	Fasa Penilaian Model	63
3.7.1	Penilaian Model: Fungsi Kehilangan	64
3.7.2	Penilaian Model: Ketepatan	65
3.7.3	Penilaian Model: Jadual Kontigensi, Ketepatan, Kepersisan dan Ingatan Semula	65

	3.7.4 Penilaian Model: Penilaian Perbandingan	67
3.8	Bahan Dan Kaedah	68
	3.8.1 Perisian dan Perkakasan	69
BAB IV	DAPATAN KAJIAN	
4.1	Pengenalan	71
4.2	Keputusan Model CNN	71
	4.2.1 Ketepatan dan Fungsi Kehilangan Model Pengelasan	73
	4.2.2 Jadual Kontigensi Pelbagai Kelas, Kepersisian dan Ingatan Semula	77
4.3	Prestasi Model Dengan Data Penanda Aras	80
4.4	Perbincangan	83
4.5	Kesimpulan	85
BAB V	RUMUSAN DAN CADANGAN	
5.1	Pengenalan	86
5.2	Rumusan Kajian	86
5.3	Sumbangan Kajian	88
5.4	Batasan Kajian	89
5.5	Cadangan Kajian	90
	RUJUKAN	92
	LAMPIRAN	
Lampiran A	Jadual Keputusan Eksperimen 1	
Lampiran B	Jadual Keputusan Eksperimen 2 & 3	
Lampiran C	Jadual Keputusan Eksperimen 4	
Lampiran D	Jadual Keputusan Eksperimen 5	
Lampiran E	Jadual Keputusan Eksperimen 6	
Lampiran D	Jadual Keputusan Eksperimen Validasi	

SENARAI JADUAL

No. Jadual		Halaman
Jadual 3.1	Senarai Lapisan Sequential dalam Model 2DCNN dengan dua lapisan konvolusi	54
Jadual 3.2	Senarai Lapisan Sequential dalam Model 2DCNN dengan tiga lapisan konvolusi	61
Jadual 3.3	Senarai eksperimen dan gabungan tetapan serta set data	63
Jadual 3.4	Jadual kontigensi pelbagai kelas	66
Jadual 3.5	Senarai perpustakaan Python serta fungsi	69
Jadual 4.1	Tetapan eksperimen	72
Jadual 4.2	Nilai kehilangan validasi dan ketepatan validasi eksperimen	74
Jadual 4.3	Keputusan keseluruhan Eksperimen 1-6	78
Jadual 4.4	Perbandingan pencapaian antara model kajian serta model bandingan	82

PUSAT SUMBER FITSM

SENARAI ILUSTRASI

No. Rajah		Halaman
Rajah 1.1	Perbandingan pandangan kamera	5
Rajah 2.1	Model asas perseptron	13
Rajah 2.2	Model MLP	14
Rajah 2.3	Konvolusi kepada matriks melalui operasi darab	20
Rajah 2.4	Senibina rangkaian <i>neocognitron</i>	21
Rajah 2.5	Contoh penggunaan penapis bersaiz 2x2 dalam 2 langkah	22
Rajah 2.6	Operasi menggunakan pengumpulan maksimum serta pengumpulan purata	23
Rajah 2.7	Strok-strok utama bagi renangan gaya bebas	36
Rajah 2.8	Strok-strok utama bagi renangan kuak kupu-kupu	37
Rajah 2.9	Strok-strok kunci bagi renangan kuak dada	38
Rajah 3.1	Fasa rekabentuk kajian	42
Rajah 3.2	Paparan antara muka perisian 4KDownloader	44
Rajah 3.3	Contoh jujukan video sebelum kerja suntingan	46
Rajah 3.4	Contoh jujukan pergerakan berciri dalam pergerakan renangan	47
Rajah 3.5	Pemilihan pandangan mengikut kriteria ditetapkan	48
Rajah 3.6	Proses penyuntingan dan pemilihan jujukan bingkai melalui proses <i>extraction</i> dan <i>cropping</i>	49
Rajah 3.7	Proses mirroring pada jujukan bingkai	50
Rajah 3.8	Tangkapan skrin sebahagian daripada set <i>dataswimClass0</i>	52
Rajah 3.9	Tangkapan skrin sebahagian daripada set <i>dataswimClass</i>	52
Rajah 3.10	CNN bingkai tunggal	53
Rajah 3.11	Kod pseudo bagi model cadangan dua lapisan konvolusi	56
Rajah 3.12	Carta alir eksperimen 2DCNN dengan tiga lapisan konvolusi	57
Rajah 3.13	Transformasi imej contoh melalui tiga lapisan konvolusi	59
Rajah 3.14	Kod pseudo bagi model cadangan tiga lapisan konvolusi	61
Rajah 3.15	Carta alir eksperimen 2DCNN dengan tiga lapisan konvolusi	62
Rajah 4.1	Fungsi kehilangan eksperimen (1) – (6)	75
Rajah 4.2	Fungsi kehilangan setiap eksperimen (1) – (6)	76

Rajah 4.3	Jadual kontigensi Eksperimen 3 beserta label	77
Rajah 4. 4	Jadual kontigensi Eksperimen (1) – (6)	79
Rajah 4.5	Ketepatan Model Terhadap Data UCF15	81
Rajah 4.6	Fungsi Kehilangan Model Terhadap Set Data UCF15	81
Rajah 4.7	Jadual kontigensi bagi eksperimen terhadap set data UCF15	83

PUSAT SUMBER FTSM

SENARAI SINGKATAN

2D	2 dimensi
2D CNN	2 dimensi convolutional neural network
3D	3 dimensi
3D CNN	3 dimensi convolutional neural network
AI	Artificial intelligence
ANN	Artificial neural network
API	Application programming interface
BoW	Bags of words
CCD	Charge-couples device
CCTV	Closed-circuit television
CMOS	Complementary metal oxide semiconductor
CNN	Convolutional neural network
Fps	Frame per second
FPU	Floating point unit
GPU	Graphic processing unit
HMM	Hidden Markov model
HOG	Histogram of gradients
HSV	Hue saturation value
IoT	Internet of things
JSON	JavaScript object notation
KDD	Knowledge discovery in databases
LSTM	Long sort-term memory
MLP	Multi-layer perceptron
RGB	Red-green-blue
RNN	Recurrent neural network

SEM	Scanning electron microscope
SIFT	Scale-invariant feature transform
SVM	Support vector machine
tMHI	Timed motion history image
TPU	Tensor processing unit
UAV	Un-manned autonomous vehicle
UCF	University of Central Florida

PUSAT SUMBER FTSM

BAB I

PENGENALAN

1.1 PENDAHULUAN

Penglihatan komputer mengalami peningkatan yang pesat dalam beberapa dekad terakhir kebelakangan ini. Ia menyediakan satu keupayaan kepada manusia dalam menjalankan pemerhatian visual menggantikan tugas-tugas berkaitan dan menyerahkannya kepada komputer untuk memerhati seterusnya melakukan penilaian terhadap skop pemandangan yang tersedia. Pendekatan analisis melalui kaedah pembelajaran mesin, sehinggalah kaedah pembelajaran mendalam, walaupun telah berlaku sejak pertengahan 1950-an lagi tetapi hanya baru-baru ini berlaku secara mendadak dengan keupayaan komputer yang tinggi. Ketersediaan perkakasan seperti GPU dan TPU untuk pengiraan samada secara terus atau secara maya melalui penyedia perkhidmatan pengkomputeran awan dan capaian kepada penggunaannya menyebabkan peningkatan dalam kemajuan penglihatan komputer. Begitu juga pembangunan perisian dan bahasa atur cara untuk kerja-kerja berkaitan juga semakin meluas berserta berkembangnya pertambahan perpustakaan perisian. Gabungan faktor-faktor tersebut memudahkan lagi kerja-kerja penyelidikan dan pembangunan model analisis penglihatan komputer sekali gus membantu dalam mempercepatkan proses pembangunan model.

Penglihatan manusia berlaku apabila cahaya memasuki mata manusia, melalui retina dan kemudiannya akan ditukar kepada isyarat elektrik lalu dihantar ke otak untuk diproses. Segala persepsi terbina melalui proses ini, termasuk keupayaan manusia untuk bertindak, berfikir serta menganalisis. Ianya berlainan dengan penglihatan menggunakan komputer di mana imej mempunyai atribut tertentu, disimpan dalam bentuk piksel, piksel-piksel diwakili dalam bentuk angka-angka yang

akan diproses dalam kerja-kerja pemrosesan imej. Imej-imej pula mungkin terdiri dari jenis statik iaitu foto dan juga dalam bentuk jujukan imej seperti video.

Dengan ketersediaan pelbagai peranti dari CCTV, dron, imej satelit sehinggalah peranti kegunaan peribadi seperti telefon pintar dan kamera disertakan pula dengan storan awan memerlukan kajian yang berterusan dalam bidang analisis imej. Penglihatan komputer kini digunakan dalam pelbagai bidang dari media sosial, keselamatan, ketenteraan, geo-informasi sehinggalah beberapa pencapaian tinggi dalam bidang perubatan untuk terutama pengimejan diagnostik. Kajian ini dicadangkan sebagai satu pendekatan yang berpotensi untuk dikembangkan dalam bidang keselamatan yang berasaskan pemantauan yang dibantu teknologi komputer.

1.2 MOTIVASI

Berenang merupakan satu aktiviti rekreasi dengan intensiti yang setanding dengan aktiviti senaman lain seumpama berlari, berbasikal dan seumpamanya. Ia menjadi alternatif memandangkan manfaat yang boleh diperolehi seperti ketahanan kardiovaskular, kekuatan otot dan ketahanan fizikal. Di sebalik manfaat-manfaat tersebut, berenang juga adalah satu aktiviti fizikal yang berimpak rendah menyebabkan ia menjadi semakin popular tambahan ia boleh dilakukan dengan disertai ahli keluarga. Aktiviti berenang boleh dilakukan di kawasan terbuka seperti laut dan sungai, manakala secara tertutup di kolam renang awam juga kolam renang yang terdapat di pusat peranginan seperti hotel.

Bagi tujuan keselamatan, kebanyakan premis ada menyediakan khidmat pegawai penyelamat yang dipertanggungjawabkan untuk memantau aktiviti pengguna kolam renang dan melakukan tugas penyelamatan sekiranya berlaku insiden tidak diingini. Akan tetapi, terdapat juga premis yang tidak menyediakan perkhidmatan pegawai penyelamat lalu ia menjadi tanggungjawab sepenuhnya pada pengguna untuk memastikan keselamatan diri serta ahli keluarga. Sekiranya pegawai penyelamat disediakan pula, ia masih lagi tidak dapat menjamin bahawa insiden lemas tidak akan berlaku sama sekali sekiranya dipertimbangkan faktor lain berkaitan manusia.

Bagi menutupi kelemahan sistem pengawalan yang sepenuhnya dilakukan oleh manusia, sistem pemantauan dibantu komputer perlu dibangunkan dengan lebih giat. Beberapa kajian telah diketengahkan dengan pelbagai pendekatan yang berbeza. Akan tetapi, pendekatan kajian berdasarkan pembelajaran mendalam masih lagi kurang dijalankan terutama penggunaan pembelajaran mendalam berdasarkan pemrosesan imej serta video belum pernah dijalankan. Maka kajian ini dicadangkan untuk mempelopori pembangunan model untuk tujuan pemantauan perenang secara visual menggunakan pendekatan pembelajaran mendalam. Maka, satu model perlu dibangunkan di mana model mempunyai keupayaan mengkelas pergerakan renang berdasarkan jujukan imej yang dirakam.

1.3 PERMASALAHAN KAJIAN

Kajian berkaitan pergerakan renang samada dalam bentuk pembelajaran mesin serta pembelajaran mendalam telah dilakukan dalam beberapa kajian dalam khalayak pengkaji. Skop pengkajian merangkumi dari pengkelasan jenis renang, pembangunan sistem pemantauan sehinggalah pemantauan prestasi renang. Akan tetapi, pembangunan model pengkelasan tidak mendapat liputan meluas di kalangan pengkaji dengan menggunakan beberapa pendekatan melalui pembelajaran mesin dan pembelajaran mendalam. Seperti kajian oleh Tarasevicius (2020) dan Brunner (2019) menggunakan pendekatan pembelajaran mendalam tetapi menggunakan data dari penderia dengan merakam ritma pergerakan perenang serta bacaan tanda vital untuk membangun model. Pendekatan lain pula seperti yang dicadangkan Liao (2003) menggunakan imej perenang dan menggunakan pendekatan pembelajaran mesin dalam membina model klasifikasi renang. Maka, setakat kajian ini dibuat, belum didapati usaha untuk membangunkan model pengkelasan renang berdasarkan imej atau jujukan imej menggunakan pendekatan pembelajaran mendalam.

Bagi tujuan pembangunan model pengkelasan renang berdasarkan imej atau jujukan imej pula memerlukan kepada penggunaan set video yang berkait renang. Dinyatakan di sini, kajian-kajian tentang pergerakan tubuh badan manusia, contohnya melibatkan set data UCF50 dan UCF101 (Soomro et al., 2012) ada disertakan pengelasan bagi pergerakan secara berenang. Akan tetapi, berdasarkan set data video yang berkaitan, ia tidak dapat membantu bagi membina model penglihatan komputer

yang dapat mengawasi pergerakan berlaku di dalam air. Selain itu, video dari set data tersebut adalah video yang diambil dari luar air juga dari pandangan atas, kajian ini mencadangkan penggunaan set data menggunakan set data dengan video yang diambil pada sisi perenang dan di dalam air di mana set data adalah sebaliknya. Justeru, dalam kajian ini, kami menumpukan dalam pengkajian terhadap analisis pergerakan manusia dalam aktiviti berenang di mana set data terdiri dari video aktiviti berenang mengikut kriteria pemilihan video yang ditetapkan. Di sini wujudnya keperluan untuk mewujudkan set data yang terdiri dari tiga kelas berdasarkan jenis strok renang – gaya bebas (*freestyle*), kuak dada (*breast stroke*), dan kuak kupu-kupu (*butterfly stroke*) berdasarkan kriteria. Set data dianalisis untuk membantu dalam membangunkan model yang berupaya mengklasifikasi jenis-jenis renang berdasarkan tiga kelas renang.

1.4 OBJEKTIF KAJIAN

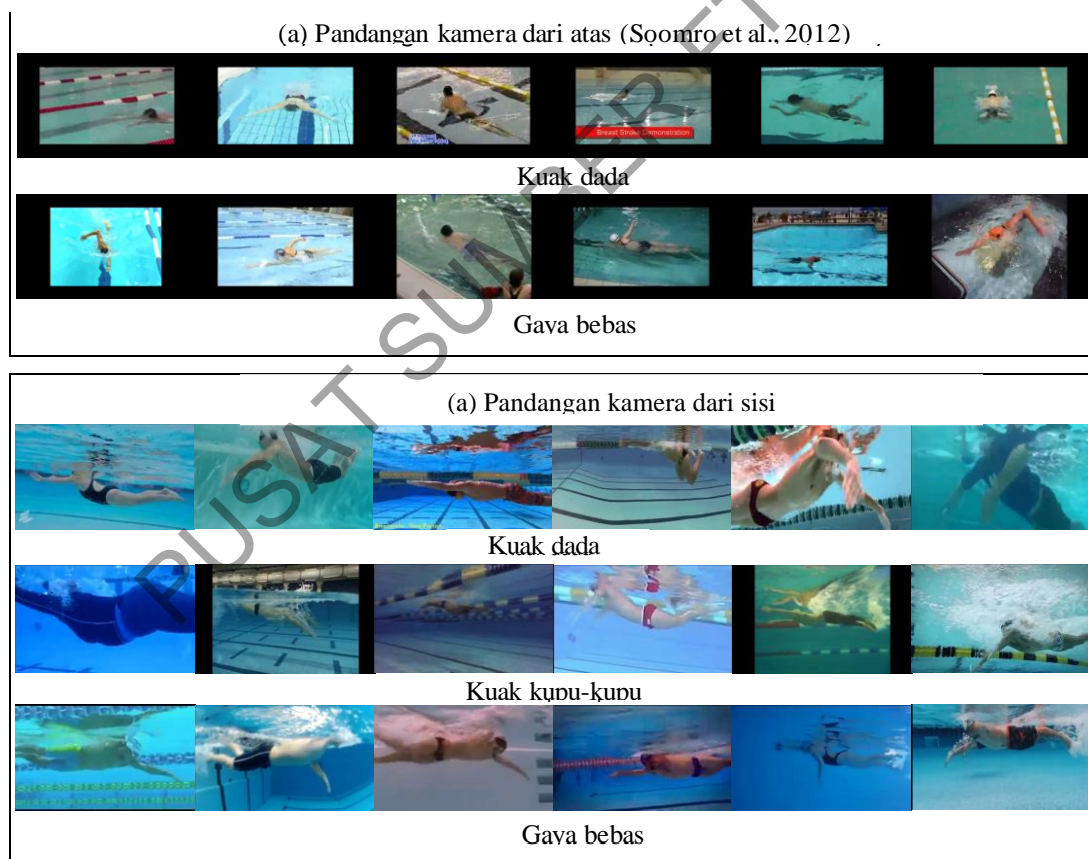
Berdasarkan permasalahan kajian yang dinyatakan sebelum ini, beberapa objektif kajian perlu ditetapkan dan menjadi panduan di sepanjang kajian. Objektif-objektif kajian adalah seperti berikut:

1. Menyedia dan menyusun jujukan imej dari video yang memenuhi kriteria serta skop kajian.
2. Membangunkan model rangkaian neural 2DCNN yang berupaya mengklasifikasi video berdasarkan tiga jenis renang.

1.5 SKOP KAJIAN

Bagi memastikan objektif kajian dapat dicapai, serta mempertimbangkan keupayaan dari segi sumber, kemahiran, serta pengetahuan, skop kajian perlu lebih spesifik dan menjadi had sempadan kajian. Skop kajian ini adalah menggunakan set data video tiga kelas renang yang utama iaitu gaya bebas, kuak dada, serta kuak kupu-kupu. Set data adalah terdiri dari kompilasi video yang juga melibatkan kerja-kerja muat turun serta penyuntingan. Bagi memastikan model yang dibina mampu membuat generalisasi dengan baik di samping berupaya disesuaikan dengan jumlah data yang

agak kecil, pandangan video perlu dipilih mengikut kriteria yang ditentukan. Kriteria-kriteria tersebut adalah – pertama, video adalah pada pandangan sisi perenang sahaja; kedua, video adalah pada pandangan dalam air sahaja; ketiga, video menunjukkan pergerakan renang untuk seorang perenang sahaja pada sesuatu masa. Kriteria ini ditetapkan adalah berbeza daripada video berkaitan yang boleh didapati dalam set data penanda aras seperti UCF101 oleh Soomro et al., (2012) seperti ditunjukkan dalam Rajah 1.1. Set data UCF101 terdiri dari video renang pandangan atas dan ia tidak dibina khusus untuk tujuan pengelasan jenis renang. Kriteria ini perlu ditetapkan untuk memastikan model 2DCNN dapat melihat elemen-elemen yang perlu sahaja, elemen selain yang diperlukan akan memperkenalkan hingar kepada model dan sekaligus akan menjejaskan prestasi model.



Rajah 1.1 Perbandingan pandangan kamera (a) UCF101 Soomro et al., (2012) dan (b) kriteria cadangan

1.6 KEPENTINGAN KAJIAN

Model hasil dari kajian ini dapat digunakan untuk membangunkan sistem pengawasan keselamatan kolam, terutamanya kolam renang dengan memaksimum keselamatan dalam kolam renang. Selain dari itu, model 2DCNN yang dibangunkan diharap dapat menutupi kekurangan tugas pengawasan sepenuhnya yang selama ini dilakukan oleh manusia tambahan pula seiring dengan peningkatan jumlah kolam renang baharu serta bertambah minat dalam aktiviti kesukanan intensiti tinggi.

1.7 ORGANISASI KAJIAN

Bab 1: Latar belakang kajian dinyatakan di dalam bab ini bagi memberi gambaran tentang kepentingan kajian perlu dijalankan. Bahagian objektif menerangkan hala tuju kajian secara keseluruhan di mana ia menjadi salah satu kayu ukur penting dalam kejayaan sesuatu kajian. Bahagian skop pula meletakkan had kajian, membantu dalam memandu arah kajian serta membezakan kajian ini berbanding dengan kajian-kajian yang sedia ada.

Bab 2: Bab ini bertujuan untuk membina pemahaman asas dalam konteks kajian ini. Teori-teori berkaitan penglihatan komputer, asas rangkaian saraf tiruan serta rangkaian saraf konvolusi. Selain dari itu, ia juga menerangkan tentang pembezaan ciri-ciri bagi tiga renangan yang asas. Juga disertakan tentang kajian ilmiah berkenaan kerja-kerja oleh penyelidik yang sebelum ini sebagai panduan dalam membina rangka kajian serta melakukan sumbangan dari segi penambahbaikan terhadap kajian-kajian yang sedia ada.

Bab 3 : Segala rekabentuk serta rancangan dalam menjalani kajian dinyatakan dalam bab ini. Ia menerangkan juga susun atur perjalanan kajian yang melibatkan penglihatan komputer serta penggunaan rangkaian saraf konvolusi melibatkan video.

Bab 4: Bab ini melaporkan hasil kajian berdasarkan model yang dibina terhadap set data yang dihasilkan. Melalui keputusan ini, perbandingan dapat dilakukan terhadap model yang sedia ada di mana ia membuka ruang untuk mencari

kelemahan dalam model serta membantu dalam mencari idea untuk penambahbaikan terhadap model dalam kajian ini.

Bab 5: Bab ini merumuskan segala dapatan kajian berdasarkan keseluruhan kajian dan analisis yang telah dijalankan. Bagi memastikan kajian relevan serta memberi sumbangan dalam bidang akademik, bahagian ini mencadangkan penambahbaikan yang boleh dilakukan untuk kajian akan datang.

PUSAT SUMBER FTSM

BAB II

KAJIAN LITERATUR

2.1 PENGENALAN

Bagi tujuan mengukuhkan pemahaman serta membantu dalam mengembangkan kajian, bahagian ini menerangkan latar belakang kepada asas-asas disiplin kepintaran buatan (AI). Dengan menjelaskan silang kaitan antara bidang seperti pembelajaran mesin, matematik komputer, rangkaian neural sehingga menerangkan bagaimana mereka akan berkaitan antara satu sama lain. Selanjutnya ditunjukkan perkaitan antara bidang-bidang tersebut sehingga terhasilnya satu bentuk rangkaian yang khusus digunakan secara meluas dalam bidang klasifikasi imej serta video iaitu Rangkaian Konvolusi Neural atau juga dikenali sebagai CNN. Bagi tujuan memudahkan pemahaman bagaimana penggunaan CNN adalah berkaitan dengan kajian ini, disertakan juga secara ringkas berkaitan kaedah-kaedah renangan yang digunakan secara rekreasi mahupun kompetitif.

2.2 LATAR BELAKANG PERKEMBANGAN KEPINTARAN BUATAN

Kepintaran buatan dikatakan bermula secara giat pada 1950, apabila Alan Turing memperkenalkan kaedah pengujian terhadap sesuatu mesin dalam mengukur kepintaran sesebuah mesin. Sejak itu pengkajian AI bermula dengan model-model yang dibangunkan menggunakan sekumpulan peraturan untuk mengambil manfaat dari data yang telah disediakan bagi tujuan melatih mesin untuk berfikir. Bagaimanapun, pendekatan sebegini mempunyai had di mana ia tidak berupaya dalam membuat keputusan secara lebih umum, terutama melibatkan data yang tidak berstruktur, dan mengakibatkan pembangun perlu sentiasa mengubah peraturan-peraturan setiap kali data baru diperkenalkan.

Pembelajaran mesin beransur-ansur mengambil tempat dalam kajian kepintaran buatan di mana kaedah statistik serta kebarangkalian digunakan oleh penyelidik dalam membangunkan algoritma pembelajaran terhadap data. Mesin memperbaiki keupayaannya dalam mempelajari data sekaligus memperbaiki keupayaan pembelajaran secara berterusan. Banyak teori berkaitan diperkenalkan jauh lebih lama sebelum istilah pembelajaran mesin itu sendiri muncul, akan tetapi, beberapa halangan seperti keupayaan penyelidik dalam melakukan kerja-kerja pengiraan yang panjang lagi remeh serta amaun data yang amat besar menyebabkan ia kurang mendapat sambutan pada masa ia mula diperkenalkan. Akan tetapi, selari dengan peningkatan dalam pembangunan perkakasan komputer berdasarkan Hukum Moore (Schaller, 1997), ia membuka semula laluan untuk penyelidik menerokai semula pilihan-pilihan tersebut dan mengembangkan lagi kaedah pengiraan yang baru yang menjadikan pembelajaran mesin suatu pendekatan yang mantap dalam pembangunan bidang kepintaran buatan hingga ke hari ini.

2.3 PENEMUAN PENGETAHUAN DAN DATA RAYA

Penggunaan komputer secara meluas dalam kehidupan seharian menjadikan data tersedia untuk tujuan penggunaan dalam analisis meningkat. Penggunaan internet dengan data yang disimpan dalam pangkalan data melonjakkan lagi amaun data yang disimpan dan diarkibkan dengan kadar yang lebih pantas. Mutakhir ini pula, penerimaan masyarakat terhadap internet benda (IoT) menjadikan kadar peningkatan bertambah secara lebih pesat lagi. Ketersediaan data dalam amaun yang besar, atau lebih dikenali sebagai data raya, memberi kelebihan dari segi statistik untuk memahami sesuatu fenomena melalui penemuan pengetahuan pangkalan data (Fayyad et al., 1996).

Data-data seperti transaksi pembelian, profil pengguna, rekod kewangan, data metrologi, dan sebagainya, disimpan dalam storan dan dapat digunakan oleh penyelidik untuk analisis. Seterunya kerja-kerja penyelidikan dalam bentuk ramalan seperti meramal tindak balas pengguna terhadap perubahan sesuatu produk ataupun kelakuan tindak balas trafik terhadap perubahan cuaca – data sebegini kemungkinan disimpan dalam bentuk berstruktur. Potensi yang besar, serta sama penting dalam bentuk semi berstruktur serta tidak berstruktur seperti yang dicadangkan (Rusu et al.,

2013), di mana data-data contohnya dalam bentuk teks, audio, imej, dan video yang dirakam, juga disimpan dalam storan, perlu diberi perhatian yang dapat memberikan sumbangan terhadap pelbagai bidang selain dari tujuan perniagaan. Ia juga membantu dalam disiplin lain seperti kesihatan seperti pengecaman sel kanser pada satu sel tumor, ataupun keselamatan dalam membangunkan sistem bantuan kenderaan bagi mengelakkan kemalangan.

Dapat disimpulkan bahawa, kewujudan data raya yang tersimpan, samada berstruktur ataupun tidak berstruktur, kedua-duanya mempunyai potensi yang besar sekiranya dapat dimanfaatkan. Penggunaan pelbagai kaedah penemuan pengetahuan melalui perlombongan data, seterusnya penggunaan kaedah-kaedah pembelajaran mesin, memberikan peluang yang lebih luas bagi manusia untuk mencapai kesejahteraan hidup serta kemajuan dalam pelbagai bidang.

2.4 KATEGORI ALGORITMA PEMBELAJARAN MESIN

Pembelajaran mesin merupakan tulang belakang dalam bidang kepintaran buatan yang terkini di mana pembelajaran mesin menggunakan pendekatan kaedah statistik terhadap set data yang besar. Melalui pembelajaran corak dari set data secara menyeluruh, mengenalpasti corak dari data, dan seterusnya pemilihan algoritma yang bersesuaian dapat meramal hasil secara prediktif dan kemudiannya memberikan cadangan yang berbentuk preskriptif. Takrifan pembelajaran mesin (Samuel, 1988) sebagai suatu bidang di mana sesuatu mesin mempunyai kebolehan untuk belajar berdasarkan pengalaman iaitu data tanpa campur tangan dalam pengubahsuaian program secara berterusan. Ini diungkapkan secara lebih jelas (M. I. Jordan & T. M. Mitchell, 2015) dengan menyatakan pembelajaran mesin sebagai suatu program komputer yang mempelajari dari pengalaman E terhadap sesuatu tugas T dan capaian prestasi P , yang mana capaian prestasi P akan menjadi bertambah baik dengan pertambahan dalam pengalaman E . Ia memberi penyelesaian kepada model-model kepintaran buatan yang awal yang lebih sebagai kepintaran buatan simbolik. Senario pembelajaran mesin menjadi lebih menarik dengan kepelbagaian algoritma yang setiap satunya terbahagi kepada sepuluh kumpulan. Secara amnya, pembelajaran

mesin dibahagikan kepada dua kelompok, iaitu pembelajaran diselia atau pembelajaran tidak diselia.

2.4.1 Pembelajaran diselia

Pembelajaran diselia berlaku apabila mesin didedahkan kepada jawapan di mana mesin perlu memeriksa melalui perbandingan sama ada ia telah melakukan pengiraan menghampiri jawapan tersedia. Secara ungkapan, fungsi pembelajaran diselia mengambil bentuk:

$$Y = f(X) \quad \dots(1)$$

Sekiranya X ialah jujukan nilai pemboleh ubah masukan kepada algoritma, Y ialah nilai yang sepatutnya diperolehi hasil dari pengiraan. Pembelajaran mula berlaku apabila mesin akan mengukur sejauh mana nilai yang didapati menumpu kepada nilai yang dijangka iaitu Y . Perbezaan terhadap nilai keluaran berbanding nilai data dipanggil sebagai ralat. Mesin akan melakukan semula proses pengiraan secara berjujuk sehinggalah nilai ralat dapat diminimumkan. Kebanyakan algoritma pembelajaran mesin tergolong di dalam kumpulan ini seperti regresi linear, hutan rawak serta mesin sokongan vektor (SVM).

2.4.2 Pembelajaran tak diselia

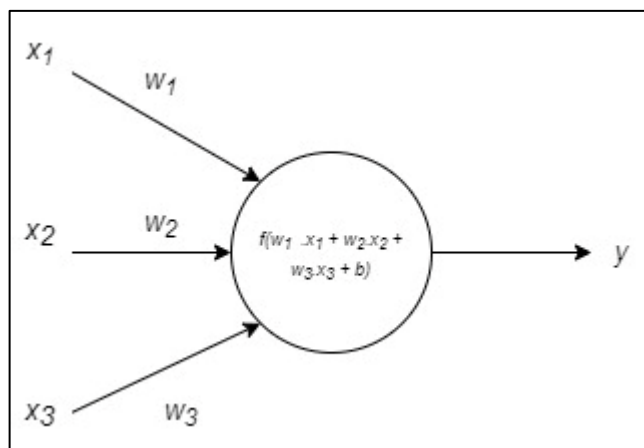
Pembelajaran tak diselia menggunakan set data berdasarkan sesuatu peristiwa, akan tetapi ia tidak dibekalkan dengan jawapan sebagai rujukan betul atau salah. Akan tetapi, ia memperolehi pengalaman dengan mencari corak tersembunyi melalui set data yang dibekalkan. Maka ia seringkali digunakan oleh penyelidik dalam tugas-tugas seperti mencari corak taburan dalam data belum berlabel, membeza kumpulan melalui ciri-ciri, maka ia sesuai digunakan dalam penerokaan pengetahuan baharu. Dua jenis pembelajaran tak diselia adalah pengelompokan serta penyekutuan.

2.4.3 Pembelajaran perkukuhan

Pembelajaran perkukuhan ialah salah satu dari sub-bidang pembelajaran mesin melibatkan agen yang dilatih untuk membuat keputusan dalam persekitaran tertentu dengan matlamat mendapat ganjaran yang maksimum (Geron, 2019). Ia berdasarkan idea percubaan dan kesilapan di mana agen belajar dengan berinteraksi dengan persekitaran dan menerima maklum balas dalam bentuk ganjaran atau penalti untuk tindakannya. Matlamat agen adalah untuk mempelajari dasar yang merupakan pemetaan daripada keadaan kepada tindakan yang memaksimumkan ganjaran terkumpul yang dijangkakan dari semasa ke semasa. Algoritma pembelajaran perkukuhan menggunakan pelbagai teknik untuk meneroka persekitaran dan mempelajari dasar yang optimum, seperti Q-pembelajaran, kecerunan dasar dan pengkritik aktor. Pembelajaran perkukuhan telah digunakan dalam banyak domain, termasuk robotik, permainan dan sistem pengesyoran.

2.5 RANGKAIAN NEURAL BUATAN

Rangkaian neural buatan mendapat tempat yang meluas dalam kajian bidang pembelajaran mesin masa kini, ianya telah bermula lebih awal lagi kajian mengenai perseptron oleh seorang neurofisiologi (McCulloch & Pitts, 1990). McCulloch, telah mencadangkan model neuron dari otak manusia yang dipermudah. Binaan asas perseptron terdiri dari beberapa bahagian asas seperti ditunjukkan Rajah 2.1. Model ini juga dikenali sebagai perseptron lapisan tunggal, terdiri dari masukan (x), pemberat (w) dan bias (b), jumlah wajaran (y) serta fungsi pengaktifan.



Rajah 2.1 Model asas perseptron (McCulloch & Pitts, 1990)

Model ini boleh juga dapat diwakilkan oleh persamaan

$$y = f(w_1x_1 + w_2x_2 + w_3x_3 + b) \quad \dots(2)$$

di mana,

y ialah keluaran dari perseptron

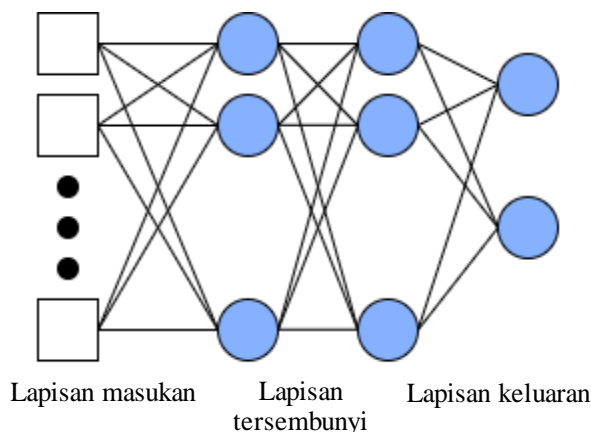
f ialah fungsi pengaktifan

w_1, w_2, w_3 , ialah pemberat bagi masukan

x_1, x_2, x_3 , ialah nilai masukan

b ialah terma bias

Penambahbaikan terhadap model awal McCulloch (White & Rosenblatt, 1963) dibuat dengan menunjukkan bahawa neuron buatan berupaya untuk belajar dari data sekiranya data dapat memenuhi persamaan (2). Dengan mencadangkan perseptron berbilang lapis (MLP), Rajah 2.2, bagi tujuan memperbaiki kekurangan tersebut. MLP terdiri dari gabungan pelbagai perseptron tunggal yang terhubung antara satu sama lain pada setiap lapisan. MLP mempunyai sekurang-kurangnya tiga lapisan iaitu lapisan masukan, lapisan tersembunyi serta lapisan keluaran.



Rajah 2.2 Model MLP

Selanjutnya, satu konsep perambatan balik dicadangkan (Rumelhart & Hintont, 1986) dan ia menjadikan rangkaian neural kembali menarik perhatian khalayak penyelidik. Perambatan balik menjadikan kitaran pengiraan dalam satu kitaran tertutup dengan pemberatan akan diubah suai berdasarkan ralat yang diperolehi pada penghujung kitaran pengiraan. Perambatan balik menjadikan pembelajaran model rangkaian neural boleh dilakukan dengan wujudnya mekanisme pembetulan nilai pemberatan supaya model menumpu dalam mewakili data dengan ralat yang paling minimum. Model boleh melakukan pengubahsuaian pada nilai pemberat, w , serta bias, b , supaya pembetulan boleh berlaku sepanjang masa pembelajaran. Pendekatan kalkulus pembezaan diperkenalkan dalam model dikenali sebagai jujukan penurunan kecerunan sehingga model mencapai nilai minimum ralat dapat dicapai. Nilai ralat yang minimum, menunjukkan model dapat menjangka nilai keluaran dengan tepat daripada pengiraan yang dilakukan di dalam lapisan-lapisan yang tersembunyi dengan ralat yang rendah. Semua langkah pengiraan dalam perambatan ke hadapan, perambatan balik serta penurunan kecerunan berlaku secara berterusan sehinggalah keadaan mantap dicapai. Keseluruhan kitaran memerlukan kuasa komputeran tinggi, di mana, mesin perlu melakukan kerja-kerja pengiraan serta kemaskini memori dalam kapasiti yang tinggi secara berterusan. Peningkatan kuasa komputeran yang semakin meningkat mengikut hukum Moore, menjadikan penggunaan rangkaian neural semakin mendapat tempat di kalangan penyelidik dari pelbagai bidang dari kajian sains tulen, kejuruteraan sehinggalah bidang ekonomi dan politik.

2.5.1 Pembelajaran Mendalam

Banyak pilihan kaedah pembelajaran mesin boleh dilakukan untuk data-data berstruktur dan berjadual. Bagi data tidak berstruktur, tidak banyak pilihan pendekatan pembelajaran mesin seperti mesin sokongan vektor. Seperti contoh, mesin sokongan vektor, SVM berkeupayaan melakukan tugas pengelasan dengan baik, akan tetapi, SVM memerlukan pengguna untuk melakukan kerja-kerja kejuruteraan ciri terlebih dahulu menyebabkan tempoh bekerja dengan data menjadi panjang. Akan tetapi, bagi pembelajaran mendalam, kerja-kerja kejuruteraan ciri tidak diperlukan di mana ia berlaku secara dalam rangkaian, sekaligus menjimatkan masa di samping mengelakkan kesilapan-kesilapan yang mungkin boleh berlaku semasa kerja-kerja kejuruteraan ciri.

Pembelajaran mendalam mengambil asas rangkaian neural tiruan (ANN) di mana model pembelajaran mendalam akan mempunyai beberapa lapisan rangkaian. Tiada panduan yang muktamad dalam menyatakan berapa lapisan supaya sesuatu model itu boleh dipanggil model pembelajaran mendalam, tetapi, secara umum, model yang mempunyai lebih dari dua lapisan boleh diistilahkan sebagai rangkaian pembelajaran mendalam. Mengambil langkah pengiraan yang sama dengan rangkaian asas ANN, pembelajaran mendalam mengambil kesempatan dalam pembangunan pesat perkakasan komputer dengan membenarkan pengiraan dalam jumlah unit titik terapung (FPU) yang tinggi. Pembelajaran mendalam telah digunakan dalam pelbagai bidang seperti perubatan, terjemahan bahasa, keselamatan siber, pasaran saham sehinggalah astrofizik. Pembelajaran mendalam terdiri dari beberapa binaan seperti rangkaian saraf berulang (rNN), ingatan jangka pendek-panjang (LSTM), pembelajaran peneguhan mendalam serta rangkaian neural konvolusi (CNN) yang akan digunakan dalam kajian ini.

2.6 LATARBELAKANG PENGLIHATAN KOMPUTER

Manusia melihat dengan cahaya dalam bentuk gelombang elektromagnet yang dipancarkan kepada retina lalu ditukar kepada isyarat elektrik. Isyarat elektrik tersebut akan dipindahkan melalui saraf untuk ditafsirkan oleh otak. Pengkajian mengenai penglihatan menggunakan mata telah bermula dengan pengkajian awalan terhadap

mamalia dimulai kucing (Hubel, 1982) dan diikuti pula kajian terhadap monyet. Dengan meletakkan sensor pada saraf tertentu mereka mendapati penglihatan pada mamalia secara amnya termasuk manusia ditafsirkan oleh korteks visual yang menerima denyutan isyarat elektrik dari retina. Kajian ini membantu dalam asas penglihatan komputer, dengan teknologi sensor seperti CCD dan CMOS digunakan untuk menangkap cahaya yang dikesan. Cahaya dikesan pula akan diterjemahkan kepada isyarat elektrik yang seterusnya akan diproses oleh pemproses lalu ditukarkan ke dalam bentuk piksel yang akan menjadi bentuk data yang difahami oleh mesin.

Bagi penyelidik bidang komputer pula, pelbagai kajian dilakukan untuk memahami keupayaan penglihatan manusia sambil menyesuaikan pengetahuan mereka kepada sistem komputer dalam bidang dikenali sebagai pemprosesan imej digital. Pemprosesan imej digital (Gonzalez & Woods, 2018) bermula dalam makmal di mana kerja-kerja penyelidikan dijalankan untuk sistem pendorongan jet. Ia digunakan untuk mendapatkan imej gambar yang terbaik dari sebuah prob angkasa lepas. Masalah berlaku apabila penghantaran imej dari prob angkasa lepas tersebut tidak memberikan imej yang memuaskan hasil dari keterhadan dalam keupayaan pemindahan data dan ia memerlukan pendekatan pemprosesan imej digital pada zaman tersebut digunakan. Sejak dari itu, bidang pemprosesan imej data berkembang pesat meliputi pelbagai bidang seperti perubatan, ketenteraan, pengimejan satelit dan pelbagai lagi.

Keupayaan manusia dalam melihat berlaku secara semula jadi dan ia membantu manusia mempelajari alam sekelilingnya. Melalui keupayaan melihat melalui mata lalu ditafsir oleh otak menyimpan memori objek yang dilihat, keupayaan ini menjadikan manusia sebagai makhluk yang mampu belajar. Tetapi, manusia mempunyai kekurangan dalam melihat sesuatu corak dari sejumlah data. Ia disebabkan otak manusia tidak berkeupayaan mengingat segala sesuatu secara kekal, apatah lagi mentafsir corak dari apa yang dilihatnya. Di sini, penggunaan komputer berpotensi untuk menampung kekurangan manusia dalam menyimpan memori yang tidak relevan dalam jangka masa yang panjang, dalam kuantiti yang besar. Memetik pandangan:

“Pada tahap yang lebih abstrak, matlamat dalam masalah penglihatan komputer adalah digunakan untuk membuat kesimpulan tentang dunia sekeliling.” (Prince, 2012)

Pemahaman yang terbina bermula dari 1960-an telah membantu manusia dalam memahami penglihatan lalu menggabungkannya dengan kaedah-kaedah terkini melibatkan pendekatan pembelajaran mesin serta pembelajaran mendalam. Ia kemudiannya dipindahkan kepada teknologi-teknologi seperti pengimejan perubatan, penderiaan jauh, *scanning electron microscope* (SEM), dan sebagainya.

2.6.1 Algoritma Pengesanan Ciri Penglihatan Komputer

Pada peringkat permulaan, khalayak penyelidik menggunakan pendekatan pengesanan ciri bagi tujuan mewakili bentuk serta tekstur imej menggunakan algoritma komputer. Algoritma komputer tersebut dibina supaya pengesanan objek dapat dilakukan dengan mengenalpasti garisan sempadan sesebuah objek. Lowe (1999) mencadangkan penggunaan pemindahan ciri skala-invarian (SIFT) untuk mengesan, ciri tempatan dalam sesebuah imej. SIFT mengesan titik-titik penting menggunakan nilai vektor yang mewakili imej rangkuman. Pendekatan ini dirujuk serta digunakan secara meluas oleh penyelidik penglihatan komputer selepasnya memandangkan keupayaannya dalam mengenalpasti objek walaupun dalam keadaan objek diputar pada sudut yang berlainan.

Selain SIFT, histogram kecerunan orientasi (HOG) pula diperkenalkan oleh Dalal (2005), untuk mewakili bentuk dan tekstur objek pada sesebuah imej. Ia membahagikan imej kepada bahagian-bahagian yang kecil lalu pengiraan histogram kecerunan terhadap setiap bahagian. Histogram yang terhasil kemudiannya disatukan bagi membentuk vektor ciri untuk mengesan dan mengenali imej. HOG mempunyai kelebihan dalam mengesan tubuh badan manusia di dalam imej serta video. Kedua-dua SIFT dan HOG digunakan dalam pelbagai aplikasi seperti pengesanan imej, penjejakan objek serta mengenali wajah sebelum peningkatan penggunaan CNN.

2.6.2 Pengesanan Ciri Melalui Penapis yang Dipelajari

Lowe (2004) dan Dalal (2005) menyatakan kelemahan penggunaan kedua-dua algoritma pengesanan ciri SIFT dan HOG di mana kedua-duanya direka untuk mengesan ciri yang terhad pada sesuatu imej seperti titik dan sempadan. Keduanya tidak berkeupayaan untuk mengesan corak atau ciri pada tahap yang lebih kompleks. Selain itu, kedua-dua SIFT dan HOG sukar untuk digunakan terhadap set data yang lebih besar terutama dalam tugas pengkelasan imej.

Pengenalan kepada CNN oleh LeCun et al., (2010) memberikan penyelesaian terhadap masalah yang dihadapi oleh penyelidik dalam bidang komputer. Beberapa set penapis yang dipelajari (Ji, 2013), ataupun kernel konvolusi digunakan untuk pengesanan ciri terhadap imej dan menghasilkan peta ciri dalam satu rangkaian neural buatan. Pengesanan ciri berlaku melalui beberapa siri lapisan konvolusi dan diikuti oleh fungsi pengaktifan bukan linear dan lapisan pengumpulan. Fungsi pengaktifan memperkenalkan bukan lineariti kepada rangkaian membenarkan rangkaian untuk mempelajari kaitan antara ciri. Peta ciri yang diperolehi melalui operasi konvolusi memberikan corak yang lebih difahami komputer dan ciri-ciri spesifik akan memberikan peta ciri dengan corak yang hampir sama. Ini adalah seperti yang dibuktikan oleh Derpanis (2012), di mana melalui operasi dekonvolusi yang dilakukan terhadap 20,000 klip pada lapisan-lapisan yang dipilih, klip-klip tersebut menghasilkan peta ciri yang hampir sama bagi konteks yang sama.

Disebabkan pengesanan ciri mengumpulkan ciri-ciri penting daripada imej yang perlu dipelajari oleh rangkaian, beberapa parameter dikenalpasti mempengaruhi dalam keupayaan pembelajaran oleh lapisan konvolusi. Sebelum penggunaan CNN, Lowe (2003) telah menjelaskan keupayaan pengesanan objek dipengaruhi oleh beberapa faktor seperti pengherotan *affine*, kewujudan hingar, dan perubahan pencahayaan. Berdasarkan Veeriah et al., (2016), pengecaman pergerakan pula adalah mencabar memandangkan variasi yang pelbagai seperti sudut kamera, keadaan pencahayaan dan latar belakang bersepadu merupakan faktor yang perlu dipertimbangkan dalam pembinaan sesuatu model. Ini juga disokong oleh Wu (2019) dengan menyatakan gabungan beberapa faktor menyebabkan kekeliruan dalam pengkelasan lalu

menyebabkan pengkelasan yang salah. Maka, dapat dinyatakan bahawa pembinaan model pengkelasan imej serta jujukan imej mementingkan kepada proses penyediaan set data secara teliti untuk memastikan lapisan pembelajaran dapat mempelajari ciri-ciri yang penting dengan baik. Faktor-faktor seperti sudut pandangan kamera, keadaan latar belakang, hingar, pencahayaan dan sebagainya perlu dipertimbangkan untuk memperolehi model pengkelasan yang berkesan. Maka untuk kajian ini, beberapa kriteria akan ditetapkan terutama untuk meminimalkan kesan dari faktor-faktor seperti pandangan kamera, latar belakang, serta hingar yang akan menjejaskan prestasi model yang dibina.

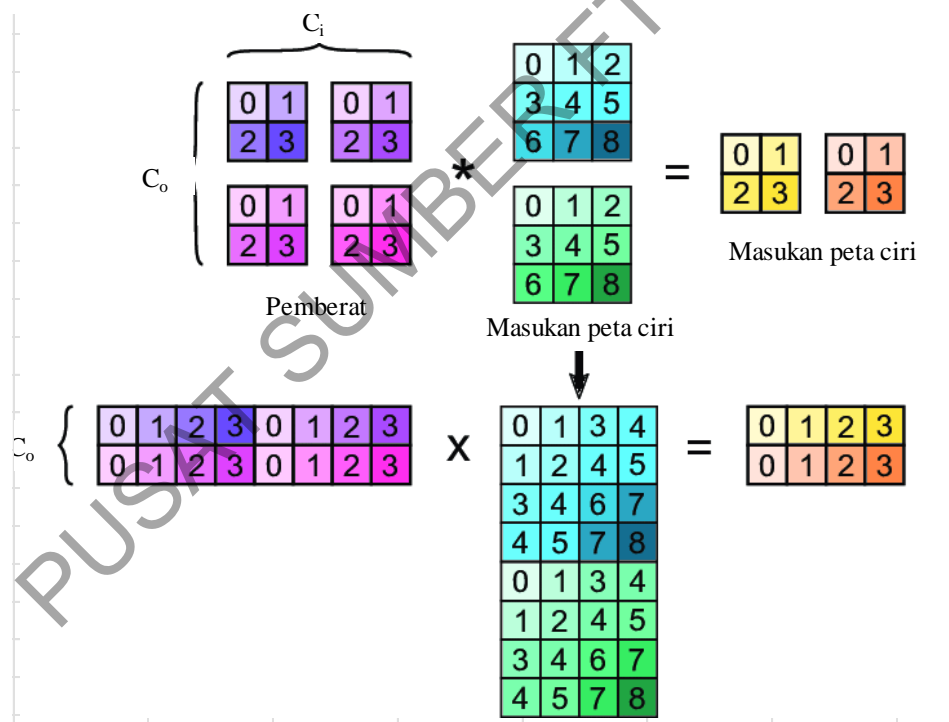
2.7 TEORI RANGKAIAN NEURAL KONVOLUSI (CNN)

Peradaban manusia menyaksikan perkembangan teknologi yang semakin pesat setiap hari. Manusia bijak dalam menggunakan sumber serta pengetahuan untuk meneroka bidang yang baru dan salah satunya adalah keupayaan meniru fenomena-fenomena kejadian alam lalu menterjemahkannya kepada teknologi atau lebih dikenali sebagai *biomimetics*. Salah satu idea bidang yang masih lagi dikagumi sehingga kini adalah dalam kajian berkaitan saraf neuron, serta penglihatan melalui saraf korteks dan penyesuaian terhadap pengetahuan dalam penglihatan komputer. Rangkaian neural konvolusi CNN adalah satu hasil dari kajian berkaitan pengetahuan biologi rangkaian saraf dan penglihatan komputer. Rangkaian neural konvolusi mempunyai latar belakang dalam gabungan beberapa konsep matematik fungsi.

2.7.1 Asas model rangkaian neural buatan

Konvolusi dalam matematik secara amnya merujuk kepada gabungan dua fungsi matematik dengan menjalankan satu operasi terhadap kedua-dua fungsi dan menghasilkan satu fungsi baru. Misalnya sekiranya dipertimbangkan dua fungsi $f(x)$ dan $g(x)$, selepas dilakukan konvolusi terhadap kedua-dua fungsi, ia akan menghasilkan fungsi ketiga boleh juga ditulis sebagai $g[f(x)]$. Pemetaan kepada fungsi yang baharu melalui hasil darab ataupun tambah bagi kedua-dua fungsi awalan seperti $(f + g)$ dan $(f * g)$. Operasi konvolusi banyak digunakan dalam kajian pemrosesan isyarat, pemrosesan data digital, kajian akustik, dan juga pemrosesan imej.

Konsep kedua berkait rapat dengan CNN adalah konsep algebra linear, di mana algebra linear digunakan sebagai matematik asas dalam pemprosesan imej di mana imej yang mempunyai unit asas piksel matriks, Rajah 2.3, diungkapkan dalam bentuk piksel matriks. Matriks pemberat, dengan saiz $C_o \times C_i$, sekiranya didarab dengan satu matriks masukan, akan menukar nilai-nilai kedua-dua matriks kepada satu matriks keluaran. Begitu juga operasi-operasi algebra linear boleh digunakan terhadap matriks piksel untuk pelbagai tujuan dari penapisan, pengesanan, termasuk juga efek-efek khas. Antara operasi yang terlibat terhadap matriks piksel adalah penambahan, penolakan, dan pendaraban matriks. Maka, konvolusi, dalam konteks ini adalah bila dua vektor yang boleh diwakilkan sebagai fungsi matriks $f(x)$ dan $g(x)$ dikatakan telah dikonvolusi, apabila operasi-operasi algebra linear dikenakan ke atas mereka.

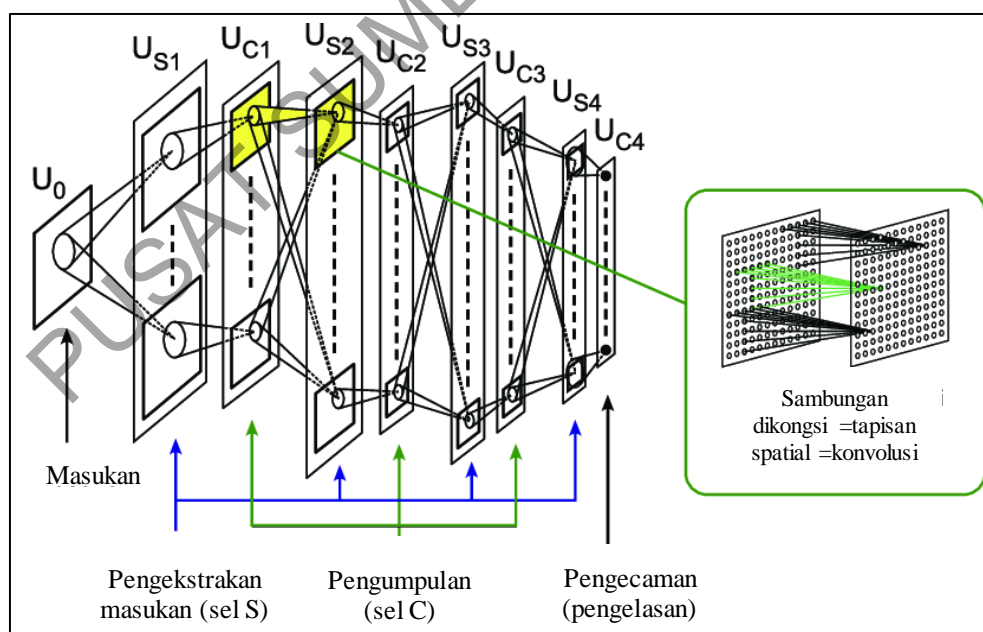


Rajah 2.3 Konvolusi kepada matriks melalui operasi darab (Capra et al., 2020)

2.7.2 Latar belakang serta seni bina rangkaian CNN

Menggunakan struktur seni bina rangkaian telah dicadangkan satu rangkaian *neocognitron* (Fukushima, 1980), Rajah 2.4, yang terdiri dari sel mudah, S , dan sel rumit, C , disusun secara melata dan mempunyai keupayaan untuk mengenali rangsangan dalam bentuk isyarat visual. Kajian ini telah merintis kepada kajian-kajian

lain seperti yang disusuli oleh satu seni bina rangkaian (LeCun et al., 2010) dibangunkan dikenali sebagai LeNet bagi tujuan mengenal pasti tulisan tangan pada cek-cek bank. Sejak itu, penggunaan CNN semakin berkembang dalam pengecaman imej, (Krizhevsky, 2007) menggunakan AlexNet bagi 1.2 juta imej dari 1000 kelas berbeza bagi pertandingan ILSVRC-2012. VGG14 pula diperkenalkan (Simonyan & Zisserman, 2015) di mana satu rangkaian CNN yang lebih mendalam telah dicadangkan tetapi menggunakan penapis bersaiz kecil dengan 3x3 pada kedalaman sehingga 19 lapisan. Walaupun dengan saiz tapisan yang kecil, VGG14 mampu membuat generalisasi dengan baik terhadap set data berlainan dari ImageNet. Szegedy et al. (2015) mencadangkan seni bina bernama *Inception* yang kemudiannya digunakan dalam ILSVRC14 juga dikenali sebagai GoogLeNet. GoogLeNet terdiri dari rangkaian dengan kedalaman sebanyak 22 lapisan. Ia dibangunkan dengan tujuan membina rangkaian yang berkeupayaan untuk menandingi rangkaian yang lebih 'mahal' dalam masa yang sama memberikan hasil yang lebih baik dalam tugas pengecaman imej.



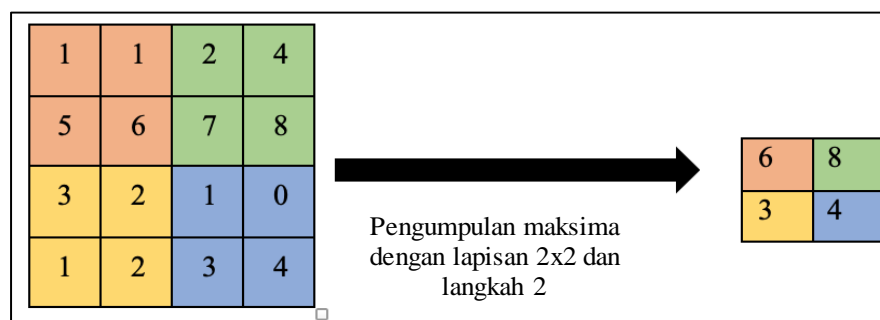
Rajah 2.4 Senibina rangkaian *neocognitron* (Fukushima, 1980)

Dalam cubaan untuk mengurangkan kos dalam melatih rangkaian, He et al. (2016) telah mengatur semula lapisan pembelajaran dengan memperkenalkan ResNet. Satu blok lebih diperkenalkan bagi mengatasi masalah kecerunan melenyap sisa dalam kebanyakan rangkaian mendalam. ResNet telah mendapat perhatian dalam

kalangan penyelidik dan digunakan secara meluas dalam penglihatan komputer di mana ia juga menjadi asas kepada rangkaian CNN yang baharu. Rangkaian neural konvolusi (CNN) diwakili dalam bentuk aturan piksel-piksel dalam matriks serta tensor.

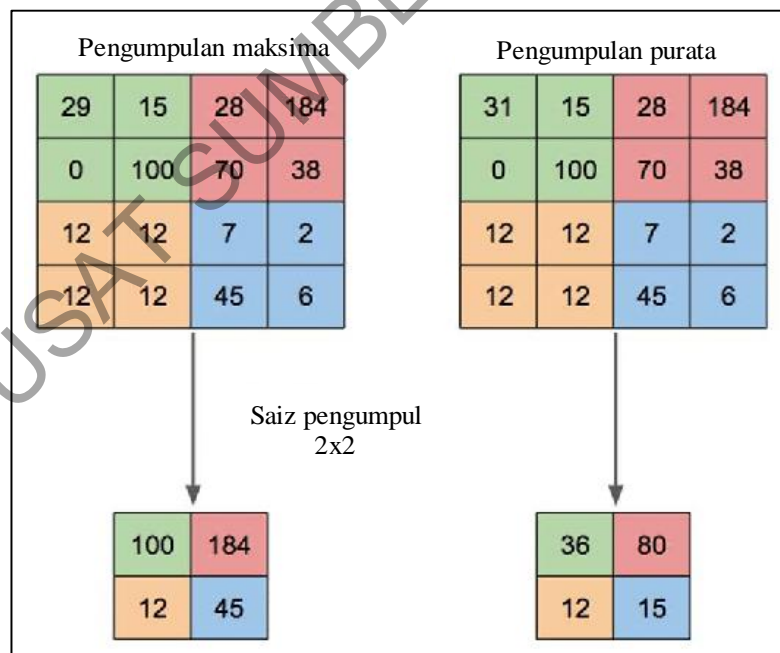
2.7.3 Lapisan konvolusi

Rangkaian saraf konvolusi terdiri dari pelbagai seni bina, akan tetapi, secara asasnya, rangkaian saraf konvolusi perlu digabungkan dengan lapisan terdiri dari empat elemen asas iaitu bahagian masukan, bahagian lapisan konvolusi, lapisan pengumpulan, serta lapisan tersambung sepenuhnya. Bahagian masukan adalah bahagian awalan di mana rangkaian menerima masukan data dalam bentuk jujukan matriks sebelum melalui lapisan-lapisan yang seterusnya. Lapisan konvolusi pula adalah teras bagi sebuah rangkaian saraf konvolusi. Ia terdiri daripada parameter berserta beberapa siri penapis yang akan digunakan sepanjang proses pembelajaran oleh model rangkaian. Lapisan konvolusi melakukan operasi berdasarkan saiz, F , serta langkah, S , seperti di Rajah 2.5, dengan tapisan pengumpulan maksima saiz 2×2 dan langkah sebanyak 2, memberikan satu matriks keluaran. Secara amnya, masukan akan melalui lapisan konvolusi seterusnya keluaran akan diperolehi pada bahagian peta ciri pada bahagian penghujung sesebuah rangkaian konvolusi ataupun lapisan tersambung sepenuhnya. Beberapa siri penapis akan berlaku di sepanjang sesuatu kerangka rangkaian konvolusi bergantung kepada kedalaman sesuatu rangkaian ataupun asas seni bina rangkaian tersebut.



Rajah 2.5 Contoh penggunaan penapis bersaiz 2×2 dalam 2 langkah (Syulistyo et al., 2020)

Lapisan pengumpulan mengurangkan saiz masukan dari lapisan konvolusi sebelum darinya. Terdapat dua jenis operasi pengumpulan yang dilakukan bagi tujuan tersebut, iaitu pengumpulan maksima dan pengumpulan purata berdasarkan Rajah 2.6. Pengiraan pengumpulan maksimum adalah dilakukan dengan mengambil nilai maksimum pada satu bingkai dalam sesuatu langkah, S . Merujuk kepada contoh di dalam rajah, dari keempat-empat nilai iaitu 29, 15, 0, 100 bagi set langkah yang pertama, nilai maksimum dari set tersebut adalah 100, maka nilai 100 akan diambil sebagai nilai maksimum yang akan dimajukan. Manakala, bagi pengumpulan purata, nilai purata akan dikira dari keempat-empat nilai dari sesuatu langkah, S . Contohnya, purata bagi empat nilai 31, 15, 0, 100 adalah 36, maka 36 adalah nilai yang akan dimajukan sekiranya menggunakan pengumpulan purata. Juga, diperhatikan di sini, dimensi akan dapat dikurangkan melalui penapisan yang seterusnya. Penggunaan penapis pengumpulan dapat memudahkan pengiraan yang perlu dilakukan semasa proses pembelajaran.



Rajah 2.6 Operasi menggunakan pengumpulan maksimum serta pengumpulan purata

Seperti model-model algoritma pembelajaran mesin yang lain yang mempunyai lapisan tersambung sepenuhnya berperanan untuk menggabungkan masukan di mana keluaran dari lapisan yang sebelumnya mengambil bentuk matriks 3x3 yang perlu diubah kepada vektor satu dimensi, input lapisan-lapisan yang

sebelumnya diratakan. Selepas dijumlahkan kepada matriks 1-dimensi pengiraan kebarangkalian kelas keluaran akan dilakukan dalam bentuk pengelasan, pengecaman dan seumpamanya.

2.7.4 RANGKAIAN 2DCNN & 3DCNN

Kedua-dua 2DCNN dan 3DCNN mendapat perhatian meluas khususnya berkaitan imej. Operasi konvolusi berlaku dalam lapisan konvolusi menurut Ji (2012), memberi keupayaan 2DCNN dan 3DCNN untuk mengekstrak ciri dari lapisan setempat pada peta ciri dari lapisan sebelumnya. Di samping itu, penambahan dan pengubahan bias akan berlaku sebelum melalui fungsi sigmoid. Menurut Ji (2012) juga, nilai bagi satu unit pada posisi (x, y) di dalam peta ciri ke- j dalam lapisan ke- i , ditulis sebagai v^{xy}_{ij} diberikan oleh persamaan

$$v^{xy}_{ij} = \tanh \left(b_{ij} + \sum_m \sum_{p=0}^{P_i-1} \sum_{q=0}^{Q_i-1} w_{ijm}^{pq} v^{(x+p)(y+q)}_{(i-1)m} \right) \quad \dots(3)$$

di mana,

$\tanh(\)$ ialah fungsi tangen hiperbola

$b_{i,j}$ ialah bias pada peta ciri pada lapisan ke- i tersambung

w_{ijm}^{pq} ialah nilai pada kedudukan (p, q) bagi kernel tersambung pada peta ciri ke- k

P_i dan Q_i adalah saiz kernel

Manakala bagi 3DCNN, berbanding imej yang dinyatakan dalam 2 dimensi iaitu ketinggian dan kelebaran imej, jujukan imej berturutan pula mempunyai dimensi tambahan, dan ia dinyatakan dalam konteks ruang dan tempoh ataupun kedalaman sekiranya melibatkan objek dalam 3-dimensi. Menurut Ji (2012) lagi, lapisan konvolusi 3D adalah terdiri dari lapisan gabungan konvolusi dalam bentuk kiub yang ditandakan antara satu sama lain. Ianya diberikan oleh persamaan

$$v_{ij}^{xyz} = \tanh \left(b_{ij} + \sum_m \sum_{p=0}^{P_i-1} \sum_{q=0}^{Q_i-1} \sum_{r=0}^{R_i-1} w_{ijm}^{pqr} v_{(i-1)m}^{(x+p)(y+q)(z+r)} \right) \quad \dots(4)$$

di mana,

$\tanh()$ ialah fungsi tangen hiperbola

$b_{i,j}$ ialah bias pada peta ciri pada lapisan ke- i tersambung

v_{ijm}^{pqr} ialah nilai pada kedudukan (p, q, r) bagi kernel tersambung pada peta ciri ke- m

P_i dan Q_i adalah saiz kernel

Kedua-dua 2DCNN dan 3DCNN adalah variasi CNN terdiri dari jenis rangkaian neural yang digunakan dalam pengkelasan data dan pemprosesan isyarat. Berbanding rangkaian neural tersambung sepenuhnya, kedua-duanya menggunakan operasi konvolusi terhadap piksel sama ada dalam imej, jujukan imej juga isyarat. Kedua-dua seni bina berkongsi asas yang sama, akan tetapi, 3DCNN adalah berdasarkan seni bina 2DCNN yang dilanjutkan dan kebiasaanya digunakan untuk pemprosesan data volumetrik seperti imej perubatan, video dan awan titik 3D berbanding 2DCNN yang digunakan kebiasaanya dalam pengkelasan imej. Bagi tujuan pengkelasan video, 3DCNN secara umumnya mempunyai kelebihan berbanding 2DCNN di mana 3DCNN mempunyai keupayaan mengenal dalam konteks ruang dan masa berbanding 2DCNN (Burton, 2019). Walaupun begitu, penggunaan 3DCNN memerlukan keupayaan pengiraan yang tinggi dengan mempertimbangkan saiz data serta operasi konvolusi yang memerlukan pergerakan kernel dalam 3 dimensi berbanding 2DCNN.

Kedua-dua seni bina juga, selain dari lapisan konvolusi, masukan dan keluaran, akan terbina dari beberapa lapisan yang lain seperti lapisan pengumpulan, lapisan rata, juga lapisan tersambung sepenuhnya. Ini merupakan ciri yang khusus

dalam satu-satu rangkaian neural buatan. Lapisan konvolusi menggunakan beberapa penapis upaya belajar yang melakukan operasi konvolusi lalu menjana peta fitur. Lapisan pengumpulan pula berfungsi dalam mengurangkan saiz dari masukan dengan mengambil nilai maksima atau purata pada satu-satu kawasan. Masukan-masukan dengan vektor pelbagai dimensi ini perlu diratakan kepada vektor satu dimensi, ini dilakukan oleh lapisan rata. Lapisan terakhir, sebelum lapisan keluaran adalah lapisan tersambung sepenuhnya, di mana semua neuron dalam lapisan ini tersambung kepada semua neuron dalam lapisan terdahulu.

2.8 PENGGUNAAN CNN DALAM PEMROSESAN IMEJ DAN VIDEO

CNN mempunyai sejarah pembangunan yang asasnya adalah diinspirasi oleh pemahaman-pemahaman fundamental dalam bidang biologi yang kemudiannya dijelmakan ke dalam aplikasi elektronik serta komputer. Dari beberapa jenis seni bina rangkaian saraf buatan setakat ini, CNN adalah seni bina yang paling banyak digunakan dalam kajian berkaitan imej dan video disebabkan kelebihan pada lapisan konvolusi dalam keupayaannya dalam memproses imej. Rangkaian saraf tiruan diketahui mempunyai kelebihan dalam bidang pembelajaran mesin dan kebanyakan masa mengatasi model-model pembelajaran mesin konvensional terutama bagi data yang lebih kompleks. Berdasarkan sumbangan oleh Fukushima (1980), di mana beliau telah memperkenalkan *neocognitron* dengan rangkaian saraf buatan pelbagai lapisan yang berupaya melakukan tugas pengesanan corak. Fukushima 1980 menggunakan model tersebut bagi tujuan pengesanan tulisan kanji Jepun yang ditulis menggunakan tangan. Ia memberi inspirasi pada kajian-kajian selepasnya seperti sumbangan LeCun (2010) yang mempelopori binaan asas rangkaian neural konvolusi yang digunakan secara meluas dan menjadi model asas bagi kajian-kajian yang menggunakan rangkaian konvolusi sehingga kini.

Sejak dari itu, semakin banyak kerja-kerja dilakukan dalam membangunkan seni bina CNN untuk tujuan pengelasan imej seperti AlexNet (Krizhevsky, 2007), GoogLeNet (Szegedy et al., 2015), VGGNet (Simonyan 2015). Ianya masih lagi berterusan sehingga hari ini di mana setiap satu mungkin berbeza antara satu sama lain dari segi seni bina dari segi kedalaman, susunan serta penambahan lapisan buatan tangan di mana setiap satu memberikan kelebihan yang tertentu bergantung kepada

kes atau set data. Selain dari pengelasan imej, CNN juga digunakan dalam bidang perubatan, seperti pengecaman sel-sel kanser (Wang et al., 2016), pengesanan kes Covid19 melalui imej X-ray dan CT-scan (Khattak et al., 2021), pengecaman wajah (Y. X. Yang et al., 2018), pengimejan geospasial (AL-Alimi et al., 2020) & (Tong et al., 2006). Secara ringkasnya, CNN mempunyai aplikasi yang meluas meliputi pelbagai bidang terutamanya berkaitan pemprosesan imej serta isyarat di mana kerja-kerja seumpama adalah di luar keupayaan manusia dalam konteks penerokaan pengetahuan yang kompleks.

2.9 PENGECAMAN PERGERAKAN MANUSIA

Tindak balas manusia terhadap sesuatu rangsangan dapat diperhatikan dengan pelbagai cara seperti pertuturan, penulisan, mimik muka, serta pergerakan tubuh. Pengkajian dalam pergerakan tubuh manusia diberi perhatian oleh pengkaji kognitif dan neuro-sains. Menurut Andersen (2003), pergerakan tubuh badan adalah selari dengan emosi yang dialami seseorang, maka penekanan diberikan khusus dalam pergerakan tubuh badan manusia di samping pengkajian terhadap ekspresi muka semata-mata. Tubuh badan menerima isyarat terus dari otak dan bertindak balas terhadap jenis-jenis rangsangan yang diterima. Maka, dengan memerhatikan pergerakan badan manusia, dapat dijangka tindak balas yang sedang dialami melalui perasaan berdasarkan pengalaman yang dilalui lalu menjangkakan tindakan-tindakan yang akan menyusul hasil dari tindak balas tersebut. Kelebihan pengetahuan tentang pergerakan-emosi, tingkah laku manusia boleh diperhatikan dan memberi kelebihan dalam pelbagai bidang seperti pengawasan keselamatan, kawalan kelompok, serta peramalan tingkah laku. Kaedah yang biasa dilakukan oleh pengkaji neuro-sains adalah dengan memerhati tindak balas manusia terhadap rangsangan sambil memantau tindak balas manusia pada bahagian otak.

Memandangkan keupayaan serta keberkesanan CNN dalam pemprosesan imej, aplikasi CNN dilanjutkan lagi kepada tugas-tugas berkaitan pemprosesan video dengan menggunakan premis bahawa video adalah terdiri dari jujukan imej-imej berturutan secara berterusan dalam tempoh-tempoh masa yang tertentu. Sebelum penggunaan CNN, pemprosesan serta analisa video telah bermula sejak sekian lama terutama bagi video berkaitan pergerakan tubuh badan manusia. Melalui pendigitalan

filem serta video. Perwakilan secara simbolik bagi pergerakan boleh dilakukan melalui pendekatan pengecaman corak serta analisa babak (Badler & Smoliar, 1979). Beliau mencadangkan untuk menggunakan model *Labanotation* dengan mewakili tubuh manusia sebagai satu kumpulan rangkaian pemprosesan di mana setiap bahagian sendi mewakili satu pemproses dan setiap pemproses menafsirkan pergerakan yang berbeza. Webb (1982) pula merujuk pergerakan jasad tegar serta gabungan jasad-jasad dan menggabungkan kedua-duanya bagi mewakili pergerakan tubuh badan manusia. Dengan mewakili tubuh badan manusia menggunakan lidi dan seterusnya silinder. Akita (1984) pula menggunakan model tersebut, lalu menghubungkannya dengan pergerakan bingkai yang berturutan. Ketiga-tiga kajian ini adalah antara usaha terawal dalam mendigitalkan pergerakan tubuh badan manusia sebagai persiapan ke arah analisa corak pergerakan. Ianya dilanjutkan dengan cubaan mewakili tubuh badan manusia supaya dapat digunakan dengan teknik-teknik analisis berdasarkan teknologi yang sesuai pada masa tersebut.

Kajian pengecaman pergerakan tubuh badan manusia menurut Aggarwal (1997) berkaitan halaju serta kedudukan boleh digunakan untuk penjodohan dengan urutan bingkai beserta intensiti imej terdiri dari teknik seperti pendekatan menggunakan penjodohan templat. Selain dari itu, Polana (1994) mencadangkan pergerakan tubuh manusia diwakili kepada jaringan 2D di samping mengira aliran optik sepanjang bingkai dan mewakilkannya dalam satah Cartesian. Berdasarkan pemahaman yang diperolehi dari kajian-kajian sebelumnya, Bobick (1996) telah mewakili pergerakan manusia menggunakan imej-sejarah-pergerakan (MHI) dalam bentuk imej statik, lalu templat dibina berdasarkan pergerakan tertentu. Pergerakan yang dianalisis akan dipadankan dengan templat yang telah dibina. Satu model Markov tersembunyi (HMM) pula dicadangkan dengan menggunakan imej berturutan dan mewakilkannya kepada bentuk urutan vektor (Yamato et al., 1992). Satu kategori pergerakan diwakili oleh set urutan vektor yang tertentu, ini akan digunakan untuk mengira kebarangkalian sesuatu pergerakan dapat dipadankan kepada set vektor mengikut klasifikasi yang telah dikelaskan terlebih dahulu.

Penentuan pergerakan dan pengecaman gaya juga dicadangkan (Bradski & Davis, 2000) di mana pergerakan diterjemahkan kepada bayang berlapisan. Ia

kemudiannya dikumpulkan menggunakan algoritma kecerunan sejarah pergerakan imej bermasa (*timed motion history image*, tMHI) yang telah diubahsuai (Bobick & Davis, 1996). Perwakilan titik-titik terpilih pada gambar atau partikel menggunakan algoritma peranjakan-purata dan mengumpulkan setiap partikel supaya dapat mewakili sesuatu objek dalam bingkai, kemudian ianya dipetakan ke dalam ruang tempoh-masa (Silva & Scharcanski, 2010). Satu cadangan algoritma yang berkeupayaan mengasingkan objek yang bergerak terhadap latar belakang yang statik dengan menentukan nilai ambang bagi setiap piksel supaya imej dalam bentuk skala kelabu (Zhu et al., 2008).

Pergerakan manusia semasa bersukan adalah antara analisis yang popular, seperti (Ballan et al., 2009) menggunakan video perlawanan bola sepak dengan menggunakan pendekatan beg perkataan (BoW) dengan terlebih dahulu mewakili setiap bingkai video sebagai satu karakter, kemudian setiap bingkai akan diwakili oleh satu histogram. Pengelasan dilakukan menggunakan SVM dengan teras perkataan di mana kaedah ini mengatasi pengelas kNN. Kaedah pembelajaran mesin menggunakan mesin vektor sokongan (SVM) diperbaiki pula dicadangkan (Xue-jun et al., 2012) menggunakan pendekatan pembelajaran aktif. Model yang dicadangkan berupaya mengekstrak video dengan baik walaupun dengan tahap hingar yang tinggi. Pengesanan peristiwa pula telah dicadangkan menggunakan rangkaian saraf berulang (RNN) jenis ingatan jangka pendek-panjang (LSTM) sejenis rangkaian buatan yang mempunyai memori dalaman untuk tujuan pengelasan aksi dalam video perlawanan bola sepak (Baccouche et al., 2010). Aksi-aksi dalam perlawanan dinyatakan menggunakan kaedah beg-perkataan (BoW) dan pergerakan dominan yang mewakili elemen pergerakan yang terbesar dalam sesuatu bingkai.

Kejayaan yang dicapai dalam kerja-kerja pemprosesan imej telah menginspirasi penyelidik untuk meneroka secara lebih lagi, ianya diikuti pula oleh kerja-kerja pemprosesan video menggunakan CNN. Berdasarkan fakta bahawa video adalah terdiri dari susunan gambar yang dimainkan secara berturutan, ini menunjukkan bahawa pemprosesan video akan berkongsi asas yang sama dengan pemprosesan imej. Sekiranya imej diwakili oleh nilai piksel, piksel video juga diungkapkan dengan mengaitkan pergerakan piksel berdasarkan ruang dan masa atau

juga dikenali sebagai maklumat *spatio-temporal*. Walaupun kedua-dua disiplin pemprosesan imej berkongsi asas yang sama, cabaran dalam pemprosesan video adalah lebih sukar, selain dari saiz data yang besar, pemprosesan video perlu mempertimbangkan banyak lagi parameter seperti perubahan intensiti, perubahan bentuk akibat pergerakan, pergerakan kamera, dan sebagainya. Pengenalan terhadap penggunaan CNN dalam pemprosesan video telah banyak membantu dalam kerja-kerja penyelidikan serta pembangunan teknologi berkaitan pemprosesan video.

Antara usaha awal yang diketahui dalam pengesanan pergerakan manusia dengan mencadangkan model konvolusi 3D yang mempunyai keupayaan dalam mempelajari ciri-ciri *spatio-temporal* dan menggabungkannya dengan rangkaian RNN-LSTM (Baccouche et al., 2010). Model ini telah digunakan untuk melakukan pengelasan video set data KTH yang mempunyai enam aktiviti pergerakan yang berbeza iaitu berjalan, berjoging, berlari, bertinju, melambai, serta menepuk tangan. Selain itu, satu rangkaian neural 3DCNN yang berupaya mengasingkan ciri-ciri dari dimensi ruang dan masa menggunakan lapisan konvolusi untuk menganalisis set data TRECVID, yang mengandungi video dari CCTV di lapangan terbang (Ji et al., 2013). Kajian ini mencadangkan penggunaan seni bina konvolusi 3D dengan saiz $7 \times 7 \times 3$ secara berselang-seli dengan lapisan yang lain.

Dalam kajian yang berlainan pula gabungan rangkaian CNN dan RNN juga digunakan tetapi untuk video dengan tempoh yang lebih panjang berbanding sebelumnya menggunakan set data Sports-1M dan UCF101 (Ng et al., 2015). Bagi rangkaian konvolusi, seni bina GoogLeNet (Szegedy et al., 2015) dan AlexNet (Krizhevsky, 2007) digunakan memandangkan keupayaan kedua-dua seni bina yang terbukti mempunyai pencapaian yang baik dalam pengelasan imej. Dengan mengambil satu bingkai sahaja bagi setiap saat video, aliran optik dari imej dihasilkan dan disua ke dalam rangkaian konvolusi dengan cara pelakuran lambat. Manakala penggunaan rangkaian CNN bagi set data video yang besar (Karpathy et al., 2014) dicadangkan yang menggunakan satu juta video dari laman perkongsian video, YouTube, dengan set data dikelaskan kepada 487 kelas. Beberapa jenis pelakuran digunakan untuk melakur maklumat melalui rangkaian – lakuran awal, lakuran lewat, serta lakuran perlahan, didapati lakuran lambat memberikan prestasi yang lebih baik

berbanding lakuran yang lain. Juga didapati, pendekatan satu bingkai per saat juga membantu dalam memberikan hasil yang memuaskan. Bagi set data yang sebegitu besar, rangkaian CNN pelbagai-resolusi dicadangkan di mana ianya mempunyai dua aliran iaitu aliran *fovea* aliran konteks. Masukan akan merambat melalui kedua-dua aliran secara selari di mana aliran *fovea* memproses bahagian tengah masukan yang beresolusi tinggi, manakala aliran konteks memproses imej resolusi rendah. Kemudian, keluaran dari kedua-dua aliran akan digabungkan sebelum memasuki lapisan tersambung sepenuhnya.

Menggunakan satu varian rangkaian 3DCNN juga dikenali 3D ConvNets (Tran et al., 2015), ianya mempunyai keupayaan yang lebih baik dalam mempelajari ciri *spatiotemporal* berbanding seni bina 2DCNN. Mereka mencadangkan penggunaan rangkaian 3DCNN dengan saiz kernel konvolusi yang kecil dengan meningkatkan kedalaman *kernel* konvolusi 3D adalah seni bina yang terbaik. Untuk tujuan itu, kedalaman *kernel* dipelbagaikan dan diuji dengan set data UCF101 bagi mengenal pasti kedalaman yang paling optimum. Manakala lapisan pengelasan pula terdiri dari jenis pengelasan linear yang ringkas.

Menurut Simonyan & Zisserman (2014), berdasarkan korteks penglihatan manusia terdiri dari dua laluan berasingan (Goodale & Milner, 2018) iaitu aliran *ventral* dan aliran *dorsal* – satu bertugas dalam mengenali objek dan yang kedua pula berperanan untuk mengesan pergerakan. Menurut mereka lagi

“Video can naturally be decomposed into spatial and temporal components. The spatial part, in the form of individual frame appearance, carries information about scenes and objects depicted in the video. The temporal part, in the form of motion across the frames, conveys the movement of the observer (camera) and the objects”

Maka satu rangkaian CNN dengan dua aliran selari dicadangkan untuk digabungkan dengan rangkaian *spatiotemporal* di mana aliran pertama adalah untuk mengesan objek, dan aliran yang kedua adalah terdiri dari aliran optik bertindan bingkai demi bingkai. Didapati kaedah ini lebih baik dibandingkan pendekatan

menggunakan bingkai-bingkai berturutan semata-mata. Kajian ini memberi inspirasi kepada kajian Yudistira & Kurita (2017) dengan mencadangkan adunan kepakaran melalui gerbang rangkaian neural konvolusi. Memandangkan dua modaliti iaitu *spatial* dan *temporal* dipisahkan kepada dua gerbang kepada dua aliran yang berbeza, kedua-duanya mempunyai nilai pemberatan yang berasingan, maka persoalan cuba dijawab dalam menentukan bagaimana nilai pemberatan dapat ditentukan untuk kedua-dua aliran yang berbeza. Melalui kajian ini, tiga model dicadangkan iaitu pelakuran dengan purata, pelakuran dengan SVM dan rangkaian gerbang. Untuk konvolusi, seni bina VGG-16 dan ResNet-50 (He et al., 2016) digunakan. Pada gerbang *spatial* nilai-nilai RGB akan dirambat melalui rangkaian, manakala pada gerbang *temporal* pula nilai-nilai mewakili pergerakan menggunakan aliran optik akan dirambat. Kedua-dua hasil dari setiap aliran akan digabungkan selepas lapisan tersambung sepenuhnya untuk memberikan keputusan pengelasan. Menggunakan set data UCF101 dan HMDB-51, didapati rangkaian gerbang CNN mencapai prestasi terkini. Rangkaian ini juga mungkin sesuai digunakan untuk penggabungan rangkaian yang mempunyai pelbagai mod, pelbagai resolusi, dan pelbagai sumber. Berdasarkan cadangan ini juga, Yudistira & Kurita (2020) mencadangkan *Correlation Net* yang menangkap hubung kait *spatiotemporal* berdasarkan tanda masa yang tertentu. Lapisan tersambung sepenuhnya menerima suapan dari lapisan *spatial* (RGB) dan *temporal* (aliran optik) yang dihubungkan terlebih dahulu sebelum suapan dengan hubungan dilakukan melalui pelakuran Shannon dengan mencari hubungan semantik antara kedua-dua aliran *spatial* dan *temporal*.

Rangkaian neural konvolusi mempunyai kelebihan dalam mengesan fitur samada dalam bentuk statik, dinamik ataupun gabungan antara keduanya. Selain itu, ia juga mempunyai aplikasi penting dalam perubatan seperti kajian oleh Ker et al., (2017), Khattak et al., (2021) dan Wang et al., (2016), di mana imej imbasan CT pesakit yang berturutan ditukarkan kepada fail 3-dimensi yang kemudiannya digunakan untuk mengesan ciri abnormal yang digunakan untuk membantu pengesanan pakar. Vrskova et al., (2022) mencadangkan penggunaan rangkaian saraf konvolusi tiga dimensi atau dikenali sebagai 3DCNN yang kebiasaannya digunakan untuk tujuan pengesanan dalam imej perubatan. Idea yang sama digunakan dengan menyusun setiap bingkai dari video menjadi sekelompok objek tiga dimensi. Dengan

menggunakan lapisan konvolusi 3 dimensi, ia mempunyai penapis yang bergerak dalam tiga dimensi (x,y,z) . Lapisan pengumpulan juga mempunyai binaan tiga dimensi, mempunyai fungsi yang sama dengan lapisan pengumpulan asas, akan tetapi, ia menjalankan fungsi persampelan menurun bagi vektor bentuk 3-dimensi. Seni bina yang dicadangkan mempunyai enam lapisan konvolusi 3-dimensi dan empat lapisan pengumpulan 3-dimensi. Saiz masukan dikawal dengan meringkaskan saiz masukan kepada $32 \times 32 \times 3$ RGB sebelum dirambat melalui rangkaian saraf, melalui rangkaian-rangkaian konvolusi dan lapisan pengumpulan maksima. Ia kemudiannya diikuti oleh lapisan normalisasi sebelum lapisan-lapisan terakhir seperti lapisan padat dan akhirnya lapisan rata.

Kerja penyelidikan juga dijalankan dengan menggunakan pendekatan-pendekatan samada menggunakan algoritma buatan tangan mahupun pembelajaran mesin untuk mengesan insiden lemas serta mencadangkan system penggeraan bagi tujuan tersebut. Seperti yang dilakukan oleh Salehi et al., (2016), mereka mencadangkan system pengesanan insiden lemas melalui pengesanan kontur bersama analisis ruang-warna HSV terhadap warna air dan warna badan perenang. Sistem akan memantau kehadiran perenang secara berterusan, sekiranya sistem mengesan perbezaan dengan badan perenang tidak dikesan akan mencetus penggeraan dihantar. Hayat et al., (2019) pula mencadangkan penggunaan pemasangan lengan robotik secara automatik dengan dikawal oleh sistem pemantauan berdasarkan aplikasi yang telah dibangunkan. Algoritma Camshift beserta penapis Kalman digunakan untuk mengesan kehadiran perenang dengan mewakili imej seorang perenang dengan bentuk elips. Perubahan bentuk dari elips kepada bentuk yang lain menandakan kebarangkalian kelemahan berlaku adalah tinggi, ia seterusnya mencetuskan sistem amaran untuk penyelamatan. Liu et al., (2019) pula menggunakan alat ultrasonik beserta penerima hidraulik bagi memantau keadaan di dalam kolam renang. Data yang diperolehi diwakilkan kepada koordinat tiga dimensi melalui pemodelan matematik dengan data dari penerima ultrasonik mengesan kedudukan perenang dan data hidraulik untuk memantau kedalaman perenang yang menunjukkan samada perenang mungkin sedang tenggelam ke dasar. Manakala Alotaibi (2020) mencadangkan system IOT menggunakan pengawal mikro, Raspbery Pi 3 disambung dengan penerima pergerakan dan kamera untuk memantau kehadiran objek di dalam kolam

renang. Bagi tujuan pengesanan serta pengecaman objek, kaedah pembelajaran pindah digunakan dengan menggunakan model yang tersedia, iaitu model ResNet50.

Berdasarkan pemerhatian di dalam kajian literasi ini, bilangan literasi berkaitan pengelasan jenis renangan sangat terhad lebih-lebih lagi pengelasan berdasarkan imej berurutan ataupun video. Bagi penggunaan pendekatan pengelasan menggunakan rangkaian konvolusi CNN, kajian yang terhampir adalah kajian-kajian yang menggunakan set data UCF50 (Vrskova et al., 2022), (Vrskova et al., 2022), (Reddy & Shah, 2013), (Vrskova et al., 2020) dan UCF101 (Soomro et al., 2012). Kedua-dua set tersebut tidak menumpukan secara spesifik terhadap jenis-jenis renangan di mana ia mengandungi sejumlah besar kelas lain menjadikan model yang dibina adalah sangat umum. Tambahan lagi, video-video dalam kedua-dua set data berkenaan kategori renangan tidak memenuhi dua kriteria yang ditetapkan dalam kajian ini, yang pertama video pada kebanyakan masa menunjukkan pandangan di luar air atau pandangan atas; kedua – video tidak menunjukkan pandangan renangan dari bahagian sisi di mana ciri-ciri pergerakan yang membezakan setiap renangan dapat dinyatakan dengan lebih lengkap jika mengambil pandangan sisi.

Bagi tujuan membangunkan model pengelasan jenis renangan spesifik untuk analisis video bagi pandangan sisi perenang dari dalam air, kajian ini mencadangkan pembangunan model pengelasan jenis renangan berasaskan rangkaian neural konvolusi yang praktikal dan mudah. Pendekatan ini diharap dapat membantu dalam penyediaan model yang boleh digunakan terutamanya untuk membangunkan sistem keselamatan dalam kolam renang sebagai satu sistem sokongan terhadap pemantauan oleh manusia sepenuhnya.

2.9.1 GAYA RENANGAN

Bahagian ini akan menerangkan mengenai elemen-elemen pergerakan yang menjadi asas dalam memahami perbezaan antara setiap jenis renangan supaya model dapat dibina melalui pemahaman yang spesifik dalam domain bidang. Berenang merupakan satu kaedah pergerakan semasa berada di dalam air samada untuk tujuan rekreasi, senaman ataupun kompetitif. Berenang juga biasa dikaitkan dengan teknik ikhtiar hidup manusia di mana ia melibatkan pelbagai variasi melibatkan kombinasi

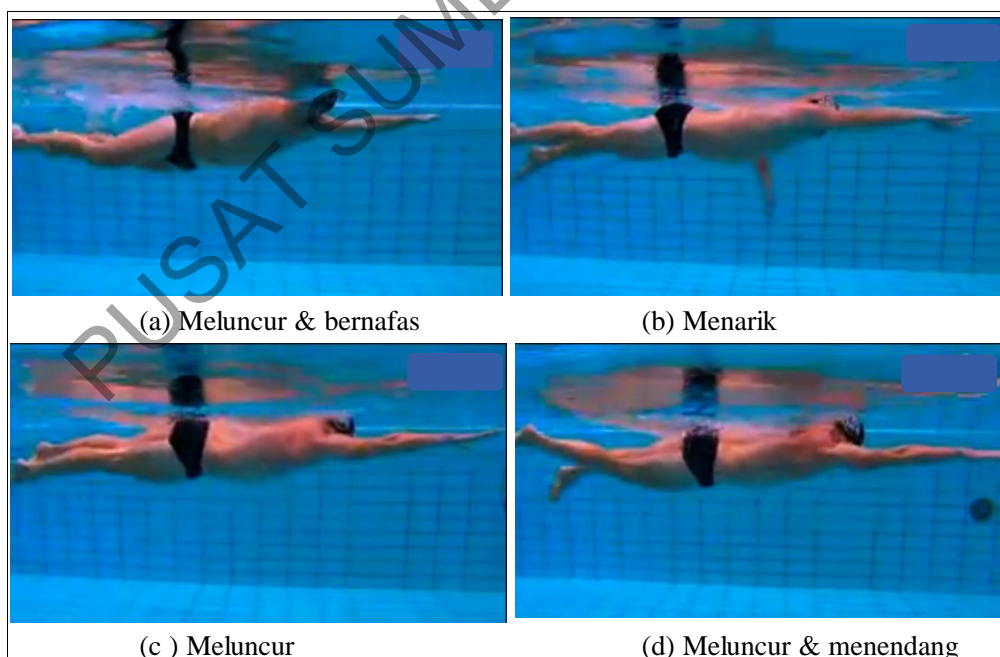
pergerakan tangan, kaki serta tubuh untuk meluncurkan badan ke hadapan, di samping memastikan pemasaan untuk mengambil nafas dengan memastikan hidung atau mulut berada di atas permukaan air dalam masa yang mencukupi sambil menarik nafas. Ia melibatkan gabungan set kemahiran pergerakan yang kompleks, ia perlu dipelajari dan memerlukan latihan untuk mencapai bentuk dan pergerakan yang sempurna.

Gaya renang mempunyai pelbagai variasi, bergantung kepada gabungan pergerakan kaki, tangan, tubuh serta pengambilan nafas. Setiap satu memberikan kesan yang berlainan samada dari segi kelajuan, penggunaan tenaga, ketahanan, dan sebagainya. Contohnya seperti gaya bebas diketahui memberikan kelajuan yang paling maksimum, akan tetapi cepat memenatkan, berbanding kuak dada, tidak memberikan saingan pada kelajuan seperti gaya bebas, tetapi memberikan ketahanan untuk renang dalam jarak yang jauh. Walaupun banyak gaya renang yang dikenalpasti, hanya tiga jenis renang yang utama berdasarkan kategori yang biasa dipertandingkan dalam kejohanan serta biasa digunakan oleh perenang rekreasi – empat gaya renang tersebut adalah gaya bebas, kuak dada serta kuak kupu-kupu. Gaya renang selain dari tiga jenis yang tersebut, pada kebanyakan masa menggunakan gabungan dari salah satu gerakan dari empat-empat gerakan seperti yang dinyatakan.

Literasi berkaitan kaedah renang diperolehi melalui beberapa sumber untuk memperihalkan setiap gaya renang bersama gerakan-gerakan yang terkait. Antaranya seperti tulisan oleh Bay (2016), memberikan panduan kepada gerakan-gerakan asas dalam jenis-jenis renang di dalam kajian ini. Newsome (2012) pula menumpukan secara spesifik kepada renang gaya bebas, dengan tujuan memberikan panduan kepada perenang dalam memperbaiki teknik renang gaya bebas. Beberapa video (Speedo International, 2013) juga menunjukkan gerakan-gerakan yang berkait dengan setiap jenis renang. Melalui kajian literasi yang dijalankan, bahagian yang seterusnya akan menghuraikan setiap jenis renang berdasarkan ciri-ciri tersebut. Melalui kepelbagaian dalam elemen unik pergerakan bagi setiap gaya renang, ini secara tidak langsung membantu dalam membezakan setiap kategori dalam pendekatan menggunakan kaedah-kaedah penglihatan komputer.

2.10.1 Gaya Renangan: Gaya Bebas

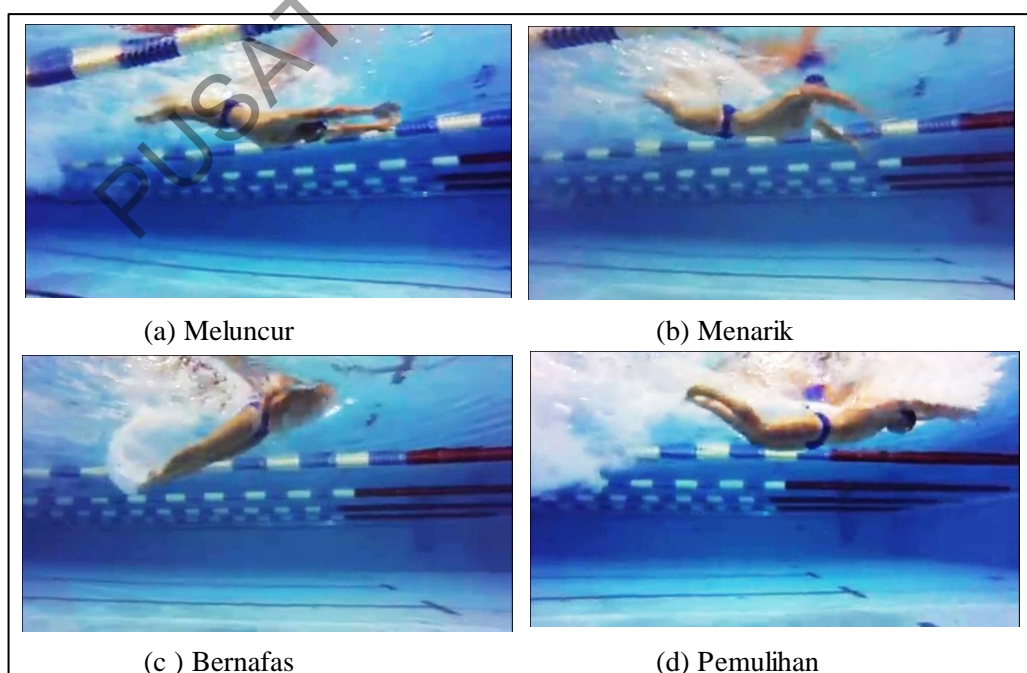
Gaya bebas seperti ditunjukkan Rajah 2.7 melibatkan tindakan “tangkap dan tarik” oleh tangan dengan cara memusingkan tangan ke hadapan secara berselang-seli, ini memberikan daya tujahan ke hadapan sambil kaki melakukan gerakan *flutter kick* – ini pula memberikan tolakan ke hadapan yang berlaku secara berterusan sambil memastikan kaki bersalingan dengan tubuh untuk memberikan badan gerakan meluncur. Gabungan kesemua pergerakan ini memberikan satu gaya renangan yang efisien dengan tenaga yang dikeluarkan. Untuk pengambilan nafas pula, kepala akan dipusingkan serentak dengan mana-mana belah tangan diangkat keluar dari permukaan air, pada pertengahan gerakan ini, perenang akan memusingkan kepala sedikit ke atas sehingga mulut dan hidung berada di luar permukaan air. Pada masa yang singkat ini, perenang akan menarik nafas dengan cepat dan dalam dan memusing semula kepala ke bawah permukaan air. Nafas pula dihembus keluar semasa berada di dalam air melalui mulut menggunakan kaedah pernafasan yang telah dilatih.



Rajah 2.7 Strok-strok utama bagi renangan gaya bebas

2.10.2 Gaya Renangan: Kuak kupu-kupu

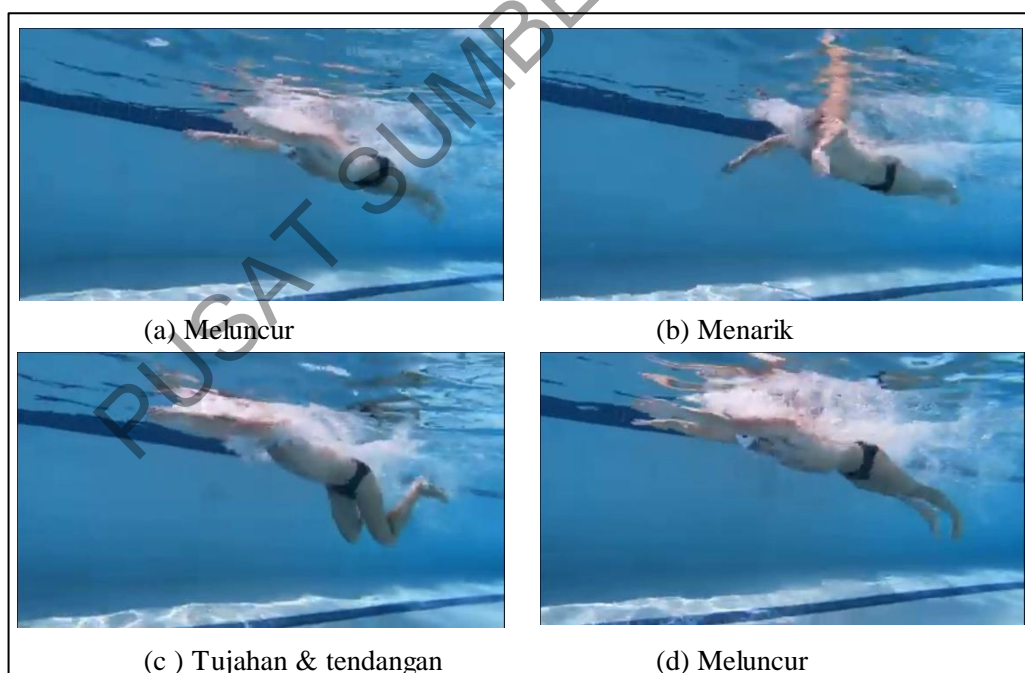
Kuak kupu-kupu mendapat nama berdasarkan pergerakan tangan perenang seolah-olah gerakan kupu-kupu melibaskan sayapnya semasa terbang. Gaya kuak kupu-kupu seperti dalam Rajah 2.8 merupakan gaya renangan yang dipercayai mempunyai kepantasan yang hampir serupa dibandingkan dengan gaya bebas. Akan tetapi, ia menjadi pilihan perenang rekreasi memandangkan tenaga yang digunakan untuk kuak kupu-kupu adalah lebih dibandingkan dengan jenis yang lain di mana ia menggunakan lebih banyak kumpulan otot menjadikannya pilihan untuk menurunkan berat badan. Gabungan gerakan tubuh dan kaki adalah berturutan di mana bahagian otot perut dan pinggul menghantar daya supaya kedua-dua belah kaki dapat melibas ke atas dan ke bawah secara serentak melalui tendangan yang dikenali sebagai *dolphin kick*. Tendangan ini memberikan tolakan seluruh badan ke hadapan. Dalam keadaan badan menghadap ke bahagian bawah kolam, kedua-dua belah tangan dihayun ke hadapan dalam bentuk putaran, mewakili pergerakan sayap kupu-kupu. Badan serta kepala akan dibenam ke dalam air sambil menghembuskan nafas, kemudian kepala dan badan akan di bawa ke atas permukaan air untuk mendapatkan udara dalam masa yang singkat.



Rajah 2.8 Strok-strok utama bagi renangan kuak kupu-kupu

2.10.3 Gaya Renangan: Kuak dada

Gaya renangan ini, berdasarkan Rajah 2.9 adalah antara gaya renangan yang rendah intensiti dibandingkan dengan gaya-gaya renangan yang lain. Bagaimanapun, ia menjadi pilihan untuk renangan rekreasi memandangkan ia kurang menggunakan tenaga dan kurang memenatkan di samping ia adalah kaedah renangan yang paling mudah dipelajari. Gerakan dimulakan dengan membenarkan badan untuk meluncur melalui air sambil kedua tangan dirapatkan untuk beberapa ketika dan kemudiannya tangan dihayunkan ke bawah, ke dalam air untuk memberikan gerakan tarikan ke hadapan. Kemudian kaki akan dilibaskan ke bawah dan ke belakang, gerakan ini juga dikenali sebagai *frog kick* di mana pergerakannya seolah-olah menyerupai pergerakan kaki belakang katak. Kedudukan badan adalah mendatar pada keseluruhan masa, mata pula mudah untuk melihat dan memberikan orientasi yang jelas pada perenang bagi kedudukan sebenar berbanding arah tujuannya.



Rajah Rajah 2.9 Strok-strok utama bagi renangan kuak dada

2.9.2 ANALISIS PERGERAKAN RENANGAN MENGGUNAKAN PEMBELAJARAN MESIN

Analisis pergerakan renangan melibatkan tiga tujuan, yang pertama adalah untuk mengkaji pergerakan tubuh, fisiologi bagi tujuan memperbaiki teknik renangan ini adalah bidang spesifik dalam kajian sains sukan. Tujuan yang kedua adalah untuk melakukan pengelasan dalam pergerakan tubuh badan manusia. Tujuan yang ketiga adalah untuk menyediakan sistem sokongan pengesanan kelemasan samada dari segi perkakasan serta pembangunan algoritma untuk aplikasi pengesanan insiden lemas. Di sini dinyatakan beberapa kajian-kajian lepas, yang dilakukan bagi setiap tujuan tersebut untuk memahami kerja-kerja yang telah dilakukan oleh penyelidikan yang lepas.

Penggunaan urutan imej renangan dengan mengesan pergerakan bahu dan tangan untuk memilih fitur yang penting dijalankan oleh Liao et al., (2003). Beberapa pengukuran kuantitatif dijalankan, berdasarkan data tersebut analisis regresi dijalankan dan satu struktur pohon keputusan dibina di mana turutan pergerakan membantu dalam kerja pengelasan jenis renangan berdasarkan turutan imej dari video. Bagi mendapatkan fitur yang dapat memperihalkan sesuatu corak renangan, penulis menggunakan segmentasi berdasarkan warna di mana tubuh perenang boleh dikenalpasti memandangkan perbezaan warna antara warna tubuh perenang dengan warna air kolam. Melalui analisis regresi, dua nilai diperolehi iaitu *large slope* (a) dan *small aspect ratio* (AR). Satu set pohon keputusan dibina berdasarkan nilai-nilai a dan AR yang diperolehi di mana setiap cabang menentukan jenis renangan yang sedang diperhatikan. Kaedah ini mengambil langkah yang panjang untuk dijalankan tambahan pula, perubahan warna serta pencahayaan yang sentiasa berubah boleh menjejaskan prestasi model.

Omae et al., (2017) mencadangkan model pengelasan jenis renangan menggunakan data melalui penderia yang dipasang pada perenang. Data yang direkodkan adalah seperti pecutan dan halaju sudut bagi setiap paksi x, y dan z . Sebelum membina model pengelasan, data ditukar kepada nilai pemboleh ubah yang mewakili fitur bagi objek-objek yang akan ditentukan. Model pembelajaran *ensemble* diperkenalkan bagi mengatasi masalah yang dihadapi penyelidik sebelumnya di mana

model pengelasan memberi keputusan bercampur dalam pengelasan renangan jenis kuak dada serta kuak kupu-kupu. Penulis juga menentukan nilai pemberat, w , yang dapat dikaitkan bagi setiap jenis renangan yang dikaji.

Model pembelajaran dalam dicadangkan oleh Tarasevicius & Serackis (2020) untuk mengenal pasti jenis renangan berdasarkan data yang diperolehi dari jam pintar. Rangkaian Bi-LSTM digunakan sebagai rangkaian pembelajaran dalam untuk memisahkan jenis renangan berdasarkan data yang dikumpul melalui meter pecutan, giroskop, dan magnetometer. Ohgi et al., (2014) pula menggunakan penderia yang mengesan pecutan pada setiap paksi x,y,z dipasang pada badan perenang bagi mengesan fasa rehat, mula, pusing dan tujuan dinyatakan oleh pecutan pada paksi y . Penulis menggunakan model pengelasan rangkaian saraf pelbagai lapisan dan pohon keputusan C4.5 dengan ketepatan sebanyak 91.1% telah dicapai. Menggunakan pendekatan sebegini, ianya adalah kurang praktikal, memandangkan ia memerlukan setiap pengguna untuk memakai sensor serta *wearables* pada setiap sesi renangan. Bagi pendekatan

Secara keseluruhannya, berdasarkan kajian literatur yang telah dilakukan secara menyeluruh, didapati tiada lagi kajian yang dilakukan melibatkan penggunaan imej juga jujukan imej untuk pengelasan jenis renangan melalui pendekatan pembelajaran mendalam. Oleh yang demikian, kajian ini mencadangkan pembinaan model pengelasan jenis renangan menggunakan pembelajaran mendalam, menggunakan rangkaian neural buatan CNN terhadap data dalam bentuk jujukan imej berdasarkan video renangan. Kajian sebegini amat wajar untuk diterokai memandangkan peluang yang boleh digunakan serta manfaat serta potensi yang akan diperolehi sekiranya model sedemikian dapat dibangunkan.

BAB III

METODOLOGI KAJIAN

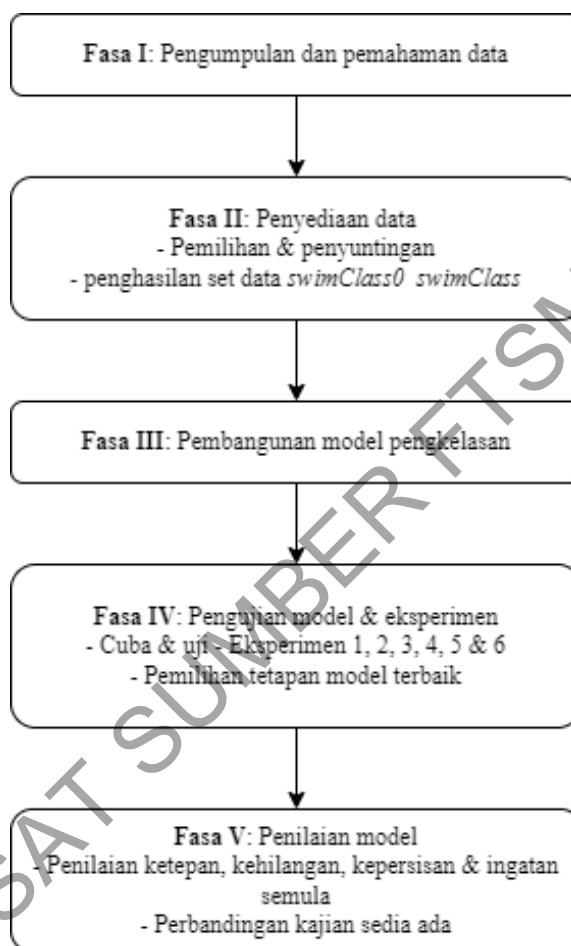
3.1 PENGENALAN

Bagi memastikan objektif kajian dapat dicapai, metodologi kajian perlu dirancang secara sistematik. Perancangan metodologi kajian dijalankan dengan mempertimbangkan kaedah-kaedah kajian yang berkaitan yang telah dijalankan berdasarkan kerja-kerja penyelidik terdahulu diikuti dengan konsultasi oleh penyelia dari semasa ke semasa. Kajian dilakukan dalam beberapa fasa utama disertai sub-fasa yang berkenaan.

3.2 REKABENTUK KAJIAN

Rekabentuk kajian yang digunakan dalam kajian ini ialah merujuk kepada rekabentuk eksperimen bagi pengkelasan video menggunakan kaedah pembelajaran dalam atau lebih spesifik menggunakan rangkaian neural konvolusi (CNN). Rekabentuk serta perancangan kajian adalah kritikal memandangkan penyediaan data akan mempengaruhi hasil akhir kajian. Analisis video adalah menggunakan rangkaian neural yang terdiri dari sejumlah lapisan tersembunyi dengan setiap satu mempunyai nilai pemberat, w serta bias, b , masing-masing, di mana nilai-nilai tersebut adalah dipengaruhi oleh nilai-nilai masukan, pemberat dan bias pula, ianya adalah di luar kawalan manusia. Selain dari itu, kerja-kerja pengiraan akan melibatkan masa pengiraan yang panjang, maka adalah penting untuk perancangan yang rapi bagi mengelakkan pembaziran sumber serta masa disebabkan ralat yang boleh dielakkan. Titik buta juga mengundang kesilapan dan akan menjejaskan keputusan eksperimen, ini juga boleh dielakkan dengan pendekatan mengkaji kerja-kerja oleh penyelidik

terdahulu dalam mengenal pasti titik-titik buta seperti yang dinyatakan dalam kerja-kerja terdahulu. Rajah 3.1 menunjukkan rekabentuk kajian yang utama yang terbahagi kepada lima fasa, setiap peringkat akan diperjelas dalam bahagian-bahagian selanjutnya.



Rajah 3.1 Rekabentuk kajian

3.3 FASA PENGUMPULAN DAN PEMAHAMAN DATA

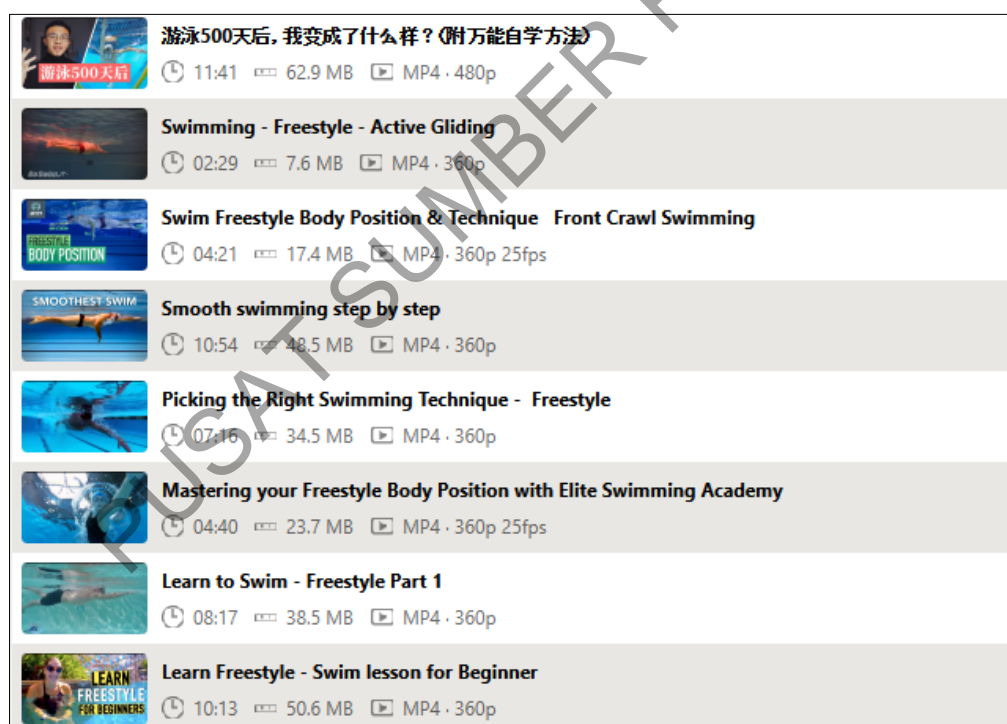
Untuk mencapai objektif setiap penyelidik harus memastikan agar persoalan kajian dapat dijawab dengan sempurna. Salah satu disiplin yang perlu dijaga untuk memastikan tujuan tersebut tercapai adalah melalui penyediaan data yang sesuai dengan pembolehubah-pembolehubah yang relevan. Pengumpulan data adalah langkah awalan penyediaan data eksperimen di mana data dikutip dari sumber

sebelum disusuli proses seterusnya. Selain dari itu, penyelidik perlu memahami domain kajian supaya konteks kajian berada pada landasan yang betul. Semua ini membantu dalam menyediakan data yang bermakna bagi tujuan menguji hipotesis serta menilai hasil kajian.

Data video dalam kajian-kajian seumpama ini kebiasaannya menggunakan samada dari set data penanda aras yang tersedia dari kerja-kerja penyelidik terdahulu ataupun set data disediakan sendiri untuk kajian yang lebih spesifik. Beberapa contoh set data penanda aras yang biasa digunakan di dalam kerja-kerja yang lepas adalah seperti *UCF50*, *UCF101* (Soomro et al., 2012) bawah seliaan *Center for Research in Computer Vision, University of Central Florida*. *UCF50* adalah set data yang mengandungi video pergerakan tubuh manusia dalam 50 kategori pergerakan, manakala *UCF101* pula adalah set data yang terdiri dari 101 jenis pergerakan tubuh manusia. Manakala *Deepmind* oleh Google pula menyedia dan menyelia set data *Kinetics 400* (Kay et al., 2017), *Kinetics 600*, *Kinetics 700* (Kay et al., 2017) serta *AVA Kinetics* (A. Li et al., 2020) di mana kesemuanya set data adalah dalam bentuk fail JSON yang mengandungi pautan kepada video-video suapan dari Youtube berdasarkan jenis-jenis pergerakan yang telah dilabel. Penggunaan model terhadap set data penanda aras dapat memberi gambaran mengenai keberkesanan model yang digunakan. Bagaimanapun, untuk kajian spesifik berkaitan analisa video renangan, tiada data penanda aras boleh diperolehi setakat kajian literatur yang telah dijalankan, kecuali data dari *UCF101* mempunyai video renangan gaya bebas dan kuak dada. Akan tetapi, video dari *UCF101* tidak selari dengan kriteria pemilihan video yang ditetapkan untuk digunakan dalam kajian ini.

Bagi memastikan kajian ini mengikuti skop, set data kompilasi disediakan bagi disesuaikan dengan skala model yang dibina serta perkakasan yang digunakan untuk kajian. Antara kriteria yang perlu dipenuhi adalah yang pertama – video memaparkan renangan dalam pandangan sisi sahaja; kedua – video memaparkan seorang sahaja perenang berada dalam bingkai pada sesuatu masa; dan ketiga – video renangan adalah pemandangan kamera dari dalam air. Kesemua kriteria ini tidak dipenuhi dalam set data penanda aras seperti yang dinyatakan sebelum ini. Oleh itu, video-video perlu dipilih secara manual untuk memastikan kesemua video yang akan

digunakan memenuhi kriteria-kriteria tersebut. Video dimuat turun dari laman perkongsian video terbesar iaitu Youtube. Carian dilakukan menggunakan kata-kunci berkaitan seperti *frontcrawl* dan *freestyle* untuk gaya bebas, *butterfly stroke* untuk kuak kupu-kupu, *breast stroke* untuk kuak dada. Cadangan dari carian Youtube menggunakan kata-kunci mengandungi bilangan hasil yang banyak. Akan tetapi tidak semua adalah berkaitan dengan kehendak carian dan ini memerlukan pemilihan dilakukan secara manual berdasarkan pemahaman dalam domain bidang. Beberapa hasil cadangan yang terawal perlu disemak terlebih secara kasar untuk memastikan video berpotensi memenuhi kriteria. Sekiranya sesebuah video didapati dapat memenuhi kriteria, ia akan dimuat turun dan diasingkan ke dalam folder mengikut kelas masing-masing; ini akan menjadi fail mentah yang perlu diproses lebih lanjut semasa penyuntingan.



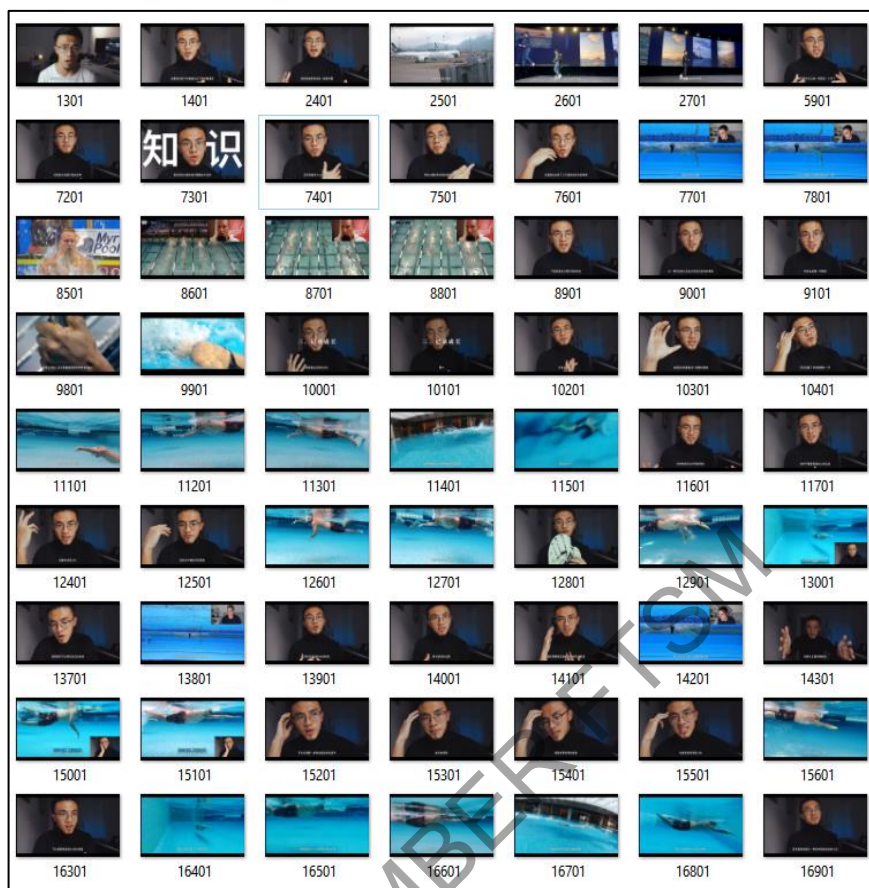
Rajah 3.2 Paparan antara muka perisian 4KDownloader

Muat turun dilakukan menggunakan perisian *4K Downloader* seperti Rajah 3.2. Sebelum muat turun dilakukan, video diimbas secara manual untuk kesesuaian sama ada ia mengandungi konteks yang berpotensi untuk kegunaan analisis atau sebaliknya. Video dipilih berdasarkan kandungan yang memenuhi kriteria yang telah ditetapkan sebelum ini. Setelah meneliti sejumlah video-video cadangan dari Youtube

berdasarkan kata kunci, secara kasar kurang dari 150 video didapati memenuhi kriteria dan sesuai untuk digunakan dalam pembinaan set data. Pada peringkat awal, berdasarkan bilangan video yang telah dimuat turun didapati untuk setiap satu kategori mengandungi kurang dari 50 video; jumlah yang agak kecil untuk sebuah kajian pembelajaran mesin, tetapi memandangkan video adalah jujukan dari sekumpulan gambar yang berturutan, bagi video dengan 25fps, sebuah video berdurasi 2 minit akan mengandungi 3000 bingkai, dan setiap bingkai pula mengandungi bilangan 3 saluran setiap satu mempunyai piksel dengan nilai yang tersendiri. Jadi beberapa strategi dirancang untuk mengurangkan keperluan kepada pengiraan yang ekstensif, sekali gus meningkatkan kejituan model.

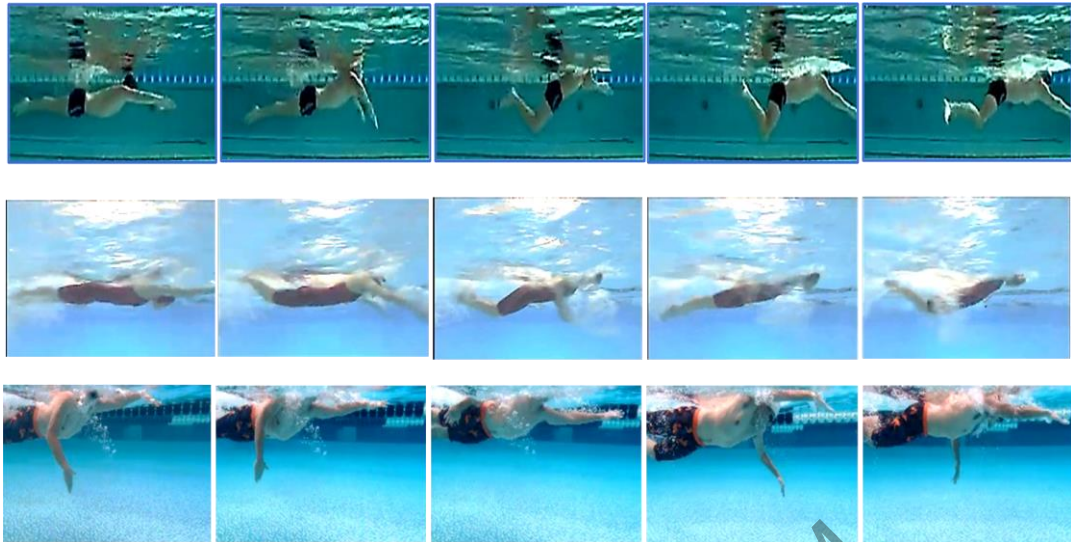
3.4 FASA PENYEDIAAN DATA

Penyediaan data dalam kajian ini secara umumnya adalah menggunakan perisian serta teknik-teknik dalam penyuntingan video untuk memilih konteks video yang sesuai. Video-video yang akan digunakan perlulah memenuhi kriteria berikut – pertama, video adalah pada pandangan sisi perenang sahaja; kedua, video adalah pada pandangan dalam air sahaja; ketiga, video menunjukkan pergerakan renangan untuk seorang perenang sahaja pada sesuatu masa. Memandangkan kesemua video yang dimuat turun bukanlah 100% terdiri dari video renangan semata di mana kebanyakan video yang memenuhi kriteria adalah video tutorial latihan renangan menggunakan teknik yang betul – jadi ada percampuran antara kandungan video renangan diselangi video penerangan oleh penyedia kandungan seperti ditunjukkan oleh Rajah 3.3. Video-video perlu melalui kerja-kerja suntingan secara manual melalui pemilihan bahagian yang berkaitan serta membuang bahagian yang tidak perlu. Bahagian yang tidak berkaitan pergerakan akan memperkenalkan hingar kepada model rangkaian, sekali gus menurunkan prestasi model dengan ketara. Manakala, pemilihan bahagian dengan berkaitan pergerakan yang sesuai pula akan membantu model mempelajari data secara lebih berkesan, terutama dengan set data dengan bilangan yang agak kecil.



Rajah 3.3 Contoh jujukan video sebelum kerja suntingan

Pergerakan berciri adalah gerakan yang boleh digunakan untuk membezakan ciri gerakan. Ia digunakan untuk membezakan setiap kelas dengan tepat pada momen-momen yang tertentu seperti ditunjukkan dalam Rajah 3.4. Contohnya, tendangan bagi setiap strok adalah berbeza – gaya bebas dengan *flutter kick*, kuak kupu-kupu mempunyai *dolphin kick*, dan kuak dada dengan *frog kick*. Selain dari itu, gerakan tangan, torso serta pemasaan dalam pengambilan nafas atau gerakan kepala juga menjadi unsur yang memberikan setiap kategori strok ciri-ciri pembeza. Gabungan setiap gerakan tadi juga memberikan satu kitaran yang tetap; setiap kitaran pula dibahagi kepada fasa yang tersendiri. Maka pergerakan berciri, membantu membezakan setiap kategori dalam pembelajaran model konvolusi dengan lebih tepat lagi.



Rajah 3.4 Contoh jujukan pergerakan berciri dalam pergerakan renang

Kerja-kerja penyuntingan video perlu untuk memastikan set data memberi masukan yang bermakna kepada rangkaian konvolusi. Rangkaian konvolusi akan mengesan bahagian-bahagian tertentu pada data, maka perlu ditentukan bahagian yang diperlukan oleh rangkaian seperti ditunjukkan dalam Rajah 3.5. Elemen-elemen berkaitan seperti tubuh dan gerakan yang dilakukan membantu rangkaian menumpu, manakala bahagian lain seperti bahagian ulasan oleh instruktur, elemen objek dalam kolam, teks dalam bentuk anotasi, logo contohnya, berpotensi untuk menyebabkan pemberat serta bias dalam rangkaian konvolusi merosot. Beberapa elemen asing sebolehnya dikeluarkan dan ada yang tidak dapat dielakkan, contohnya seperti lampu dan pembahagi lorong kolam. Kerja-kerja pengecualian hingar dilakukan secara manual dalam fasa kerja-kerja suntingan. Penyuntingan video melibatkan kerja-kerja seperti *extraction*, *cropping*, dan *mirroring*. Penerangan mengenai setiap proses dalam fasa penyuntingan akan dihuraikan secara lebih lanjut dalam bahagian seterusnya.



(a) Pandangan atas, depan



(b) Pandangan atas, sisi



(c) Pandangan sisi, dalam air, tapi kelihatan logo.



(d) Pandangan sisi, dalam air, kelihatan logo dan anotasi



(e) Pandangan sisi, dalam air, kurang hingar, sebahagian badan tidak kelihatan.

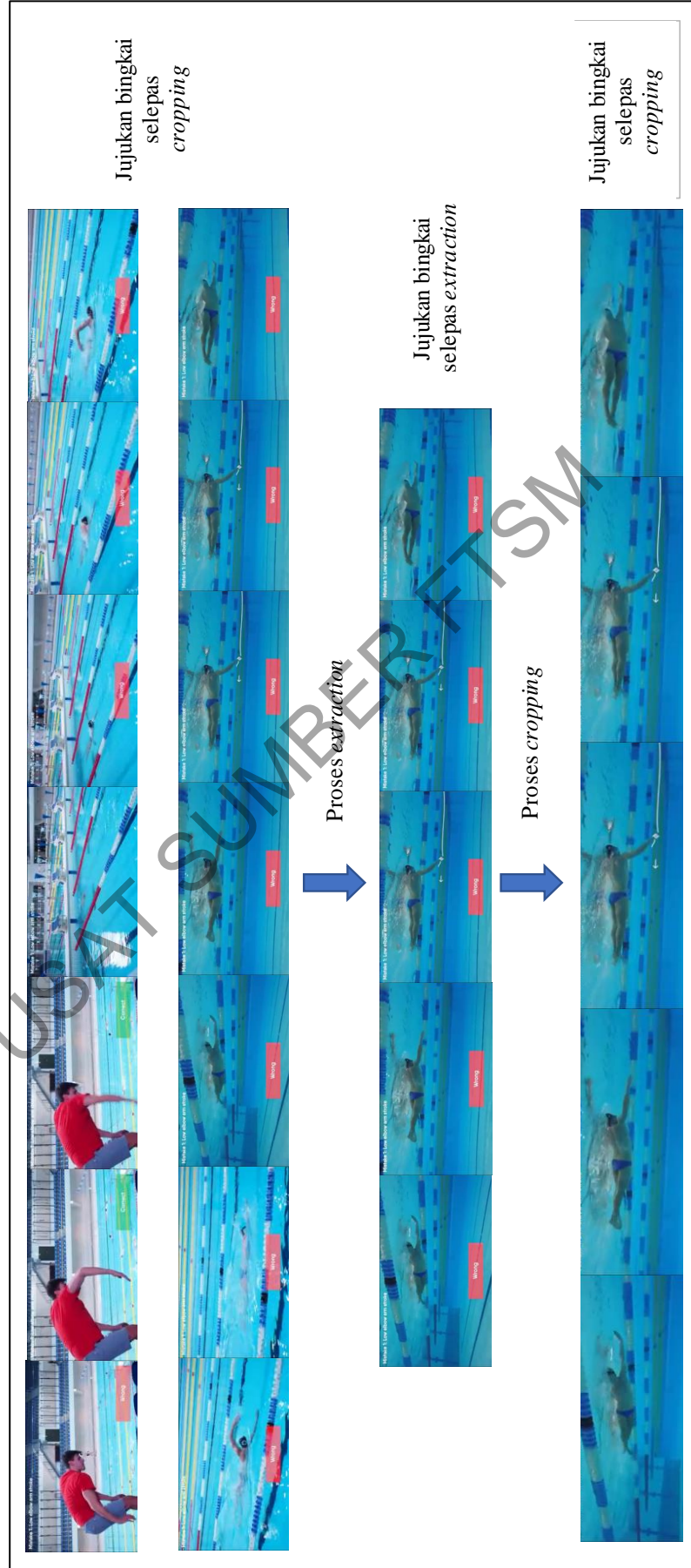


(e) Pandangan sisi, dalam air, badan sepenuhnya kelihatan

Rajah 3.5 Pemilihan pandangan mengikut kriteria ditetapkan

3.4.1 Fasa Penyediaan Data: Penyuntingan Secara *Extraction*

Proses penyuntingan melalui *extraction* adalah seperti ditunjukkan dalam Rajah 3.6. Proses ini, adalah proses yang pertama dalam keseluruhan proses penyuntingan. Proses ini dilakukan dengan memilih bahagian-bahagian dalam video yang mengungkapkan ciri renang berdasarkan skop kajian – renang terendam dan pandangan sisi. Berikutan kebanyakan video mempunyai bahagian-bahagian yang tidak diinginkan seperti renang pandangan atas atau hadapan, penerangan oleh instruktur, tayangan logo, dan sebagainya. Maka bahagian yang tidak berkenaan akan dikeluarkan dalam peringkat ini.



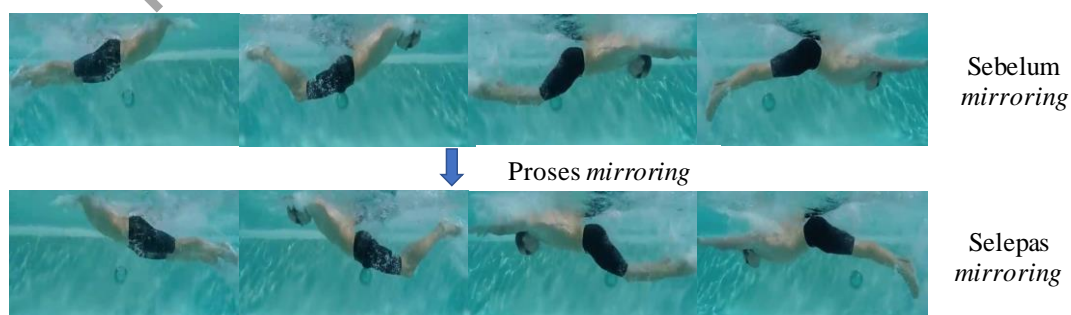
Rajah 3.6 Proses penyuntingan dan pemilihan jujukan bingkai melalui proses extraction dan cropping

3.4.2 Fasa Penyediaan Data: *Data Augmentation*

Proses *extraction* pula akan disusuli dengan proses *cropping*, seperti ditunjukkan dalam Rajah 3.6. *Cropping* pula bertujuan untuk membuang anotasi dalam video ataupun bahagian-bahagian yang tidak berkenaan dalam video seperti logo, *watermark*, serta sarikata. Proses ini sama seperti melakukan *cropping* pada sesuatu imej, akan tetapi, sedikit tetapan dilakukan dalam aplikasi penyuntingan sebelum proses dijalankan seperti memilih ukuran bingkai yang diperlukan sambil menutupi bahagian yang tidak diingini.

3.4.3 Fasa Penyediaan Data: Penyuntingan Secara *Mirroring*

Mirroring pula mempunyai tujuan yang khusus dalam meningkatkan saiz set data, seperti kajian oleh Taylor & Nitschke (2019), ia juga dikenali sebagai *data augmentation* ataupun penambahan data menggunakan set data imej sedia ada. Menurut Taylor & Nitschke (2019), penambahan data memberikan sebanyak 14% penambahbaikan pencapaian model berbanding data sedia ada dengan pendekatan seperti menterbalik, memusing, memotong dan sebagainya. Akan tetapi penambahan terhadap set data kajian ini akan hanya melibatkan teknik *mirroring* sahaja dengan setiap video renang asal akan disertai dengan imej cerminan masing-masing – contohnya seperti ditunjukkan Rajah 3.7, video asal mungkin menunjukkan renang kiri ke kanan; dengan cerminan, satu lagi salinan fail dalam bentuk renang kearah berlawanan dihasilkan.

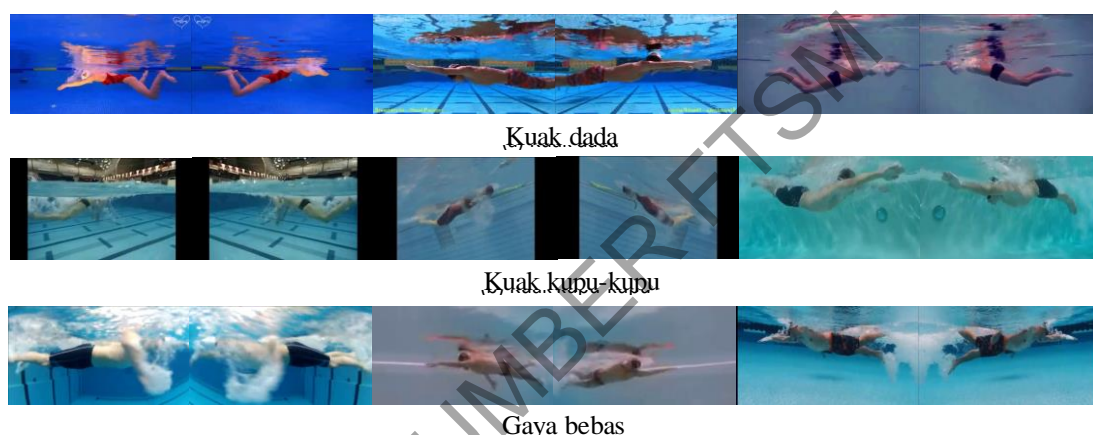


Rajah 3.7 Proses *mirroring* pada jujukan bingkai

Selepas proses *mirroring*, saiz data menjadi lebih besar dengan jumlah asal video yang telah dipilih pada awal adalah sekitar 90 video untuk setiap kelas untuk set data *swimClass0*, meningkat dua kali ganda kepada sekitar 180 video bagi setiap kelas untuk *swimClass*. Semua kerja penyuntingan ini dilakukan dengan menggunakan perisian sumber terbuka – *Blender*. *Blender* secara khusus dibangunkan untuk kerja-kerja animasi, tetapi perisian *Blender* juga dilengkapi keupayaan untuk kerja-kerja suntingan video memandangkan kerja-kerja animasi juga memerlukan kepada ciri untuk suntingan. Selain dari kelebihan *Blender* sebagai sumber terbuka, ia membolehkan penggunaan GPU dalam kerja-kerja *rendering* dengan mengubah tetapan aplikasi *Blender* dalam masa yang sama ia menjimatkan masa dalam kerja suntingan.

3.4.4 Set Data

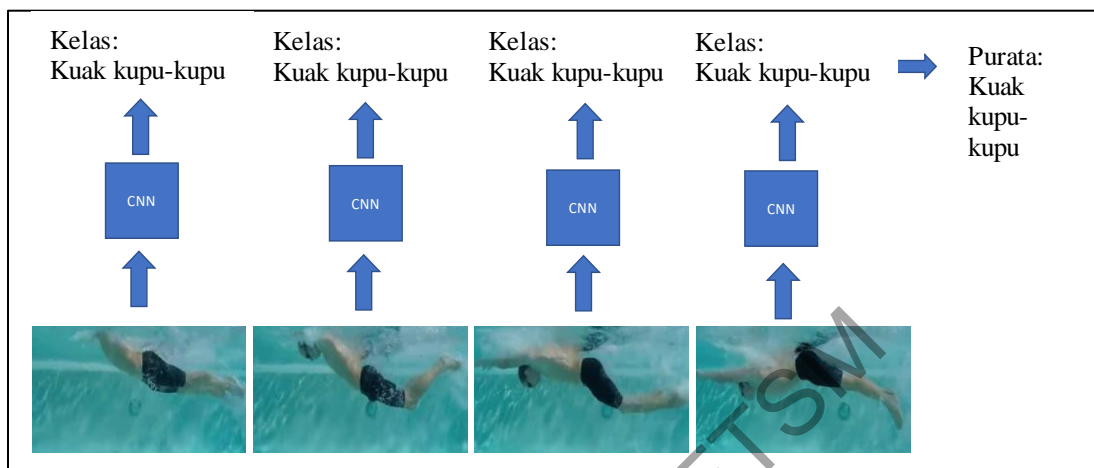
Set data yang telah dibangunkan terdiri dari dua set data *swimClass0* dan *swimClass* di mana kedua-dua set data terdiri video-video kompilasi yang dipilih serta dikelaskan berdasarkan tiga jenis renangan yang telah ditentukan iaitu renangan gaya bebas, kuak dada, serta kuak kupu-kupu. Setiap satu kelas diletakkan mengikut tiga fail yang berasingan mengikut kelas renangan masing-masing. Bagi set data *swimClass0*, Rajah 3.8, ia terdiri dari 100 fail untuk gaya bebas, kuak dada 105 fail, dan kuak kupu-kupu sebanyak 116 fail data asal. Manakala bagi *swimClass*, Rajah 3.9, ia terdiri dari dengan jumlah fail setiap satu jenis ialah gaya bebas 198 fail, kuak dada 195 fail, dan kuak kupu-kupu sebanyak 212 fail untuk *swimClass* untuk data yang ditambah terhadap *swimClass0*. Dapat diperhatikan di sini, *swimClass* adalah set data yang ditambah melalui teknik penambahan data secara *mirroring*. Untuk memastikan set data adalah seimbang untuk setiap kelas, beberapa video yang telah melalui proses *mirroring* tidak dimasukkan. Ini memandangkan set data yang tidak seimbang akan menghasilkan model yang bias, seperti yang dinyatakan oleh He et al. (2009) di mana set data yang tidak seimbang akan menyebabkan model bias kepada kelas dengan bilangan data yang lebih banyak. Semasa eksperimen, 80 peratus dari data akan digunakan untuk tujuan latihan, manakala, bakinya sebanyak 20 peratus akan digunakan untuk tujuan validasi model selepas setiap kali latihan berlangsung.

Rajah 3.8 Tangkapan skrin sebahagian daripada set data *swimClass0*Rajah 3.9 Tangkapan skrin sebahagian daripada set data *swimClass*

3.5 FASA PEMBANGUNAN MODEL

Seperti dinyatakan di dalam kajian literatur, beberapa kerja telah dilakukan penyelidik terdahulu menggunakan beberapa model cadangan. Bagi tujuan kajian ini, model yang akan dirujuk adalah dari model yang ringkas memandangkan saiz set data yang agak besar di samping limitasi dalam kuasa komputer. Dalam kajian ini, model dibangunkan menggunakan asas seni bina berdasarkan seni bina rangkaian neural konvolusi 2DCNN seperti yang dicadangkan oleh Karpathy et al., (2014) dan sebahagian pelaksanaan kod adalah seperti dicadangkan oleh Anwar, T. (2021). Seni binanya seperti ditunjukkan Rajah 3.10 yang juga diinspirasi oleh rangkaian neural *ImageNet* (Krizhevsky, 2007) dalam kerja pengecaman imej. Pendekatan ini dipanggil sebagai CNN bingkai tunggal. Melalui CNN bingkai tunggal, satu jujukan bingkai dari sebuah video, setiap kali bingkai dirambat melalui rangkaian, ia akan menentukan kelas bagi setiap bingkai. Katakan satu jujukan bingkai terdiri dari sejumlah n bingkai akan memberikan kelas bingkai secara individu. Selepas lengkap satu jujukan video,

nilai output akan dikira semula berdasarkan purata output kelas yang diperolehi. Memandangkan seni bina yang mudah serta praktikal, maka model 2DCNN dengan bingkai tunggal adalah bersesuaian untuk dilaksana dalam kajian ini.



Rajah 3.10 CNN bingkai tunggal

Model asas 2DCNN akan mempunyai 8 lapisan tersembunyi terdiri daripada rangkaian konvolusi 2D, lapisan normalisasi, lapisan pengumpulan serta lapisan tersambung sepenuhnya. Kajian akan dimulakan dengan model 2DCNN asas dengan dua lapisan konvolusi. Seterusnya, model akan dilanjutkan dengan menambah satu lagi lapisan konvolusi tambahan terhadap model asas kepada tiga lapisan konvolusi. Setiap satu seni bina diterangkan secara lebih lanjut dalam bahagian seterusnya.

3.5.1 Model 2DCNN Dengan Dua Lapisan Konvolusi

Pada dasarnya, model 2DCNN asas terdiri dari sejumlah lapisan yang membina rangkaian neural konvolusi 2DCNN dalam kajian ini. Model 2DCNN ini terdiri dari dua lapisan konvolusi pada dua lapisan awal secara berturutan seperti ditunjukkan dalam Jadual 3.1. Selain dari lapisan konvolusi, lapisan-lapisan lain disusun secara berturutan. Lapisan normalisasi kelompok yang berperanan untuk mempercepat proses pembelajaran serta menstabilkan rangkaian neural. Operasi matriks pula dijalankan pada lapisan pengumpulan maksima serta pengumpulan purata. Seterusnya, nilai yang diperolehi melalui operasi matriks sebelumnya akan dihantar kepada lapisan tersambung padat, di mana setiap keluaran akan dihantar kepada setiap cabang dalam lapisan tersambung padat. Ianya akan disusuli semula oleh satu lagi lapisan

normalisasi kelompok sebelum akhirnya semua keluaran akan dihantar kepada setiap cabang dalam lapisan padat yang terakhir di mana pengelasan akan dilakukan dalam lapisan ini.

Jadual 3.1 Senarai Lapisan *sequential* dalam Model 2DCNN dengan dua lapisan konvolusi

Bil	Lapisan	Saiz Keluaran
1	Konvolusi 2D	(Tiada, 62, 62, 64)
2	Konvolusi 2D	(Tiada, 60, 60, 64)
3	Normalisasi Kelompok	(Tiada, 60, 60, 64)
4	Pengumpulan Maksima 2D	(Tiada, 30, 30, 64)
5	Pengumpulan Purata 2D	(Tiada, 64)
6	Lapisan Padat	(Tiada, 256)
7	Normalisasi Kelompok	(Tiada, 256)
8	Lapisan Padat	(Tiada, 3)

Setiap lapisan mempunyai saiz keluaran seperti ditunjukkan dalam Jadual 3.1. Setiap nilai keluaran mewakili bilangan neuron yang melakukan tugas yang tertentu seperti mengira jumlah pemberat bagi masukan dan menghantar keputusan kepada fungsi pengaktifan. Dalam satu rangkaian neural, keluaran bagi satu lapisan dihantar sebagai masukan untuk lapisan seterusnya. Berdasarkan Jadual 3.1, dengan mempertimbangkan saiz masukan 64 x 64 x 3 (tinggi imej x lebar imej x saluran mewakili warna). Saiz kernel pula adalah 3 x 3, dengan 64 penapis, lapisan akan mempelajari sebanyak 64 penapis bersaiz 3 x 3. Nilai langkah ataupun *stride* adalah menggunakan nilai asal iaitu satu. Maka, saiz keluaran bagi setiap pemetaan fitur adalah seperti contoh berikut,

$$\text{Saiz keluaran} = \text{saiz masukan} - \text{saiz kernel} + 1 \quad \dots(1)$$

Maka, bagi dimensi ketinggian,

$$\text{Ketinggian keluaran} = \text{ketinggian masukan} - \text{ketinggian kernel} + 1 \quad \dots(2)$$

Maka,

$$\text{Ketinggian keluaran} = (64 - 3 + 1) = 62$$

Bagi dimensi kelebaran pula,

$$\text{Kelebaran keluaran} = \text{kelebaran masukan} - \text{kelebaran kernel} + 1 \quad \dots(3)$$

Maka,

$$\text{Kelebaran keluaran} = (64 - 3 + 1) = 62$$

Akhirnya diperolehi saiz keluaran menjadi (62 x 62 x 64) dan begitu juga untuk lapisan konvolusi yang seterusnya.

Lapisan berkelompok pula akan memberikan saiz keluaran yang sama dengan masukan iaitu (60 x 60 x 64). Pada lapisan pengumpulan keempat, iaitu lapisan keempat, saiz pengumpulan ditetapkan kepada (2 x 2). Saiz pengumpulan akan meringkaskan nilai masukan bergantung kepada saiz pengumpulan, iaitu 2 dalam konteks ini, akan mengecilkan saiz input kepada separuh dari nilai asal kepada (30 x 30 x 64). Seterusnya, nilai-nilai yang diperolehi akan dipermudahkan lagi dengan menukar isyarat kepada matrik satu dimensi pada lapisan pengumpulan purata 2D – saiz 64, yang akan melalui setiap lapisan padat (saiz 256) dan lapisan normalisasi kelompok (saiz 256). Lapisan padat terakhir adalah di mana pengelasan akan berlaku, iaitu saiz keluaran adalah 3, iaitu mewakili 3 kelas yang berbeza. Bagi setiap lapisan juga, didapati mengandungi parameter dengan nilai ‘tiada’. Nilai tersebut adalah mewakili saiz dimensi kelompok bagi masukan tensor, di mana, bagi sesebuah lapisan konvolusi, nilai saiz kelompok tidak ditetapkan, dan ditetapkan kepada ‘tiada’.

Untuk membantu pembinaan kod aturcara secara lebih sistematik, kod pseudo digunakan untuk memberi gambaran awalan kepada kod model yang akan dibina. Rajah 3.11 menunjukkan kod pseudo yang dirujuk untuk membina model asas 2DCNN. Selain itu, carta alir untuk eksperimen menggunakan model 2DCNN asas ditunjukkan dalam Rajah 3.12. Set data *swimClass0* dan *swimClass* akan dirambat

secara berasingan melalui rangkaian. Saiz bingkai juga akan diubah kepada dua saiz yang berbeza iaitu 32 x 32 grid serta saiz 64 x 64 grid.

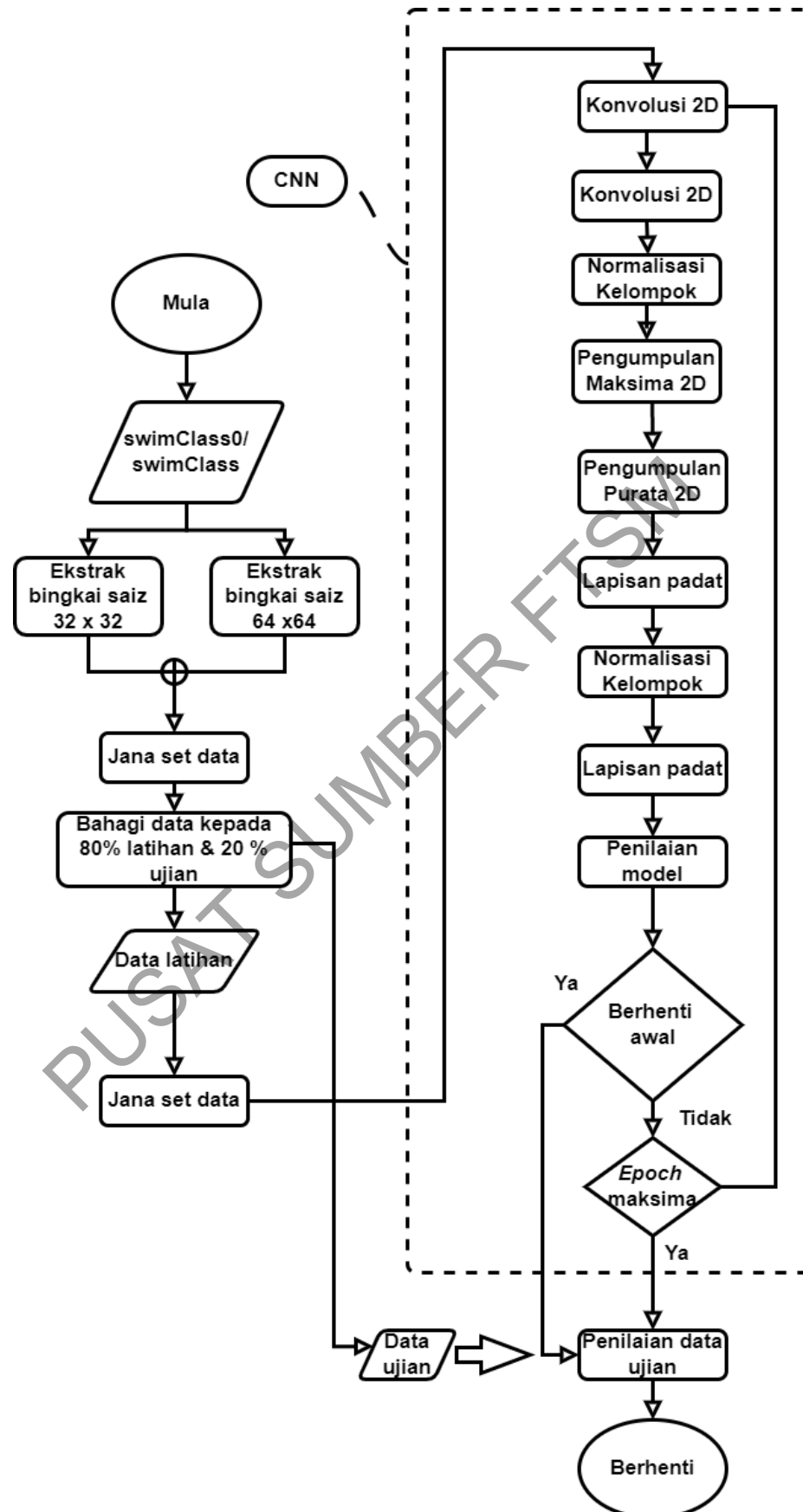
Set data yang terdiri dari tiga kelas, setiap satunya akan dipecahkan melalui tetapan *validation_split* pada fungsi `model.fit()` kepada dua bahagian, iaitu set pertama untuk tujuan latihan model, mengambil 80% dari semua data, manakala baki sebanyak 20% pula digunakan untuk tujuan validasi model.

```

1: procedure 2DCNN(inputDir)
2:   X = [ ]
3:   Y = [ ]
4:   classesList = os.listdir(inputDir)
5:   i = 0
6:   for c in classesList do
7:     Ytemp = np.zeros(shape = (numClasses))
8:     Ytemp[i] = 1
9:     print(c)
10:    fileList = os.listdir(os.path.join(inputDir, c))
11:    for f in fileList do
12:      frames = framesExtraction(os.path.join(os.path.join(inputDir, c), f))
13:    end for
14:  end for
15:  X = np.asarray(X)
16:  Y = np.asarray(Y)
17:  return X, Y
18:  X, Y = createData(dataFolder)
19:  xTrain, xTest, yTrain, yTest = trainTestSplit(X, Y)
20:  model.add(conv2D(filters=64, kernelSize = (3x3), activation = 'relu')
21:  model.add(conv2D(filters=64, kernelSize = (3x3), activation = 'relu')
22:  model.add(conv2D(filters=64, kernelSize = (3x3), activation = 'relu')
23:  model.add(BatchNormalization)
24:  model.add(MaxPooling2D(pool_size=2,2))
25:  model.add(GlobalAveragePooling2D)
26:  model.add(Dense(256, activation = 'relu'))
27:  model.add(BatchNormalization())
28:  model.add(Dense(model_output_size, activation = 'softmax'))
29: end procedure

```

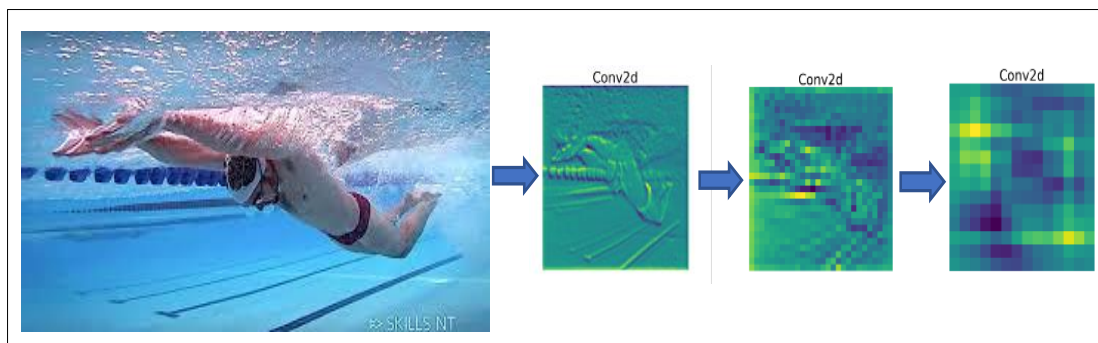
Rajah 3.11 Kod pseudo bagi model cadangan dengan dua lapisan konvolusi



Rajah 3.12 Carta alir eksperimen 2DCNN dengan dua lapisan konvolusi

Video akan ditukar kepada satu jujukan gambar yang akan diwakili dengan bingkai yang berasingan. Bergantung kepada set eksperimen, bingkai akan dibahagikan mengikut saiz grid samada 32×32 atau 64×64 . Kemudian setiap bingkai perlu ditukar kepada bentuk matriks melalui fungsi `create_dataset()` menggunakan fungsi *numpy* iaitu `np.asarray()`. Memandangkan set data adalah dari jenis kategori, ia perlu ditukar kepada bentuk numerik. *One-hot encoding* dijalankan supaya ianya dapat digunakan semula pada bahagian-bahagian tertentu dalam model nanti terutama semasa pengelasan.

Untuk operasi konvolusi, saiz *kernel* ditetapkan kepada 3×3 , dan nilai langkah ataupun *stride* adalah menggunakan nilai asal iaitu satu setiap langkah. Selepas saiz diselaraskan, kesemua nilai dalam matriks akan dibahagi pula kepada nilai 255 untuk tujuan normalisasi semua nilai sebelum dirambat melalui rangkaian konvolusi. Lapisan konvolusi dalam rangkaian yang dibina ini mempunyai dua lapisan yang berasingan dan disusun secara berturutan. Melalui lapisan konvolusi yang pertama, akan diperolehi keluaran dengan saiz $62 \times 62 \times 64$. Ini kemudian akan dihantar ke lapisan konvolusi kedua dan akan mengeluarkan keluaran dengan saiz $60 \times 60 \times 64$. Selepas melalui lapisan-lapisan konvolusi, setiap bingkai akan mengalami transformasi daripada imej yang dapat dikenalpasti, sehingga menjadi satu perwakilan imej kepada bentuk matriks. Matriks yang diperolehi sekiranya diwakili dalam bentuk grafik tidak memberikan apa-apa makna pada mata manusia, tetapi ianya lebih difahami oleh rangkaian dan berguna dalam membuat penilaian seperti perwakilan yang ditunjukkan dalam Rajah 3.13. Semakin bertambahnya bilangan lapisan konvolusi, ia akan menjadi mustahil untuk ditafsirkan oleh manusia secara langsung. Selepas melalui dua lapisan konvolusi pada peringkat awal akan memberikan nilai parameter yang besar, maka, bagi tujuan memudahkan pengiraan pada peringkat yang seterusnya, lapisan normalisasi berkelompok diperkenalkan. Lapisan normalisasi, menurut (Kohler et al., 2018), adalah bertujuan untuk mempercepatkan proses latihan, mengurangkan nilai pemberatan yang sebelumnya, serta mengawal atur model.



Rajah 3.13 Transformasi imej contoh melalui tiga lapisan konvolusi

Selepas nilai parameter dikecilkan melalui lapisan normalisasi, ia akan melalui dua lapisan pengumpulan secara berturutan iaitu lapisan pengumpulan maksimum, serta lapisan pengumpulan purata. Menurut O'Shea & Nash (2015), lapisan pengumpulan mengurangkan saiz parameter latihan dengan melakukan operasi matriks dalam mencari samada nilai maksimum untuk lapisan pengumpulan maksimum ataupun purata dalam lapisan pengumpulan purata. Nilai keluaran selepas dari kedua-dua lapisan pengumpulan akan dihantar pula kepada lapisan seterusnya iaitu lapisan tersambung sepenuhnya. Lapisan ini tersambung antara satu sama lain pada bahagian sebelum dan selepasnya. Menurut Albawi et al., (2017), lapisan tersambung sepenuhnya memperkenalkan sambungan yang kompleks yang akan menjejaskan masa pengiraan. Maka, sekali lagi, pada lapisan seterusnya diletakkan satu lagi lapisan normalisasi berkelompok yang kedua yang akan mengurangkan nilai parameter sebelumnya secara mendadak. Selepas itu, parameter tadi akan melalui lapisan terakhir dalam rangkaian CNN ini yang juga lapisan tersambung sepenuhnya yang terakhir sebelum pengkelasan dapat dibuat.

Beberapa lapisan memerlukan penggunaan fungsi pengaktifan di mana ia memainkan peranan yang penting dalam sesebuah rangkaian dengan mempengaruhi nilai keluaran sebagai hasil dari operasi antara pemberat beserta masukan sekaligus menentukan pengaktifan bagi sesuatu neuron. Dua jenis fungsi pengaktifan yang biasa digunakan iaitu *ReLU* serta *SoftMax*. *ReLU* boleh digunakan dalam kedua-dua lapisan tersembunyi mahupun lapisan keluaran di mana ia digunakan pada lapisan konvolusi serta lapisan padat, ia memberikan nilai sifar sekiranya nilai masukan adalah kurang dari sifar. Fungsi aktivasi *SoftMax* pula digunakan pada lapisan padat terakhir juga

berperanan sebagai lapisan keluaran terakhir untuk pengkelasan di mana *SoftMax* digunakan bagi tugas pengkelasan dengan lebih dari dua kelas.

Bagi memastikan model rangkaian konvolusi mencapai prestasi yang diperlukan dalam tugas pengelasan, proses latihan dijalankan secara berulang di dalam satu lelaran ataupun *epoch*. Pada penghujung setiap lelaran, nilai ketepatan akan dihitung sambil model berusaha untuk meminimakan kehilangan. Setiap kali berakhirnya satu-satu lelaran, pembolehkan seperti pemberat dan bias akan diubah yang memberikan nilai ketepatan maksimum serta nilai kehilangan yang minimum. Untuk eksperimen ini, nilai maksimum lelaran ditetapkan berdasarkan syor oleh Vrskova et al., (2022) dengan 35 kitaran sebagai memadai. Bagaimanapun, sekiranya model berhenti dari memperbaiki nilai ketepatan berbanding kehilangan, latihan akan dihentikan lebih awal dengan menetapkan sebanyak 10 kitaran. Ini bermakna, sekiranya nilai kehilangan tidak menjadi semakin baik secara berturutan dalam 10 lelaran, latihan akan dihentikan lebih awal sebelum mencapai nilai 35 lelaran dengan menetapkan fungsi `EarlyStopping()`. `EarlyStopping()` juga berperanan untuk mengelakkan berlakunya *overfitting* berlaku dalam sesebuah model pembelajaran dalam (Brownlee 2018). Bagi tujuan pengoptimaan pula, beberapa algoritma boleh dipertimbangkan, akan tetapi algoritma pengoptimaan latihan *Adaptive Moment Estimation (Adam)*, diperkenalkan oleh Kingma & Ba (2015) digunakan dalam kajian ini. *Adam* yang berasaskan pengoptimaan penurunan kecerunan menjadi algoritma pilihan disebabkan kelebihan pada masa pengiraan yang lebih pantas apabila digunakan terhadap set data yang besar dan mengandungi banyak parameter.

3.5.2 Model 2DCNN Dengan Tiga Lapisan Konvolusi

Di sini, model asas 2DCNN dilanjutkan dengan hanya menambah satu lagi lapisan konvolusi secara berturutan dengan dua lapisan konvolusi terawal seperti yang disenaraikan dalam Jadual 3.2. Selain dari itu, tiada perubahan dilakukan terhadap lapisan-lapisan yang lain, termasuk saiz keluaran bagi setiap lapisan adalah sama. Sama seperti model 2DCNN asas, aturcara di bina berdasarkan Rajah 3.14 dan carta alir eksperimen menggunakan model 2DCNN dengan tiga lapisan konvolusi adalah berdasarkan Rajah 3.15. Secara dasarnya, hampir keseluruhan seni bina berkongsi tetapan serta aliran kerja yang sama, akan tetapi perbezaan adalah pada penambahan

satu lagi lapisan konvolusi untuk menguji kesan pertambahan lapisan konvolusi dalam peningkatan prestasi model.

Jadual 3.2 Senarai Lapisan *sequential* dalam Model 2DCNN dengan tiga lapisan konvolusi

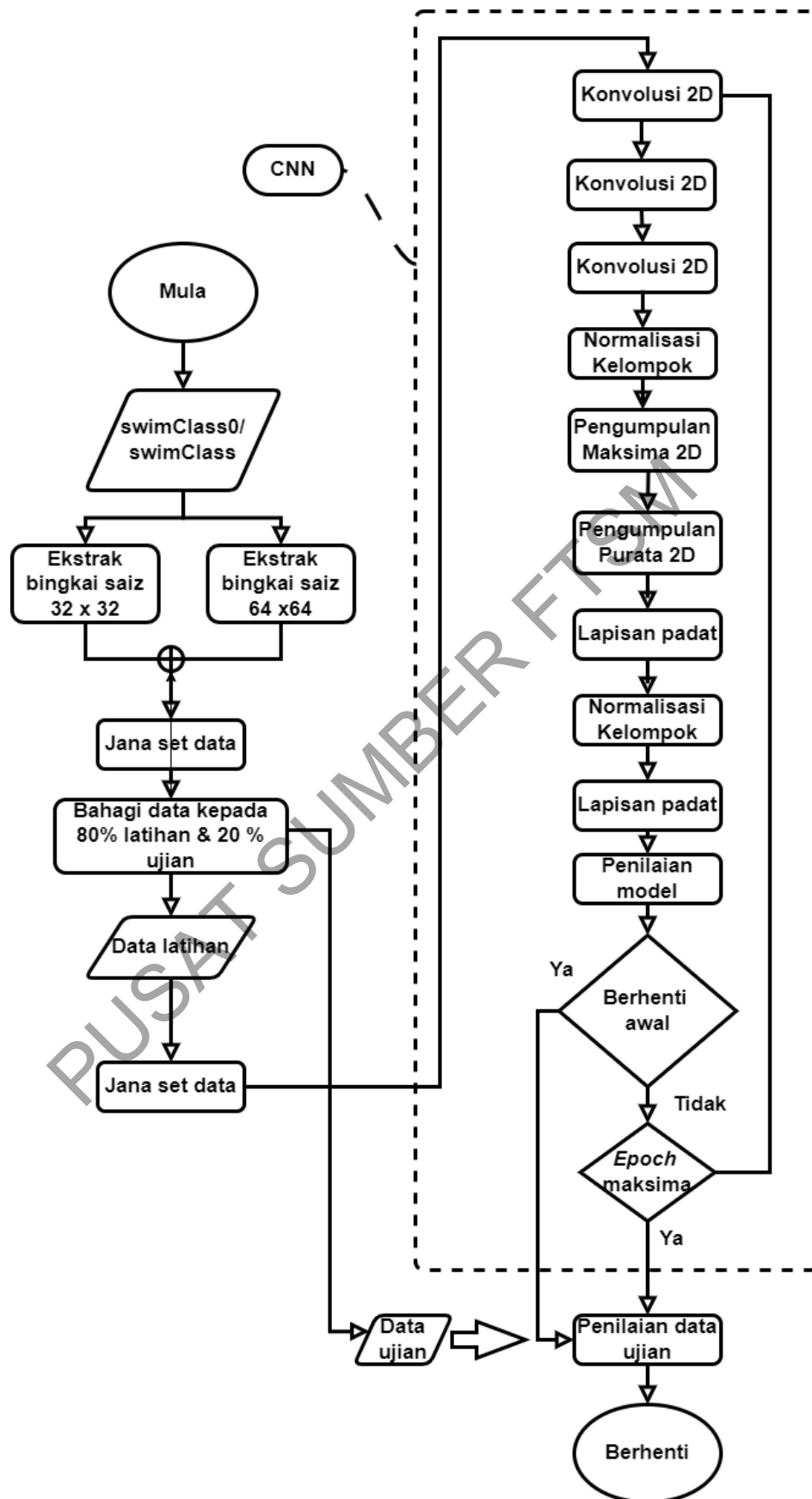
Bil	Lapisan	Saiz Keluaran
1	Konvolusi 2D	(Tiada, 62, 62, 64)
2	Konvolusi 2D	(Tiada, 60, 60, 64)
3	Konvolusi 2D	(Tiada, 60, 60, 64)
4	Normalisasi Kelompok	(Tiada, 60, 60, 64)
5	Pengumpulan Maksima 2D	(Tiada, 30, 30, 64)
6	Pengumpulan Purata 2D	(Tiada, 64)
7	Lapisan Padat	(Tiada, 256)
8	Normalisasi Kelompok	(Tiada, 256)
9	Lapisan Padat	(Tiada, 3)

```

1: procedure 2DCNN(inputDir)
2:   X = [ ]
3:   Y = [ ]
4:   classesList = os.listdir(inputDir)
5:   i = 0
6:   for c in classesList do
7:     Ytemp = np.zeros(shape = (numClasses))
8:     Ytemp[i] = 1
9:     print(c)
10:    fileList = os.listdir(os.path.join(inputDir, c))
11:    for f in fileList do
12:      frames = framesExtraction(os.path.join(os.path.join(inputDir, c), f))
13:    end for
14:  end for
15:  X = np.asarray(X)
16:  Y = np.asarray(Y)
17:  return X, Y
18:  X, Y = createData(dataFolder)
19:  xTrain, xTest, yTrain, yTest = trainTestSplit(X, Y)
20:  model.add(conv2D(filters=64, kernelSize = (3x3), activation = 'relu')
21:  model.add(conv2D(filters=64, kernelSize = (3x3), activation = 'relu')
22:  model.add(conv2D(filters=64, kernelSize = (3x3), activation = 'relu')
23:  model.add(BatchNormalization)
24:  model.add(MaxPooling2D(pool_size=2,2))
25:  model.add(GlobalAveragePooling2D)
26:  model.add(Dense(256, activation = 'relu'))
27:  model.add(BatchNormalization())
28:  model.add(Dense(model_output_size, activation = 'softmax')
29: end procedure

```

Rajah 3.14 Kod pseudo bagi model cadangan dengan tiga lapisan konvolusi



Rajah 3.15 Carta alir eksperimen 2DCNN dengan tiga lapisan konvolusi

3.6 TETAPAN EKSPERIMEN

Pencapaian sesebuah model pembelajaran dalam juga ditentukan oleh tetapan atau hiperparameter yang boleh diubah sebelum melatih sesebuah model. Antara hiperparameter yang berkaitan adalah seperti kadar pembelajaran, bilangan lapisan tersembunyi, fungsi pengaktifan dan sebagainya. Hiperparameter kebiasaannya ditetapkan berdasarkan konsep cuba jaya dan ianya memakan masa. Untuk kajian ini, tetapan serta hiperparameter hanya diubah secara minimal. Maka, untuk setiap gabungan tetapan yang berlainan, termasuk juga variasi set data memerlukan eksperimen dilakukan dalam set yang berasingan. Tetapan untuk setiap set yang berasingan adalah seperti yang dinyatakan dalam Jadual 3.3. Maka untuk eksperimen yang dijalankan, akan dipelbagaikan gabungan seperti set data *swimClass0* ataupun *swimClass*, saiz input samada 32 x 32 atau 64 x 64, dan bilangan lapisan konvolusi.

Jadual 3.3 Senarai eksperimen dan gabungan tetapan serta set data

Eksperimen	Set Data	Saiz input	Bilangan lapisan konvolusi
1	<i>swimClass0</i>	32 x 32	2
2	<i>swimClass0</i>	64 x 64	2
3	<i>swimClass0</i>	64 x 64	3
4	<i>swimClass</i>	32 x 32	2
5	<i>swimClass</i>	64 x 64	2
6	<i>swimClass</i>	64 64	3

3.7 FASA PENILAIAN MODEL

Fasa penilaian model adalah fasa yang penting untuk membuat perbandingan, seterusnya menentukan model yang terbaik dalam sesebuah kajian. Dalam kajian ini, penilaian akan dibahagikan kepada dua fasa, yang pertama adalah penilaian metrik pencapaian model, kedua adalah membuat perbandingan terhadap kajian sedia ada. Penilaian sesuatu model pembelajaran dalam melibatkan penaksiran pencapaian model untuk sesuatu tugas dan membandingkan terhadap sesuatu metrik yang telah ditetapkan. Pengujian keberkesanan model perlu dinilai dengan kaedah penilaian yang

bersesuaian untuk model pembelajaran mendalam atau model CNN secara khususnya. Beberapa metrik penilaian digunakan dalam kajian ini untuk memastikan model berupaya melakukan tugas pengelasan seperti yang dijangkakan. Metrik-metrik penilaian yang digunakan dibincang dalam bahagian selanjutnya.

3.7.1 Penilaian Model: Fungsi Kehilangan

Idea asas penilaian yang pertama adalah berdasarkan keupayaan model meneka output berdasarkan label yang betul, ianya berkaitan dengan fungsi kehilangan atau *loss function*, di mana model kajian ini menggunakan fungsi entropi silang kategori berdasarkan kelas yang telah dilabel menggunakan kaedah *one-hot encoding*. Entropi, $H(X)$ diungkap oleh satu taburan kebarangkalian dengan nilai pembolehubah X , sebagai:

$$H(X) = \sum_x y \log(\hat{y}) \quad (1)$$

di mana,

y mewakili vektor *one-hot encoding* mewakili taburan kebarangkalian benar setiap kelas

\hat{y} , ialah vektor mewakili taburan kebarangkalian jangkaan bagi setiap kelas, diperolehi dengan menggunakan fungsi *Softmax* terhadap keluaran model

Untuk sesebuah model CNN dengan pemberat dan bias sentiasa dilaraskan supaya nilai kehilangan terkecil diperolehi, kehilangan entropi silang, L_{CE} dinyatakan sebagai:

$$L_{CE} = -[\sum_{i=1}^n t_i \log(p_i)] \quad (2)$$

di mana,

t_i adalah label benar dan

p_i adalah kebarangkalian *Softmax* bagi kelas ke i .

Kehilangan silang entropi adalah fungsi kehilangan yang biasa digunakan dalam pembelajaran dalam memandangkan ia sesuai dalam mengoptima model yang menjangka kebarangkalian. Tujuan melatih model pembelajaran dalam kebiasaanya adalah bertujuan untuk mengurangkan kehilangan silang entropi.

3.7.2 Penilaian Model: Ketepatan

Ketepatan pengelasan pula menyatakan keupayaan sesebuah model dalam melakukan pengelasan, ia diukur berdasarkan nisbah bilangan jangkaan yang betul terhadap bilangan kekerapan jangkaan dibuat:

$$\text{Ketepatan} = \frac{\text{Bilangan jangkaan betul}}{\text{Bilangan keseluruhan kekerapan}} \quad (3)$$

Kedua-dua kehilangan serta ketepatan dikira pada setiap lelaran, dan nilai tersebut digunakan oleh model dalam memperbaiki pencapaian model dengan mengubah nilai pemberat.

3.7.3 Penilaian Model: Jadual Kontigensi, Ketepatan, Kepersisan dan Ingatan Semula

Bagi pengelasan kelas pelbagai, ukuran pencapaian boleh juga dinyatakan dengan cara yang lebih baik dengan menggunakan jadual kontigensi ataupun *confusion matrix*. Penggunaan ukuran ketepatan semata cuma menyatakan pencapaian model secara keseluruhan, jadual kontigensi memberikan maklumat mengenai ralat serta jenis ralat berhubung sesebuah kerja pengelasan. Selain itu, ia sesuai digunakan bagi set data yang tidak seimbang, maka penggunaan jadual kontigensi adalah bersesuaian dengan kajian ini. Ia juga membantu dalam menyediakan metrik pengukuran yang lain seperti kejituan dan ingatan semula.

Jadual 3.4 Jadual kontigensi pelbagai kelas

		Benar				
		k	1	2	...	n
Diramal	1	$C_{1,1}$	$C_{1,2}$...	$C_{1,k}$	
	2	$C_{2,1}$	$C_{i,j}$...	:	
	:	:	:	:	:	
	
	n	$C_{k,1}$	$C_{k,k}$	

Jadual kontigensi diwakili dalam bentuk jadual berdasarkan hasil ramalan oleh sesuatu model pengkelasan dengan menunjukkan bagi setiap positif benar, negatif benar, positif palsu dan negatif benar seperti di dalam Jadual 3.4. Dengan diwakili oleh satu matriks dengan saiz $n \times n$, di mana n adalah mewakili bilangan kelas. Setiap lajur mewakili label jangkaan dan setiap baris pula mewakili label benar. Daripada jadual kontigensi tersebut, bilangan ramalan positif benar TP bagi setiap kelas k adalah seperti berikut:

$$TP_k = C_{k,k} \quad (4)$$

di mana,

k adalah merujuk kepada kelas individu

Seterusnya, bilangan ramalan negative palsu FN bagi setiap kelas benar k dinyatakan seperti berikut:

$$FN_k = \sum_{i=1, i \neq k}^n \sum_{j=1, j \neq k}^n C_{i,j} \quad (5)$$

di mana,

n = jumlah bilangan kelas

i = baris kelas dijangka

j = lajur kelas benar

Bilangan ramalan positif palsu FP bagi setiap kelas k boleh dikira mengikut persamaan berikut:

$$FP_k = \sum_{j=1, j \neq i, i=k}^n C_{j,i} \quad (6)$$

Menggunakan nilai-nilai yang dirujuk kepada Jadual 3.4, setiap metrik penilaian, kepersisan dan ingatan semula boleh dinyatakan seperti berikut:

$$\text{Ketepatan kelas} = \frac{TP+TN}{TP+FN+FP+TN} \quad (7)$$

$$\text{Kepersisan} = \frac{TP}{TP+FP} \quad (8)$$

$$\text{Ingatan semula} = \frac{TP}{TP+FN} = \frac{TP}{P} \quad (9)$$

Di mana kepersisan adalah metrik dalam menilai bilangan positif benar yang diperolehi, manakala ingatan semula atau *recall* pula untuk menilai bilangan positif benar diperolehi dari nilai positif yang sebenar. Kesemua metrik penilaian tersebut, digunakan secara meluas dalam menentukan prestasi sesebuah model pembelajaran mesin dan model pembelajaran mendalam secara lebih menyeluruh dan digunakan untuk menilai prestasi model dalam kajian ini.

3.7.4 Penilaian Model: Penilaian Perbandingan

Bagi menilai keupayaan model dengan lebih lanjut, perbandingan kajian terdahulu seumpama perlu dijalankan. Walaubagaimanapun, selepas kajian literasi dibuat secara menyeluruh didapati tiada kerja seumpama yang menghampiri kajian ini, kecuali beberapa kajian seperti yang dinyatakan dalam kajian literasi seperti Omae et al., (2017) & Liao et al., (2003). Omae et al., (2017) membangunkan model menggunakan data sensor dipasang pada badan perenang untuk tujuan pengelasan renangan. Manakala, Liao et al., (2003) menggunakan analisis video untuk tujuan pengelasan, akan tetapi model berasaskan pokok keputusan berdasarkan analisis regresi terhadap

segmen warna dan corak pergerakan yang spesifik untuk gaya renang tertentu. Kedua-dua kajian adalah tidak koheren bagi tujuan perbandingan. Maka pendekatan alternatif yang koheren perlu dicadangkan supaya argumen bersesuaian dapat dibina sebagai asas perbandingan antara kajian nanti.

Antara kajian yang paling dominan dalam bidang pengelasan jujukan imej dan video adalah kajian terhadap pengelasan 101 kelas pergerakan manusia menggunakan set data UCF101. Beberapa kajian yang terkini, dengan pencapaian yang terbaik adalah seperti model SMART oleh Gowda et al., (2021) dengan 98.64% ketepatan, Duan et al., (2020) pula mencadangkan OmniSource dengan ketepatan sebanyak 98.6%, dan Y. Li et al., (2022) dengan PERF-Net memperolehi ketepatan pengelasan sebanyak 98.6%. Walaubagaimanapun, pembangunan menggunakan set data UCF101 memerlukan kepada penggunaan perkakasan komputer dengan spesifikasi yang tinggi. Ia akan menjadi limitasi untuk dilaksanakan bagi tujuan penilaian model bagi kajian ini. Vrskova et al., (2022) menjalankan pilihan penilaian alternatif dengan menggunakan sebahagian sahaja dari set data penuh iaitu menggunakan set data UCF15. Menggunakan set data tersebut, Vrskova et al., (2022) telah memperolehi ketepatan pengelasan sebanyak 84.4%.

Maka, kajian oleh Vrskova et al., (2022) sesuai digunakan untuk membuat penilaian secara perbandingan dengan model kajian ini untuk menentu sah keupayaan model. Sekiranya model boleh digunakan terhadap set data penanda aras, maka ia juga berupaya untuk digunakan terhadap set data yang berlainan. Untuk itu, model dengan salah satu dari tetapan dari keenam-enam eksperimen boleh dipilih bagi tujuan tersebut menggunakan set data UCF15. Model kemudiannya akan dinilai menggunakan metrik penilaian yang sama dengan yang sebelumnya iaitu dari segi fungsi kehilangan, ketepatan, kepersisan dan ingatan semula.

3.8 BAHAN DAN KAEDAH

Pembangunan model analisis video melibatkan serba-serbi besar, dari segi saiz data sehinggalah penggunaan perkakasan. Perkakasan komputer memerlukan penggunaan pemprosesan yang ekstensif seperti CPU pelbagai bilangan teras, saiz memori yang besar di samping penggunaan pemproses unit grafik berkeupayaan tinggi untuk

melakukan operasi komputeran selari. Bahasa pengaturcaraan pula perlu menggunakan bahasa pengaturcaraan tahap tinggi seperti C++, JavaScript ataupun Python, selain itu, ia perlu mempunyai sokongan perpustakaan, rangka kerja dan sebagainya untuk kegunaan pembangunan model rangkaian neural serta penilaian.

3.8.1 Perisian dan Perkakasan

Bagi memenuhi keperluan-keperluan tersebut, model rangkaian konvolusi ini ditulis menggunakan bahasa pengaturcaraan Python. Python menyokong keperluan-keperluan pembinaan model dengan cepat serta kesediaan pelbagai perpustakaan, rangka kerja serta API untuk pembinaan model pembelajaran dalam. Beberapa perpustakaan serta API digunakan untuk memudahkan tugas pengaturcaraan seperti OpenCV, Numpy, Keras, Tensorflow, Scikit-learn, serta Matplotlib seperti yang disenaraikan dalam Jadual 3.5.

Jadual 3.5 Senarai perpustakaan Python serta fungsi

Perpustakaan / Rangka kerja	Fungsi
Numpy	Pemprosesan pelbagai dimensi tatasusunan dengan operasi algebra linear
Scikit-learn	Mengandungi beberapa algoritma pembelajaran mesin
Tensorflow	Pengiraan tensor terutama untuk pembelajaran dalam
Keras	Antara muka untuk penggunaan rangkaian saraf buatan
Matplotlib	Perpustakaan untuk visualisasi data
Seaborn	Perpustakaan untuk visualisasi data berasaskan Matplotlib

Bagi perkakasan, pada kebanyakan masa, kerja-kerja dilakukan secara bertukar ganti di antara komputer peribadi dengan platform buku nota awan iaitu Colab yang tersedia secara percuma oleh Google. Untuk komputer peribadi, analisis dijalankan pada sebuah komputer riba yang mempunyai pemproses Intel ® Core i7-10750H @ 2.60GHz yang terdiri dari 4 teras fizikal dengan 24GB memori capaian rawak (RAM). Selain itu, untuk unit pemproses grafik (GPU), ia menggunakan