

**PENGECAMAN MIKRO-EKSPRESI RINGAN PADA
PANGKALAN DATA KOMPOSIT**

NUR AISHAH BINTI AB RAZAK

UNIVERSITI KEBANGSAAN MALAYSIA

PENGECAMAN MIKRO-EKSPRESI RINGAN PADA PANGKALAN DATA
KOMPOSIT

NUR AISHAH BINTI AB RAZAK

PROJEK YANG DIKEMUKAKAN UNTUK MEMENUHI SEBAHAGIAN
DARIPADA SYARAT MEMPEROLEHI IJAZAH SARJANA SAINS KOMPUTER
(KECERDASAN BUATAN)

FAKULTI TEKNOLOGI DAN SAINS MAKLUMAT
UNIVERSITI KEBANGSAAN MALAYSIA
BANGI

2022

LIGHTWEIGHT MICRO-EXPRESSION RECOGNITION ON COMPOSITE
DATABASE

NUR AISHAH BINTI AB RAZAK

DISSERTATION SUBMITTED IN PARTIAL FULFILMENT OF THE
REQUIREMENTS FOR THE DEGREE OF MASTER OF COMPUTER SCIENCE
(ARTIFICIAL INTELLIGENCE)

FAKULTY OF INFORMATION SCIENCE AND TECHNOLOGY
UNIVERSITI KEBANGSAAN MALAYSIA
BANGI

2022

PENGAKUAN

Saya dengan ini mengisytiharkan bahawa kerja-kerja dalam tesis ini adalah milik saya sendiri kecuali petikan dan ringkasan yang telah dijelaskan sumbernya.

11 Februari 2022

NUR AISHAH BINTI AB RAZAK

P104371

PENGHARGAAN

Saya ingin merakamkan setinggi-tinggi penghargaan kepada penyelia saya, Dr Shahnorbanun atas bimbingan dan komitmen beliau yang telah membantu sepanjang perjalanan penyelesaian tesis saya. Objektivitinya ditambah dengan kepercayaannya terhadap keupayaan saya adalah apa yang saya perlukan untuk terus mengejar matlamat saya.

Penyelidikan ini juga tidak akan dapat diselesaikan tanpa bantuan daripada Prof. Xiaolan Fu dari Institut Psikologi, Akademi Sains Cina, Dr. Moi Hoon Yap dari Manchester Metropolitan University dan Prof. Xiaobai Li dari University of Oulu kerana memberi kebenaran mendapatkan set-set data yang diperlukan untuk tesis ini. Kepada mereka, saya ucapkan ribuan terima kasih.

Di kesempatan ini saya juga berterima kasih kepada Madhumita Takalkar, dari University of Sydney kerana cukup baik hati berkongsi kertas kajiannya yg berakses terhad dan menjawab soalan yang berkaitan dengannya.

Akhirnya, kepada keluarga dan anak-anak saya yang saya kerana memahami dan bersabar semasa saya menjalankan penyelidikan dan menulis laporan - terima kasih yang tidak terhingga.

Yang paling penting, kepada Allah Yang Maha Kuasa kerana menunjukkan kepada saya jalan melalui segala kesukaran dan memberi saya kekuatan dan ketabahan.

ABSTRAK

Potensi memanfaatkan ekspresi mikro dalam pelbagai bidang seperti keselamatan, penjagaan kesihatan dan pendidikan telah meningkatkan minat dalam bidang ini. Tidak seperti ekspresi muka biasa, ekspresi mikro adalah kabur dan berlaku dengan pantas menjadikannya tidak mudah dikecam. Selain itu, kebanyakan pendekatan pengecaman ekspresi mikro (Micro Expression Recognition, MER) hari ini adalah kompleks dan memerlukan sumber pengiraan yang banyak dalam pra-pemprosesan. Prestasi pengecaman ekspresi mikro dalam Penilaian Pangkalan Data Komposit (Composite Database Evaluation, CDE) yang mengandungi data dari pelbagai etnik, jantina dan kumpulan umur juga masih boleh ditambah-baik dengan cabaran yang berpunca daripada set data yang tidak seimbang dan kecil. Penyelidikan ini bertujuan untuk menunjukkan bahawa pemindahan pembelajaran daripada pangkalan data ekspresi makro melalui pembelajaran mendalam ringan boleh mencapai prestasi MER yang tinggi dengan hanya menggunakan imej statik sebagai input bagi mengurangkan kerumitan pra-pemprosesan. Masalah set data tidak seimbang juga didekati sebagai masalah penalaan algoritma dan bukan hanya masalah kejuruteraan dan penghasilan data. Model MER yang dicadangkan dibangunkan daripada model asas EfficientNet-B0 yang dipangkas lalu dilatih pada pangkalan data ekspresi makro sebelum diperhalusi pada set data CDE. Model yang terhasil terdiri daripada 5 blok pembalikan mobil (MBConv) dengan sebanyak 15 lapisan serta mempunyai hanya 867K parameter. Penalaan algoritma mudah yang memanipulasi fungsi kehilangan yang menekankan kelas minoriti dicadangkan untuk menangani set data yang tidak seimbang. Keputusan eksperimen menggunakan pengesahan silang LOSO (Leave-One-Subject-Out) menunjukkan peningkatan prestasi yang ketara berbanding dengan metod terdahulu dengan skor (Purata F1) UF1 0.98 dan skor (Purata Recall) UAR 1.77.

Kata kunci: Pengecaman ekspresi mikro, penilaian pangkalan data komposit, EfficientNet, pemindahan pembelajaran, set data tidak seimbang, rangkaian ringan, fungsi kehilangan

ABSTRACT

The potential of leveraging micro-expression in various areas such as security, health care and education has intensified interests in this area. Unlike normal facial expression, micro-expression is subtle and occurs rapidly making it not easily discernible. Moreover, most micro-expression recognition (MER) approaches today are complex and require computationally expensive pre-processing. Micro-expression recognition performance in Composite Database Evaluation (CDE) with variety of ethnicities, gender and age group also still leaves a lot to be desired with challenges stemming from imbalanced and small dataset. This work seeks to demonstrate how transfer learning from macro-expression database (FER 2013) in a lightweight deep-learning network can achieve high MER performance using only static images as input, reducing pre-processing complexity. This research also re-defined the imbalanced dataset problem as an algorithm-tuning problem instead of data engineering and generation problem to lighten the pre-processing steps. The proposed MER model is developed from truncated EfficientNet-B0 model that is pre-trained on a macro-expression database (FER 2013) before fine-tuned on the CDE dataset. The resultant model consists of 5 mobile inverted bottleneck, MBConv blocks with a total of 15 layers with only 867k parameters. A simple algorithm-tuning that manipulates the loss function to place more importance on minority classes is suggested to deal with the imbalanced dataset. Experimental results using Leave-One-Subject-Out cross validation show substantial performance increase compared to the state-of-the-arts with (Unweighted F1) UF1 score of 0.98 and (Unweighted Average Recall) UAR score of 1.77.

Keywords: Micro-expression recognition, composite database evaluation, EfficientNet, transfer learning, imbalanced dataset, lightweight network, loss function

KANDUNGAN

		Halaman
PENGAKUAN		iii
PENGHARGAAN		iv
ABSTRAK		v
ABSTRACT		vi
KANDUNGAN		vii
SENARAI JADUAL		x
SENARAI ILLUSTRASI		xi
SENARAI SINGKATAN		xiii
BAB I	PENDAHULUAN	
1.1	Latar Belakang	1
1.2	Permasalahan Kajian	4
1.3	Skop Kajian	5
1.4	Objektif Kajian	6
1.5	Persoalan Kajian	6
1.6	Sumbangan	6
BAB II	KAJIAN LITERATUR	8
2.1	Pengenalan	8
2.2	Definisi Ekspresi Mikro	8
2.3	Pangkalan Data Ekspresi Mikro	9
	2.3.1 SMIC	10
	2.3.2 CASME II	12
	2.3.3 SAMM	13
	2.3.4 Perbincangan	14
2.4	Pendekatan Traditional	22
	2.4.1 Pendekatan Corak Perduaan Tempatan (LBP)	23
	2.4.2 Pendekatan Histogram-of-Gradient (HOG)	23
	2.4.3 Pendekatan Aliran Optik (OF)	23
2.5	Pendekatan Pembelajaran Mendalam	24

2.5.1	Ciri-ciri Input	25
2.5.2	Seni bina rangkaian	30
2.6	Pra-Pemrosesan	36
2.6.1	Pencarian bingkai puncak	36
2.6.2	Pengesanan Titik-Titik Wajah dan Penjajaran Imej	37
2.7	Penilaian dan Metrik	38
2.7.1	Penilaian Pangkalan Data Tahanan (HDE)	38
2.7.2	Penilaian Data Komposit (CDE)	38
2.7.3	Metrik	39
2.8	Kesimpulan	41
BAB III		43
3.1	Cadangan Kaedah	43
3.1.1	Mengesan Bingkai Puncak	44
3.1.2	Penggabungan Set Data	45
3.1.3	Pengesanan Muka/Mata	45
3.1.4	Penjajaran Muka dan Pemotongan Imej	47
3.1.5	Augmentasi Data	48
3.1.6	Pindah Belajar	48
3.1.7	Model cadangan	50
3.2	Uji Kaji	53
3.2.1	Persediaan Uji Kaji	53
3.2.2	Pengoptimum	56
3.2.3	Fungsi Kehilangan	58
3.2.4	Metrik Penilaian	60
3.3	Kesimpulan	60
BAB IV		62
4.1	Pengenalan	62
4.2	Keputusan Penilaian Pangkalan Data Gabungan (CDE)	62
4.3	Keputusan Penilaian Pembelajaran Pemindahan	67
4.4	Keputusan Penilaian Fungsi Kehilangan	69
4.5	Kesimpulan	71

BAB V		74
5.1	Pengenalan	74
5.2	Jurang dan kajian masa depan	75

SENARAI RUJUKAN		77
------------------------	--	-----------

LAMPIRAN

A	Pengiraan Bilangan Parameter Model	84
B	Set Data CDE	86

Pusat Sumber
FTSM

SENARAI JADUAL

No. Jadual		Halaman
Jadual 2.3	Pangkalan data ekspresi mikro	3
Jadual 2.3.4(d)	Pelabelan semula dan kelas asal untuk set data CDE	13
Jadual 2.5.2	Kaedah Pembelajaran Mendalam dalam MER	23
Jadual 2.7.2	Kajian lepas yang menggunakan penilaian CDE.	38
Jadual 3.1.6	Label dan bilangan sampel dalam pangkalan data FER 2013.	49
Jadual 3.2.1	Matlamat uji kaji	53
Jadual 4.2	Keputusan Penilaian Pangkalan Data Komposit (CDE) terhadap kajian dasar	63
Jadual 4.3	Keputusan penilaian menggunakan pangkalan data sumber ImageNet dan FER 2013	67
Jadual 4.4.1	Skor UF1 dan UAR bagi penilaian fungsi kehilangan	70
Jadual 4.4.2	Penilaian fungsi kehilangan untuk model yang telah dilatih pada FER 2013 yang tidak seimbang	71

SENARAI ILLUSTRASI

No Rajah.		Halaman
Rajah 2.3.1	Contoh imej dari pangkalan data SMIC	5
Rajah 2.3.2	Contoh imej dari set data CASME II	6
Rajah 2.3.3	Contoh imej dari set data SAMM	7
Rajah 2.3.4(b)	Contoh imej dari pangkalan data FER2013	11
Rajah 2.5.1	Contoh bingkai-bingkai untuk kelas 'Surprise' dalam set data CASME II	21
Rajah 2.5.2(b)	Blok sisa	25
Rajah. 2.5.2(d)	Modul perhatian	26
Rajah. 2.5.2(e)	Rangkaian Konvolusi Graf	27
Rajah. 2.5.2(f)-1	Blok MBConv yang digunakan dalam seni bina EfficientNet.	28
Rajah 2.5.2(f)-2	Senibina EfficientNet-B0	29
Rajah. 3.1	Cadangan kaedah kajian	43
Rajah. 3.1.3(a)	Pengesanan muka dan mata menggunakan Haar-Cascade	46
Rajah. 3.1.3(b)	Pengesanan muka dan mata menggunakan MTCNN	46
Rajah. 3.1.4	Penjajaran muka dan pemotongan imej.	47
Rajah. 3.1.8	Cadangan seni bina model EfficientNet-ME	52
Rajah. 3.2.1	Senibina Efficient-ME(t3)	54
Rajah 3.3.2(i)	2 pengoptimum teratas dalam setiap kitaran.	57
Rajah. 3.3.2(ii)	Statistik ketepatan untuk setiap pengoptimum	58
Rajah. 3.2.3	Skor UF1 dan F1 setiap kelas menggunakan persamaan pemberat kelas seimbang	59
Rajah 4.2.1	Peta ciri output Efficient-ME(tr3)	64

Rajah 4.2.2	Peta ciri output Efficient-ME	65
Rajah. 4.2.3	Ketepatan dan Skor UF1 Set Data Pengesahan dan Ujian bagi model Efficient-B0 dan Efficient-ME	66
Rajah. 4.3.2	Skor UF1 dan UAR untuk model EfficientNet-B0 pra-dilatih dengan ImageNet, Efficient-ME pra-dilatih dengan ImageNet dan Efficient-ME pra-dilatih dengan FER 2013.	69

Pusat Sumber
FTSM

SENARAI SINGKATAN

CNN	Convolutional Neural Network
MER	Micro-Expression Recognition
CDE	Composite Database Evaluation
OF	Optical Flow
LBP	Local Binary Pattern
HOG	Histogram of Gradient
FL	Focal loss
CB	Class-Balanced Loss
KNN	K-Nearest Neighbour
AU	Action Unit
MEGC	Micro-Expression Grand Challenge
FACS	Facial Action Codes

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Minat dalam pengesanan dan pengecaman ekspresi mikro telah meningkat sejak beberapa tahun kebelakangan ini kerana potensi aplikasinya dalam pelbagai bidang seperti keselamatan, pendidikan dan kesihatan awam. Ekspresi mikro atau ekspresi pergerakan mikro seperti yang dirujuk oleh (Haggard dan Isaacs 1966) ialah ekspresi pergerakan wajah yang ditahan dan disembunyikan - menganjurkan bahawa ekspresi mikro menawarkan tafsiran yang lebih tulen tentang keadaan emosi manusia daripada ekspresi muka biasa atau perkataan lisan (Ekman & Friesen 1969).

Ekspresi mikro digambarkan sebagai ekspresi wajah yang samar, spontan, dan berlaku dengan pantas serta berlangsung tidak lebih dari 0.5 saat dan hanya melibatkan otot-otot wajah tertentu (Yan et al. 2013). Oleh kerana itu, ianya tidak mudah bagi manusia untuk mengecam ekspresi mikro. Contohnya, penjaga pantai Amerika Syarikat yang dilatih mengecam ekspresi mikro menggunakan Alat Latihan Ekspresi Mikro (METT) hanya mencapai ketepatan tidak lebih dari 50% (Frank et al. 2009).

Penyelidikan-penyelidikan terdahulu mengenai Pengecaman Ekspresi Mikro (MER) lebih tertumpu kepada mencipta ciri-ciri yang boleh mewakili pergerakan otot halus dalam ekspresi mikro. Variasi Corak Binari Tempatan (LBP) (Pfister et al. 2011; Wang et al. 2014) dan *Histogram of Gradients* (HOG) (Chen et al. 2016) adalah antara ciri-ciri yang telah digunakan untuk menggambarkan maklumat tekstur dan bentuk ekspresi mikro. Untuk meningkatkan ketepatan MER, ramai penyelidik kemudiannya beralih kepada ciri-ciri berasaskan Aliran Optik (OF) seperti HOOOF (Happy & Routray 2016) dan Bi-WOOF (Liong et al. 2016b) dengan tujuan untuk mendapatkan maklumat berasaskan temporal daripada pergerakan muka yang pantas di dalam ekspresi mikro

(Song et al.2019; Sun et al.2020). Walau bagaimanapun, ciri-ciri ini perlu diekstrak terlebih dahulu sama ada secara manual atau automatik yang meningkatkan kerumitan MER dan oleh itu tidak praktikal untuk sistem pengecaman dalam kehidupan sebenar.

Seiring dengan kejayaan aplikasi kaedah pembelajaran mendalam dalam Pengecaman Ekspresi Wajah (FER), kecenderungan semasa di bidang MER kini lebih menjurus kepada pendekatan pembelajaran mendalam namun kebanyakan pelaksanaannya di dalam MER adalah menggunakan ciri-ciri Aliran Optik (OF) yang di pra-estrah dan disuap ke rangkaian pembelajaran mendalam (Khor et al. 2019; Liong et al. 2019; Xia, Z. et al. 2020a) yang lebih kompleks. Pelaksanaan lain yang menggunakan Imej Dinamik (DI) (Nie et al. 2021; Verm et al. 2020; Xia, B. et al. 2020) memerlukan pembinaan imej-imej dinamik ini terlebih dahulu sebelum digunakan sebagai input kepada rangkaian pembelajaran mendalam.

Oleh sebab pernyataan oleh (Liong et al. 2016b; Sun et al. 2020) bahawa bingkai puncak dalam video ekspresi mikro – yang merupakan bingkai apabila ekspresi wajah mencapai kemuncaknya - adalah mencukupi untuk mengekod maklumat ekspresi mikro, beberapa kerja telah cuba untuk menggunakan hanya bingkai puncak sebagai input kepada rangkaian pembelajaran mendalam untuk mengurangkan kerumitan model dan meningkatkan kecekapannya. Kebanyakan kaedah ini menggunakan pemindahan pembelajaran kerana sampel yang tidak mencukupi dalam pangkalan data ekspresi mikro (Peng et al. 2018; Sun et al. 2020; Takalkar & Xu 2017) tetapi kesan-kesan pemindahan pembelajaran dalam MER tidak diterokai dan dijelaskan, seperti kesan pemilihan pangkalan data tertentu kepada prestasi MER. Lebih-lebih lagi, ketepatan kaedah ini masih lebih rendah berbanding dengan pendekatan lain.

Selain itu, beberapa penyelesaian seperti di dalam penyelidikan (Reddy et al.2019; Verm et al. 2020) adalah bukan penyelesaian menyeluruh yang baik kerana ianya dilatih dan disahkan menggunakan sebahagian daripada pangkalan data ekspresi mikro sahaja yang mempunyai demografi dan cara pengumpulan data yang sama. Kaedah penilaian yang lebih baik yang dicadangkan dalam MEGC 2019 (See et al. 2019) melibatkan pengesahan silang *Leave-One-Subject-Out* (LOSO) pada set data ekspresi mikro komposit menggunakan metrik F1-Score tanpa pemberat (UF1) yang

direka untuk mengurangkan ketidakseimbangan kelas dan masalah generalisasi model. Kaedah ini dipanggil Penilaian Pangkalan Data Komposit (CDE) dan ia masih merupakan masalah penyelidikan yang berterusan akibat daripada set data gabungan itu yang lebih mencabar. Set data CDE adalah satu senarai nama-nama sampel video dari set data SMIC, SAMM dan CASME II serta label kelasnya yang sepadan. Pengelasan untuk kaedah CDE adalah untuk 3 kelas. Keputusan terbaik setakat ini mencapai skor UF1 0.864 dan skor UAR 0.85 (Xia, B. et al. 2020).

Demi memahami cara untuk menambah-baik MER pada CDE, (Xia, Z. et al. 2020b) telah menemui kepentingan model dan input yang lebih ringkas apabila menggunakan pembelajaran mendalam pada CDE manakala model yang kompleks lebih sesuai untuk penilaian pangkalan data tunggal. Kerumitan model merujuk kepada bilangan parameter dalam model yang terdiri daripada semua berat dan bias yang akan dipelajari oleh model manakala kerumitan input adalah resolusi input (Liong et al. 2019). Walau bagaimanapun, penyelidikan untuk mengurangkan kerumitan kedua-dua input dan model masih lagi kurang setakat ini.

Oleh itu, walaupun terdapat minat dan usaha yang besar dalam MER sejak beberapa tahun kebelakangan ini, bidang ini masih mempunyai cabaran berterusan dan perlu ditangani seperti:

1. Kaedah MER semasa menghasilkan ketepatan yang agak rendah pada Pangkalan Data Komposit yang mempunyai pelbagai demografi untuk digeneralisasi dengan baik dalam keadaan yang realistik
2. Kaedah MER memerlukan sumber pengiraan yang banyak (mahal) serta kompleks menjadikannya tidak praktikal dalam aplikasi kehidupan sebenar dan berkesan dalam Penilaian Set Data Gabungan (CDE)

Sehubungan dengan itu, kajian ini mencadangkan model Pengecaman Ekspresi Mikro (MER) yang lebih ringkas (bilangan parameter yang lebih kecil) dengan ketepatan yang lebih tinggi pada set data CDE dengan mengekstrak ciri secara automatik dari bingkai puncak sahaja demi mengurangkan kerumitan input melalui pemindahan pembelajaran dari pangkalan data ekspresi makro. Model ringkas yang

dicadangkan ini dibangunkan dengan memangkas model berasaskan EfficientNet-B0 yang diusulkan oleh (Tan & Quoc 2019) lalu menghasilkan model dengan hanya 867k parameter. Kajian ini ini adalah kajian pertama yang diketahui setakat ini yang menggunakan EfficientNet dalam MER. Selanjutnya, di dalam Bab 4 akan ditunjukkan bahawa kaedah yang dicadangkan mencapai hasil unggul dengan skor 0.98 UF1 dan skor 1.178 UAR berbanding kaedah-kaedah MER yang lain.

Seterusnya, di dalam Bab 2, kajian terdahulu yang telah dilakukan dalam MER akan digariskan. Kemudian, kaedah yang dicadangkan dan penetapan uji kaji akan dibentangkan dalam Bab 3. Selepas itu, dalam Bab 4, keputusan dan penemuan dari semua uji kaji akan dibincangkan dan dibahaskan. Akhirnya, kesimpulan di buat serta jurang dan kerja masa depan dikemukakan dalam Bab 5.

1.2 PERMASALAHAN KAJIAN

Kaedah Pengecaman Ekspresi Mikro (MER) terkini yang menggunakan pembelajaran mendalam mengesyorkan model yang kompleks, data latihan yang banyak serta memerlukan sumber pengiraan yang banyak menjadikannya kurang sesuai untuk set data ekspresi mikro. Untuk mengurangkan kerumitan model pembelajaran yang mendalam, mengekstrak ciri-ciri terlebih dahulu biasanya dicadangkan untuk mendapatkan ciri-ciri peringkat rendah untuk disuap kepada rangkaian pembelajaran yang mendalam. Walau bagaimanapun, ini memerlukan pengiraan pra-pemprosesan tambahan yang meningkatkan kerumitan input. Sebaliknya, kaedah pembelajaran mendalam yang kurang kompleks menggunakan hanya bingkai puncak tidak banyak dilakukan dalam kajian lepas serta cuma berjaya mencapai ketepatan sederhana untuk boleh dipercayai.

Tambahan pula, walaupun pemindahan pembelajaran telah digunakan secara meluas untuk mengatasi masalah set data yang kecil, kesan pemindahan pembelajaran dan pangkalan data sumber yang digunakan dalam MER belum cukup diterokai dan difahami dengan baik. Proses pemilihan pangkalan data sumber untuk dipelajari dalam pemindahan pembelajaran juga tidak dijelaskan.

Selain daripada masalah set data yang kecil, MER juga mencabar kerana set data yang tidak seimbang yang menyebabkan model berat sebelah dan sukar untuk digeneralisasi. Oleh kerana mengumpul lebih banyak sampel ekspresi mikro bukanlah tugas yang mudah, set data yang tidak seimbang mesti dikendalikan dengan cara yang berbeza. Walau bagaimanapun, teknik seperti penjanaan sampel sintetik dan pemprosesan semula data memerlukan kejuruteraan data dan pra-pemprosesan tambahan sebelum klasifikasi MER.

1.3 SKOP KAJIAN

Skop penyelidikan ini terhad kepada tugas pengesanan ekspresi mikro iaitu tugas klasifikasi label emosi dan tidak termasuk tugas lain yang berkaitan iaitu mengesan ekspresi mikro (*Micro-Expression Spotting*) yang bertujuan untuk mencari kedudukan bingkai puncak dengan tepat dalam video.

Protokol set data dan penilaian dihadkan kepada apa yang dicadangkan dalam Micro Expression Grand Challenge (MEGC) 2019 (See et al. 2019) iaitu menggunakan pengesahan silang LOSO dan metrik skor UFI dan UAR untuk pengesahan. Jenis penilaian lain seperti penilaian pangkalan data tunggal dan penilaian pangkalan data penahanan tidak akan dijalankan.

Set data yang digunakan adalah dari set data komposit MEGC 2019, yang merupakan senarai nama-nama sampel video dari set-set-data SAMM, SMIC dan CASME serta label-label kelasnya. Dari senarai ini, sampel-sampel dikumpul dari ketiga-tiga pangkalan data asal untuk pra-pemprosesan.

Kaedah yang dicadangkan di dalam kajian ini juga hanya akan dinilai dan dibandingkan dengan kajian-kajian terkini yang juga dinilai menggunakan Penilaian Set Data Gabungan (CDE) dalam (See et al. 2019). Untuk eksperimen sumber pangkalan data yang digunakan di dalam pemindahan pembelajaran, hanya pangkalan data ekspresi wajah, FER 2013 dan pangkalan data umum, ImageNet dipertimbangkan untuk perbandingan. Pangkalan-pangkalan data imej lain set data adalah di luar skop penyelidikan ini.

1.4 OBJEKTIF KAJIAN

Berikut ialah objektif-objektif kajian ini:

1. Untuk menangani masalah mengenali ekspresi mikro dengan tepat dengan hanya bingkai puncak sebagai input menggunakan model pembelajaran mendalam yang ringkas pada set data ekspresi mikro gabungan (set data CDE).
2. Untuk menyiasat pengaruh pangkalan data sumber yang digunakan dalam pemindahan pembelajaran dan menganalisis faktor-faktor yang mencetuskan perbezaan yang menjadikan pangkalan data sumber tertentu lebih berkesan di MER.
3. Untuk mengetahui sama ada masalah set data yang tidak seimbang dapat ditangani dari sudut pandangan penalaan algoritma dan bukannya mengumpul lebih banyak data (sebenar atau sintetik).

1.5 PERSOALAN KAJIAN

Q1. Adakah mungkin untuk mengekalkan atau meningkatkan ketepatan pengecaman ekspresi mikro sambil mengurangkan parameter rangkaian?

Q2. Sejauh manakah pengaruh pangkalan data sumber semasa menjalankan pemindahan pembelajaran yang disasarkan kepada set data ekspresi mikro?

Q3. Bolehkah prestasi MER yang tinggi dapat dicapai dengan menggunakan hanya bingkai puncak sebagai input tanpa langkah pengekstrakan ciri terlebih dahulu?

Q4. Adakah mungkin untuk mencapai keputusan prestasi yang lebih baik pada set data Penilaian Pangkalan Data Komposit (CDE)?

Q5. Bolehkah masalah set data yang tidak seimbang ditangani tanpa menjana lebih banyak sampel atau kejuruteraan semula set data?

1.6 SUMBANGAN

Sumbangan kajian ini adalah seperti berikut:

1. Penyelidikan ini mencadangkan rangka kerja MER baru yang ringkas berdasarkan blok EfficientNet dan menggunakan hanya bingkai puncak sebagai input. Ini adalah kajian pertama yang diketahui setakat ini yang menggunakan blok EfficientNet untuk tugas MER.
2. Penyelidikan ini menggariskan kriteria pangkalan data sumber yang paling sesuai untuk digunakan dalam pemindahan pembelajaran bagi tugas MER dan kepentingan pengaruh pilihan pangkalan data sumber terhadap prestasi pengecaman ekspresi mikro (MER)
3. Penyelidikan ini mencadangkan pendekatan masalah set data yang tidak seimbang sebagai masalah penalaan algoritma dan bukan hanya masalah kejuruteraan data yang boleh memudahkan langkah-langkah pra-pemrosesan dan menghapuskan keperluan untuk sampel tambahan.

Pusat Sumber
FTSM

BAB II

KAJIAN LITERATUR

2.1 PENGENALAN

Pengecaman ekspresi mikro (MER) adalah tugas mengklasifikasikan label emosi kepada sampel ekspresi mikro yang dikesan. Kebanyakan kajian terkini MER tertumpu kepada pengecaman ekspresi mikro spontan berbanding yang disengajakan. Terdapat 3 pangkalan data ekspresi mikro spontan, SMIC, CASME II dan SAMM, yang digunakan secara meluas untuk MER. Dari segi kaedah MER, terdapat 2 pendekatan umum untuk MER: Pendekatan tradisional - yang memerlukan ciri-ciri *handcrafted* yang di pra-ekstrak dan kedua, pendekatan Pembelajaran Mendalam biasanya menampilkan Rangkaian Neural Konvolusional (CNN). Pendekatan pembelajaran yang mendalam telah melepasi prestasi pendekatan tradisional dan kajian-kajian di bidang ini semakin meningkat. Ciri-ciri yang biasanya digunakan dalam pendekatan pembelajaran mendalam adalah jujukan imej, imej dinamik dan imej statik. Oleh kerana set data ekspresi mikro yang tidak seimbang, teknik seperti penambahan data, pemindahan pembelajaran dan menggunakan fungsi kehilangan yang berbeza sering dieksploitasi. Bahagian-bahagian seterusnya akan menerangkan dan menganalisis kajian-kajian yang berkaitan.

2.2 DEFINISI EKSPRESI MIKRO

Pada tahun 1966, Haggard dan Isaacs dalam usaha mereka untuk memahami petunjuk bukan lisan (contohnya, ekspresi wajah) pesakit psikiatri mereka, telah memeriksa rakaman video bingkai-ke-bingkai pesakit mereka dan menemui keadaan bilamana ekspresi wajah pesakit tiba-tiba berubah dengan jelas dalam satu perlatan hingga satu perlima saat (Haggard & Isaacs 1966). Mereka juga mendapati bahawa ekspresi wajah yang mereka gelar, mikro-sebentar tidak semestinya selari dengan apa yang dikatakan oleh pesakit yang menunjukkan hadirnya perasaan terpendam.

Kemudian, pada tahun 1969, Ekman dan Friesen mencadangkan bahawa ekspresi wajah yang terjelma ini terdapat kunci untuk memahami keadaan emosi tersembunyi sebenar seseorang yang mungkin tidak jelas dari hanya ucapan lisan (Ekman & Friesen 1969). Kajian mereka kebanyakannya tertumpu kepada mengesan helah tingkah laku bukan lisan yang mungkin tidak selari dengan tingkah laku lisan tetapi idea mereka telah diterima pakai dalam bidang lain seperti menyesuaikan kaedah pengajaran berdasarkan keadaan emosi pelajar dalam bidang pendidikan (Ning & Wang 2020) dan mengenal pasti orang yang berpotensi mempunyai niat jahat di dalam bidang keselamatan (Frank et al. 2009).

Dari segi definisi, ekspresi mikro adalah ekspresi wajah yang tidak jelas dan pantas dengan tempoh tidak lebih daripada 0.5 saat (Yan et al. 2013). Ia berbeza daripada ekspresi wajah biasa dari segi tempoh (ekspresi mikro mempunyai tempoh yang lebih pendek) dan penampilan di mana ekspresi mikro adalah kabur dan tidak jelas sebagaimana ekspresi wajah biasa atau nama lainnya, ekspresi makro (Yan et al. 2013). Terdapat 3 cara ekspresi mikro boleh dimunculkan: ekspresi mikro yang dilakukan di mana seseorang memaparkan emosi yang dia tidak rasakan, ekspresi mikro yang disamarkan di mana seseorang menyamar emosinya dengan memaparkan ekspresi wajah untuk emosi yang berbeza, ekspresi mikro yang dineutralkan di mana seseorang tidak menunjukkan apa-apa ekspresi wajah walaupun ia merasakan sesuatu (Porter & Brinke 2008).

2.3 PANGKALAN DATA EKSPRESI MIKRO

Demi mengkaji ekspresi mikro, beberapa pangkalan data telah dibangunkan. Pangkalan data ini terbahagi kepada dua kategori - pangkalan data tidak spontan dan pangkalan data spontan. Pangkalan data ekspresi mikro seperti Polikovsky (Polikovsky et al. 2009) dan USD-HD (Shreve et al. 2011) dicipta dengan meminta subjek untuk melakukan ekspresi wajah tanpa mengira perasaan sebenar mereka. Manakala pangkalan data ekspresi mikro spontan pula menghasilkan ekspresi wajah yang terbit tanpa sengaja atau yang cuba di benam oleh subjek tersebut. Teknik-teknik yang digunakan untuk merangsang ekspresi mikro spontan biasanya melibatkan membangkitkan perasaan tertentu dengan menunjukkan kepada peserta klip video yang

sangat beremosi dan menggalakkan penyembunyian ekspresi wajah menggunakan penalti (Pfister et al. 2011; Li et al. 2013) atau sistem insentif (Yan et al. 2013; Davison et al. 2018).

Salah satu kriteria penting ekspresi mikro adalah ia terbit secara sukarela (Hess & Kleck 1990). Oleh itu, kajian di MER memberi tumpuan kepada set data ekspresi mikro spontan. Pada masa ini, terdapat tiga set data ekspresi mikro spontan yang digunakan secara meluas iaitu SMIC (Li et al. 2013; Pfister et al. 2011), CASME (Qu et al. 2018; Yan et al. 2013; Yan et al. 2014) dan SAMM (Davison et al. 2018; Yap et al. 2019). Berbanding dengan pangkalan data ekspresi wajah biasa, set data ini secara keseluruhan jauh lebih kecil kerana cabaran mengumpul sampel ekspresi mikro spontan dan pelabelan manual yang diperlukan.

Jadual 2.3 Pangkalan data ekspresi mikro.

Set Data	Subjek	Sampel	Indeks bingkai	Kelas	Taburan kelas emosi
SMIC	16	164	Tidak dilabel	3	51 P, 70 N, 43 Su
CASME II	26	247	On, Apex, Off	5	32 H, 64 D, 25 Su, 27 R, 99 O
SAMM	32	159	On, Apex, Off	7	57 A, 12 C, 9 D, 8 F, 26 H, 6 S, 15 Su, 26 O

Kelas emosi – P: Positive, N: Negative, O: Others, H: Happiness, D: Disgust, Su: Surprised, R: Repression, T: Tense, F: Fear, C: Contempt, Sa: Sad, A: Anger. Frame Index – On: frame mula, Off: frame akhir, Apex: frame puncak

2.3.1 SMIC

Pangkalan data ekspresi mikro spontan (SMIC) (Li et al. 2013) adalah versi baharu SMIC-Sub (Pfister et al. 2011) yang merupakan salah satu perintis pangkalan data ekspresi mikro spontan yang mengandungi hanya 6 subjek. SMIC, sebaliknya mengandungi 16 subjek yang menghasilkan 164 sampel ekspresi mikro. Subjek dalam SMIC terdiri daripada 2 etnik dengan separuh daripada subjek daripada etnik Caucassian dan separuh lagi, etnik Asia. 6 daripada subjek adalah perempuan dengan selebihnya adalah lelaki dengan purata umur 28 tahun (Li et al. 2013).

SMIC meniru keadaan yang berkepentingan tinggi untuk memaksa penyembunyian emosi seperti yang disebutkan oleh (Ekman & Friesen 1969) di mana para peserta akan diberi hukuman yang keras (soal selidik yang membosankan dan panjang untuk di isi) sekiranya mereka gagal menyembunyikan emosi mereka semasa menonton klip video yang sangat merangsang emosi (Yan et al. 2014). Kelas emosi akhir ditentukan oleh semakan silang laporan peserta dengan pelabel pakar untuk memastikan keputusan yang selaras. Hanya sampel yang konsisten antara pelabel pakar dengan laporan peserta yang disimpan dan selebihnya dibuang (Yan et al. 2014).

Sampel-sampel ekspresi mikro di dalam set data ini dikelaskan kepada 3 kelas - Positif, Negatif dan *Surprise* - dan bukannya menggunakan label emosi yang biasa digunakan dalam pangkalan data lain seperti yang ditunjukkan dalam Jadual 2.3. Ini adalah berdasarkan penemuan mereka bahawa peserta melaporkan perasaan bercampur-baur semasa menonton klip video yang bertujuan untuk membangkitkan satu emosi tertentu. Untuk memastikan data boleh dipercayai, sampel-sampel untuk emosi-emosi berikut: *Sad*, *Anger*, *Fear* dan *Disgust* telah dikumpulkan ke dalam 1 kelas iaitu Negatif. Sampel-sampel dengan kelas *Happiness* dan *Surprise* masing-masing dipetakan ke kelas Positif dan *Surprise* kerana pengelasan sampel-sampel ini jelas.

Berbanding dengan pangkalan data ekspresi mikro yang lain, SMIC juga mengandungi sebahagian sampel yang di rekod dengan tetapan yang berbeza - menggunakan kamera visual biasa (VIS) dan kamera *Near-InfraRed* (NIR) - sebagai tambahan kepada kamera berkelajuan tinggi (HS) (Yan et al. 2014) lalu menghasilkan tambahan 142 sampel ekspresi mikro.

Satu kekurangan set data SMIC adalah ia tidak menyediakan label bingkai puncak, yang merupakan bingkai apabila ekspresi wajah memuncak dan paling berbeza dalam urutan bingkai. Turut tiada adalah kod FACS Unit Tindakan (AU) muka seperti CASME II dan SAMM. Contoh imej dari pangkalan data SMIC ditunjukkan dalam Rajah 2.3.1.



Rajah. 2.3.1: Contoh bingkai dari pangkalan data SMIC

2.3.2 CASME II

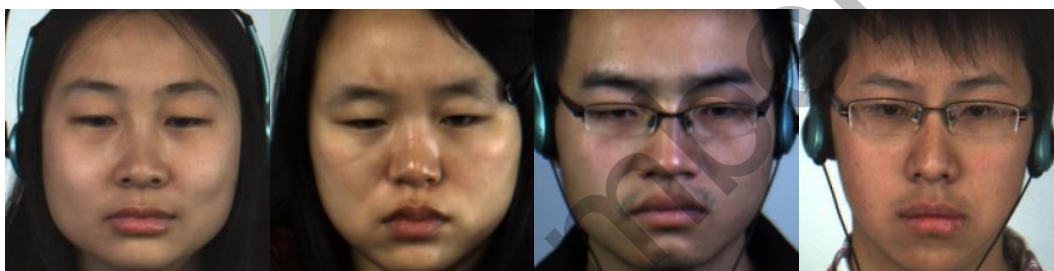
CASME II (Yan et al. 2014) adalah versi baharu set data CASME (Yan et al. 2013) yang menambahbaik saiz dan kualiti imej. Ia mengandungi 247 sampel, lebih 52 daripada pendahulunya. Set data ini juga mempunyai kadar pensampelan yang lebih tinggi sebanyak 200fps dan saiz imej 280 piksel x 340 piksel, membolehkan lebih banyak butiran diekstrak yang penting untuk ekspresi mikro yang bersifat halus.

Tidak seperti pangkalan data lain, CASME II mempertimbangkan 2 aspek yang berbeza untuk menjana ekspresi mikro berdasarkan pemerhatian Ekman: ekspresi mikro akibat menyembunyikan emosi sebenar kepada orang lain dan ekspresi mikro akibat menipu diri sendiri (Ekman & Friesen 1969) untuk meningkatkan kepelbagaian data serta perwakilan kehidupan sebenar yang lebih baik. Ia mencapai matlamat ini dengan memberikan arahan yang berbeza kepada separuh daripada peserta selepas mereka diminta untuk melihat video yang sangat beremosi. Untuk mencapai aspek pertama, peserta diarahkan untuk mengekalkan ekspresi neutral sepanjang sesi; dan untuk aspek kedua, peserta diarahkan untuk menyembunyikan ekspresi mereka jika mereka menyedarinya (Yan et al. 2014).

Terdapat 35 peserta yang terlibat dalam persediaan eksperimen mereka, di mana 26 dipilih sebagai subjek yang sah di mana 11 daripada mereka adalah perempuan dan selebihnya adalah lelaki. Semua peserta berketurunan Cina dengan purata umur 22.03 tahun (Yan et al. 2014) yang tidak mewakili demografi dunia sebenar.

Setelah sampel ekspresi mikro direkodkan, sampel ditapis secara manual untuk memenuhi kriteria tempoh pendek (kurang daripada 500ms). Kemudian, bingkai awal

(*onset*) dan akhir(*offset*) yang menandakan permulaan dan pengakhiran ekspresi mikro telah ditentukan secara manual. Semasa proses ini, Unit Tindakan Muka (AU) berdasarkan Sistem Pengekodan Tindakan Muka (FACs) yang diperkenalkan oleh Ekman (Ekman & Friesen 1978) juga telah ditandakan. Kelas label untuk setiap sampel akhirnya ditentukan berdasarkan gabungan kod AU, kenyataan peserta dan kandungan video (Yan et al. 2014). Jadual 2.3 di atas memperincikan kelas dalam CASME II. Contoh imej dari pangkalan data CASME II dibentangkan dalam Rajah 2.3.2.



Rajah 2.3.2: Contoh frame dari set data CASME II

2.3.3 SAMM

Set data SAMM telah dibangunkan dengan matlamat untuk mewujudkan set data ekspresi mikro berkod FACS. Idea ini bertujuan untuk mempunyai cara yang objektif untuk menggambarkan ekspresi mikro dari segi pergerakan mikro muka dan bukannya oleh kelas emosi. Kod FACS ditentukan untuk setiap sampel tanpa pengetahuan atau berkaitan dengan laporan peserta selepas eksperimen mahupun kandungan video rangsangan (Davison et al. 2018). Pendekatan ini menyingkirkan subjektiviti pelabelan emosi dan mencadangkan standard penentuan ekspresi mikro.

Walaupun FACS dikodkan tanpa hubungan dengan kelas emosi, (Davison et al. 2018) telah menunjukkan bahawa terdapat korelasi yang signifikan antara Unit Tindakan FACS (AU) dengan rangsangan emosi dari video. Walau bagaimanapun, bertentangan dengan ekspresi wajah biasa, mungkin hanya ada satu AU yang dipamerkan dan bukannya beberapa kerana arahan kepada peserta untuk menyekat emosi mereka dan sifat ekspresi mikro yang kurang jelas. Walaupun hubungan antara FACS dan emosi mungkin boleh mendorong MER yang lebih cekap (Davison et al. 2018), pelabelan kod FACS masih merupakan kerja manual yang intensif.

Sehingga kini, set data SAMM adalah set data ekspresi mikro yang paling pelbagai dari segi demografi dengan nisbah jantina seimbang (16 lelaki, 16 perempuan), 12 etnik dan min umur yang lebih besar iaitu 33.24 tahun. Sebelum eksperimen dijalankan peserta diminta untuk menjawab soal selidik yang akan digunakan untuk menyesuaikan rangsangan video yang lebih sesuai dengan keadaan emosi peserta untuk meningkatkan peluang mendapatkan ekspresi mikro.

Pangkalan data yang terhasil mengandungi 159 sampel dari 32 subjek yang dikelaskan kepada 7 kelas emosi. Kelas '*Anger*' mempunyai sampel paling banyak (57 sampel) manakala kelas '*Sadness*' mengandungi sampel paling sedikit (6 sampel) seperti yang dapat dilihat dalam Jadual 2.3. Ini menunjukkan taburan kelas yang tidak seimbang dalam set data. Contoh imej dari set data SAMM digambarkan dalam Rajah 2.3.3.



Rajah 2.3.3: Contoh bingkai dari set data SAMM

2.3.4 Perbincangan

Bilangan sampel di dalam 3 set data di atas yang telah digabungkan termasuk dari kelas '*Others*' hanyalah sebanyak 570 sampel sahaja. Bilangan ini jauh lebih kecil berbanding dengan pangkalan data ekspresi makro contohnya FERA 2013 (lebih daripada 35k imej) (Goodfellow et al. 2013) atau Extended Cohn-Kanade Dataset (CK+) (Lucey et al. 2010) dengan sejumlah 5876 imej. Ini menjadikannya mencabar untuk menggunakan teknik pembelajaran mendalam yang memerlukan set data yang besar untuk dilatih, tanpanya akan terhasil model yang tidak dapat digeneralisasi ('*overfit*') (Peng et al. 2017; Xia, Z. et al. 2020b). Teknik semasa yang digunakan untuk memerangi masalah ini termasuk strategi augmentasi data dan pemindahan pembelajaran serta menggunakan rangkaian yang lebih kecil.

Ketiga-tiga set data ekspresi mikro ini mempunyai taburan demografi dan cara pengumpulan data yang berbeza. Malahan, resolusi video dan pencahayaan yang digunakan juga berbeza-beza. Paling penting sekali, pengelasan label antara ketiga-tiga set data ini berbeza yang mana SMIC menggunakan 3 label, SAMM 7 label dan CASME II 5 label. Dalam MEGC 2019, set data komposit (CDE) dicadangkan yang dapat menghasilkan set data realistik dengan pelbagai latar belakang subjek (See et al. 2019).

Set data ekspresi mikro yang ada hari ini juga mengalami ketidakseimbangan kelas dengan nisbah ketidakseimbangan 3:1.3:1 untuk 3 kelas dalam set data CDE dengan kelas 'Surprise' yang hanya mempunyai 83 sampel (See et al. 2019). Set data yang tidak seimbang akan menghasilkan sistem MER yang kurang mantap yang tidak umum kerana kelas majoriti akan menguasai pembelajaran model (Lin et al. 2017). Selain daripada strategi augmentasi data, pelarasan fungsi kehilangan turut dicadangkan untuk menangani isu ini (Xia, B. et al. 2020).

a. Augmentasi Data

Matlamat augmentasi data adalah untuk menambah set data sedia ada dengan menghasilkan sampel baru secara buatan dari sampel asal. Strategi pembesaran data yang mudah biasanya melibatkan kaedah manipulasi imej seperti membalikkan imej secara mendatar dan putaran imej seperti yang digunakan oleh (Zhi & Xu 2019) atau memotong kepingan kecil dari tempat yang berbeza dalam imej seperti yang digunakan dalam (Peng et al. 2017). Selain pembalikan imej, (Takalkar & Xu 2017) meningkatkan bilangan sampel dengan menggunakan bingkai-bingkai lain dalam video yang diekstrak setiap 0.2 saat. (Takalkar et al. 2020) telah menunjukkan bagaimana pembesaran data menggunakan transformasi imej mudah boleh meningkatkan prestasi MER sekurang-kurangnya sebanyak 15%.

Kaedah augmentasi data yang lebih canggih terdiri daripada penjanaan data sintetik seperti yang dilihat dalam kajian (Xu et al. 2021) dengan menyeraskan AU muka dalam sampel asal kepada yang sintetik menggunakan Rangkaian Adversarial Generatif (GAN) dan dalam (Pan & Yang 2010) yang mencadangkan penggantian

beberapa ciri muka (mata, hidung, mulut dan lain-lain) sesuatu sampel untuk menghasilkan sampel yang baru. Keberkesanan GAN dalam MER masih belum dapat dilihat tetapi mempunyai potensi yang cukup besar. Sementara itu kaedah (Pan & Yang 2010) mungkin boleh diterokai untuk sampel ekspresi mikro.

b. Pemindahan Pembelajaran

Pemindahan pembelajaran adalah bertujuan untuk memanfaatkan pengetahuan yang dipelajari dari bidang sumber untuk menyelesaikan tugas sasaran yang berkaitan (Kadokia et al. 2021). Ia biasanya digunakan untuk menyelesaikan masalah data yang tidak mencukupi. (Peng et al. 2018; Wang et al. 2018) menggunakan pemindahan pembelajaran dengan melatih model mereka pada set data ekspresi wajah yang lebih besar seperti *Karolinska Directed Emotional Faces* (KDEF) dan CK+ dahulu dan kemudian memperhalusi model pada set data ekspresi mikro. Sebaliknya, (Xia, B. et al. 2020) mencadangkan 2 rangkaian (MacroNet dan MicroNet) yang dilatih pada set data ekspresi makro dan mikro secara berasingan untuk memisahkan ciri-ciri ekspresi daripada identiti subjek sebelum memperhalusi MicroNet yang dipandu oleh MacroNet. MTM-NET yang dicadangkan dalam (Xia, B. et al. 2020) telah melepasi semua kaedah lain dalam penilaian CDE dengan skor UF1 0.864.

Selain daripada memanfaatkan pangkalan data ekspresi makro, (Quang et al. 2019) telah mencadangkan model CapsuleNet yang terdiri daripada blok ResNet-18 yang telah dilatih pada pangkalan data ImageNet sebelum diperhalusi pada set data ekspresi mikro. ImageNet adalah pangkalan data yang besar dengan imej dari pelbagai bidang dan sering digunakan untuk latihan model dalam pengesanan objek dan aksi (Deng et al. 2009). (Quang et al. 2019) telah mencapai prestasi sederhana dengan skor UF1 0.65 dalam Penilaian Pangkalan Data Komposit MEGC 2019 (See et al. 2019).

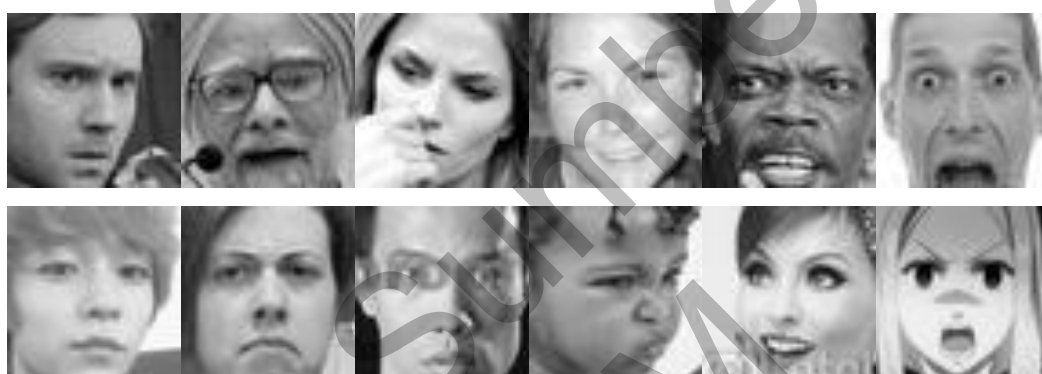
ImageNet (Deng et al. 2009) adalah pangkalan data umum yang mengandungi secara purata 1000 imej untuk setiap perkara atau konsep seperti yang ditakrifkan dalam WordNet. ImageNet digunakan secara meluas dalam pendekatan pemindahan pembelajaran kerana sokongan yang disediakan oleh modul pembelajaran mesin seperti Keras dan PyTorch yang menyediakan model yang biasa digunakan seperti ResNet dan

VGGNet dengan berat pra-dilatih dari klasifikasi ImageNet. ImageNet juga popular kerana mempunyai set data yang besar yang terdiri daripada pelbagai subjek yang pelbagai.

Walau bagaimanapun, (Patel et al. 2016) mendakwa bahawa ciri-ciri yang diekstrak dari ImageNet tidak sesuai untuk digunakan secara langsung dalam pemindahan pembelajaran untuk MER kerana set data ImageNet yang mengandungi imej umum dan bukannya imej muka terlalu berbeza daripada set data ekspresi mikro. Di samping itu, (Xia, B. et al. 2020) telah mendapati bahawa pangkalan data sumber yang dipilih untuk pemindahan pembelajaran sangat memberi kesan kepada MER. Sebagai contoh, mereka menunjukkan prestasi yang lebih baik pada set data CASME II apabila model mereka dilatih terlebih dahulu dengan pangkalan data Oulu-CASIA yang mempunyai profil demografi yang sama, berbanding penilaian yang dilakukan pada SMIC dan SAMM. Ini adalah selaras dengan dakwaan yang di buat oleh (Patel et al. 2016) bahawa persamaan pangkalan data sumber dan sasaran meningkatkan prestasi MER.

Dalam penilaian silang pangkalan data ekspresi wajah yang ekstensif yang dijalankan oleh (Li & Deng 2020) dengan melatih rangkaian neural pada 1 pangkalan data dan mengujinya pada 6 pangkalan data yang lain untuk pengelasan emosi 7 kelas, pangkalan data FER2013 terbukti mempunyai jurang prestasi yang paling sedikit berbanding dengan keputusan ujian dalam set data yang sama. Ini menandakan pemindahan dan generalisasi yang baik dari set data ini ke pengelasan emosi pangkalan data lain. Walau bagaimanapun, penulis gagal untuk menyerlahkan fakta ini tetapi sebaliknya memberi tumpuan lebih kepada keupayaan set data CK+ untuk mencapai skor ketepatan ujian yang tinggi menggunakan model yang telah dilatih pada set data lain. Set data AffectNet dan RAF-DB 2.0 disebut-sebut lebih pelbagai dan dapat digeneralisasi dengan baik walaupun peratusan jurang prestasi rendah yang konsisten untuk set data FER2013 kecuali dalam satu tugas – pengelasan ‘Happiness’ di mana jurang prestasi AffectNet adalah 1% dan FER2013 adalah perbezaan kecil sebanyak 2%.

FER 2013 (Goodfellow et al. 2013) adalah set data yang tersedia secara terbuka yang digunakan dalam Cabaran Pengecaman Ekspresi Wajah 2013. Ia terdiri daripada 35,887, imej ekspresi wajah yang dipotong untuk menunjukkan hanya kawasan muka yang dilabelkan ke dalam 7 kelas: *Angry*, *Disgust*, *Fear*, *Happy*, *Sad*, *Surprise* dan *Neutral*. Imej wajah diperoleh daripada carian istilah emosi di Google Image. Set data mengandungi lebih daripada 30k imej, masih jauh lebih kecil daripada saiz ImageNet yang mempunyai lebih daripada 1 juta imej. Imej-imej adalah skala kelabu dengan saiz 48x48. Contoh imej dari pangkalan data ini ditunjukkan dalam Rajah 2.3.4.



Rajah 2.3.4 (b). Contoh imej dari pangkalan data FER2013

c. Model yang lebih cetek dan ringan

Semakin banyak kaedah dalam MER menggunakan pembelajaran mendalam kebelakangan ini namun teknik ini memerlukan data latihan yang banyak untuk mengelakkan masalah *overfitting* di mana model pembelajaran mendalam itu telah tersesuai dengan data latihan dan tidak dapat meramal data-data baru yang belum pernah didedahkan kepada model itu. Set-set data ekspresi mikro yang digabungkan masih lagi terlalu kecil dan pastinya akan menimbulkan masalah ini. Salah satu cara untuk mengatasinya adalah menggunakan model yang lebih cetek dan ringan.

(Liong et al. 2019) mencadangkan model ringan dengan hanya 2 lapisan yang mengambil input kiub OF. Mereka mencapai hasil memberangsangkan pada masa penerbitan dengan skor UFI 0.7353 walaupun kaedah mereka hanya memerlukan 1670 parameter. Namun, kaedah mereka bukanlah penyelesaian menyeluruh kerana input

kiub OF yang terdiri daripada galur optik (OS), OF menegak dan OF melintang perlu di ekstrak terlebih dahulu dan memerlukan input data resolusi tinggi.

(Xia, Z. et al. 2020b) pula dalam satu siri eksperimen untuk memahami impak saiz model pembelajaran mendalam dan saiz input menemui bahawa rangkaian cetek dan lebih ringan berfungsi dengan lebih baik pada set data gabungan (CDE) manakala rangkaian lebih dalam sesuai untuk penilaian pada set data tunggal. Mereka mendakwa ini adalah kerana set data komposit (CDE) dengan pelbagai etnik, julat umur dan cara peendapatan data memiliki lebih gangguan antara set data (*domain shift*) yang akan dipelajari oleh model yang lebih dalam kerana kapasitinya yang tinggi lebih sensitif kepada butiran-butiran kecil dalam set data. Input yang ringkas dengan resolusi rendah serta model yang lebih cetek dapat mengatasi masalah ini dengan menyuling hanya cukup maklumat untuk tugas pengelasan tanpa terganggu dengan setiap perincian.

Selain itu, dalam analisis yang dijalankan oleh (Kadokia et al. 2021) pada set data SAMM untuk mengkaji keberkesanan 5 model pembelajaran mendalam yang telah dilatih dengan set data ImageNet, MobileNet telah muncul sebagai berprestasi tertinggi dalam tugas MER menggunakan pemindahan pembelajaran walaupun ianya model yang terkecil. Tiada analisis mengapa MobileNet v1 mencapai prestasi terbaik dalam (Kadokia et al. 2021) tetapi ia boleh diperhatikan bahawa MobileNet v1 jauh lebih kecil daripada model lain yang dibandingkan - kira-kira 5 kali lebih kecil daripada rangkaian terkecil seterusnya, Xception. Ini menunjukkan potensi yang menggalakkan dalam menggunakan rangkaian pembelajaran mendalam yang lebih ringkas untuk tugas pengecaman ekspresi mikro.

(Kadokia et al. 2021) juga mendapati bahawa pengoptimum Adam adalah pilihan terbaik untuk model berasaskan blok MBConv dengan kadar pembelajaran 0.0001 kerana kadar pembelajaran terlalu tinggi seperti 0.001 menghasilkan prestasi lebih rendah. Antara manfaat pengoptimum Adam termasuk kadar pembelajaran individu bagi setiap parameter dalam rangkaian yang dihadkan oleh kadar pembelajaran maksimum yang boleh ditetapkan (Kingma & Ba 2014). Ini mengelakkan perlunya untuk mengoptimumkan parameternya untuk menghasilkan prestasi terbaik kerana Adam menyesuaikan kadar pembelajaran dengan sewajarnya (Kingma & Ba 2014).

d. Set Data Komposit

Di dalam *Micro-Expression Grand Challenge* (MEGC) 2019 yang merupakan pertandingan pengecaman ekspresi mikro yang kedua selepas pertama kali diperkenalkan pada tahun 2018, set data komposit yang menggabungkan sampel-sampel dari set-set data SMIC, SAMP dan CASMEII telah dicadangkan (See et al. 2019). Tujuan penggabungan ini adalah untuk mendapatkan set data yang mencerminkan realiti di mana subjek-subjek adalah dari pelbagai latar belakang dan data-data dikumpulkan dengan pelbagai cara. Dengan menggabungkan 3 set data ini juga dapat menambah jumlah keseluruhan bilangan sampel (See et al. 2019).

Namun, ketiga-tiga set data ini mempunyai sistem pengelasan yang berbeza. Jadi, dalam MEGC 2019 pengelasan semula sampel-sampel ini dilakukan. Kelas-kelas emosi asal di petakan ke 3 kelas iaitu '*Negative*', '*Positive*', dan '*Surprise*' seperti di dalam Jadual 2.3.4.d. Sampel yang mempunyai kelas asal '*Others*' pula disisihkan (See et al. 2019). Manfaat pelabelan semula adalah ia menyediakan piawai untuk label kelas-kelas yang berbeza di dalam 3 set data ekspresi mikro. Walau bagaimanapun, kelemahannya ialah set data yang tidak seimbang kerana kebanyakan kelas asal dilabelkan semula sebagai kelas '*Negative*'. Sistem MER juga kehilangan keupayaan untuk membezakan emosi yang berbeza yang dikumpulkan ke dalam kelas '*Negative*'.

Jadual 2.3.4(d). Pelabelan semula dan kelas asal untuk set data CDE

Kelas baharu	Kelas asal	Jumlah
Negative	CASME II: Disgust and Repression, SMIC: Negative, SAMP: Anger, Contempt, Disgust, Fear and Sadness	250
Positive	CASME II: Happiness, SMIC: Positive, SAMP: Happiness	109
Surprise	CASME II: Surprise, SMIC: Surprise, SAMP: Surprise	83

Jumlah keseluruhan sampel yang digunakan adalah sebanyak 442 sampel (Lampiran B). Hasil dari penggabungan ini adalah senarai nama-nama sampel video dari ketiga-tiga set data asal dan label kelas baharunya dan set data ini biasanya dipanggil set data CDE. Sampel-sampel video ini mempunyai sejumlah bingkai dari permulaan terbitnya ekspresi mikro hingga ia berakhir. Namun, oleh kerana set data ini hanyalah satu senarai, sampel-sampel video asal perlu diperolehi sendiri dari pangkalan

data asal. Ini juga bermakna, set data CDE ini tidak menjalankan sebarang pemrosesan awal atau menetapkan samada keseluruhan sampel video harus digunakan atau hanya bingkai-bingkai tertentu dalam video sahaja. Teknik-teknik yang menggunakan hanya bingkai puncak di dalam sampel-sampel video misalannya haruslah mendapatkan kedudukan bingkai puncak itu sendiri samada melalui anotasi asal bagi CASME II dan SAMM atau melalui algoritma pengesanan untuk SMIC yang tiada anotasi.

Set data CDE ini menggunakan penilaian CDE (*Composite Database Evaluation*) iaitu memakai pengesanan silang LOSO (*Leave-One-Subject-Out*) dan metrik purata F1, UF1 (*Unweighted-F1*) dan purata Recall (*Unweighted Average Recall*), UAR untuk penilaian (See et al. 2019).

e. Fungsi Kehilangan

Set data gabungan, CDE mempunyai taburan kelas yang tidak seimbang dengan kira-kira 56% sampel berada dalam kelas 'Negatif'. Model yang dilatih pada set data yang tidak seimbang cenderung untuk mempelajari ciri-ciri kelas dominan dan mengakibatkan generalisasi yang lemah untuk kelas minoriti. Untuk menangani isu ini, beberapa kajian telah dipetik menggunakan fungsi kehilangan yang disesuaikan untuk set data yang tidak seimbang dan bukannya fungsi kehilangan entropi silang (Nie et al. 2021; Lai et al. 2020; Xia, B. et al. 2020).

Antara contoh-contoh fungsi kehilangan yang lebih sesuai untuk taburan kelas yang tidak seimbang adalah fungsi kehilangan kelas seimbang (CB) yang mengira kehilangan berdasarkan faktor pemberat (Cui et al. 2019) dan fungsi kehilangan fokus (FL) yang mengurangkan kehilangan dari sampel dominan untuk memberi tumpuan lebih kepada kelas yang sukar digolongkan (Lin et al. 2017). Gabungan fungsi kehilangan juga disebut dalam kesusasteraan, contohnya fungsi kehilangan Triplet digunakan dalam (Xia, B. et al. 2020) yang menggabungkan kehilangan dari input triplet manakala (Nie et al. 2021) menggunakan fungsi kehilangan fokus pelbagai kelas yang seimbang. (Nie et al. 2021) telah menunjukkan prestasi unggul menggunakan fungsi kehilangan fokus pelbagai kelas yang seimbang pada penilaian CDE (See et al. 2019) dan oleh itu, pendekatan yang sama digunakan dalam kajian ini.

Objektif fungsi kehilangan fokus adalah untuk mengelakkan pengemaskinian berat yang besar dari kehilangan sampel-sampel mudah. Di dalam kajian (Lin et al. 2017), fungsi kehilangan fokus klasifikasi binari diperkenalkan dengan menggunakan faktor fokus γ , yang akan mengurangkan sumbangan daripada sampel yang mudah dikelaskan. Ini boleh diperluaskan kepada kehilangan fokus pelbagai kelas dengan memasukkan parameter pemberat kelas, α yang menandakan kepentingan setiap kelas dalam pengiraan kerugian (Liu et al. 2018). Nilai α yang lebih tinggi meningkatkan sumbangan kelas. Formula untuk kehilangan fokus berbilang kelas ditakrifkan dalam Eq. (3.1.4).

$$FL(\check{p}) = - \sum_{i=1}^C \alpha_i (1 - \check{p})^\gamma \log(\check{p}) \quad (3.1.4)$$

di mana C adalah bilangan kelas dan \check{p} merupakan vektor kebarangkalian ramalan ke atas semua kelas. Dari formula ini, dapat dilihat bahawa kerugian yang dihasilkan akan menjadi semakin kecil jika \check{p} menjadi lebih besar, menandakan pengurangan kerugian ramalan yang yakin.

Fungsi kehilangan kelas seimbang (CB) juga menggunakan konsep pemberat yang sama seperti dalam kehilangan fokus tetapi menawarkan pendekatan matematik untuk menetapkan nilai faktor pemberat yang dikira sebagai $\alpha(1 - \beta)/(1 - \beta^{n_i})$ di mana β ialah hyperparameter dengan nilai antara 0 dan 1, dan n_i adalah bilangan sampel untuk kelas i (Cui et al. 2019).

2.4 PENDEKATAN TRADITIONAL

Teknik tradisional dalam MER menggunakan ciri-ciri *handcrafted* berdasarkan penampilan (warna, tekstur) atau geometri (tanda-tanda muka) yang diekstrak daripada imej atau urutan imej. Teknik-teknik ini memerlukan pra-pemprosesan yang ketara dan penalaan untuk mengekstrak ciri-ciri yang diperlukan sebelum klasifikasi boleh dilakukan. Pendekatan tradisional dalam MER digariskan dalam bahagian-bahagian seterusnya.

2.4.1 Pendekatan Corak Perduaan Tempatan (LBP)

Corak Perduaan Tempatan (LBP) mewakili maklumat tekstur dengan melawat setiap piksel dalam imej dan membandingkan nilainya dengan piksel-piksel di sekelilingnya. Variasi yang lebih biasa digunakan dalam MER adalah Corak Perduaan Tempatan-Tiga Satah Orthogonal (LBP-TOP) yang mempertimbangkan dimensi temporal sebagai tambahan kepada lokasi piksel dan juga Corak Perduaan Tempatan dengan Persimpangan Enam Titik (LBP-SIP) yang mengurangkan maklumat berlebihan dalam LBP-TOP seperti yang dicadangkan oleh (Wang et al. 2014). (Pfister et al. 2011) menggunakan satu lagi varian LBP – corak seragam LBP (U-LBP) - untuk mengekstrak ciri-ciri muka yang kemudiannya digunakan untuk mengira perbezaan ciri untuk menentukan bingkai puncak. Ciri-ciri diekstrak yang berasaskan LBP biasanya disuap kepada pengelas seperti SVM, bagaimanapun (Takalkar et al. 2019) mencadangkan menyatukan ciri-ciri LBP-TOP yang mendalam dengan model CNN. Walaupun kaedah berasaskan LBP telah menunjukkan prestasi cemerlang dalam MER dan pengecaman ekspresi wajah tetapi kos pengiraannya memerlukan sumber yang banyak serta sensitif terhadap perbezaan pencahayaan antara bingkai (Wang et al. 2014).

2.4.2 Pendekatan Histogram-of-Gradient (HOG)

Histogram-of-Gradients (HOG), seperti namanya adalah histogram orientasi kecerunan dan amplitud di setiap bahagian imej yang boleh digunakan untuk mewakili maklumat tekstur dan rupa bentuknya. Variasi HOG, Histogram Orientasi Kecerunan 3D (HOD3D) yang menangkap ciri-ciri HOG spatial-temporal telah digunakan dalam (Chen et al. 2016) dengan beberapa kaedah skim pemberat berdasarkan kepentingan kawasan muka untuk ekspresi mikro. (Chen et al. 2016) mencapai ketepatan yang munasabah tetapi ia masih mahal dari segi pengiraan.

2.4.3 Pendekatan Aliran Optik (OF)

Ciri Aliran Optik (OF) mewakili anggaran pergerakan antara urutan video atau bingkai yang menjadikannya sesuai untuk mengekstrak perubahan ekspresi wajah halus dalam MER. Variannya, Histogram Aliran Optik Berorientasi (HOOF) dikembangkan oleh (Happy & Routray 2016) kepada HOOF Kabur (FHOOF) dan Histogram Aliran

Optik Berorientasi Kabur (FHOFO) untuk menangani isu kebergantungan antara vektor HOOF dengan kepekaan terhadap intensiti pergerakan muka dan pencahayaan. Sebaliknya, (Liong et al. 2016b) mencadangkan satu lagi varian - Aliran Optik Berorientasi dengan Dua Pemberat (Bi-WOOF) yang menampilkan deskriptor untuk menggambarkan pergerakan muka antara permulaan dan bingkai puncak menggunakan tiga ciri: orientasi, magnitud dan galur optik. Kedua-dua (Song et al.2019) dan (Happy & Routray 2016) menggunakan pengelas tradisional seperti SVM dan KNN.

Banyak kajian juga menggunakan ciri OF bersama-sama dengan teknik pembelajaran mendalam seperti (Liong et al. 2018) yang mengekstrak perbezaan antara bingkai permulaan dan puncak untuk disuap kepada CNN, (Li et al. 2019) yang mencadangkan penggunaan CNN dalam untuk pengesanan mercu tanda muka dan pengekstrakan ciri HOOF untuk digunakan dalam pengelas SVM, (Khor et al. 2019) yang mencadangkan rangkaian dwi-aliran berdasarkan AlexNet yang dipangkas yang mengambil input kiub aliran optik dan (Peng et al. 2017) yang mencadangkan rangkaian dua aliran untuk mengendalikan kadar bingkai yang berbeza dengan urutan aliran optik sebagai input.

Sama seperti ciri-ciri *handcrafted* yang lain, pendekatan berasaskan OF memerlukan pra-pemprosesan untuk mengekstrak ciri-ciri terlebih dahulu. Tambahan lagi, ia memerlukan sumber pengiraan lebih banyak dan lebih perlahan dari LBP-TOP (Li et al. 2019). Ia juga memerlukan pengendalian khas untuk mengurangkan kepekaan terhadap pencahayaan dan kecerahan yang menjadikannya tidak sesuai untuk MER masa nyata. Tambahan lagi, langkah pra-pemprosesan memakan masa untuk mendapatkan ciri-ciri OF (Lei et al. 2021). (Liong et al. 2019) pula mendapati untuk mendapat ciri-ciri OF ekspresi mikro yang lebih tepat memerlukan sampel data ber-resolusi tinggi yang tidak mudah didapati untuk set data MER.

2.5 PENDEKATAN PEMBELAJARAN MENDALAM

Dengan munculnya perkakasan dipercepatkan GPU, teknik pembelajaran yang mendalam telah menyaksikan kebangkitan minat sejak beberapa tahun kebelakangan ini. Kaedah MER berasaskan pembelajaran mendalam berbeza daripada kaedah

tradisional kerana ia secara automatik mengekstrak ciri-ciri ekspresi mikro menggunakan rangkaian saraf yang mendalam dari imej statik atau urutan bingkai dalam video yang boleh mengelakkan keperluan untuk mengekstrak ciri terlebih dahulu. Tugas-tugas menggunakan kaedah pembelajaran mendalam juga telah melampaui prestasi kaedah tradisional, memperhebat perhatian di bidang ini.

Walau bagaimanapun, kerana rangkaian saraf yang mendalam biasanya memerlukan data latihan yang besar, pertimbangan yang teliti mesti digunakan apabila mereka bentuk rangkaian untuk mengelakkan '*overfitting*' dan generalisasi rendah (Li et al. 2021; Xia, Z. et al. 2020b). Sub-bahagian seterusnya akan menerangkan ciri input dan seni bina rangkaian kajian-kajian terdahulu di dalam MER yang berdasarkan pembelajaran mendalam. Jadual 2.5.2 meringkaskan kaedah pembelajaran mendalam yang diterangkan dalam bahagian ini.

2.5.1 Ciri-ciri Input

a. Imej/jujukan bingkai

Kaedah menggunakan urutan imej atau bingkai dalam video ekspresi mikro secara amnya mengekstrak ciri *spatio-temporal* bagi memodelkan perubahan minit dalam ekspresi mikro dari masa ke masa (Reddy et al. 2019; Xia, Z. et al. 2020a) tetapi ini meningkatkan kos pengiraan dan kerumitan model (Li et al. 2021; Xia, Z. et al. 2020a). Namun, kajian-kajian terkini cuba memanfaatkan maklumat ekspresi setempat untuk mengurangkan kerumitan model dan mengesan ciri-ciri berguna dengan lebih baik (Hashmi et al. 2021; Song et al. 2019; Yang et al. 2021) berdasarkan pengetahuan bahawa ekspresi mikro biasanya hanya melibatkan beberapa pergerakan otot di bahagian-bahagian tertentu wajah.

Dalam kajian (Song et al. 2019), gabungan bingkai puncak dan urutan bingkai mula-puncak-akhir digunakan untuk mengekstrak ciri berasaskan penampilan setempat dan global bersama-sama dengan ciri aliran optik menggunakan rangkaian tiga aliran. Sama seperti (Song et al. 2019), (Yang et al. 2021) juga menggunakan rangkaian 3-aliran untuk mengekstrak ciri-ciri *spatio-temporal* (*spatial*, *OF* dan *galur optik*) tetapi mereka mengeksploitasi modul perhatian untuk memberi tumpuan kepada kawasan muka yang

tertentu. Di samping itu, (Hashmi et al. 2021) mencadangkan rangkaian pelbagai perhatian untuk mempelajari ciri-ciri pilihan dan kaitannya dengan bahagian muka. Walau bagaimanapun, (Hashmi et al. 2021) memerlukan imej resolusi tinggi untuk menentukan kawasan muka yang penting lalu meningkatkan lagi kerumitan rangkaian.

(Liong et al. 2019) juga menggunakan ciri-ciri OF 3-D iaitu galur optik, OF menegak dan OF melintang yang di ekstrak dari bingkai puncak dan bingkai awal namun ciri 3D ini disuap ke model ringan yang hanya mempunyai kedalaman 2 lapisan bagi menunjukkan model ringan dan cetak masih boleh menjana dan mengekalkan perwakilan ciri. Model yang dihasilkan hanya mempunyai 1.67k parameter namun tidak dapat digeneralisasi dengan baik untuk semua set data ekspresi mikro. Malah, walaupun model yang dihasilkan ringan, ia memerlukan ciri-ciri tahap rendah yang sudah di pra-estrawak terlebih dahulu.

Berbeza dengan yang lain-lain, (Xia, B. et al. 2020) mencadangkan rangkaian yang boleh mendapatkan ciri-ciri ekspresi yang berasingan daripada penampilan subjek untuk digunakan sebagai input kepada model MER. Walaupun mereka mendakwa ini membolehkan mereka untuk mengeksploitasi pangkalan data ekspresi makro yang lebih besar dengan lebih cekap, hasil eksperimen mereka masih menunjukkan kesan berat sebelah identiti dari set data sumber itu apabila digunakan untuk memandu pengestrakan ciri dari set data ekspresi mikro. Sementara itu, kaedah yang mempelajari deskriptor graf muka dari imej yang pergerakan mukanya diserlahkan dan digabung dengan maklumat AU sebagai input kepada klasifikasi ekspresi mikro telah dicadangkan oleh (Lei et al. 2021). Walau bagaimanapun, kaedah ini memerlukan AU yang telah dianotasi dalam set data.

Seperti yang digariskan di atas, dengan menggunakan urutan bingkai sebagai input dapat meningkatkan ruangan ciri dengan memasukkan dimensi temporal tetapi ia datang dengan peningkatan kos kerumitan dan pengiraan. Modul perhatian dan memanfaatkan pangkalan data ekspresi wajah adalah beberapa cara untuk menangani masalah ini.

b. Imej Dinamik (DI)

Ciri-ciri ruang-temporal sering diekstrak secara berasingan dalam rangkaian berbilang aliran menyebabkan kaitan antara rupa bentuk dan maklumat pergerakan tidak dapat diperolehi (Verm et al. 2020). Oleh itu, penggunaan imej dinamik yang boleh mewakili pergerakan wajah ekspresi mikro dan maklumat rupa bentuk di dalam sebuah video ke dalam satu imej telah dicadangkan. Ini mengurangkan kerumitan dan kos pengiraan model namun ia memerlukan lebih usaha pada pra-pemprosesan data untuk mendapatkan imej dinamik.

Dalam kajian oleh (Verma et al. 2019), imej dinamik diperolehi dengan menggunakan peringkat vektor ciri untuk semua bingkai di dalam video untuk menukarnya menjadi vektor tunggal. Mereka mencadangkan LEARNNet yang menggunakan lapisan pengumpulan (*accretion layer*) sebagai tambahan kepada lapisan konvolusi berurutan biasa dengan tujuan untuk menggabungkan maklumat yang relevan dari lapisan sebelumnya yang mereka dakwa dapat meningkatkan pembelajaran ciri-ciri terperinci. Begitu juga (Verm et al. 2020) yang turut menggunakan peringkat pergerakan bingkai sebelum menggabungkannya. Walau bagaimanapun, proses penggabungan adalah menggunakan skim pemberat bingkai namun kajian itu tidak menjelaskan bagaimana nilai pemberat diperolehi dan ia boleh diandaikan bahawa ini dipilih secara manual. Seni bina yang dicadangkan memberi tumpuan kepada mengekstrak kedua-dua pergerakan mikro dan ciri-ciri global menggunakan pelbagai saiz *kernel* dan 2 blok utama untuk menyuling ciri-ciri selanjutnya.

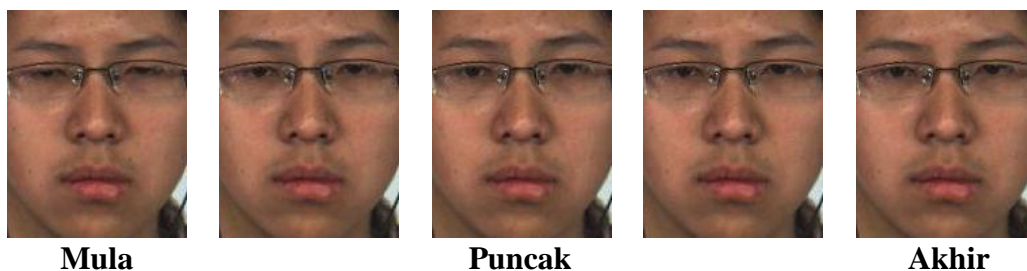
Selanjutnya, (Nie et al. 2021) juga membina imej dinamik terlebih dahulu menggunakan pendekatan yang sama dengan (Verma et al. 2019) dan (Verm et al. 2020) yang kemudiannya disampaikan ke rangkaian dwi-aliran yang mengesan jantina dan ekspresi mikro secara berasingan. Respons ciri-ciri dari aliran jantina disuap kepada aliran emosi di setiap blok konvolusi untuk menggunakan pengetahuan jantina dalam MER. Walau bagaimanapun, tanpa kajian perbandingan tentang tindak balas ciri jantina dengan model yang tidak menggunakannya, sukar untuk mengatakan bahawa prestasi yang lebih baik berbanding (Khor et al. 2019; Song et al. 2019; Quang et al. 2019) adalah disebabkan oleh ciri-ciri jantina atau sebaliknya disebabkan oleh fungsi

kehilangan kelas seimbang (CB) fokus yang digunakan di sini berbanding dengan kajian-kajian lain yang dibandingkan.

c. Bingkai puncak

Bingkai puncak di dalam video adalah bingkai bilamana ekspresi mikro yang terbit mencapai kemuncak dan ia mewakili maklumat ekspresi yang paling jelas manakala bingkai awal (*onset*) dan akhir (*offset*) adalah bingkai apabila ekspresi mikro bermula dan berakhir (Liong et al. 2016a; Sun et al. 2020). Oleh kerana sifat ekspresi mikro yang halus dan pantas, perbezaan antara bingkai mula-puncak-akhir hampir tidak dapat dilihat oleh manusia dengan kadar pengecaman tidak lebih daripada 50% untuk pengawal pantai AS yang terlatih menggunakan Alat Latihan Ekspresi Mikro (METT) yang dicadangkan oleh Ekman (Frank et al. 2009). Contoh bingkai mula, puncak dan akhir digambarkan dalam Rajah. 2.5.1.

Kaedah yang hanya menggunakan bingkai puncak sebagai input biasanya menggunakan kaedah pemindahan pembelajaran disebabkan oleh bilangan data yang tidak mencukupi (iaitu bilangan bingkai yang boleh digunakan) (Peng et al. 2018; Quang et al. 2019; Sun et al. 2020). Rangkaianya juga biasanya lebih mudah untuk mengelakkan ‘*overfitting*’ apabila menggunakan data terhad tetapi juga untuk menerangkan ciri-ciri penting secara am (Xia, Z. et al. 2020b). Kelebihan kaedah ini adalah kerumitan model yang rendah dan membolehkan pemanfaatan pangkalan data imej wajah statik yang tersedia secara meluas. Kelemahannya pula adalah model mungkin kehilangan maklumat penting berkaitan dengan pergerakan muka dan perlu mengendalikan data puncak yang tidak mencukupi.



Rajah. 2.5.1 Contoh bingkai-bingkai untuk kelas ‘Surprise’ dalam set data CASME II. Bingkai awal, puncak dan akhir ditunjukkan mengikut urutan.

Kajian (Quang et al. 2019) adalah satu-satunya kaedah berasaskan bingkai puncak dalam pesaing 4 teratas di MEGC 2019 (See et al. 2019) dan kajian pertama yang menggunakan CapsuleNet untuk MER. Model ini dilatih di ImageNet (Deng et al. 2009) dahulu sebelum dinilai pada set data CDE. Walau bagaimanapun, (Quang et al. 2019) meletakkan banyak penekanan kepada proses mendapatkan lokasi bingkai puncak yang tepat demi meningkatkan ketepatan model. Ini memerlukan pra-pemprosesan tambahan.

Sebaliknya, (Sun et al. 2020) mencadangkan pembelajaran guru-pelajar (*Teacher-Student learning*) di mana rangkaian guru yang di pra-latih pada set data ekspresi wajah FERA 2017 digunakan untuk memindahkan pengetahuan guru ke rangkaian pelajar yang cetek (1 lapisan konvolusi dan 3 lapisan yang bersambung sepenuhnya, FC) dengan memandu fungsi kehilangan dalam rangkaian pelajar. (Sun et al. 2020) mendapati bahawa dengan hanya menggunakan bingkai puncak dapat memberi hasil yang lebih baik daripada menggunakan urutan bingkai dalam rangkaian pelajar cetek, selaras dengan penemuan oleh (Xia, Z. et al. 2020b). (Sun et al. 2020) juga menunjukkan bahawa pendekatan pemindahan pembelajaran lebih baik berbanding latihan rangkaian dari awal. Walau bagaimanapun, penilaian yang dilakukan pada setiap set data ekspresi mikro adalah secara berasingan.

Kajian yang terkini, ICE-GAN (Yu et al. 2020) mendekati MER dari perspektif data dan model. Rangkaian hujung ke hujung yang termasuk janaan bingkai puncak ekspresi mikro yang sintetik dengan menggunakan Rangkaian Adversarial Generatif (GAN) yang dapat memelihara identiti asal dan juga rangkaian berasaskan kapsul yang tidak sensitif terhadap perbezaan kedudukan atau putaran - telah mereka kaitkan dengan peningkatan besar prestasi model mereka ke atas semua model lain yang menggunakan bingkai puncak setakat ini, mencatatkan skor UFI 0.874 pada set data CDE. Pendekatan mereka adalah bertentangan dengan (Xia, B. et al. 2020) di mana ciri-ciri ekspresi mikro telah diekstrak berasingan daripada penampilan wajah subjek. Walaupun kaedah di dalam (Yu et al. 2020) akan membolehkan janaan lebih banyak sampel untuk seseorang subjek, walaupun untuk emosi yang tidak dipamerkan oleh subjek itu, kaedah (Xia, B. et al. 2020) dapat membuat sampel untuk identiti baru yang akan meningkatkan generalisasi model dengan lebih baik.

Setakat ini, kajian yang menggunakan bingkai puncak secara terus untuk set data gabungan CDE cuma ada 2 iaitu CapsuleNet (Quang et al. 2019) dan ICE-GAN (Yu et al. 2020). Hasil CapsuleNet hanya sederhana sahaja dengan skor UFI 0.65 dan ICE-GAN pula memerlukan janaan data tambahan yang menambah kerumitan kaedah mereka. Kedua-dua kajian menggunakan blok kapsul sebagai blok asas rangkaian.

2.5.2 Seni bina rangkaian

Seperti yang ditunjukkan dalam Jadual 2.5.2, ada beberapa jenis blok yang biasa digunakan untuk mewujudkan rangkaian pembelajaran mendalam untuk MER masing-masing dengan kualiti dan faedah tersendiri. Blok-blok ini merupakan tambahan kepada Rangkaian Neural Konvolutional (CNN) yang biasa digunakan dalam bidang penglihatan komputer.

Jadual 2.5.2: Kaedah Pembelajaran Mendalam di dalam MER

Kaedah	Tahun	Input	Seni bina	Blok	Pra-latih
3-stream (Song et al. 2019)	2019	Jujukan imej	3S CNN	-	FER2013
LEARNet (Verma et al. 2019)	2019	DI	CNN	Hibrid	-
CapsuleNet (Quang et al. 2019)	2019	B. Puncak	CNN	Kapsul	ImageNet
Distill (Sun et al. 2020)	2020	B. Puncak	CNN	-	FERA2017
MTMNet (Xia, B. et al. 2020)	2020	Jujukan imej	2S CNN + DA + GAN	RES	CK+, MMI, Oulu-CASIA
AffectiveNet (Verm et al. 2020)	2020	DI	4S CNN	MFL	-
ICE-GAN (Yu et al. 2020)	2020	B. Puncak	CNN+GAN	Kapsul	ImageNet
LARNet (Hashmi et al. 2021)	2021	Jujukan imej	CNN	Attention	-
MERTA (Yang et al. 2021)	2021	Jujukan imej	3S CNN	Attention	-
GEME (Nie et al. 2021)	2021	DI	2S CNN + DA	RES	-
AUGCN (Lei et al. 2021)	2021	Jujukan imej	2S CNN + GCN	Graf, Transformer	-
MimaNet (

S: Aliran (*Stream*). DA: Penyesuaian domain, RES: Blok sisa, GAN: *Generative Adversarial Learning*, DI: Imej Dinamik, GCN: Rangkaian Konvolusi Graf, B. Puncak: Bingkai Puncak

a. CNN

Rangkaian Neural Konvolusional 2D, 2D-CNN adalah inovasi pelopor yang memperluaskan algoritma Rangkaian Neural Buatan (ANN) ke medan Visi Komputer. 2D-CNN biasanya terdiri daripada beberapa lapisan, direka untuk mengekstrak maklumat dari input imej secara berperingkat (Gogul & Sathiesh 2017). Ini membolehkan 2D-CNN secara automatik mempelajari perwakilan imej dari butiran peringkat rendah kepada butiran peringkat tinggi secara berurutan.

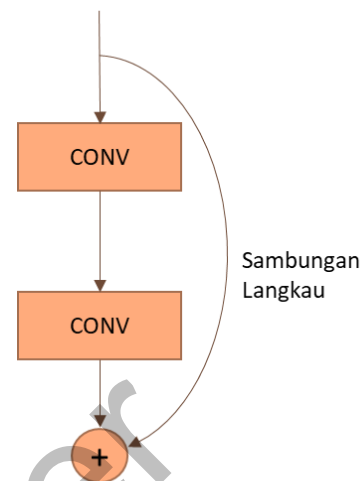
Terdapat tiga jenis lapisan utama dalam 2D-CNN iaitu lapisan Konvolusional (CONV), lapisan *Pooling* dan lapisan Bersambung Sepenuhnya (FC). Lapisan CONV bertanggungjawab untuk mengekstrak ciri spatial dengan menggunakan penapis (*filter*) merentasi matriks piksel input imej menggunakan operasi matematik yang dipanggil *convolution*. Penapis boleh diset menjadi saiz dan bilangan yang berbeza bergantung kepada jenis ciri-ciri yang akan diekstrak. Lapisan *Pooling*, sebaliknya, digunakan untuk mengurangkan dimensi input demi mengurangkan pengiraan serta untuk mendapatkan ciri-ciri am. Lapisan *Pooling* yang biasa digunakan adalah *Max Pooling* dan *Average Pooling*. Lapisan jenis terakhir, lapisan FC adalah lapisan neuron yang padat dengan sambungan penuh ke output lapisan sebelumnya (Gogul & Sathiesh 2017).

Walaupun 2D-CNN digunakan untuk input 2 dimensi, seperti imej, konsepnya telah diperluaskan ke 3D-CNN untuk membolehkan pengekstrakan ciri-ciri untuk dimensi ketiga, contohnya, dimensi temporal seperti yang dilihat dalam kaedah pembelajaran mendalam MER menggunakan ciri aliran optik (Zhi & Xu 2019; Liang et al. 2019).

b. Blok Sisa

Dalam kajian (Peng et al. 2018) dan (Sun et al. 2020), ResNet digunakan sebagai model tulang belakang untuk pelaksanaan MER. Seni bina ResNet dicipta oleh (He et al. 2015) berdasarkan blok sisa yang direka untuk menangani masalah kecerunan yang hilang yang biasa ditemui dalam rangkaian dalam. Blok sisa terdiri daripada sambungan langkau antara output sesebuah lapisan dan input ke lapisan berikutnya yang lebih dalam. Ia boleh melangkau sebanyak 1 lapisan perantara atau lebih.

Dengan menambah blok sisa ke CNN yang mendalam, (He et al. 2015) telah menunjukkan bahawa rangkaian yang dihasilkan dapat memanfaatkan sepenuhnya seni bina rangkaian yang lebih mendalam tanpa menghadapi masalah kehilangan kecerunan yang menyebabkan peningkatan kesilapan latihan setelah kedalaman rangkaian tertentu dicapai. Blok sisa membolehkan ciri-ciri sisa (ciri identiti) dieksploitasi dalam lapisan berikutnya yang sangat berguna semasa propagasi balik.



Rajah. 2.4.2(b): Blok Sisa

c. Modul Kapsul

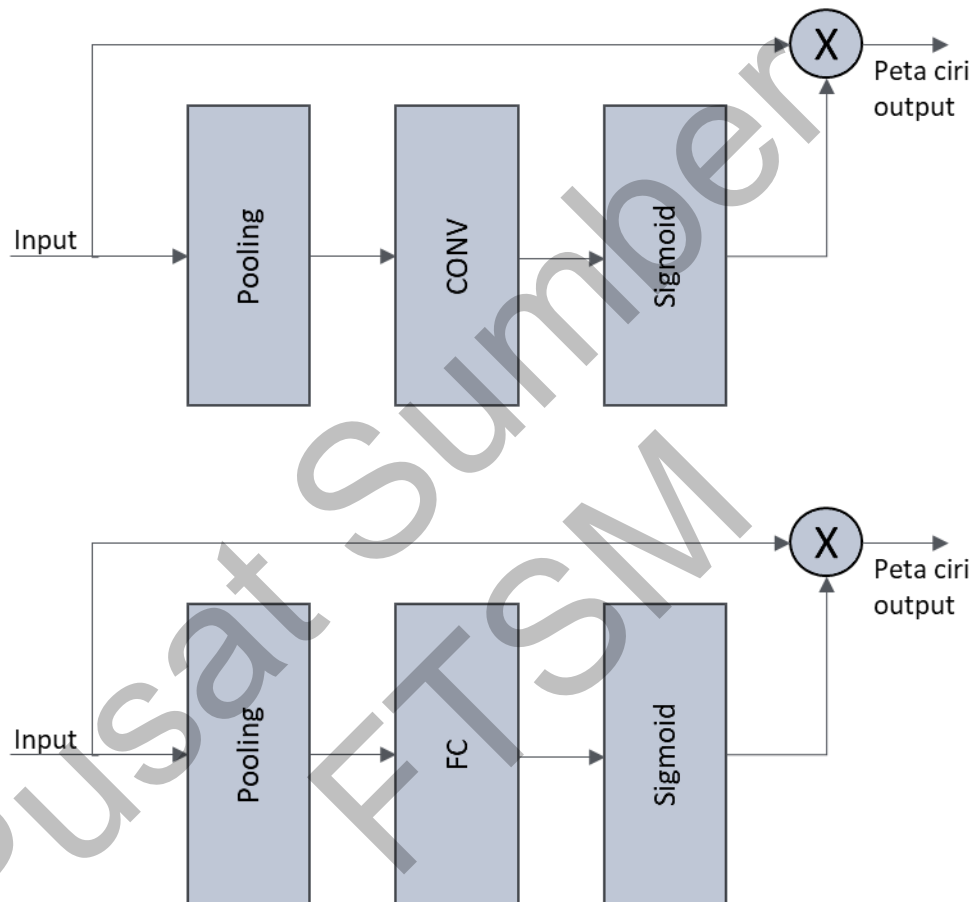
(Sabour et al. 2017) memperkenalkan CapsuleNet untuk menyelesaikan masalah invarian putaran (*rotational invariance*) dan hierarki entiti yang dikaitkan dengan CNN. Model CNN adalah invarian putaran jika ia cuma dapat mengesan objek di mana sahaja mereka berada dalam imej jika imej tidak diputar. Ini bermakna CNN memerlukan semua perwakilan objek yang mungkin semasa latihan untuk mengenali objek yang orientasinya diputar. CNN juga tidak mempunyai maklumat hubungan antara entiti yang berbeza dalam imej yang mentakrifkan objek.

CapsuleNet menangani masalah ini dengan menggunakan modul kapsul yang dapat merangkumi kaitan antara setiap ciri bersama-sama dengan sifatnya (putaran, orientasi dll.) secara hierarki (Sabour et al. 2017). Kelemahan CapsuleNet adalah kos pengiraannya yang tinggi. Dalam tugas MER, (Quang et al. 2019) dan (Yu et al. 2020) telah menggunakan modul kapsul dalam kajian mereka.

d. Modul perhatian

Dari pemahaman bahawa ekspresi wajah dapat dijelaskan oleh perubahan di bahagian muka tertentu, beberapa kaedah MER telah mengeksploitasi modul perhatian untuk memberi tumpuan kepada bahagian wajah yang disasarkan (Hashmi et al. 2021; Yang et al. 2021). Modul perhatian terdiri daripada lapisan CONV, lapisan padat dan fungsi sigmoid. Berdasarkan input yang diberi, modul perhatian menjana peta output

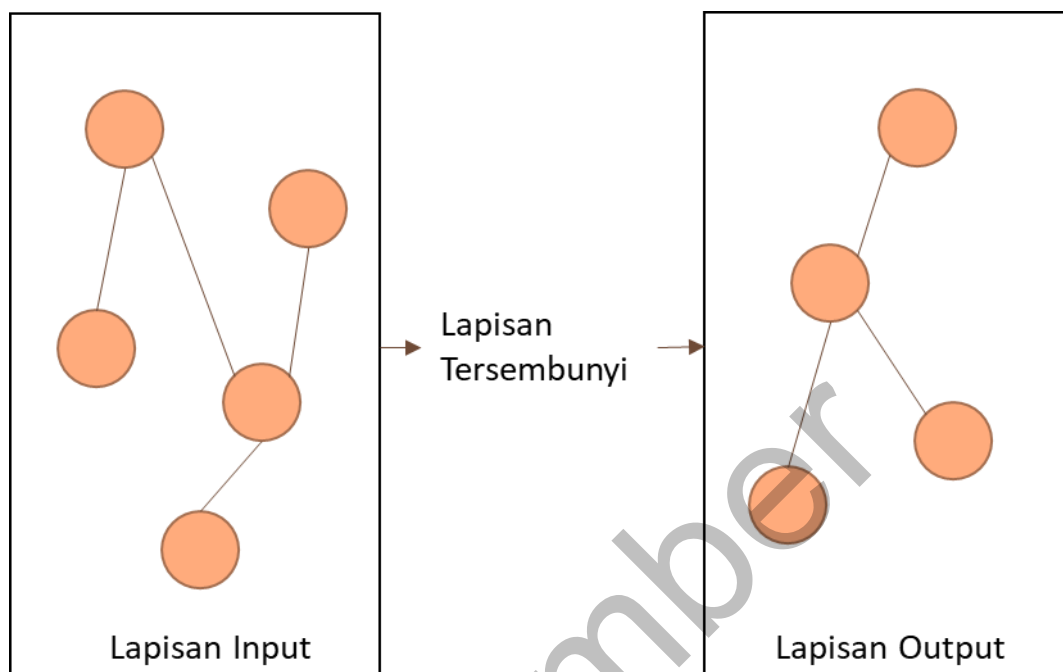
yang diperhalusi. Terdapat 2 jenis modul perhatian, modul perhatian saluran dan modul perhatian spatial bergantung kepada dimensi yang digunakan seperti yang digambarkan dalam Rajah 2.4.2(d). Walaupun modul perhatian membantu dalam memberi tumpuan kepada ciri-ciri penting, ia menyebabkan peningkatan pengiraan kerana penjanaaan peta ciri perhatian.



Rajah. 2.4.2(d): Modul perhatian. Atas: modul perhatian spatial. Bawah: modul perhatian saluran

e. Graf

Rangkaian Konvolusi Graf (GCN) adalah satu lagi lelaran CNN yang beroperasi atas struktur graf yang boleh digunakan untuk mewakili data asas. Untuk tugas MER, graf boleh digunakan untuk mewakili titik mercu tanda kening dan mulut (Lei et al. 2021). Walaupun mempunyai perwakilan data yang kuat, pembinaan struktur graf selalunya memerlukan penghasilan ciri manual (Zhang et al. 2019).



Rajah. 2.4.2(e): Rangkaian Konvolusi Graf

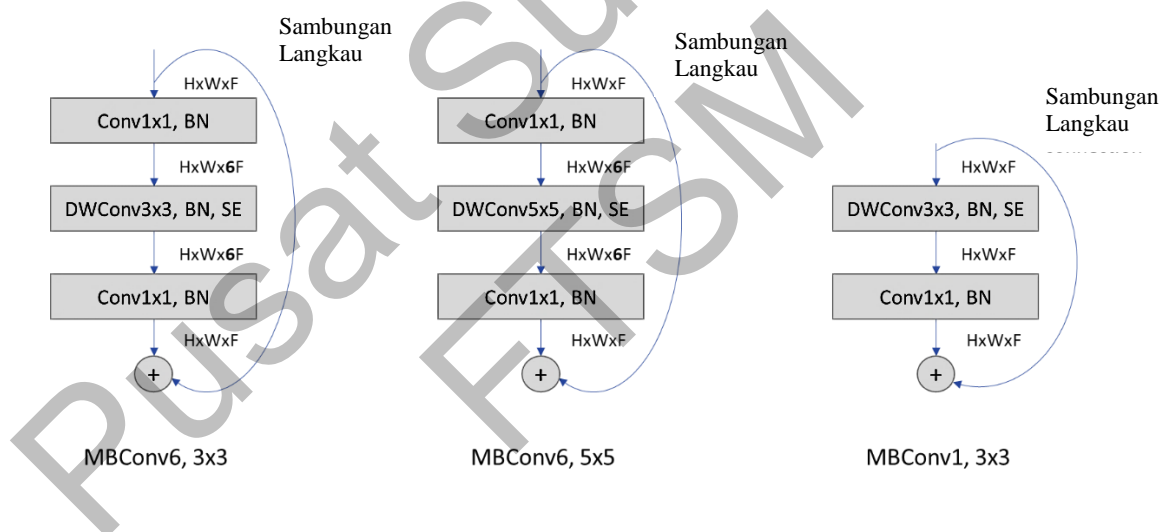
f. MBConv

Hambatan Terbalik Mobil or MBConv telah dicadangkan dalam rangka kerja seni bina MobileNet v2 oleh (Sandler et al. 2018). MBConv menggunakan sambungan langkau seperti dalam blok sisa tetapi dengan bilangan parameter yang jauh lebih kecil. Ia melakukan ini dengan terlebih dahulu memperluaskan bilangan saluran input sebelum melakukan konvolusi secara lebar boleh pisah (*depth-wise separable convolution*). Kemudian bilangan saluran menyusut lagi untuk membolehkan penambahan dengan sisa input (lihat Rajah 2.4.2(f)). MBConv dapat menggunakan bilangan saluran yang diperluaskan dalam konvolusi yang mendalam kerana proses konvolusi dilakukan secara berasingan untuk setiap saluran.

Blok MBConv kemudiannya diguna pakai oleh seni bina EfficientNet (Tan & Quoc 2019) serta dimantapkan dengan pengoptimuman '*Squeeze and Excitation*' (Hu et al. 2018). Walau bagaimanapun, peningkatan hebat prestasi model ini berbanding MobileNet dan ResNet sebenarnya terletak pada kaedah penskalaan kompaun untuk mencari kombinasi terbaik resolusi input, kedalaman dan lebar rangkaian. Mereka telah mencadangkan 8 varian gabungan resolusi, kedalaman dan lebar dari yang paling ringan

kepada versi yang paling kompleks. Versi yang paling ringan, EfficientNet-B0 melewati ResNet-50 dalam klasifikasi ImageNet walaupun mempunyai 4.9 kali kurang parameter (Tan & Quoc 2019).

EfficientNet-B0, varian paling ringan, menggunakan 3 jenis blok MBConv seperti yang digambarkan dalam Rajah 2.4.2(f).1. Yang pertama, MBConv6, 3x3 adalah susunan lapisan konvolusi, konvolusi secara lebar dengan saiz kernel 3 dan yang terakhir adalah lapisan konvolusi yang mengecilkan semula output konvolusi sebelum menambah kembali input sisa. Peta ciri input diperluaskan kepada 6 saluran sebelum konvolusi secara lebar (operasi pengembangan). Begitu juga, MBConv6, 5x5 mempunyai susunan lapisan yang sama tetapi menggunakan saiz kernel 5 dalam lapisan konvolusi secara lebar. MBConv1, 3x3 sebaliknya, terdiri daripada lapisan konvolusi secara lebar dan lapisan konvolusi biasa tanpa pengembangan saluran.



Rajah. 2.4.2(f).1: Blok MBConv yang digunakan dalam seni bina EfficientNet. Setiap blok konvolusi mempunyai lapisan BN (*Batch Normalization*). DWConv: Konvolusi secara lebar. SE: Squeeze-and-Excitation

EfficientNet mencari kombinasi terbaik susunan-susunan blok ini melalui carian seni bina saraf dengan mengoptimumkan ketepatan dan FLOPS (Tan & Quoc 2019) menghasilkan model yang mempunyai ketepatan terbaik dan pertukaran FLOPS. Dalam kajian mereka, (Tan & Quoc 2019) juga menekankan keunggulan model mereka dalam 8 tugas pemindahan pembelajaran berbanding dengan model canggih seperti ResNet-152, model DenseNet-201 dan Inception, mencapai ketepatan tertinggi dalam

lebih daripada separuh tugas walaupun ianya lebih kecil. Keputusan yang dilaporkan pada tugas klasifikasi ImageNet serta tugas pemindahan pembelajaran menunjukkan potensi EfficientNet dalam tugas lain yang serupa seperti MER. Rajah 2.5.2(f).2 menunjukkan seni bina rangkaian EfficientNet-B0 yang merupakan varian EfficientNet yang paling ringan.



Rajah. 2.5.2(f).2 .Seni bina EfficientNet-B0. Ia terdiri daripada 7 blok MBConv dengan kedalaman (bilangan lapisan) dan lebar (saiz saluran) yang berbeza. Bentuk input ialah 224x224x3

2.6 PRA-PEMROSESAN

Terdapat beberapa langkah pra-pemprosesan yang biasa digunakan untuk menyediakan set data latihan atau pengesahan silang untuk MER yang akan diterangkan dengan lebih lanjut di dalam bahagian ini. Pra-pemprosesan penting di lakukan untuk meningkatkan ketepatan MER dengan mengurangkan gangguan pada data. Pra-pemprosesan juga penting untuk mendapatkan ciri-ciri yang akan digunakan contohnya bingkai puncak atau OF.

2.6.1 Pencarian bingkai puncak

Teknik-teknik yang menggunakan ciri input bingkai puncak perlu mendapatkan kedudukan bingkai puncak dari sampel video untuk set data SMIC kerana set data ini tiada anotasi bingkai puncak. Set data CDE pula hanya mempunyai senarai nama sampel video dan label-labelnya, jadi sekiranya menggunakan set data CDE dengan ciri input bingkai puncak, langkah ini masih perlu dilakukan.

Antara cara-cara yang digunakan untuk mencari bingkai puncak adalah teknik D&C-RoI (Liong et al. 2019) yang menggunakan cara pembahagian dan penaklukan bahagian-bahagian wajah, (Li et al. 2017; Liong et al. 2016a) yang mencadangkan penggunaan ciri-ciri OF dan HOOFF untuk mencari perbezaan terbesar antara bingkai dan (Song et al. 2019) yang menggunakan ciri LBP untuk mengesan bingkai puncak. Teknik lain, seperti (Zhang et al. 2018) bergantung kepada CNN untuk mengesan bingkai puncak. Sebaliknya, (Zhou et al. 2019) menganggarkan lokasi bingkai puncak sebagai bingkai tengah di antara bingkai permulaan dan akhir dan tetap mendapat skor UF1 yang munasabah dalam MEGC 2019 sebagai salah satu daripada 3 pesaing teratas (See et al. 2019).

2.6.2 Pengesanan Titik-Titik Wajah dan Penjajaran Imej

Langkah penjajaran wajah dilakukan untuk mengatasi masalah pergerakan kepala di dalam sampel video yang boleh mengganggu keberkesanan pengekstrakan ciri-ciri dalam model MER. Untuk melakukan penjajaran wajah, titik-titik wajah dikesan terlebih dahulu untuk menentukan pemusingan imej bagi mendapatkan imej wajah yang tegak.

(Peng et al. 2017; Song et al. 2019; Xia, Z. et al. 2018) menggunakan ASM (*Active Shape Model*) untuk mengesan 68 titik-titik wajah yang digunakan untuk menjajarkan kedudukan muka. (Reddy et al. 2019, Takalkar & Xu 2017) pula menggunakan modul DLib untuk mendapatkan titik tengah mata kiri dan kanan yang digunakan untuk mengira sudut putaran.

(Lai et al. 2020; Xu et al. 2021) pula melakukan penjajaran wajah menggunakan Rangkaian Konvolusi Bertingkat Pelbagai Tugas (MTCNN) (Zhang et al. 2016) yang mengesan sempadan muka bersama-sama dengan 5 titik utama: titik pusat mata kanan, titik pusat mata kiri, titik tengah hidung, sudut kanan mulut dan sudut kiri mulut. MTCNN menggunakan hierarki (atau tingkatan) yang mengskalakan semula imej, menapis objek bukan wajah dan menyaring calon-calon muka untuk mendapatkan titik mercu tanda (Zhang et al. 2016). Setelah titik mata kiri dan kanan didapati, sudut putaran dikira untuk menjajar imej wajah.

2.7 PENILAIAN DAN METRIK

Beberapa cara telah digunakan untuk menilai keberkesanan MER seperti penilaian HDE dan CDE. Metrik yang biasa digunakan pula ialah UF1 dan UAR.

2.7.1 Penilaian Pangkalan Data Tahanan (HDE)

Penilaian pangkalan data tahanan (*Holdout Database Evaluation*) telah dicadangkan untuk digunakan dalam MEGC 2018 (Yap et al. 2018). Cara penilaian ini dilakukan ialah dengan melatih model MER pada set data berlainan dengan set data ujian. Contohnya, dalam MEGC 2018, model MER perlu dilatih terlebih dahulu pada set data CASME II sebelum dinilai pada set data SAMM dan kemudiannya yang sebaliknya pula dijalankan (latih pada set data SAMM dan nilai pada set data CASME II). Tujuan penilaian ini ialah untuk menilai pemindahan (*transferability*) ciri-ciri ekspresi mikro dari satu set data ke set data yang lain.

2.7.2 Penilaian Data Komposit (CDE)

Penilaian CDE dipelopori oleh (See et al. 2019) untuk digunakan dalam MEGC 2019 menggunakan LOSO CV. Penilaian LOSO CV dicadangkan untuk set data ekspresi mikro kerana ia dapat mengesahkan keupayaan model untuk berfungsi pada subjek baru yang tidak pernah lagi dilihat oleh model

LOSO CV, adalah lanjutan pengesahan silang Leave-One-Out (LOO) dan dilaksanakan dengan menyisihkan sampel dari satu subjek untuk ujian dan selebihnya digunakan untuk melatih model pada setiap kitaran pengesahan silang. Oleh kerana set data gabungan mempunyai sejumlah 68 subjek unik (28 dari SAMM, 24 dari CASMEII, 16 dari SMIC), proses latihan dan ujian diulang 68 kali menghasilkan 68 model yang berbeza. Pada setiap pengesahan silang, bilangan sampel yang diketepikan untuk ujian akan berbeza untuk setiap subjek kerana bilangan sampel yang diperoleh oleh setiap subjek berbeza-beza. Pada akhir setiap kitaran LOSO CV, akan ada 68 keputusan ujian yang akan digunakan untuk menilai prestasi model. LOSO CV akan menjamin penilaian bebas subjek yang sama seperti aplikasi kehidupan sebenar.

Kaedah-kaedah MER yang menggunakan penilaian dan set data CDE adalah seperti STSTNet yang mempunyai seni bina pembelajaran mendalam MER yang paling ringan (Liong et al. 2019) namun skor UFInya lebih rendah, GEME yang menggunakan Imej Dinamik sebagai input (Nie et al. 2021), ICE-GAN (Yu et al. 2020) yang setakat ini merupakan kaedah terbaik apabila hanya menggunakan imej statik bingkai puncak dan MTM-NET yang telah mencapai skor UF1 tertinggi dalam Penilaian CDE (Xia, B. et al. 2020). MTM-NET menjalankan pemindahan pembelajaran dari set data ekspresi makro (CK +, Oulu-CASIA, MMI) tetapi memerlukan kedua-dua bingkai permulaan dan puncak sebagai input.

Skor UF1 dan UAR dalam Jadual 2.7.2 adalah skor penilaian kaedah-kaedah yang telah disebutkan mengikut apa yang diterbitkan dalam penerbitan masing-masing. Kiraan parameter juga diambil dari penerbitan untuk STSTNet, dikira menggunakan saiz memori parameter model untuk ICE-GAN dan dari konfigurasi lapisan dalam model GEME. Untuk maklumat mengenai pengiraan, sila rujuk Lampiran A. Kiraan parameter untuk MTM-NET tidak dapat disimpulkan kerana penerbitannya tidak mempunyai butiran yang diperlukan.

Jadual 2.7.2: Kajian lepas yang menggunakan penilaian CDE. OF: Aliran Optik. DI: Imej dinamik. Apex: Bingkai puncak, Onset: Bingkai mula

Kaedah	Tahun	Input	Skor UF1	Skor UAR	# Parameter
STSTNet (Liong et al. 2019)	2019	OF	0.735	0.760	1.67k
ICE-GAN (Yu et al. 2020)	2020	Apex	0.845	0.841	21.6 juta
MTM-Net (Xia, B. et al. 2020)	2020	Onset-Apex	0.86	0.85	n/a
GEME (Nie et al. 2021)	2021	DI	0.7221	0.7303	53 juta

2.7.3 Metrik

Skor UF1 dan UAR sering digunakan untuk memberikan perwakilan prestasi model yang lebih tepat dalam set data yang tidak seimbang seperti set data ekspresi mikro. Berikut ialah langkah-langkah pengiraan metrik UF1 dan UAR.

a. Skor F1 Tidak Berpemberat (UFI)

Ketepatan (*Precision*) (Eq. 3.3.4(i)) dan *Recall* (Eq. 3.3.4(ii)) adalah dua metrik yang biasa digunakan untuk menilai prestasi model pembelajaran mendalam tetapi ini tidak menggambarkan prestasi keseluruhan dengan tepat. Sebaliknya, skor F1 (Eq. 3.3.4(iii)) biasanya digunakan untuk mewakili keseimbangan antara ketepatan - nisbah sampel yang diramalkan dengan betul ke atas semua ramalan untuk sesuatu kelas, dan nilai *Recall* - nisbah ramalan yang betul untuk sesuatu kelas ke atas bilangan sampel dalam kelas tersebut (Johnson & Khoshgoftaar 2019). Formula untuk metrik-metrik ini ditunjukkan di bawah:

$$Precision = \frac{TP}{TP + FP} \quad (3.3.4(i))$$

$$Recall = \frac{TP}{TP + FN} \quad (3.3.4(ii))$$

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} = \frac{2TP}{2TP + FN + FP} \quad (3.3.4(iii))$$

di mana TP adalah benar positif, FP adalah positif palsu dan FN adalah nilai negatif palsu yang boleh diambil dari matriks kekeliruan (Johnson & Khoshgoftaar 2019).

Skor UF1, atau purata makro F1 dikira dengan mengambil purata skor F1 setiap kelas (See et al. 2019). Metrik ini menganggap semua kelas sama (tidak berpemberat) dan akan membolehkan kita menilai keupayaan model untuk meramalkan semua jenis kelas dan bukannya hanya kelas tertentu dengan baik. Formula untuk UF1 dibentangkan dalam Eq. 3.3.4(iv).

$$UF1 = \frac{\sum_{i=1}^c F1_i}{N_c} \quad (3.3.4(iv))$$

dimana C ialah label kelas ('Positive', 'Negative', 'Surprise') dan N_c ialah jumlah kelas iaitu 3.

b. Purata Recall Tidak Berpemberat (UAR)

Sama seperti UF1, UAR adalah skor purata ketepatan setiap kelas yang boleh memberitahu kita tentang ketepatan keseluruhan untuk kesemua kelas (See et al. 2019). Formula UAR ditunjukkan dalam Eq. 3.3.4(v):

$$UAR = \frac{1}{C} \sum_{i=1}^C \frac{TP_i}{N_i} \quad (3.3.4(v))$$

Di mana C ialah jumlah kelas, TP_i jumlah benar positif untuk kelas i dan N_i adalah bilangan sampel dalam kelas i .

2.8 KESIMPULAN

Ketepatan kajian MER lepas yang menggunakan penilaian dan set data CDE masih boleh ditingkatkan lagi. Kaedah yang menggunakan hanya bingkai puncak sebagai input pada penilaian CDE pula belum banyak diterokai dengan hanya 2 kajian lepas yang menggunakannya. Justeru itu, kajian ini akan mencadangkan kaedah MER menggunakan bingkai puncak pada penilaian dan set data CDE.

Berdasarkan dari kajian lepas, hasil yang dilaporkan telah menunjukkan keputusan yang menggalakkan dalam MER apabila menggunakan teknik pemindahan pembelajaran (Peng et al. 2018; Quang et al. 2019; Wang et al. 2018; Xia, B. et al. 2020). (Patel et al. 2016) dan (Xia, B. et al. 2020) pula mendapati bahawa pangkalan data sumber dan sasaran yang serupa dapat meningkatkan prestasi MER, serta prestasi yang menggalakkan dalam kajian (Xia, B. et al. 2020) apabila menggunakan penyesuaian domain dari pangkalan data ekspresi makro. Maka, penyelidikan ini memilih kaedah pemindahan pembelajaran dari pangkalan data ekspresi makro.

Pangkalan data FER 201 dipilih sebagai set data sumber pemindahan pembelajaran dan bukannya secara langsung dari pangkalan data imej umum, ImageNet berdasarkan kajian (Li & Deng 2020) yang menunjukkan prestasi FER 2013 sebagai data sumber untuk pemindahan pembelajaran ke set-set data ekspresi wajah yang lain. Pangkalan data FER 2013 sudah pastinya lebih mirip dengan set data ekspresi

mikro yang mengandungi imej ekspresi wajah berbanding ImageNet. Oleh kerana matlamat pemindahan pembelajaran adalah untuk belajar dari pangkalan data sumber untuk digunakan pada tugas sasaran (Patel et al. 2016), kajian ini menganjurkan bahawa pangkalan data dengan imej wajah, seperti pangkalan data ekspresi makro seperti FER 2013, jauh lebih sesuai di MER seperti yang akan dibuktikan dalam kajian ini. Namun, setakat ini FER 2013 belum diterokai untuk tugas MER.

Dilhamkan oleh (Kadokia et al. 2021), penyelidikan ini mencadangkan model ringan berdasarkan EfficientNet (Tan & Quoc 2019) yang menggunakan blok binaan MBConv yang sama seperti MobileNet v2 (Sandler et al. 2018). EfficientNet mempunyai kelebihan model yang ringan seperti MobileNet tetapi lebih ampuh akibat penskalaan kompaun kedalaman, lebar dan resolusi input rangkaian. EfficientNet juga telah melepasi MobileNet dan ResNet dalam tugas klasifikasi ImageNet.

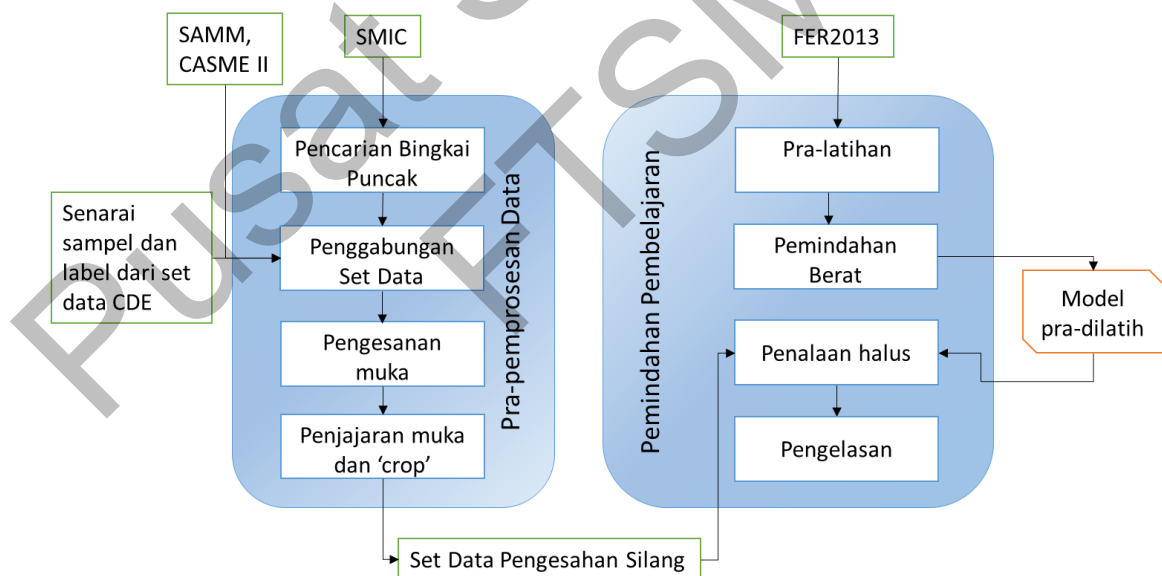
Namun, lebih penting lagi, (Tan & Quoc 2019) juga telah menunjukkan bahawa model EfficientNet dengan pemberat awal dari pra-latihan dengan ImageNet menunjukkan prestasi yang baik dan mencapai hasil unggul untuk 5 daripada 8 tugas pemindahan pembelajaran pada set data yang biasa digunakan seperti CIFAR-100 dan Flowers walaupun ianya jauh lebih kecil. Oleh itu, EfficientNet dipilih sebagai model asas untuk kaedah yang dicadangkan.

BAB III

KAEDAH KAJIAN

3.1 CADANGAN KAEDAH

Kaedah yang dicadangkan adalah seperti yang digambarkan dalam Rajah 3.1 yang terdiri daripada langkah-langkah pra-pemprosesan data untuk menyediakan dan menggabungkan set-set data ekspresi mikro diikuti dengan langkah-langkah pemindahan pembelajaran yang merangkumi cadangan model, pra-latihan dan langkah-langkah penalaan. Langkah-langkah ini digariskan secara terperinci dalam seksyen-seksyen berikutnya.



Rajah. 3.1: Cadangan Kaedah Kajian

3.1.1 Mengesan Bingkai Puncak

Kaedah yang dicadangkan dalam penyelidikan ini menggunakan hanya satu bingkai iaitu bingkai puncak, untuk setiap sampel video dalam set data sebagai input. Penyelidikan ini mengambil inspirasi daripada penemuan (Liong et al. 2016b) dan (Sun et al. 2020) bahawa menggunakan bingkai puncak adalah mencukupi untuk pengecaman ekspresi mikro kerana tidak ada jaminan ketepatan pengecaman yang lebih baik dengan menggunakan lebih atau semua bingkai dalam sampel. Bingkai puncak adalah bingkai apabila ekspresi mikro mencapai puncaknya, manakala *onset* dan *offset* adalah bingkai apabila ekspresi mikro bermula dan berakhir.

Set data CASME II dan SAMM kedua-duanya mempunyai anotasi lokasi bingkai puncak. Walau bagaimanapun, set data SMIC tidak mempunyai maklumat ini. Maka, langkah mengesan kedudukan bingkai puncak perlu dilakukan untuk set data SMIC.

Penyelidikan ini memutuskan untuk mengambil pendekatan (Zhou et al. 2019) berdasarkan andaian bahawa memandangkan video pendek setiap sampel memang telah diketahui mengandungi satu contoh ekspresi mikro sahaja, perbezaan antara bingkai puncak dan beberapa bingkai yang dekat dengannya boleh diabaikan kerana sifat tak ketara dan pantas ekspresi mikro. Sebagai contoh, SMIC-HS yang digunakan dalam set data CDE telah dirakam dengan kamera berkelajuan tinggi 100fps (Yan et al. 2013) yang bermaksud setiap bingkai hanya merangkumi 0.01 saat yang menjadikan perbezaan antara beberapa bingkai bersebelahan satu sama lain, kecil.

Pengiraan lokasi bingkai puncak yang digunakan, yang dianggarkan sebagai bingkai tengah ditunjukkan di bawah:

Berdasarkan senarai bingkai yang merupakan disusun dari bingkai mula ke bingkai akhir dalam video sampel:

```
lokasiFramePuncak = panjang(bilanganBingkai)/2
```

```
framePuncak = senaraiBingkai[lokasiFramePuncak]
```

3.1.2 Penggabungan Set Data

Eksperimen di dalam kajian ini dijalankan menggunakan protokol Penilaian Pangkalan Data Komposit (CDE) yang digunakan dalam MEGC 2019 (See et al. 2019). Ini adalah pendekatan yang realistik di mana set data mengandungi sampel dari subjek pelbagai etnik dan jantina serta persekitaran tangkapan (seperti resolusi kamera yang berbeza dll).

Memandangkan set data CDE hanya mempunyai senarai nama sampel video dan label-labelnya sedangkan kajian ini menggunakan bingkai puncak bukan keseluruhan video, untuk membina set data yang akan digunakan semasa latihan dan pengesanan silang, bingkai-bingkai puncak dikumpulkan dan dilabelkan semula menurut senarai nama sampel video dan label dalam set data CDE. Bingkai-bingkai puncak ditentukan oleh kaedah dalam seksyen 3.1.1 di atas. Secara keseluruhan terdapat 442 sampel dari gabungan SMIC, CASMEII dan SAMM dalam set data akhir.

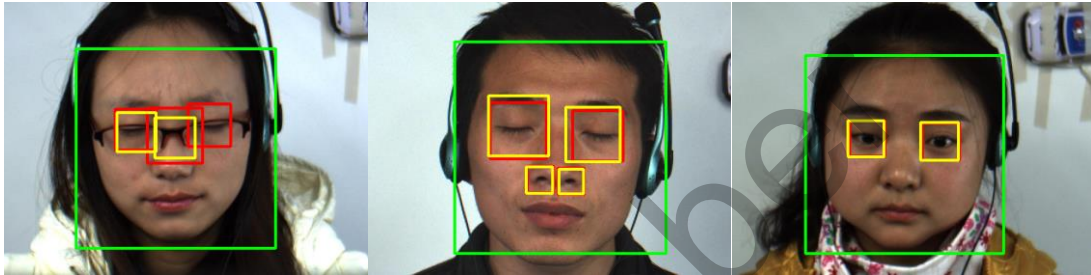
3.1.3 Pengesanan Muka/Mata

Dalam langkah ini, pengesanan wajah dilakukan pada imej bingkai puncak dengan matlamat untuk mendapatkan titik-titik utama yang akan digunakan dalam penjajaran muka dan pemotongan imej. Titik-titik utama ini adalah koordinat mata kanan dan mata kiri serta koordinat sempadan muka dalam imej.

Kajian ini menilai 2 kaedah pengesanan muka untuk menentukan cara yang paling sesuai untuk set data CDE. Kaedah pertama adalah menggunakan model pra-latih Haar-Cascade (Chandrappa et al. 2019) untuk pengesanan muka dan mata yang datang bersama-sama dengan modul OpenCV (Patel et al. 2016). Model yang digunakan untuk pengesanan wajah adalah 'haarcascade_frontalface_default.xml' dan untuk pengesanan mata, dua model telah digunakan: satu dilatih untuk mengesan mata terbuka ('haarcascade_eye.xml') dan satu untuk mengesan mata dengan cermin mata ('haarcascade_eye_tree_eyeglasses.xml').

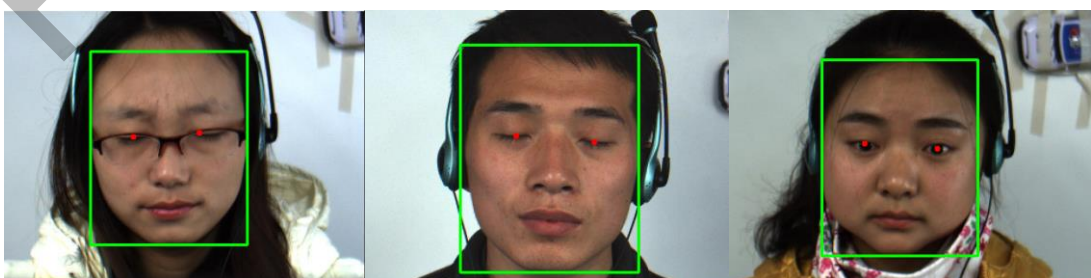
Seperti yang dilihat dalam Rajah 3.1.3, Haar-Cascade tidak berfungsi dengan baik untuk sampel dengan mata tertutup dan bercermin mata. Walaupun parameter

algoritma Haar-Cascade seperti saiz objek yang akan dikesan dan bilangan minimum jiran untuk setiap kemungkinan pengesanan, boleh diperhalusi untuk mengesan kes-kes tepi ini, ia tidak dapat generalisasi dengan baik untuk setiap sampel dalam set data dan memerlukan pemilihan parameter manual untuk setiap kes pengesanan (seperti mengesan mata tertutup berbanding mata dengan cermin mata).



Rajah. 3.1.3(a): Pengesanan muka dan mata menggunakan Haar-Cascade. Kotak hijau: Sempadan muka yang dikesan. Kotak kuning: Mata yang dikesan. Kotak merah: Mata dikesan dengan cermin mata.

Kaedah kedua yang dinilai adalah menggunakan Rangkaian Konvolusi Bertingkat Pelbagai Tugas (MTCNN) (Zhang et al. 2016). Seperti yang ditunjukkan dalam Rajah 3.1.3(b), sempadan muka yang dikesan adalah lebih dekat dengan garis sebenar muka dan algoritma ini berjaya mengesan lokasi mata dengan sangat baik tanpa mengira kehadiran cermin mata atau mata tertutup. Berdasarkan penilaian ini, kajian ini memilih MTCNN sebagai kaedah yang digunakan untuk pengesanan muka dan mata untuk digunakan pada set data.



Rajah. 3.1.3(b): Pengesanan muka dan mata menggunakan MTCNN. Kotak hijau: Sempadan muka yang dikesan. Titik merah: Titik tengah mata yang dikesan.

3.1.4 Penjajaran Muka dan Pemotongan Imej

Penjajaran muka dan pemotongan imej biasanya digunakan di MER untuk memastikan bahawa algoritma pengesanan tidak terpengaruh dengan pergerakan kepala dan kawasan di sekitar muka (contohnya rambut, latar belakang). Oleh kerana ekspresi mikro adalah halus, maka ianya amat penting untuk membuang pergerakan besar dan pencemaran dalam imej yang tidak berkaitan dengan ekspresi wajah.

Penjajaran muka dilakukan dengan memutar imej berpaksikan titik tengahnya. Sudut putaran dikira menggunakan formula matematik tangen songsang ke atas jarak antara lokasi pusat mata kiri dan kanan. Pengiraan ditunjukkan dalam Eq. 3.1.4.

$$\begin{aligned} \text{delta_x} &= \text{right_eye_x} - \text{left_eye_x} \\ \text{delta_y} &= \text{right_eye_y} - \text{left_eye_y} \\ \text{rotation_angle} &= (\text{atan}(\text{delta_y}/\text{delta_x}) * 180) / \text{Pi} \end{aligned} \quad (3.1.4)$$

Selepas itu, imej dipotong menggunakan lokasi sempadan muka yang diperolehi dari langkah sebelum ini. Potongan imej dilakukan untuk menghapuskan butiran yang tidak perlu yang tidak berkaitan dengan ekspresi wajah dalam imej (Quang et al. 2019; Xia, Z. et al. 2020a). Wajah yang dipotong mungkin mempunyai saiz yang berbeza, jadi dalam kajian ini imej diubah saiz kepada 224 x 224 sambil memastikan imej diskalakan sewajarnya. Dimensi imej (224 x 224) dipilih mengikut keperluan input model yang dicadangkan.



Rajah. 3.1.4: Penjajaran muka dan pemotongan imej. Dari kiri ke kanan: Imej asal, Imej yang dijajarkan wajah, Imej yang dipotong (*crop*).

3.1.5 Augmentasi Data

Dalam kajian ini, kaedah transformasi imej mudah dilakukan diilhamkan oleh (Takalkar et al. 2020) yang memerhatikan bahawa walaupun dengan hanya menggunakan transformasi mudah, prestasi MER bertambah baik sebanyak 15%. Transformasi imej yang digunakan adalah pembalikan mendatar dan kecerahan rawak pada sampel rawak dalam data latihan. Langkah ini dijalankan sebahagian daripada pemisahan data ujian dan latihan. Benih (*seed*) yang digunakan untuk transformasi adalah nombor yang dijana secara rawak antara 0 dan 10.

3.1.6 Pindah Belajar

Oleh kerana set data ekspresi mikro bersaiz kecil dan tidak mencukupi untuk melatih rangkaian pembelajaran yang mendalam tanpa menyebabkan masalah 'overfitting', penyelidikan ini akan menggunakan kaedah pemindahan pembelajaran untuk memanfaatkan pangkalan data ekspresi makro yang lebih besar.. Sub-seksyen berikut akan menerangkan langkah-langkah yang dijalankan dalam kajian ini untuk pemindahan pembelajaran .

a. Berat Permulaan

Pertama, berat model yang dicadangkan diset dengan berat yang telah dilatih dengan ImageNet. Menggunakan berat permulaan boleh membantu dalam mengurangkan masa latihan dan membuat model berpusat lebih cepat daripada dengan latihan dari awal (Sun et al. 2020). Berat pra-dilatih yang digunakan datang bersama-sama dengan modul Keras yang digunakan.

b. Pra-latihan dari pangkalan data FER 2013

Set data pra-latihan yang digunakan adalah set data FER 2013. Sama seperti set data ekspresi mikro, label sampel dalam pangkalan data FER 2013 dilabelkan semula ke dalam 3 kelas seperti yang dibentangkan dalam Jadual 3.1.6: Positif (8989 sampel), Negatif (16698 sampel) dan *Surprise* (4002 sampel) dengan imej berlabel 'Neutral' dibuang. Ini membolehkan model pra-latihan untuk mengklasifikasikan bilangan kelas yang sama seperti set data CDE.

Jadual 3.1.6: Label dan bilangan sampel dalam pangkalan data FER 2013.

Kelas baharu	Kelas asal	Jumlah
Negative	Angry: 4953, Disgust: 547, Fear: 5121, Sad: 6077	16698
Positive	Happiness: 8989	8989
Surprise	Surprise: 4002	4002
N/A	Neutral: 6198	N/A

Selain daripada pelabelan semula, tiada langkah pra-pemprosesan lain seperti penjajaran muka dilakukan. Intuisi untuk melakukan ini adalah untuk membolehkan rangkaian untuk mempelajari seberapa banyak ciri-ciri muka umum yang mungkin boleh dipindahkan ke tugas MER. Set data kemudiannya dibahagikan kepada set latihan dan pengesahan dengan pembahagian 80:20 sebelum pra-latihan. Memandangkan imej dalam FER 2013 mempunyai dimensi yang lebih kecil (48x48), lapisan input dalam model yang dicadangkan diselaraskan untuk menampung saiz yang lebih kecil semasa pra-latihan.

Untuk mengelakkan model daripada mengekstrak terlalu banyak maklumat yang akan menyebabkan ia terlalu sesuai dengan pangkalan data FER 2013, pra-latihan dilakukan hanya sebanyak 5 ulangan dan 100 lelaran dengan mekanisme berhenti awal selepas 10 lelaran tanpa penurunan kepada fungsi kehilangan set data pengesahan. Dalam setiap ulangan, lapisan kepala model dilatih terlebih dahulu dengan membekukan lapisan-lapisan asas dengan kadar pembelajaran awal (LR) 0.001 dan saiz kelompok 16. Saiz kelompok 16 digunakan kerana had memori GPU.

Selepas itu, keseluruhan rangkaian dilatih dengan kadar pembelajaran awal yang lebih rendah sebanyak 0.0001. Ini memastikan bahawa berat permulaan dari pra-latihan ImageNet tidak diganti tetapi sebaliknya ia dikekalkan dan diperhalusi untuk disesuaikan kepada set data ekspresi makro.

c. Penalaan halus pada set data pengesahan silang

Oleh kerana saiz imej pangkalan data FER 2013 dan set data pengesahan silang adalah berbeza, dalam langkah ini pada mulanya model yang dicadangkan dibina dengan seni bina yang sama digunakan untuk pra-latihan FER 2013 tetapi dengan

dimensi input asal 243x243. Kemudian, berat dari model pra-dilatih disalin ke model yang akan digunakan dalam langkah penalaan halus.

Penalaan halus mengikuti langkah-langkah latihan yang sama seperti pra-latihan. Pertama, kepala model dilatih dengan membekukan lapisan asas-lapisan sebanyak 5 ulangan, saiz kelompok 16 dan 15 lelaran dengan mekanisme berhenti awal selepas 2 lelaran tanpa penurunan fungsi kehilangan pada set data pengesahan. Kemudian, seluruh rangkaian dilatih untuk 5 ulangan lagi, saiz kelompok 16 dan 20 lelaran dengan mekanisme berhenti awal yang sama. Kadar pembelajaran (LR) awal untuk latihan kepala adalah 0.001 dan 0.0001 untuk keseluruhan rangkaian.

Bilangan lelaran yang lebih rendah dipilih dalam langkah ini berdasarkan penilaian awal bilangan optimum lelaran dalam setiap ulangan untuk set data pengesahan silang. Semasa fasa penyiasatan, penyelidikan ini mendapati bahawa untuk set data kecil seperti set data pengesahan silang bilangan lelaran yang lebih rendah dapat menghalang '*overfitting*'. Untuk membolehkan rangkaian terus belajar, lebih banyak ulangan adalah lebih baik daripada lebih banyak epoch kerana ini membolehkan rangkaian untuk memulakan penurunan kecerunan dari ruang carian lain.

Langkah penalaan halus pada set data pengesahan silang dijalankan dari awal sebanyak 3 kali menggunakan pengesahan silang LOSO dan hasilnya di puratakan untuk mendapat skor akhir.

3.1.7 Model cadangan

Dalam bahagian ini, intuisi di sebalik pemilihan model asas dan seni bina model yang dicadangkan dijelaskan.

f. EfficientNet sebagai model asas

Dalam kajian perbandingan oleh (Kadokia et al. 2021) yang disahkan pada set data SAMM, didapati bahawa pemindahan pembelajaran dengan MobileNet v1 mencapai skor F1 tertinggi dan kerugian silang terendah diikuti oleh model VGG16 dan ResNet50 walaupun ia adalah model terkecil.

Terdapat 7 varian model EfficientNet dari yang terkecil hingga yang terbesar. Berdasarkan penemuan oleh (Liong et al. 2019; Xia, Z. et al. 2020b) bahawa rangkaian cetek dan lebih ringan berfungsi dengan lebih baik pada set data MER, kajian ini memutuskan untuk menggunakan varian yang paling ringan, model EfficientNet-B0 yang mempunyai 5.3 juta parameter.

g. Seni bina Model

Dalam kajian (Xia, Z. et al. 2020b; Liong et al. 2019; Khor et al. 2019), telah ditampikan bahawa rangkaian yang lebih ringan (bilangan parameter yang lebih kecil) dan cetek (kurang kedalaman) menunjukkan prestasi yang lebih baik untuk set data yang kecil dan tidak seimbang. Walau bagaimanapun, model asas yang telah dipilih kajian ini, EfficientNet-B0, walaupun merupakan salah satu model CNN paling ringan masih agak besar berbanding rangkaian cetek lain yang dicadangkan untuk MER dengan kira-kira 5.3 juta parameter. Sebagai perbandingan, (Khor et al. 2019) hanya memerlukan 0.97 juta parameter dan (Liong et al. 2019), rangkaian yang sangat kecil dengan hanya 1.67k parameter. Namun, kaedah-kaedah itu diberikan ciri-ciri tahap rendah yang sudah diekstrak terlebih dahulu maka tidak memerlukan lapisan tambahan untuk pengekstrakan ciri.

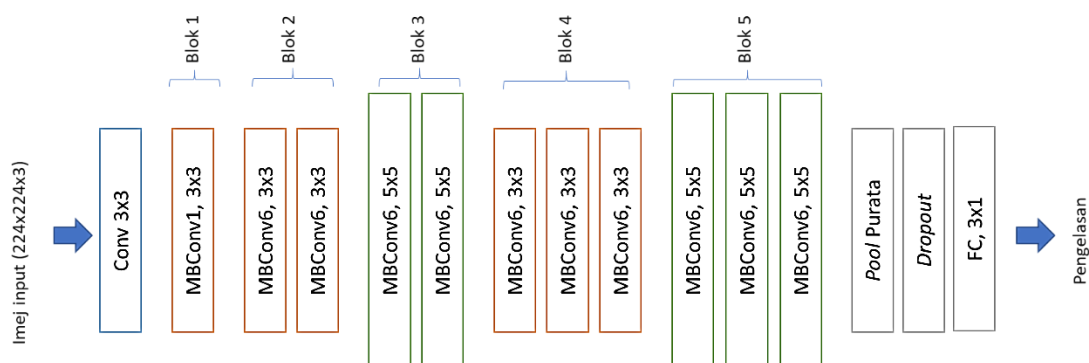
Untuk itu, kajian ini mencadangkan versi EfficientNet-B0 yang dipangkas, dipanggil Efficient-ME yang memiliki keseimbangan antara kerumitan model (bilangan parameter) dan keupayaan pengekstrakan ciri. Matlamatnya adalah untuk menunjukkan bahawa rangkaian ringan di bawah satu juta parameter boleh berfungsi dengan baik jika tidak lebih baik daripada rangkaian yang lebih berat.

Model yang dicadangkan mengekalkan 5 blok pertama EfficientNet-B0 dan mengesampingkan 2 blok terakhir. Selepas blok ke-5 EfficientNet-B0, lapisan *Pooling* purata global dan lapisan *Dropout* ditambah sebelum diteruskan ke lapisan padat dengan 3 nod untuk klasifikasi. Lapisan *Pooling* purata global telah digunakan untuk menjana peta ciri output untuk digunakan dalam lapisan klasifikasi padat. Kadar keciciran 0.5 digunakan dalam lapisan *Dropout* untuk mengelakkan 'overfitting' seperti yang digunakan oleh (Liong et al. 2018; Peng et al. 2017; Nie et al. 2021). Dimensi

input dikekalkan pada 224x224 dengan 3 saluran. Imej input tidak diubah kepada skala kelabu atau dinormalkan tidak seperti kajian-kajian lain kerana beberapa lapisan pertama EfficientNet yang dilaksanakan oleh modul Keras sudah mengerjakan pra-pemrosesan input. Lapisan klasifikasi padat menggunakan fungsi pengaktifan *Softmax* bersama-sama dengan *regularizer* kernel L1 dan L2 dengan kedua-dua faktor regularisasi ditetapkan kepada 0.01 seperti (Hashmi et al. 2021). *Regularizer* kernel berfungsi mengurangkan berat kernel dan biasanya digunakan untuk mengelakkan ‘*overfitting*’.

Dengan tetapan yang dicadangkan, jumlah parameter dalam Efficient-ME adalah 867k yang berada di bawah sasaran 1 juta dan lebih kecil daripada (Khor et al. 2019). Seni bina model ini digambarkan dalam Rajah 3.1.7(b).

Justifikasi untuk memangkas EfficientNet-B0 hingga blok ke-5 berpunca daripada pemerhatian (Tan et al. 2019) bahawa pelbagai lapisan MBConv dalam satu rangkaian memberikan hasil yang lebih baik daripada menggunakan satu jenis lapisan sahaja. Berdasarkan ini, memangkas sehingga blok ke-5 akan mengekalkan pasangan MBConv6 dengan konvolusi 3x3 dan MBConv6 dengan konvolusi 5x5 (pasangan Blok 2 dan Blok 3 serta blok 4 dan Blok 5). Satu lagi motivasi untuk memilih Blok 5 sebagai titik pangkasan adalah saiz model sasaran di bawah 1 juta yang boleh dicapai dengan memotong di blok ke-5. Ianya juga penting untuk memastikan lapisan konvolusi terakhir dapat mengekstrak ciri-ciri yang mencukupi demi klasifikasi yang tepat. Memotong rangkaian terlalu cetek akan menyebabkan kurangnya ciri dan butiran untuk difahami oleh model.



Rajah. 3.1.8: Cadangan seni bina model berdasarkan EfficientNet-B0.

3.2 UJI KAJI

Bahagian ini menerangkan eksperimen yang telah dijalankan, tetapan konfigurasi seperti pengoptimum dan fungsi kehilangan yang digunakan, kaedah penilaian dan kajian asas untuk dibandingkan semasa penilaian.

3.2.1 Persediaan Uji Kaji

Semua eksperimen dilakukan dengan Windows 11 (x64) yang mempunyai 1 NVIDIA GeForce RTX 3070 GPU dengan memori khusus 8GB. Model ini dibangunkan menggunakan modul Keras pada rangka kerja Tensorflow 2.7 dan bahasa pengaturcaraan Python 3.9.

Eksperimen yang dijalankan berfungsi untuk menjawab soalan penyelidikan yang ditimbulkan dalam Seksyen 1.5 dan boleh dibahagikan kepada 3 eksperimen berasingan mengikut Jadual 3.2.1 di bawah yang menilai sama ada model ringan di bawah satu juta parameter dengan kerumitan input yang rendah dapat mencapai prestasi MER yang tinggi, kesan set data sumber yang digunakan dalam pemindahan pembelajaran, dan akhirnya menangani masalah set data yang tidak seimbang tanpa menghasilkan sampel baru. Persediaan terperinci untuk menjalankan setiap percubaan dijelaskan dalam sub-bahagian seterusnya.

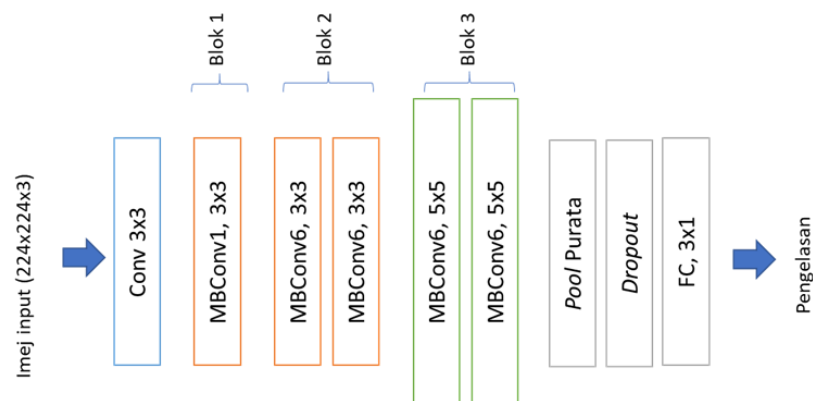
Jadual 3.2.1: Matlamat Uji Kaji

Soalan Kajian	Matlamat Uji Kaji
Q1 & Q3	1. Menunjukkan bahawa model ringan dengan kerumitan input yang rendah boleh mencapai prestasi MER yang tinggi
Q2	2. Menunjukkan bahawa pangkalan data sumber yang digunakan dalam pemindahan pembelajaran memainkan peranan besar dalam MER
Q4 & Q5	3. Menunjukkan bahawa masalah set data yang tidak seimbang boleh diatasi dengan penalaan algoritma

a. **Uji Kaji 1 – Model ringan dengan kerumitan input yang rendah boleh mencapai prestasi MER yang tinggi**

Untuk menunjukkan bahawa model ringan dengan bilangan parameter yang lebih kecil dan menggunakan input tidak kompleks boleh mencapai prestasi MER yang tinggi berbanding dengan kajian-kajian unggul semasa dalam CDE, penyelidikan ini akan menilai model yang dicadangkan, Efficient-ME, yang mempunyai parameter sebanyak 867k dan hanya mengambil bingkai puncak tanpa pengekstrakan ciri-ciri dengan menggunakan protokol penilaian MEGC, iaitu LOSO CV dengan skor UF1 dan UAR sebagai metrik. Menggunakan protokol penilaian MEGC memastikan konsistensi dan menyediakan perbandingan yang adil dan piawai dengan kajian bandingan.

Tambahan pula, varian yang lebih ringan daripada model yang dicadangkan, Efficient-ME, yang akan dipanggil Efficient-ME(t3) yang dibina dengan memangkas model EfficientNet-B0 dengan mengekalkan hanya 3 blok MBConv pertama bersama-sama dengan lapisan *Pooling* purata global, lapisan *Dropout* dan lapisan padat ditambah dengan tetapan yang sama dengan Efficient-ME juga dinilai untuk mewajarkan seni bina Efficient-ME (Rujuk rajah 3.2.1(a)). Model ini jauh lebih kecil dengan hanya 68k parameter dan mengikuti proses pemindahan pembelajaran yang sama yang diterangkan dalam seksyen 3.1.6 iaitu berat permulaan diset dari berat pra-dilatih ImageNet, pra-latihan dengan set data FER 2013 dan penalaan terakhir pada set data pengesahan silang menggunakan kaedah penilaian CV LOSO. Seni bina Efficient-ME(t3) digambarkan dalam Rajah 3.2.1.



Rajah. 3.2.1: Seni bina Efficient-ME(t3). Model ini dibina dengan memotong EfficientNet-B0 dengan mengekalkan hanya 3 blok MBConv pertama.

b. Uji Kaji 2 – Pangkalan data sumber yang digunakan dalam pemindahan pembelajaran memainkan peranan besar dalam MER

Untuk eksperimen ini, kajian ini membandingkan prestasi model yang dicadangkan, Efficient-ME, apabila pra-dilatih dengan pangkalan data ekspresi makro, FER 2013 dan apabila menggunakan berat yang telah dilatih dari pangkalan data umum, ImageNet secara langsung.

Untuk mengurangkan kesan set data yang tidak seimbang dalam FER 2013, kajian ini menggunakan bilangan sampel yang sama dari 3 kelas FER 2013 iaitu semua 4002 sampel dalam kelas 'Surprise' dan secara rawak memilih 4002 sampel dari setiap kelas 'Negatif' dan 'Positif' untuk mendapatkan sejumlah 12006 sampel. Ini memberikan perbandingan yang tidak berat sebelah kerana ImageNet jauh lebih besar daripada FER 2013 memberikannya kelebihan dan kurang terjejas oleh kelas yang tidak seimbang. 2 model pra-dilatih kemudian diperhalusi pada set data pengesahan silang dengan tetapan yang sama dan dinilai menggunakan protokol pengesahan silang LOSO.

h. Uji Kaji 3 – Masalah set data yang tidak seimbang boleh diatasi dengan penalaan algoritma

Di samping mendekati masalah set data yang tidak seimbang dari perspektif penjanaan data atau pensampelan, kajian ini berpendapat bahawa masalah ini dapat juga ditangani dari perspektif penalaan algoritma. Apa yang dimaksudkan dengan penalaan algoritma adalah apa-apa kaedah yang mengubah atau menala bagaimana algoritma atau rangkaian mengendalikan set data yang tidak seimbang.

Dalam eksperimen ini, kajian ini mendakwa bahawa penalaan algoritma dengan cara menyesuaikan fungsi kehilangan untuk mengambil kira pengagihan kelas yang tidak seimbang adalah mencukupi untuk mengurangkan masalah set data yang tidak seimbang. Untuk membuktikan dakwaan ini, penyelidikan ini membandingkan prestasi Efficient-ME dengan mengubah fungsi kehilangan yang digunakan semasa penalaan halus pada set data CDE yang tidak seimbang. Fungsi kehilangan yang dicadangkan adalah kehilangan fokus pelbagai kelas dengan berat kelas seperti yang diperincikan dalam Seksyen 3.2.3, yang direka untuk memberi lebih banyak kepentingan kepada