

**PENGENALAN EMOSI MANUSIA MENGGUNAKAN *LOG-GABOR*
CONVOLUTIONAL NETWORKS MELALUI PENDEKATAN *FACIAL*
*REGION SEGMENTATION***

SKRIPSI

diajukan untuk memenuhi sebagian dari
syarat memperoleh gelar Sarjana S-1
Program Studi Ilmu Komputer



Oleh
Naufan Rusyda Faikar
NIM 1607645

**PROGRAM STUDI ILMU KOMPUTER
DEPARTEMEN PENDIDIKAN ILMU KOMPUTER
FAKULTAS PENDIDIKAN MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS PENDIDIKAN INDONESIA
BANDUNG
2020**

PENGENALAN EMOSI MANUSIA MENGGUNAKAN *LOG-GABOR*
CONVOLUTIONAL NETWORKS MELALUI PENDEKATAN *FACIAL REGION*
SEGMENTATION

Oleh
Naufan Rusyda Faikar
1607645

Sebuah skripsi yang diajukan untuk memenuhi sebagian dari syarat memperoleh
gelar Sarjana S-1 Program Studi Ilmu Komputer di Fakultas Pendidikan
Matematika dan Ilmu Pengetahuan Alam

© Naufan Rusyda Faikar 2020
Universitas Pendidikan Indonesia
Agustus 2020

Hak Cipta Dilindungi Undang-Undang
Skripsi ini boleh diperbanyak seluruhnya atau sebagian, dengan dicetak ulang,
difotokopi atau cara lainnya tanpa izin dari penulis dengan tetap mencantumkan
halaman ini apa adanya

NAUFAN RUSYDA FAIKAR
1607645

PENGENALAN EMOSI MANUSIA MENGGUNAKAN *LOG-GABOR*
CONVOLUTIONAL NETWORKS MELALUI PENDEKATAN *FACIAL REGION*
SEGMENTATION

disetujui dan disahkan oleh pembimbing:

Pembimbing I,

Prof. Dr. H. Wawan Setiawan, M.Kom
NIP. 196601011991031005

Pembimbing II,

Yaya Wihardi, S.Kom., M.Kom.
NIP. 198903252015041001

Mengetahui,
Ketua Departemen Pendidikan Ilmu Komputer

Dr. Lala Septem Riza, M.T.
NIP. 197811262008121001

PERNYATAAN

Saya menyatakan bahwa skripsi berjudul “Pengenalan Emosi Manusia Menggunakan *Log-Gabor Convolutional Networks* Melalui Pendekatan *Facial Region Segmentation*” ini adalah benar karya saya sendiri. Dan saya tidak melakukan tindakan plagiat yang menyalahi etika dalam karya tulis ilmiah. Apabila saya terbukti bersalah, maka saya bersedia untuk memperbaiki diri, meminta maaf kepada pihak yang bersangkutan dan menanggung setiap sanksi yang berlaku.

Bandung, Agustus 2020

Yang membuat pernyataan,

Naufan Rusyda Faikar

*Untuk ibu, bapak
dan adik-adikku tercinta.*

KATA PENGANTAR

Segala puji bagi Allah *Subhānahu wa Ta'āla*, yang dengan nikmat-Nya maka sempurnalah segala kebaikan. Tiada daya dan upaya kecuali hanya dari-Nya. Hanya dengan memohon pertolongan-Nya, penulis dapat menyelesaikan skripsi berjudul “Pengenalan Emosi Manusia dengan *Log-Gabor Convolutional Networks* Melalui Pendekatan *Facial Region Segmentation*” ini tepat waktu. Skripsi ini disusun dalam rangka memenuhi sebagian syarat memperoleh gelar Sarjana S-1 Jurusan Ilmu Komputer di Universitas Pendidikan Indonesia.

Pada kenyataannya, skripsi ini bukan merupakan kredit tersendiri bagi penulis. Melainkan merupakan upaya murni kolaboratif dengan berbagai pihak selama penulis belajar di bangku perkuliahan. Setelah memulai dengan mengucapkan syukur kepada Allah *Subhānahu wa Ta'āla* di atas segalanya, penulis ingin mengucapkan terima kasih yang tulus kepada kedua pembimbing skripsi ini, bapak Yaya Wihardi dan bapak Wawan Setiawan, karena telah bersedia dengan sepenuh hati membimbing penulis dalam menyelesaikan skripsi ini.

Sejak diberikan kesempatan oleh bapak Yaya untuk bergabung bersama beliau, bapak Wawan, ibu Enjun Junaeti dan keempat anggota lain dalam riset *smart classroom*, penulis merasa lebih beruntung dari kebanyakan teman-teman lain. Selama melakukan riset bersama-sama, penulis telah banyak belajar dari berbagai tahap yang telah dilalui.

Penulis ingin mengucapkan terima kasih sebanyak-banyaknya kepada bapak Yaya sebagai mentor terbaik. Selama melakukan bimbingan, beliau telah memberikan sebuah advis kepada penulis bahwa skripsi yang bagus adalah yang selesai. Ketika skripsi itu harus tertunda penyelesaiannya karena ingin serba perfek, maka akan banyak peluang yang mungkin terlewatkan. Terus terang, penulis sangat menyukai bagaimana beliau mengumpulkan setiap mahasiswanya di ruangan yang sama untuk melakukan bimbingan pekanan. Dengan begitu, penulis telah men-

dapatkan berbagai masukan dan pandangan yang berbeda dari beliau sendiri dan teman-teman saat itu. Selama mengikuti kelas, penulis telah banyak belajar dari beliau khususnya mengenai *computer vision* dan *image processing*. Bagi penulis, beliau termasuk salah satu dosen yang paling cakap dalam mengajar. Sebagai salah satu anggota laboratorium Kecerdasan Buatan dan Robotika, beliau telah menela-dani penulis untuk memiliki loyalitas dan kedisiplinan yang tinggi.

Juga terima kasih kepada bapak Wawan yang telah mempercayai penulis se-bagai salah satu anggota riset *smart classroom* yang beliau ketuai. Tanpa dukungan yang besar dari beliau, penulis tidak akan memiliki kesempatan untuk dapat berma-lam di kampus dan menuntaskan proyek akhir robotika.

Penulis merasa sangat bersyukur telah diberikan kesempatan dan kepercayaan dalam mengajar sebagai asisten dari ibu Rani Megasari di kelas Basis Data, bapak Yudi Wibisono di kelas Sistem Basis Data, bapak Herbert Siregar di kelas Pem-rograman Visual dan bapak Eddy Prasetyo Nugroho di kelas Rekayasa Perangkat Lunak. Terima kasih juga kepada teman-teman yang telah menjadi partner mengajar yang kompeten. Dengan mengajar, penulis tidak hanya mendapatkan pengetahuan yang lebih luas dan mendalam mengenai bahan ajar yang disampaikan, namun ju-ga mendapatkan keluasan untuk meningkatkan kemampuan mengajar dan berbicara di depan kelas. Penulis menjadi mengerti bahwa menempatkan diri bukan sebagai pengajar, akan tetapi sebagai partner bagi para studen, penting dilakukan dalam mengajar di kelas. Hubungan emosional yang baik sedikit banyak mempengaruhi motivasi mereka dalam mengikuti kelas.

Penulis merasa sangat beruntung telah memiliki dosen-dosen yang istimewa. Sangat menyenangkan mendengarkan mereka saling menceritakan kisah inspiratif satu sama lain. Selain yang sudah disebutkan di atas, penulis ingin berterima ka-sih lebih khusus kepada ibu Rosa Ariani Sukamto yang telah mengajarkan penulis untuk tidak perlu menjadi orang lain, untuk selalu jujur dalam berlaku dan untuk selalu berjuang juga tidak malu dalam belajar. Juga kepada bapak Yudi yang telah

mengajarkan penulis untuk tidak takut berbuat kesalahan dalam belajar dan untuk tidak berhenti belajar sebelum mampu menghasilkan buah karya. Juga kepada bapak Herbert Siregar yang telah mengajarkan penulis untuk selalu belajar memahami sesuatu secara mendasar dan untuk selalu memiliki etos juga etika dalam bekerja. Juga kepada bapak Eddy yang telah mengajarkan penulis untuk selalu menilai sesuatu secara adil juga lugas. Juga kepada ibu Enjun yang telah mengajarkan penulis untuk selalu mengutamakan disiplin dalam berlaku dan untuk selalu menghargai usaha orang lain. Juga kepada dosen-dosen lain yang tidak dapat disebutkan satu per satu. Namun, satu hal yang dapat dipastikan bahwa penulis telah belajar banyak sekali makna dari mereka semua. Jika diperbolehkan, penulis ingin selalu dapat duduk setidaknya satu kali lagi di hadapan mereka untuk mendengarkan dan mencatat beberapa pelajaran terakhir.

Tidak lupa, penulis juga ingin mengucapkan terima kasih kepada teman-teman yang telah memberikan warna dan menyulap setiap kebersamaan kami dalam belajar di kelas menjadi sangat menyenangkan dan tidak akan pernah tergantikan. Yang telah membuat kenangan dalam ruang-ruang kelas tidak akan pernah sama jika tanpa mereka. Secara istimewa, penulis ingin berterima kasih sebanyak-banyaknya kepada Muhammad Faris Muzakki dan Yahya Firdaus yang telah menjadi partner dalam banyak pekerjaan. Juga kepada Ammar Ashshiddiqi, Reyhan Fikri Dzikriansyah, Teguh Arianto,, Adnan Khairi As., Asep Saepul Ahmad, I. G. N. Agung A. A. W., dan Genta Satria A. P. sebagai teman-teman terdekat penulis dalam menjalani kehidupan di kampus. Juga kepada kakak-kakak tingkat yang telah bersedia menjawab dan memandu penulis dengan begitu tulus dan tanpa pamrih dalam belajar.

Setiap momen yang penulis habiskan bersama teman-teman seperjuangan adalah menyenangkan dan tidak akan tergantikan. Di sisi lain, setiap momen duduk mencatat dan mengacungkan tangan bertanya mengenai setiap pelajaran yang disampaikan oleh dosen-dosen yang berdedikasi juga tidak akan terlupakan. Bersama dengan mereka semua, kami telah saling berbagi banyak pengetahuan dan cerita.

Tanpa henti-hentinya, penulis juga bersyukur telah diberikan keluarga yang selalu menjadi orang-orang yang paling pertama dan paling setia dalam mendukung setiap keputusan penulis. Mereka adalah abi Ahmad Djunaedi Sastradinata, umi Desy Rosika Natalia, Ahmad Faaiz Al-Auza'i, Salma Kaisan Syauqi dan Kaisa Rifqa Ghassani. Tanpa doa dan dukungan dari mereka semua, penulis tidak akan mampu berdiri dan melangkah di atas kaki sendiri menuju perjalanan yang penuh dengan kebahagiaan.

Terakhir, penulis ingin berterima kasih kepada bapak Lala Septem Riza selaku Ketua Departemen Pendidikan Ilmu Komputer Universitas Pendidikan Indonesia, ibu Rani selaku Ketua Program Studi Ilmu Komputer Universitas Pendidikan Indonesia serta semua dosen penguji proposal juga laporan akhir skripsi ini.

Demikian pengantar ini dibuat dengan sungguh-sungguh. Penulis berharap bahwa pekerjaan ini dapat bermanfaat bagi penulis sendiri dan seluruh pembaca budiman. *At last but not least*, penulis menyatakan secara terbuka untuk menerima segala masukan dalam menyempurnakan skripsi ini.

Bandung, Agustus 2020

Penulis

PENGENALAN EMOSI MANUSIA MENGGUNAKAN *LOG-GABOR*
CONVOLUTIONAL NETWORKS MELALUI PENDEKATAN *FACIAL REGION*
SEGMENTATION

Oleh
Naufan Rusyda Faikar
1607645

ABSTRAK

Pengenalan emosi manusia secara otomatis dapat bermanfaat pada sektor-sektor terkait komputasi afektif. Penelitian ini merupakan penelitian pertama yang mengadopsi teknik *facial region segmentation* (FRS) pada arsitektur *Log-Gabor Convolutional Networks* (Log-GCNs) dalam membangun model menggunakan set data gambar wajah nonfrontal, FER-2013. Dengan menggunakan deteksi *facial landmark*, daerah fitur wajah tertentu dapat disegmentasi menjadi dua-tiga bagian. Setiap bagian dapat dilatih baik secara individu maupun bersamaan menggunakan teknik *network ensemble*, di mana sejumlah arsitektur GCN yang identik tergabung di dalamnya. Hasil eksperimen membuktikan bahwa Log-GCN dengan FRS berhasil mengungguli *baseline* dengan augmentasi data melalui peningkatan akurasi sebesar 6,07%.

Kata kunci: Rekognisi emosi; rekognisi ekspresi wajah; FER; segmentasi daerah wajah; *deep convolutional neural network*; jaringan ansambel.

*HUMAN EMOTION RECOGNITION USING LOG-GABOR CONVOLUTIONAL
NETWORKS THROUGH FACIAL REGION SEGMENTATION APPROACH*

*Arranged by
Naufan Rusyda Faikar
1607645*

ABSTRACT

Automatic recognition of human emotions can be useful in sectors related to affective computing. We believe that it is the first study to adopt facial region segmentation (FRS) techniques on the Log-Gabor Convolutional Networks (Log-GCNs) architecture in order to build a model that using the non-frontal face dataset images, FER-2013. By using facial landmarks detection, certain facial feature areas can be segmented into two-three parts. Each region can be trained either individually or together using network ensemble techniques, where a number of identical GCN architectures are combined. The experimental results prove that Log-GCN with FRS successfully outperformed the baseline with data augmentation through an increase in accuracy of 6.07%.

Keywords: Emotion recognition; facial expression recognition; FER; facial region segmentation; deep convolutional neural network; network ensemble.

DAFTAR ISI

HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	ix
ABSTRAK.....	x
DAFTAR ISI.....	xii
DAFTAR TABEL	xiv
DAFTAR GAMBAR	xviii
 BAB I PENDAHULUAN	 1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah.....	5
1.3 Rumusan Masalah.....	5
1.4 Tujuan Penelitian	5
1.5 Batasan Penelitian.....	6
1.6 Manfaat Penelitian	6
1.7 Struktur Organisasi Skripsi	7
 BAB II KAJIAN PUSTAKA	 9
2.1 Pengenalan Emosi.....	9
2.2 Penelitian Terkait.....	14
2.3 Filter Gabor.....	22
2.4 Filter Log-Gabor	25

2.5	<i>Machine Learning</i>	25
2.5.1	<i>Artificial Neural Network</i>	27
2.5.2	<i>Convolutional Neural Network</i>	33
2.5.3	<i>Gabor Convolutional Networks</i>	39
2.6	<i>Facial Region Segmentation</i>	39
BAB III METODE PENELITIAN		41
3.1	Perancangan Penelitian	41
3.1.1	Perumusan Masalah	41
3.1.2	Studi Literatur	42
3.1.3	Pengumpulan Data	42
3.1.4	Perancangan Metode	44
3.1.5	Pengembangan Metode	47
3.1.6	Evaluasi Sistem	48
3.2	Alat dan Bahan Penelitian	49
BAB IV TEMUAN DAN PEMBAHASAN		50
4.1	<i>Data Cleansing</i>	50
4.2	Praproses Data	51
4.3	Eksekusi Skenario Pemodelan	58
4.3.1	Implementasi <i>CNN Baseline</i>	58
4.3.2	Modifikasi <i>CNN Baseline</i> Menjadi GCN	63
4.3.3	Modifikasi GCN Menjadi <i>Ensemble GCNs</i>	67
4.3.4	Modifikasi Log-GCN Menjadi <i>Ensemble Log-GCNs</i>	75
4.4	Evaluasi Akhir	84
BAB V KESIMPULAN DAN SARAN		98
DAFTAR PUSTAKA		100

DAFTAR TABEL

Tabel 2.1	Penelitian Terkait.....	19
Tabel 3.1	Spesifikasi Perangkat Keras	49
Tabel 3.2	Spesifikasi Perangkat Lunak.....	49
Tabel 4.1	Perbandingan Teknik-Teknik <i>Facial Landmark Detection</i>	54
Tabel 4.2	Perbandingan Performa Model <i>Baseline</i> dengan dan tanpa Augmentasi Data	58
Tabel 4.3	Perbandingan Performa Model CNN <i>Baseline</i> dan GCN	64
Tabel 4.4	Perbandingan Performa Berbagai Kombinasi Model <i>Ensemble</i> GCNs	68
Tabel 4.5	Perbandingan Performa Model GCN dan Log-GCN	79
Tabel 4.6	Perbandingan Performa Berbagai Kombinasi Model <i>Ensemble</i> Log-GCNs	81
Tabel 4.7	Perbandingan Performa Prediksi Model CNN <i>Baseline</i> dan Log-GCN (EN + NM + ENM) pada Set Data Tes FER-2013	87

DAFTAR GAMBAR

Gambar 2.1	Pembagian Model-Model Komunikasi	11
Gambar 2.2	Perbedaan Rerata Pola Ekspresi Wajah Orang (a) Barat dan (b) Asia dalam Enam Kelas Emosi Berbeda	12
Gambar 2.3	Otot-Otot Wajah Terkait Ekspresi Wajah	13
Gambar 2.4	Diagram Klasifikasi Emosi	14
Gambar 2.5	Taksonomi Pengenalan Emosi	15
Gambar 2.6	Contoh <i>Facial Electromyography</i>	15
Gambar 2.7	Contoh <i>Facial Action Units</i>	16
Gambar 2.8	Pratinjau Set Data JAFFE	17
Gambar 2.9	Arsitektur CNN <i>Baseline</i>	20
Gambar 2.10	Segmentasi Delapan Belas Bagian Wajah	22
Gambar 2.11	Fungsi Gabor; Hasil Modulasi Sinyal Sinusoidal oleh Fungsi <i>Gaussian</i>	23
Gambar 2.12	Contoh Filter Gabor dan Hasil Konvolusi oleh Filter Gabor	24
Gambar 2.13	Struktur Neuron	27
Gambar 2.14	Model <i>Artificial Neural Network</i>	29
Gambar 2.15	Kasus-Kasus Konvolusi 2D	36
Gambar 2.16	Contoh-Contoh Hasil ZCA	37
Gambar 2.17	Contoh-Contoh Hasil Augmentasi Data	38
Gambar 2.18	Contoh-Contoh Hasil Random Erasing	39
Gambar 2.19	Modulasi dan Konvolusi GoFs di GCNs	39
Gambar 2.20	Arsitektur GCNs	40
Gambar 3.1	Alur Penelitian	41
Gambar 3.2	Pratinjau Set Data FER-2013	43
Gambar 3.3	Distribusi Set Data FER-2013 per Label	43

Gambar 3.4	Variasi Set Data FER-2013	45
Gambar 3.5	Rancangan Metode	46
Gambar 3.6	Subtahap Praproses Data.....	46
Gambar 3.7	Subtahap Eksekusi Empat Skenario Pemodelan	47
Gambar 4.1	Beberapa Contoh Data Gambar Rusak pada FER-2013.....	52
Gambar 4.2	Beberapa Contoh Hasil Implementasi ZCA terhadap FER-2013 .	53
Gambar 4.3	Beberapa Contoh Hasil Implementasi <i>Histogram Equalization</i> terhadap FER-2013	53
Gambar 4.4	Beberapa Contoh Hasil Deteksi <i>Facial Landmark</i> terhadap FER- 2013 Menggunakan FAN	55
Gambar 4.5	Beberapa Contoh Hasil Deteksi <i>Facial Landmark</i> terhadap FER- 2013 Menggunakan <i>RetinaFace</i> pada ResNet50	56
Gambar 4.6	Beberapa Contoh Hasil per Subtahap <i>Face Alignment</i>	56
Gambar 4.7	Beberapa Contoh Hasil per Subtahap <i>Facial Region Segmentation</i>	57
Gambar 4.8	Performa Model CNN <i>Baseline</i> tanpa Augmentasi Data.....	59
Gambar 4.9	Performa Model CNN <i>Baseline</i> dengan Augmentasi Data.....	59
Gambar 4.10	Perbandingan Performa Model dengan dan tanpa Augmentasi Data Per Kelas Emosi	60
Gambar 4.11	Beberapa Contoh Hasil Per Subtahap <i>Augmentasi Data</i>	61
Gambar 4.12	Beberapa Contoh Kemiripan Data Gambar Wajah Berlabel Emo- si <i>Disgust</i> (Kiri) dan <i>Angry</i> (Kanan) pada FER-2013	61
Gambar 4.13	Rerata Seluruh Data Gambar Wajah pada FER-2013 Per Label Emosi	62
Gambar 4.14	Beberapa Contoh Kemiripan Filter Gabor (Kiri) dengan Kernel dari Lapisan Konvolusi Ke-8 (Kanan Atas) dan Ke-9 (Kanan Bawah) pada Model CNN <i>Baseline</i>	63
Gambar 4.15	Performa Model GCN	65

Gambar 4.16 Perbandingan Performa Model CNN <i>Baseline</i> dan GCN Per Kelas Emosi	66
Gambar 4.17 Performa Model GCNs Menggunakan Fitur EN + NM	69
Gambar 4.17 Performa Model GCNs Menggunakan Fitur EN + NM (Lanjutan)	70
Gambar 4.18 Performa Model GCNs Menggunakan Fitur E + N + M	71
Gambar 4.18 Performa Model GCNs Menggunakan Fitur E + N + M (Lanjutan)	72
Gambar 4.19 Performa Model GCNs Menggunakan Fitur E + N	73
Gambar 4.20 Performa Model GCNs Menggunakan Fitur N + M	74
Gambar 4.21 Performa Model GCNs Menggunakan Fitur E + M	75
Gambar 4.22 Performa Model GCNs Menggunakan Fitur EN + NM (<i>Concat.</i>)	76
Gambar 4.23 Performa Model GCNs Menggunakan Fitur E + N + M (<i>Concat.</i>)	77
Gambar 4.24 Performa Model GCNs Menggunakan Fitur ENM	77
Gambar 4.25 Beberapa Contoh Perbedaan Filter log-Gabor (Kanan) dan Gabor (Kiri)	78
Gambar 4.26 Performa Model Log-GCN	79
Gambar 4.27 Perbandingan Performa Model GCN dan Log-GCN Per Kelas Emosi	80
Gambar 4.28 Performa Model Log-GCNs Menggunakan Fitur EN + NM	82
Gambar 4.28 Performa Model Log-GCNs Menggunakan Fitur EN + NM (Lanjutan)	83
Gambar 4.29 Performa Model Log-GCNs Menggunakan Fitur EN + NM (<i>Concat.</i>)	83
Gambar 4.30 Performa Model Log-GCNs Menggunakan Fitur EN + NM (<i>Concat.</i>)	84
Gambar 4.30 Performa Model Log-GCNs Menggunakan Fitur EN + NM + ENM	84
Gambar 4.31 Perbandingan Performa Tiap-Tiap Model Terbaik dari Seluruh Skenario Eksperimen	85

Gambar 4.32 <i>Confusion Matrix</i> Performa Model CNN <i>Baseline</i> dengan Augmentasi Data pada Set Data Wajah Penuh	86
Gambar 4.33 Beberapa Contoh Kesalahan Prediksi Emosi <i>Angry</i> oleh Model Terbaik (Label Prediksi Label Sebenarnya)	89
Gambar 4.34 Beberapa Contoh Kesalahan Prediksi Emosi <i>Fear</i> oleh Model Terbaik (Label Prediksi Label Sebenarnya)	90
Gambar 4.35 Beberapa Contoh Kesalahan Prediksi Emosi <i>Happy</i> oleh Model Terbaik (Label Prediksi Label Sebenarnya)	91
Gambar 4.36 Beberapa Contoh Kesalahan Prediksi Emosi <i>Sad</i> oleh Model Terbaik (Label Prediksi Label Sebenarnya)	92
Gambar 4.37 Beberapa Contoh Kesalahan Prediksi Emosi <i>Surprise</i> oleh Model Terbaik (Label Prediksi Label Sebenarnya)	93
Gambar 4.38 Beberapa Contoh Kesalahan Prediksi Emosi <i>Neutral</i> oleh Model Terbaik (Label Prediksi Label Sebenarnya)	94
Gambar 4.39 Beberapa Contoh Keberhasilan Prediksi Emosi <i>Angry</i> oleh Model Terbaik	95
Gambar 4.40 Beberapa Contoh Keberhasilan Prediksi Emosi <i>Fear</i> oleh Model Terbaik	95
Gambar 4.41 Beberapa Contoh Keberhasilan Prediksi Emosi <i>Happy</i> oleh Model Terbaik	96
Gambar 4.42 Beberapa Contoh Keberhasilan Prediksi Emosi <i>Sad</i> oleh Model Terbaik	96
Gambar 4.43 Beberapa Contoh Keberhasilan Prediksi Emosi <i>Surprise</i> oleh Model Terbaik	97
Gambar 4.44 Beberapa Contoh Keberhasilan Prediksi Emosi <i>Neutral</i> oleh Model Terbaik	97

BAB I

PENDAHULUAN

Pada bab ini disajikan latar belakang penelitian pengenalan emosi manusia meliputi jawaban atas pertanyaan-pertanyaan mengapa domain penelitian ini menarik dan penting untuk dilakukan, bagaimana peluang aplikasi dari hasil penelitian ini, bagaimana alur kerja pemodelan pengenalan emosi manusia, bagaimana pencapaian terkini dari penelitian terkait, bagaimana pendekatan yang diusulkan serta peluang keberhasilannya. Kemudian dipaparkan mengenai identifikasi masalah, rumusan masalah, tujuan penelitian, batasan penelitian dan manfaat penelitian yang relevan. Diakhiri dengan penjelasan struktur organisasi penulisan skripsi ini dari Bab I hingga Bab V dan bagian-bagian pelengkap lainnya.

1.1 Latar Belakang

Pengenalan emosi manusia sangat banyak manfaatnya, terutama pada sektor-sektor terkait komputasi afektif¹. Pada periklanan umpamanya, pengenalan emosi dapat diterapkan berdasarkan hubungan berbanding lurus antara kualitas respons emosional masyarakat terhadap iklan dengan peningkatan penjualan (Sujata dkk., 2018). Pada rekayasa perangkat lunak, pengenalan emosi dapat dimanfaatkan, baik untuk menilai efisiensi kerja dan kualitas kode tiap-tiap karyawan maupun untuk mengukur kepuasan pelanggan guna menggantikan penggunaan kuisioner yang menyedatkan (Kołakowska dkk., 2013, 2014). Beccue & Kaul (2018) mengidentifikasi setidaknya terdapat tujuh peluang pasar dalam penggunaan teknologi ini yang tetap relevan hingga 2025 mendatang, yaitu untuk diterapkan pada layanan pelanggan, penelitian produk/pasar, pengalaman pelanggan, perawatan kesehatan, pendidikan, otomotif, dan gim.

¹Komputasi afektif adalah studi multidisipliner pada pengembangan sistem yang mampu mengenali, memahami dan menghasilkan emosi manusia.

Pengenalan emosi memberikan kemampuan tiruan kepada sistem komputer untuk dapat menafsirkan bermacam-macam emosi manusia. Untuk mencapai tujuan tersebut, berbagai pendekatan komunikasi telah dilakukan, baik itu komunikasi verbal maupun nonverbal² (Garcia-Garcia dkk., 2017). Sejauh ini, pengenalan emosi melalui ekspresi wajah diklaim paling akurat. Sebab ekspresi wajah, dalam kasus terbanyak, memiliki kontribusi terbesar dalam komunikasi (Mehrabian & Wiener, 1967; Mehrabian & Ferris, 1967; Lapakko, 2007). Disertai asumsi bahwa sejumlah ekspresi wajah manusia bersifat universal³, pengenalan emosi sangat mungkin dilakukan melalui proses analisis ekspresi wajah (P. Ekman, 1970).

Pengenalan ekspresi wajah termasuk kajian multidisipliner (Mandal & Awasthi, 2014), yang tidak hanya bersifat teoretis —tentang bagaimana ekspresi wajah menyatakan emosi manusia—, melainkan juga bersifat teknis —tentang bagaimana komputer dapat meniru kemampuan emosi manusia— (P. Ekman, 1993; Gendron dkk., 2014; Jack dkk., 2012). Atas kemajuan teknologi yang kian canggih, perbalahan teoretis tidak lagi kentara. Meskipun manusia belum mampu sepenuhnya untuk menjelaskan setiap aspek di dalamnya, akan tetapi *machine learning* (Hebb, 1949) telah terbukti berhasil menyelesaikan masalah-masalah terkait rekognisi secara akurat. Tidak hanya dapat mengenali emosi manusia, teknologi rekognisi saat ini telah mampu mengenali emosi kucing (Evangelista dkk., 2019) dan anjing (Amici dkk., 2019). Sehingga penggunaan *machine learning* pada pengenalan ekspresi wajah mengetren dewasa ini.

Hingga saat ini, keberhasilan teknologi rekognisi sangat bergantung kepada kuantitas dan kualitas set data yang digunakan. Set data ekspresi wajah yang sangat bervariasi, meliputi variasi karakteristik subjek —seperti ras, etnis, jenis kelamin dan usia— dan potret —seperti kualitas gambar; rotasi, sudut dan jarak

²Komunikasi verbal melibatkan kata-kata, baik lisan maupun tulisan. Sebaliknya, komunikasi nonverbal melibatkan perilaku fisik (bahasa tubuh), meliputi ekspresi wajah, postur tubuh dan isyarat.

³Set emosi yang terdiri dari enam kelas emosi (*six basic emotions*), yang dianggap universal oleh P. Ekman (1970), meliputi *angry*, *disgust*, *fear*, *happy*, *sad*, dan *surprise*.

pengambilan— merupakan sebuah tantangan besar bagi penelitian pengenalan ekspresi wajah. Untuk menyeragamkan distribusi probabilitas set data ekspresi wajah, pengambilan set data sering kali dilakukan pada kondisi lingkungan yang sudah diatur sedemikian rupa. Pengambilan dengan cara demikian mengacu kepada pemotretan ekspresi wajah menggunakan sisi depan wajah untuk menghasilkan set data wajah frontal (Lucey dkk., 2010). Melalui cara ini, penelitian mutakhir pada pengenalan ekspresi wajah frontal mampu mencapai akurasi sempurna pada data wajah frontal (Zhou & Wang, 2019).-

Terlepas dari banyaknya usaha dan mahalanya biaya yang harus diberikan untuk mengakuisisi set data wajah frontal, penggunaan set data frontal memiliki beberapa kerugian yang berarti. Dalam banyak skenario praktis, model rekognisi yang dilatih menggunakan set data wajah frontal menjadi tidak relevan untuk penerapan dalam kondisi liar. Sebab data input aplikasi diperoleh di bawah pengaturan lingkungan dan peralatan yang berbeda. Pada sebagian besar kasus, pengambilan potret wajah subjek tidak dilakukan dalam posisi yang sejajar dengan sensor kamera. Contohnya pada aplikasi pendeteksi emosi pengemudi kendaraan bermotor, sensor kamera diletakkan pada sudut rendah (Assari & Rahmati, 2011). Sementara pada aplikasi pendeteksi emosi studen di kelas nyata, sensor kamera ditempatkan pada sudut tinggi (Jing dkk., 2020). Untuk alasan yang sama, pengembangan model akan menjadi sulit dilakukan melalui prosedur penambahan set data wajah frontal dari sumber yang berbeda (Yan dkk., 2018). Oleh karena itu, penelitian pengenalan ekspresi wajah nonfrontal menjadi urgen untuk dilaksanakan.

Pengembangan model pengenalan ekspresi wajah nonfrontal hingga saat ini belum mencapai kepuasan berarti. Melalui peningkatan kompleksitas arsitektur *Convolutional Neural Network* (CNN) (LeCun dkk., 1989) serta pemanfaatan ekstraksi fitur tradisional, model *state-of-the-art* terkini hanya mampu mencapai akurasi sebesar 75,42% (Georgescu dkk., 2019). Melalui pengurangan tingkat variasi set data *learning*, kinerja model pun dapat ditingkatkan. Misalnya melalui teknik

face frontalization yang baru-baru ini dikembangkan, di mana set data wajah non-frontal dibangkitkan dari set data wajah frontal (Lai & Lai, 2018). Akan tetapi teknik-teknik ini memerlukan komputasi yang relatif mahal.

Dengan mengetahui bahwa ekspresi wajah manusia terbentuk dari lebih dari sepuluh ribu kombinasi gerakan relatif lima belas otot bagian wajah (Ekman & Rosenberg, 1997; P. Ekman, 2004; Westbrook & Varacallo, 2019), di mana gerakan-gerakan tersebut menyebabkan perubahan bentuk-bentuk yang kasat mata pada kulit wajah, penerapan ekstraksi fitur tekstur gambar menjadi sangat relevan. Filter Gabor, yang mana juga dianggap mirip dengan sistem visual beberapa mamalia (Sivalingamaiah & Reddy, 2012), telah terbukti sangat cocok untuk kasus-kasus yang memerlukan proses analisis tekstur (Vijay & Indumathi, 2019; Mohammed, 2019). Dengan sifatnya yang kuat terhadap perubahan rotasi, distorsi dan variasi iluminasi pada sinyal gambar (Sisodia dkk., 2013), filter ini telah dimanfaatkan oleh banyak penelitian terdahulu pada pengenalan ekspresi wajah (Lyons dkk., 1998; Islam, Mahmud, & Hossain, 2018a; Qin dkk., 2020).

Pada penelitian ini, diajukan sebuah pendekatan baru pada set data wajah non-frontal, di mana daerah-daerah wajah disegmentasikan secara otomatis menggunakan algoritma tertentu untuk mempersempit *region of interest* (RoI) sebelum pada akhirnya dilatih menggunakan *Gabor Convolutional Networks* (GCNs) (Luan dkk., 2018). GCNs merupakan CNN yang dimasuki oleh filter Gabor (Gabor, 1946), yang mana filter ini dapat meningkatkan kemampuan CNN dalam mengekstraksi informasi tekstur terhadap perubahan orientasi dan skala. Tiap-tiap bagian wajah ini dilatih secara individu dan kemudian dihitung berdasarkan nilai akurasi per bagian untuk mendapatkan nilai akurasi akhir. Sementara itu, dilakukan eksperimen modifikasi terhadap GCNs, yaitu dengan menggantikan filter Gabor menjadi log-Gabor (Field, 1987a). Gagasan ini muncul setelah pembuktian bahwa filter log-Gabor lebih unggul daripada filter Gabor untuk tekstur yang kompleks (Nava dkk., 2011).

1.2 Identifikasi Masalah

Dari latar belakang di atas, penulis mengidentifikasi beberapa permasalahan spesifik yang diangkat di penelitian ini sebagai berikut.

1. Rancangan pemodelan pengenalan emosi melalui ekspresi wajah menggunakan *Gabor Convolutional Networks* (GCNs) melalui pendekatan *facial region segmentation* (FRS).
2. Analisis kinerja model menggunakan GCNs dengan dan tanpa melalui pendekatan FRS.
3. Analisis kinerja sebelum dan sesudah modifikasi instruksi pada GCNs melalui penggantian filter Gabor menjadi log-Gabor.

1.3 Rumusan Masalah

Untuk memperjelas permasalahan yang akan diteliti maka penulis merumuskan permasalahan tersebut sebagai berikut.

1. Bagaimana rancangan pemodelan pengenalan emosi melalui ekspresi wajah menggunakan GCNs melalui pendekatan FRS?
2. Bagaimana analisis kinerja model menggunakan GCNs dengan dan tanpa melalui pendekatan FRS?
3. Bagaimana analisis kinerja sebelum dan sesudah modifikasi instruksi pada GCNs melalui penggantian filter Gabor menjadi log-Gabor?

1.4 Tujuan Penelitian

Berdasarkan rumusan di atas, tujuan yang hendak dicapai dari penelitian ini terkumpul ke dalam poin-poin berikut ini.

1. Menemukan rancangan pemodelan pengenalan emosi melalui ekspresi wajah menggunakan GCNs melalui pendekatan FRS yang optimal.
2. Mengevaluasi analisis kinerja model menggunakan GCNs dengan dan tanpa melalui pendekatan FRS.
3. Mengevaluasi analisis kinerja sebelum dan sesudah modifikasi instruksi pada GCNs melalui penggantian filter Gabor menjadi log-Gabor.

1.5 Batasan Penelitian

Agar penelitian dapat dilaksanakan sesuai dengan bidang fokus permasalahan yang telah ditentukan, maka berikut ini adalah batasan penelitian ini.

1. Diasumsikan bahwa ekspresi wajah adalah representasi paling akurat dari emosi manusia.
2. Diasumsikan bahwa ekspresi wajah adalah universal untuk tujuh kelas emosi berbeda, yaitu *angry*, *disgust*, *fear*, *happy*, *sad*, *surprise*, dan *neutral*.
3. Digunakan set data berupa gambar-gambar dua dimensi (2D) dari pose wajah manusia, bukan berupa video atau gambar-gambar sekuensial terhadap waktu dari perubahan tampilan wajah manusia.

1.6 Manfaat Penelitian

Hasil analisis dari eksperimen ini diharapkan dapat memberikan manfaat-manfaat sebagai berikut.

1. Memberikan pengetahuan dan pengalaman pertama bagi penulis di penelitian pengenalan emosi otomatis melalui ekspresi wajah.
2. Memberikan kontribusi kepada masyarakat di penelitian pengenalan ekspresi wajah.

3. Memberikan dorongan kepada industri dalam pemanfaatan pengenalan ekspresi wajah sebagai salah satu usaha mencapai tujuan bisnis mereka.

1.7 Struktur Organisasi Skripsi

Struktur organisasi skripsi berisi perincian tentang urutan skripsi yang terorganisasi ke dalam lima bab, mulai dari Bab I hingga Bab V. Bab I memuat uraian tentang pendahuluan dan merupakan bab awal skripsi yang secara terurut memuat latar belakang penelitian, identifikasi masalah penelitian, rumusan masalah penelitian, tujuan penelitian, batasan masalah, manfaat penelitian, dan struktur organisasi skripsi. Bab I ini berperan besar dalam pengembangan bab-bab selanjutnya pada skripsi ini.

Bab II memuat kajian pustaka dari sumber-sumber kredibel yang membahas mengenai sejumlah teori, konsep dan gagasan terhadap studi kasus yang diangkat dari berbagai penelitian sebelumnya yang relevan dengan penelitian ini. Diawali dengan pembahasan mengenai studi pengenalan emosi dan kaitannya dengan ekspresi wajah. Dilanjutkan dengan pembahasan penelitian terkait yang menjadi baseline bagi penelitian ini. Dan diakhiri dengan pembahasan landasan teori seperti filter Gabor, filter log-Gabor GCNs dan FRS. Kajian pustaka ini memiliki peran yang sangat penting sebagai landasan teoritik dalam penyusunan rumusan masalah, tujuan serta hipotesis penelitian.

Bab III berisi penjabaran terperinci mengenai metode penelitian sebagai landasan ilmiah penelitian ini. Diuraikan secara komprehensif mengenai prosedur penelitian yang meliputi data sampel penelitian, desain, metode dan rancangan penelitian, instrumen penelitian serta teknik analisis data.

Bab IV menguraikan implementasi metodologi dan analisis hasil penelitian berikut pembahasannya dalam bentuk deskripsi naratif. Termasuk di dalamnya pembahasan mengenai perbandingan antara metodologi yang dikembangkan dengan penelitian sebelumnya. Data hasil penelitian ditampilkan dalam bentuk tabel dan grafik statistik serta alternatif bentuk relevan lainnya. Pembahasan keputusan

yang diambil dan hasil eksperimen pada penelitian ini terhadap penelitian terkait juga ikut disertakan pada bab ini.

Bab V merupakan bab terakhir dalam skripsi ini yang menyajikan kesimpulan dari penafsiran dan pemaknaan subjektif peneliti serta saran perbaikan terhadap hasil analisis temuan penelitian. Terdapat dua alternatif cara penyajian kesimpulan, yaitu dengan cara menyajikannya dalam bentuk butiran atau uraian padat.

Selain memuat sejumlah bab inti yang telah disebutkan di atas, skripsi ini dilengkapi dengan daftar pustaka sebagai sumber rujukan bagi pembaca budiman dengan sumber-sumber relevan dan kredibel sesuai dengan kriteria penulis.

BAB II

KAJIAN PUSTAKA

Pada bab ini disajikan penjelasan sejumlah teori, konsep dan gagasan terkait dengan domain penelitian pengenalan emosi manusia. Diawali dengan penjelasan lebih lengkap atas latar belakang yang disarikan dari sumber-sumber multidisiplin yang relevan dari bidang psikologi, sosial, biologi, statistika, dan lain-lain. Dilanjutkan dengan pemaparan penelitian-penelitian terkait meliputi alur berpikir dalam penentuan berbagai pendekatan yang diusulkan serta set data publik yang digunakan pada penelitian ini, perbandingan pendekatan serta hasil eksperimen dari sejumlah penelitian terkait mutakhir dan sedikit mengenai teknik-teknik yang diadopsi pada penelitian ini. Kemudian dipaparkan beberapa metode, teknik dan formula yang disoroti pada penelitian ini meliputi filter Gabor, filter Log-Gabor, GCNs dan FRS. Diakhiri dengan deskripsi singkat mengenai beberapa pustaka utama yang dipakai dalam pelaksanaan eksperimen.

2.1 Pengenalan Emosi

Emosi merupakan bagian dari komunikasi. Komunikasi berperan penting dalam interaksi sosial di beragam sistem sosial. Studi komunikasi merupakan studi multidisipliner yang melibatkan psikologi, sosiologi, antropologi dan linguistik (Mandal & Awasthi, 2014). Pada teknologi informasi, komunikasi spesifik dipelajari dalam interaksi manusia dan komputer (Cowie dkk., 2001; Fragopanagos & Taylor, 2005).

Komunikasi adalah untuk berbagi realitas bersama (Cobley, 2008). Membaca pikiran orang lain —atau membuat kesimpulan tentang keadaan pikiran orang lain— dalam komunikasi penting dilakukan dalam bertindak kooperatif. Ada banyak cara untuk setiap orang menyampaikan suatu makna. Sementara itu, sebuah kalimat dapat memiliki makna yang berbeda dalam konteks yang lain (Miller, 1999). Selain kemampuan memahami konteks, penting untuk seseorang memiliki

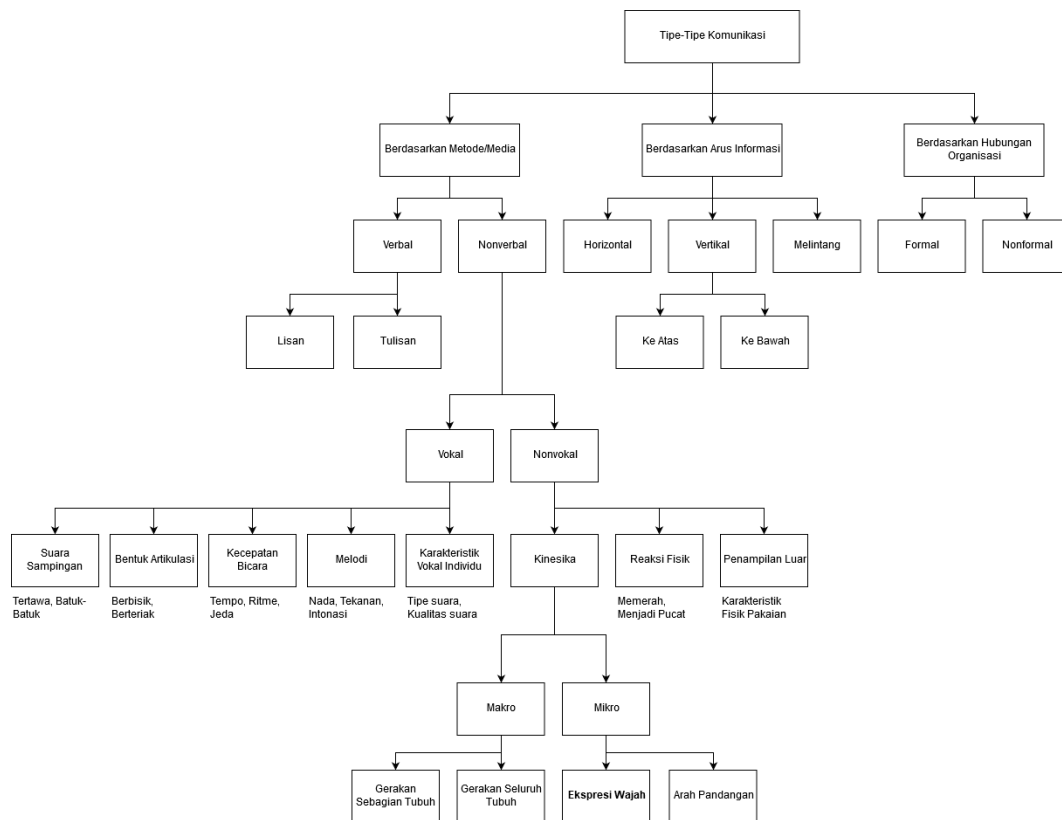
kemampuan mengetahui sifat dan latar belakang dari lawan bicara mereka dalam komunikasi. Sebab apa yang orang sampaikan mungkin berbeda dari apa yang mereka pikirkan. Lingkungan juga berkontribusi dalam mempengaruhi bagaimana orang berpikir (Entman, 1989).

Membaca pikiran yang tidak akurat dapat berdampak pada kebingungan, kesalahpahaman hingga berujung kepada konflik hidup dan mati. Penipuan adalah salah satu dampaknya, sehingga pengenalan tanda-tanda kebohongan menjadi masalah yang menantang (P. Ekman, 2009; Vrij dkk., 2019). Masalah komunikasi ini menjadi semakin rumit dalam komunikasi antar kultur. Di mana kompleksitas dan keanekaragaman dialek yang dimiliki oleh masing-masing bahasa menyulitkan dalam proses interpretasi dan penerjemahan ke dalam bahasa lain (Purnell, 2018).

Komunikasi terbagi menjadi tiga model (Gambar 2.1), yaitu berdasarkan metode atau media, arus informasi dan hubungan organisasi. Model kedua dan ketiga mengacu pada komunikasi di lingkungan organisasi fungsional (Postmes dkk., 2001; Elving, 2005; Lunenburg, 2010; Kandlousi dkk., 2010). Sedangkan model pertama lebih generik, mengacu pada komunikasi perseorangan di mana faktor emosi menjadi lebih diperhatikan.

Antara emosi, *mood* dan perasaan, memiliki definisi yang berbeda di kalangan para ahli. Beberapa di antara mereka membedakannya. Dalam membandingkan emosi dan *mood*, P. E. Ekman & Davidson (1994) menyatakan bahwa 1) *mood* berlangsung lebih lama, 2) *mood* lebih mudah terprovokasi, 3) *mood* lebih sulit untuk dimodulasi, 4) *mood* tidak memiliki ekspresi wajah unik sendiri, dan 5) *mood* lebih sulit disadari penyebabnya dibandingkan dengan emosi.

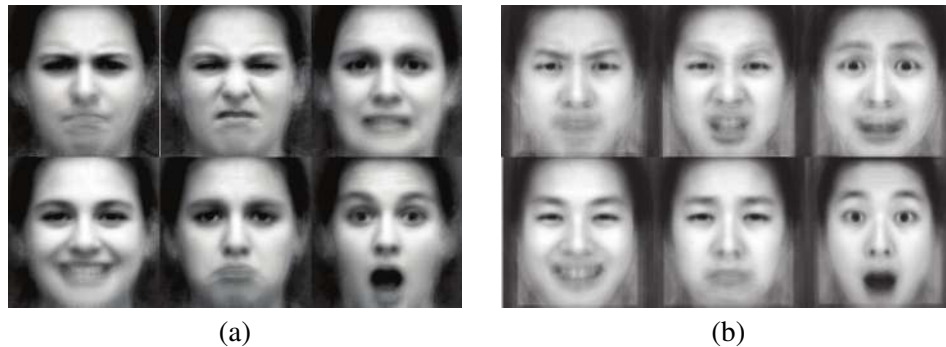
Dalam kaitan emosi dengan perasaan/sentimen, Freedman (2017) menyatakan bahwa perasaan merupakan campuran dari emosi dan bertahan lebih lama. Perasaan muncul ketika membiarkan emosi, yang biasanya hanya bertahan selama enam detik. Menurut Pérez (2018), sentimen akan tetap bertahan selama disadari dan dipikirkan. Berbeda dengan emosi, sentimen dikendalikan secara sadar oleh pikiran.



Gambar 2.1 Pembagian Model-Model Komunikasi (Surkamp, 2014;
www.businesscommunicationarticles.com)

Sementara sebagian ahli membedakan perasaan dengan sentimen (Munezero dkk., 2014). Meskipun konvensi dalam masalah ini belum terjadi, mereka bersepakat bahwa emosi bertahan dalam rentang waktu relatif lebih pendek.

Emosi merupakan sesuatu yang amat kompleks untuk dipahami bahkan secara mendalam. Emosi seseorang terbentuk atas berbagai faktor, seperti kepribadian (Arnold, 1960), bahasa (Lutz & Abu-Lughod, 1990), kemampuan adaptasi (Smith dkk., 1990), kultur (Kitayama & Markus, 1994), kondisi kesehatan (Pennebaker, 1995), kemampuan pengambilan keputusan (Schwarz, 2000), harapan (Kenny, 2003), motivasi (Bradley & Lang, 2007) dan lain-lain. Seseorang dapat memahami signifikansi dan nilai dari sebuah emosi jika dia pernah mengalaminya sendiri (Mun, 2019).



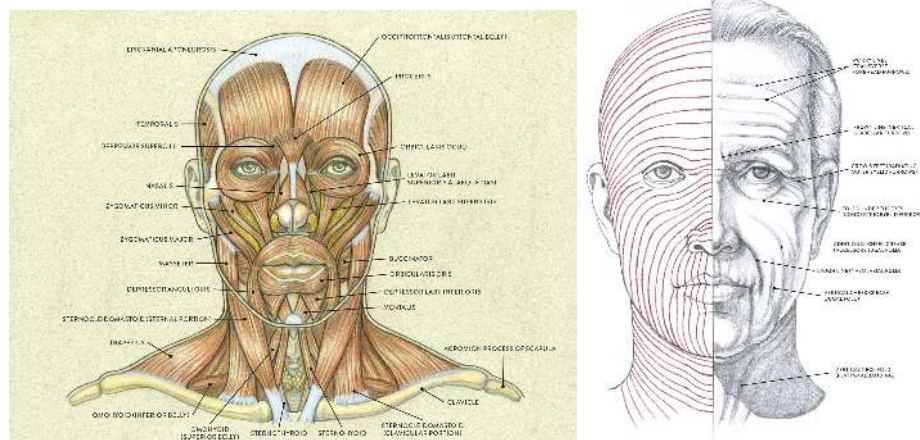
Gambar 2.2 Perbedaan Rerata Pola Ekspresi Wajah Orang (a) Barat dan (b) Asia dalam Enam Kelas Emosi Berbeda (Benitez-Garcia dkk., 2017)

Emosi merupakan sesuatu yang abstrak bagi orang lain. Akan tetapi, pada umumnya, emosi memiliki bentukan yang dapat diamati. Secara alami, emosi membentuk perilaku temporal seseorang (Baumeister dkk., 2007). Bentuk tersebut hanya diakui akurat jika hal itu bersifat evaluatif (Mun, 2019).

Pengenalan emosi dapat dilakukan dengan berbagai teknik seperti melalui pengenalan ekspresi wajah (Canedo & Neves, 2019), ucapan (Khalil dkk., 2019), tulisan (Seyeditabari dkk., 2018) atau sinyal-sinyal lain (Wagh & Vasanth, 2019). Masing-masing pendekatan tersebut dianggap representatif dalam pemodelan pengenalan emosi yang presisi dan akurat.

Pengenalan emosi melalui ekspresi wajah sendiri belum mampu mencerminkan emosi secara tepat dan utuh (Fernández-Dols & Crivelli, 2013). Namun sejauh ini, pengenalan emosi melalui ekspresi wajah dinilai paling akurat. Sebab sebagian besar komunikasi adalah bersifat nonverbal (Lapakko, 2007; Mehrabian & Wiener, 1967; Mehrabian & Ferris, 1967).

Keuniversalan emosi masih terus diperdebatkan hingga sekarang ini (Gendron dkk., 2014; Hall dkk., 2019; Jack dkk., 2012). P. Ekman (1970) berpendapat bahwa sejumlah ekspresi wajah (*six basic emotion*) bersifat universal, yaitu *angry*, *disgust*, *fear*, *happy*, *sad*, *surprise*, dan *neutral*. Sebagian besar ilmuwan yang memenuhi kualifikasi P. Ekman (2016) menyetujui adanya bukti kuat pada keuniversalan emo-

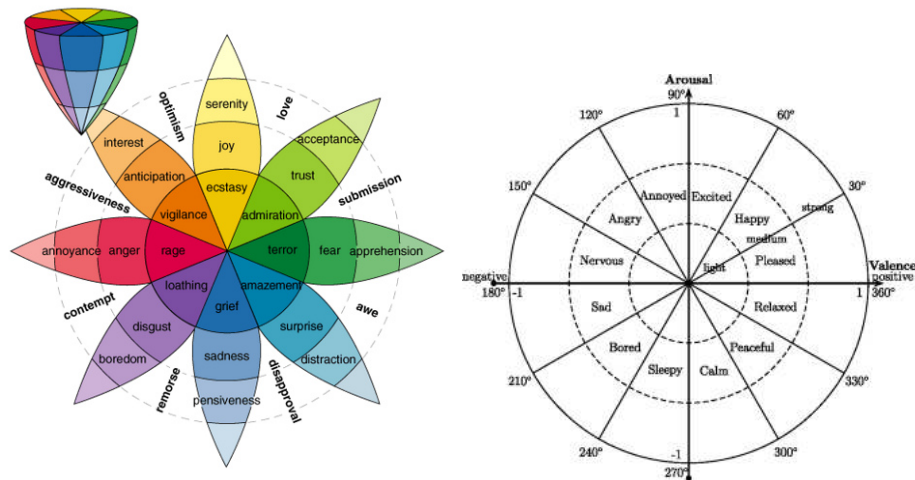


Gambar 2.3 Otot-Otot Wajah Terkait Ekspresi Wajah (Winslow, 2015)

si. Baru-baru ini terbukti bahwa *six basic emotion* tidak berlaku antar kultur Asia dan Barat (Benitez-Garcia dkk., 2017, 2018). Gambar 2.2 menunjukkan perbedaan yang cukup mendasar antara rerata pola ekspresi wajah dari kedua kultur tersebut.

Ekspresi wajah dapat dibedakan melalui lebih dari 10.000 kombinasi gerakan relatif lima belas otot bagian wajah (Gambar 2.3) (P. Ekman, 2004; Westbrook & Varacallo, 2019). Kerutan kulit wajah karena usia, rona merah wajah, arah lirik dan kedipan mata serta ukuran pupil juga terlibat dalam pengenalan ekspresi wajah (G. Guo dkk., 2013; Kret, 2015).

Emosi tidak biner atau terdefinisi dengan baik, melainkan setiap emosi saling terhubung dan mempengaruhi satu sama lain menurut kedekatannya (Paiva dkk., 2001). Klasifikasi emosi tidak hanya terbatas pada model diskrit yang menghasilkan lima, enam atau 23 jenis emosi berbeda. Melalui pendekatan dimensional, Plutchik & Kellerman (2013) memperkenalkan diagram emosi yang menghubungkan delapan jenis emosi dasar (Gambar 2.4a). Melalui pendekatan *fuzzy*, Kowalczyk & Czubenko (2016) memperkenalkan lingkaran emosi dengan dua belas jenis emosi dasar (Gambar 2.4b). Diagram pohon terstruktur Parrott (2001) bahkan mampu mendefinisikan hingga lebih dari seratus jenis emosi spesifik.

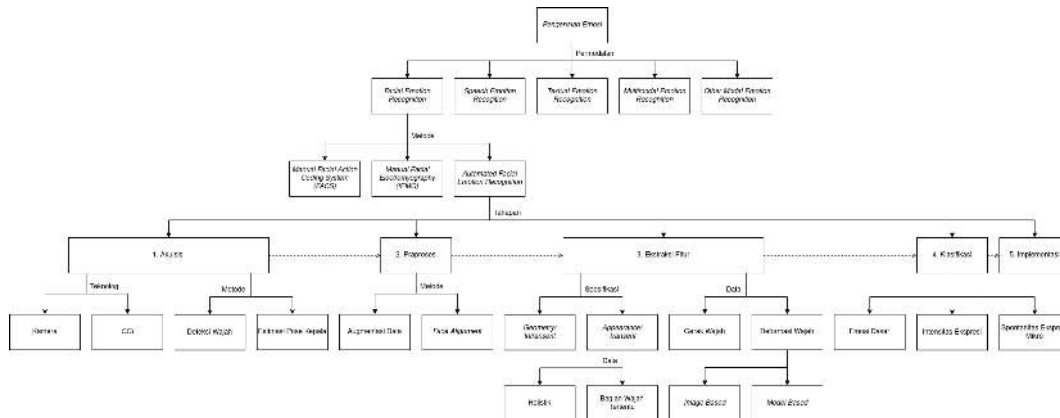


Gambar 2.4 Diagram Klasifikasi Emosi

2.2 Penelitian Terkait

Pengenalan emosi, terkait bidang ilmu komputer, memiliki domain penelitian yang sangat luas (Gambar 2.5). Berdasarkan jenis modal, pengenalan emosi dapat dilakukan baik melalui sinyal ucapan, tulisan, ekspresi wajah, sinyal digital —seperti sinyal haptik dari *drawing tablet* (Schrader & Kalyuga, 2020), sinyal aktifitas otak dari *electroencephalogram* (Wagh & Vasanth, 2019) dan sinyal tekanan dari *keyboard* (Lv dkk., 2008)— maupun kombinasi dari itu semua (Avots dkk., 2019; D. Li dkk., 2019; Mittal dkk., 2019).

Pengenalan emosi melalui ucapan dapat dilakukan baik dengan pendekatan linguistik (tentang apa yang diucapkan) maupun paralinguistik (tentang bagaimana mengucapkannya) dalam komunikasi verbal. Dengan teknik-teknik pemrosesan sinyal, karakteristik unik tiap-tiap emosi dapat dikenali. Namun, belum ada konsensus mengenai set fitur standar untuk pengenalan ini. Sehingga menyulitkan pengevaluasian akurasi model pengenalan emosi dalam ucapan (Hook dkk., 2019; Roy dkk., 2020). Bagaimanapun, hal ini merupakan tantangan terbesar bagi penelitian



Gambar 2.5 Taksonomi Pengenalan Emosi (Sumathi dkk., 2012)

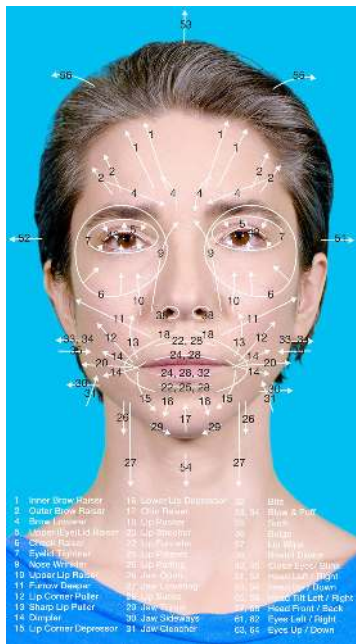
pengenalan emosi secara umum.

Pengenalan emosi melalui tulisan juga tergolong ke dalam kajian linguistik komputasi. Jika pengenalan emosi dalam ucapan cenderung meneliti mengenai gelombang bunyi dari wicara, maka pengenalan emosi dalam tulisan meneliti tentang tata bahasa yang terkandung dalam vektor-vektor kata baik dalam kumpulan frasa, kalimat maupun dokumen (Collobert dkk., 2011; Baali & Ghneim, 2019).

Pengenalan emosi melalui ekspresi wajah menurut Wolf (2015) terbagi menjadi tiga metode, yaitu pelacakan aktivitas elektromiografi wajah, pengkodean aksi wajah dan rekognisi ekspresi wajah otomatis. Pelacakan aktivitas elektromiografi wajah atau *facial electromyography* melibatkan penggunaan elektroda-elektroda yang dilekatkan pada permukaan wajah di titik-titik tertentu yang dianggap representatif (Gambar 2.6). Sayangnya metode ini tidak dapat digunakan dalam situasi sosial sebab kompleksitas teknisnya. Namun, untuk mendeteksi aktivitas yang halus pada otot-otot wajah, metode ini sangat dapat diandalkan. Kesulitan yang dihadapi adalah menempatkan elektroda-elektroda tersebut pada posisi yang benar-benar tepat.



Gambar 2.6 Contoh *Facial Electromyography* (Gibert dkk., 2010)



Gambar 2.7 Contoh *Facial Action Units*
(www.coralievogelaar.com)

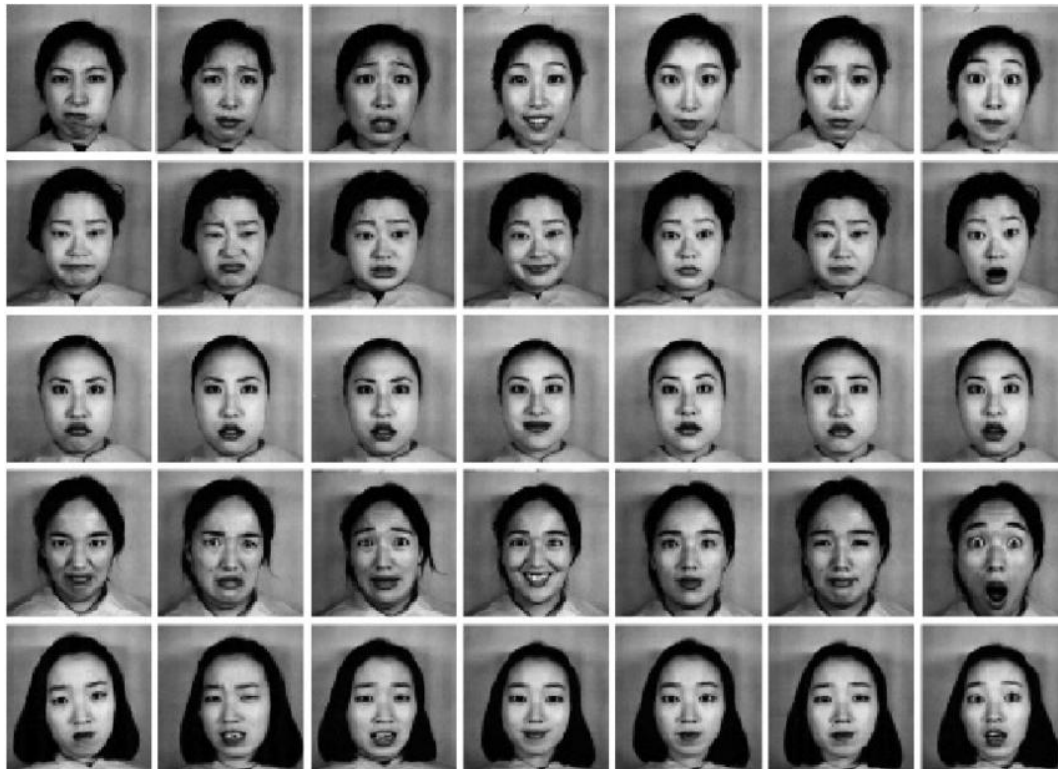
Pengkodean aksi wajah atau *facial action coding* (Ekman & Rosenberg, 1997) merupakan pengenalan emosi yang melibatkan analisis perubahan perilaku-perilaku otot-otot wajah yang kasat mata disebut sebagai unit aksi *action units*. Unit-unit aksi ini didefinisikan secara manual, di antaranya adalah bagian kiri, kanan, luar dan dalam untuk tiap-tiap alis, mata, hidung, bibir dan mulut (Gambar 2.7). Metode ini dapat meminimalkan bias oleh pengamat emosi. Namun proses analisis metode ini memerlukan bentuk emosi yang relatif kuat dan usaha manual yang banyak (Wolf, 2015). Dalam usaha meminimalkan waktu analisis pada perlakuan manual, pelacakan unit-unit aksi dikenali secara otomatis dengan pengukuran jarak relatif antar titik-titik wajah hasil

deteksi tengara wajah atau *facial landmarks* (Valstar & Pantic, 2006).

Pengenalan ekspresi wajah otomatis bekerja dengan cara menyandikan informasi ekspresi dari representasi wajah, baik informasi spasial dari tiap-tiap gambar statis maupun informasi hubungan temporal dari gambar-gambar sekuens, dengan bantuan *machine learning*. Teknik *machine learning* telah berkembang dari *shallow learning* (di mana pemerolehan fitur-fitur wajah masih dilakukan secara manual) menjadi *deep learning* (di mana pemerolehan fitur-fitur wajah diberikan seutuhnya kepada mesin). Teknik *deep learning* telah terbukti jauh mengungguli metode-metode tradisional (S. Li & Deng, 2018).

Pengenalan ekspresi wajah otomatis sangat bergantung pada kualitas dan kuantitas set data *learning*¹. Maka dari itu, usaha memperoleh set data yang bagus

¹Penggunaan istilah set data *learning* dalam konteks *machine learning* mengacu kepada pengeritian yang luas, yaitu mencakup set data *training*, *validation* dan *test*.



Gambar 2.8 Pratinjau Set Data JAFFE (Zhao & Zhang, 2011)

menjadi tantangan tersendiri. Namun kekhawatiran pada pemerolehan set data ini tidak diperlukan lagi, sebab saat ini telah banyak orang dermawan yang bersusah payah untuk melakukannya. Set data tersebut sudah dilabeli dan disediakan gratis secara publik; penulis secara khusus bersyukur kepada para dermawan itu. Tinjauan mendetail mengenai beragam set data publik untuk pengenalan ekspresi wajah dapat ditemukan di S. Li & Deng (2018).

Sebagian besar pengenalan ekspresi wajah otomatis memanfaatkan set data wajah frontal, contohnya JAFFE (Lyons dkk., 1998), CK+ (Lucey dkk., 2010) dan EmotioNet (Fabian Benitez-Quiroz dkk., 2016). Gambar 2.8 memperlihatkan beberapa contoh gambar wajah frontal dari JAFFE. Mereka telah berhasil mencapai performa akurasi lebih dari 95%. Model pengenalan ekspresi wajah melalui pendekatan teknik yang diusulkan oleh Zhou & Wang (2019) bahkan mencapai akurasi sebesar 100% untuk CK+. Namun, di samping kesulitan akuisisi data wajah fron-

tal, model yang dikembangkan dengan set data wajah frontal menjadi tidak relevan pada penerapan di kondisi alam liar (S. Li & Deng, 2018).

Pengenalan ekspresi wajah menggunakan set data wajah nonfrontal, contohnya FER-2013² (Goodfellow dkk., 2013), hingga sekarang belum mencapai kepuasan berarti. Hasil survei penulis menuturkan bahwa akurasi tertinggi yang berhasil diraih saat ini untuk FER-2013 adalah 75,42% (Georgescu dkk., 2019). Untuk menemukan metode optimal yang mampu meningkatkan akurasi model pengenalan ekspresi wajah nonfrontal, maka penulis melakukan survei pada penelitian-penelitian terbaru. Penulis membatasi survei tersebut pada praproses, ekstraksi fitur dan model CNN berikut performanya. Tabel 2.1 memberikan gambaran mengenai hasil penerapan berbagai usulan metode penelitian dalam pengenalan emosi otomatis melalui ekspresi wajah.

Perancangan desain model CNN untuk pengenalan ekspresi wajah menghadapi dilema dalam upaya menghasilkan model yang efektif (dalam konteks akurasi) dan efisien (dalam konteks kompleksitas). Sehubungan dengan hal ini, Agrawal & Mittal (2020) menjadi *baseline* bagi penelitian penulis, yang mana model pengenalan ekspresi wajah yang diusulkan dapat menghasilkan akurasi seperti manusia sebesar 65,23% (Goodfellow dkk., 2013) untuk FER-2013. Model yang diusulkan merupakan sebuah *Deep Convolutional Neural Networks* (DCNNs) murni dan sederhana —tidak menggunakan *dropout layer* dan *fully connected layer*—. Dari dua model yang diberikan, penulis memilih *model2* yang memiliki 464.183 parameter *training*. Sebenarnya, perbedaan dari kedua model hanya terletak pada jumlah filter pada masing-masing lapisan konvolusi. Gambaran singkat arsitektur model yang digunakan dapat dilihat pada Gambar 2.9.

Network ensemble adalah model jaringan yang terdiri beberapa jaringan. Model ini telah terbukti mengungguli model jaringan tunggal. Implementasi model

²Set data FER-2013 merupakan set data wajah nonfrontal publik yang paling populer digunakan dalam banyak penelitian pengenalan ekspresi wajah. Sejauh pengetahuan terbaik penulis, belum ada set data wajah nonfrontal lain yang tersedia secara publik.

Tabel 2.1
Penelitian Terkait

Referensi	Basis Data	Tipe Jaringan	Seleksi Data	Praproses	Ekstraksi Fitur	Classifier	Performa (%)
Georgescu dkk., 2019	FER-2013	CNN, <i>NE</i>	k-NN	<i>DA</i>	CNNs + BOVW	SVM	75,42
Agrawal & Mittal, 2020		CNN	80% set <i>training</i> , 20% set <i>test</i>		CNN (16; 0,46M)*		65,23
Engin dkk., 2018			28.709 set <i>training</i> , 3.589 set <i>validation</i> , 3.589 set <i>test</i>	<i>PN</i>	VGG-face		67,60
Kim dkk., 2016		<i>DA + FA + IN</i>		CNN (5; 2,4M)*		73,73	
Pramerdorfer & Kampel, 2016				CNN (10/16/33; 1,8M/1,2M/5,3M)*		75,20	
Z. Zhang dkk., 2015		CNN, <i>MN</i>		<i>FA</i>	CNN (6; 21,3M)*		75,10
Devries dkk., 2014				<i>IN</i>	CNN (4; 12M)*		67,21
Qin dkk., 2020	CK+	CNN, <i>NE</i>	327 sekuens	<i>FA + IN</i>	Filter Gabor + CNNs		96,81
Adil dkk., 2019				<i>FRS</i>	Filter Gabor	SVM	92,19
Ye dkk., 2019					CNNs		98,70
Majumder dkk., 2016					CNN + LBP	SOM	98,95

DA—Data Augmentation; FA—Face Alignment; FRS—Facial Region Segmentation IN—Illumination Normalization;

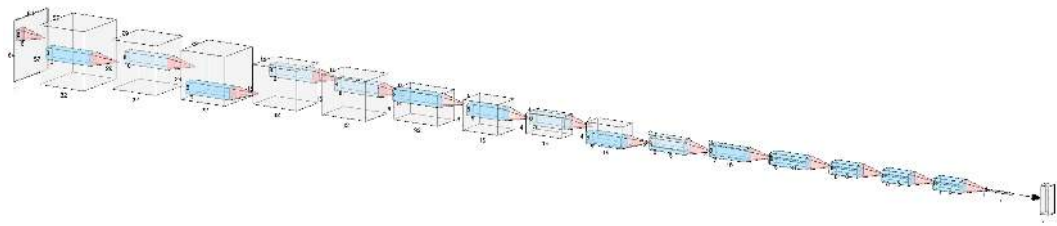
PN—Pose Normalization; NE—Network Ensemble; MN—Multitask Network, CN—Cascaded Network;

*(Jumlah total lapisan jaringan; jumlah total parameter *training*).

ini mengharuskan tiap-tiap jaringan memiliki keragaman yang cukup dan didukung oleh metode pemaduan yang tepat, yaitu antara lain:

1. *Majority voting*: kelas data prediksi ditentukan oleh suara terbanyak dari tiap-tiap jaringan;
2. *Simple average*: kelas data prediksi ditentukan oleh perhitungan rerata probabilitas dengan bobot yang sama dari tiap-tiap jaringan;
3. *Weighted average*: kelas data prediksi ditentukan oleh perhitungan rerata probabilitas dengan bobot yang berbeda dari tiap-tiap jaringan (S. Li & Deng, 2018).

Model ini diterapkan pada pengenalan ekspresi wajah oleh Kim dkk. (2016) dan Pramerdorfer & Kampel (2016) demi mendapatkan akurasi sebesar 73,73% dan 75,2% untuk FER-2013.



Gambar 2.9 Arsitektur CNN *Baseline* (Agrawal & Mittal, 2020)

Multitask network adalah model jaringan yang diusulkan untuk mentransfer pengetahuan dari tugas-tugas relevan dan mengurai faktor-faktor gangguan. Model ini telah digunakan oleh Devries dkk. (2014) dan Z. Zhang dkk. (2015) pada pengenalan ekspresi wajah demi mendapatkan akurasi sebesar 67,21% dan 75,10% untuk FER-2013.

Cascaded network adalah model jaringan yang memadukan berbagai modul untuk tugas berbeda secara sekuensial. Modul-modul tersebut masing-masing berkontribusi pada penyaringan faktor-faktor variasi berbeda yang tidak dibutuhkan pada pengenalan ekspresi wajah (S. Li & Deng, 2018).

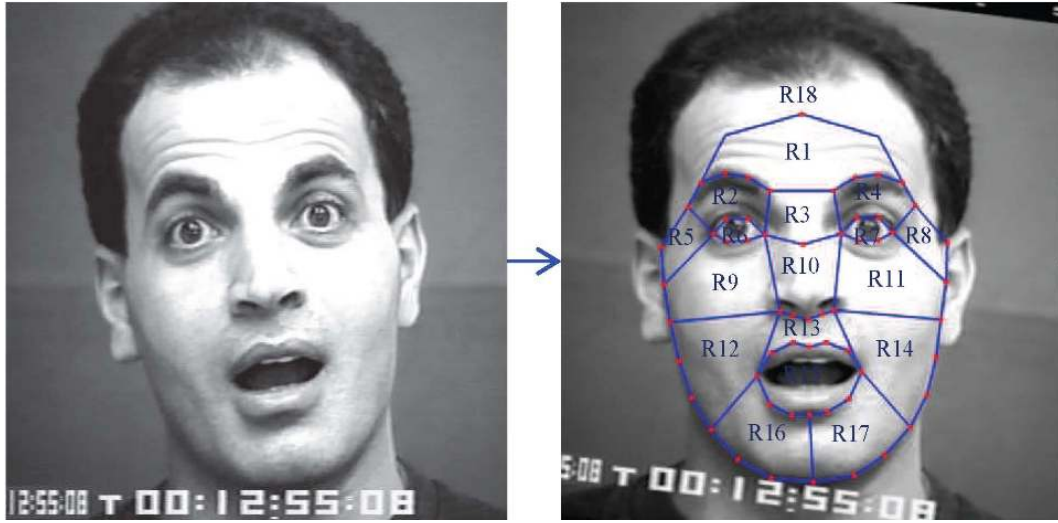
Generative Adversarial Network (GAN) adalah model jaringan khusus untuk membangkitkan gambar-gambar sintesis yang realistis. Pada pengenalan ekspresi wajah, model ini digunakan oleh Lai & Lai (2018) untuk membangkitkan gambar wajah frontal dari set data wajah nonfrontal. Penjelasan mendetail tentang model-model jaringan ini dapat ditemukan di S. Li & Deng (2018).

Strategi memadukan CNN dan ekstraksi fitur konvensional belakangan ini telah banyak dilakukan. Strategi ini terbukti dapat meningkatkan akurasi model pengenalan ekspresi wajah. Georgescu dkk. (2019), melalui usaha memadukan fitur-fitur yang diperoleh secara otomatis dan manual, saat ini menjadi *state-of-the-art* bagi pengenalan ekspresi wajah untuk FER-2013. Eksperimen dimulai dengan pemilihan sampel data menggunakan *k-Nearest Neighbor* (k-NN) (Fix, 1951). Secara manual, fitur-fitur diperoleh menggunakan model *bag-of-visual-word* (BOVW) (Ionescu dkk., 2013) yang terdiri dari tahap ekstraksi fitur menggunakan *Scale-*

Invariant Feature Transform (SIFT) (Lowe, 2004) dan tahap *clustering* menggunakan *k-means* (Leung & Malik, 2001). Secara otomatis, ekstraksi fitur dilakukan menggunakan tiga model CNN, yaitu VGG-13 (Barsoum dkk., 2016), VGG-f (Chatfield dkk., 2014) dan VGG-face (Parkhi dkk., 2015). Setelah itu, vektor representasi fitur dari CNN dan BOVW dirangkai, dinormalisasikan dan dilatih menggunakan model *Dense-Sparse-Dense* (DSD) (Han dkk., 2016) dengan *Support Vector Machine* (SVM) (Cortes & Vapnik, 1995) sebagai *classifier*.

Facial region segmentation merupakan salah satu teknik *cropping*, yang mana memisahkan beberapa daerah fitur gambar wajah sehingga menghasilkan set data baru (Gambar 2.10). *Facial region segmentation* telah diterapkan untuk set data wajah frontal karena kelengkapan fitur-fiturnya. Pada Majumder dkk. (2016), *facial region segmentation* digunakan untuk memisahkan daerah mata kiri, mata kanan, hidung dan mulut sehingga menghasilkan set data baru. Set data baru tersebut diduplikasi dan dikenakan *local binary pattern* (LBP) (Ojala dkk., 1996). Kemudian set data baru sebelum dan setelah perkenaan LBP dilatih menggunakan *network ensemble* dengan *self-organizing map* (SOM) *classifier* (Kohonen, 1990). Model tersebut dilatih pada CK+ demi menghasilkan akurasi sebesar 98,95%. Pada Ye dkk. (2019), bagian-bagian wajah yang disegmentasi adalah dahi (pertengahan daerah kedua mata), kedua mata dan mulut. Bagian-bagian tersebut dilatih menggunakan *network ensemble* secara individu dan menghasilkan akurasi sebesar 98,70%.

Filter Gabor merupakan salah satu filter konvolusi 2D yang populer digunakan dalam ekstraksi fitur berbagai tugas klasifikasi gambar menggunakan CNN. Pada Qin dkk. (2020), pengenalan sejumlah filter Gabor pada set data input CK+ dibuktikan mampu meningkatkan kemampuan generalisasi CNN dua kanal dan mencapai akurasi sebesar 96,81%. Pada Adil dkk. (2019), penggunaan filter Gabor pada SVM *classifier* menghasilkan akurasi sebesar 92,19% untuk set data yang sama.



Gambar 2.10 Segmentasi Delapan Belas Bagian Wajah (Ghimire dkk., 2015)

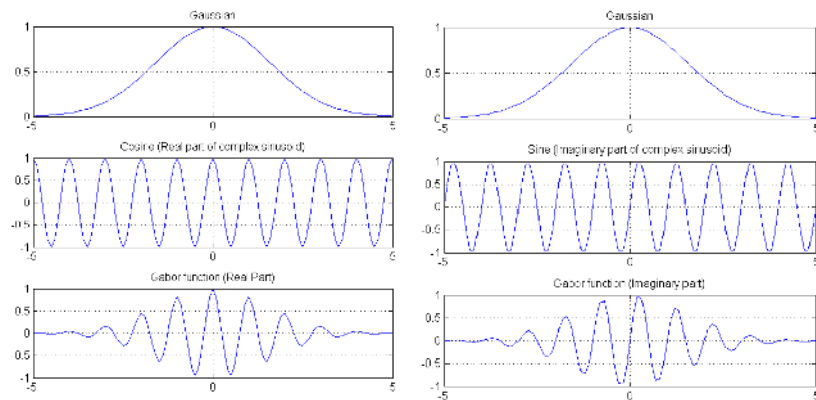
2.3 Filter Gabor

Filter Gabor adalah sebuah filter linear dari fungsi Gabor yang telah berhasil diterapkan di berbagai tugas pemrosesan gambar (Dunn & Higgins, 1993), misalnya untuk sistem autentikasi biometrik (Aggrawal, 2016; Vijay & Indumathi, 2019), deteksi kendaraan spesifik (Sara dkk., 2019) dan identifikasi tulisan tangan (Baati & Kanoun, 2018; Mohammed, 2019).

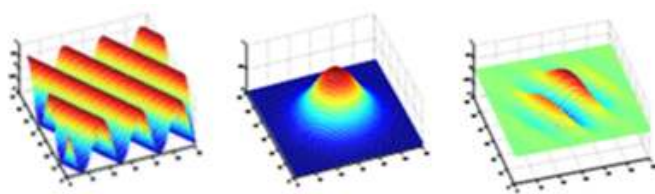
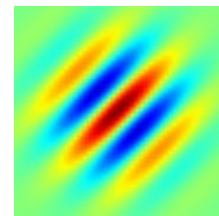
Fungsi Gabor awalnya diperuntukkan dalam analisis sinyal saluran komunikasi, sebagai fungsi frekuensi terhadap waktu, guna mendapatkan esensi informasi (Gabor, 1946). Fungsi ini diperluas menjadi fungsi 2D oleh Daugman (1985), sebagai sebuah sinyal sinusoidal kompleks yang dimodulasi oleh fungsi Gaussian (Gambar 2.11), yang dinyatakan (Grigorescu & Petkov, 2006) dalam (2.1) dan (2.2),

$$g(x, y, \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(\frac{-x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \exp\left(2\pi \frac{x'}{\lambda} + \psi\right) \quad (2.1)$$

$$b = \log_2 \left(\frac{\frac{\sigma}{\lambda} \pi + \sqrt{\frac{\ln 2}{2}}}{\frac{\sigma}{\lambda} \pi - \sqrt{\frac{\ln 2}{2}}} \right), \frac{\sigma}{\lambda} = \frac{1}{\pi} \sqrt{\frac{\ln 2}{2}} \left(\frac{2^b + 1}{2^b - 1} \right) \quad (2.2)$$



(a) Fungsi Gabor pada Dimensi Dua (Viswanathan dkk., 2014)

(b) Fungsi Gabor pada Dimensi Tiga
(<https://medium.com>)(c) Filter Gabor
(<https://think-piece.tistory.com>)Gambar 2.11 Fungsi Gabor; Hasil Modulasi Sinyal Sinusoidal oleh Fungsi *Gaussian*

dimana,

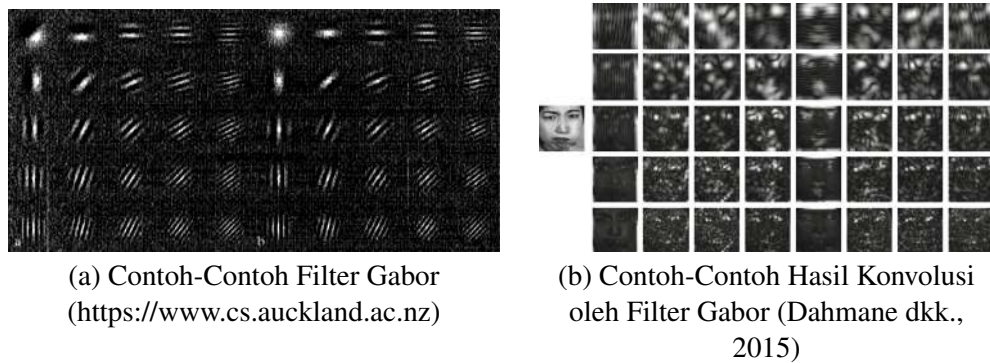
$$x' = x \cos(\theta) + y \sin(\theta),$$

$$y' = -x \sin(\theta) + y \cos(\theta),$$

λ = panjang gelombang faktor cosinus dari fungsi Gabor dengan nilai bilangan riil yang valid 2–256 piksel,

θ = orientasi normal terhadap garis paralel fungsi Gabor dengan nilai bilangan riil yang valid 0–80°,

ψ = fase *offset* faktor kosinus dari fungsi Gabor dengan nilai bilangan riil yang valid -180° hingga 180°,



Gambar 2.12 Contoh Filter Gabor dan Hasil Konvolusi oleh Filter Gabor

- σ = standar deviasi faktor *Gaussian* yang nilainya ditentukan oleh parameter λ dan b ,
- γ = rasio aspek spasial atau eliptisitas faktor *Gaussian* dengan nilai terletak antara 0,23 dan 0,92,
- b = *bandwidth* frekuensi spasial dari filter dengan nilai terletak antara 0,4 dan 2,5 oktaf.

Dalam pemrosesan gambar, pertama-tama, bank filter Gabor (Gambar 2.12a) dibuat melalui konfigurasi unik berbagai parameter dalam (2.1) dan (2.2). Selanjutnya konvolusi dilakukan melalui perkalian dot matriks antara m filter Gabor dengan n input gambar sehingga menghasilkan $m \times n$ *feature maps* (Gambar 2.12b).

Representasi frekuensi dan orientasi dari filter Gabor ditemukan mirip dengan sistem visual beberapa mamalia, sehingga filter ini sangat cocok untuk analisis tekstur (Sivalingamaiah & Reddy, 2012). Oleh karena itu, komputasi transformasi kompleks sebab tidak ortogonal menjadi kelemahan utama filter ini. Sifatnya yang demikian menimbulkan redundansi koefisien dan informasi (Vyas & Rege, 2006). Sehingga tidak cocok digunakan pada sistem waktu nyata. Hal ini memotivasi berbagai penelitian untuk mencoba melakukan optimasi tanpa kehilangan akurasi yang berarti (Ilonen dkk., 2005; Amayeh dkk., 2009; Kugaevskikh, 2017).

2.4 Filter Log-Gabor

Nava dkk. (2011) menyebutkan bahwa fungsi log-Gabor (Field, 1987b) dapat mengungguli performa fungsi Gabor pada pengenalan gambar yang memiliki tekstur yang kompleks. Dengan asumsi bahwa ekspresi wajah terbentuk dari kombinasi berbagai otot wajah yang menghasilkan tekstur kulit wajah yang kompleks, fungsi log-Gabor diharapkan dapat meningkatkan performa model rekognisi emosi. Fungsi log-Gabor dapat dirumuskan (Walia & Verma, 2016) ke dalam (2.3),

$$g(r, \theta) = \exp \left(\frac{-\log^2 \left(\frac{r}{f_0} \right)}{2\sigma_r^2} \right) \exp \left(\frac{-(\theta - \theta_0)^2}{2\sigma_\theta^2} \right), \quad (2.3)$$

$$f_0 = \text{wav} \times \text{scaleFactor}^n \quad (2.4)$$

dimana,

(r, θ)	= koordinat polar
f_0	= frekuensi pusat filter
wav	= panjang gelombang dari skala terkecil
scaleFactor	= faktor skala antar filter terurut
θ	= sudut orientasi filter
σ_r	= skala <i>bandwidth</i>
σ_θ	= sudut <i>bandwidth</i>

2.5 Machine Learning

Learning adalah fenomena —yang disimpulkan dari perilaku— yang dapat diamati di seluruh makhluk hidup, termasuk manusia, hewan (Beran, 2020) dan tanaman

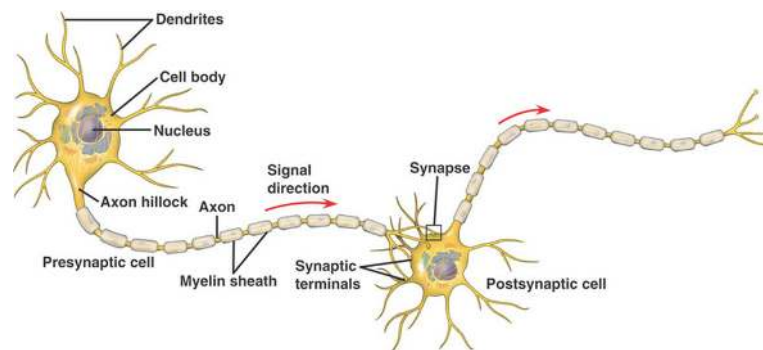
(Parise dkk., 2020), baik saat mereka terjaga maupun saat tidur. *Learning* dapat terjadi sekalipun tanpa keterlibatan dari orang lain (Gross, 2015). *Learning* berarti perubahan potensi perilaku yang relatif permanen akibat keteraturan di lingkungan pelaku (De Houwer dkk., 2013; Haselgrove, 2016).

Hewan dan tanaman belajar melalui hubungan antara peristiwa dan respons yang berdekatan (*associative learning*). Sebagai contoh, seekor kucing yang mana selalu diberi makan oleh pemiliknya setiap kali mengeong di dapur belajar jika mengeong adalah cara untuk meminta dan mendapatkan makanan (Ramos, 2009). Contoh lain, tanaman putri malu akan mengatup sedikit dalam menanggapi sentuhan ringan dan akan mengatup rapat dalam menanggapi stimulus taktil. Setelah beberapa kali dikenakan sentuhan ringan dan stimulus taktil, maka tanaman putri malu akan mulai mengatup rapat dalam menanggapi sentuhan ringan (Abramson & Chicas-Mosier, 2016).

Machine learning (Hebb, 1949) muncul akibat keinginan untuk mengenakan *learning* kepada mesin, sehingga dia mampu berinteraksi dengan lingkungannya berdasarkan pengetahuan yang diperoleh dari pengalaman. Untuk itu, berbagai disiplin terkait dipelajari seperti statistika, model interaksi sel-sel otak, teori kontrol adaptif, model-model psikologi, kecerdasan buatan, dan model-model evolusi (Nilsson, 2005). Akan tetapi, bagaimanapun, hingga saat ini, konsep-konsep dan teknik-teknik *machine learning* lebih banyak diperoleh dari statistika daripada disiplin lain (Mitchell, 2006).

Sebuah mesin dikatakan belajar jika sistem secara andal meningkatkan kinerja (yang diukur oleh metrik tertentu) secara otomatis pada tugas tertentu mengikuti pengalaman tertentu. Mesin tersebut harus mampu bekerja menurut arsitektur komputasi dan algoritma tertentu dalam memproses set data statistik untuk aplikasi dimana (Nilsson, 2005; Mitchell, 2006):

1. Beberapa tugas memerlukan algoritma yang rumit untuk didesain secara manual. Atau dengan kata lain, hanya dapat ditentukan oleh pasangan input dan



Gambar 2.13 Struktur Neuron (<https://mikerbio.weebly.com>)

keluaran. Sebagai contoh, tidak ada seorang pun mampu menulis algoritma untuk melabeli foto-foto yang memuat dirinya, namun melakukannya adalah mudah bagi semua orang.

2. Korelasi antar data tersembunyi dalam himpunan data yang besar dan kompleks. Sebagai contoh, pada kasus penipuan kartu kredit, akan sangat sulit bagi orang untuk menemukan pola penipuan secara manual dari data transaksi bank.
3. Lingkungan operasional tidak terdefinisi secara komplet pada waktu desain. Sebagai contoh, toko buku daring yang mampu menyesuaikan diri dengan preferensi pembelian setiap akun pelanggan.

Dengan syarat-syarat di atas, pemanfaatan *machine learning* menjadi sangat luas dan beragam. *Machine learning* dapat diterapkan dengan bijak hampir di setiap lini kehidupan manusia.

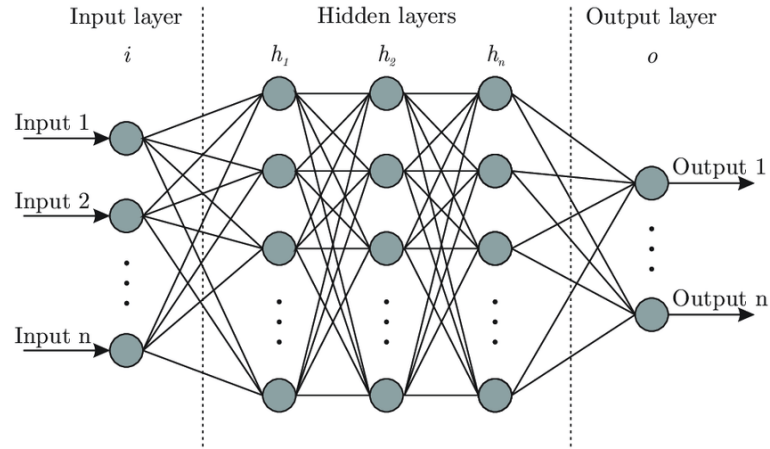
2.5.1 Artificial Neural Network

Otak manusia tersusun atas miliaran neuron yang saling berhubungan. Masing-masing neuron memiliki sel tubuh (soma), sejumlah dendrit dan sejumlah akson (Gambar 2.13). Dendrit berfungsi untuk menerima sinyal dari akson lain pada neuron tetangga. Sinyal tersebut akan ditransmisikan melalui dari akson ke dendrit

yang terkoneksi melalui sinapsis jika kekuatan sinyal tersebut di atas ambang batas tertentu. Melalui penyebaran dan interaksi sinyal-sinyal ini dalam anatomi neuron yang rumit, di mana melibatkan mekanisme nonlinear pada berbagai skala temporal dan spasial, pemrosesan informasi di otak berlangsung (Hines & Carnevale, 1997; Singh dkk., 2001).

Artificial Neural Network (ANN) mengacu pada model matematika yang memiliki arsitektur terdistribusi, yaitu terdiri atas soma (dapat dianalogikan) sebagai simpul/unit komputasi, sejumlah dendrit sebagai input, sejumlah akson sebagai keluaran dan sinapsis sebagai penyimpan informasi berupa bobot/parameter. Setiap simpul memiliki sebuah fungsi tertentu yang disebut *perceptron* (Singh dkk., 2001; An dkk., 2017). Secara umum, jaringan ini dapat digambarkan sebagai graf terarah yang terdiri dari lapisan input, lapisan tersembunyi dan lapisan target. Lapisan input terdiri dari i buah simpul ditambah dengan sebuah simpul konstan yang selalu menghasilkan nilai bobot sama dengan satu. Di mana lapisan tersembunyi dapat terdiri dari satu atau lebih lapisan yang memiliki sejumlah simpul yang saling terhubung. Setiap lapisan tersembunyi dapat memiliki jumlah simpul yang beragam ditambah sebuah simpul bias yang tidak memiliki input dan menghasilkan sebuah konstanta. Kedalaman jaringan diukur dengan menjumlahkan banyak lapisan tersembunyi dan lapisan target. Besar jaringan diukur dengan menjumlahkan seluruh simpul pada setiap lapisan jaringan (Shalev-Shwartz & Ben-David, 2014).

Salah satu bentuk dari jaringan ini adalah *feedforward multilayer perceptron*, di mana *feedforward* berarti aliran data bergerak satu arah, dari input ke target/keluaran. Gambar 2.14 memperlihatkan jaringan *feedforward* yang memiliki sejumlah input, tiga lapisan tersembunyi dan sejumlah target. Misalkan (x_1, x_2, \dots, x_i) merupakan vektor input dan (y_1, y_2, \dots, y_m) merupakan vektor target. Setiap simpul input terkoneksi dengan setiap simpul pada lapisan tersembunyi pertama dan setiap koneksi ini memiliki sebuah bobot $w_{i,j}$. Maka simpul ke- j pada lapisan tersembunyi pertama, melalui fungsi aktivasi/transfer nonlinear tanh, melewati hasil



Gambar 2.14 Model *Artificial Neural Network* (Bre dkk., 2018)

penjumlahan bobot dari setiap inputnya, yaitu

$$p_j = \tanh \left(\sum_i w_{i,j} x_i \right) \quad (2.5)$$

Melalui proses yang sama, keluaran setiap simpul pada lapisan tersembunyi kedua dan ketiga berturut-turut adalah

$$q_k = \tanh \left(\sum_j w_{j,k} x_j \right) \quad (2.6)$$

$$r_l = \tanh \left(\sum_k w_{k,l} x_k \right) \quad (2.7)$$

Akhirnya, lapisan target melakukan penjumlahan sederhana dari setiap inputnya dengan

$$y_m = \sum_l w_{l,m} x_l \quad (2.8)$$

Fungsi tanh dapat digantikan oleh fungsi nonlinear lain, misalnya fungsi sigmoid: $1/(1 + \exp[-\sum wx])$. Kedua fungsi ini memetakan rentang input yang tidak terbatas ke rentang keluaran terbatas (Singh dkk., 2001).

Supervised learning melatih jaringan dengan menginterpolasi data pelatihan

untuk memberikan generalisasi hubungan antar set data input dan target. Sebab hampir tidak mungkin untuk memperoleh hubungan yang konvergen di hampir setiap kasus. Interpolasi dilakukan melalui penyesuaian setiap nilai bobot sebanyak n *epoch* atau hingga mencapai model yang optimal. Model optimal diperoleh dengan meminimalkan fungsi error untuk semua bobot jaringan. Di antara fungsi tersebut adalah fungsi error *sum-of-squares*, yaitu (Singh dkk., 2001)

$$e^{\{z\}} = \frac{1}{2} \sum_m \beta (y_m^{\{z\}} - T_m^{\{z\}})^2 \quad (2.9)$$

dimana,

z = input vektor tunggal

T_m = nilai keluaran target

m = indeks simpul target

β_m = nilai prioritas bobot

Backpropagation (Rumelhart dkk., 1986) merupakan salah satu algoritma populer untuk melatih jaringan dengan menentukan gradien fungsi error tiap-tiap bobot. Bobot-bobot ini mulanya diinisialisasi dengan nilai acak kecil. Dalam kalkulus, gradien garis singgung sebuah fungsi nonlinear di sebuah titik dapat diperoleh melalui substitusi nilai koordinat titik tersebut ke dalam derivatif/turunan parsial fungsi nonlinear tersebut (Shalev-Shwartz & Ben-David, 2014). Pada lapisan bobot target, dengan mensubstitusikan (2.8) ke dalam (2.9), gradien fungsi error setiap bobot diperoleh dari

$$\frac{\partial e}{\partial w_{k,l}} = \sum_m \beta (y_m - T_m) \frac{\partial y_m}{\partial w_{k,l}} = \frac{\partial r_l}{\partial w_{k,l}} \sum_m \beta (y_m - T_m) w_{l,m} \quad (2.10)$$

Proses perhitungan ini berlanjut dari lapisan bobot target ke lapisan-lapisan bobot tersembunyi sehingga memberikan vektor gradien error komplit, $\partial e / \partial w$, di mana w

adalah set seluruh bobot jaringan. Evaluasi vektor gradien eror ini dapat dilakukan melalui proses *gradien descent*, yaitu dengan menyesuaikan vektor bobot ke arah negatif terhadap vektor gradien, contohnya

$$\Delta w = -\mu \frac{\partial e}{\partial w} \quad (2.11)$$

di mana μ adalah besaran langkah yang diambil yang umumnya bernilai 0,1–1,0. Gradien ini dihitung ulang menggunakan nilai bobot yang baru sebanyak n *epoch* atau hingga mencapai nilai eror mendekati nol. Nilai bobot yang baru akan diperbarui setiap kali vektor dari set data pelatihan melewati sebuah simpul. Untuk mengurangi waktu yang diperlukan dalam memperbarui setiap nilai bobot, pelatihan *batch* dipraktikkan, yaitu dengan memperbarui setiap nilai bobot hanya setelah mendapatkan rerata eror untuk seluruh vektor tersebut,

$$E = \frac{1}{Z} \sum_{z=1}^{z=Z} e^{\{z\}} \quad (2.12)$$

dan vektor gradien eror yang bersesuaian (Singh dkk., 2001).

Dalam konteks permasalahan pada penelitian ini, penulis menggunakan *loss function* yaitu *categorical crossentropy* yang cocok untuk klasifikasi multikelas (jumlah kelas > 2) berlabel tunggal pada kasus ini. Untuk itu, aktivasi *softmax* (Grave dkk., 2016) digunakan pada lapisan terakhir pada jaringan. Fungsi *softmax* didefinisikan oleh (2.13),

$$f(z_w) = \frac{\exp(z_w)}{\sum_{w' \in V} \exp(z_{w'})} \quad (2.13)$$

dimana,

z_w = representasi kelas yang dihitung

z'_w = representasi kelas selain yang dihitung

w = kelas yang dihitung

w' = kelas selain yang dihitung

V = himpunan kelas

Softmax bekerja dengan menentukan probabilitas sebuah kasus untuk setiap kelas prediksi. Nilai probabilitas tersebut berkisar antara 0 dan 1. Jika semua probabilitas dijumlahkan maka akan sama dengan 1.

Sementara itu, *learning rate* diperoleh menggunakan *Adaptive Momentum* (Adam), yang merupakan peningkatan dari metode *Gradient Descent* standar, sebagai *optimizer*. Adam bekerja dengan menghitung *learning rate* yang adaptif untuk setiap parameter (Prilianti dkk., 2019). Adam dipilih bukan berarti paling optimal, melainkan karena lebih mudah dikonfigurasi (Schneider dkk., 2019). Adam dapat diformulasikan ke dalam (2.14)–(2.18),

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad (2.14)$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (2.15)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t \quad (2.16)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (2.17)$$

$$\theta_{t-1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t + \epsilon}} \hat{m}_t \quad (2.18)$$

dimana,

\hat{v}_t = gradien kuadrat sebelumnya dengan bias yang dikoreksi

\hat{m}_t = gradien rata-rata sebelumnya dengan bias yang dikoreksi

v_t, m_t = keduanya memperkirakan peluruhan momen pertama dan kedua dari gradien yang dihitung

β = konstanta peluruhan

θ_{t+1} = Parameter pembaruan untuk $t + 1$

2.5.2 *Convolutional Neural Network*

Pengenalan tulisan tangan merupakan sebuah permasalahan klasik pada topik irisan antara *machine learning* dan *computer vision*, di mana pengaplikasian ANN untuk kasus ini menghadapi dua tantangan besar. Secara sederhana, ANN dapat menerima set data input berlabel manapun selama itu valid dan representatif untuk kemudian melatih model yang secara umum mampu mengenali bentuk tulisan apapun. Dalam kasus ini, set data input yang digunakan adalah berupa set matriks dua dimensi yang mengandung setiap piksel dari sebuah data gambar. Sesuai dengan sifat dari matriks, transformasi geometri tertentu dimungkinkan terjadi sehingga akan memperbanyak kondisi yang perlu ditangani oleh jaringan. Sebagai akibatnya, set data *learning* yang diperlukan menjadi sangat banyak dan variatif. Untuk kasus pengenalan yang lebih rumit, pengumpulan data akan menjadi semakin mustahil dilakukan. Dalam konteks pemrosesan gambar, setiap piksel merupakan representasi dari makna yang terkandung dalam sebuah gambar. Oleh karena itu, setiap piksel dari setiap data gambar input akan dimasukkan sebagai inputan jaringan. Semakin besar ukuran piksel dan semakin banyak data gambar input akan membuat proses pelatihan model memerlukan semakin banyak waktu dan semakin besar memori.

Untuk menjawab tantangan di atas, LeCun dkk. (1989) memperkenalkan sebuah jaringan spesial yang secara statistik terbukti lebih andal dalam memproses set data gambar, yaitu *Convolutional Neural Network* (CNN). Sesuai dengan sebutannya, jaringan ini mengharuskan terjadinya proses konvolusi minimal di salah satu lapisan tersembunyi. Secara umum, konvolusi adalah sebuah operasi perka-

lian matematika terhadap dua buah matriks dengan aturan dan parameter tertentu. Konvolusi 1D dapat didefinisikan oleh,

$$y(t) = \int x(a)w(t-a) da \quad (2.19)$$

di mana t adalah indeks waktu dan a adalah umur pengukuran. Konvolusi biasanya dinotasikan dengan lambang bintang (*asterisk*):

$$y(t) = (x * w)(t) \quad (2.20)$$

di mana fungsi x adalah input, fungsi w adalah bobot/kernel dan fungsi y adalah peta fitur/*feature map*. Sementara waktu t diasumsikan dengan nilai integer, maka konvolusi diskrit didefinisikan sebagai,

$$y(t) = (x * w)(t) = \sum_{a=-\infty}^{\infty} x(1)w(t-a) \quad (2.21)$$

Input di sini pada umumnya adalah larik data multidimensi, sedangkan kernel adalah larik bobot multidimensi. Larik multidimensi ini sering disebut sebagai tensor. Konvolusi 2D dapat ditulis sebagai,

$$Y(i, j) = (I * K)(i, j) = \sum_m \sum_n I(m, n)K(i-m, j-n) \quad (2.22)$$

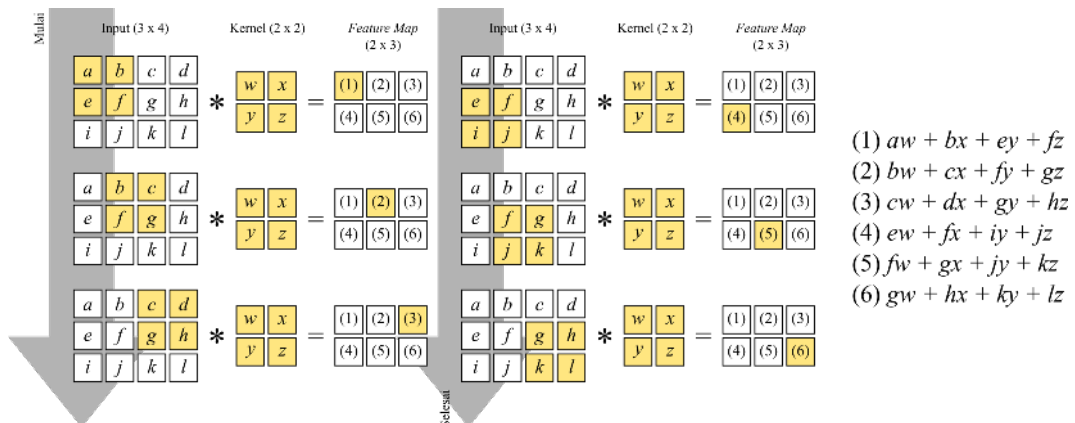
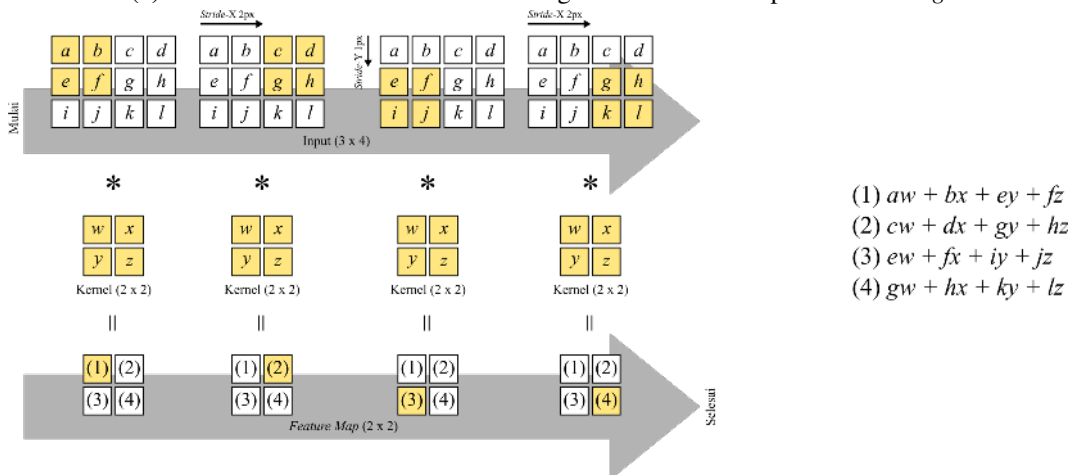
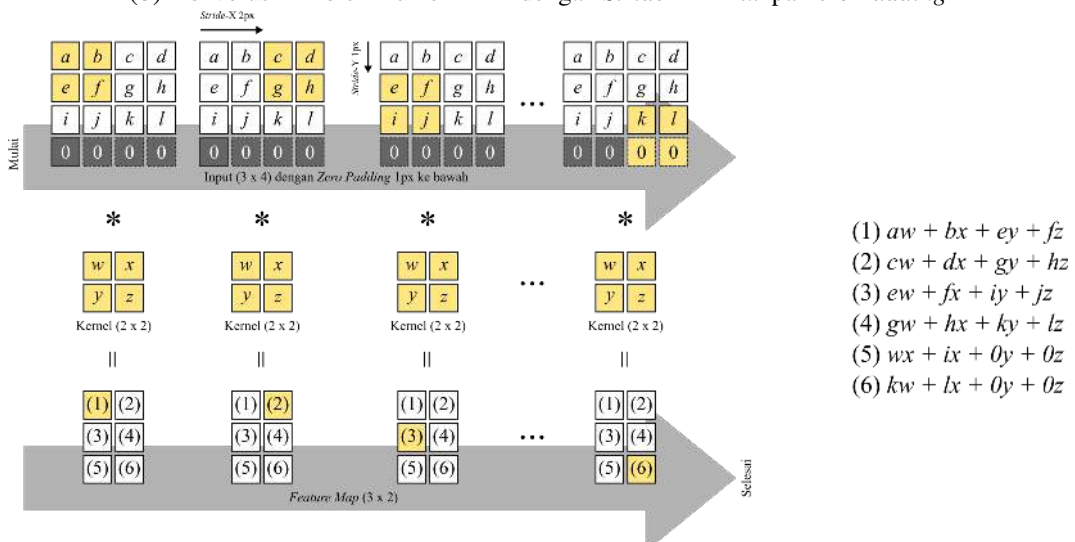
di mana I adalah larik gambar 2D dan K adalah larik kernel 2D. Konvolusi ini bersifat komutatif akibat pembalikan larik kernel guna mendapatkan pembuktian. Namun, hal ini tidak penting dalam implementasi ANN. Sehingga banyak pustaka ANN mengimplementasikan *cross-correlation*, konvolusi tanpa membalikkan kernel, dan menyebutnya sebagai konvolusi (Goodfellow dkk., 2013):

$$Y(i, j) = (K * I)(i, j) = \sum_m \sum_n K(i+m, j+n)I(m, n) \quad (2.23)$$

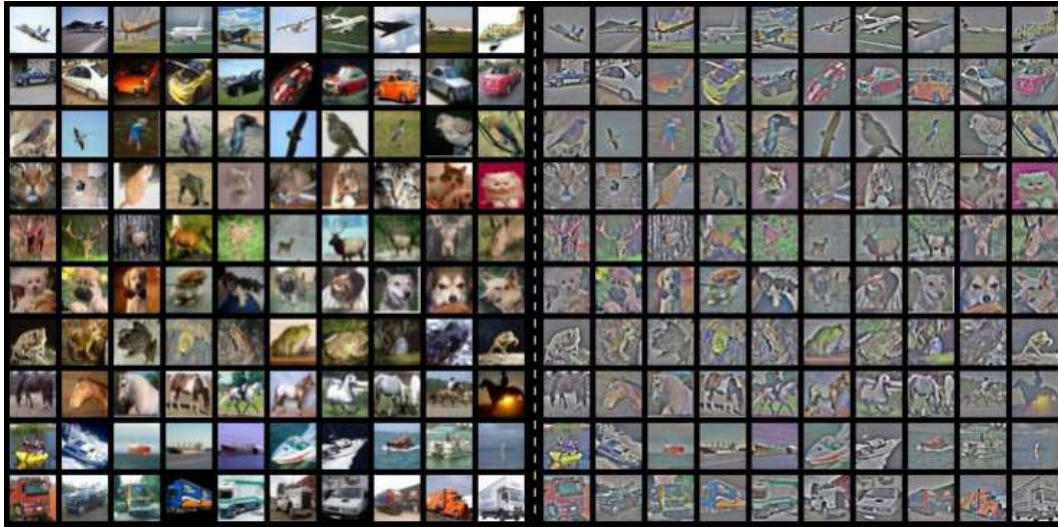
Gambar 2.15a memperlihatkan contoh konvolusi input $I(3, 4)$ oleh kernel $K(2, 2)$ menghasilkan *feature map* $F(2, 3)$. CNN biasanya memiliki interaksi atau konektivitas *sparse* ketika ukuran kernel lebih kecil daripada ukuran input. Tujuannya adalah mengurangi jumlah memori yang dibutuhkan untuk menyimpan parameter-parameter dari fitur-fitur gambar yang berarti. Jika ukuran kernel diperbesar, maka ukuran *feature map* akan lebih kecil. Dan begitu pula sebaliknya. Di lapangan, konvolusi memiliki parameter tambahan yang disebut sebagai *stride*, yang mana parameter ini mengatur seberapa besar pergeseran operasi perkalian oleh kernel terhadap input. *Stride* yang berukuran lebih besar akan menghasilkan *feature map* yang lebih kecil. Gambar 2.15b menunjukkan bagaimana ukuran *stride* yang berbeda, dari contoh pada Gambar 2.15a, dapat mempengaruhi pemetaan dari input ke *feature map* melalui proses konvolusi. Pemetaan dari input ke *feature map* tanpa penambahan apapun terhadap input sering disebut sebagai konvolusi yang valid. Kita juga dapat mengendarai ukuran *feature map* dengan melakukan zero padding pada input, di mana input diperluas dengan menambahkan aksis tertentu yang semua pikselnya bernilai nol seperti yang ditunjukkan pada Gambar 2.15c (Goodfellow dkk., 2016).

Praproses data merupakan tahap yang cukup penting untuk pemodelan klasifikasi gambar. Praproses harus dilakukan secara hati-hati melalui analisis karakteristik data inputan. Pemilihan praproses yang tepat menghasilkan tahap ekstraksi fitur yang lebih baik. Sebaliknya, pemilihan yang tidak tepat dapat merusak atau bahkan menghilangkan fitur. Beragam praproses telah diterapkan oleh penelitian-penelitian terkait terdahulu, di antaranya adalah *data augmentation*, *illumination normalization*, *face alignment* dan *pose normalization*.

Zero Component Analysis (ZCA) merupakan teknik *image whitening* yang terbukti mengungguli performa model klasifikasi gambar yang dikenakan oleh teknik normalisasi rerata dan standarisasi piksel gambar. Beberapa contoh hasil penggunaan teknik ini ditunjukkan pada Gambar 2.16. Sementara fungsi ZCA (Pal & Sudeep,

(a) Konvolusi 2D oleh Kernel 2×2 dengan *Stride* 1×1 tanpa *Zero Padding*(b) Konvolusi 2D oleh Kernel 2×2 dengan *Stride* 2×1 tanpa *Zero Padding*(c) Konvolusi 2D oleh Kernel 2×2 dengan *Stride* 2×1 dan *Zero Padding*

Gambar 2.15 Kasus-Kasus Konvolusi 2D (Goodfellow dkk., 2016)



Gambar 2.16 Contoh-Contoh Hasil ZCA (Pal & Sudeep, 2016)

2016) dijelaskan pada (2.24),

$$X_{ZCA} = U \cdot \text{diag}(1/\sqrt{\text{diag}(S) + \epsilon}) \cdot U^T \cdot X' \quad (2.24)$$

$$X' = X/255 \quad (2.25)$$

dimana,

U = matriks vektor Eigen,

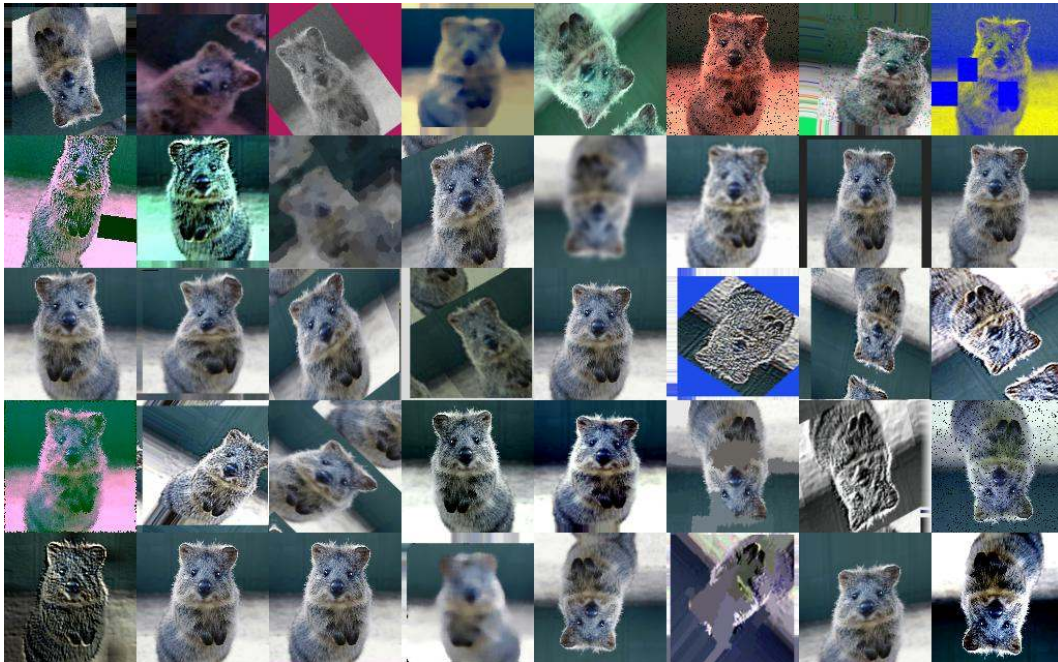
U^T = matriks transpos dari U ,

$\text{diag}(M)$ = diagonal matriks M ,

S = matriks nilai Eigen dari dekomposisi nilai singular matriks *covariance*,

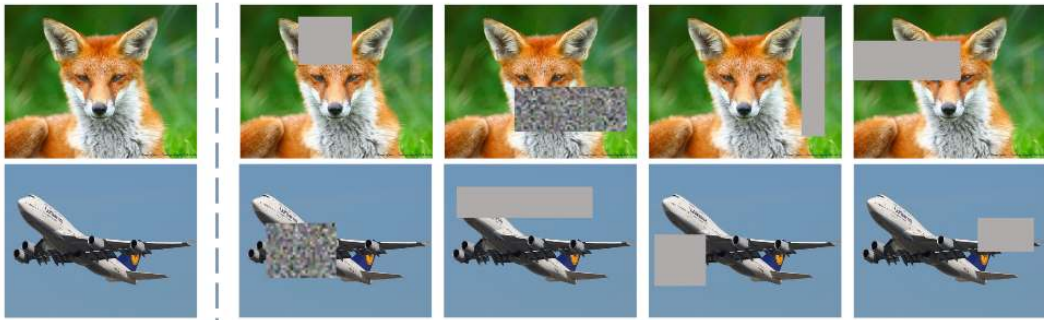
ϵ = koefisien *whitening*,

X' = hasil normalisasi oleh penskalaan fitur terhadap data X .

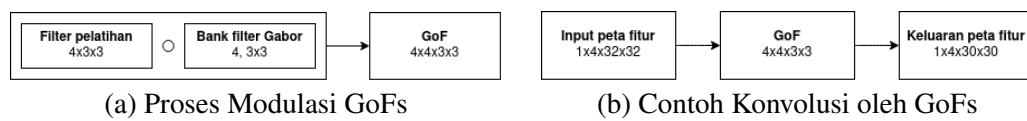


Gambar 2.17 Contoh-Contoh Hasil Augmentasi Data
(<https://github.com/aleju/imgaug>)

Data augmentation biasa diterapkan untuk mengatasi *overfitting* pada pelatihan model CNN dengan set data yang terbatas. *Data augmentation* dapat dilakukan baik menggunakan pendekatan konvensional (filter kernel, transformasi geometri, penghapusan acak, transformasi ruang warna dan pembauran gambar-gambar) maupun *deep learning* (*adversarial learning*, *neural style transfer* dan *GAN data augmentation*). *Data augmentation* juga dapat dilakukan menggunakan *meta learning*, yaitu paduan dari kedua pendekatan tersebut. Gambar 2.17 menunjukkan beberapa contoh hasil augmentasi data. Sejauh ini berdasarkan hasil eksperimen, pendekatan tradisional yang paling efektif dalam meningkatkan akurasi model adalah teknik *cropping*, yaitu sebuah teknik pemrosesan gambar yang bekerja dengan menghilangkan sekumpulan piksel tertentu dari gambar (Shorten & Khoshgoftaar, 2019). Di antara teknik *cropping* yang paling populer dalam meningkatkan ketahanan model CNN terhadap set data kecil adalah *random erasing*, di mana teknik ini menghilangkan satu atau dua area piksel gambar secara acak (Gambar 2.18). Area-



Gambar 2.18 Contoh-Contoh Hasil Random Erasing (Zhong dkk., 2017)



Gambar 2.19 Modulasi dan Konvolusi GoFs di GCNs (Luan dkk., 2018)

area gambar yang dihapus digantikan oleh piksel-piksel yang serupa atau sebarang (Zhong dkk., 2017; O’Gara & McGuinness, 2019). Penjelasan mendetail mengenai *data augmentation* dapat dipelajari pada Shorten & Khoshgoftaar (2019).

2.5.3 Gabor Convolutional Networks

Gabor Convolutional Networks (GCNs), yang pertama kali diperkenalkan oleh Luan dkk. (2018), merupakan DCNNs yang dimasuki oleh filter konvolusi *Gabor Orientation Filters* (GoFs) untuk meningkatkan ketahanan fitur yang dipelajari dalam perubahan skala dan orientasi. GoFs merupakan lapisan konvolusi dari modulasi filter pelatihan oleh filter Gabor dengan variasi skala dan orientasi. Gambar 2.19 menunjukkan bagaimana proses modulasi GoFs berlangsung. Sedangkan Gambar 2.20 menunjukkan bagaimana GoFs menempati setiap lapisan konvolusi pada jaringan tersebut.

2.6 Facial Region Segmentation

Pada studi segmentasi bagian wajah dalam konteks pengenalan ekspresi, tidak ada kesepakatan tentang bagaimana segmentasi yang optimal itu. Namun berdasarkan



C—Konvolusi Spasial; MP—Max-Pooling; R—ReLU; M—Max; BN—BatchNormalization; D: Dropout

Gambar 2.20 Arsitektur GCNs (Luan dkk., 2018)

studi literatur, penulis menyimpulkan bahwa mayoritas penelitian menggunakan tiga bagian wajah yang paling menonjol pada pengenalan emosi. Ketiga bagian tersebut secara berurutan adalah mata, hidung, dan mulut (K. Guo, 2012). Proporsi kontribusi ketiga bagian tersebut hampir sama untuk setiap emosi *angry*, *disgust*, *fear*, *happy*, *sad*, *surprise*, dan *neutral*. Benitez-Garcia dkk. (2018) membuktikan bahwa pengenalan ekspresi dengan tiga bagian ini mengungguli pengenalan menggunakan empat bagian, yaitu mata, hidung, mulut, dan dahi.

Segmentasi bagian wajah dapat dilakukan melalui berbagai cara, di antaranya melalui pendekatan *grid/block based* dan *local region based*. Dalam penelitian tersebut, mereka membagi wajah menjadi delapan belas bagian (Gambar 2.10) di mana area mata dibagi menjadi enam bagian, area hidung dibagi menjadi dua bagian dan sepuluh bagian sisanya untuk dahi, pipi dan dagu. Segmentasi area mulut pun dibuat sangat spesifik dengan membaginya berdasarkan garis luar bibir. Hasil eksperimen menunjukkan bahwa representasi *local region based* lebih baik daripada representasi holistik dalam *face registration* (Ghimire dkk., 2015). Ini menunjukkan bahwa mendefinisikan RoI secara spesifik dapat meningkatkan kualitas ekstraksi fitur.

BAB III

METODE PENELITIAN

Pada bab ini disajikan detail perancangan metode penelitian yang perlu dan sangat berpeluang untuk dilakukan dalam rangka menjawab sejumlah permasalahan utama yang diangkat meliputi deskripsi set data publik yang akan digunakan, tahap-tahap pelaksanaan eksperimen dalam berbagai skenario yang diusulkan, metode serta metrik yang digunakan dalam evaluasi dan alat serta bahan penelitian yang diperlukan.

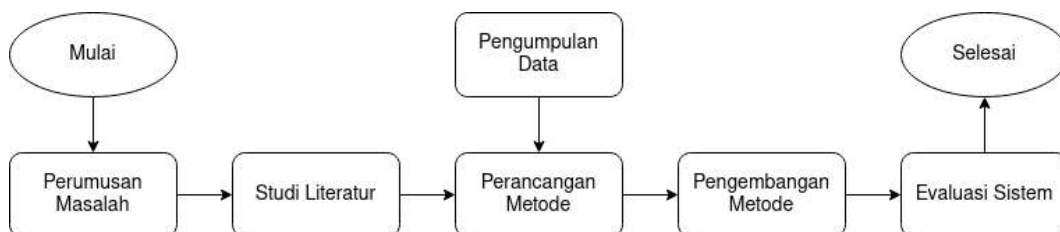
3.1 Perancangan Penelitian

Perancangan penelitian merupakan gambaran utuh mengenai tahap-tahap kerja penulis dalam menemukan solusi permasalahan yang diangkat sebagaimana yang ditunjukkan pada Gambar 3.1.

3.1.1 Perumusan Masalah

Beranjak dari penentuan bidang penelitian (*problem area*) yang ditekuni dan penemuan masalah (*problem finding*) yang muncul —akibat adanya kesenjangan antara harapan dan kenyataan; sebagai bahan utama dalam latar belakang penelitian— maka selanjutnya masalah tersebut diuraikan dengan berpijak kepada penelitian-penelitian terdahulu. Tahap ini memiliki peran yang sangat penting karena mendasari atas pelaksanaan penelitian.

Dengan perumusan masalah ini akan menghasilkan deskripsi tujuan dan man-



Gambar 3.1 Alur Penelitian

faat penelitian secara jelas. Tujuan yang dimaksud adalah keberhasilan penelitian dalam menemukan usulan solusi yang tepat atas permasalahan yang telah dirumuskan. Sedangkan manfaat yang dimaksud menjelaskan bagaimana penelitian ini dapat berkontribusi kepada kehidupan masyarakat.

3.1.2 Studi Literatur

Tahap ini mencakup usaha peneliti dalam memahami akar dari rumusan masalah yang dihadapi dan mempelajari bagaimana penelitian-penelitian terdahulu dalam menyelesaikan persoalan terkait. Dengan memahami akar permasalahan, peneliti mengetahui teori-teori dan/atau konsep-konsep yang harus dikuasai untuk memulai penelitian. Dengan mempelajari penelitian terkait, peneliti mengetahui wilayah kontribusinya pada bidang penelitian yang sama untuk membuat usulan solusi baru.

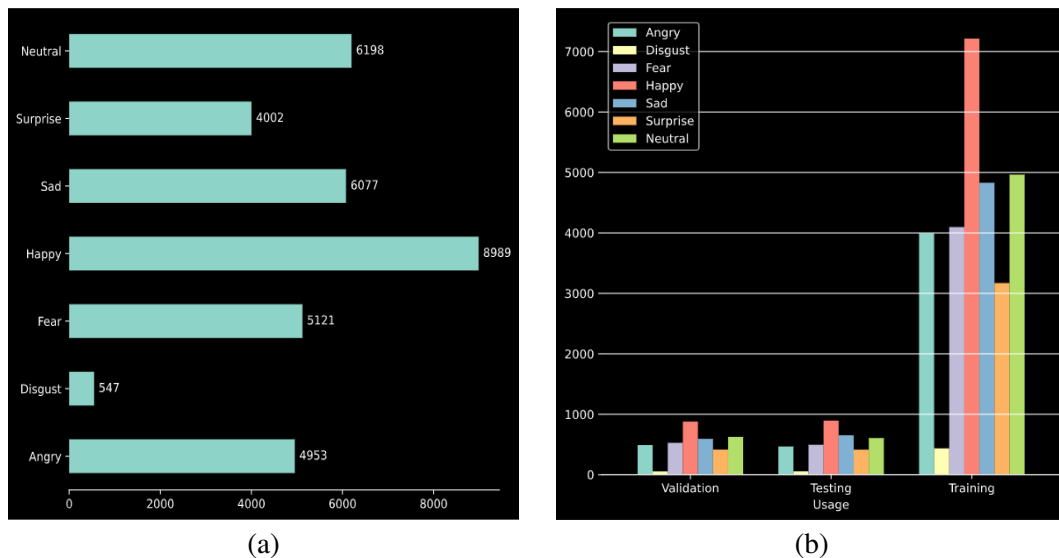
Dengan studi literatur ini, penulis dapat memilih untuk mengembangkan solusi yang diusulkan oleh penelitian lain atau mengembangkan solusi baru dalam menyelesaikan persoalan yang sama. Juga penulis dapat mengevaluasi seberapa andal, akurat dan efektif solusi yang diusulkan terhadap penelitian terdahulu. Adapun literatur yang dijadikan sebagai studi berkenaan dengan teori, konsep atau teknik *facial expression recognition*, GCNs, *facial region segmentation* serta berbagai disiplin ilmu terkait lain.

3.1.3 Pengumpulan Data

Pada penelitian ini, penulis menggunakan set data publik FER-2013 diambil dari <https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data>. Set data ini terdiri dari 35.887 gambar pose wajah yang dilabeli manual berdasarkan tujuh kelas emosi, yaitu *angry*, *disgust*, *fear*, *happy*, *sad*, *surprise*, dan *neutral*. Masing-masing gambar berukuran 48 x 48 piksel dalam skala warna abu-abu (*grayscale*). Pada Gambar 3.2, diperlihatkan beberapa contoh data gambar FER-2013 per label.



Gambar 3.2 Pratinjau Set Data FER-2013



Gambar 3.3 Distribusi Set Data FER-2013 per Label

Set data ini juga dikelompokkan berdasarkan fungsinya ke dalam set data *training*, *validation* dan *testing* berturut-turut menurut rasio 8:1:1. Pada Gambar 3.3a, diperlihatkan bahwa distribusi data per label tidak merata (*imbalanced dataset*). Yang menjadi kekhawatiran terbesar adalah data berlabel *disgust* hanya ditemukan sejumlah kurang lebih 1,5% dari keseluruhan. Hal ini dapat menyebabkan model tidak akan mampu mengenali ekspresi wajah *disgust*. Kendati pun demikian, jika mengesampingkan hal tersebut, distribusi set data FER-2013 terbilang cukup

merata. Di sisi lain, persebaran data per label untuk tiap-tiap fungsi seimbang sebagaimana yang ditunjukkan oleh Gambar 3.3b.

Disebutkan bahwa set data ini dikumpulkan menggunakan teknik *web crawling* dengan kata kunci terkait emosi, sehingga memiliki variasi yang sangat tinggi terhadap populasi dan lingkungan. Beberapa contoh variasi tersebut ditunjukkan pada Gambar 3.4. Variasi yang sangat beragam ini penting sekali dalam pertimbangan pemilihan teknik pada sejumlah proses berikutnya meliputi *image enhancement*, *face detection*, *facial landmark detection* dan *facial alignment* yang cocok digunakan.

3.1.4 Perancangan Metode

Pada penelitian ini, penulis memiliki tiga tahap utama sebagaimana yang tertera pada Gambar 3.5. Setelah set data FER-2013 dimuat, penulis akan menganalisis apakah terdapat data bising atau tidak. Jika ditemukan data bising atau tidak bersesuaian dengan konteks, maka akan dibersihkan. Tahap ini disebut *data cleansing* atau *cleaning*. Setelah dibersihkan, maka akan dilakukan penyusunan ulang set data jika diperlukan.

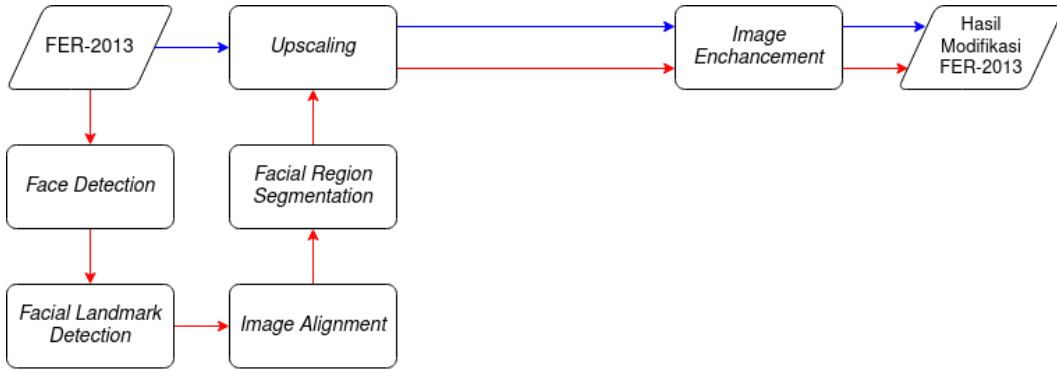
Tahap kedua merupakan praproses data, di mana dilakukan dua skenario penyesuaian sebagaimana yang ditunjukkan oleh Gambar 3.6. Kedua skenario ini diperlukan berbagai skenario pada tahap ketiga. Pada skenario pertama, praproses hanya melalui tahap *upscaling* —penyesuaian ukuran dimensi data gambar terhadap arsitektur CNN yang digunakan— dan *image enhancement* (jika ada). Pada skenario kedua, ada empat tahap tambahan sebelum pada akhirnya memasuki kedua tahap tadi. Pertama, seluruh data gambar akan dilakukan pendeteksian wajah. Jika terdapat wajah dalam gambar, maka akan diteruskan ke tahap kedua, yaitu pendeteksian *facial landmark*. *Facial landmark* adalah titik-titik koordinat relatif pada setiap gambar yang menunjukkan sekitar komponen wajah seperti mata, hidung dan mulut. Setelah itu, dilakukan penjajaran (*alignment*) area wajah yang terdeteksi



Gambar 3.4 Variasi Set Data FER-2013



Gambar 3.5 Rancangan Metode



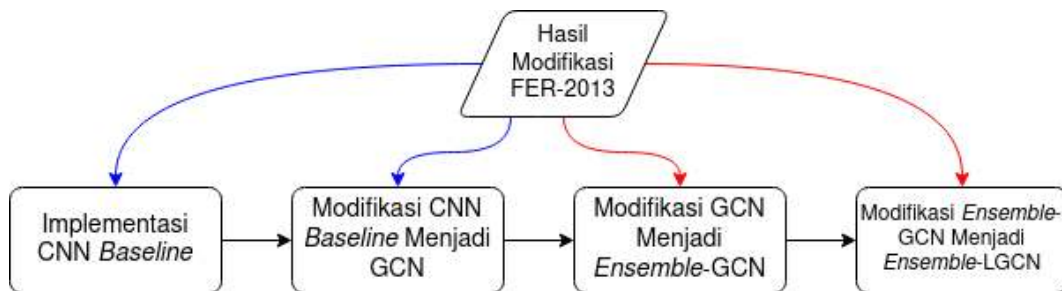
Gambar 3.6 Subtahap Praproses Data

terhadap rotasi. Penjajaran yang di sini memiliki dua subtahap, yaitu penjajaran *facial landmark* terhadap *bounding box* dari hasil deteksi wajah dan penjajaran *facial landmark* yang saling bersesuaian. Penjajaran terhadap rotasi ini dihitung menggunakan rumus trigonometri (Zill & Dewar, 2011) pada (3.1) dengan mengambil tiga titik sampel dari *facial landmark* —atau dua titik sampel dan sebuah titik hasil pemetaan salah satu dari kedua titik tersebut—,

$$\theta = \tan^{-1} \left(\frac{a}{b} \right) = \tan^{-1} \left(\frac{B_y - C_y}{A_x - C_x} \right) \quad (3.1)$$

Sebelum ke tahap keempat, tiap-tiap gambar wajah dilakukan pemotongan menurut *bounding box* yang dihasilkan dan perhitungan ulang *facial landmark*. Terakhir, akan dilakukan segmentasi bagian-bagian wajah untuk tiap-tiap gambar berdasarkan *facial landmark* yang dihasilkan.

Tahap ketiga merupakan tahap eksekusi empat buah skenario pemodelan pengenalan ekspresi wajah. Seperti yang tercantum pada Gambar 3.7, dua skenario awal merupakan implementasi dan pengembangan model *baseline* dengan mengadopsi ide dari arsitektur GCNs tanpa melibatkan set data hasil dari proses *facial*



Gambar 3.7 Subtahap Eksekusi Empat Skenario Pemodelan

region segmentation. Skenario berikutnya merupakan pengembangan model terbaik sebelumnya menjadi *network ensemble*. Dilanjutkan dengan skenario terakhir, yaitu uji coba penggantian filter Gabor yang diadopsi pada arsitektur GCNs menjadi filter log-Gabor.

3.1.5 Pengembangan Metode

Hasil rancangan metode penelitian akan mengalami pengembangan melalui lima tahap prosedur. Tahap pertama, verifikasi terhadap metode penelitian akan dilakukan sebelum metode tersebut diterapkan. Verifikasi ini akan menilai apakah setiap persyaratan teknis dan teoretis yang ditentukan telah dipenuhi atau tidak. Setelah berhasil diverifikasi, metode penelitian akan dieksekusi untuk pertama kalinya.

Tahap kedua merupakan validasi awal terhadap penerapan pertama kali atas metode penelitian. Proses validasi ini akan melibatkan hasil eksperimen dari berbagai literatur yang menjadi rujukan penelitian ini.

Tahap ketiga merupakan inisialisasi konfigurasi pemodelan pengenalan ekspresi wajah seperti menentukan metrik pengukuran yang cocok untuk kasus klasifikasi emosi manusia, menentukan fungsi aktivasi lapisan keluaran yang cocok untuk klasifikasi multikelas, menentukan *seed* untuk beberapa proses yang melibatkan *random number generator*, menentukan berapa kali nilai bobot pada model akan diperbarui ketika *training*, menentukan kapan model perlu disimpan ketika *training* dan menentukan teknik pengukuran kinerja model untuk evaluasi.

Tahap keempat merupakan perbaikan atas metode penelitian yang diusulkan jika pada tahap kedua ditemukan adanya kekurangan atau kesalahan dalam implementasi. Perbaikan ini tidak terbatas pada metode-metode yang telah direncanakan sebelumnya. Melainkan fleksibel mengikuti hasil eksperimen dan hipotesis yang akan muncul dari setiap tahap eksperimen yang dilalui. Metode yang sudah diperbaiki selanjutnya akan dijalankan dan divalidasi kembali.

Setelah metode penelitian telah dinyatakan valid, pada tahap kelima, metode akan dievaluasi dan dioptimalkan berulang kali mengacu pada performa model prediksi yang dihasilkan. Model yang paling optimal akan diambil sebagai hasil akhir dari penelitian ini. Setiap perubahan terkendali pada pengembangan metode ini akan selalu dicatat pada bab berikutnya sebagai bahan evaluasi akhir.

3.1.6 Evaluasi Sistem

Evaluasi sistem dilakukan untuk menyimpulkan performa model pengenalan ekspresi wajah untuk tiap-tiap skenario yang berhasil dieksekusi. Kinerja model diukur menggunakan *f1-score* (Goutte & Gaussier, 2005) pada (3.2)–(3.4),

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} = \frac{\text{True Positive}}{\text{Total Predicted Positive}} \quad (3.2)$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} = \frac{\text{True Positive}}{\text{Total Actual Positive}} \quad (3.3)$$

$$\text{F-1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3.4)$$

Precision memberikan gambaran mengenai proporsi identifikasi positif yang benar, sehingga akan bernilai sama dengan 1,0 jika tidak ada *false positive* (kesalahan prediksi pada kelas positif). Sementara *recall* memberikan gambaran mengenai proporsi positif aktual yang diidentifikasi dengan benar, sehingga akan bernilai sama dengan 1,0 jika tidak ada *false negative* (kesalahan prediksi pada kelas negatif). Log pengukuran performa model untuk setiap *epoch* saat *training* juga akan disimpan.

Tabel 3.1
Spesifikasi Perangkat Keras

Jenis Komponen	Merek	Memori	Kecepatan
CPU	Intel Core i5-8250U	-	1,6GHz–3,4GHz
RAM	Corsair Vengeance DDR4	2 × 8GB	2.400MHz
GPU	Nvidia GeForce GTX 1050M	4GB	-

Tabel 3.2
Spesifikasi Perangkat Lunak

Nama Program	Versi	Nama Program	Versi
CUDA <i>driver</i>	10.2	jupyter-client	6.1.3
Nvidia <i>driver</i>	440.100	LogGabor	20191113
Visual Studio Code	1.47.1	matplotlib	3.2.1
Python	3.8.3	notebook	6.0.3
GCC	10.1.1-1	numpy	1.18.5
gcn	1.0	opencv-python	4.2.0.34
imgaug	0.4.0	pandas	1.0.4
jupyter	1.0.0	scikit-learn	0.23.1
jupyter-core	4.6.3	torch	1.5.0

Di akhir, evaluasi juga akan dilakukan melalui analisis secara kualitatif.

3.2 Alat dan Bahan Penelitian

Alat yang digunakan pada penelitian ini adalah seperangkat komputer untuk melatih model pengenalan ekspresi wajah. Spesifikasi perangkat keras dan lunak yang digunakan ditunjukkan berturut-turut pada Tabel 3.1 dan Tabel 3.2. Sistem operasi yang digunakan adalah Fedora 32 (*Workstation Edition*) dengan kernel Linux 5.7.8-200. Sementara bahan yang digunakan adalah set data publik pengenalan ekspresi wajah nonfrontal, FER-2013.

BAB IV

TEMUAN DAN PEMBAHASAN

Pada bab ini disajikan berbagai temuan dan pembahasan secara komprehensif terhadap tahap-tahap pelaksanaan eksperimen dan pengujian yang berhasil dilakukan pada setiap skenario yang telah dirancang.

4.1 *Data Cleansing*

Set data FER-2013 didistribusikan resmi dalam sebuah *file* teks CSV¹ berukuran 301,1MB. Sementara itu, pemuatan *file* menggunakan pustaka *pandas* hanya memerlukan memori sebesar 291,2MB. *File* tersebut terdiri dari 35.887 baris dan 3 kolom, yaitu *emotion*, *pixels* dan *Usage*. Setiap kolom berturut-turut memuat 7 label emosi dalam rentang nilai integer 0–6 (0: *angry*, 1: *disgust*, 2: *fear*, 3: *happy*, 4: *sad*, 5: *surprise* dan 6: *neutral*), 2.304 piksel gambar dalam rentang nilai 0–255 dan 3 jenis penggunaan yang unik (*Training*: set data *training*, *PrivateTest*: set data *validation* dan *PublicTest*: set data *testing*). Adapun distribusi data telah dijelaskan pada bab yang sebelumnya.

Berdasarkan pengecekan secara hati-hati dan menyeluruh, setiap entri data adalah unik. Sehingga dapat dipastikan bahwa tidak ada irisan entri data yang sama untuk penggunaan yang berbeda. Meskipun dalam pemeriksaan singkat, ditemukan beberapa gambar yang terlihat sangat mirip. Namun hal ini tidak akan menjadi masalah, sebab tidak ada data gambar yang memiliki piksel-piksel yang tepat sama.

Data cleansing dimulai dengan pemeriksaan data rusak, tidak relevan atau tidak lengkap. Pemeriksaan khusus data gambar dilakukan secara manual satu per satu supaya menyeluruh dan akurat. Alhasil disimpulkan bahwa setiap entri data adalah lengkap. Sayangnya, sejumlah data gambar sebanyak kurang lebih 85 buah

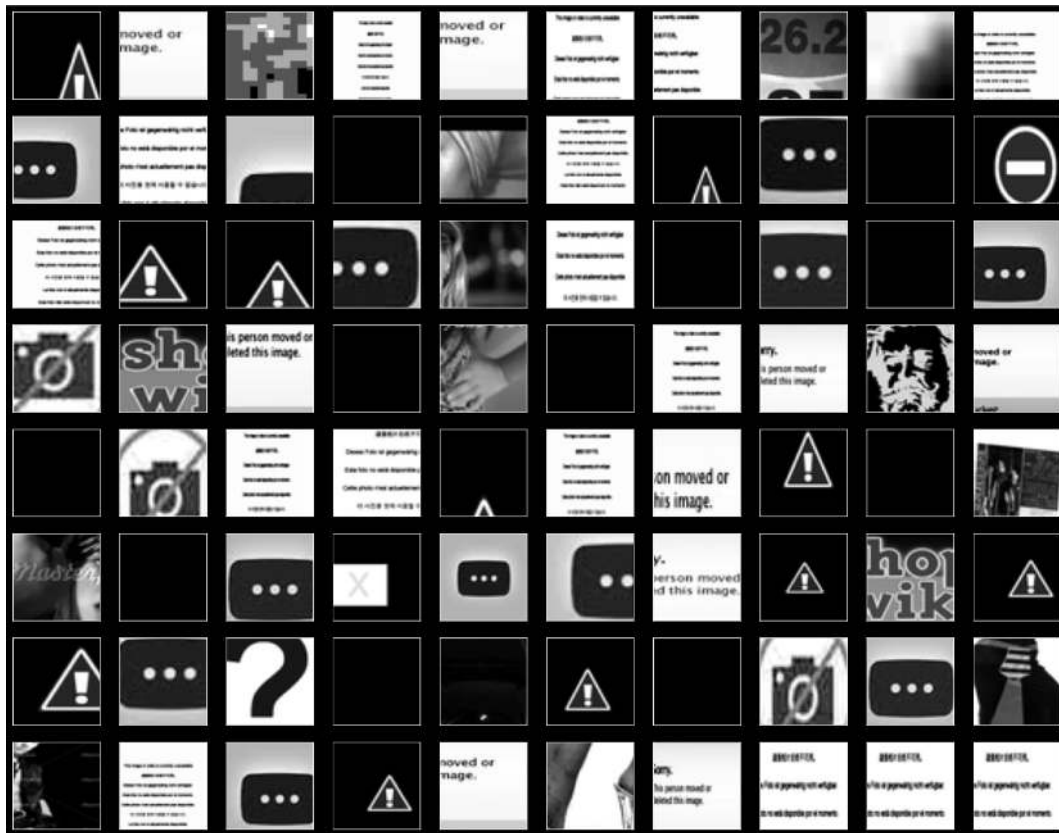
¹CSV atau *Comma-Separated Value*, merupakan format *file* teks khusus berbentuk tabel di mana setiap nilai dipisahkan menggunakan tanda koma.

ditemukan rusak, yaitu data gambar ke-59, 2059, 2171, 2809, 3262, 3931, 4275, 5082, 5274, 5439, 5722, 5881, 6102, 6458, 6699, 7172, 7496, 7527, 7629, 8030, 8737, 8856, 9026, 9500, 9673, 9679, 10423, 11244, 11286, 11295, 11846, 12289, 12352, 13402, 13697, 13988, 14279, 15144, 15553, 15838, 15894, 17081, 19238, 19632, 20222, 20712, 20817, 21817, 22198, 22314, 22407, 22927, 23596, 23894, 24053, 24441, 24891, 25219, 25603, 25909, 26383, 26860, 26897, 28601, 29447, 29557, 30002, 30705, 30981, 31127, 31825, 32662, 32683, 34334, 35121, 35469, 35632, 35743, 5509, 10023, 11826, 17620, 10219, 14550, dan 15389. Sehingga setiap entri data terkait perlu dihapus. Pada Gambar 4.1, ditampilkan beberapa contoh data gambar yang rusak. Yang banyak terlihat di sana adalah gambar sudah tidak dapat diakses kembali dari alamat web semula. Namun, yang membingungkan adalah semua itu masih merupakan *file* gambar, yang mana seolah-olah gambar-gambar tersebut berupa hasil tangkapan layar (*screenshot*).

4.2 Praproses Data

Pada skenario pertama, praproses data diawali dengan memperbesar (*upscaling*) dimensi seluruh data gambar dari 48×48 piksel menjadi 64×64 piksel menggunakan *opencv-python* dalam rangka penyesuaian ukuran set data gambar dengan dimensi lapisan input arsitektur model CNN *baseline*. Selanjutnya, penulis mencoba menemukan teknik *image enhancement* yang paling cocok untuk meningkatkan kualitas atau keterbacaan tiap-tiap gambar. Teknik-teknik yang akan penulis uji coba hanya berkisar pada teknik *illumination normalization*. Namun, sebelum menerapkan teknik-teknik tersebut, penulis mengambil beberapa sampel gambar yang memiliki karakteristik yang cukup jauh. Dengan harapan bahwa sejumlah gambar tersebut dapat mewakili setiap variasi karakteristik gambar yang akan dikenakan teknik *enhancement*.

Teknik pertama adalah *image whitening* menggunakan ZCA untuk menggantikan proses normalisasi yang disebutkan pada penelitian *baseline*. Kode program-



Gambar 4.1 Beberapa Contoh Data Gambar Rusak pada FER-2013

nya diambil dari <https://github.com/mwv/zca>. Sayangnya, sebagaimana yang ditunjukkan pada Gambar 4.2, pengenalan teknik ini ternyata menghilangkan banyak sekali informasi spasial mencakup informasi tekstur (seperti kerutan kulit di dahi dan pipi) dan garis tepi. Penulis menduga hal ini disebabkan oleh kecilnya resolusi gambar serta adanya limitasi dari *grayscale*. Kekurangan ini tentunya sangat bertentangan dengan alasan mengapa penulis mengadopsi filter Gabor pada metode penelitian penulis.

Teknik kedua adalah *histogram equalization* yang cukup populer dan dapat dipercaya terutama untuk meningkatkan kualitas area gambar yang relatif terlalu gelap. Hasilnya dapat dilihat pada Gambar 4.3. Secara keseluruhan, teknik ini sangat andal dalam meningkatkan kualitas gambar baik yang terlalu gelap maupun terlalu terang. Namun, teknik ini gagal terhadap gambar yang memiliki kontras



Gambar 4.2 Beberapa Contoh Hasil Implementasi ZCA terhadap FER-2013



Gambar 4.3 Beberapa Contoh Hasil Implementasi *Histogram Equalization* terhadap FER-2013

yang cukup jauh antara objek dan latar. Alhasil, kedua teknik ini tidak akan penulis gunakan pada penelitian ini.

Pada skenario kedua, penulis menambahkan empat langkah ekstra tepat sebelum memasuki tahap *upscaling*. Tahap pertama merupakan pendeteksian area wajah pada setiap gambar dengan batas ambang (*threshold*) sama dengan 0,5. Setiap area wajah yang ditandai dan berhasil melewati *threshold* akan diteruskan ke tahap berikutnya, yaitu pendeteksian atau prediksi *facial landmark*. Kedua tahap ini dapat ditangani baik secara terpisah maupun tergabung oleh berbagai pustaka yang telah tersedia publik. Penulis telah mencoba beberapa di antaranya dan mem-

Tabel 4.1
Perbandingan Teknik-Teknik Facial Landmark Detection

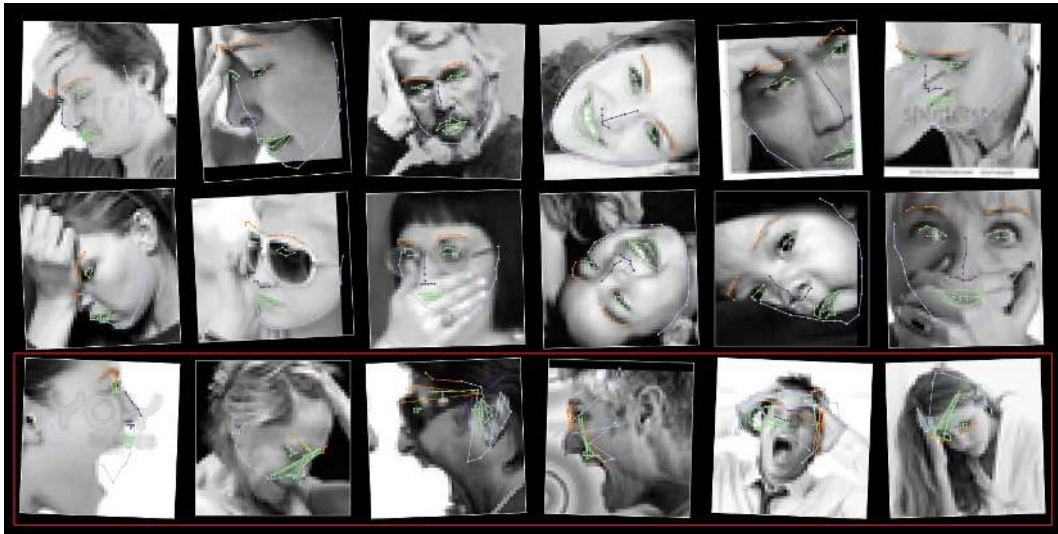
Pustaka	$Loss$ (%) ∇	Waktu (s) ∇	n^{Δ}	Threshold	GPU
<i>face_alignment</i> (face_detector=sfd)	3,10	2.786	68	0,5	✓
<i>RetinaFace</i> (net=mobilenet0.25)	42,23	334	5		
<i>RetinaFace</i> (net=resnet50)	7,05	476			

Loss—persentase banyak data gambar yang tidak terdeteksi wajah; Waktu—waktu total eksekusi pendeteksian *facial landmark* untuk seluruh data gambar; n —banyak *facial landmark* per wajah; ^ΔLebih tinggi lebih baik; [▽]Lebih rendah lebih baik.

bandingkan metode mana yang paling kredibel untuk diimplementasikan pada set data wajah nonfrontal, FER-2013. Namun, di sini penulis hanya akan membahas hasil dari dua metode terbaik, yaitu *Face Alignment Network* (FAN) dalam pustaka *face_alignment* (Bulat & Tzimiropoulos, 2017) dan *RetinaFace* (Deng dkk., 2019).

Sebagaimana yang disimpulkan pada Tabel 4.1, metode FAN unggul dalam kecilnya persentase kehilangan data dan banyaknya *facial landmark* yang mampu diprediksi. Kendati pun waktu eksekusinya sangat lama, hal ini tidak akan menjadi masalah yang penting. Sementara itu, metode *RetinaFace* memiliki waktu eksekusi relatif cukup cepat, namun memiliki persentase kehilangan data yang relatif besar. Terutama metode *RetinaFace* yang menggunakan arsitektur MobileNet0.25, metode *RetinaFace* kurang bisa diandalkan. Akan tetapi, masih perlu validasi tiap-tiap metode dengan mengambil beberapa sampel hasil deteksi *facial landmark*.

Berdasarkan Gambar 4.4 di baris pertama dan kedua, metode FAN sangat presisi dalam memprediksi *facial landmark*. Bahkan berhasil memprediksi *facial landmark* pada objek-objek wajah dengan rotasi yang sangat ekstrem dan/atau halangan yang cukup mengganggu. Namun, pendeteksian tidak akurat bahkan sangat berantakan untuk sampel gambar wajah di baris ketiga. Sayangnya, penulis pun tidak dapat menyimpulkan gambar-gambar wajah dengan karakteristik bagaimana yang gagal diprediksi. Oleh karena penulis kesulitan jika harus memeriksa



Gambar 4.4 Beberapa Contoh Hasil Deteksi *Facial Landmark* terhadap FER-2013 Menggunakan FAN

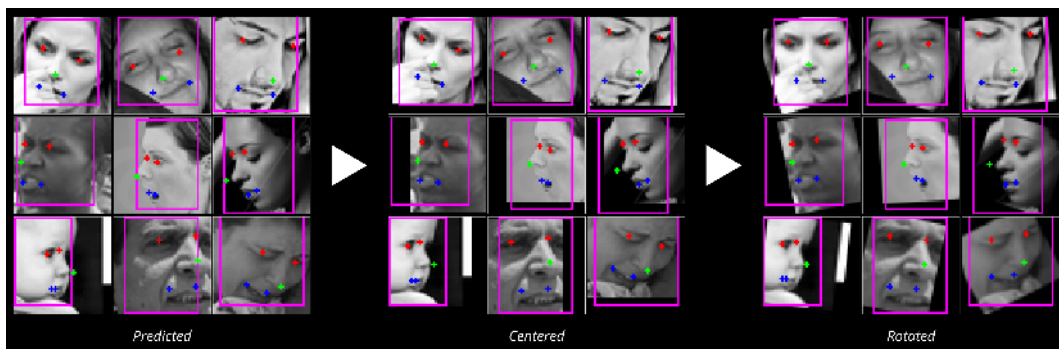
satu per satu hasilnya, maka metode FAN dinilai tidak bisa diandalkan.

Kemudian penulis mencoba meninjau hasil prediksi metode *RetinaFace* dengan ResNet50 yang ditunjukkan pada Gambar 4.5. Di sini penulis mengambil sampel dari seratus gambar wajah pertama yang berhasil dikenali dan diprediksi dengan skor yang mendekati *threshold*. Hasilnya, prediksi *facial landmark* menggunakan metode ini cukup akurat dan presisi. Disertai kenyataan bahwa entri data terkait yang gagal dikenali tersebar cukup merata pada setiap label emosi, maka metode ini menjadi pilihan penulis yang terbaik dalam mendeteksi *facial landmark*.

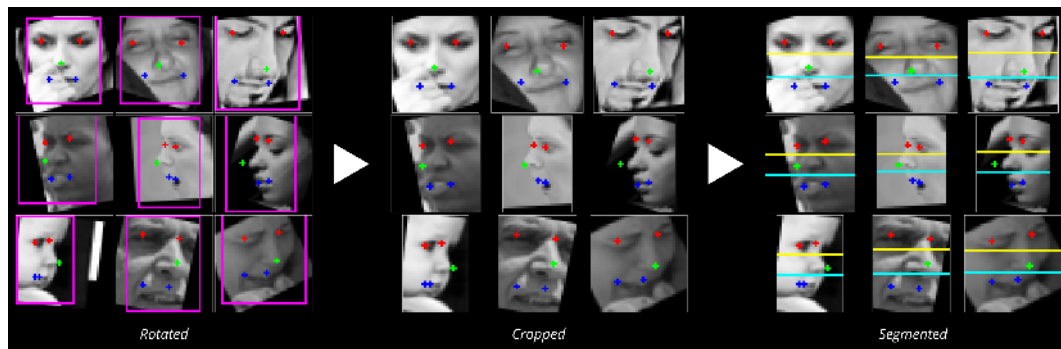
Tahap ketiga adalah *face alignment*, di mana dilakukan penjajaran setiap objek wajah menggunakan pustaka *opencv* berdasarkan *bounding box* wajah dan *facial landmark* yang bersesuaian. Pertama-tama, sentralisasi setiap objek wajah dilakukan melalui penggeseran seluruh piksel gambar sejauh perpindahan antara titik senter semua *facial landmark* ke titik senter *bounding box* wajah. Lalu perotasi-an setiap objek wajah dilakukan terhadap titik senter *bounding box* melalui teknik tertentu yang sesuai dengan salah satu dari tiga situasi yang telah didefinisikan. Situasi pertama adalah jika koordinat pada sumbu x tengara hidung berada di tengah-



Gambar 4.5 Beberapa Contoh Hasil Deteksi *Facial Landmark* terhadap FER-2013 Menggunakan *RetinaFace* pada ResNet50



Gambar 4.6 Beberapa Contoh Hasil per Subtahap *Face Alignment*



Gambar 4.7 Beberapa Contoh Hasil per Subtahap *Facial Region Segmentation*

tengah antara tengara sisi kiri dan kanan bibir, maka perotasian dilakukan untuk menjajarkan koordinat pada sumbu y kedua tengara mata. Situasi kedua adalah jika koordinat pada sumbu x tengara hidung berada di kiri terhadap tengara sisi kiri bibir, maka perotasian dilakukan untuk menjajarkan koordinat pada sumbu x tengara mata kanan dan sisi kanan bibir. Situasi ketiga adalah jika koordinat pada sumbu x tengara hidung berada di kanan terhadap tengara sisi kanan bibir, maka perotasian dilakukan untuk menjajarkan koordinat pada sumbu x tengara mata kiri dan sisi kiri bibir. Setiap proses pada tahap ini direpresentasikan ke dalam Gambar 4.6, di mana setiap baris gambar mencontohkan hasil dari masing-masing situasi secara berurutan dari atas ke bawah.

Tahap keempat adalah *facial region segmentation*, di mana tiap-tiap bagian wajah disegmentasi guna menghasilkan set data sekunder. Pada tahap ini, diusulkan dua konfigurasi segmentasi yang berbeda. Konfigurasi pertama adalah pembagian setiap objek wajah menjadi tiga area, yaitu kedua mata, hidung dan mulut. Sedangkan konfigurasi kedua adalah pembagian dua area yang saling tumpang tindih (*overlapping*), yaitu area kedua mata hingga hidung dan area hidung hingga mulut. Proses segmentasi ini dilakukan dengan memotong gambar sesuai dengan *bounding box*, melakukan kalkulasi ulang untuk tiap-tiap *facial landmark* dan memotong gambar per area yang ditentukan. Setiap area yang dipotong didefinisikan oleh sisi-sisi *bounding box* dan garis horizontal yang melalui titik tengah antar dua

Tabel 4.2
Perbandingan Performa Model Baseline dengan dan tanpa Augmentasi Data

	Akurasi (%) [△]	Epoch	Waktu (jam) [▽]
Tanpa data augmentasi	54,24	48	0,31
Dengan data augmentasi	60,27	174	1,64

Akurasi—akurasi model pada *testing*; *Epoch*—banyak *epoch* yang dibutuhkan agar model menjadi optimal; Waktu—waktu yang dibutuhkan agar model menjadi optimal; [△]Lebih tinggi lebih baik; [▽]Lebih rendah lebih baik.

area yang bersinggungan. Misalnya untuk area kedua mata, dibatasi oleh sisi atas, kiri dan kanan *bounding box* serta garis horizontal batas area kedua mata dan hidung. Contoh-contoh setiap proses pada tahap ini dapat dilihat pada Gambar 4.7.

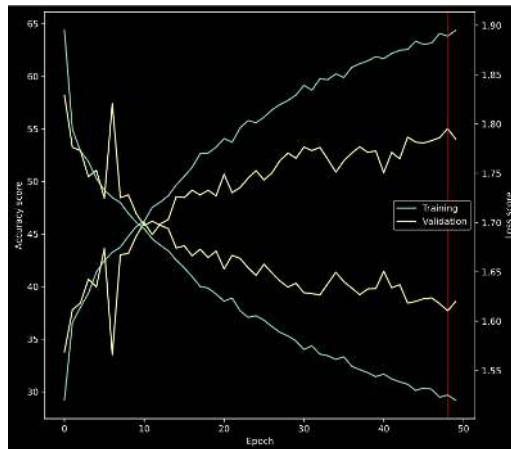
Secara khusus, melalui algoritma segmentasi yang dikembangkan pada penelitian ini, penulis berupaya menjawab permasalahan segmentasi bagian-bagian wajah pada set data wajah nonfrontal. Yang mana set data tersebut memuat berbagai variasi stuktur wajah yang berbeda (Islam, Mahmud, & Hossain, 2018b).

4.3 Eksekusi Skenario Pemodelan

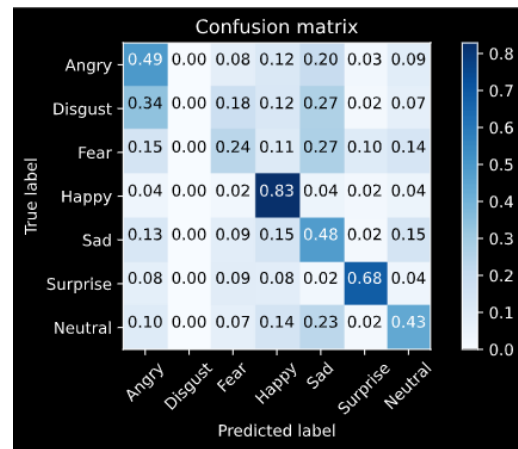
Pada subbab ini, penulis akan menjelaskan proses dan hasil eksperimen secara mendetail untuk setiap skenario pemodelan yang telah direncanakan. Setiap skenario dijalankan sebanyak tiga kali percobaan tanpa mengubah konfigurasi apapun untuk lalu diambil model dengan performa yang terbaik. Penggunaan statistik pada percobaan berulang merupakan solusi tradisional dalam menyimpulkan kinerja model. Meskipun pada dasarnya penulis telah mengeset *seed* untuk setiap *random number generator* ke nilai tertentu, performa model tetap terus berubah pada pengulangan percobaan yang berikutnya.

4.3.1 Implementasi CNN *Baseline*

Eksperimen dimulai dengan mengimplementasikan model CNN *baseline* pilihan untuk pengenalan ekspresi wajah pada set data FER-2013. Model yang diklaim dapat mencapai akurasi tes 65,23% tersebut, berhasil penulis rekonstruksi ulang.

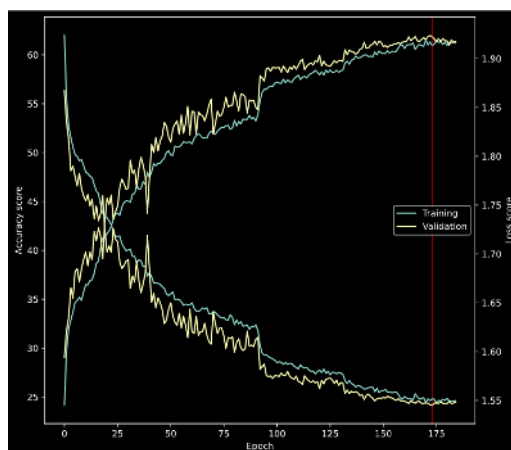


(a) Grafik Performa Per *Epoch* untuk Model tanpa Augmentasi Data

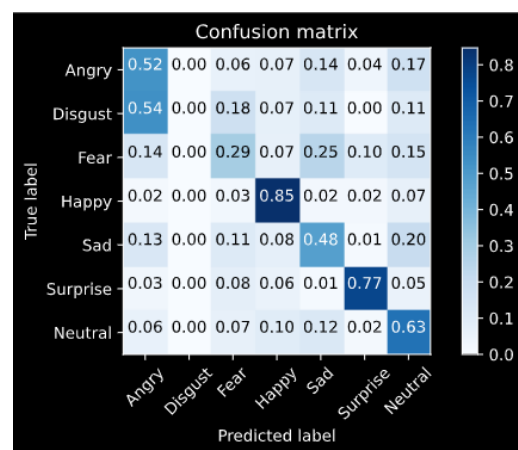


(b) *Confusion Matrix* Performa Model tanpa Augmentasi Data

Gambar 4.8 Performa Model CNN *Baseline* tanpa Augmentasi Data



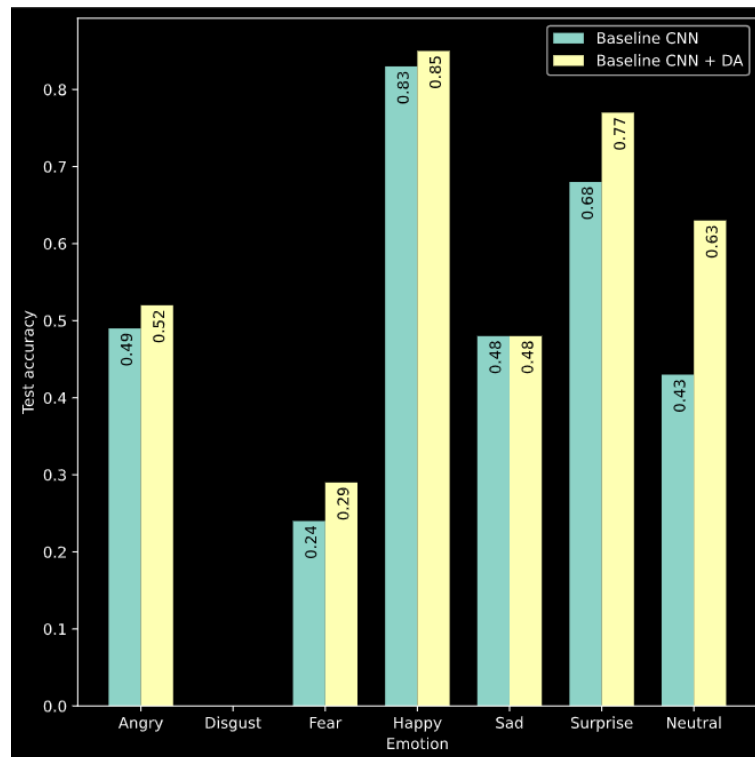
(a) Grafik Performa Per *Epoch* untuk Model dengan Augmentasi Data



(b) *Confusion Matrix* Performa Model dengan Augmentasi Data

Gambar 4.9 Performa Model CNN *Baseline* dengan Augmentasi Data

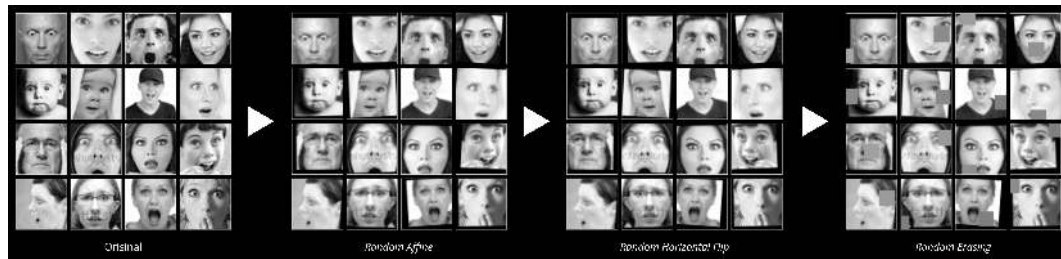
Namun, penulis kesulitan bahkan hanya untuk mendekati akurasi tersebut. Penulis menduga terdapat beberapa alasan untuk hal ini. Pertama, yang mana merupakan hal yang paling penting, pembagian distribusi data *training*, *validation* dan *testing* berbeda. Jika pada penelitian *baseline* menggunakan rasio 80:20 untuk data *training* dan *testing*, penulis menggunakan rasio 80:10:10 berturut-turut untuk data *training*, *validation* dan *testing* sesuai dengan *default* dari set data FER-2013 sendiri. Dengan mempertimbangkan bahwa hampir seluruh penelitian pada FER-2013



Gambar 4.10 Perbandingan Performa Model dengan dan tanpa Augmentasi Data Per Kelas Emosi

menggunakan rasio 80:10:10, akan menjadi lebih adil bagi penulis dalam membandingkan hasil metode usulan penulis dengan penelitian lain. Kedua, penelitian *baseline* tidak memerinci bagaimana cara mereka melakukan augmentasi data. Menurut pengamatan penulis, penggunaan teknik augmentasi data dapat meningkatkan performa model pada FER-2013. Seperti yang terlihat pada Tabel 4.2, augmentasi data telah meningkatkan performa model secara cukup signifikan, yaitu sebesar 6,03%. Performa model meningkat untuk setiap label emosi kecuali *disgust* dan *sad*. Jika membandingkan *confusion matrix* pada Gambar 4.8b dan 4.9b, data berlabel *disgust* malah lebih banyak dikenali sebagai *angry*. Meskipun waktu *training*-nya meningkat sebesar 5,29 kali lipat.

Pada tahap augmentasi data, penulis mengadopsi tiga macam teknik berturut-turut adalah transformasi *affine* acak, *horizontal flipping* acak dan *random erasing*.



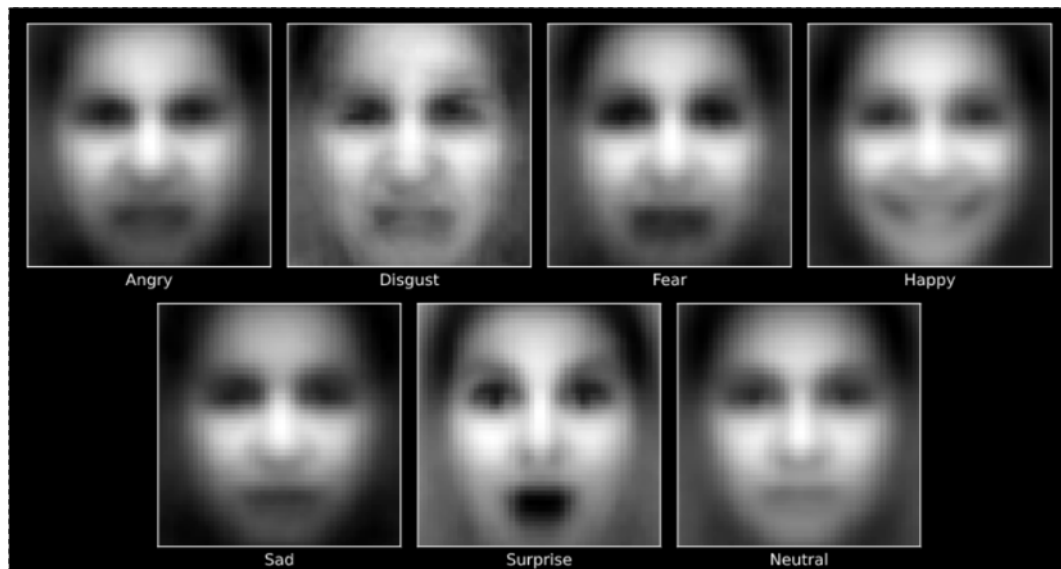
Gambar 4.11 Beberapa Contoh Hasil Per Subtahap *Augmentasi Data*



Gambar 4.12 Beberapa Contoh Kemiripan Data Gambar Wajah Berlabel Emosi *Disgust* (Kiri) dan *Angry* (Kanan) pada FER-2013

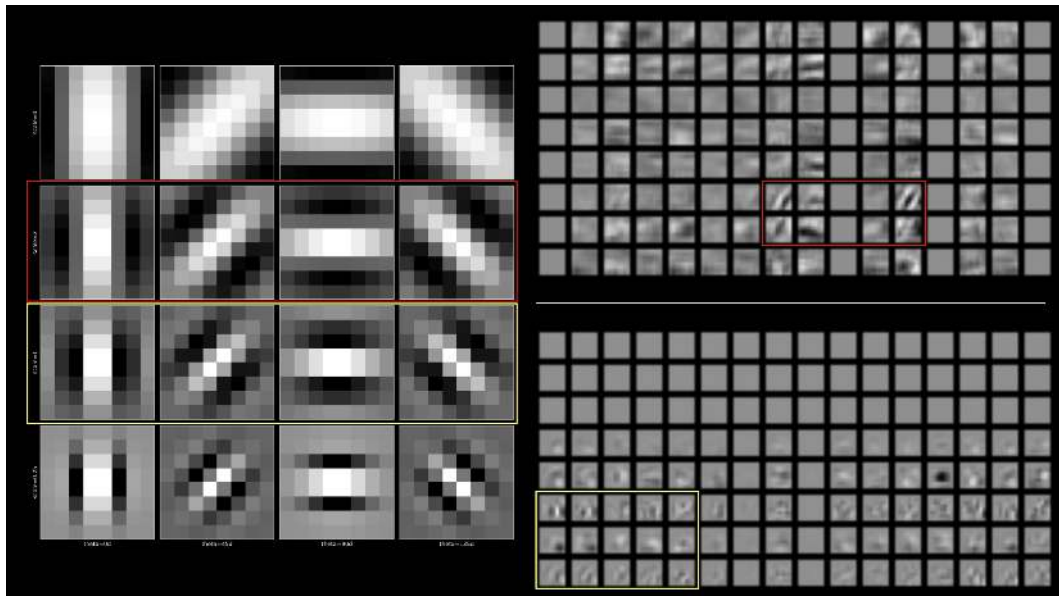
Khusus untuk eksperimen yang melibatkan *facial region segmentation*, teknik *random erasing* tidak digunakan sebab arsitektur CNN kesulitan untuk belajar. Sebab memang untuk kasus tertentu, pengombinasian lebih dari dua teknik augmentasi data dapat mempengaruhi kemampuan model dalam generalisasi. Proses augmentasi data ini dilakukan secara berulang per *batch*, atau biasa disebut sebagai *online data augmentation*, agar proses *training* tidak terlalu berat pada komputer berspesifikasi yang telah disebutkan pada bab sebelumnya. Gambar 4.11 memperlihatkan beberapa contoh untuk setiap proses data augmentasi.

Ada beberapa pertanyaan yang muncul setelah melihat *confusion matrix* pada Gambar 4.9a. Pertama, yang mana merupakan hal yang paling mencolok, model



Gambar 4.13 Rerata Seluruh Data Gambar Wajah pada FER-2013 Per Label Emosi

tidak pernah berhasil dalam mengenali emosi berlabel *disgust*. Apakah karena data *training*-nya terlalu sedikit? Lalu jika dilihat secara saksama, emosi *disgust* lebih banyak dikenali sebagai *angry*. Apakah karena emosi *disgust* sulit dibedakan dari *angry*? Untuk itu, penulis melakukan perbandingan beberapa sampel data dari kedua label emosi tersebut yang dapat dilihat pada Gambar 4.12. Di sana terlihat beberapa data gambar yang telah ditandai ternyata memang agak sulit untuk dibedakan oleh penulis secara manual sekalipun. Namun hanya dengan ini, penulis tidak bisa serta-merta menyimpulkan bahwa terdapat kesalahan dalam pelabelan set data FER-2013. Kemudian penulis mencoba melakukan pendekatan yang berbeda untuk menjawab persoalan ini, yaitu dengan menghitung rerata gambar wajah per label emosi. Berdasarkan Gambar 4.13, penulis menyimpulkan bahwa sebenarnya meskipun terdapat beberapa gambar yang terlihat sangat mirip, namun secara keseluruhan data gambar per label masih dapat dibedakan. Kedua, data berlabel *fear* belum mampu seperempatnya dikenali oleh model. Malahan lebih banyak dikenali sebagai emosi *sad*. Sedangkan untuk data berlabel lain *sad* dan *neutral* masih belum mampu dikenali separuhnya. Penulis menduga bahwa hal ini akibat dari ku-



Gambar 4.14 Beberapa Contoh Kemiripan Filter Gabor (Kiri) dengan Kernel dari Lapisan Konvolusi Ke-8 (Kanan Atas) dan Ke-9 (Kanan Bawah) pada Model CNN *Baseline*

rangnya kompleksitas model *learning*. Oleh karena itu, penulis bersemangat untuk melanjutkan eksperimen ke skenario yang berikutnya. Secara keseluruhan, performa model rekognisi untuk tiap-tiap label emosi mengalami kenaikan. Gambar 4.10 merangkum peningkatan tersebut berdasarkan Gambar 4.8b dan 4.9b.

4.3.2 Modifikasi CNN *Baseline* Menjadi GCN

Pada skenario ke-2 ini, penulis melakukan modifikasi arsitektur CNN *baseline* menjadi GCN melalui: 1) pengubahan setiap lapisan konvolusi biasa menjadi lapisan konvolusi Gabor, 2) penambahan sebuah *max layer* sebelum lapisan akhir dan 3) penambahan sebuah lapisan *dense (fully connected layer)* di bagian paling akhir dalam arsitektur model.

Pada arsitektur GCN, terdapat dua parameter tambahan yang perlu ditentukan, yaitu M dan $scale$. M adalah parameter yang menentukan banyaknya variasi rotasi filter Gabor yang akan dipakai dalam rentang $0-180^\circ$. Jika $M = 1$, maka

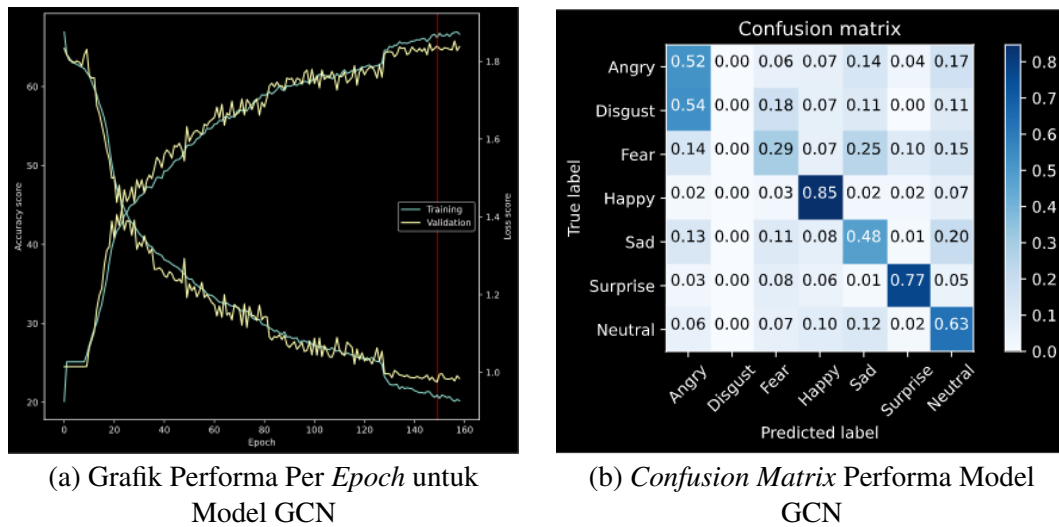
Tabel 4.3
Perbandingan Performa Model CNN Baseline dan GCN

	Akurasi (%) [△]	Epoch	Waktu (jam) [▽]
Model CNN <i>baseline</i>	60,27	174	1,64
Model GCN	63,51	150	7,95

Akurasi—akurasi model pada *testing*; Epoch—banyak *epoch* yang dibutuhkan agar model menjadi optimal; Waktu—waktu yang dibutuhkan agar model menjadi optimal; [△]Lebih tinggi lebih baik; [▽]Lebih rendah lebih baik.

hanya sebuah filter Gabor pada sudut 0° yang dipakai. Jika $M > 2$, maka filter Gabor yang dipakai adalah sebanyak M buah meliputi sebuah filter Gabor pada sudut 0° ditambah $M - 1$ buah filter Gabor pada rotasi yang dihitung secara kumulatif menurut $\theta = (i/M) \times 180^\circ$ di mana i adalah indeks urutan filter Gabor yang dimulai dari $i = 1$. Sedangkan *scale* adalah skala filter Gabor relatif terhadap ukuran kernel yang dipakai pada setiap lapisan konvolusi, yaitu 8×8 . Dari enam belas lapisan konvolusi pada arsitektur *baseline*, penulis mengelompokkannya menjadi empat grup secara berurutan. Kemudian penulis melakukan *training* pada model CNN *baseline* yang telah diubah menjadi GCN melalui pencacahan konfigurasi parameter M dalam rentang nilai 1–4 dan *scale* dalam rentang nilai yang sama. Dari situ penulis dapat menyimpulkan bahwa parameter yang optimal untuk kasus ini adalah $M = 4$ dengan *scale* = 2 untuk lapisan konvolusi ke-1 hingga ke-8 dan *scale* = 1 untuk lapisan konvolusi yang lain. Hal ini dapat dijelaskan melalui visualisasi kernel pada setiap lapisan konvolusi, di mana kernel pada lapisan konvolusi ke-8 ke bawah mirip dengan filter Gabor pada *scale* = 2 dan kernel pada lapisan konvolusi ke-9 ke atas mirip dengan filter Gabor pada *scale* = 1. Pada Gambar 4.14 diperlihatkan beberapa contoh kemiripan tersebut. Pada bagian kiri, ditampilkan empat baris filter Gabor dengan parameter *scale* yang berbeda pada empat rotasi yang berbeda. Pada bagian kanan, dari atas ke bawah, menunjukkan separuh kernel dari lapisan konvolusi ke-8 dan ke-9 secara berurutan.

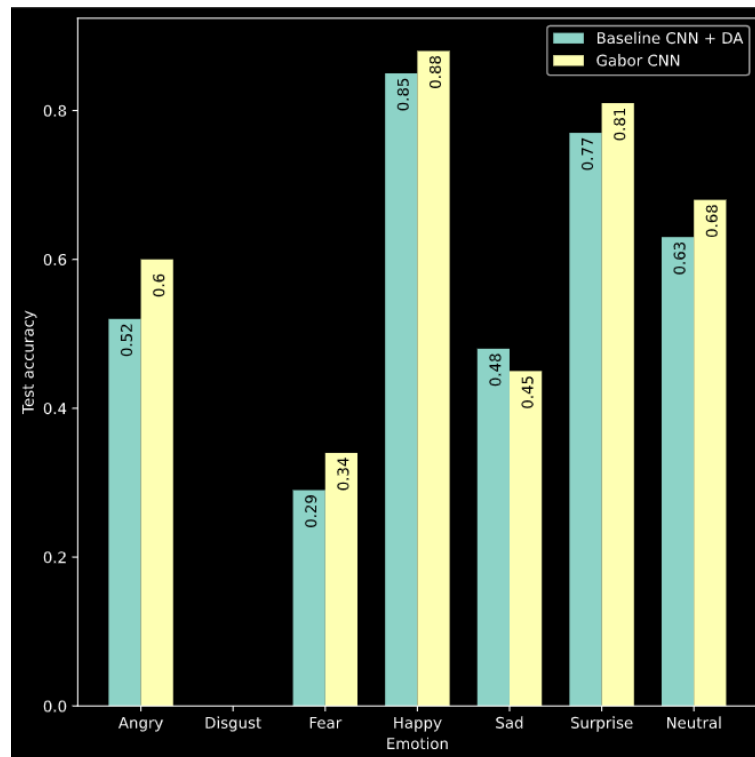
Dari eksperimen ini dibuktikan bahwa model GCN, yang merupakan hasil modifikasi dari CNN *baseline*, memiliki performa yang lebih unggul dari pendahu-



Gambar 4.15 Performa Model GCN

lunya. Sebagaimana yang terangkum dalam Tabel 4.3, performa model meningkat sebesar 3,24% disertai oleh peningkatan waktu *training* sebesar 4,84 kali lipat. Di sisi lain, banyak *epoch* yang harus dilalui untuk melatih model GCN lebih rendah daripada sebelumnya untuk *batch size* yang sama.

Sebelum membahas lebih lanjut mengenai perbandingan performa kedua model, ada dua hal yang menarik bagi penulis ketika membandingkan grafik log performa model pada Gambar 4.9a dan 4.15a. Pertama, melalui pengamatan penulis pada pengubahan parameter M , penggunaan arsitektur model GCN selalu dimulai dengan grafik akurasi untuk *training* dan *validation* yang sama sekali konstan pada *epoch* ke-1 hingga ke-10. Sementara itu, grafik *loss* untuk *training* dan *validation* selalu mengalami perbaikan. Baru pada *epoch* ke-10 hingga ke-25, tiap-tiap grafik mengalami perbaikan yang sangat signifikan. Sedangkan grafik log performa model yang sebelumnya relatif lebih halus setelah *epoch* ke-5. Hal ini menerangkan bahwa pengenalan filter Gabor pada lapisan konvolusi CNN menyulitkan mesin untuk belajar, namun memberikan wawasan yang lebih baik dalam rekognisi kelas emosi. Kedua, yang mana mengherankan penulis, grafik performa *validation* hampir selalu lebih baik daripada grafik performa *training* hingga *epoch* tertentu di mana terlihat



Gambar 4.16 Perbandingan Performa Model CNN *Baseline* dan GCN Per Kelas Emosi

loncatan peningkatan yang cukup dengan jelas sesaat setelah grafik performa *training* relatif mulai stabil. Jika dibandingkan dengan grafik log performa model yang paling awal pada Gambar 4.8a, hal ini menyatakan bagaimana data augmentasi dan penggunaan *Gabor convolutional layer* dapat meningkatkan kemampuan generalisasi model rekognisi. Adapun kejadian di mana grafik *validation* mulai konstan setelah mengalami loncatan peningkatan yang signifikan, menjelaskan bahwa hal ini terjadi akibat penulis menggunakan teknik pengecilan parameter *learning rate* menggunakan fungsi *reduce learning rate on plateau* dalam *training* dengan faktor pengali sebesar $\sqrt{0,1}$ ketika menemukan bahwa tidak ada peningkatan senilai tertentu pada grafik *validation loss* dalam sepuluh *epoch* terakhir.

Secara umum, pengadopsian arsitektur GCN telah berhasil meningkatkan performa model rekognisi per kelas emosi, seperti yang diperlihatkan pada Gambar

4.16. Performa model meningkat untuk setiap label emosi kecuali *disgust* dan *fear*. Berulang lagi tidak ada perbaikan apapun dalam pengenalan emosi *disgust*. Bahkan jika membandingkan *confusion matrix* pada Gambar 4.9b dan 4.15b, data berlabel *disgust* malah lebih banyak lagi dikenali sebagai *angry*. Sehingga penulis berakhir pada kesimpulan bahwa ketimpangan yang sangat jelas terjadi pada banyak data berlabel emosi *disgust* telah menyebabkan kegagalan model dalam mengenali emosi tersebut. Sementara kemampuan model dalam mengenali emosi *sad* menjadi sedikit berkurang.

4.3.3 Modifikasi GCN Menjadi *Ensemble* GCNs

Pada skenario ini, penulis mencoba membangun *ensemble network* menggunakan dua teknik yang berbeda. Teknik yang pertama merupakan teknik *ensemble* konvensional, yaitu penggabungan hasil prediksi secara terpisah dari setiap model yang dilatih menggunakan bagian wajah tertentu. Hasil prediksi akhir untuk sebuah input gambar baru dihitung menggunakan rumus statistik tertentu, yaitu *simple average* dan *weighted average*. *Simple average* diperoleh melalui operasi perhitungan rata-rata biasa pada (4.1),

$$\bar{y} = \frac{y_i + y_j}{2} \quad (4.1)$$

sedangkan *weighted average* diperoleh dari (4.2),

$$\bar{y} = \frac{y_i * a_i + y_j * a_j}{a_i + a_j} \quad (4.2)$$

di mana y adalah larik probabilitas berdimensi 1×7 hasil prediksi sebuah input baru dan a adalah akurasi tes untuk masing-masing model i dan j yang berbeda. Sejujurnya, alangkah lebih baik jika *weighted average* dihitung dengan mempertimbangkan akurasi tes per kelas pada *confusion matrix*. Namun penulis kesulitan menghitungnya, sebab model sama sekali tidak bisa memprediksi data berlabel *disgust*. Sementara teknik yang kedua bekerja dengan menggabungkan arsitektur

Tabel 4.4
Perbandingan Performa Berbagai Kombinasi Model Ensemble GCNs

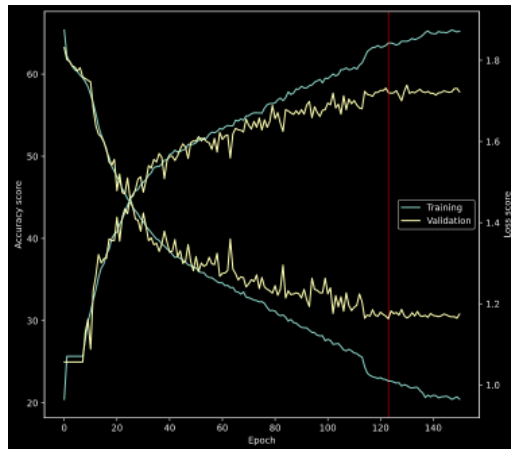
Fitur	Waktu (jam) [▽]		Epoch	Akurasi Testing (%) [△]		
				Single	Simple Average	Weighted Average
EN + NM	4,17	8,87 [▽]	124	58,09	62,20	62,20
	4,70		141	58,06		
E + N + M	3,44	10,34	153	17,41	44,10	53,22
	4,05		181	25,45		
	2,85		129	53,88		
E + N	Ibid.	7,49	Ibid.		17,74	17,62
N + M		6,90			52,77	53,61
E + M		6,29			44,91	53,28
EN + NM (concat.)	7,34	103	59,71	-		
E + N + M (concat.)	8,36	131	54,12			
ENM	6,80	131	62,08	-		

E—(*Eyes*) area bagian kedua mata; N—(*Nose*) area bagian hidung; M—(*Mouth*) area bagian mulut; Epoch—banyak *epoch* yang dibutuhkan agar model menjadi optimal; Waktu—waktu yang dibutuhkan agar model menjadi optimal; [△]Lebih tinggi lebih baik; [▽]Lebih rendah lebih baik.

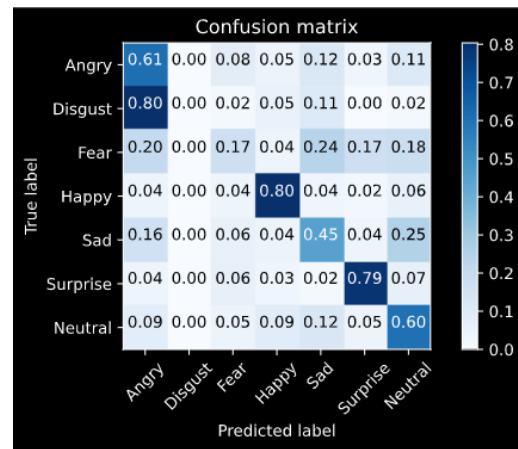
jaringan sebanyak n menjadi jaringan bertingkat (*cascaded network*), di mana n adalah banyak bagian wajah yang disegmentasi.

Sebelum membahas hasil eksperimen pada skenario ini, perlu diketahui bahwa kali ini penulis meninggalkan fungsi *reduce learning rate on plateau*. Karena menurut hasil pengamatan penulis, proses *training* pada skenario ini relatif lambat dan kurang stabil. Sehingga jika penulis tetap menggunakan fungsi tersebut, proses *training* hingga mencapai optimal akan sangat lambat akibat beberapa kali pengecilan *learning rate* dalam rentang waktu yang sebentar.

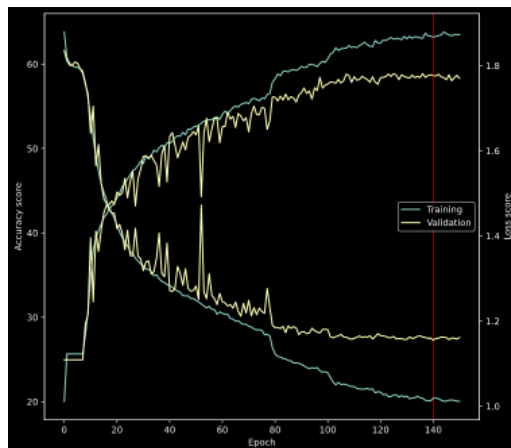
Secara khusus, salah satu tujuan dari skenario ini adalah menjawab sebuah *future work* penelitian terkait demi mengembangkan model rekognisi yang andal menggunakan teknik FRS untuk mampu mengenali emosi dari gambar wajah manusia yang diputar pada sudut berapa pun (Islam, Mahmud, Hossain, Goala, & Mia, 2018). Hasil dari tiap-tiap eksperimen pada skenario ini terangkum pada Tabel 4.4.



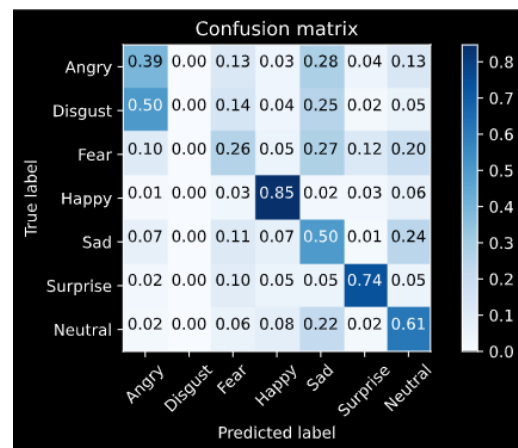
(a) Grafik Performa Per *Epoch* untuk Model EN



(b) *Confusion Matrix* Performa Model EN



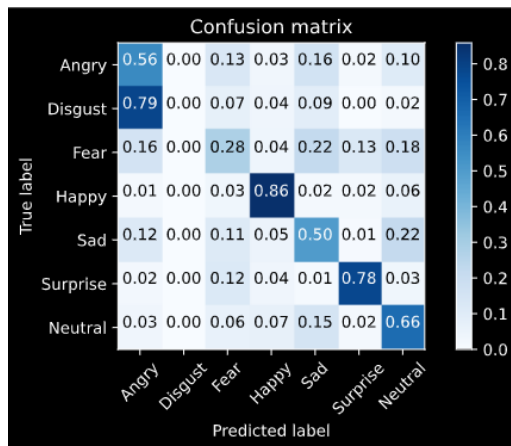
(c) Grafik Performa Per *Epoch* untuk Model NM



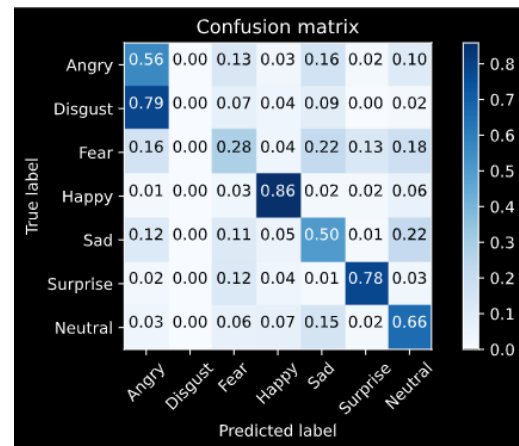
(d) *Confusion Matrix* Performa Model NM

Gambar 4.17 Performa Model GCNs Menggunakan Fitur EN + NM

Pada kolom akurasi *testing*, terlihat bahwa pemodelan menggunakan set data hasil FRS dua bagian wajah (EN + NM) memiliki akurasi yang paling tinggi; jauh lebih baik daripada pemodelan menggunakan tiga bagian wajah (E + N + M) dan kombinasi dua-dua (E + N, N + M dan E + M), baik menggunakan teknik *ensemble* yang pertama maupun yang kedua. Kendati pun akurasinya masih lebih rendah ketimbang model yang dilatih tanpa memanfaatkan teknik FRS (model terbaik pada skenario sebelumnya), namun penurunannya hanya sebesar 1,31%. Sementara pemodelan menggunakan tiga area wajah dan kombinasi dua-dua tidak bisa dian-



(e) *Confusion Matrix* Performa Model EN + NM Berdasarkan Perhitungan *Simple Average*



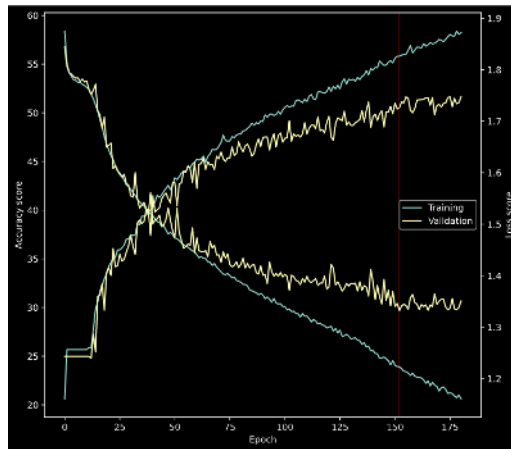
(f) *Confusion Matrix* Performa Model EN + NM Berdasarkan Perhitungan *Weighted Average*

Gambar 4.17 Performa Model GCNs Menggunakan Fitur EN + NM (Lanjutan)

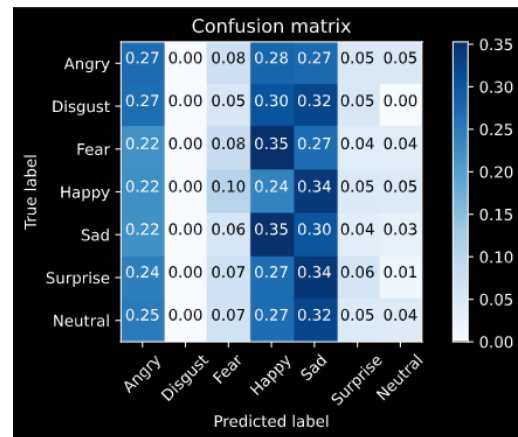
dalkan, bahkan performanya jauh lebih rendah daripada model *baseline*.

Lebih jauhnya, jika ditilik setiap *confusion matrix* yang ditunjukkan pada Gambar 4.17b–4.24b, maka dapat disimpulkan beberapa poin. Dalam membandingkan masing-masing dari performa model EN dan NM, tampak bahwa model EN memiliki akurasi yang lebih tinggi pada pengenalan emosi *angry* dan *surprise*. Untuk pengenalan emosi *angry*, model EN jauh lebih baik dari model NM dengan selisih 22%. Sehingga jelas bahwa emosi *angry* lebih mudah dikenali menggunakan bagian wajah atas manusia. Sedangkan model NM relatif sedikit lebih baik pada sisanya. Khusus untuk *disgust*, model EN lebih banyak melakukan kesalahan prediksi sebagai *angry* dengan selisih 30%. Hal ini menjelaskan bahwa emosi *disgust* lebih sulit dikenali menggunakan bagian wajah atas manusia. Untuk emosi *fear*, dapat disimpulkan bahwa emosi ini lebih sulit dikenali dengan melihat bagian atas dan bawah wajah secara terpisah.

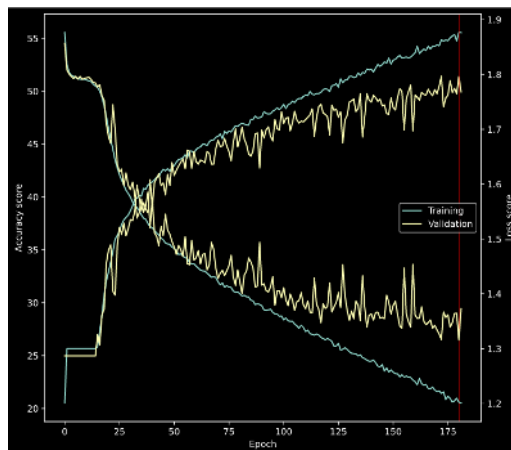
Di sisi lain, performa model EN + NM dalam perhitungan dua teknik rerata memberikan *confusion matrix* yang tepat sama seperti yang terlihat pada Gambar 4.17e dan 4.17f. Untuk setiap label emosi, kombinasi kedua model ini mampu



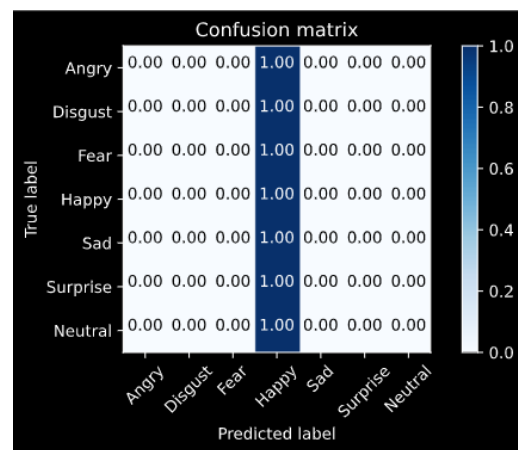
(a) Grafik Performa Per *Epoch* untuk Model E



(b) *Confusion Matrix* Performa Model E



(c) Grafik Performa Per *Epoch* untuk Model N

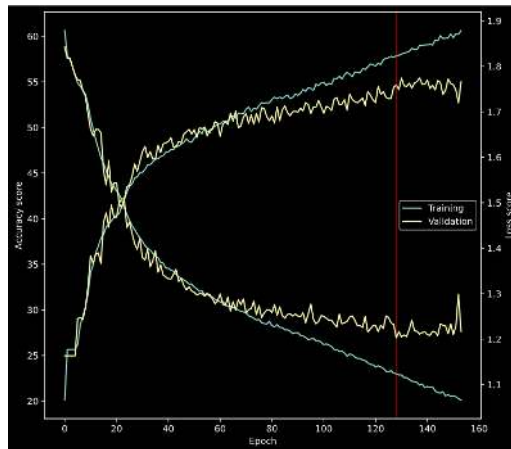


(d) *Confusion Matrix* Performa Model N

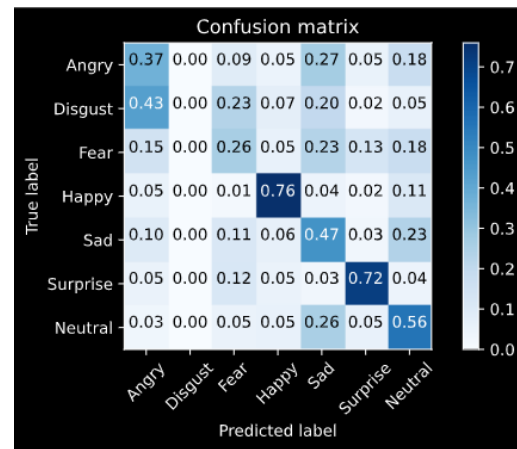
Gambar 4.18 Performa Model GCNs Menggunakan Fitur E + N + M

memberikan nilai akurasi yang sedikit lebih baik dibandingkan model NM. Namun untuk label *angry*, akurasinya 5% lebih rendah daripada model EN. Sementara penyimpangan prediksi emosi *disgust* mendekati kemampuan model EN.

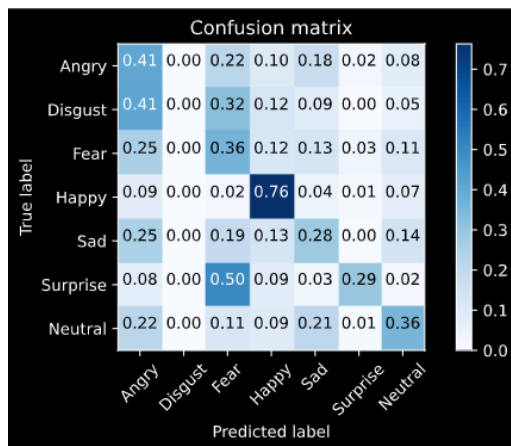
Dalam membandingkan masing-masing dari performa model E, N dan M, terlihat pada Gambar 4.18b–4.18f bahwa untuk model E dan N memiliki *confusion matrix* yang tidak menarik dan masuk akal. Model E hanya mampu mengenali emosi *angry*, *happy* dan *sad* pada akurasi yang kurang dari 30%. Sedangkan untuk emosi sisanya tersebar dengan cukup merata pada ketiga label emosi tersebut. Emosi



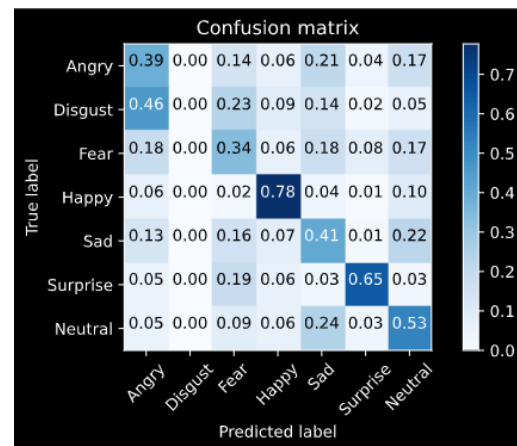
(e) Grafik Performa Per *Epoch* untuk Model M



(f) *Confusion Matrix* Performa Model M



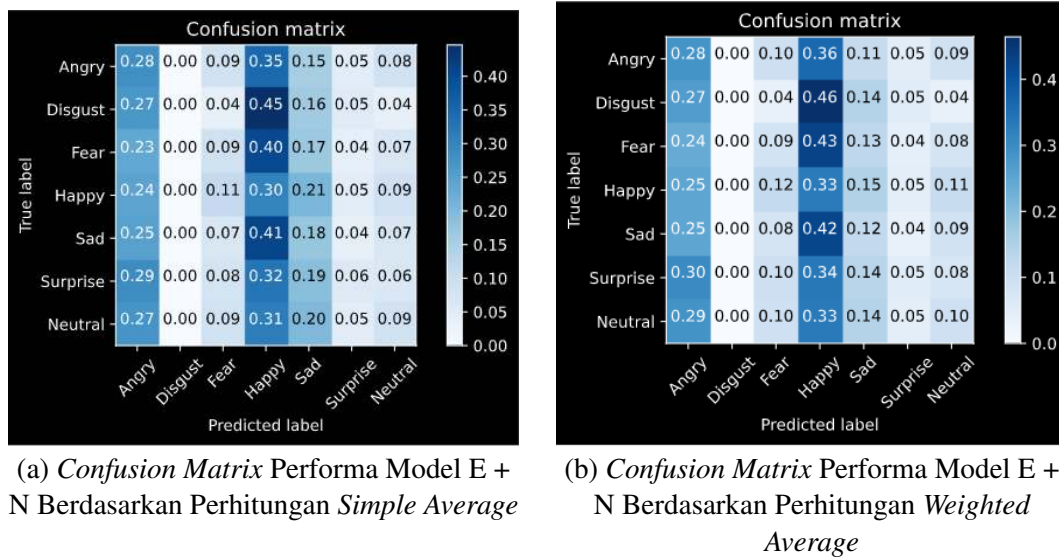
(g) *Confusion Matrix* Performa Model E + N + M Berdasarkan Perhitungan *Simple Average*



(h) *Confusion Matrix* Performa Model E + N + M Berdasarkan Perhitungan *Weighted Average*

Gambar 4.18 Performa Model GCNs Menggunakan Fitur E + N + M (Lanjutan)

surprise dan *neutral* yang umumnya termasuk kelas yang paling mudah untuk dikenali, tidak mampu diprediksi oleh model E. Selanjutnya yang sangat mengejutkan adalah model N hanya mampu mengenali emosi *happy* secara 100% akurat. Namun, label emosi lainnya juga 100% gagal diprediksi dan malah dianggap sebagai *happy*. Kemudian untuk model M, performanya dalam mengenali setiap emosi cukup dapat diterima. Sebab memberikan *confusion matrix* yang mirip dengan model *baseline* tanpa augmentasi data.

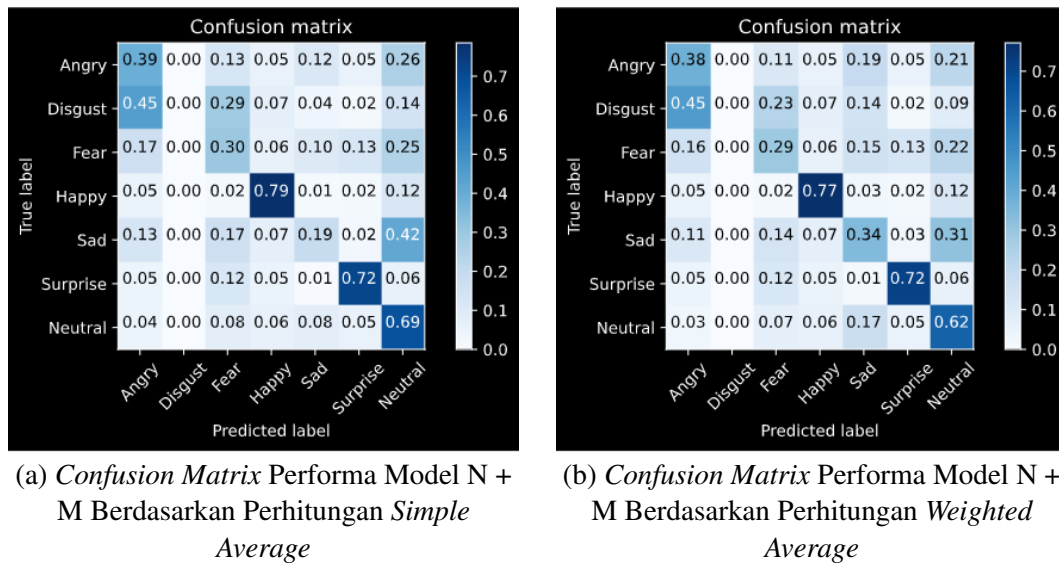


Gambar 4.19 Performa Model GCNs Menggunakan Fitur E + N

Di sisi lain, rerata performa model E + N + M memberikan *confusion matrix* yang cukup mirip dengan performa model M. Terutama dalam perhitungan *weighted average*, model memberikan akurasi yang sedikit lebih baik pada pengenalan emosi *angry* dan *fear*. Sedangkan untuk sisanya relatif menurun sedikit.

Dari analisis di atas, diketahui bahwa ketiga fitur wajah tersebut ternyata saling berkaitan satu sama lain pada pengenalan emosi manusia. Namun area bagian mulut memang memberikan kontribusi paling besar dari keseluruhan skenario yang dilakukan. Di sisi lain, rendahnya akurasi model rekognisi emosi untuk tiap-tiap bagian wajah terpisah mengungkapkan bahwa di antara kemungkinan sebabnya adalah adanya perbedaan yang kuat antara orang barat dan timur dalam bagaimana mereka mengekspresikan suatu emosi (Benitez-Garcia dkk., 2017).

Meskipun penggunaan fitur bagian mulut memberikan akurasi yang relatif lebih baik daripada yang lain, namun penggabungan dengan fitur lain (area bagian mata dan hidung) malah mengurangi sedikit performa model rekognisi. Di mana model N + M memberikan penurunan yang cukup signifikan pada pengenalan emosi *sad* dengan selisih akurasi 13%. Akan tetapi untuk emosi *angry*, *fear*, *happy* dan

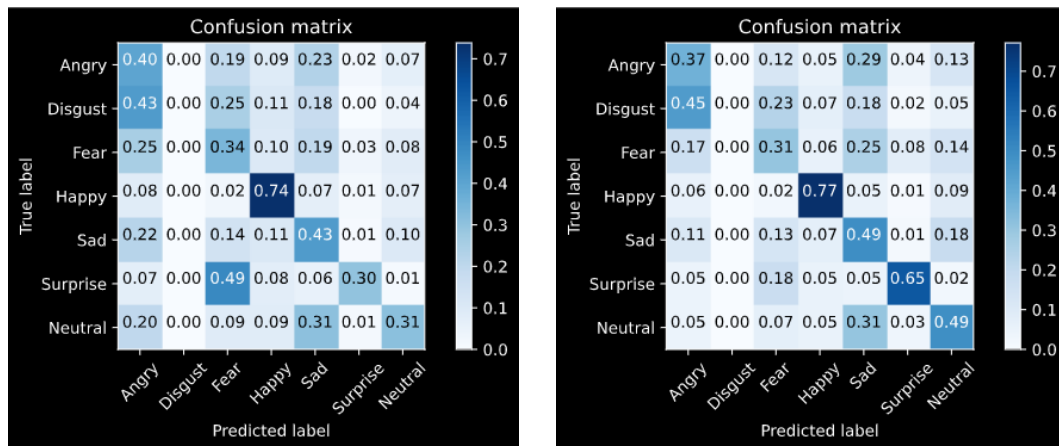


Gambar 4.20 Performa Model GCNs Menggunakan Fitur N + M

neutral meningkat sedikit berturut-turut sebesar 1%, 3%, 1% dan 6%. Sementara itu, model E + M hanya memberikan peningkatan yang sedikit pada emosi *fear*, *happy* dan *sad* berturut-turut sebesar 5%, 1% dan 2%. Namun mengalami penurunan sebesar 7% untuk emosi *surprise* dan *neutral*.

Hal ini mungkin terjadi sebagai akibat dari sangat jeleknya akurasi model pada set data bagian mata dan hidung secara terpisah, terutama untuk bagian mata. Sehingga meskipun akurasi model dihitung secara proporsional, performa model menjadi sedikit menurun. Fenomena penurunan performa ini diperkuat oleh perbandingan hasil perhitungan akurasi menggunakan dua teknik yang berbeda, di mana perhitungan *simple average* yang tidak proporsional hampir selalu memberikan nilai akurasi yang lebih rendah dari *weighted average*. Sementara penggabungan fitur bagian mata dan hidung menghasilkan performa yang tidak bisa diterima.

Selanjutnya, model EN + NM yang dilatih menggunakan *network ensemble* pun tidak memiliki keunggulan yang menarik dibandingkan model-model yang sebelumnya. Malahan model E + N + M yang dilatih pada *network ensemble* sangat mengecewakan dalam mengenali emosi *angry*. Namun satu hal yang membeda-



(a) *Confusion Matrix* Performa Model E + M Berdasarkan Perhitungan *Simple Average*

(b) *Confusion Matrix* Performa Model E + M Berdasarkan Perhitungan *Weighted Average*

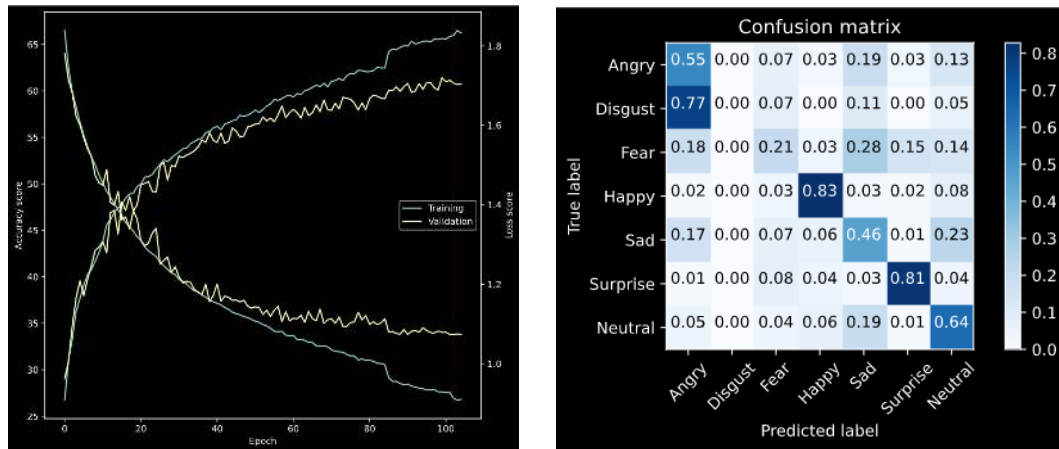
Gambar 4.21 Performa Model GCNs Menggunakan Fitur E + M

kannya dari semua model sebelumnya adalah kebanyakan data tes berlabel emosi *disgust* dikenali sebagai *fear*. Hal ini membuka sedikit peluang peningkatan performa model dalam mengenali *disgust*, yaitu bahwa sebetulnya tidak benar jika *disgust* ini tidak dapat dibedakan dari *angry*.

Sebagai pertimbangan lain, pada skenario ini dilakukan *training* pada set data hasil *cropping* area wajah penuh, yaitu set data yang disimpan sesaat sebelum masuk ke dalam proses FRS. Sayangnya, teknik *cropping* ini ternyata tidak membuat performa model lebih baik ketimbang tanpa *cropping*. Di sisi lain, tanpa harus mengorbankan performa yang cukup berarti, teknik ini lebih disukai daripada pemodelan menggunakan teknik FRS, sebab membutuhkan waktu *training* yang relatif lebih cepat.

4.3.4 Modifikasi Log-GCN Menjadi *Ensemble* Log-GCNs

Pada eksperimen ini, penulis mencoba melakukan modifikasi GCN melalui perubahan filter Gabor yang dimanfaatkan dalam proses ekstraksi fitur menjadi filter log-Gabor. Namun sebelum mengeksekusi skenario eksperimen terakhir ini, pe-



(a) Grafik Performa Per *Epoch* untuk Model EN + NM (*Concat.*)

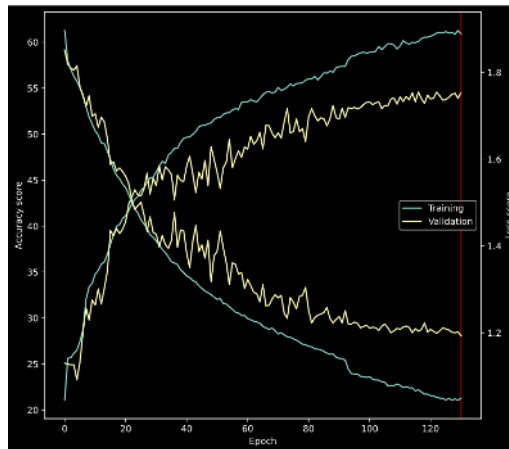
(b) *Confusion Matrix* Performa Model EN + NM (*Concat.*)

Gambar 4.22 Performa Model GCNs Menggunakan Fitur EN + NM (*Concat.*)

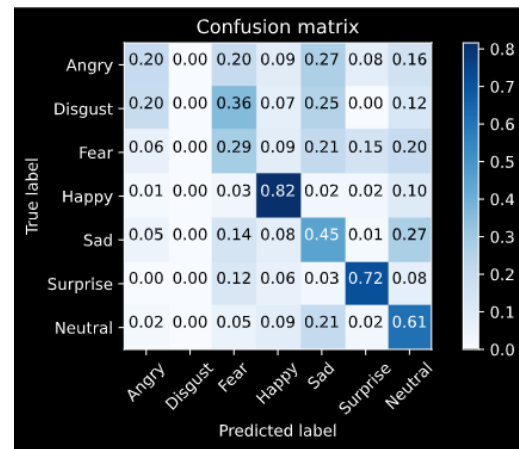
nulis ingin melihat apakah pengubahan filter tersebut dapat memberikan dampak terhadap kinerja model GCN, sehingga diputuskan bahwa penulis kembali menggunakan set data primer.

Konfigurasi parameter filter log-Gabor yang dipakai pada eksperimen ini dibuat menyerupai konfigurasi filter Gabor terbaik yang sebelumnya. Bedanya, untuk filter log-Gabor, penulis menentukan frekuensi spasial $sf_0 = 0,0075/scale$ sebagai ganti atas variabel *scale* pada filter Gabor. Di mana untuk lapisan konvolusi ke-8 ke bawah menggunakan $scale = 1,25$, sedangkan sisanya menggunakan $scale = 1,0$. Perbedaan yang mendasar dari kedua filter ini ditunjukkan pada Gambar 4.25. Di mana bagian daerah dengan frekuensi spasial yang rendah (daerah berwarna hitam) pada filter log-Gabor cenderung cembung mengarah ke dalam. Sedangkan pada filter Gabor, daerah ini cenderung cembung mengarah ke luar.

Hasil dari eksperimen pada skenario ini ditampilkan pada Tabel 4.5 di mana terbukti bahwa model log-GCN memiliki performa yang lebih baik daripada GCN dengan perbaikan yang cukup jelas pada data berlabel *sad* (Gambar 4.27). Untuk mendapatkan peningkatan sebesar 1,62% pada akurasi *testing*, model log-GCN memerlukan waktu *training* 1,05 kali lebih lama. Model ini menjadi model

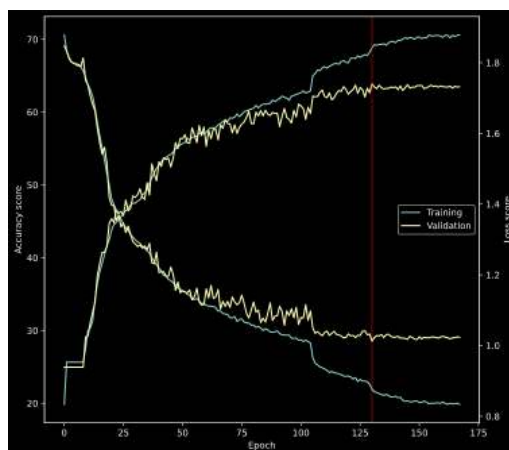


(a) Grafik Performa Per *Epoch* untuk Model E + N + M (*Concat.*)

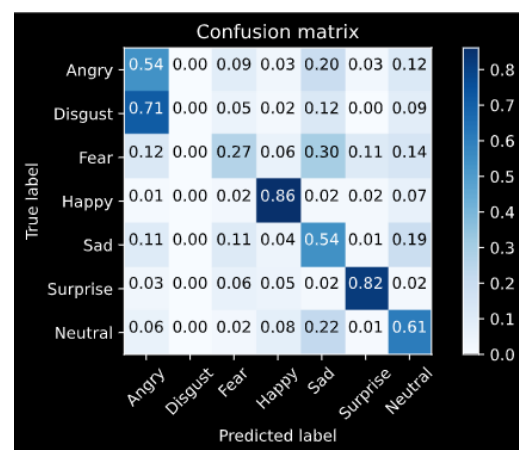


(b) *Confusion Matrix* Performa Model E + N + M (*Concat.*)

Gambar 4.23 Performa Model GCNs Menggunakan Fitur E + N + M (*Concat.*)



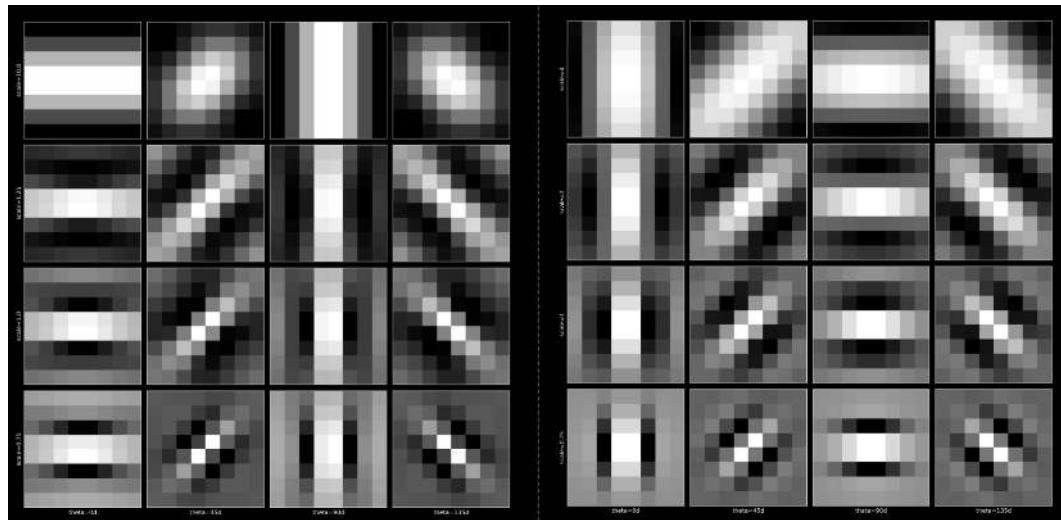
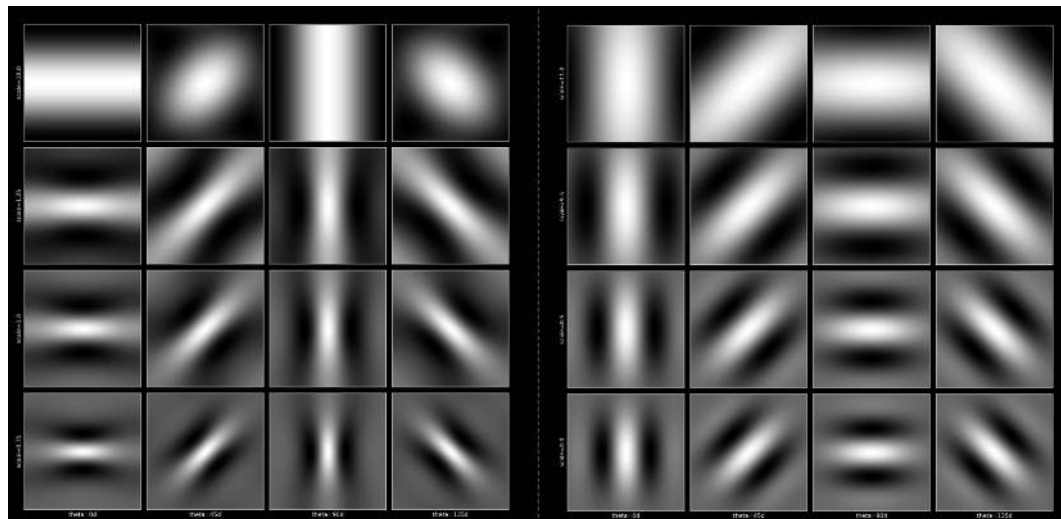
(a) Grafik Performa Per *Epoch* untuk Model ENM



(b) *Confusion Matrix* Performa Model ENM

Gambar 4.24 Performa Model GCNs Menggunakan Fitur ENM

dengan performa terbaik di atas semua skenario pemodelan pada penelitian ini. Peningkatan akurasi ini bukan akibat dari kebetulan melakukan *training* pada kondisi mesin sedemikian rupa. Sebab berdasarkan dari percobaan tiga kali *training* untuk masing-masing skenario utama pada penelitian ini, biasanya perbedaan akurasi antara model paling buruk dan paling baik berkisar di bawah angka 1,0%. Kemudian yang amat disayangkan di sini adalah model yang dilatih hingga saat ini masih belum mampu untuk mengenali data berlabel *disgust* dengan benar sama sekali se-

(a) Filter log-Gabor Versus Gabor pada 8×8 Piksel(b) Filter log-Gabor Versus Gabor pada 100×100 Piksel

Gambar 4.25 Beberapa Contoh Perbedaan Filter log-Gabor (Kanan) dan Gabor (Kiri)

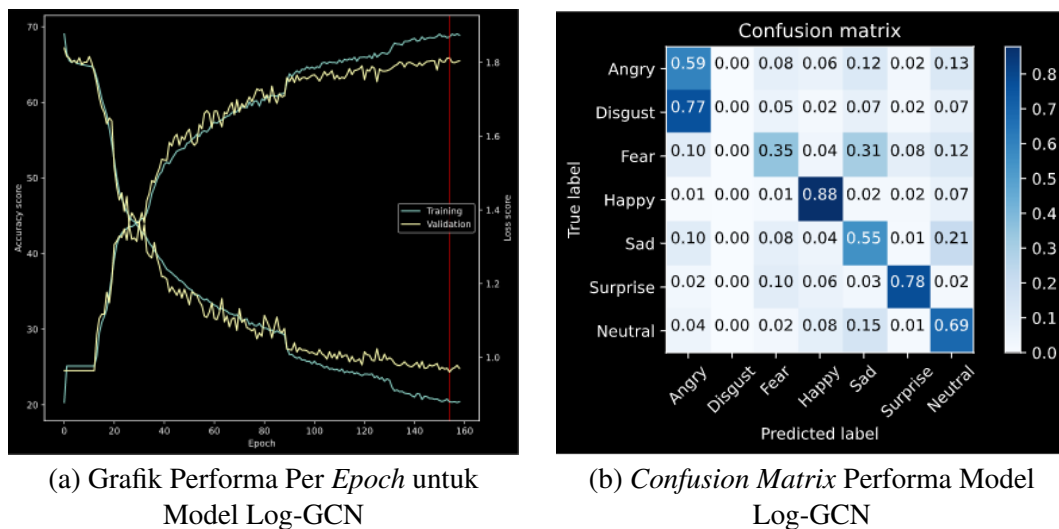
bagaimana yang ditunjukkan oleh *confusion matrix* pada Gambar 4.26b.

Setelah mengetahui bahwa pengubahan filter Gabor menjadi log-Gabor berhasil meningkatkan performa model, penulis melakukan modifikasi log-GCN menjadi *ensemble* log-GCNs, yaitu pengenalan teknik FRS terhadap arsitektur log-GCN. Dari berbagai skenario implementasi teknik FRS pada Tabel 4.4, penulis mengangkat tiga pendekatan yang menghasilkan kinerja model terbaik untuk dilakukan kembali

Tabel 4.5
Perbandingan Performa Model GCN dan Log-GCN

	Akurasi (%) [△]	Epoch	Waktu (jam) [▽]
Model GCN	63,51	150	7,95
Model Log-GCN	65,13	155	8,38

Akurasi—akurasi model pada *testing*; *Epoch*—banyak *epoch* yang dibutuhkan agar model menjadi optimal; Waktu—waktu yang dibutuhkan agar model menjadi optimal; [△]Lebih tinggi lebih baik; [▽]Lebih rendah lebih baik.

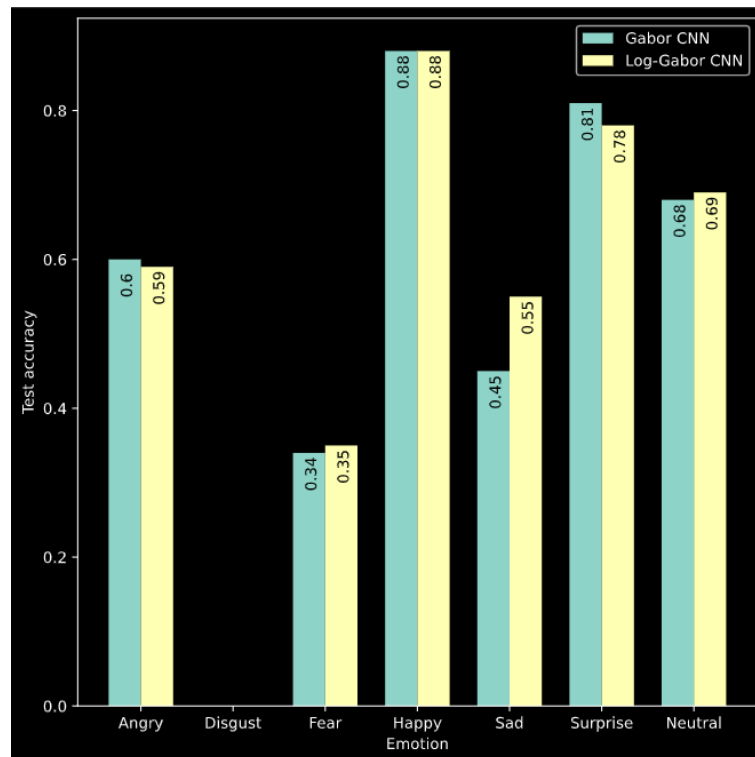


Gambar 4.26 Performa Model Log-GCN

pada eksperimen tahap ini. Hasilnya diperlihatkan pada Tabel 4.6.

Dari tiga pendekatan FRS ini yang dilakukan pada arsitektur log-GCN, seluruhnya membutuhkan waktu *training* yang relatif lebih cepat dalam *epoch* yang relatif lebih sedikit dan menghasilkan akurasi pengenalan yang lebih tinggi daripada sebelumnya. Adapun kinerja yang diperoleh oleh tiga pendekatan tersebut memiliki urutan yang sama seperti sebelumnya. Yaitu kinerja terbaik berhasil dicapai oleh model EN + NM yang menggunakan teknik *ensemble* konvensional. Dilanjutkan dengan model EN + NM yang dilatih pada *network ensemble* dan model ENM berturut-turut menempati urutan kedua dan ketiga.

Berdasarkan *confusion matrix* pada Gambar 4.28b dan 4.28d, terlihat bahwa model EN memiliki akurasi yang lebih tinggi pada pengenalan emosi *angry* dengan



Gambar 4.27 Perbandingan Performa Model GCN dan Log-GCN Per Kelas Emosi

selisih 14% dan *sad* dengan selisih 5%. Ini kembali menekankan bahwa emosi *angry* lebih mudah dikenali menggunakan bagian wajah atas manusia. Adapun model NM relatif cukup lebih baik pada sisanya. Kecuali untuk *disgust*, model EN kembali lebih banyak melakukan kesalahan prediksi sebagai *angry*. Namun kali ini, selisihnya hanya sebesar 2%. Untuk emosi *fear*, model ini jauh lebih baik dalam pengenalan emosi tersebut dengan kembali lagi model NM memiliki akurasi yang lebih baik ketimbang model EN. Secara umum, sebesar 2,01% performa *testing* untuk model EN + NM berhasil ditingkatkan.

Model EN + NM memberikan hasil perhitungan dua teknik rerata berbeda dalam *confusion matrix* yang tepat sama seperti yang tampak pada Gambar 4.28e dan 4.28f. Untuk setiap label emosi, kombinasi kedua model ini mampu memberikan nilai akurasi yang sedikit lebih baik dibandingkan model NM. Namun untuk label

Tabel 4.6
Perbandingan Performa Berbagai Kombinasi Model Ensemble Log-GCNs

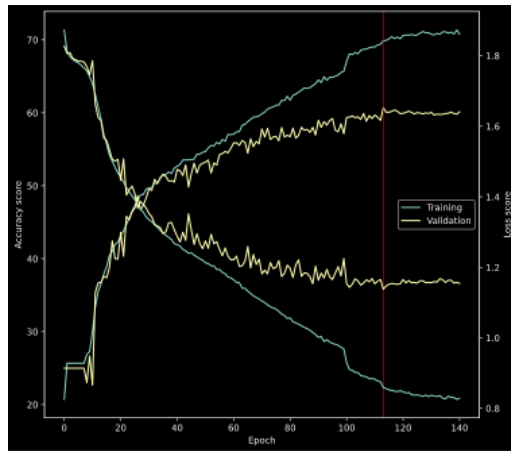
Fitur	Waktu (jam)▽		Epoch	Akurasi Testing (%)△		
				Single	Simple Average	Weighted Average
EN + NM	3,71	6,91	114	59,23	64,21	64,21
	3,20		102	59,83		
EN + NM (concat.)	8,69		136	61,27	-	
ENM	4,34		86	63,07		
EN + NM + ENM	11,25		Ibid.	Ibid.	66,25	66,34

E—(Eyes) area bagian kedua mata; N—(Nose) area bagian hidung; M—(Mouth) area bagian mulut; Epoch—banyak epoch yang dibutuhkan agar model menjadi optimal; Waktu—waktu yang dibutuhkan agar model menjadi optimal; [△]Lebih tinggi lebih baik; [▽]Lebih rendah lebih baik.

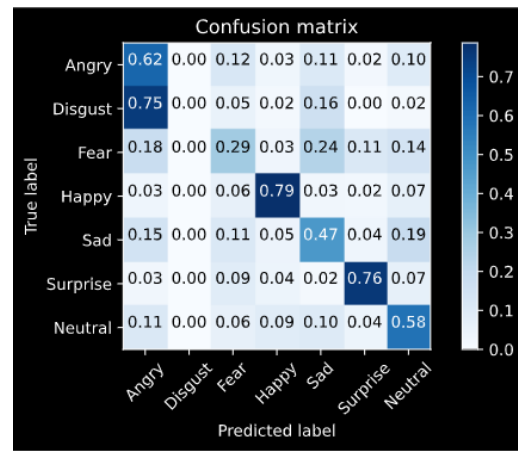
angry, akurasinya turun sebesar 1% terhadap model EN. Sementara penyimpangan prediksi emosi *disgust* relatif cukup lebih besar daripada model EN dan NM secara terpisah.

Model EN + NM yang dilatih menggunakan *network ensemble* memiliki akurasi rekognisi yang relatif lebih rendah terhadap model di atas untuk semua kelas emosi kecuali *sad*. Selain dari peningkatan akurasi 2% untuk emosi *sad*, tidak ditemukan adanya hal-hal yang menarik untuk dibahas baik dari log performa model per *epoch* pada Gambar 4.29a maupun dari *confusion matrix* pada Gambar 4.29b. Bahkan jika dibandingkan dengan model EN + NM yang menggunakan GCN sebelumnya, tidak ada peningkatan yang cukup signifikan. Meskipun secara keseluruhan, akurasinya meningkat sebesar 1,56%.

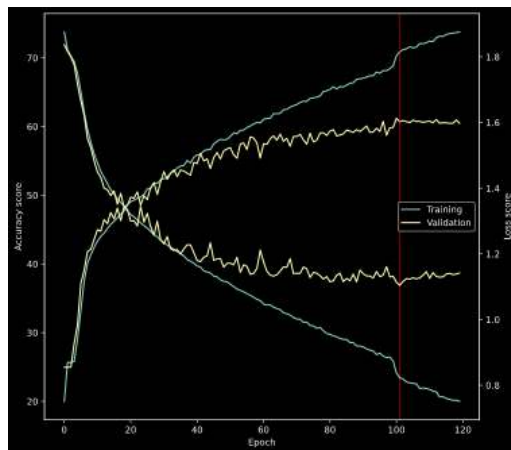
Model ENM pada eksperimen ini hanya mendapatkan peningkatan akurasi sebesar 0,99% terhadap eksperimen sebelumnya. Meskipun demikian, akurasi pengenalan emosi *fear* pada model ini meningkatkan sebesar 8% sebagaimana yang terlihat dalam *confusion matrix* pada Gambar 4.30b. Performa tersebut dapat diperoleh hanya dalam *epoch* yang relatif jauh lebih sedikit daripada model yang pernah dilatih pada penelitian ini. Dari informasi tersebut, penulis merasa bila masih ter-



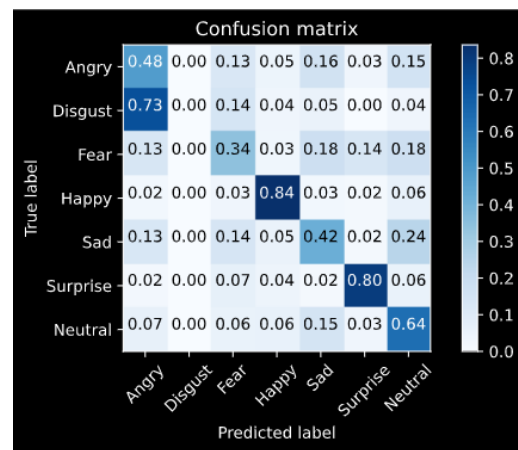
(a) Grafik Performa Per *Epoch* untuk Model EN



(b) *Confusion Matrix* Performa Model EN



(c) Grafik Performa Per *Epoch* untuk Model NM

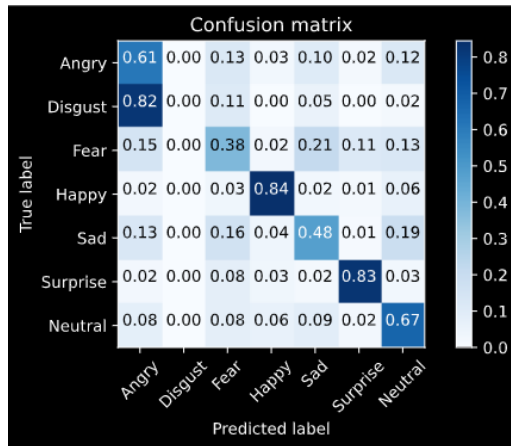


(d) *Confusion Matrix* Performa Model NM

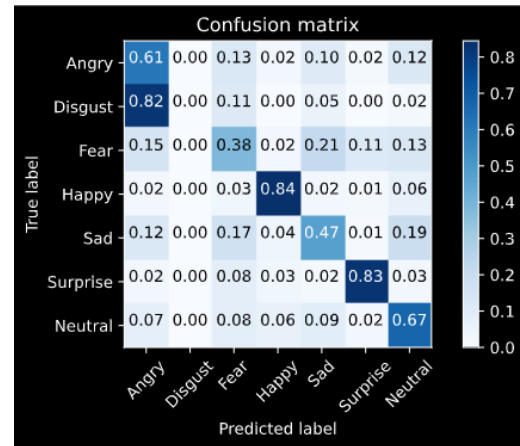
Gambar 4.28 Performa Model Log-GCNs Menggunakan Fitur EN + NM

dapat peluang untuk meningkatkan akurasi yang diperoleh.

Pada eksperimen ini, penulis juga mencoba melakukan penggabungan terhadap dua pendekatan yang berbeda, yaitu pendekatan holistik dan parsial. Di mana model prediksi ENM dan EN + NM berturut-turut merupakan representasi dari pendekatan holistik dan parsial. Alhasil, model yang dihasilkan mampu mengungguli setiap model yang pernah ada pada penelitian ini dengan peningkatan relatif terbesar pada akurasi pengenalan emosi *fear* dan *neutral*. Sementara itu, kesalahan pengenalan emosi *disgust* sebagai *angry* juga sangat memprihatinkan. Sebagaimana yang

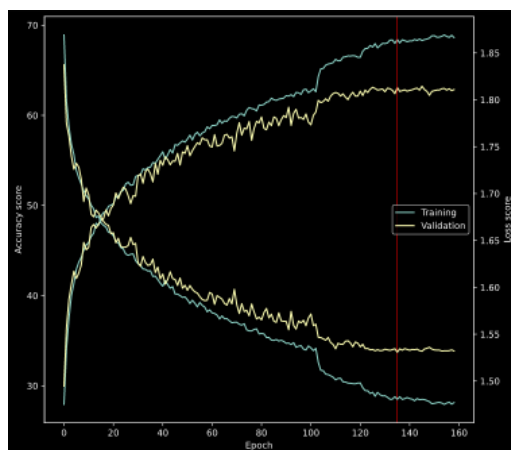


(e) *Confusion Matrix* Performa Model EN + NM Berdasarkan Perhitungan *Simple Average*

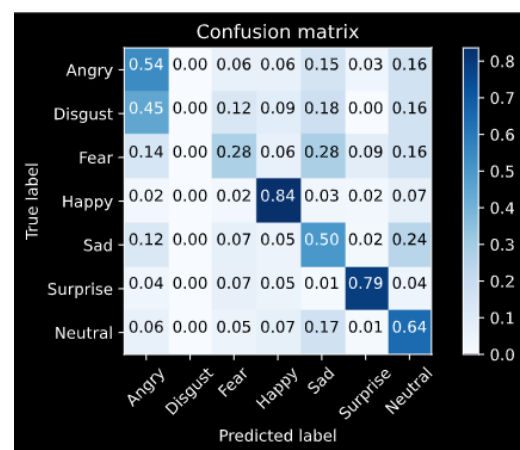


(f) *Confusion Matrix* Performa Model EN + NM Berdasarkan Perhitungan *Weighted Average*

Gambar 4.28 Performa Model Log-GCNs Menggunakan Fitur EN + NM (Lanjutan)



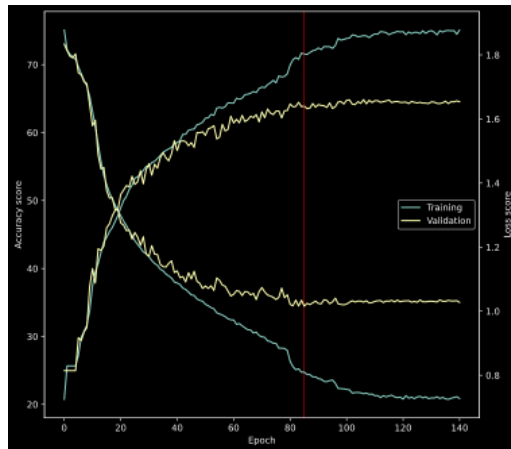
(a) Grafik Performa Per *Epoch* untuk Model EN + NM (*Concat.*)



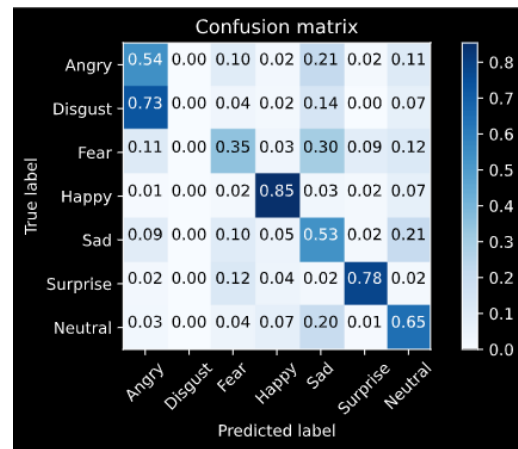
(b) *Confusion Matrix* Performa Model EN + NM (*Concat.*)

Gambar 4.29 Performa Model Log-GCNs Menggunakan Fitur EN + NM (*Concat.*)

tampak melalui *confusion matrix* pada Gambar 4.30c dan 4.30d, perhitungan dua teknik *averaging* tersebut menghasilkan hasil yang hampir serupa. Mungkin sifat ini diturunkan dari sifat model EN + NM itu sendiri. Pencapaian ini membuktikan bahwa pengenalan menggunakan kombinasi data holistik dan parsial lebih akurat ketimbang menggunakan salah satu dari kedua data tersebut. Secara umum, per-

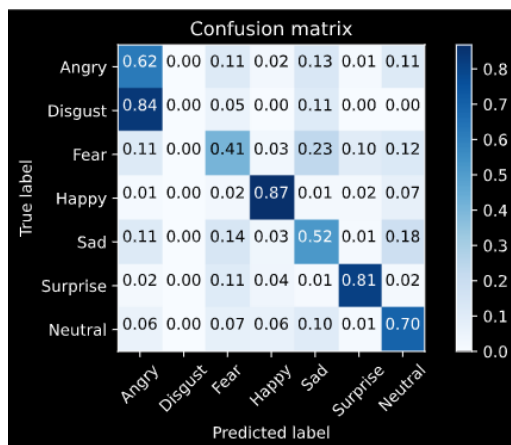


(a) Grafik Performa Per Epoch untuk Model EN + NM (Concat.)

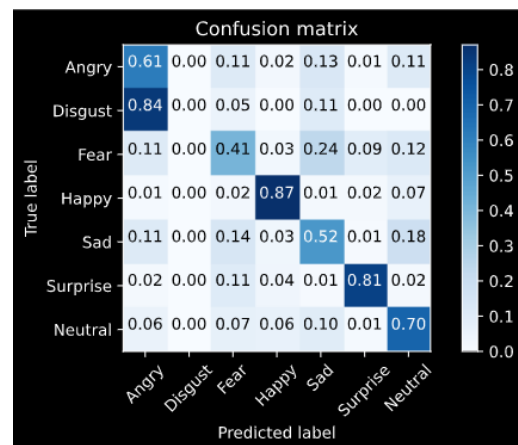


(b) Confusion Matrix Performa Model EN + NM (Concat.)

Gambar 4.30 Performa Model Log-GCNs Menggunakan Fitur EN + NM (Concat.)



(c) Confusion Matrix Performa Model EN + NM + ENM Berdasarkan Perhitungan Simple Average



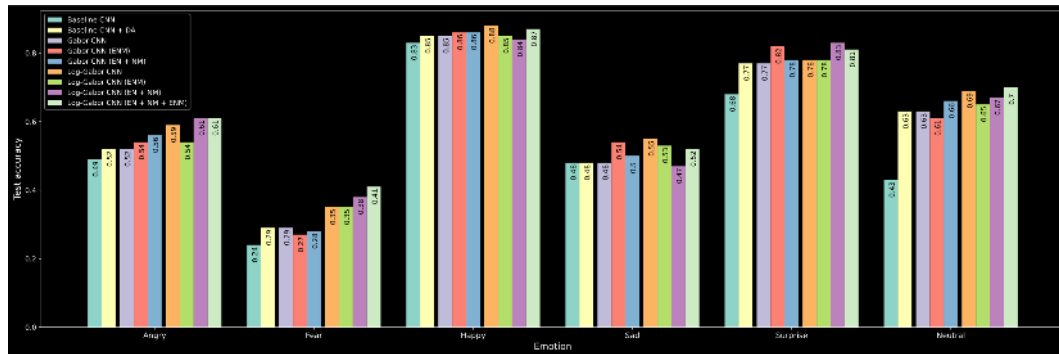
(d) Confusion Matrix Performa Model EN + NM + ENM Berdasarkan Perhitungan Weighted Average

Gambar 4.30 Performa Model Log-GCNs Menggunakan Fitur EN + NM + ENM

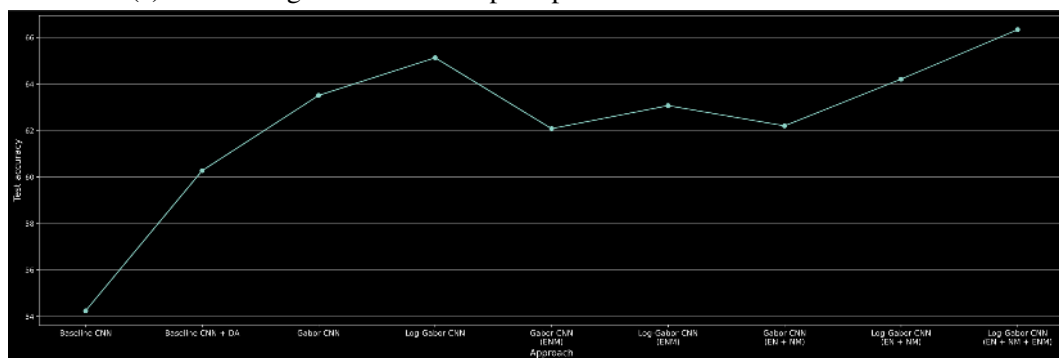
bandingan performa tiap-tiap model terbaik dari seluruh skenario eksperimen yang dilakukan per kelas emosi ditunjukkan pada Gambar 4.31.

4.4 Evaluasi Akhir

Analisis secara kuantitatif telah dibahas dengan cukup detail pada bagian-bagian sebelumnya. Oleh karena itu, pada bagian ini akan memaparkan analisis secara ku-



(a) Perbandingan Performa Tiap-Tiap Model Terbaik Per Kelas Emosi

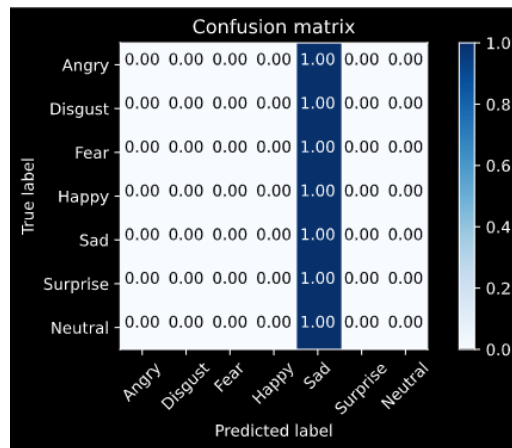


(b) Perbandingan Performa Tiap-Tiap Model Terbaik Secara Keseluruhan

Gambar 4.31 Perbandingan Performa Tiap-Tiap Model Terbaik dari Seluruh Skenario Eksperimen

alitatif mengenai peningkatan model terbaik terhadap model CNN *baseline* yang menggunakan augmentasi data. Penulis berpikir bahwa evaluasi akhir yang terbaik dapat dilakukan melalui penyelidikan kombinasi *facial action unit* per kelas emosi yang berhasil dan tidak berhasil dikenali dengan benar oleh model terbaik secara saksama terhadap beberapa sampel data tes yang diberikan pada Gambar 4.33–4.44. Penyelidikan yang dimaksud di sini mengacu kepada P. Ekman dkk. (2002) dan L. Zhang dkk. (2018). Namun, penulis merasa terlalu berlebihan jika evaluasi dilakukan dengan cara demikian. Sebab model terbaik yang berhasil diperoleh secara umum memiliki kinerja yang belum cukup dapat diandalkan, baik diukur secara kuantitatif maupun secara kualitatif.

Pada sistem pengenalan ekspresi wajah di dunia nyata, di mana biasanya kamera diletakkan pada posisi dan sudut yang mampu menjangkau banyak wajah



Gambar 4.32 *Confusion Matrix* Performa Model CNN *Baseline* dengan Augmentasi Data pada Set Data Wajah Penuh

manusia di sebuah tempat kerumunan, set data wajah yang akan diprediksi dalam waktu-waktu tertentu diakuisisi menggunakan teknik deteksi wajah. Sehingga model terbaik yang diusulkan pada penelitian ini lebih cocok diterapkan daripada model *baseline*. Sebab model yang diusulkan dilatih menggunakan set data wajah hasil dari proses deteksi wajah yang serupa. Sementara model *baseline* dilatih menggunakan set data wajah yang juga mengandung selain dari area wajah, sehingga sejumlah informasi yang tidak diperlukan juga ikut terlibat dalam pelatihan model. Dibuktikan dengan pengujian model *baseline* menggunakan set data wajah penuh hanya memberikan nilai akurasi 17,14% dengan *confusion matrix* yang tertera pada Gambar 4.32.

Selain model yang diusulkan mampu memberikan nilai akurasi yang relatif lebih baik baik pada pengenalan secara umum maupun pada masing-masing kelas emosi kecuali *disgust*, model tersebut juga memberikan nilai *precision* dan *recall* yang secara umum relatif lebih baik ketimbang model *baseline*. Secara berturut-turut, *precision* yang diberikan oleh model *baseline*, yaitu 0,365, 0,359, 0,658, 0,426, 0,801, dan 0,457 untuk emosi *angry*, *fear*, *happy*, *sad*, *surprise* dan *neutral*. Sedangkan *precision* yang diberikan oleh model yang diusulkan secara berturut-

Tabel 4.7
*Perbandingan Performa Prediksi Model CNN Baseline dan Log-GCN
 (EN + NM + ENM) pada Set Data Tes FER-2013*

	Memori (MiB)	Waktu (sekon)
Model CNN <i>Baseline</i>	577	1,55
Model Log-GCN (EN + NM + ENM)	1.523	15,98

Waktu yang diukur adalah waktu prediksi seluruh data tes FER-2013 menggunakan spesifikasi komputer dan sistem dalam penelitian ini.

turut, yaitu 0,349, 0,446, 0,829, 0,466, 0,850, dan 0,579. Adapun *precision* untuk emosi *disgust* tidak bisa dihitung karena memberikan perhitungan yang tidak terdefinisi. Jelas terlihat bahwa selain untuk emosi *angry*, *precision* yang diberikan oleh model yang diusulkan relatif lebih baik ketimbang model *baseline*. Penurunan nilai *precision* yang diberikan oleh model yang diusulkan terhadap model *baseline* pada prediksi emosi *angry* terjadi akibat meningkatnya kesalahan prediksi emosi *disgust* menjadi *angry*. Jika tanpa melibatkan kelas *disgust*, perhitungan rata-rata dari *precision* untuk model *baseline* dan model yang diusulkan secara berturut-turut menghasilkan 0,511 dan 0,586.

Model yang diusulkan membutuhkan sumber daya komputasi yang relatif lebih besar, baik ketika *training* maupun prediksi, ketimbang model *baseline*. Di luar dari waktu *training* yang lama sebagaimana yang telah dipaparkan sebelumnya, performa kedua model tersebut dalam memprediksi set data tes FER-2013 berjumlah 3.331 data gambar wajah menggunakan spesifikasi komputer dan sistem yang telah disebutkan dalam penelitian ini tertera pada Tabel 4.7. Tampak bahwa model yang diusulkan memerlukan konsumsi memori yang lebih boros tiga kali lipat dan waktu prediksi yang lebih lama sepuluh kali lipat. Data tabel tersebut tidak bisa dijadikan sebagai acuan dalam estimasi perhitungan kecepatan prediksi dalam *frame per second* (FPS) ketika produksi, di mana berbagai macam faktor terlibat seperti ukuran gambar input, kecepatan algoritma *face detection*, kecepatan algoritma FRS dan lain-lain.

Emosi berlabel *disgust* tidak mampu sama sekali dikenali dengan benar ba-

ik oleh model *baseline* maupun oleh model yang diusulkan. Bahkan model yang diusulkan lebih banyak salah memprediksi emosi tersebut sebagai *angry*. Meskipun kesalahan prediksi ini telah menghancurkan angka presisi dalam pengenalan emosi *angry*, penulis tidak melihat signifikansi dalam hal ini yang mengakibatkan menurunnya performa model secara umum dalam memprediksi setiap emosi terutama *angry* dan *disgust* dengan benar. Terlebih lagi, kesalahan prediksi data berlabel *disgust* menjadi *angry* masih mendapatkan toleransi. Widen & Russell (2013) menyimpulkan bahwa mayoritas anak-anak pada usia delapan tahun menyatakan bahwa ekspresi wajah *disgust* mengindikasikan emosi *angry*. Dan hanya setengah dari anak-anak berumur sekitar sembilan tahun yang menganggap bahwa ekspresi wajah *disgust* mengindikasikan emosi *disgust*. Bahkan sebagian dari orang-orang dewasa pun salah mengartikan ekspresi wajah *disgust* sebagai *angry* (Widen & Russell, 2010). Setengah dari mereka justru memasukkan ekspresi wajah *disgust* sebagai emosi *angry* dan seperempat dari mereka memasukkan *angry* sebagai *disgust* (Widen dkk., 2004). Hal ini membuktikan bahwa sebenarnya orang-orang mempertanyakan mengenai perbedaan emosi *disgust* dan *angry* itu sendiri. Bagaimana pun juga, kedua emosi tersebut sama-sama tergolong ke dalam kelompok emosi negatif. Menurut orang-orang yang memiliki kecemasan sosial, emosi *disgust* dinilai lebih negatif daripada *angry* (Amir dkk., 2010).

Secara umum, akurasi model yang diusulkan dapat mengungguli akurasi model *baseline* seperti yang telah dibahas sebelumnya. Dengan akurasi yang diperoleh tersebut, model yang diusulkan berhasil mencapai *human accuracy* pada set data FER-2013, yaitu $65 \pm 5\%$ (Goodfellow dkk., 2013). Namun, jika ditinjau dari kemampuan rekognisi per label emosi, model hanya mendapatkan *human accuracy* untuk label *angry*, *happy*, *surprise* dan *neutral*.



Gambar 4.33 Beberapa Contoh Kesalahan Prediksi Emosi *Angry* oleh Model Terbaik (Label Prediksi | Label Sebenarnya)



Gambar 4.34 Beberapa Contoh Kesalahan Prediksi Emosi *Fear* oleh Model Terbaik (Label Prediksi | Label Sebenarnya)



Gambar 4.35 Beberapa Contoh Kesalahan Prediksi Emosi *Happy* oleh Model Terbaik (Label Prediksi | Label Sebenarnya)



Gambar 4.36 Beberapa Contoh Kesalahan Prediksi Emosi *Sad* oleh Model Terbaik
(Label Prediksi | Label Sebenarnya)



Gambar 4.37 Beberapa Contoh Kesalahan Prediksi Emosi *Surprise* oleh Model Terbaik (Label Prediksi | Label Sebenarnya)



Gambar 4.38 Beberapa Contoh Kesalahan Prediksi Emosi *Neutral* oleh Model Terbaik (Label Prediksi | Label Sebenarnya)



Gambar 4.39 Beberapa Contoh Keberhasilan Prediksi Emosi *Angry* oleh Model Terbaik



Gambar 4.40 Beberapa Contoh Keberhasilan Prediksi Emosi *Fear* oleh Model Terbaik



Gambar 4.41 Beberapa Contoh Keberhasilan Prediksi Emosi *Happy* oleh Model Terbaik



Gambar 4.42 Beberapa Contoh Keberhasilan Prediksi Emosi *Sad* oleh Model Terbaik



Gambar 4.43 Beberapa Contoh Keberhasilan Prediksi Emosi *Surprise* oleh Model Terbaik



Gambar 4.44 Beberapa Contoh Keberhasilan Prediksi Emosi *Neutral* oleh Model Terbaik

BAB V

KESIMPULAN DAN SARAN

Setelah berhasil mengeksekusi semua skenario yang telah diusulkan, berikut ini beberapa kesimpulan yang dirangkum untuk menjawab permasalahan utama yang diangkat pada penelitian ini.

1. Pemodelan pengenalan emosi menggunakan Log-GCNs melalui pendekatan FRS telah berhasil dirancang dan dieksekusi untuk set data wajah nonfrontal, FER-2013.
2. Dibuktikan bahwa arsitektur GCN, sebagai hasil modifikasi dari arsitektur *baseline*, baik dengan maupun tanpa FRS dapat meningkatkan performa model *baseline*.
3. Dibuktikan bahwa modifikasi instruksi pada arsitektur GCNs melalui penggantian filter Gabor menjadi filter Log-Gabor dapat meningkatkan performa model.

Menurut pengamatan kami, metode yang diusulkan pada penelitian ini masih dapat dikembangkan lebih baik lagi dengan mengikuti beberapa saran berikut ini.

1. Beberapa peluang peningkatan metode yang dapat diterapkan di antaranya adalah —namun tidak terbatas pada— 1) penggunaan teknik *facial landmark detection* yang lebih andal untuk set data wajah nonfrontal, 2) penggunaan teknik *image enhancement* yang andal terutama untuk data gambar wajah yang *underbright* atau *overbright*, 3) pemanfaatan teknik *upsampling* terutama untuk data berlabel *disgust*, dan 4) penggunaan teknik *network architecture search* untuk menemukan konfigurasi *hyperparameter* dan arsitektur jaringan yang lebih cocok untuk konteks pengenalan emosi menggunakan set data FER-2013.

2. Perbaiki algoritma dan praproses pada teknik FRS, terutama untuk area wajah bagian atas (dari mata hingga hidung). Sebab pengenalan emosi pada set data wajah dengan sebagian area yang tertutup sesuatu seperti masker akan cukup relevan diterapkan pada masa pandemi COVID-19 saat ini. Salah satu teknik yang dapat diadopsi adalah *face frontalization*.

DAFTAR PUSTAKA

- Abramson, C. I., & Chicas-Mosier, A. M. (2016). Learning in plants: lessons from *mimosa pudica*. *Frontiers in psychology*, 7, 417.
- Adil, B., Nadjib, K. M., & Yacine, L. (2019). A novel approach for facial expression recognition. In *2019 international conference on networking and advanced systems (icnas)* (pp. 1–5).
- Aggrawal, A. (2016). Fingerprint authentication system using log-gabor filter. *International Journal of Computer Applications*, 145(9).
- Agrawal, A., & Mittal, N. (2020). Using cnn for facial expression recognition: a study of the effects of kernel size and number of filters on accuracy. *The Visual Computer*, 36(2), 405–412.
- Amayeh, G., Tavakkoli, A., & Bebis, G. (2009). Accurate and efficient computation of gabor features in real-time applications. In *International symposium on visual computing* (pp. 243–252).
- Amici, F., Waterman, J., Kellermann, C. M., Karimullah, K., & Bräuer, J. (2019). The ability to recognize dog emotions depends on the cultural milieu in which we grow up. *Scientific reports*, 9(1), 1–9.
- Amir, N., Najmi, S., Bomyea, J., & Burns, M. (2010). Disgust and anger in social anxiety. *International Journal of Cognitive Therapy*, 3(1), 3–10.
- An, H., Ehsan, M. A., Zhou, Z., & Yi, Y. (2017). Electrical modeling and analysis of 3d synaptic array using vertical rram structure. In *2017 18th international symposium on quality electronic design (isqed)* (pp. 1–6).
- Arnold, M. B. (1960). Emotion and personality.
- Assari, M. A., & Rahmati, M. (2011). Driver drowsiness detection using face expression recognition. In *2011 ieee international conference on signal and image processing applications (icsipa)* (pp. 337–341).
- Avots, E., Sapiński, T., Bachmann, M., & Kamińska, D. (2019). Audiovisual

- emotion recognition in wild. *Machine Vision and Applications*, 30(5), 975–985.
- Baali, M., & Ghneim, N. (2019). Emotion analysis of arabic tweets using deep learning approach. *Journal of Big Data*, 6(1), 89.
- Baati, K., & Kanoun, S. (2018). Towards a hybrid system for the identification of arabic and latin scripts in printed and handwritten natures. In *International conference on hybrid intelligent systems* (pp. 294–301).
- Barsoum, E., Zhang, C., Canton Ferrer, C., & Zhang, Z. (2016). Training deep networks for facial expression recognition with crowd-sourced label distribution. In *Acm international conference on multimodal interaction (icmi)*.
- Baumeister, R. F., Vohs, K. D., Nathan DeWall, C., & Zhang, L. (2007). How emotion shapes behavior: Feedback, anticipation, and reflection, rather than direct causation. *Personality and social psychology review*, 11(2), 167–203.
- Beccue, M., & Kaul, A. (2018). *Executive summary emotion recognition and sentiment analysis enterprise use cases for customer service, product/market research, customer experience, healthcare, automotive, education, and gaming*.
- Benitez-Garcia, G., Nakamura, T., & Kaneko, M. (2017). Analysis of differences between western and east-asian faces based on facial region segmentation and pca for facial expression recognition. In *Aip conference proceedings* (Vol. 1807, p. 020025).
- Benitez-Garcia, G., Nakamura, T., & Kaneko, M. (2018). Multicultural facial expression recognition based on differences of western-caucasian and east-asian facial expressions of emotions. *IEICE TRANSACTIONS on Information and Systems*, 101(5), 1317–1324.
- Beran, M. (2020). Animal learning and cognition. In *Oxford research encyclopedia of psychology*.
- Bradley, M. M., & Lang, P. J. (2007). Emotion and motivation.
- Bre, F., Gimenez, J. M., & Fachinotti, V. D. (2018). Prediction of wind pressure

- coefficients on building surfaces using artificial neural networks. *Energy and Buildings*, 158, 1429–1441.
- Bulat, A., & Tzimiropoulos, G. (2017). How far are we from solving the 2d & 3d face alignment problem? (and a dataset of 230,000 3d facial landmarks). In *International conference on computer vision*.
- Canedo, D., & Neves, A. J. (2019). Facial expression recognition using computer vision: A systematic review. *Applied Sciences*, 9(21), 4678.
- Chatfield, K., Simonyan, K., Vedaldi, A., & Zisserman, A. (2014). Return of the devil in the details: Delving deep into convolutional nets. *arXiv preprint arXiv:1405.3531*.
- Cobley, P. (2008). Communication: Definitions and concepts. *The international encyclopedia of communication*.
- Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural language processing (almost) from scratch. *Journal of machine learning research*, 12(Aug), 2493–2537.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273–297.
- Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., & Taylor, J. G. (2001). Emotion recognition in human-computer interaction. *IEEE Signal processing magazine*, 18(1), 32–80.
- Dahmane, M., Cossette, S., & Meunier, J. (2015). Conditional gabor phase-based disparity estimation applied to facial tracking for person-specific facial action recognition: a preliminary study. *Multimedia Tools and Applications*, 74(17), 7111–7130.
- Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *JOSA A*, 2(7), 1160–1169.
- De Houwer, J., Barnes-Holmes, D., & Moors, A. (2013). What is learning? on the

- nature and merits of a functional definition of learning. *Psychonomic bulletin & review*, 20(4), 631–642.
- Deng, J., Guo, J., Yuxiang, Z., Yu, J., Kotsia, I., & Zafeiriou, S. (2019). Retinaface: Single-stage dense face localisation in the wild..
- Devries, T., Biswaranjan, K., & Taylor, G. W. (2014). Multi-task learning of facial landmarks and expression. In *2014 canadian conference on computer and robot vision* (pp. 98–103).
- Dunn, D. F., & Higgins, W. E. (1993). Optimal gabor-filter design for texture segmentation. In *1993 ieee international conference on acoustics, speech, and signal processing* (Vol. 5, pp. 37–40).
- Ekman, & Rosenberg. (1997). *What the face reveals: Basic and applied studies of spontaneous expression using the facial action coding system (facs)*. Oxford University Press, USA.
- Ekman, P. (1970). Universal facial expressions of emotions. *California mental health research digest*, 8(4), 151–158.
- Ekman, P. (1993). Facial expression and emotion. *American psychologist*, 48(4), 384.
- Ekman, P. (2004). Emotions revealed: Understanding faces and feelings.
- Ekman, P. (2009). *Telling lies: Clues to deceit in the marketplace, politics, and marriage (revised edition)*. WW Norton & Company.
- Ekman, P. (2016). What scientists who study emotion agree about. *Perspectives on Psychological Science*, 11(1), 31–34.
- Ekman, P., Friesen, W., & Hager, J. (2002). The facial action coding system: A technique for the measurement of facial movement. a human face. *I-Tech Education and Publishing, Vienna*.
- Ekman, P. E., & Davidson, R. J. (1994). *The nature of emotion: Fundamental questions*. Oxford University Press.
- Elving, W. J. (2005). The role of communication in organisational change. *Corpo-*

rate communications: an international journal.

- Engin, D., Ecabert, C., Ekenel, H. K., & Thiran, J.-P. (2018). Face frontalization for cross-pose facial expression recognition. In *2018 26th european signal processing conference (eusipco)* (pp. 1795–1799).
- Entman, R. M. (1989). How the media affect what people think: An information processing approach. *The journal of Politics*, 51(2), 347–370.
- Evangelista, M. C., Watanabe, R., Leung, V. S., Monteiro, B. P., O'Toole, E., Pang, D. S., & Steagall, P. V. (2019). Facial expressions of pain in cats: the development and validation of a feline grimace scale. *Scientific Reports*, 9(1), 1–11.
- Fabian Benitez-Quiroz, C., Srinivasan, R., & Martinez, A. M. (2016). Emotionet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 5562–5570).
- Fernández-Dols, J.-M., & Crivelli, C. (2013). Emotion and expression: Naturalistic studies. *Emotion Review*, 5(1), 24–29.
- Field, D. J. (1987a). Relations between the statistics of natural images and the response properties of cortical cells. *Josa a*, 4(12), 2379–2394.
- Field, D. J. (1987b, Dec). Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A*, 4(12), 2379–2394. Retrieved from <http://josaa.osa.org/abstract.cfm?URI=josaa-4-12-2379> doi: 10.1364/JOSAA.4.002379
- Fix, E. (1951). *Discriminatory analysis: nonparametric discrimination, consistency properties*. USAF school of Aviation Medicine.
- Fragopanagos, N., & Taylor, J. G. (2005). Emotion recognition in human–computer interaction. *Neural Networks*, 18(4), 389–405.
- Freedman, J. (2017). *Emotions, feelings and moods*. Retrieved from <https://www.6seconds.org/2017/05/15/emotion-feeling-mood/>

- Gabor, D. (1946). Theory of communication. part 1: The analysis of information. *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering*, 93(26), 429–441.
- Garcia-Garcia, J. M., Penichet, V. M., & Lozano, M. D. (2017). Emotion detection: a technology review. In *Proceedings of the xviii international conference on human computer interaction* (pp. 1–8).
- Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014). Perceptions of emotion from facial expressions are not culturally universal: evidence from a remote culture. *Emotion*, 14(2), 251.
- Georgescu, M.-I., Ionescu, R. T., & Popescu, M. (2019). Local learning with deep and handcrafted features for facial expression recognition. *IEEE Access*, 7, 64827–64836.
- Ghimire, D., Jeong, S., Yoon, S., Choi, J., & Lee, J. (2015). Facial expression recognition based on region specific appearance and geometric features. In *2015 tenth international conference on digital information management (icdim)* (pp. 142–147).
- Gibert, G., Fordyce, A., & Stevens, C. J. (2010). Role of form and motion information in auditory-visual speech perception of mcgurk combinations and fusions. In *Auditory-visual speech processing 2010*.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press. Retrieved from <http://www.deeplearningbook.org>
- Goodfellow, I., Erhan, D., Carrier, P.-L., Courville, A., Mirza, M., Hamner, B., ... Bengio, Y. (2013). Challenges in representation learning: A report on three machine learning contests. In *International conference on neural information processing* (pp. 117–124).
- Goutte, C., & Gaussier, E. (2005). A probabilistic interpretation of precision, recall and f-score, with implication for evaluation. In *European conference on information retrieval* (pp. 345–359).

- Grave, E., Joulin, A., & CISS, E. (2016). M., grangier, d., and jegou, h. efficient softmax approximation for gpus. *arXiv preprint arXiv:1609.04309*.
- Grigorescu, C., & Petkov, N. (2006). *2-d gabor function–interactive visualization*.
- Gross, R. (2015). *Psychology: The science of mind and behaviour 7th edition*. Hodder Education.
- Guo, G., Guo, R., & Li, X. (2013). Facial expression recognition influenced by human aging. *IEEE Transactions on Affective Computing*, 4(3), 291–298.
- Guo, K. (2012). Holistic gaze strategy to categorize facial expression of varying intensities. *PLoS One*, 7(8).
- Hall, J. A., Horgan, T. G., & Murphy, N. A. (2019). Nonverbal communication. *Annual review of psychology*, 70, 271–294.
- Han, S., Pool, J., Narang, S., Mao, H., Gong, E., Tang, S., ... others (2016). Dsd: Dense-sparse-dense training for deep neural networks. *arXiv preprint arXiv:1607.04381*.
- Haselgrove, M. (2016). *Learning: A very short introduction*. Oxford University Press.
- Hebb, D. O. (1949). *The organization of behavior*. na.
- Hines, M. L., & Carnevale, N. T. (1997). The neuron simulation environment. *Neural computation*, 9(6), 1179–1209.
- Hook, J., Noroozi, F., Toygar, O., & Anbarjafari, G. (2019). Automatic speech based emotion recognition using paralinguistics features. *Bulletin of the Polish Academy of Sciences. Technical Sciences*, 67(3).
- Ilonen, J., Kämäräinen, J.-K., & Kälviäinen, H. (2005). *Efficient computation of gabor features*. Lappeenranta University of Technology.
- Ionescu, R. T., Popescu, M., & Grozea, C. (2013). Local learning to improve bag of visual words model for facial expression recognition. In *Workshop on challenges in representation learning, icml*.
- Islam, B., Mahmud, F., & Hossain, A. (2018a). Facial expression region seg-

- mentation based approach to emotion recognition using 2d gabor filter and multiclass support vector machine. In *2018 21st international conference of computer and information technology (iccit)* (pp. 1–6).
- Islam, B., Mahmud, F., & Hossain, A. (2018b). Facial expression region segmentation based approach to emotion recognition using 2d gabor filter and multiclass support vector machine. In *2018 21st international conference of computer and information technology (iccit)* (pp. 1–6).
- Islam, B., Mahmud, F., Hossain, A., Goala, P. B., & Mia, M. S. (2018). A facial region segmentation based approach to recognize human emotion using fusion of hog & lbp features and artificial neural network. In *2018 4th international conference on electrical engineering and information & communication technology (iceeict)* (pp. 642–646).
- Jack, R. E., Garrod, O. G., Yu, H., Caldara, R., & Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences*, 109(19), 7241–7244.
- Jing, W., Tian, F., Zhang, J., Chao, K.-M., Hong, Z., & Liu, X. (2020). Feature super-resolution based facial expression recognition for multi-scale low-resolution faces. *arXiv preprint arXiv:2004.02234*.
- Kandlousi, N. S. A. E., Ali, A. J., & Abdollahi, A. (2010). Organizational citizenship behavior in concern of communication satisfaction: The role of the formal and informal communication. *International Journal of Business and Management*, 5(10), 51.
- Kenny, A. (2003). *Action, emotion and will*. Routledge.
- Khalil, R. A., Jones, E., Babar, M. I., Jan, T., Zafar, M. H., & Alhussain, T. (2019). Speech emotion recognition using deep learning techniques: A review. *IEEE Access*, 7, 117327–117345.
- Kim, B.-K., Dong, S.-Y., Roh, J., Kim, G., & Lee, S.-Y. (2016). Fusing aligned and non-aligned face information for automatic affect recognition in the wild: a

- deep learning approach. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 48–57).
- Kitayama, S. E., & Markus, H. R. E. (1994). *Emotion and culture: Empirical studies of mutual influence*. American Psychological Association.
- Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, 78(9), 1464–1480.
- Kořakowska, A., Landowska, A., Szwoch, M., Szwoch, W., & Wróbel, M. R. (2013). Emotion recognition and its application in software engineering. In *2013 6th international conference on human system interactions (hsi)* (pp. 532–539).
- Kořakowska, A., Landowska, A., Szwoch, M., Szwoch, W., & Wrobel, M. R. (2014). Emotion recognition and its applications. In *Human-computer systems interaction: Backgrounds and applications 3* (pp. 51–62). Springer.
- Kowalczyk, Z., & Czubenko, M. (2016). Computational approaches to modeling artificial emotion—an overview of the proposed solutions. *Frontiers in Robotics and AI*, 3, 21.
- Kret, M. E. (2015). Emotional expressions beyond facial muscle actions. a call for studying autonomic signals and their impact on social perception. *Frontiers in psychology*, 6, 711.
- Kugaevskikh, A. (2017, 08). Comparison gabor filter parameters for efficient edge detection. *Informatsionnye tekhnologii*, 23, 598-605.
- Lai, Y.-H., & Lai, S.-H. (2018). Emotion-preserving representation learning via generative adversarial network for multi-view facial expression recognition. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)* (pp. 263–270).
- Lapakko, D. (2007). Communication is 93% nonverbal: An urban legend proliferates. *Communication and Theater Association of Minnesota Journal*, 34(1), 2.

- LeCun, Y., dkk. (1989). Generalization and network design strategies. connectionism in perspective. *Zurich, Switzerland, Elsiever*.
- Leung, T., & Malik, J. (2001). Representing and recognizing the visual appearance of materials using three-dimensional textures. *International journal of computer vision*, 43(1), 29–44.
- Li, D., Wang, Z., Wang, C., Liu, S., Chi, W., Dong, E., ... Song, Y. (2019). The fusion of electroencephalography and facial expression for continuous emotion recognition. *IEEE Access*, 7, 155724–155736.
- Li, S., & Deng, W. (2018). Deep facial expression recognition: A survey. *arXiv preprint arXiv:1804.08348*.
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2), 91–110.
- Luan, S., Chen, C., Zhang, B., Han, J., & Liu, J. (2018). Gabor convolutional networks. *IEEE Transactions on Image Processing*, 27(9), 4357–4366.
- Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010). The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In *2010 IEEE computer society conference on computer vision and pattern recognition-workshops* (pp. 94–101).
- Lunenburg, F. C. (2010). Formal communication channels: Upward, downward, horizontal, and external. *Focus on Colleges, Universities, and Schools*, 4(1), 1–7.
- Lutz, C. A., & Abu-Lughod, L. E. (1990). Language and the politics of emotion. In *This book grew out of a session at the 1987 annual meeting of the American anthropological association called "emotion and discourse"*.
- Lv, H.-R., Lin, Z.-L., Yin, W.-J., & Dong, J. (2008). Emotion recognition based on pressure sensor keyboards. In *2008 IEEE international conference on multimedia and expo* (pp. 1089–1092).
- Lyons, M., Akamatsu, S., Kamachi, M., & Gyoba, J. (1998). Coding facial expres-

- sions with gabor wavelets. In *Proceedings third ieee international conference on automatic face and gesture recognition* (pp. 200–205).
- Majumder, A., Behera, L., & Subramanian, V. K. (2016). Automatic facial expression recognition system using deep network-based data fusion. *IEEE transactions on cybernetics*, 48(1), 103–114.
- Mandal, M. K., & Awasthi, A. (2014). *Understanding facial expressions in communication: Cross-cultural and multidisciplinary perspectives*. Springer.
- Mehrabian, A., & Ferris, S. R. (1967). Inference of attitudes from nonverbal communication in two channels. *Journal of consulting psychology*, 31(3), 248.
- Mehrabian, A., & Wiener, M. (1967). Decoding of inconsistent communications. *Journal of personality and social psychology*, 6(1), 109.
- Miller, G. A. (1999). On knowing a word. *Annual Review of psychology*, 50(1), 1–19.
- Mitchell, T. M. (2006). *The discipline of machine learning* (Vol. 9). Carnegie Mellon University, School of Computer Science, Machine Learning . . .
- Mittal, T., Bhattacharya, U., Chandra, R., Bera, A., & Manocha, D. (2019). M3er: Multiplicative multimodal emotion recognition using facial, textual, and speech cues. *arXiv preprint arXiv:1911.05659*.
- Mohammed, I. S. (2019). Handwritten signature recognition with gabor filters and neural network. In *Aip conference proceedings* (Vol. 2096, p. 020005).
- Mun, C. (2019). *Knowing emotions: Truthfulness and recognition in affective experience*. Oxford University Press.
- Munezero, M. D., Montero, C. S., Sutinen, E., & Pajunen, J. (2014). Are they different? affect, feeling, emotion, sentiment, and opinion detection in text. *IEEE transactions on affective computing*, 5(2), 101–111.
- Nava, R., Escalante-Ramírez, B., & Cristóbal, G. (2011). A comparison study of gabor and log-gabor wavelets for texture segmentation. In *2011 7th international symposium on image and signal processing and analysis (ispa)* (pp.

189–194).

- Nilsson, N. J. (2005). Introduction to machine learning an early draft of a proposed textbook department of computer science. *Standorf University*.
- O’Gara, S., & McGuinness, K. (2019). Comparing data augmentation strategies for deep image classification.
- Ojala, T., Pietikäinen, M., & Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. *Pattern recognition*, 29(1), 51–59.
- Paiva, A., André, E., Arafa, Y., Botelho, L., Costa, M., Figueiredo, P., . . . others (2001). Safira: Supporting affective interactions in real time applications. *CAST 2001: Living in Mixed Realities*, 227–230.
- Pal, K. K., & Sudeep, K. (2016). Preprocessing for image classification by convolutional neural networks. In *2016 ieee international conference on recent trends in electronics, information & communication technology (rteict)* (pp. 1778–1781).
- Parise, A. G., Gagliano, M., & Souza, G. M. (2020). Extended cognition in plants: is it possible? *Plant Signaling & Behavior*, 1710661.
- Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep face recognition.
- Parrott, W. G. (2001). *Emotions in social psychology: Essential readings*. Psychology Press.
- Pennebaker, J. W. (1995). *Emotion, disclosure, & health*. American Psychological Association.
- Plutchik, R., & Kellerman, H. (2013). *Theories of emotion* (Vol. 1). Academic Press.
- Postmes, T., Tanis, M., & De Wit, B. (2001). Communication and commitment in organizations: A social identity approach. *Group Processes & Intergroup Relations*, 4(3), 227–246.
- Pramerdorfer, C., & Kampel, M. (2016). Facial expression recognition using convo-

- lutional neural networks: state of the art. *arXiv preprint arXiv:1612.02903*.
- Prilianti, K., Brotosudarmo, T., Anam, S., & Suryanto, A. (2019). Performance comparison of the convolutional neural network optimizer for photosynthetic pigments prediction on plant digital image. In *Aip conference proceedings* (Vol. 2084, p. 020020).
- Purnell, L. (2018). Cross cultural communication: Verbal and non-verbal communication, interpretation and translation. In *Global applications of culturally competent health care: Guidelines for practice* (pp. 131–142). Springer.
- Pérez, J. F. (2018). *Emotions vs sentiments: Why you should know the difference*. Retrieved from <https://aleteia.org/2018/08/17/emotions-vs-sentiments-why-you-should-know-the-difference/>
- Qin, S., Zhu, Z., Zou, Y., & Wang, X. (2020). Facial expression recognition based on gabor wavelet transform and 2-channel cnn. *International Journal of Wavelets, Multiresolution and Information Processing*, 18(02), 2050003.
- Ramos, D. (2009). *Animal learning: Basic practice and application in dogs and cats*. Retrieved from <https://www.vin.com/apputil/content/defaultadv1.aspx?pId=11290&id=4252593>
- Roy, T., Marwala, T., & Chakraverty, S. (2020). Speech emotion recognition using neural network and wavelet features. In *Recent trends in wave mechanics and vibrations* (pp. 427–438). Springer.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *nature*, 323(6088), 533–536.
- Sara, B., Noureddine, A., & Hajar, B. (2019). Ambulance detection system. In *International conference on artificial intelligence and symbolic computation* (pp. 19–25).
- Schneider, F., Balles, L., & Hennig, P. (2019). Deepobs: A deep learning optimizer benchmark suite. *arXiv preprint arXiv:1903.05499*.
- Schrader, C., & Kalyuga, S. (2020). Linking student's emotions to engagement and

- writing performance when learning japanese letters with a pen-based tablet: An investigation based on individual pen pressure parameters. *International Journal of Human-Computer Studies*, 135, 102374.
- Schwarz, N. (2000). Emotion, cognition, and decision making. *Cognition & Emotion*, 14(4), 433–440.
- Seyeditabari, A., Tabari, N., & Zadrozny, W. (2018). Emotion detection in text: a review. *arXiv preprint arXiv:1806.00674*.
- Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms*. Cambridge university press.
- Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1), 60.
- Singh, H. P., Bailer-Jones, C. A., & Gupta, R. (2001). An introduction to artificial neural networks.
- Sisodia, P., Verma, A., & Kansal, S. (2013). Human facial expression recognition using gabor filter bank with minimum number of feature vectors. *International Journal of Applied Information Systems*, 5(9), 9–13.
- Sivalingamaiah, M., & Reddy, B. V. (2012). Texture segmentation using multichannel gabor filtering. *IOSR Journal of Electronics and Communication Engineering*, 2(6), 22–26.
- Smith, C. A., Lazarus, R. S., dkk. (1990). Emotion and adaptation. *Handbook of personality: Theory and research*, 609–637.
- Sujata, J., Madhavi, D., & Vishal, K. (2018). Facial emotion detection technology: concept, marketing applications, business drivers and challenges. *International Journal of Engineering & Technology*, 7(1.9), 149–156.
- Sumathi, C., Santhanam, T., & Mahadevi, M. (2012). Automatic facial expression analysis a survey. *International Journal of Computer Science and Engineering Survey*, 3(6), 47.
- Surkamp, C. (2014). Non-verbal communication: Why we need it in foreign la-

- language teaching and how we can foster it with drama activities. *Scenario*, 2014(02), 12–27.
- Valstar, M., & Pantic, M. (2006). Fully automatic facial action unit detection and temporal analysis. In *2006 conference on computer vision and pattern recognition workshop (cvprw'06)* (pp. 149–149).
- Vijay, M., & Indumathi, G. (2019). A highly secure multi-factor authentication system using biometrics to enhance privacy in internet of things (iot). *International Research Journal of Multidisciplinary Technovation*, 1(6), 26–34.
- Viswanathan, K., Gioia, P., & Morin, L. (2014). Morlet wavelet transformed holograms for numerical adaptive view-based reconstruction. In *Optics and photonics for information processing viii* (Vol. 9216, p. 92160G).
- Vrij, A., Hartwig, M., & Granhag, P. A. (2019). Reading lies: nonverbal communication and deception. *Annual review of psychology*, 70, 295–317.
- Vyas, V. S., & Rege, P. (2006). Automated texture analysis with gabor filter. *GVIP journal*, 6(1), 35–41.
- Wagh, K. P., & Vasanth, K. (2019). Electroencephalograph (eeg) based emotion recognition system: A review. In *Innovations in electronics and communication engineering* (pp. 37–59). Springer.
- Walia, E., & Verma, V. (2016). Boosting local texture descriptors with log-gabor filters response for improved image retrieval. *International Journal of Multimedia Information Retrieval*, 5(3), 173–184.
- Westbrook, K. E., & Varacallo, M. (2019). *Anatomy, head and neck, facial muscles*. Retrieved from <https://www.ncbi.nlm.nih.gov/books/NBK493209/>
- Widen, S. C., & Russell, J. A. (2010). The “disgust face” conveys anger to children. *Emotion*, 10(4), 455.
- Widen, S. C., & Russell, J. A. (2013). Children’s recognition of disgust in others. *Psychological Bulletin*, 139(2), 271.

- Widen, S. C., Russell, J. A., & Brooks, A. (2004). Anger and disgust: Discrete or overlapping categories. In *2004 aps annual convention, boston college, chicago, il*.
- Winslow, V. L. (2015). *Classic human anatomy in motion: The artist's guide to the dynamics of figure drawing*. Watson-Guption.
- Wolf, K. (2015). Measuring facial expression of emotion. *Dialogues in clinical neuroscience*, 17(4), 457.
- Yan, K., Zheng, W., Zhang, T., Zong, Y., & Cui, Z. (2018). Cross-database non-frontal facial expression recognition based on transductive deep transfer learning. *arXiv preprint arXiv:1811.12774*.
- Ye, Y., Zhang, X., Lin, Y., & Wang, H. (2019). Facial expression recognition via region-based convolutional fusion network. *Journal of Visual Communication and Image Representation*, 62, 1–11.
- Zhang, L., Verma, B., Tjondronegoro, D., & Chandran, V. (2018). Facial expression analysis under partial occlusion: A survey. *ACM Computing Surveys (CSUR)*, 51(2), 1–49.
- Zhang, Z., Luo, P., Loy, C.-C., & Tang, X. (2015). Learning social relation traits from face images. In *Proceedings of the ieee international conference on computer vision* (pp. 3631–3639).
- Zhao, X., & Zhang, S. (2011). Facial expression recognition based on local binary patterns and kernel discriminant isomap. *Sensors*, 11(10), 9573–9588.
- Zhong, Z., Zheng, L., Kang, G., Li, S., & Yang, Y. (2017). Random erasing data augmentation. *arXiv preprint arXiv:1708.04896*.
- Zhou, J., & Wang, T. (2019). Facial feature recognition based on asnmf method.
- Zill, D., & Dewar, J. (2011). *Algebra and trigonometry*. Jones & Bartlett Publishers.