

---

## Deteksi Ekspresi Wajah Pada Scene Film Menggunakan *Residual Masking Network*

Bhanu Sri Nugraha<sup>1</sup>, Fahma Inti Ilmawati<sup>2</sup>, Dhani Ariatmanto<sup>3</sup>, Lukman<sup>4</sup>

bhanu@amikom.ac.id<sup>1</sup>, fahma.inti.ilmawati@students.amikom.ac.id<sup>2</sup>, dhaniari@amikom.ac.id<sup>3</sup>, masman@amikom.ac.id<sup>4</sup>

<sup>1,2,3,4</sup> Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta

---

### Informasi Artikel

Diterima : 4 Okt 2024

Direvisi : 24 Apr 2025

Disetujui : 30 Apr 2025

---

### Kata Kunci

*Facial Expression Recognition, Residual Masking Network, film, scene, wajah*

---

### Abstrak

Pada sebuah film, ada berbagai hal yang harus di perhatikan di dalamnya, salah satunya ekspresi aktor dalam mendalami perannya. Hal itu yang dapat membuat permisa terlarut dalam alur cerita dari film tersebut dan dapat memberikan nilai ketepatan pendalaman peran bagi orang yang menyaksikannya. Dengan populernya deep learning, khususnya CNN (*Convolutional Neural Network*) dapat secara otomatis mengekstraksi dan mempelajari untuk sistem pengenalan ekspresi wajah yang baik. Dalam eksperimen ini, kami menggunakan *Residual Masking Network* (RNM). Dibangun di atas pemahaman ini, kami mengevaluasi kumpulan data ini dengan model klasifikasi gambar standar untuk menganalisis kelayakan penggunaan ekspresi wajah dalam menentukan ketepatan muatan emosional di dalam mendalami peran aktor pada sebuah film. Hasil akurasi dalam penelitian ini 99% yang mendeteksi ekspresi marah.

---

### Keywords

*Facial Expression Recognition, Residual Masking Network, film, scene, face*

---

### Abstract

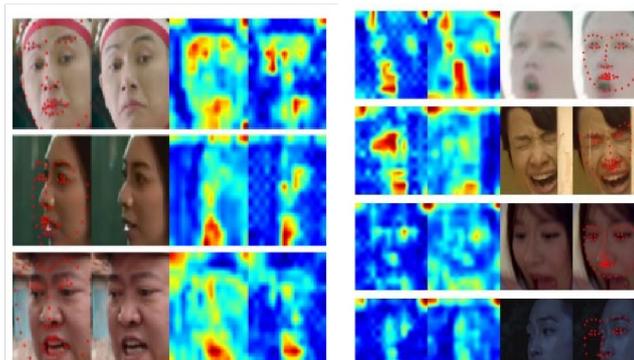
*In a film, there are various things that must be paid attention to, one of which is the actor's expression in deepening his role. This is what can make the audience immersed in the storyline of the film and can provide value for the accuracy of deepening the role for the people who watch it. With the popularity of deep learning, especially CNN (Convolutional Neural Network) can automatically extract and learn for a good facial expression recognition system. In this experiment, we use Residual Masking Network (RNM). Building on this understanding, we evaluate this dataset with standard image classification models to analyse the feasibility of using facial expressions in determining the appropriateness of emotional content in an actor's role in a film. The accuracy results in this study were 99% for detecting angry expressions.*

## A. Pendahuluan

Pada sebuah film, ada berbagai hal yang harus di perhatikan di dalamnya, salah satunya ekspresi aktor dalam mendalami perannya. Hal itu yang dapat membuat pemirsa terlarut dalam alur cerita dari film tersebut dan dapat memberikan nilai ketepatan pendalaman peran bagi orang yang menyaksikannya. Terdapat beberapa hal yang aktor yang pertanyakan, apakah mereka sudah layak mendalami peran yang dijalankan, atau pemaparan ekspresi gerakan wajah mereka sudah sesuai dengan emosi yang diharapkan pemirsa.

Memahami bagaimana ikatan antara emosi yang diwakili dan emosi yang dirasakan ini dapat memberi kita informasi konkret tentang interaksi manusia dengan media digital yang berkembang pesat. Ini dapat mengintegrasikan platform streaming film besar dan digunakan untuk pengambilan informasi mengenai pengalaman pemirsa, meninjau kualitas, dan untuk peningkatan sistem rekomendasi yang canggih. Oleh karena itu, persepsi film emosional juga bisa menjadi faktor pendukung untuk menciptakan platform streaming film yang efektif [1].

Belakangan ini, *Deep Learning* menjadi populer, khususnya *Convolutional Neural Networks (CNN)*. *Deep learning* telah muncul sebagai teknik yang ampuh untuk *Facial Expression Recognition (FER)* karena kemampuannya mengekstraksi dan mempelajari fitur diskriminatif dari gambar wajah secara otomatis [2]. Namun, dalam hal ekspresi wajah, perlu dicatat bahwa sebagian besar isyarat berasal dari beberapa daerah wajah seperti mata, mulut. Sebaliknya, daerah lain memiliki kontribusi kecil untuk output, misalnya rambut, rahang. Metode ini mencoba untuk fokus pada wilayah penting ini dengan menggunakan fitur perantara bernama *landmark* wajah [2]. Deteksi *landmark* wajah bisa mendapatkan hasil yang mencolok dalam kondisi labcontrolled. Namun, di kondisi gambar resolusi rendah, biasanya tidak tampil dengan baik karena variasi dalam pose kepala, mata dst berkurang. (lihat Gambar 1).



**Gambar 1.** Contoh deteksi *landmark* dan fitur *Masking Block*

Tujuan penelitian ini, kami menghasilkan kumpulan data dengan model klasifikasi gambar standar untuk menganalisis kelayakan penggunaan ekspresi wajah untuk menentukan muatan emosional sebuah film khususnya ketepatan emosi wajah aktor dalam mendalami perannya. Untuk mengatasi masalah kurangnya kumpulan data untuk ruang lingkup yang dianalisis, kami menunjukkan kelayakan menggunakan kumpulan data umum untuk proses training dan mengusulkan cara baru untuk melihat emosi dengan membuat presentase emosi berdasarkan bukti

yang diperoleh dalam proses deteksi. Dalam penelitian ini, kami memberikan tiga kontribusi signifikan: Mengusulkan Ide *Masking* terbaru untuk meningkatkan kinerja, membangun *Residual Masking Network* untuk pengenalan ekspresi wajah, dan membuat kumpulan data baru.

Menurut João Almeida dkk, untuk menjembatani teori emosi dan pengamatan visual dari ekspresi wajah, berpendapat bahwa ekspresi wajah bersifat universal dan memberikan informasi yang cukup untuk memprediksi emosi. Mereka menyajikan uji komprehensif dari kumpulan data paling relevan yang digunakan untuk FER 2013, namun kekurangan dan keterbatasan dalam kumpulan data ini, mulai dari ketidakseimbangan beberapa gambar kosong yang tidak mengandung wajah. Untuk mengatasi permasalahan tersebut maka peneliti menghasilkan deteksi emosi melalui seleksi alam menjadi serangkaian emosi dasar yang terbatas dan terpisah melalui analisis ekspresi wajah dengan akurasi 90%[1].

Salah satu pendekatan yang telah dieksplorasi adalah penggunaan model pembelajaran mendalam, khususnya jaringan saraf konvolusional (CNN), untuk mengekstrak fitur diskriminatif dari gambar wajah. Zhang et al mengusulkan algoritma pengenalan ekspresi wajah berdasarkan jaringan saraf residual yang ditingkatkan. Penulis mengatasi tantangan pengenalan ekspresi wajah dengan memanfaatkan arsitektur jaringan residual, yang memungkinkan pelatihan jaringan yang lebih dalam tanpa mengalami masalah hilangnya gradien. Algoritma mereka mencapai hasil yang menjanjikan dalam hal akurasi dan ketangguhan [3].

Survei lain yang dilakukan oleh Li & Deng memberikan gambaran tentang pengenalan ekspresi wajah mendalam. Penulis membahas berbagai teknik dan model yang digunakan dalam bidang ini, termasuk model diskriminatif dan jaringan mendalam multitugas. Mereka menyoroti pentingnya menggabungkan pengenalan keadaan dan lokalitas titik kunci untuk meningkatkan pengenalan ekspresi wajah. Survei ini memberikan wawasan berharga tentang kemajuan dan tantangan dalam pengenalan ekspresi wajah mendalam [2].

Selain kemajuan dalam pengenalan ekspresi wajah, juga ada studi yang menyelidiki batasan yang diperkenalkan oleh masker wajah. Abate dkk. menganalisis dan membandingkan tiga metode pengenalan ekspresi wajah berbasis frame statis teratas, termasuk *Residual Masking Network*. Penulis memeriksa dampak masker wajah pada pengenalan ekspresi dan membahas tantangan yang timbul dari daerah wajah yang terhalangi. Penelitian ini memberikan wawasan tentang implikasi praktis masker wajah pada sistem pengenalan ekspresi wajah [4].

Selanjutnya, Pazhoohi dkk. melakukan studi tentang bagaimana masker wajah memengaruhi pengenalan emosi pada populasi umum dan individu dengan ciri-ciri autistik. Hasil mereka mendukung penelitian sebelumnya yang menunjukkan bahwa pemakaian masker wajah mengurangi pengenalan ekspresi wajah dan keyakinan dalam identifikasi ekspresi. Studi ini menyoroti pentingnya mempertimbangkan faktor eksternal, seperti masker wajah, saat merancang sistem pengenalan ekspresi wajah [5].

Terakhir, Lou mengusulkan pendekatan pengenalan ekspresi wajah berbasis fusi fitur multi-skala berdasarkan mekanisme perhatian. Penulis menggabungkan mekanisme perhatian untuk fokus selektif pada daerah wajah yang informatif dan menggabungkan fitur pada skala yang berbeda. Pendekatan ini bertujuan untuk

mengurangi variasi intra-kelas dan meningkatkan akurasi pengenalan ekspresi wajah [6].

Pengenalan ekspresi wajah telah mengalami kemajuan signifikan melalui penggunaan model pembelajaran mendalam, seperti jaringan saraf residual dan jaringan saraf konvolusional. Para peneliti juga telah mengeksplorasi dampak faktor eksternal, seperti masker wajah, pada pengenalan ekspresi. Penelitian masa depan dalam bidang ini harus terus mengatasi tantangan yang timbul dari penutupan dan faktor eksternal untuk meningkatkan ketangguhan dan aplikabilitas dunia nyata dari sistem pengenalan ekspresi wajah.

Berdasarkan penelitian pertama, tidak digunakannya proses pendeteksian gambar melalui *masking block* pada *residual masking network* dan pada penelitian kedua tidak adanya dataset berupa cuplikan film. Maka dari itu dipenelitian ini kami akan mengupayakan proses deteksi menggunakan *residual network* dengan dataset cuplikan film yang tepat.

## B. Metode Penelitian

### 1. Alur Penelitian

Alur utama dari metode yang diusulkan adalah *Residual Masking Network* yang diilustrasikan pada Gambar 2. Jaringan ini berisi empat *Residual Masking Blocks* (*Resmasking Blocks*) utama [7]. Setiap *Residual Masking Block*, yang beroperasi pada ukuran fitur yang berbeda, berisi *Residual Layer* dan *Masking Block* (lihat Tabel I).

**Tabel 1.** Konfigurasi *Residual Masking Network* yang temperance

Layer name	Output size	Detail
Conv1	$64 \times 112 \times 112$	$7 \times 7$ , stride 2
MaxPooling	$64 \times 56 \times 56$	$3 \times 3$ , stride 2
Resmasking Block 1	$64 \times 56 \times 56$	RL 1, MB 1
Resmasking Block 2	$128 \times 28 \times 28$	RL 2, MB 2
Resmasking Block 3	$256 \times 14 \times 14$	RL 3, MB 3
Resmasking Block 4	$512 \times 7 \times 7$	RL 4, MB 4
Average pooling	$512 \times 7 \times 7$	
FC, Softmax	7	

Sebuah gambar masukan ukuran  $224 \times 224$  akan melalui yang pertama  $3 \times 3$  lapisan konvolusi dengan langkah 2 sebelum melewati  $2 \times 2$  lapisan *max-pooling*, mengurangi ukuran spasialnya menjadi  $56 \times 56$ . Selanjutnya, peta fitur yang diperoleh setelah lapisan penyatuan sebelumnya ditransformasikan oleh empat *Blok Masking Residual* berikut dengan peta fitur yang dihasilkan dari empat ukuran spasial termasuk  $56 \times 56$ ,  $28 \times 28$ ,  $14 \times 14$ , dan  $7 \times 7$ . Jaringan diakhiri dengan lapisan penyatuan rata-rata dan lapisan 7- arah yang terhubung penuh dengan softmax untuk menghasilkan output yang sesuai dengan tujuh keadaan ekspresi wajah (6 emosi dan satu keadaan netral).

### 2. *Residual Masking Block*

Dalam penelitian ini, kami mengusulkan *Masking Block*, yang melakukan operasi penilaian. Kemudian *Residual Masking Block* yang berisi *Residual Layer* dan

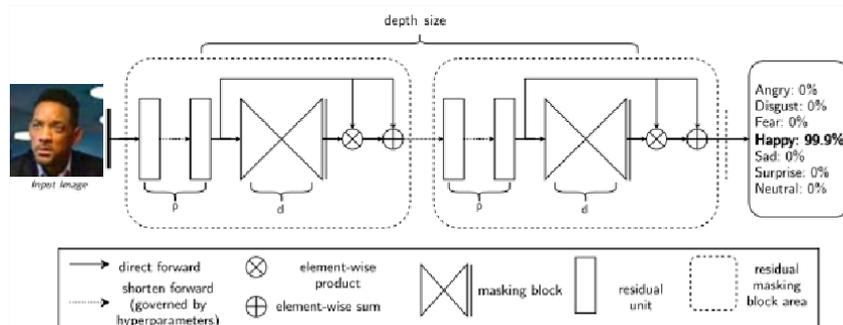
Masking Block, dengan yang pertama bertanggung jawab atas pemrosesan fitur dan yang terakhir menghasilkan bobot untuk peta fitur yang sesuai, seperti pada Gambar 2. Diberikan peta fitur input

$$F \in \mathbb{R}^{C \times W \times H},$$

$$F R = R(F), FR \in \mathbb{R}$$

Kemudian, peta aktivasi dengan ukuran yang sama FM memiliki nilai dalam rentang  $[0,1]$  dihitung melalui Masking Block dengan rumus  $FM = M(FR)$ . Akhirnya, peta fitur yang disempurnakan - output dari Residual Masking Block akan dihasilkan oleh Formula 1. Dengan cara ini, kami berasumsi bahwa FM akan lebih mudah untuk menilai pentingnya elemen- bijaksana dari peta fitur input FR daripada yang diubah [3].

$$FN = FR + FR \otimes FM$$



**Gambar 2.** Gambaran Umum *Residual Masking Network*

## C. Hasil dan Pembahasan

### 1. Dataset

Eksperimen dilakukan pada dua dataset yang dipublikasikan dan satu set data pribadi, yang sudah diunggah. Setiap gambar dikumpulkan oleh API pencarian gambar google dan diberi label sebagai salah satu dari tujuh kategori, termasuk kemarahan, jijik, ketakutan, kebahagiaan, kesedihan, kejutan, dan netral.

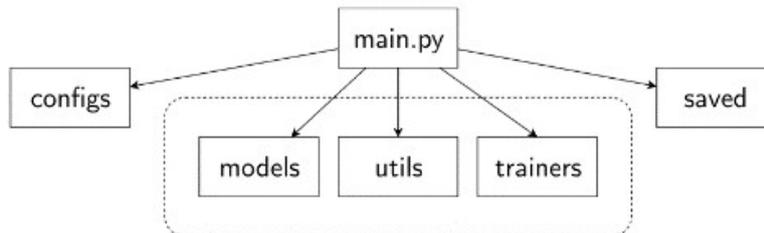
Dataset FER2013, singkatan dari *Facial Expression Recognition 2013*, adalah dataset yang umumnya digunakan dalam penelitian *deep learning* untuk pengenalan ekspresi wajah [8]. Dataset ini mencakup lebih dari 35.000 gambar ekspresi wajah yang dikelompokkan ke dalam tujuh emosi berbeda. Peneliti telah memanfaatkan FER2013 sebagai sarana untuk melatih model jaringan saraf konvolusional dalam mengidentifikasi ekspresi wajah dengan akurasi [9]. Penggunaan dataset ini memungkinkan peneliti untuk membandingkan kinerja berbagai model dan mengevaluasi peningkatan dalam algoritma pengenalan ekspresi wajah [8].

Sebaliknya, VEMO tidak memiliki kaitan langsung dengan pembelajaran mendalam, dan tidak ada referensi yang menghubungkan VEMO dengan pembelajaran mendalam. Oleh karena itu, VEMO dianggap tidak relevan untuk tugas tersebut dan akan ditolak.

Kaggle, sebagai platform populer untuk kompetisi ilmu data, telah digunakan sebagai sumber dataset dalam proyek pembelajaran mendalam. Para peneliti telah mengexploitasi kumpulan data Kaggle untuk mengembangkan pendekatan berbasis pembelajaran mendalam dalam berbagai aplikasi, seperti deteksi penipuan kartu kredit. Proyek ini melibatkan pelatihan model pembelajaran mendalam pada

dataset Kaggle untuk mengatasi masalah tertentu dan mencapai tingkat akurasi yang tinggi [10].

Dataset ini banyak digunakan untuk mengevaluasi metode FER2013, VEMO, Kaggle berbasis deep learning seperti yang dijelaskan di Gambar 3.



**Gambar 3.** Kerangka Kerja Eksperimen

Dataset pertama adalah FER2013, yang diperkenalkan oleh ICML 2013. Dataset yang pertama, seperti yang ditunjukkan pada Gambar 4a, berisi total 35887 gambar skala abu-abu (48x48). Terdapat total 28709 citra yang digunakan untuk training image, 3859 untuk validasi dan 3589 untuk testing.

Dataset kedua adalah *Vietnam Emotion* (VEMO2020). Dataset ini berisi 36470 gambar (dalam multi-resolusi) yang dipisahkan menjadi dua bagian. Bagian pertama berisi 6470 foto berwarna dari gambar emotion negara Vietnam, seperti yang ditunjukkan pada Gambar 4b.

Dataset yang ketiga pada Gambar 4c, diambil dari dataset pribadi yang diunggah di Kaggle : ceofaila/scene-movie-face-expression Dataset berjumlah 176 gambar berwarna berisi cuplikan atau penggalan film yang menunjukkan ekspresi wajah aktor terkenal.



**Gambar 4.** Kumpulan dataset FER2013



**Gambar 5.** Kumpulan dataset VEMO



**Gambar 6.** Kumpulan dataset Kaggle

## 2. Pengaturan Eksperimen

Gambar pelatihan asli ditingkatkan menjadi  $224 \times 224$  dan dikonversi ke RGB sebelum proses pelatihan untuk beradaptasi dengan model pra-pelatihan ImageNet. Selain itu, gambar pelatihan ditambah untuk mencegah *over-fitting*. Metode augmentasi termasuk membalik kiri-kanan dan memutar dalam kisaran  $[-30,30]$ .

Setiap eksperimen berlangsung selama maksimum 50 epoch, dan berhenti ketika akurasi validasi tidak ditingkatkan lebih dari delapan langkah. Ukuran batch 48 dan tingkat pembelajaran awal 0,0001, scheduler mengurangi learning rate sepuluh kali lipat jika akurasi validasi tidak meningkat dalam dua epoch yang berkelanjutan. Momentum diatur ke 0,9 dan penurunan berat badan menjadi 0,001. Metrik evaluasi untuk tugas klasifikasi adalah:

$$\text{Ketepatan} = \frac{TP + TN}{TP + TN + FP + FN}$$

Dimana TP: Benar Positif, TN: Benar Negatif, FP: Positif Salah, FN: Negatif Palsu. Eksperimen dari jaringan yang berbeda juga dilakukan menggunakan lingkungan pengaturan serta evaluasi metrik.

**Tabel 2.** Evaluasi kinerja jaringan dan metode pada FER2013

Model	Top-1 Accuracy	Top-5 Accuracy
Resnet34	70.59	90.23
CBAM Resnet34	71.67	90.72
RNM	72.16	90.91

**Tabel 3.** Evaluasi kinerja jaringan dan metode pada VEMO

Emotion	VGG19	ResMaskingNet	ResMaskingNet + 6
Accuracy	70.80	74.14	76.82

**Tabel 4.** Evaluasi kinerja jaringan dan metode pada dataset Kaggle

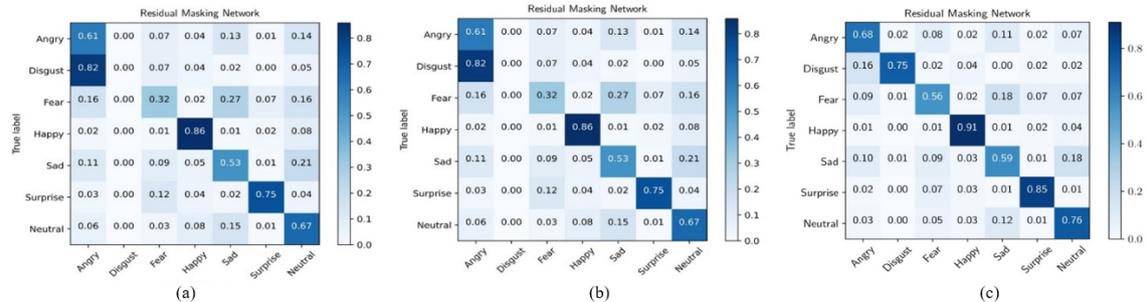
Emotion	Accuracy
Angry	99
Disgust	29
Fear	57
Happy	34
Sad	47
Surprise	66
Neutral	23

Eksperimen dilakukan menggunakan Collaboratory CPU AMD A10-7860K Radeon R7, 12 Compute Cores 4C, VGA, GTX 1050Ti, RAM 8GB. Untuk menguji waktu pemrosesan dalam aplikasi real time. Dengan infrastruktur ini, jaringan yang diusulkan dapat memproses frame per detik dan setiap frame berisi satu wajah. Dengan hasil ini, kami dapat menjamin aplikasi real-time.

### 3. Evaluasi dan Analisis

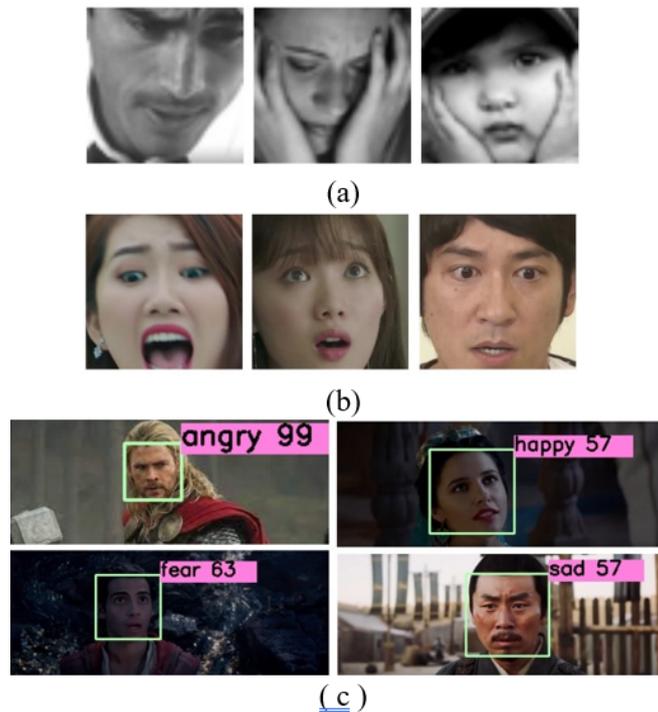
Metode yang diusulkan telah diuji pada beberapa aspek dan menghasilkan hasil yang positif. Untuk dataset public FER2013, pertama, kami memilih sejumlah jaringan klasifikasi yang terkenal dan kuat seperti Resnet151 [11], Densenet121

[12], Cbam resnet50 [13], dan Bam resnet50 [14]. Untuk penjelasan dari VEMO dan Kaggle hasilnya seperti yang diuraikan table II dan III menjelaskan bahwa akurasi penelitian pada dataset Kaggle menunjukkan pada xmin: 98, ymin: 20, xmax: 147, ymax: 69, emo\_label = angry, emo\_proba = 0.99.



**Gambar 7.** Hasil konfusi matrix (a)dataset FER2013 (b)dataset VEMO (c)dataset Kaggle

Selanjutnya, berikut hasil dari residual masking network dari ketiga datasets. (Lihat Gambar 8 )



**Gambar 8.** Hasil Ekspresi (a)ekspresi sedih di dataset FER2013 (b)ekspresi terkejut di dataset VEMO (c)persentase ekspresi di dataset Kaggle

#### D. Simpulan

Dalam penelitian ini, kita dapat memperoleh distribusi emosi yang sebagian besar berfokus pada pertimbangan emosi dasar lainnya, dengan sistem yang lebih kompleks. yang dapat membedakan kemarahan. kesedihan dan emosi lain yang ironis, menuju emosi lain yang nyata.

Hasil presentase akurasi ekspresi gerakan wajah yang tertinggi yaitu marah 99%, ketakutan 57%, jijik 29%, senang 34%, sedih 47%, terkejut 66 % dan netral 23 %. Selain ekspresi wajah, ada karakteristik lain dalam film yang dapat digunakan untuk memperkirakan muatan emosional aktor. Oleh karena itu, sebagai penelitian di masa mendatang, kami berharap dapat menggunakan landmark wajah yang tertutup, menghadap kesamping atau tidak terlihat jelas dengan menggunakannya sebagai masukan untuk model proses selanjutnya. Informasi ini dapat dimanfaatkan sebagai regularisasi yang disematkan untuk menimbang informasi gerakan wajah aktor dalam klasifikasi emosi film yang disampaikan.

## E. Referensi

- [1] J. Almeida, L. Vilaça, I. N. Teixeira, and P. Viana, "Emotion Identification in Movies through Facial Expression Recognition," *Applied Sciences*, vol. 11, no. 15, p. 6827, Jul. 2021, doi: 10.3390/app11156827.
- [2] S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," *IEEE Trans Affect Comput*, vol. 13, no. 3, pp. 1195–1215, Jul. 2022, doi: 10.1109/TAFFC.2020.2981446.
- [3] W. Zhang, X. Zhang, and Y. Tang, "Facial expression recognition based on improved residual network," *IET Image Process*, vol. 17, no. 7, pp. 2005–2014, May 2023, doi: 10.1049/ipr2.12743.
- [4] A. F. Abate, L. Cimmino, B.-C. Mocanu, F. Narducci, and F. Pop, "The limitations for expression recognition in computer vision introduced by facial masks," *Multimed Tools Appl*, vol. 82, no. 8, pp. 11305–11319, Mar. 2023, doi: 10.1007/s11042-022-13559-8.
- [5] F. Pazhoohi, L. Forby, and A. Kingstone, "Facial masks affect emotion recognition in the general population and individuals with autistic traits," *PLoS One*, vol. 16, no. 9, p. e0257740, Sep. 2021, doi: 10.1371/journal.pone.0257740.
- [6] Y. Lou, "Multi-scale feature fusion facial expression recognition based on attention mechanism," in *International Conference on Image, Signal Processing, and Pattern Recognition (ISPP 2023)*, P. Batista and R. Bilas Pachori, Eds., SPIE, Jun. 2023, p. 49. doi: 10.1117/12.2680974.
- [7] K. Xu and A. Yao, "Accelerating Video Object Segmentation with Compressed Video." [Online]. Available: <https://github.com/kai422/CoVOS>
- [8] M. E. Irhebhude, A. O. Kolawole, and G. Nenbunmwa Amos, "Perspective on Dark-Skinned Emotion Recognition Using Deep-Learned and Handcrafted Feature Techniques," in *Emotion Recognition - Recent Advances, New Perspectives and Applications*, IntechOpen, 2023. doi: 10.5772/intechopen.109739.
- [9] B. Gan and C. Zhang, "Target Detection and Network Optimization: Deep Learning in Face Expression Feature Recognition," *J Sens*, vol. 2022, pp. 1–10, Jun. 2022, doi: 10.1155/2022/5423959.
- [10] A. Alharbi *et al.*, "A Novel text2IMG Mechanism of Credit Card Fraud Detection: A Deep Learning Approach," *Electronics (Basel)*, vol. 11, no. 5, p. 756, Mar. 2022, doi: 10.3390/electronics11050756.
- [11] I. A. Ahmed, E. M. Senan, H. S. A. Shatnawi, Z. M. Alkhraisha, and M. M. A. Al-Azzam, "Hybrid Techniques for the Diagnosis of Acute Lymphoblastic

- Leukemia Based on Fusion of CNN Features," *Diagnostics*, vol. 13, no. 6, p. 1026, Mar. 2023, doi: 10.3390/diagnostics13061026.
- [12] W. Islam, M. Jones, R. Faiz, N. Sadeghipour, Y. Qiu, and B. Zheng, "Improving Performance of Breast Lesion Classification Using a ResNet50 Model Optimized with a Novel Attention Mechanism," *Tomography*, vol. 8, no. 5, pp. 2411–2425, Sep. 2022, doi: 10.3390/tomography8050200.
- [13] C. M. Bhuma and R. Kongara, "Virus Texture Classification of TEM Images Using Fusion of Chebyshev Moments and Resnet50 Features," *Brazilian Archives of Biology and Technology*, vol. 65, 2022, doi: 10.1590/1678-4324-2022210636.
- [14] A. M. Al-Hejri, R. M. Al-Tam, M. Fazea, A. H. Sable, S. Lee, and M. A. Al-antari, "ETECADx: Ensemble Self-Attention Transformer Encoder for Breast Cancer Diagnosis Using Full-Field Digital X-ray Breast Images," *Diagnostics*, vol. 13, no. 1, p. 89, Dec. 2022, doi: 10.3390/diagnostics13010089.