

Klasifikasi Emosi Menggunakan Convolutional Neural Networks

Emotion Classification Based on Convolutional Neural Networks

Ripan Septian^{1*}, Dede Irawan Saputra², Susanto Sambasri³

^{1,2,3}Teknik Elektro Universitas Jenderal Achmad Yani

Jl. Terusan Jend. Sudirman, Cibeber, Cimahi Sel., Kota Cimahi, Jawa Barat 40525

Telp/Fax : 022 – 6656190/022 - 6652069

ripanseptian@gmail.com^{1*}, dedeirawan.saputra@lecture.unjani.ac.id², santo.sambasri@gmail.com³

Abstrak – Berbagai studi mengenai klasifikasi emosi telah dilakukan dengan menggunakan beberapa metode tertentu, namun masih terdapat permasalahan dalam performansi untuk proses data yang banyak, termasuk permasalahan dalam hal sistem yang overfit. Pada paper ini diimpelentasikan metode convolutional neural networks (CNN) untuk mengidentifikasi emosi pada citra wajah dengan tujuh jenis emosi yakni marah, senang, sedih, jijik, terkejut, takut dan netral. Metode CNN mampu mengekstraksi ciri tersembunyi pada citra dengan melakukan proses konvolusi, selain itu dalam pemrosesan data, metode ini dapat mengurangi overfit dibandingkan metode lainnya, dengan menggunakan augmentasi pada data serta penambahan regulasi seperti dropout. Dengan menggunakan dataset facial expression recognition (FER) dilakukan proses training dengan menganalisis perubahan nilai akurasi berdasarkan parameter batch, epoch, dan dropout. Hasil dari analisis di dapat parameter optimal batch 32, epoch 100 dan dropout 0.6 dengan akurasi training 61.51%, akurasi validasi 64.54%, training loss 4.29% dan validation loss 3.96%. Dari parameter optimal tersebut diujikan pada data asing sebanyak 3589 data dan menghasilkan nilai akurasi 67%, presisi 67% dan recall 66%.

Kata Kunci: overfit, convolutional neural networks, facial emotion recognition.

Abstract – Researchs about emotion classification have been done used many different methods, yet there was still problem about sistem performance when detecting amount of data, especially the overfit problems. In this paper convolutional neural networks (CNN) was implemented in order to identify seven kind of emotion, such as angry, happy, disgust, surprise, scared and neutral. CNN able to extract hidden features on the images using convolution method, other than that CNN able to reduce overfit by implementing data augmentation process and add regulation such as dropout. Facial expression recognition (FER) are used as training dataset. The acquired accuracy results then analysed based on the alteration of batch, epoch and dropout parameter in order to achieved optimum model. According to the analysis result, system gained optimum accuracy of training 61.51%, validation 64.54%, training loss 4.29% and validation loss 3.96% from parameter values of batch 32, epoch 100 and dropout 0.6. The optimum model then tested with 3589 foreign data and gained 67% of accuracy, 67% of precision and 66% of recall.

Keywords: over fit, convolutional neural networks, facial emotion recognition.

1. Pendahuluan

Emosi pada manusia khususnya ditunjukkan dengan ekspresi atau mimik pada wajah yang dimana ekspresi tersebut merupakan hasil dari pergerakan otot – otot pada wajah. Ekspresi pada wajah merupakan salah satu komunikasi non-verbal yang mengungkapkan pikiran yang sedang melintas pada diri seseorang. Sebagai contoh, sebuah senyuman menunjukkan keramah-tamahan, mengangkat alis menunjukkan ekspresi heran, mengernyitkan dahi menunjukkan ketakutan dan kegelisahan [1]. Saat ini teknologi biometrika sudah banyak digunakan, dimana teknologi ini memiliki dua karakteristik pembeda yaitu karakter fisik dan perilaku. Ekspresi wajah merupakan karakteristik perilaku sehingga penggunaan teknologi biometrika memungkinkan digunakan untuk pengenalan emosi pada seseorang.

Deteksi ekspresi wajah berbasis komputer sendiri merupakan satu ilmu yang selalu aktif dikaji sepanjang waktu. Beberapa penelitian yang berkaitan dengan pengenalan ekspresi wajah diantaranya mengembangkan sebuah penelitian yang membandingkan metode *Local Gabor Filter Bank and PCA plus LDA* [2] dengan menggunakan 138 citra data latih, berikutnya menggunakan *Fisherface* sebagai ekstraksi ciri dan *Backpropagation Neural Network* sebagai *Classifier* [3] dengan objek deteksi 240 citra, selanjutnya analisis ekspresi wajah dengan menggunakan *Gabor Wavelet Transform (GWT)* dan *DCT*, serta jaringan *Radial Basis Function (RBF)* sebagai *Classifier* [4], selain itu penelitian dengan menggunakan *2-D discrete cosine transform (DCT)* dan jaringan syaraf *FeedForward* dengan satu lapisan tersembunyi [5]. Beberapa penelitian yang tersebut rata-rata menggunakan *neural network* yang dikombinasikan beberapa metode lainnya. Aplikasi deteksi wajah tersebut dapat digunakan dalam berbagai bidang, salah satunya di bidang elektronika. Beberapa objek deteksi untuk penelitian diatas tidak lebih dari 500 citra data latih. Meskipun hasil dari penelitian tersebut mencapai akurasi 80% - 90% dengan data latih yang telah disebutkan, namun objek secara nyata memiliki variasi kompleksitas yang berbeda-beda. Maka dari itu dibutuhkan jumlah data yang banyak untuk meningkatkan ke-akuratan hasil deteksi, sehingga dapat meminimalisir *overfit* atau hasil akurasi *training* lebih besar dari hasil akurasi validasi dimana sistem tidak dapat mendeteksi data asing. *Convolutional Neural Network (CNN)* memiliki kemampuan untuk mengatasi masalah jumlah data yang banyak. *Alexnet* menggunakan *Deep Convolutional Neural Networks* untuk mengenali objek lebih dari 15 juta citra dengan lebih dari 22 ribu kategori [6]. Kemampuan CNN dapat dikendalikan dengan memvariasikan kedalaman dan lebarnya, dan juga CNN lebih banyak menduga sebuah objek dengan benar. Dibandingkan dengan standar *Feedforward Neural Network* dengan layer yang sama, CNN memiliki lebih sedikit koneksi dan parameter, sehingga lebih mudah di latih [6].

Berdasarkan permasalahan tersebut pada penelitian ini digunakan metode CNN yang dianggap lebih cocok untuk deteksi emosi pada citra ekspresi wajah. Hal ini karena CNN dapat menangani banyak data dengan kompleksitas tinggi serta beberapa variabel yang dapat dikendalikan untuk meningkatkan akurasi. Akurasi yang didapat dari sistem CNN ini adalah akurasi hasil perbandingan model dengan data asing yang belum pernah dilatih sebelumnya.

2. Metode Penelitian

Klasifikasi emosi menggunakan CNN diimplementasikan menjadi dua bagian, yaitu proses *training* dan deteksi. Proses *training* merupakan pembentukan model yang akan digunakan sebagai acuan klasifikasi, sedangkan proses deteksi merupakan penggunaan model terhadap data asing yang belum pernah digunakan. Pada proses perancangan digunakan bahasa pemrograman *python* dengan menggunakan beberapa *library* seperti *Keras*, *Open-CV*, *numPy*, dan *matplotlib*. *Backend* sistem untuk *neural network* menggunakan *Tensor Flow*.

Dataset untuk proses *training* menggunakan dataset *Facial Expression Recognition (FER)* dengan 35.888 total data citra *grayscale* dan ukuran 64 x 64 piksel dalam 7 kategori emosi yang sudah diklasifikasikan yakni marah, jijik, takut, senang, sedih, tekejut dan netral. Data sudah berupa nilai piksel setiap citra dalam bentuk *array* dengan format *.csv*, sehingga dapat lebih mempercepat proses *training*.

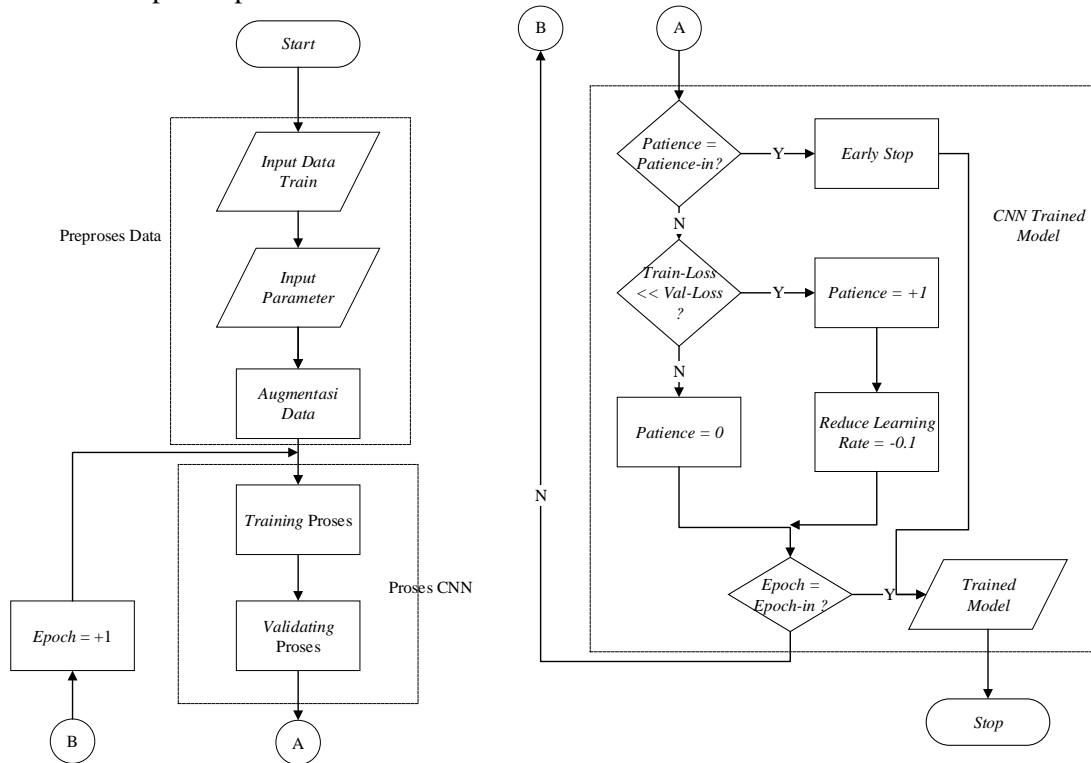
Pada Tabel 1 merupakan arsitektur CNN yang digunakan untuk melakukan proses *training* dengan standar *fully convolutional neural network* yang terdiri dari 10 lapisan konvolusi, ReLu,

Batch Normalization, *Dropout* dan *Global Max Pooling*. Arsitektur ini memiliki sekitar 600.000 parameter *training*. Nilai insialisasi *dropout* pada arsitektur diberikan dengan nilai 0.5 sebagai titik tengah antara 0 dan 1 [15]. *Adam optimizer* digunakan untuk mengontrol *learning rate* pada arsitektur dikarenakan nilai *input* sudah ditentukan sebelum *training* sehingga dibutuhkan *learning rate* yang *adaptive* atau terus berubah [7]. Selain itu digunakan *categorical cross entropy* untuk menghitung *loss* dari hasil *training* dikarenakan dataset yang digunakan memiliki beberapa kategori [7].

Tabel 1. Arsitektur CNN

Proses	Parameter
Konvolusi	Filters = 16, kernel = (7, 7), Padding = 1
Batch Normalization	
Konvolusi	Filters = 16, kernel = (7, 7), Padding = 1
Batch Normalization	
Activation Relu	
Max Pooling	Pool size = (2,2), padding = 1
Dropout	0.5
Konvolusi	Filters = 32, kernel = (5, 5), Padding = 1
Batch Normalization	
Konvolusi	Filters = 32, kernel = (5, 5), Padding = 1
Batch Normalization	
Activation Relu	
Max Pooling	Pool size = (2,2), padding = 1
Dropout	0.5
Konvolusi	Filters = 64, kernel = (3, 3), Padding = 1
Batch Normalization	
Konvolusi	Filters = 64, kernel = (3, 3), Padding = 1
Batch Normalization	
Activation Relu	
Max Pooling	Pool size = (2,2), padding = 1
Dropout	0.5
Konvolusi	Filters = 128, kernel = (3, 3), Padding = 1
Batch Normalization	
Konvolusi	Filters = 128, kernel = (3, 3), Padding = 1
Batch Normalization	
Activation Relu	
Max Pooling	Pool size = (2,2), padding = 1
Dropout	0.5
Konvolusi	Filters = 256, kernel = (3, 3), Padding = 1
Batch Normalization	
Konvolusi	Filters = 7, kernel = (3, 3), Padding = 1
Global Max Pooling	
Activation Softmax	

Arsitektur diimplementasikan pada proses *training* dengan alur proses seperti pada Gambar 1. Pada alur proses tersebut, proses *training* dibagi menjadi tiga blok proses yakni, preproses data, proses CNN dan proses pembentukan model.



Gambar 1. Flowchart proses *training*

Pada bagian preproses data dilakukan standarisasi atau normalisasi data agar nilai data yang diproses berada diantara -1 dan 1 dengan menggunakan persamaan (1) sehingga lebih mudah diproses oleh *neural network*. Dimana nilai x merupakan nilai normalisasi dan nilai p merupakan nilai awal dari data [8].

$$x = ((p \div 255) - 0,5) \times 2 \tag{1}$$

Selain itu dilakukan proses augmentasi data dengan melakukan alterasi berupa, rotasi citra, pergeseran citra pada sumbu x , pergeseran citra pada sumbu y , memperbesar dan memperkecil citra, serta membalikan gambar secara horizontal. Penentuan nilai parameter awal seperti *batch*, *epoch* dan *dropout* ditentukan pada tahap preproses data, untuk selanjutnya diubah dengan tujuan menemukan parameter optimal.

Para bagian proses CNN, data dibagi menjadi dua yakni 28.704 data untuk proses *training* dan 7.184 untuk proses validasi. Proses ini menghasilkan nilai akurasi dan nilai *loss* untuk proses *training* dan validasi yang selanjutnya akan dianalisa untuk mencari model optimal dengan akurasi validasi yang tinggi dan nilai *loss* yang rendah [10].

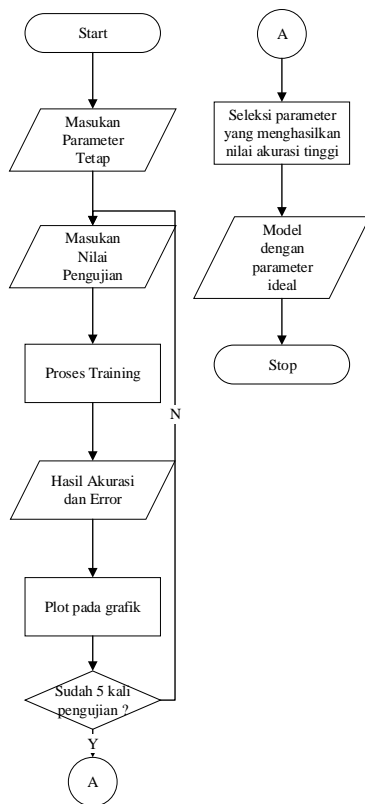
Pembentukan model pada proses CNN *trained model* melewati beberapa kondisi untuk mencegah model akhir mencapai *overfit* atau stagnansi. Variabel *patience* digunakan sebagai kondisi untuk mencegah *overfit*, dimana nilai ini merupakan maksimum *epoch* yang digunakan jika tidak ada perubahan pada nilai *loss* validasi [11]. Variabel ini terhubung dengan fungsi *early stop*, jika nilai *patience* mencapai maksimum maka fungsi *early stop* akan memaksa proses *training* berhenti. Selain itu jika nilai *patience* mencapai seperempat nilai maksimumnya fungsi lain bernama *reduce learning rate* akan mengurangi nilai *learning rate* sebesar 0,1 dari *adam optimizer* [12].

Proses *training* dilakukan berulang dengan melakukan perubahan pada parameter berdasarkan Tabel 2. Proses ini dilakukan dengan langkah pada Gambar 2 untuk menemukan nilai parameter optimal dari model yang akan dihasilkan. Proses *training* dilakukan pada server dengan

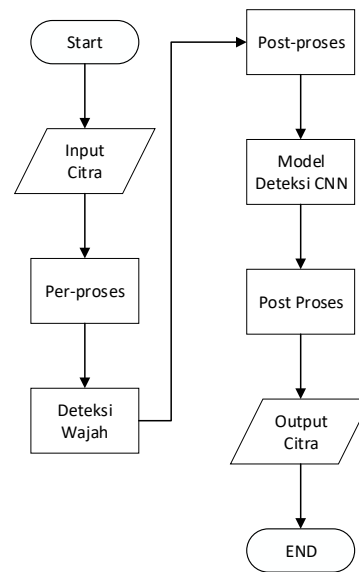
spesifikasi CPU intel i7 dengan GPU NVIDIA GTX 1080 Ti 16GB dan RAM 16GB 2400Mhz DDR4.

Tabel 2. Parameter Uji

Parameter Uji	Parameter Tetap			Nilai Percobaan				
	Batch Time	Epoch	Dropout	1	2	3	4	5
Batch Time	-	50	0	16	32	64	128	256
Epoch	Best Batch time	-	0	30	50	70	90	100
Drop Out	Best Batch time	Best Epoch	-	0,2	0,4	0,5	0,6	0,8



Gambar 2. Flowchart Akurasi



Gambar 3. Flowchart proses deteksi

Proses deteksi menggunakan *feed forward* dengan data yang dimasukkan pada model yang sudah optimal ditunjukkan pada *flowchart* Gambar 3. Proses ini terbagi menjadi dua yaitu proses deteksi wajah dan proses deteksi emosi. Pada proses deteksi wajah, digunakan metode *haar cascade classifier* untuk menghasilkan posisi wajah pada data yang akan diidentifikasi jenis emosinya [13][14]. Selanjutnya hasil deteksi wajah diidentifikasi dengan keluaran berupa label emosi yang sudah diputuskan oleh sistem.

Analisa nilai akurasi, presisi dan *recall*, dilakukan menggunakan metode iterasi data dengan 3.589 jumlah data citra asing yang belum dipakai pada proses *training* dan terbagi dalam 7 variasi emosi seperti ditunjukkan pada Tabel 3 [18]. Hasil deteksi pada setiap iterasinya akan masukan pada sebuah tabel *confussion matix* untuk selanjutnya dihitung nilai akurasi, presisi, dan *recall* menggunakan persamaan (2)(3) dan (4). Proses akuisisi data ditunjukkan pada *flowchart* Gambar 4.

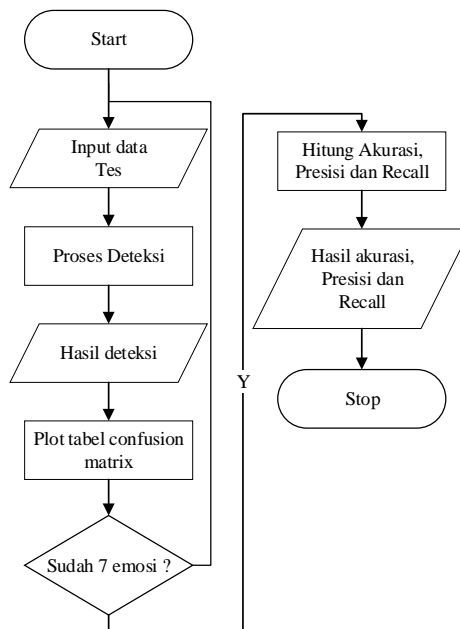
$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \tag{2}$$

$$Presisi = \frac{TP}{TP+FP} \tag{3}$$

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

Tabel 3. Jenis emosi

Emosi	Jumlah Data
Marah	467
Jijik	56
Takut	469
Senang	869
Sedih	653
Terkejut	415
Netral	607



Gambar 4. Flowchart akurasi deteksi

3. Hasil dan Pembahasan

Normalisasi data pada tahap preproses dengan menggunakan Persamaan (1) ditunjukkan pada Tabel 4. Nilai piksel data yang sebelumnya besar dinormalisasi agar berada pada jarak -1 sampai 1.

Tabel 4. Normalisasi Nilai Awal

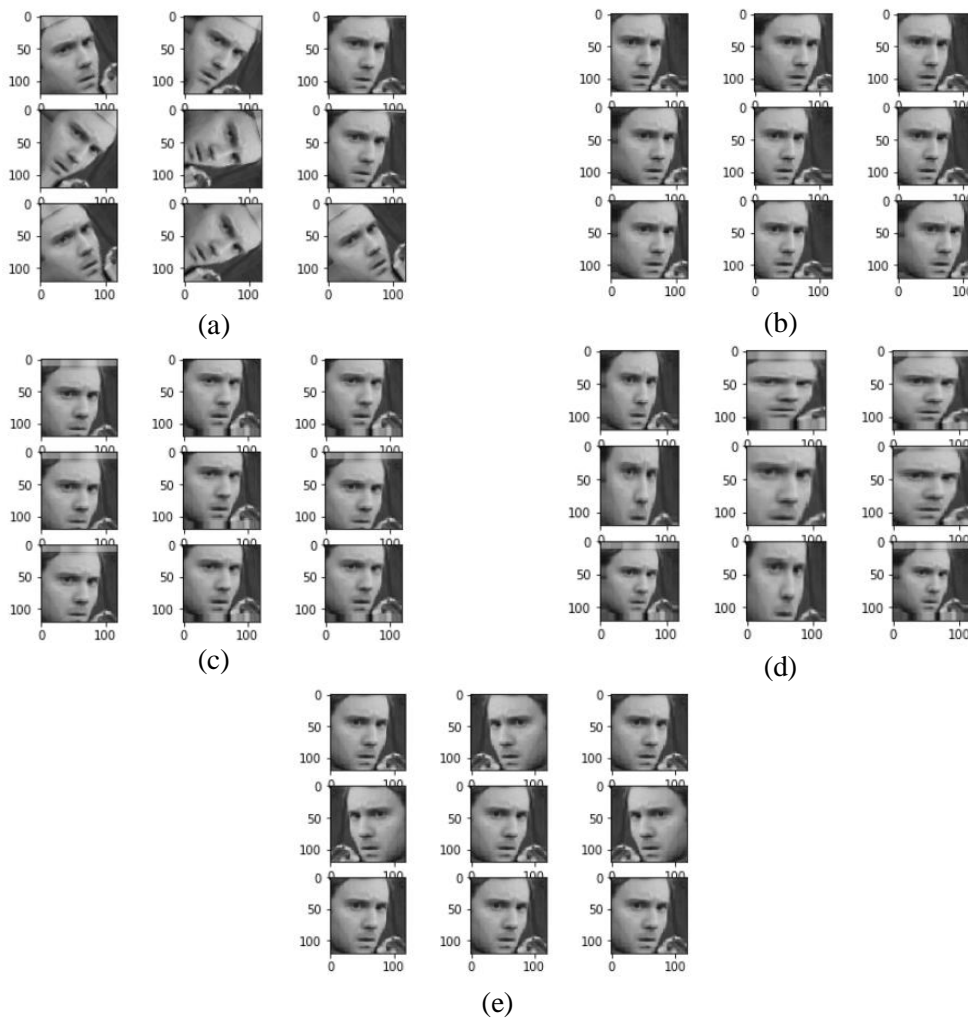
Nilai p	70	80	82	72	58	58	60	63	54	151	147	155
Persamaan	$x = ((p/255) - (0,5)) * 2$											
Niali x	-0,45	-0,37	-0,35	-0,4	-0,54	-0,5	-0,5	-0,5	-0,57	0,18	0,15	0,21

Augmentasi data memasukan nilai parameter untuk masing-masing augmentasi ditunjukkan pada Tabel 5 dan hasil dari augmentasi ditunjukkan pada Gambar 5. Untuk nilai rotasi dimasukan

nilai 90, dimana sistem akan melakukan rotasi citra secara acak pada data dengan rotasi antara -90° sampai 90°. Nilai pergeseran pada sumbu x diberikan dengan nilai -10,10 dimana citra akan bergeser dengan nilai 10 piksel ke kiri dan 10 piksel ke kanan, begitu pula dengan pergeseran sumbu y dengan parameter yang sama dimana citra akan bergeser keatas dan kebawah. Nilai pembesaran citra diberikan dengan nilai 0,7 dan 1,3 dimana data akan diubah perbesarannya dengan jarak antara 70% sampai 130%. Terakhir citra akan dibalikan secara horizontal, sehingga dengan menggunakan proses augmentasi didapat 5 variasi data citra *input* yang dihasilkan dari 1 citra *input*.

Tabel 5. Parameter Augmentasi

Parameter	Nilai
<i>Rotation_range</i>	90
<i>Width_shift_range</i>	-10,10
<i>Height_shift_range</i>	-10,10
<i>Zoom_range</i>	0,7,1,3
<i>Horizontal_flip</i>	TRUE



Gambar 5. Hasil Augmentasi data (a) Augmentasi Rotasi; (b) Pergeseran Sumbu x; (c) Pergeseran Sumbu y; (d) Perbersaran; (e) Pembalikan Horisontal

Setelah data disiapkan kemudian dilakukan proses *training* terhadap data dengan menggunakan variasi parameter uji pada Tabel 2. Hasil pelatihan terhadap beberapa parameter uji tersebut, didapatkan hasil akurasi dan *loss* yang dianggap paling baik untuk masing-masing skenario ditunjukkan pada Tabel 6 sampai dengan Tabel 8. Dari data tersebut dapat diambil parameter paling optimal dari tiga parameter *batch*, *epoch* dan *dropout* yang menghasilkan nilai paling baik dari semua skenario yakni dengan nilai parameter *batch* 32, *epoch* 100, dan *dropout* 0,6. Parameter tersebut menghasilkan nilai akurasi *training* 61.15%, akurasi validasi 64.54%, nilai *training loss* 4.29% dan *validation loss* 3.96%.

Tabel 6. Akumulasi Skenario I

Batch	acc	val_acc	loss	val_loss
16	54,91%	60,35%	4,96%	4,38%
32	55,15%	60,59%	4,93%	4,36%
64	55,11%	59,92%	4,91%	4,40%
128	52,14%	57,91%	5,21%	4,62%
256	55,03%	59,36%	4,97%	4,49%

Tabel 7. Akumulasi Skenario II

Epoch	acc	val_acc	loss	val_loss
70	56,20%	60,42%	4,82%	4,30%
90	56,61%	62,15%	4,77%	4,20%
100	57,60%	62,80%	4,65%	4,11%
120	54,71%	60,41%	4,93%	4,31%
150	55,52%	60,46%	4,90%	4,31%

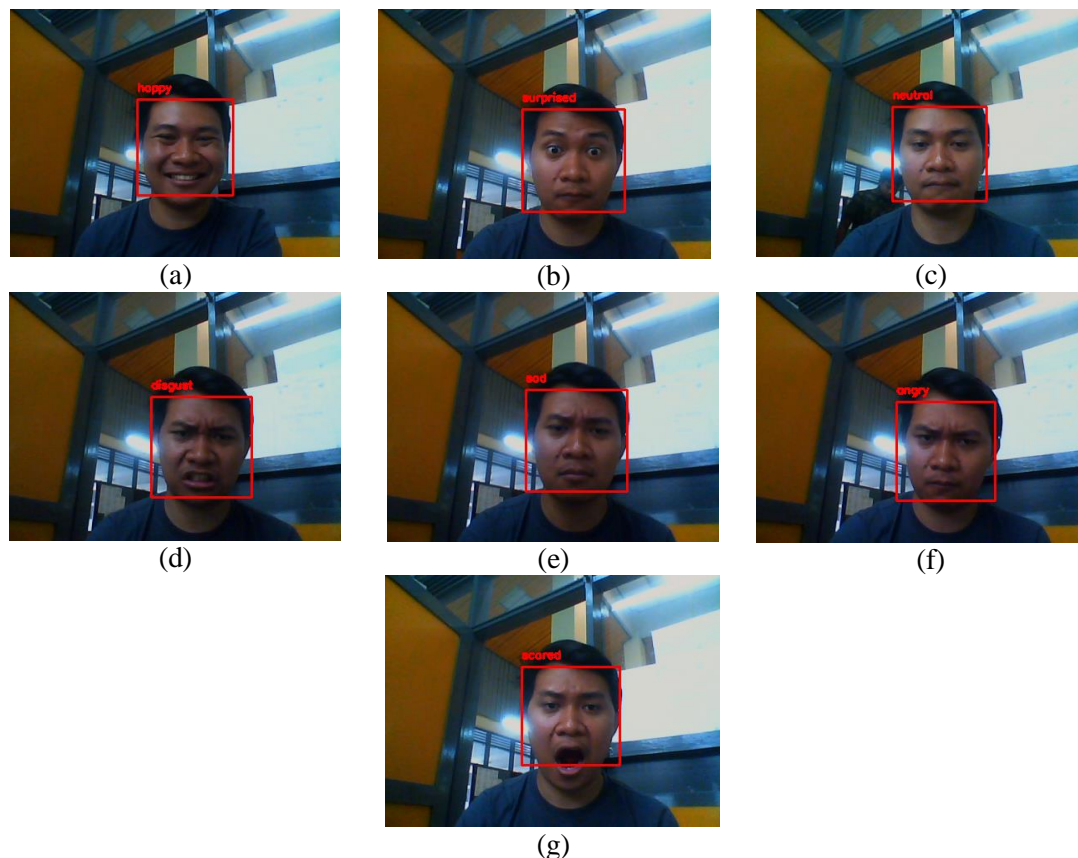
Tabel 8. Akumulasi Skenario III

Dropout	acc	val_acc	loss	val_loss
0.2	58,06%	62,78%	4,64%	4,14%
0.4	60,63%	64,22%	4,32%	4,00%
0.6	61,15%	64,54%	4,29%	3,96%
0.8	45,65%	52,01%	5,91%	5,15%
1	75,31%	66,72%	2,76%	4,12%

Hasil parameter optimal diujikan kembali pada data asing dengan hasil ditunjukkan pada Tabel 9. Nilai persentasi *cell* yang ditandai merupakan nilai *true positive* atau persentasi keberhasilan sistem dalam menebak emosi yang benar. Nilai persentasi pada baris dari tabel merupakan nilai *false positive* yakni nilai dimana sistem menebak emosi yang salah dengan benar, sedangkan pada nilai kolom merupakan nilai *false negative* yakni nilai dimana sistem menebak emosi yang salah dengan salah. Dari nilai-nilai persentasi tersebut dapat dihitung menggunakan persamaan (2), (3), dan (4) sehingga menghasilkan nilai akurasi deteksi 67%, nilai presisi 67% dan nilai *recall* 66%. Gambar 6 menunjukkan hasil dari deteksi pada 7 emosi yang ditentukan dengan menggunakan citra *input* secara *realtime*.

Tabel 9. Confussion Matrix

	Marah	Jijik	Takut	Senang	Sedih	Terkejut	Netral
Marah	70%	1%	10%	4%	2%	3%	10%
Jijik	26%	55%	5%	4%	5%	2%	3%
Takut	12%	1%	45%	5%	20%	7%	11%
Senang	2%	0%	2%	87%	2%	2%	5%
Sedih	9%	1%	10%	6%	55%	1%	19%
Terkejut	3%	0%	10%	5%	2%	80%	0%
Netral	5%	0%	5%	9%	3%	2%	75%



Gambar 6. Hasil Deteksi 7 Emosi (a) Senang; (b) Terkejut; (c) Netral; (d) Jijik; (e) Sedih; (f) Marah; (g) Terkejut.

4. Kesimpulan

Berdasarkan hasil dari penelitian, klasifikasi emosi menggunakan metode *convolutional neuralnetwork* (CNN) berhasil diimplementasikan untuk mendeteksi ketujuh emosi yakni marah, senang, sedih, jijik, terkejut, dan neutral. Dihasilkan nilai parameter optimal *batch* 32, *epoch* 100, dan *dropout* 0,6; dengan akurasi *training* 61,15%; akurasi validasi 64,54%; *training loss* 4,29%; dan validasi *loss* 3,96%; dari hasil akurasi tersebut dapat disimpulkan bahwa sistem tidak mengalami *overfit* karena nilai akurasi validasi lebih besar dari nilai akurasi *training*. Parameter Optimal tersebut diujikan kepada data asing yang belum pernah di deteksi sebelumnya dan menghasilkan nilai akurasi deteksi 67%, presisi 67% dan *recall* 66%.

Ucapan Terima Kasih

Terimakasih kepada Universitas Jenderal Achmad Yani yang memberikan kesempatan untuk penulisan dan pengembangan dari penelitian ini.

Referensi

- [1] Herlina, *Isyarat Wajah*. 2010.
- [2] Deng, H.B., Jin, L.W., Zhen, L.X., dan Huang, J.C., "A New Facial Expression Recognition Method Based On Local Gabor Filter Bank and PCA plus LDA", *international Journal of Information Technology*, 11, 11, 86 – 96, 2005.
- [3] Abidin, Z, dan Harjoko, A., "A Neural Network based Facial Expression Recognition using Fisherface". *International Journal of Computer Applications*, 59, 11, 30-44, 2012.

- [4] Lekshmi V., P., dan Sasikumar, M., “A Neural Network Based Facial Expression Analysis using Gabor Wavelets”, *World Academy of Science, Engineering and Technology*, Vol. 42, 563-567, 2008.
- [5] Kulkarni, S.S., “Facial Image Based Mood Recognition Using Committee Neural Networks”, *Thesis*, Department of Biomedical Engineering University of Akron, 2006.
- [6] Krizhevsky Alex, Sutskever Ilya, dan Hinton Geoffrey E, “ImageNet Classification with Deep Convolutional Neural Networks”, in *ImageNet Large Scale Visual Recognition Competition*, 2012.
- [7] Arriaga Octavio, Ploger Paul G., dan Valdenegro Matias, “Real-time Convolutional Neural Networks for Emotion and Gender Classification”, arXiv:1710.07557v1, 2017.
- [8] Fausett Lauren V., *Fundamentals of Neural Networks*, 1994.
- [9] Bishop, C. H., *Pattern Recognition and Machine Learning*, 2006.
- [10] Goodfellow Ian, Bengio Yoshua and Courville Aaron, *Deep Learning*, 2016.
- [11] Albawi Saad, Mohammed Tareq Abed dan Alzawi Saad, “Understanding of a Convolutional Neural Network”, *ICET*, 2017.
- [12] LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P., “Gradient-based learning applied to document recognition”. *Proceedings of the IEEE*, 86(11), pp.2278-2324, 1998.
- [13] Zufar Muhamad dan Setiyono Budi, “Convolutional Neural Networks untuk Pengenalan Wajah Secara Real-Time”, *Institut Teknologi Sepuluh Nopember, Jurnal Sains dan Seni ITS* Vol. 5, No.2, 2016.
- [14] P. I. Wilson dan D. J. Fernandez, “Facial Feature Detection Using Haar Classifiers”, *Journal of Computing Sciences in Colleges*, pp. 127-133, 2006.
- [15] Wu Haibing, Xiaodong Gu, “Max-Pooling Dropout for Regularization of Convolutional Neural Networks”, di *ICONIP*, 2015, Lecture Notes in Computer Science vol. 9489.
- [16] Powers David M. W., “Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation”, *Finders University of South America*, Technical Report SIE-07-001, 2007.