



## Implementasi *Convolutional Neural Network* untuk Klasifikasi Variasi Intensitas Emosi pada *Dynamic Image Sequence*

Lia Farokhah

Teknik Informatika, Fakultas Teknologi dan Desain, Institut Teknologi dan Bisnis ASIA Malang  
[lia.farokhah@gmail.com](mailto:lia.farokhah@gmail.com)

### Abstract

*Facial emotion recognition (FER) is a research topic that focuses on the analysis of human facial expressions. There are many FER research has been conducted on single images or photo. Emotion analysis on single images has many disadvantages compared to dynamic image sequences or videos. This is due to human emotions or expressions within a certain time. The classification of emotions becomes complicated when considering different emotions. There are some people who are very expressive, there are some people who have low or moderate expressions. Predictions of emotion with variety intensities has decreased error due to data sets that provide only a few emotions intensities. Data annotation is a major problem in recognition fields that require a lot of time and effort to annotate new data. This study aims to find information about facial emotions with emotional intensity from subtle to sharp in a sequence images or videos. The dataset will be trained using Convolutional neural network by augmentation to add data annotations. The proposed method was evaluated using the public BP4D-Spontaneous dataset. The evaluation results show that the average emotion recognition in video sequences using the holdout method is 18%. Evaluation of the loss function parameter shows overfitting where the curve generalization gap is too high. The last evaluation is the evaluation of the emotion class between the real class and the prediction class in 14.28%. This shows that the classification of emotion recognition in dynamic image sequences is quite low.*

*Keywords: Convolutional Neural Network, Face Emotion Recognition, Expression Intensities, Dynamic Image Sequence*

### Abstrak

*Face emotion recognition (FER) merupakan salah satu topik penelitian yang berfokus pada analisis pada ekspresi emosi wajah manusia. Penelitian FER sudah banyak dilakukan pada single image atau pembacaan emosi pada foto. Analisis emosi pada single image memiliki banyak kekurangan dibandingkan pada *dynamic image sequence* atau video. Hal ini dikarenakan karena manusia mengeluarkan emosi atau ekspresi dalam tempo waktu tertentu. Klasifikasi emosi manusia menjadi rumit ketika berhadapan dengan intensitas emosi yang berbeda-beda. Ada beberapa orang yang sangat ekspresif, ada beberapa orang yang memiliki ekspresi yang rendah atau sedang. Prediksi akurasi emosi manusia mengalami penurunan atau kesalahan karena dataset yang hanya menyediakan beberapa intensitas emosi. Anotasi data menjadi satu masalah besar bidang pengenalan yang membutuhkan banyak waktu dan tenaga untuk meng-anotasi data baru. Penelitian ini bertujuan untuk mencari akurasi klasifikasi emosi wajah dengan tantangan intensitas emosi dari halus sampai tajam dalam sebuah image sequence atau video. Datasets akan dilatih menggunakan *Convolutional neural network* dengan melakukan augmentasi untuk menambah anotasi data. Metode yang diusulkan dievaluasi menggunakan dataset publik BP4D-Spontaneous. Hasil evaluasi menunjukkan rata-rata pengenalan emosi pada video *sequence* menggunakan metode holdout sebesar 18%. Evaluasi parameter loss function menunjukkan overfitting dimana gap generalisasi kurva tinggi. Evaluasi terakhir adalah evaluasi kelas emosi antara kelas sebenarnya dan kelas prediksi sebesar 14,28%. Hal ini menunjukkan klasifikasi pada pengenalan emosi di *dynamic image sequence* cukup rendah.*

Kata kunci: Convolutional Neural Network, Pengenalan Ekspresi Wajah, Variasi ekspresi, Urutan gambar dinamis

### 1. Pendahuluan

Setiap orang terkadang memiliki keinginan atau perasaan yang terpendam. Ada banyak orang yang tidak mau menyampaikan secara langsung atau bahkan

sengaja menutupinya. Salah satu cara untuk membaca keinginan manusia secara tidak langsung adalah melihat ekspresi yang ditunjukkan melalui wajahnya. Ekspresi atau emosi wajah setiap orang memiliki makna yang bisa diterjemahkan. Ekspresi wajah pada seseorang menjadi

penting karena merupakan salah satu cara penyampaian keinginan atau perasaan secara non verbal [1-6]

Penelitian terkait pembacaan keinginan seseorang melalui ekspresi wajah memiliki kemudahan lebih besar dibandingkan menganalisis sikap seseorang dari sebuah kalimat yang diucapkan atau melalui gestur badan [1][7][5]. Hal ini karena ucapan bisa jadi tidak sesuai dengan apa yang diinginkan sedangkan gestur tubuh masih terbatas penelitiannya. Setiap manusia memiliki emosi atau ekspresi yang sama dalam mengekspresikan 6 emosi dasar seperti marah, bahagia, sedih, jijik, kaget dan takut [8][9][10]. Hal ini menunjukkan tema ini bisa diteliti lebih lanjut untuk bisa diimplementasikan diberbagai bidang.

Adapun salah satu masalah dalam klasifikasi emosi manusia adalah perbedaan intensitas ekspresi. Intensitas ekspresi berhubungan erat dengan emosi rendah dan halus. Dalam kehidupan nyata, Walaupun setiap orang memiliki ekspresi yang sama namun mereka memiliki intensitas ekspresi yang berbeda [8][11-14]. Ada orang yang sangat ekspresif atau istilahnya *peak expression*, ada orang yang memiliki ekspresi yang sedang atau rendah dimana disebut *non peak expression*. Cukup sulit melakukan klasifikasi emosi dengan intensitas emosi yang berbeda karena dataset yang ada tidak mendukung dan menyediakan keseluruhan emosi dari segala intensitas yang ada.

Penelitian yang banyak dilakukan adalah penelitian dengan melibatkan ekspresi tinggi atau *peak expression*. *Peak expression* menghasilkan akurasi yang bagus. Dalam kehidupan nyata, *non peak expression*/ ekspresi halus dan rendah lebih banyak terjadi daripada *peak expression* [11]. Metode *non peak expression* atau permasalahan intensitas emosi mulai dikembangkan untuk mengikuti kejadian di kehidupan nyata STM-Explet[15], LOMo[16], IACNN[17], BDBN[18], Facenet2expnet[19], DTAGN[20], PPDN[21]. Metode ini menangani tantangan bidang *Face Emotion Recognition* (FER) yaitu anotasi data. Anotasi data menjadi salah satu masalah yang cukup serius. Anotasi data menghabiskan banyak tenaga dan waktu dalam pelabelan [22-25]. Disisi lain, akurasi pengenalan emosi wajah dengan variasi intensitas pada *dynamic image sequence* masih memiliki tantangan[26].

Selain itu, berdasarkan fitur ekspresi, banyak penelitian yang membahas pengenalan emosi pada pada *single image* atau foto tunggal daripada sebuah *image sequence* atau video. Hal ini menjadi masalah karena emosi manusia terjadi dalam waktu atau tempo tertentu. Emosi manusia bisa dibaca lebih valid dalam sebuah video daripada *single image* atau foto [27][28][29][30][31].

Dalam *Face Emotion Recognition* (FER) evolusi dataset terjadi dari dataset *lab controlled* ke *wild condition*. Datasets pengenalan ekspresi wajah *lab controlled* lebih mudah dideteksi karena kondisi wajah terlihat jelas dari depan. Sedangkan dataset pengenalan wajah *wild condition* lebih cenderung menggambarkan ekspresi wajah yang memiliki banyak tantangan seperti

sudut pandang, pose kepala, intensitas emosi yang beragam dan masih banyak lagi. Datasets *wild condition* lebih menggambarkan ekspresi pada kehidupan nyata.

Adapun tujuan dalam penelitian ini yaitu mengklasifikasikan ekspresi atau emosi wajah pada *image sequence* dengan perbedaan intensitas yang ada serta halangan pose wajah. Tujuan implementasi jangka Panjang untuk layanan dalam berbagai bidang. Adapun contoh layanan seperti layanan produk, bincang konsultasi, keamanan transportasi dan masih banyak lagi. Untuk bertransformasi dari pengenalan emosi *single image* ke video atau *image sequence* diperlukan lebih banyak penelitian yang membahas tantangan dan akurasi yang ada sehingga pada akhirnya bisa diimplementasikan dan membawa manfaat yang besar.

Adapun metode yang digunakan dalam penelitian ini adalah menggunakan *deep learning convolutional neural network* (CNN). Pada proses pelatihan jaringan menggunakan CNN dilakukan *pre-trained* menggunakan imagenet dan augmentasi data.

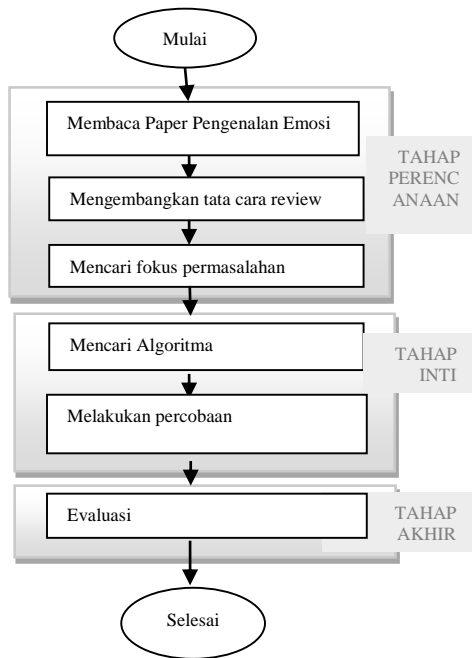
## 2. Metode Penelitian

### 2.1. Alur Penelitian

Adapun alur penelitian sangat penting digambarkan karena terkait protokol alur penelitian. Alur penelitian dibagi menjadi tiga yaitu tahap perencanaan, tahap inti dan tahap akhir. Alur penelitian dimulai dari mencari tema penelitian terkait pengenalan emosi wajah pada manusia. Kemudian dilanjutkan untuk melakukan arah penelitian pada paper rujukan yang didapatkan. Dari paper rujukan yang didapatkan, dianalisis bahwa pengenalan emosi secara fitur dibagi menjadi dua klasifikasi yaitu *single image* dan *image sequence* atau video.

Perbandingan akurasi antara penelitian *single image* dan *image sequence* atau video cukup signifikan. Pengenalan emosi pada *single image* lebih banyak dilakukan daripada pada pengenalan emosi pada sebuah video atau urutan gambar. Pada kenyataannya, tantangan pengenalan emosi terkait dalam sebuah urutan kejadian yang berkelanjutan. Pada tahap ketiga perencanaan dipilih fokus ke fitur *image sequence* dan pencarian dataset publik yang mendukung.

Pada tahap inti, dilakukan studi literatur tentang metode yang bisa menangani klasifikasi suatu objek. Metode pembelajarannya menggunakan *deep learning* dengan menerapkan *Convolutional Neural Network* atau CNN. Pada tahap evaluasi dilihat dari parameter *loss function* dan pengujian klasifikasi menggunakan *holdout* dimana dilakukan pembagian antara data training dan data testing. Selain itu, dilakukan evaluasi kelas emosi sebenarnya dibandingkan kelas prediksinya. Adapun alur penelitian secara lengkap ditunjukkan pada Gambar 1.



Gambar 1. Alur Penelitian

## 2.2 Metode Yang Diusulkan

Adapun metode yang dilakukan adalah melakukan augmentasi untuk menghasilkan data training baru dari data aslinya. Adapun tujuan augmentasi adalah meningkatkan generalisasi model. Augmentasi melebarkan distribusi normal data poin pada dataset. Selain itu pada penelitian ini menggunakan *pre-trained* imagenet. *Pre-trained* digunakan supaya jaringan kita tidak belajar terlalu dalam sehingga akan membuat hardware kita bekerja sangat keras. *Pre-trained* untuk arsitekturnya menggunakan VGG16. Pada training jaringan menggunakan adaptive learning ADAM.

### Program Arsitektur VGG16

```
vgg = VGG16(input_shape = IMAGE_SIZE + [3],
weights = 'imagenet', include_top = False)
```

```
for layer in vgg.layers:
    layer.trainable = False
```

```
x = Flatten()(vgg.output)
#x = Dense(128, activation = 'relu')(x) #
we can add a new fully connected layer but it
will increase the execution time.
x = Dense(num_classes, activation =
'softmax')(x) # adding the output layer with
softmax function as this is a multi label
classification problem.
```

```
model = Model(inputs = vgg.input, outputs =
x)
```

```
model.compile(loss='categorical_crossentropy',
optimizer='adam', metrics=['accuracy'])
```

### Adam Optimizer

```
m= beta1 * m + (1-beta1) * dw
```

$$v = \text{beta2} * v + (1-\text{beta2}) * (\text{dw}^{**2})$$

$$w += -\text{lr} * m / (\text{np.sqrt}(v) + \text{eps})$$

Adapun arsitekturnya ditunjukkan pada Gambar 2.



Gambar 2. Arsitektur VGG16

## 2.3. Datasets

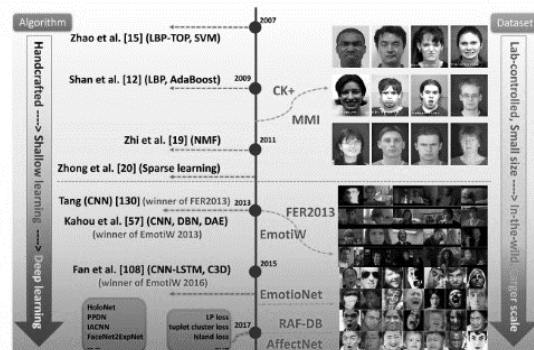
Data yang dipakai adalah dataset publik yang sudah diuji dan diduplikasi peneliti lain yaitu dataset BP4D Spontaneous [32][33]. Ada 6 Kelompok emosi yang diklasifikasikan yaitu *happy, surprise, anger, disgust, sadness, fear*. Adapun jumlah data poin tiap kelas adalah 500 dengan total data sekitar 3000 data. Adapun gambaran dataset ditunjukkan pada Gambar 3.



Gambar 3. Contoh Kondisi Dataset

## 2.4 Tantangan Dataset

Secara umum data yang dipakai adalah *dynamic image sequence* dengan *wild condition* atau mendekati dunia nyata. Dataset FER mempunyai dua jenis yaitu dataset *lab-controlled* dan dataset *wild condition*. Dataset *lab-controlled* merupakan dataset yang memperlihatkan emosi wajah seseorang secara jelas dari depan. Pada dataset *wild condition*, emosi wajah seseorang mengalami banyak tantangan seperti kurangnya emosi bisa terlihat karena pose kepala, halangan kaca yang dipakai atau posisi wajah yang tidak menghadap kedepan bahkan perbedaan variasi intensitas emosi pada kelas yang sama. Adapun perkembangan dataset ditunjukkan pada Gambar 4[26]. Selain itu, tantangan intensitas emosi ditunjukkan pada Gambar 5 [34].



Gambar 4. Tantangan Evolusi Dataset  
Sumber gambar: [26]



Gambar 5. Intensitas Emosi  
Sumber gambar: [12]

#### 2.4. Evaluasi

Adapun training menggunakan library python. Evaluasi pada penelitian ini menggunakan parameter loss function, dan metode validasinya holdout dimana membagi data menjadi dua dimana data training 80% dan data testing 20% dimana total data yang digunakan. Selain itu, dilakukan evaluasi dengan menggunakan prediksi label sebenarnya dengan prediksi dari hasil sistem. Evaluasi ini untuk melihat hasil prediksi emosi dengan kelas emosi sebenarnya.

### 3. Hasil dan Pembahasan

Adapun hasil training menggunakan Convolutional Neural Network atau CNN ditunjukkan pada Tabel 1.

Tabel 1. Hasil Akurasi Jaringan

	precision	recall	F1-score	support
Anger	0.15	0.16	0.16	102
Disgust	0.24	0.31	0.27	88
Fear	0.21	0.12	0.15	112
Happy	0.18	0.18	0.18	102
Sadness	0.17	0.24	0.20	101
surprise	0.11	0.09	0.10	69
accuracy			0.18	574
Macro	0.18	0.18	0.17	574
avg				
Weighted	0.18	0.18	0.18	574
avg				

#### Akurasi per kelas dalam klasifikasi emosi

```

y_pred = model.predict(validation_generator)
y_pred_bool = np.argmax(y_pred, axis=1)
target_names =
training_generator.class_indices

classes = list()
for i in target_names.keys():
    classes.append(i)

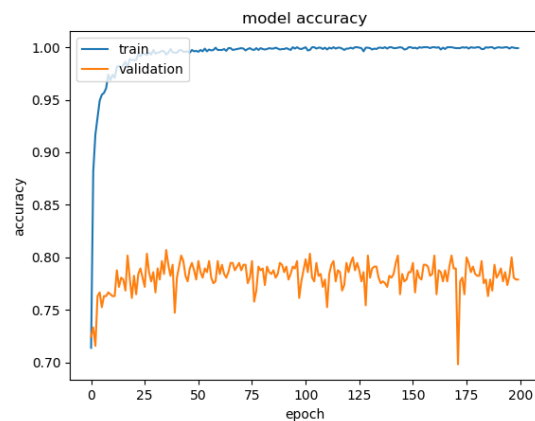
print(classification_report(y_true,
y_pred_bool, target_names=classes))

```

Pada hasil training menggunakan CNN terlihat bahwa akurasi rata-ratanya adalah 18%. Akurasi pada masing-masing kelas sebesar 15% pada kelas anger, 24% pada kelas disgust, 21% pada kelas fear, 18% pada kelas happy, 17% pada kelas sadness dan 11% pada kelas surprise. Hasil akurasi menunjukkan banyak sekali tantangan dalam pengenalan emosi pada sebuah video atau image sequence.

Akurasi yang rendah disebabkan oleh tantangan head pose pada dataset BP4D Spontaneous. Dataset BP4D merupakan transformasi dari dataset *lab-controlled* ke *wild condition*. Dataset Wild condition pada dynamic image sequence memiliki akurasi rendah dan tantangan lebih besar[26]. Hal ini dibuktikan dengan hasil akurasi yang ada publikasi terbaru systematic literature review deep learning bidang face emotion recognition[26].

Adapun evaluasi parameter loss function untuk melihat seberapa baik prediksi kelas label pada 6 emosi manusia dengan 200 *epoch* ditunjukkan pada Gambar 6 dan Gambar 7.



Gambar 6. Model Accuracy



Gambar 7. Model Loss

Pada gambar 6, model akurasi dibagi menjadi dua yaitu model akurasi data training dan data validation. Hasil percobaan menunjukkan bahwa model akurasinya mengalami overfitting dimana akurasi training jauh lebih besar daripada akurasi model validasi. Jaringan mengalami kesulitan dalam uji. Pada Data latih

generalisasi modelnya buruk sehingga tidak mempelajari pola. Menurut resep Andrew Ng's bahwa ada beberapa hal yang bisa dilakukan dengan hasil model loss seperti Gambar 7 yaitu memperbaiki generalisasi model, memperkuat *regularization*, data dan tata cara latih dan pengujian[35]. Pada penelitian ini, tidak dilakukan metode fine tuning dimana metode ini paling baik untuk mengurangi *overfitting*[35]. Adapun listing program untuk menggambarkan model akurasi sebagai berikut:

#### Program Model Akurasi

```

koi
print(history.history.keys())
# "Accuracy"
plt.plot(history.history['accuracy'])
)
plt.plot(history.history['val_accuracy'])
plt.title('model accuracy')
plt.ylabel('accuracy')
plt.xlabel('epoch')
plt.legend(['train', 'validation'],
loc='upper left')
plt.show()

```

Pada Gambar 8 yaitu parameter loss function menunjukkan overfitting dimana data training saat generalisasi model terlalu cepat drop kebawah sehingga terlihat tidak bisa mempelajari pola yang ada dalam dataset. Secara umum loss function yang baik adalah tidak overfitting atau underfitting dimana kedua kurva validasi dan training memiliki gap yang kecil [35].

#### Program Loss Function


```

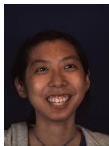
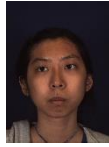


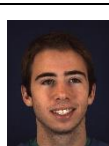

# "Loss Function"
plt.plot(history.history['loss'])
plt.plot(history.history['val_loss'])
plt.title('model loss')
plt.ylabel('loss')
plt.xlabel('epoch')
plt.legend(['train', 'validation'],
loc='upper left')
plt.show()

```

Adapun pengujian selanjutnya adalah pengujian kelas sebenarnya dengan kelas prediksi sistem. Dalam pengujian, dilakukan 6 klasifikasi emosi dari dataset BP4D Spontaneous dengan objek yang berbeda. Adapun detail pengujian ditunjukkan pada Tabel 2.

Tabel 2. Tabel Prediksi Emosi

Ekspresi	Kelas Sebenarnya	Prediksi
	Happy	Surprise

Ekspresi	Kelas Sebenarnya	Prediksi
	Happy	Surprise
	Sad	Disgust
	Sad	Sad
	Fear	Happy
	Happy	Surprise
	Fear	Surprise

Dari hasil evaluasi tabel 2 maka didapatkan akurasi prediksi sebesar 14,28%. Hasil akurasi sejalan dengan evaluasi *loss function* dan akurasi yang ada.

#### 4. Kesimpulan

Dari hasil percobaan disimpulkan bahwa klasifikasi pengenalan emosi pada *dynamic image sequence* dengan dataset BP4D Spontaneous memiliki tantangan yang cukup besar daripada *single image* atau foto. Hal ini dibuktikan dengan hasil akurasi rata-rata yang cukup rendah yaitu 18% dan pengujian berdasarkan kelas sebenarnya dan kelas prediksi sebesar 14,28%. Untuk meningkatkan akurasi dan mengecilkan gap pada loss model perlu mengikuti empat resep Andrew Ng's dalam mencari penyebab *overfitting* yaitu menguatkan model, *regularization*, penambahan data, mengecek cara pengujian dan pelatihan dataset [35]. Hal ini bertujuan

untuk menguatkan akurasi pengenalan emosi sehingga bisa di implementasikan dalam berbagai bidang. Adapun implikasi dan pekerjaan selanjutnya dari hasil penelitian ini adalah memperbanyak penelitian yang berfokus pada dataset image sequence atau video dibandingkan *single* image karena akurasi di bidang video atau image sequence terutama di dataset wild condition dengan tantangan intensitas emosi masih memiliki tantangan yang cukup besar dibandingkan pengenalan emosi pada *single* image.

## Daftar Rujukan

- [1] Y. Huang, F. Chen, S. Lv, and X. Wang, "Facial expression recognition: A survey," *Symmetry (Basel)*, vol. 11, no. 10, 2019, doi: 10.3390/sym11101189.
- [2] I. M. Revina and W. R. S. Emmanuel, "A Survey on Human Face Expression Recognition Techniques," *J. King Saud Univ. - Comput. Inf. Sci.*, 2018, doi: 10.1016/j.jksuci.2018.09.002.
- [3] O. Çelikütan, S. Ulukaya, and B. Sankur, "A comparative study of face landmarking techniques," *Eurasip J. Image Video Process.*, vol. 2013, no. 1, pp. 1–27, 2013, doi: 10.1186/1687-5281-2013-13.
- [4] S. P. Teja Reddy, S. Teja Karri, S. R. Dubey, and S. Mukherjee, "Spontaneous Facial Micro-Expression Recognition using 3D Spatiotemporal Convolutional Neural Networks," *Proc. Int. Jt. Conf. Neural Networks*, vol. 2019-July, pp. 1–8, 2019, doi: 10.1109/IJCNN.2019.8852419.
- [5] P. Ekman, "Darwin, Deception, and Facial Expression," vol. 221, pp. 205–221, 2003, doi: 10.1196/annals.1280.010.
- [6] D. A. Trevisan and E. Birmingham, "Are emotion recognition abilities related to everyday social functioning in ASD? A meta-analysis," *Res. Autism Spectr. Disord.*, vol. 32, pp. 24–42, 2016, doi: 10.1016/j.rasd.2016.08.004.
- [7] G. Sandbach, S. Zafeiriou, M. Pantic, and L. Yin, "Static and dynamic 3D facial expression recognition: A comprehensive survey," *Image Vis. Comput.*, vol. 30, no. 10, pp. 683–697, 2012, doi: 10.1016/j.imavis.2012.06.005.
- [8] S. L. Happy, A. Dantcheva, and F. Bremond, "A Weakly Supervised learning technique for classifying facial expressions," *Pattern Recognit. Lett.*, vol. 128, pp. 162–168, 2019, doi: 10.1016/j.patrec.2019.08.025.
- [9] Paul, Ekman., *The Facial Action Coding System: A Technique for Measurement of Facial Movement*. Palo Alto: Consulting Psychologist Press, 1978.
- [10] P. Ekman, "Darwin, Deception, and Facial Expression," *Ann. N. Y. Acad. Sci.*, vol. 1000, pp. 205–221, 2003, doi: 10.1196/annals.1280.010.
- [11] Q.-Y. Zhou, J. Park, and V. Koltun, "Peak-Piloted Deep Network for Facial Expression Recognition," *Eur. Conf. Comput. Vision (ECCV)*, vol. 9906, no. August, pp. 694–711, 2016, doi: 10.1007/978-3-319-46475-6.
- [12] Z. Yu, Q. Liu, and G. Liu, "Deeper cascaded peak-piloted network for weak expression recognition," *Vis. Comput.*, vol. 34, no. 12, pp. 1691–1699, 2018, doi: 10.1007/s00371-017-1443-0.
- [13] D. H. Kim, W. J. Baddar, J. Jang, and Y. M. Ro, "Multi-objective based spatio-temporal feature representation learning robust to expression intensity variations for facial expression recognition," *IEEE Trans. Affect. Comput.*, vol. 10, no. 2, pp. 223–236, 2019, doi: 10.1109/TAFFC.2017.2695999.
- [14] Y. Kim, B. Yoo, Y. Kwak, C. Choi, and J. Kim, "Deep generative-contrastive networks for facial expression recognition," in *International Conference on Multimedia Retrieval*, 2017, pp. 1–11, [Online]. Available: <http://arxiv.org/abs/1703.07140>.
- [15] M. Liu, S. Shan, R. Wang, and X. Chen, "Learning expressionlets on spatio-temporal manifold for dynamic facial expression recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1749–1756, doi: 10.1109/CVPR.2014.226.
- [16] K. Sikka, G. Sharma, and M. Bartlett, "LOMo: Latent ordinal model for facial analysis in videos," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, vol. 2016-Decem, pp. 5580–5589, doi: 10.1109/CVPR.2016.602.
- [17] Z. Meng, P. Liu, J. Cai, S. Han, and Y. Tong, "Identity-Aware Convolutional Neural Network for Facial Expression Recognition," in *Proceedings - 12th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2017 - 1st International Workshop on Adaptive Shot Learning for Gesture Understanding and Production, ASL4GUP 2017, Biometrics in the Wild, Bwild 2017, Heteroge*, 2017, pp. 558–565, doi: 10.1109/FG.2017.140.
- [18] P. Liu, S. Han, Z. Meng, and Y. Tong, "Facial expression recognition via a boosted deep belief network," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1805–1812, doi: 10.1109/CVPR.2014.233.
- [19] H. Ding, S. K. Zhou, and R. Chellappa, "FaceNet2ExpNet: Regularizing a Deep Face Recognition Net for Expression Recognition," in *Proceedings - 12th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2017 - 1st International Workshop on Adaptive Shot Learning for Gesture Understanding and Production, ASL4GUP 2017, Biometrics in the Wild, Bwild 2017, Heteroge*, 2017, pp. 118–126, doi: 10.1109/FG.2017.23.
- [20] H. Jung, S. Lee, J. Yim, S. Park, and J. Kim, "Joint fine-tuning in deep neural networks for facial expression recognition," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, vol. 2015 Inter, pp. 2983–2991, doi: 10.1109/ICCV.2015.341.
- [21] Q.-Y. Zhou, J. Park, and V. Koltun, "Peak-Piloted Deep Network for Facial Expression Recognition," *Eur. Conf. Comput. Vision (ECCV)*, vol. 9906, no. August, pp. 694–711, 2016, doi: 10.1007/978-3-319-46475-6.
- [22] C. F. Benitez-Quiroz, R. Srinivasan, and A. M. Martinez, "EmotioNet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 5562–5570, 2016, doi: 10.1109/CVPR.2016.600.
- [23] S. Li, W. Deng, and J. P. Du, "Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017, vol. 2017-Janua*, pp. 2584–2593, doi: 10.1109/CVPR.2017.277.
- [24] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild," *IEEE Trans. Affect. Comput.*, vol. 10, no. 1, pp. 18–31, 2019, doi: 10.1109/TAFFC.2017.2740923.
- [25] E. Barsoum, C. Zhang, C. C. Ferrer, and Z. Zhang, "Training deep networks for facial expression recognition with crowd-sourced label distribution," in *ICMI 2016 - Proceedings of the 18th ACM International Conference on Multimodal Interaction*, 2016, pp. 279–283, doi: 10.1145/2993148.2993165.
- [26] S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," *IEEE Trans. Affect. Comput.*, vol. 3045, no. c, pp. 1–1, 2020, doi: 10.1109/taffc.2020.2981446.
- [27] S. W. Chew, P. Lucey, S. Lucey, J. Saragih, J. F. Cohn, and S. Sridharan, "Person-independent facial expression detection using Constrained Local Models," in *2011 IEEE International Conference on Automatic Face and Gesture Recognition and Workshops, FG 2011, 2011*, pp. 915–920, doi: 10.1109/FG.2011.5771373.
- [28] A. Sanin, C. Sanderson, M. T. Harandi, and B. C. Lovell, "Spatio-temporal covariance descriptors for action and gesture recognition," in *Proceedings of IEEE Workshop on Applications of Computer Vision*, 2013, pp. 103–110, doi: 10.1109/WACV.2013.6475006.
- [29] Y. Peng, L. Qingshan, and D. N. Metaxas, "Facial expression recognition using encoded dynamic features," in *Proceedings of the 2007 IEEE International Conference on Multimedia and Expo, ICME 2007, 2007*, pp. 1107–1110, doi: 10.1109/icme.2007.4284848.

- [30] Z. Guoying, "Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, p. 9, 2007.
- [31] G. Zhao, X. Huang, M. Taini, S. Z. Li, and M. Pietikäinen, "Facial expression recognition from near-infrared videos," *Image Vis. Comput.*, vol. 29, no. 9, pp. 607–619, 2011, doi: 10.1016/j.imavis.2011.07.002.
- [32] Xing Zhang, Lijun Yin, Jeff Cohn, Shaun Canavan, Michael Reale, Andy Horowitz, Peng Liu, and Jeff Girard, "BP4D-Spontaneous: A high resolution spontaneous 3D dynamic facial expression database", *Image and Vision Computing*, 32 (2014), pp. 692-706 (special issue of the Best of FG13)
- [33] Xing Zhang, Lijun Yin, Jeff Cohn, Shaun Canavan, Michael Reale, Andy Horowitz, and Peng Liu, "A high resolution spontaneous 3D dynamic facial expression database", *The 10th IEEE International Conference on Automatic Face and Gesture Recognition (FG13)*, April, 2013.
- [34] Z. Yu, Q. Liu, and G. Liu, "Deeper cascaded peak-piloted network for weak expression recognition," *Vis. Comput.*, vol. 34, no. 12, pp. 1691–1699, 2018, doi: 10.1007/s00371-017-1443-0
- [35] A. Rosebrock, *Deep Learning for Computer Vision With Python*, 2nd ed. United States of America: PyImageSearch, 2018.