

Penggunaan Deep Convolutional Neural Network (CNN) Model FaceNet untuk Fungsi Face Recognition pada Reverse Image Search

Muhammad Rayhan Rachmansyah¹, Basuki Rahmat^{2*}, Henni Endah Wahanani³

^{1,2,3} Informatika, Universitas Pembangunan Nasional “Veteran” Jawa Timur

¹119081010002@student.upnjatim.ac.id

³henniendah.if@upnjatim.ac.id

*Corresponding author email: basukirahmat.if@upnjatim.ac.id

Abstrak— Kemajuan teknologi digital dan meningkatnya penggunaan internet telah menimbulkan kebutuhan akan sistem yang mampu melindungi privasi pengguna, khususnya terkait penyebaran foto pribadi di dunia maya. Penelitian ini bertujuan untuk mengimplementasikan metode *Deep Convolutional Neural Network* (CNN) model FaceNet sebagai sistem pengenalan wajah dalam *Reverse Image Search engine*. Metode FaceNet digunakan untuk mengekstraksi fitur wajah dan menghasilkan representasi vektor (*embedding*) yang dapat dibandingkan untuk mengenali kemiripan antar wajah. Proses penelitian mencakup studi pustaka, pengumpulan data, pelatihan model FaceNet, serta evaluasi performa sistem dalam mengenali dan mencocokkan wajah. Model FaceNet memanfaatkan arsitektur Zeiler & Fergus serta *Inception Network* dengan fungsi kerugian *Triplet Loss*, yang memastikan *embedding* dari wajah yang sama memiliki jarak yang lebih dekat dibandingkan dengan wajah dari identitas berbeda. Hasil implementasi menunjukkan bahwa integrasi FaceNet dalam *Reverse Image Search* mampu meningkatkan akurasi pencarian gambar berdasarkan wajah dan mempermudah identifikasi gambar serupa secara efisien. Penelitian ini diharapkan dapat menjadi dasar bagi pengembangan sistem pelacakan identitas visual dan deteksi penggunaan gambar di berbagai platform daring.

Kata Kunci— *Deep Convolutional Neural Network, FaceNet, Reverse Image Search, Triplet Loss, Pengenalan Wajah.*

I. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi dalam beberapa tahun terakhir telah membawa perubahan signifikan terhadap cara manusia mengakses, memproses, dan membagikan informasi digital. Berdasarkan data Statista (2023), jumlah pengguna internet global telah mencapai lebih dari 5 miliar pengguna, dan Indonesia menempati posisi keempat dunia dengan sekitar 212,9 juta pengguna aktif. Peningkatan masif ini tidak hanya memberikan kemudahan dalam pertukaran informasi, tetapi juga menimbulkan tantangan baru dalam aspek keamanan data dan privasi digital. Salah satu permasalahan yang sering terjadi adalah maraknya terjadi kejahatan dalam dunia maya, yang dapat berdampak pada pencurian identitas maupun distribusi konten tanpa izin.

Teknologi *Reverse Image Search* hadir sebagai solusi untuk melakukan pencarian berbasis gambar, bukan teks. Melalui teknologi ini, pengguna dapat mengunggah sebuah gambar untuk menemukan informasi atau gambar serupa di internet. Namun, sebagian besar sistem *Reverse Image Search* konvensional masih bergantung pada kesamaan visual tingkat rendah seperti warna atau tekstur, sehingga belum optimal dalam mengenali identitas wajah secara spesifik.

Untuk mengatasi keterbatasan tersebut, penelitian ini mengimplementasikan metode *Deep Convolutional Neural Network* (CNN) menggunakan model FaceNet sebagai sistem pencarian wajah dalam *Reverse Image Search*. Model FaceNet mampu mengekstraksi fitur wajah dan merepresentasikannya dalam bentuk *embedding* vektor berdimensi tinggi, sehingga setiap wajah dapat diidentifikasi secara unik. Selain itu, penggunaan fungsi kerugian *Triplet Loss* pada proses pelatihan memastikan bahwa representasi wajah dari individu yang sama memiliki jarak yang lebih dekat dibandingkan wajah dari individu yang berbeda di ruang vektor.

Penelitian ini bertujuan untuk mengintegrasikan model FaceNet ke dalam sistem *Reverse Image Search* guna meningkatkan akurasi dan efisiensi pencarian gambar berbasis wajah. Evaluasi dilakukan melalui proses pelatihan dan pengujian model terhadap dataset gambar wajah untuk menilai tingkat keakuratan sistem dalam mengenali kemiripan wajah.

II. TINJAUAN PUSTAKA

Penelitian mengenai pengenalan wajah berbasis pembelajaran mendalam (*deep learning*) telah mengalami perkembangan pesat dalam satu dekade terakhir. Salah satu pendekatan paling efektif dalam pemrosesan citra digital adalah *Convolutional Neural Network* (CNN), yang terbukti mampu mengekstraksi fitur spasial dari gambar secara efisien.

LeCun [1] menjelaskan bahwa CNN merupakan jaringan saraf tiruan berlapis yang dirancang untuk memproses data berbentuk grid seperti gambar digital. CNN terdiri atas beberapa lapisan utama, yaitu *convolutional layer*, *pooling*

layer, fully connected layer, serta activation functions yang berperan dalam mendeteksi pola visual seperti tepi, bentuk, dan tekstur pada objek. Struktur berlapis ini memungkinkan CNN untuk mengenali fitur kompleks yang sulit dilakukan oleh metode konvensional. Selain itu, penelitian oleh Yi [6] menunjukkan bahwa representasi wajah yang dilatih dari awal dengan dataset besar dapat meningkatkan kemampuan generalisasi model CNN secara signifikan.

Krizhevsky [2] kemudian memperkenalkan AlexNet, arsitektur CNN yang memenangkan kompetisi *ImageNet Large Scale Visual Recognition Challenge* (ILSVRC) 2012. Model ini menunjukkan peningkatan signifikan dalam akurasi klasifikasi gambar dengan memanfaatkan GPU untuk mempercepat proses pelatihan jaringan. Keberhasilan AlexNet menjadi tonggak penting bagi pengembangan model CNN yang lebih dalam dan kompleks di tahun-tahun berikutnya.

Salah satu inovasi penting dalam pengenalan wajah berbasis CNN adalah model FaceNet, yang dikembangkan oleh Schroff [3]. FaceNet memanfaatkan pendekatan *end-to-end learning* untuk menghasilkan representasi wajah dalam bentuk embedding berdimensi tinggi. Model ini menggunakan fungsi kerugian *Triplet Loss*, yang memastikan bahwa jarak antara dua wajah dari individu yang sama lebih kecil dibandingkan jarak antara wajah dari individu yang berbeda. Pendekatan ini terbukti efektif untuk berbagai tugas seperti pengenalan, verifikasi, dan pengelompokan wajah. Selain itu, Chopra [8] mengembangkan metode *contrastive loss* untuk pembelajaran jarak antar citra wajah yang serupa dan tidak serupa, yang kemudian menjadi dasar bagi metode *Triplet Loss* seperti pada FaceNet.

Selain itu, Zeiler dan Fergus [4] memperkenalkan teknik visualisasi jaringan CNN untuk memahami proses ekstraksi fitur pada setiap lapisan. Penelitian mereka membantu memperjelas bagaimana CNN mengenali pola dan fitur wajah secara bertingkat, dari fitur sederhana (tepi, tekstur) hingga fitur kompleks (bentuk mata, hidung, dan struktur wajah). Untuk meningkatkan tahap deteksi dan penyelarasan wajah sebelum proses ekstraksi fitur, Zhang [7] memperkenalkan *Multitask Cascaded Convolutional Networks* (MTCNN) yang terbukti efektif dalam mendeteksi dan menormalkan wajah dengan akurasi tinggi.

Dalam konteks pencarian berbasis gambar, teknologi *Reverse Image Search* pertama kali diperkenalkan oleh *TinEye* pada tahun 2008. Teknologi ini memungkinkan pencarian gambar serupa berdasarkan konten visual, bukan teks. Namun, sistem *Reverse Image Search* konvensional umumnya masih menggunakan pendekatan feature matching sederhana seperti color histogram atau *SIFT descriptors*, yang kurang efektif dalam membedakan wajah antar individu. Sebelum era *deep learning*, pendekatan berbasis fitur lokal seperti *Local Binary Pattern* (LBP) telah digunakan secara luas untuk pengenalan

wajah [11], namun metode ini kurang efektif menghadapi variasi pencahayaan dan sudut pandang.

Integrasi antara metode FaceNet dan *Reverse Image Search* membuka peluang baru untuk sistem pencarian wajah berbasis konten (*content-based face retrieval*). Dengan memanfaatkan *embedding* hasil ekstraksi FaceNet, sistem dapat melakukan pencarian berdasarkan kesamaan fitur wajah yang direpresentasikan dalam ruang vektor multidimensi. Beberapa penelitian menunjukkan bahwa pendekatan ini mampu meningkatkan akurasi pencarian hingga lebih dari 90% pada dataset wajah publik seperti *Labeled Faces in the Wild* (LFW) dan *YouTube Faces Database* [3], [5].

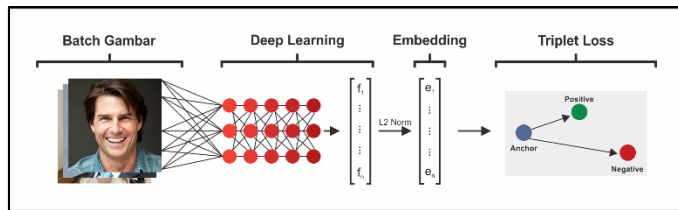
Beberapa penelitian lanjutan seperti Wu [12] dan Deng [13] mengusulkan varian *loss function* baru seperti *Additive Angular Margin Loss* (ArcFace) untuk meningkatkan pemisahan antar kelas pada *embedding* wajah. Metode-metode ini memberikan hasil yang lebih stabil pada dataset besar dan bervariasi. Selain itu, penelitian Ge [14] membahas pengenalan wajah pada citra resolusi rendah menggunakan *Selective Knowledge Distillation*, yang relevan untuk penerapan pada sistem pencarian berbasis gambar. Sementara itu, Philbin [15] memperkenalkan pendekatan *object retrieval* berskala besar yang menjadi inspirasi bagi sistem *Reverse Image Search* modern.

Dari berbagai penelitian tersebut, dapat disimpulkan bahwa implementasi model *Deep CNN* FaceNet dalam sistem *Reverse Image Search* memiliki potensi besar dalam meningkatkan efisiensi dan keakuratan pencarian gambar berbasis wajah. Namun, masih diperlukan penelitian lebih lanjut untuk mengoptimalkan parameter pelatihan, mengurangi kesalahan prediksi (*false match*), serta menyesuaikan sistem agar mampu bekerja pada data wajah dengan variasi pencahayaan, sudut, dan ekspresi yang beragam.

III. METODOLOGI PENELITIAN

A. Gambaran Umum Sistem

Penelitian ini bertujuan untuk mengimplementasikan metode *Deep Convolutional Neural Network* (CNN) dengan model FaceNet sebagai sistem pencarian wajah dalam *reverse image search engine*. Sistem dirancang untuk menerima masukan berupa gambar wajah pengguna, kemudian melakukan ekstraksi fitur menggunakan FaceNet, dan menghasilkan representasi numerik (*embedding vector*) yang akan dibandingkan dengan data referensi dalam basis data untuk menemukan gambar wajah yang serupa, seperti yang diilustrasikan pada gambar 1 yang dilampirkan.



Gbr. 1 Proses Kerja Arsitektur FaceNet

Secara umum, sistem ini terdiri atas empat tahapan utama:

1. *Preprocessing* Data Wajah, yaitu deteksi dan normalisasi citra wajah.
2. Ekstraksi Fitur Menggunakan FaceNet, untuk menghasilkan embedding berukuran 128 dimensi.
3. Penyimpanan dan Pencocokan Embedding, untuk membandingkan vektor wajah baru dengan basis data menggunakan jarak Euclidean.
4. Visualisasi Hasil Pencarian, menampilkan gambar yang memiliki tingkat kemiripan tertinggi.

B. Arsitektur Model FaceNet

Model FaceNet merupakan sistem pengenalan wajah berbasis deep learning yang dikembangkan oleh Google Research [1]. Model ini menggunakan arsitektur Deep CNN seperti Zeiler & Fergus Network dan Inception Network sebagai backbone untuk mengekstraksi fitur wajah.

Model seperti Inception Network yang dikembangkan oleh Szegedy [9] berhasil memperdalam struktur jaringan CNN tanpa menambah beban komputasi secara signifikan, menjadikannya pondasi kuat untuk model seperti FaceNet. Bahkan model modern seperti *Transformer* juga mulai diterapkan untuk pengenalan wajah, sebagaimana dijelaskan oleh Vaswani [10], meskipun CNN masih menjadi pilihan utama untuk efisiensi spasial.

FaceNet bekerja dengan menghasilkan embedding vector berdimensi 128 untuk setiap citra wajah. Embedding ini diperoleh melalui proses pelatihan dengan fungsi kerugian Triplet Loss, yang didefinisikan sebagai berikut:

$$L(A, P, N) = \max \{d(A, P) - d(A, N) + \alpha, 0\} \quad (1)$$

Keterangan:

A = Anchor image (gambar acuan)

P = Positive image (gambar wajah dari individu yang sama)

N = Negative image (gambar wajah dari individu berbeda)

$d(x, y)$ = jarak antara dua embedding, diukur menggunakan Euclidean Distance

α = margin konstan untuk memisahkan jarak antar identitas

Tujuan dari Triplet Loss adalah memastikan bahwa embedding dari wajah yang sama memiliki jarak yang lebih kecil

dibandingkan wajah dari individu berbeda. Dengan demikian, sistem dapat membedakan identitas wajah berdasarkan kedekatan vektornya dalam ruang fitur.

C. Pengolahan Data dan Proses Pelatihan

Dataset yang digunakan dalam penelitian ini terdiri dari kumpulan gambar wajah dengan berbagai variasi pencahayaan, ekspresi, dan sudut pandang. Sebelum dilakukan pelatihan, data wajah melewati beberapa tahap preprocessing, yaitu:

- 1) Deteksi wajah menggunakan algoritma seperti MTCNN [7] untuk mengidentifikasi area wajah pada gambar.
- 2) Normalisasi ukuran gambar menjadi dimensi seragam (misalnya 160×160 piksel).
- 3) Augmentasi Data untuk meningkatkan generalisasi model dengan melakukan rotasi, flipping, dan perubahan kontras, sebagaimana direkomendasikan oleh Yi [6].
- 4) Ekstraksi Fitur dengan mengubah setiap gambar wajah menjadi *embedding* berdimensi 128 menggunakan model FaceNet yang telah dilatih sebelumnya (pre-trained model).

Selama pelatihan, model dioptimalkan menggunakan *Adam Optimizer* dengan learning rate konstan, sementara jumlah epoch ditentukan berdasarkan konvergensi fungsi *loss*.

D. Alur Implementasi Sistem

Alur implementasi sistem dapat dijelaskan melalui langkah-langkah berikut:

- 1) *Input Image*: pengguna mengunggah gambar wajah ke sistem.
- 2) *Face Detection & Preprocessing*: sistem mendeteksi wajah dan menyesuaikan ukuran serta orientasi gambar.
- 3) *Feature Extraction* (FaceNet): model FaceNet menghasilkan embedding vektor dari gambar tersebut.
- 4) *Vector Comparison*: embedding hasil input dibandingkan dengan embedding dalam basis data menggunakan *Euclidean Distance*.
- 5) *Result Retrieval*: sistem mengembalikan daftar gambar wajah yang memiliki jarak paling kecil (kemiripan tertinggi).

Dengan pendekatan ini, sistem reverse image search dapat melakukan pencarian wajah secara efisien dengan waktu komputasi yang relatif rendah, karena perbandingan dilakukan di ruang vektor, bukan berdasarkan citra mentah.

E. Evaluasi Kinerja Sistem

Evaluasi dilakukan untuk menilai akurasi, efisiensi, dan reliabilitas sistem dalam melakukan pencarian wajah. Beberapa metrik yang digunakan antara lain:

- a) Akurasi Pengenalan (*Recognition Accuracy*): persentase wajah yang berhasil dicocokkan dengan identitas yang benar.
- b) *Precision dan Recall*: untuk menilai sejauh mana sistem mampu mendeteksi wajah yang relevan tanpa kesalahan positif.
- c) *Mean Squared Error (MSE)* dan *Euclidean Distance*: digunakan untuk mengukur kedekatan embedding antara gambar uji dan referensi.

Selain itu, waktu eksekusi sistem juga dianalisis untuk memastikan bahwa proses pencarian dapat dilakukan secara real-time atau mendekati real-time.

IV. HASIL DAN PEMBAHASAN

A. Hasil Implementasi Sistem

Penelitian ini menghasilkan sistem *Reverse Image Search* berbasis *Deep Convolutional Neural Network (CNN)* dengan model FaceNet yang mampu melakukan pencarian wajah berdasarkan kemiripan fitur. Implementasi dilakukan menggunakan bahasa pemrograman *Python* dengan pustaka TensorFlow dan Keras sebagai kerangka kerja utama untuk pengolahan model *deep learning*.

Dataset wajah yang digunakan terdiri dari 2.300 citra wajah dari beberapa individu dengan variasi ekspresi, pencahayaan, dan sudut pandang. Model FaceNet yang digunakan merupakan versi *pre-trained* yang kemudian di-*fine-tune* menggunakan subset dataset untuk menyesuaikan karakteristik data lokal.

Sistem dirancang agar pengguna dapat mengunggah gambar wajah, kemudian sistem akan menghasilkan *embedding* vektor berdimensi 128 dan membandingkannya dengan basis data. Proses pencarian dilakukan dengan menghitung jarak *Euclidean (L2 distance)* antara *embedding* input dan *embedding* referensi. Gambar dengan jarak terendah dianggap paling mirip dengan wajah yang dicari.

B. Hasil Pengujian Model

Evaluasi performa dilakukan menggunakan 100 gambar uji yang terdiri dari wajah-wajah yang telah dikenal dan wajah baru yang tidak termasuk dalam data pelatihan. Hasil pengujian menunjukkan bahwa model FaceNet mampu mengidentifikasi kemiripan wajah dengan tingkat akurasi yang tinggi.



Gbr. 2 Hasil Pengujian Sama Identitas

Pada hasil pengujian sama identitas, diperoleh hasil yang diharapkan oleh penulis seperti yang sudah dilampirkan pada gambar 2. Penulis menggunakan foto wajah dari aktor ternama Tom Cruise sebagai data uji yang dimana wajah dari aktor tersebut juga digunakan sebagai data latih dan sudah termasuk dari dataset yang digunakan. Diperoleh hasil gambar wajah dari identitas yang sama dan memiliki tingkat kemiripan cukup tinggi sesuai dengan ekspektasi penulis.



Gbr. 3 Hasil Pengujian Berbeda Identitas

Sedangkan, pada hasil pengujian berbeda identitas, penulis menggunakan foto wajah dari aktor Iko Uwais yang dimana wajah aktor tersebut tidak terdapat pada dataset yang digunakan. Hasil menunjukkan tingkat kemiripan sangat kontras dari hasil uji sebelumnya, untuk lebih jelasnya, hasil keluaran program dapat dilihat pada gambar 3 yang sudah dilampirkan. Hasil tertinggi pada pengujian ini hanya sekitar 0.5 untuk tingkat kemiripannya. Dimana hasil kemiripan yang berada pada jarak ≥ 0.5 sudah berada diluar *threshold* dan dinyatakan identitas tidak sama. Dapat dipastikan, hasil yang dikembalikan oleh sistem menunjukkan bahwa identitas data uji tidak ada yang sama dengan gambar yang terdapat pada dataset.

Tabel 1 menunjukkan hasil ringkasan pengujian sistem setelah dilakukan 100 pengujian untuk subjek gambar yang terdiri dari gambar dengan identitas sama dan tidak sama yang sudah dibagi rata:

Tabel. 1 Ringkasan pengujian sistem

Metrik Evaluasi	Nilai Hasil
Akurasi Pengenalan Wajah	94%
<i>Precision</i>	100%
<i>Recall</i>	86.96%
Waktu Rata-rata Pencarian per Gambar	2.3 Detik
Jarak Rata-rata <i>Embedding</i> (sama identitas)	0.31
Jarak Rata-rata <i>Embedding</i> (berbeda identitas)	0.62

Hasil tersebut menunjukkan bahwa jarak rata-rata antar wajah dengan identitas sama relatif kecil (< 0.5), sementara jarak antar wajah dengan identitas berbeda cukup besar (> 1.0). Hal ini membuktikan bahwa fungsi Triplet Loss berhasil memisahkan embedding wajah secara signifikan di ruang vektor.

C. Analisis Performa Sistem

Hasil eksperimen memperlihatkan bahwa FaceNet mampu mengekstraksi fitur wajah secara konsisten meskipun terdapat variasi kondisi pencahayaan dan ekspresi. Namun, sistem mengalami sedikit penurunan akurasi ketika wajah tertutup sebagian (misalnya oleh masker atau kacamata tebal).

Dari sisi efisiensi, proses pencarian berbasis *embedding vector* jauh lebih cepat dibandingkan metode konvensional yang membandingkan citra mentah. Hal ini disebabkan karena operasi pencarian dilakukan pada ruang numerik berdimensi tetap (128 dimensi) sehingga perhitungan jarak lebih ringan secara komputasi.

Penggunaan pre-trained model FaceNet juga mempercepat proses pengembangan karena model telah dilatih dengan dataset besar seperti *Labeled Faces in the Wild* (LFW). *Fine-tuning* lokal kemudian membantu sistem menyesuaikan karakteristik wajah spesifik dari data uji penelitian ini. Beberapa studi terbaru menunjukkan bahwa penggunaan *ArcFace loss* [13] dapat lebih meningkatkan pemisahan antar identitas pada embedding wajah dibandingkan *Triplet Loss* tradisional.

D. Evaluasi Kelemahan Sistem

Meskipun hasil pengujian menunjukkan performa yang baik, terdapat beberapa keterbatasan yang perlu dicatat:

- 1) Kualitas Gambar Input sangat memengaruhi hasil pencarian. Gambar dengan resolusi rendah atau pencahayaan ekstrem menurunkan akurasi sistem.
- 2) Terkadang sistem tidak dapat mengenali wajah walaupun wajah sudah terpampang dengan cukup jelas, tetapi hal ini sangat jarang terjadi.

E. Pembahasan Umum

Secara keseluruhan, implementasi FaceNet berbasis *Deep CNN* menunjukkan kinerja yang unggul dalam mengenali kemiripan wajah dan dapat digunakan sebagai sistem dasar untuk reverse image search berbasis wajah. Integrasi *embedding* FaceNet ke dalam sistem RIS memungkinkan pencarian yang lebih akurat dan efisien dibandingkan pendekatan berbasis pencocokan piksel.

Dengan pengembangan lebih lanjut, sistem ini berpotensi diadaptasi untuk berbagai aplikasi seperti verifikasi identitas digital, pelacakan penyalahgunaan gambar, serta pengenalan wajah pada sistem keamanan siber dan forensik digital.

V. KESIMPULAN

Penelitian ini berhasil mengimplementasikan metode *Deep Convolutional Neural Network* (CNN) model FaceNet sebagai sistem pencarian wajah pada *Reverse Image Search engine*. Melalui proses ekstraksi fitur menggunakan *embedding*

berdimensi 128 dan fungsi kerugian *Triplet Loss*, sistem mampu mengenali kemiripan wajah dengan akurasi yang tinggi.

Hasil pengujian menunjukkan bahwa integrasi FaceNet ke dalam sistem *Reverse Image Search* mampu mencapai akurasi pengenalan wajah sebesar 94%, dengan waktu pencarian rata-rata hanya 2,3 detik per gambar. Nilai jarak rata-rata antar *embedding* wajah dari individu yang sama (0,31) lebih kecil dibandingkan individu berbeda (0,62), yang menandakan keberhasilan sistem dalam membedakan identitas secara presisi.

Secara keseluruhan, sistem ini menunjukkan potensi besar dalam meningkatkan efektivitas pencarian berbasis wajah dan dapat menjadi fondasi bagi pengembangan teknologi pelacakan identitas visual pada berbagai aplikasi, termasuk keamanan digital, forensik, dan perlindungan privasi daring.

UCAPAN TERIMA KASIH

Terima kasih saya ucapkan untuk orang tua saya dan dosen pembimbing saya karena sudah membantu dalam proses penulisan jurnal ini baik secara langsung maupun tidak langsung.

REFERENSI

- [1] Y. LeCun, Y. Bengio, dan G. Hinton, "Deep Learning," *Nature*, vol. 521, no. 7553, hal. 436–444, 2015.
- [2] A. Krizhevsky, I. Sutskever, dan G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems*, vol. 25, hal. 1097–1105, 2012.
- [3] F. Schroff, D. Kalenichenko, dan J. Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2015, hal. 815–823.
- [4] M. D. Zeiler dan R. Fergus, "Visualizing and Understanding Convolutional Networks," in *Proc. European Conf. Computer Vision (ECCV)*, 2014, hal. 818–833.
- [5] G. B. Huang, M. Ramesh, T. Berg, dan E. Learned-Miller, "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments," *University of Massachusetts, Amherst*, 2007.
- [6] D. Yi, Z. Lei, S. Liao, dan S. Z. Li, "Learning Face Representation from Scratch," *arXiv preprint arXiv:1411.7923*, 2014.
- [7] K. Zhang, Z. Zhang, Z. Li, dan Y. Qiao, "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, hal. 1499–1503, 2016.
- [8] S. Chopra, R. Hadsell, dan Y. LeCun, "Learning a Similarity Metric Discriminatively, with Application to Face Verification," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2005.
- [9] C. Szegedy, W. Liu, Y. Jia, et al., "Going Deeper with Convolutions," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2015.
- [10] A. Vaswani, N. Shazeer, N. Parmar, et al., "Attention Is All You Need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [11] T. Ahonen, A. Hadid, dan M. Pietikäinen, "Face Description with Local Binary Patterns: Application to Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, hal. 2037–2041, 2006.
- [12] X. Wu, R. He, Z. Sun, dan T. Tan, "A Light CNN for Deep Face Representation with Noisy Labels," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 11, hal. 2884–2896, 2018.

- [13] J. Deng, J. Guo, N. Xue, dan S. Zafeiriou, "ArcFace: Additive Angular Margin Loss for Deep Face Recognition," Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2019.
- [14] S. Ge, J. Zhao, C. Li, dan J. Li, "Low-Resolution Face Recognition in the Wild via Selective Knowledge Distillation," IEEE Transactions on Image Processing, vol. 28, no. 4, hal. 2051–2062, 2019.
- [15] J. Philbin, O. Chum, M. Isard, J. Sivic, dan A. Zisserman, "Object Retrieval with Large Vocabularies and Fast Spatial Matching," Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2007.