



One-Shot Learning Menggunakan Siamese Neural Network Untuk Pendeteksian Wajah

Muhammad Dzealfath¹, Ardy Octavian Winarta², Prizka Rismawati Arum³

^{1,2,3} Program Studi Statistika, Universitas Muhammadiyah Semarang

¹ dzealfath.aptx@gmail.com

² ardyocta01@gmail.com

³ prizka.rismawatiarum@unimus.ac.id

Corresponding author: prizka.rismawatiarum@unimus.ac.id

Abstract: *One-Shot Learning using Siamese Neural Network is a method used for facial recognition, this method implements a deep learning algorithm which allows facial detection using one single image. This research aims to extract features from the images and calculate the dissimilarity of each compared images. This research uses Convolutional Neural Network (CNN) to perform feature extraction, where the images will be divided into two pairs and fed into two different CNN networks, then calculate the Euclidian Distance to measure similarity, calculate Contrastive Loss, and train Siamese Network to produce accurate results. The results of this research show that the algorithm can produce accurate results with high mean dissimilarity value for non-similar images and low mean dissimilarity value for similar images.*

Keywords: *Convolutional Neural Network (CNN), Face Detection, One-Shot Learning, Siamese Network*

Abstrak: *One-Shot Learning menggunakan Siamese Neural Network adalah metode yang digunakan untuk identifikasi wajah, metode ini mengimplementasikan algoritma deep learning yang memungkinkan pendeteksian wajah dengan menggunakan satu keping gambar. Penelitian ini bertujuan untuk mengekstrak fitur dari gambar dan menghitung ketidakmiripan dari gambar yang saling dibandingkan. Penelitian ini menggunakan Convolutional Neural Network (CNN) untuk melakukan ekstraksi fitur, dimana gambar akan dibagi menjadi dua pasangan dan dimasukkan kedalam dua jaringan CNN yang berbeda, kemudian menghitung Euclidian Distance untuk mengukur kesamaan, menghitung Contrastive Loss, dan melatih Siamese Network untuk menghasilkan hasil yang akurat. Hasil dari penelitian ini menunjukkan bahwa algoritma tersebut dapat menghasilkan hasil yang akurat dengan rata rata nilai ketidakmiripan yang tinggi untuk gambar yang tidak mirip dan rata-rata nilai ketidakmiripan yang rendah untuk gambar yang mirip.*

Kata kunci: *Convolutional Neural Network (CNN), Deteksi Wajah, One-Shot Learning, Siamese Network.*

I. PENDAHULUAN

Masalah yang dihadapi dalam pengenalan wajah adalah metode konvensional yang membutuhkan data pelatihan yang besar dan model pelatihan yang memakan waktu. Selain itu, model yang dilatih menggunakan metode ini masih mengalami kesulitan dalam mengenali wajah baru atau asing. Ini karena model hanya dapat mengenali wajah yang sudah ada di data pelatihan dan tidak dapat mengenali wajah baru secara akurat. Oleh karena itu diperlukan metode baru dalam pengenalan wajah untuk mengatasi masalah tersebut [1][2]. Salah satu metode yang dapat digunakan adalah *one-shot learning* menggunakan *Siamese Neural Network*. Metode ini memanfaatkan arsitektur *Siamese Neural Network* yang memungkinkan model untuk membandingkan dua citra wajah secara langsung dan memberikan skor kemiripan di antara keduanya. Dalam pelatihan model, hanya diperlukan satu gambar untuk setiap wajah, sehingga meminimalkan jumlah data yang diperlukan. Dengan menggunakan metode ini, model dapat mengenali wajah baru dengan akurasi yang lebih tinggi dibandingkan dengan metode konvensional dalam pengenalan wajah [3].

Bentuk wajah memberikan informasi visual secara multi-dimensi, yang menyediakan berbagai informasi mengenai ciri-ciri seorang individu, seperti identitas, jenis kelamin, usia, ras, suasana hati, dan niat. Proses pengenalan wajah bisa dilakukan melalui gambar diam maupun urutan video yang berawal dari sebuah gambar diam. Wajah umumnya terdiri dari dua mata, mulut, hidung, dan berbagai fitur lain yang ditempatkan pada lokasi yang sama [4][5].

Disisi lain pembelajaran mesin juga telah berhasil digunakan untuk mencapai kinerja canggih dalam berbagai implementasi seperti pencarian web, deteksi spam, pembuatan teks, serta pengenalan ucapan dan gambar [6][7][8]. Namun, algoritma ini sering tidak bekerja dengan maksimal ketika



dipaksa untuk membuat prediksi tentang data yang hanya memiliki sedikit informasi yang dipelajari. Tujuan kami adalah ingin menggeneralisasi model ke kategori yang tidak biasa ini tanpa memerlukan pelatihan ulang ekstensif yang mungkin mahal atau tidak mungkin karena data yang terbatas atau dalam data yang pengambilannya dibatasi seperti pengambilan web. Salah satu tugas yang sangat menarik adalah klasifikasi ini memiliki batasan bahwa kita hanya boleh mengamati satu contoh dari setiap kelas sebelum membuat prediksi [9]. Teknik ini disebut *one-shot learning* dan merupakan fokus utama dari model kami yang disajikan dalam karya ini. Ini harus dibedakan dari pembelajaran *zero-shot*, di mana model tidak dapat melihat contoh apa pun dari kelas target [10].

Pengenalan wajah merupakan salah satu aplikasi penting dalam teknologi kecerdasan buatan yang dapat digunakan dalam berbagai bidang, seperti keamanan, manajemen sumber daya manusia, dan pengenalan identitas [11]. Oleh karena itu, penting untuk mengatasi permasalahan dalam pengenalan wajah, seperti membutuhkan data pelatihan yang besar dan model pelatihan yang memakan waktu serta kesulitan dalam mengenali wajah baru [12]. Dengan mengatasi masalah ini, kami dapat mengembangkan model yang lebih akurat dan efisien dalam mengenali wajah, yang dapat membantu meningkatkan keamanan dan efisiensi di berbagai aplikasi. Selain itu, pengenalan wajah juga dapat membantu pengembangan teknologi canggih dan kecerdasan buatan, seperti mobil otomatis, server robotik, dan sistem pengenalan emosi manusia. Secara keseluruhan, mengatasi masalah pengenalan wajah sangat penting untuk membuka peluang baru dalam aplikasi kecerdasan buatan dan dapat membantu meningkatkan keamanan dan efisiensi di berbagai bidang. Penelitian ini bertujuan untuk mengusulkan metode baru dalam pengenalan wajah menggunakan *one-shot learning* dengan arsitektur *Siamese Neural Network* untuk mengatasi keterbatasan metode konvensional.

II. METODE PENELITIAN

Secara keseluruhan, penelitian tentang algoritme *one-shot learning* masih belum matang dan mendapat perhatian terbatas dari komunitas pembelajaran mesin. Namun demikian ada beberapa garis kunci penelitian yang mendahului makalah ini. Dalam [13], Li Fei-Fei et al. mengembangkan 1 kerangka kerja Bayesian variasional untuk klasifikasi gambar *one-shot learning* menggunakan premis bahwa kelas yang dipelajari sebelumnya dapat dimanfaatkan untuk membantu memperkirakan kelas yang akan datang ketika sangat sedikit data yang tersedia dari kelas tertentu. Informasi ini dapat berhasil dimasukkan ke dalam prior, yang diperbarui ketika data dari kelas baru diamati, dan kemudian digabungkan dengan kemungkinan untuk menghasilkan distribusi posterior kedalam kelas yang baru. Kemudian dalam [14], Xieyuanli, Chen et al. meneliti tentang bagaimana cara mengatasi masalah *overloop* dalam pendeteksian menggunakan lidar dengan mengembangkan model baru dari *OverlapNet* yang dimasukkan kedalam algoritma *monte-carlo* yang berdiri diatas jaringan *siamese network* dan menghasilkan model yang dapat mendeteksi secara dini *overloop* dan model dapat menggeneralisasi secara baik terhadap data lain yang diberikan. Sedangkan pada penelitian yang dilakukan oleh Hindy, Hanan et al. [15], menggunakan *siamese network* untuk mendeteksi pola baru dari serangan siber yang dimana peneliti menggunakan hanya satu model pelatihan pola serangan siber dan model tersebut dapat mendeteksi 3 pola serangan siber lain yang belum pernah dilihat sebelumnya.

Prosedur penelitian yang akan digunakan adalah mengumpulkan data latih dan uji dari subjek, kemudian menguji hipotesis dengan mengimplementasikan *One-Shot Learning* menggunakan *Siamese Neural Network* untuk pendeteksian wajah, dan akhirnya mengevaluasi akurasi dari pendeteksian wajah tersebut. Seperti yang diketahui bahwa penerapan dari *siamese network* adalah dengan memberikan gambar berpasangan guna membandingkan antara dua gambar dan dari gambar tersebut harus memiliki label seperti “sama” dan “tidak sama”. Setelah memiliki data berpasangan tersebut akan dibuat peta penyematan pada jaringan (*embedding mapping network*). Peta jaringan

yang telah tersedia akan dijadikan acuan untuk menentukan input jaringan pada *siamese network*. Hasil keluaran dari *siamese network* ini sendiri akan digunakan untuk menghitung jarak menggunakan *euclidian distance*, yang kemudian akan digunakan untuk menghitung *contrastive loss* yang diterapkan bersamaan dengan backpropagation neural network. Dan langkah terakhir adalah dengan melatih jaringan *contrastive loss* yang telah didapatkan.

III. HASIL DAN PEMBAHASAN

Dari keseluruhan data yang ada kami akan mengambil beberapa gambar yang dimana kami mengambil sample dari gambar yang sama dan gambar yang berbeda, kemudian kami akan memberikan label pada masing-masing gambar seperti pada contoh dibawah agar dapat diolah dan dimasukkan kedalam *Convolutional Neural Network* (CNN).

Input pair		Label
		SAMA
		BERBEDA
		SAMA
		BERBEDA

Gambar 1. Labeling data

Setelah kita mendapatkan data tersebut, kami akan membuat *siamese network* yang berdiri diatas CNN. Kami akan membagi gambar tersebut menjadi 2 pasangan, yang dimana masing-masing pasangan berisikan gambar dengan label yang sama dan berbeda. Dari pasangan gambar tersebut kami akan memasukan kedalam dua jaringan CNN yang berbeda dengan satu pasangan gambar ke dalam jaringan CNN A dan pasangan gambar lain kedalam jaringan CNN B. Perlu diingat bahwa peran CNN di sini hanya untuk mengekstrak fitur dan bukan untuk mengklasifikasikan. Seperti yang kita ketahui bahwa jaringan ini harus memiliki bobot dan arsitektur yang sama, jika Jaringan A kita adalah CNN dengan tiga layer maka Jaringan B kita juga harus CNN dengan tiga layer dan kita harus menggunakan set bobot yang sama untuk kedua jaringan ini. Jadi Jaringan A dan Jaringan B akan memberi kita penyematan untuk gambar masukan dan masing-masing. Untuk melakukan proses ekstrasi fitur ini yang dimana kami akan menggunakan dua lapisan konvolusi dengan aktivasi relu dan max pooling untuk melakukan proses ekstrasi fitur dan diakhiri dengan flat layer. Setelah mendapatkan hasil dari ekstrasi fitur yang dilakukan pada CNN, kami mendapatkan total 20.000 titik data dengan rincian 10.000 titik data untuk gambar yang sama dan 10.000 titik data untuk gambar yang berbeda. Dari data tersebut kemudian kita akan membagi data tersebut dengan proporsi 75:15 untuk masing masing data latih dan data uji, atau sekitar 15.000 untuk data latih dan 5.000 untuk data uji. Dan masing-masing akan dimasukkan kedalam jaringan A dan jaringan B sehingga menghasilkan vektor fitur yang baru. Setelah kami mendapatkan vektor fitur dari gambar tadi, kami akan menggunakan kedua vektor fitur tadi untuk mengukur kesamaan menggunakan *euclidian distance*

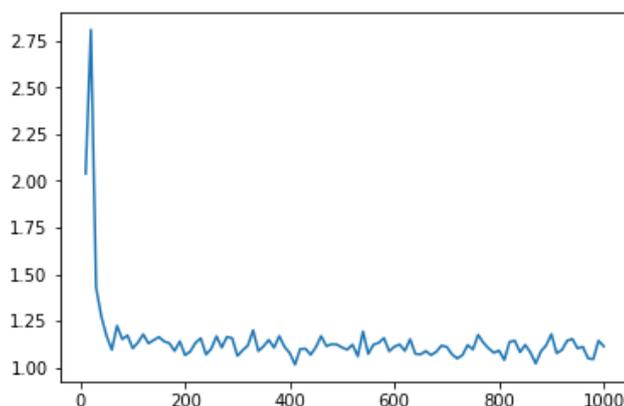
sebagai tolok ukur yang didapatkan dari jaringan A dan jaringan B yang akan digunakan sebagai masukan pada tahap *contrastive loss*, *euclidian distance* sendiri dapat dihitung dengan persamaan:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Pada tahap *contrastive loss*, loss akan dihitung dan algoritma backpropagation akan dijalankan. *Contrastive loss* sendiri dapat dihitung dengan rumus:

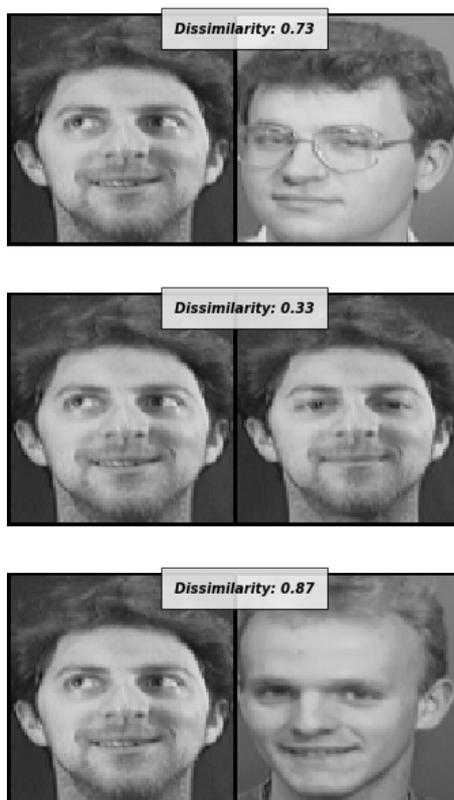
$$L_{\text{contrastive}}(x_0, x_1, y) = \frac{1}{2} y \|f(x_0) - f(x_1)\|_2^2 + \frac{1}{2} (1 - y) \{\max(0, m - \|f(x_0) - f(x_1)\|_2)\}^2 \quad (2)$$

Setelah kami mendapatkan hasil dari *contrastive loss*, kami melakukan proses pelatihan menggunakan model *siamese network*. Kami menggunakan *optimizer* adam dengan *learning rate* sebesar 0.0005, dengan total 100 epoch, dengan hasil dari training kami menghasilkan loss pada setiap titik data menurun seperti yang ditampilkan pada gambar dibawah



Gambar 2. Loss dari pelatihan

Dengan rata-rata loss yang dihasilkan adalah 1.00, kemudian kami mencoba menerapkan langsung pada data gambar yang ada yang kami pilih secara acak dan mencoba membandingkan antara kemudian kami dua menghitung gambar nilai ketidakmiripan dari dua gambar yang dibandingkan tersebut. Dari hasil tersebut kami mendapatkan hasil dengan rata-rata hasil yang akurat seperti sebagai berikut:



Gambar 3. Hasil akurasi dari *siamese network*

Angka yang berada diatas gambar menunjukkan nilai ketidakmiripan, yang berarti semakin tinggi angka tersebut maka gambar semakin tidak mirip dan semakin rendah angka tersebut maka gambar semakin mirip. Dari 3 gambar diatas hasil yang didapatkan sangat baik, dengan gambar pertama menunjukkan bahwa ada ketidak miripan yang tinggi, kemudian gambar kedua menunjukan bahwa gambar tersebut memiliki ketidakmiripan yang rendah atau dapat dikatakan gambar tersebut mirip, dan yang terakhir angka menunjukkan bahwa gambar memiliki nilai ketidak miripan yang tinggi. Hasil penelitian ini juga menunjukkan bahwa *siamese network* mencapai akurasi 98,5% pada dataset AT&T. Hasil ini mengungguli metode konvensional seperti Principal Component Analysis (PCA) dan Linear Discriminant Analysis (LDA), yang masing-masing mencapai akurasi 85,25% dan 92,75% [16]. Akurasi yang tinggi dari *siamese network* dalam pengenalan wajah menunjukkan keefektifan arsitektur jaringan dalam mempelajari representasi fitur yang menangkap kesamaan antar wajah.

IV. KESIMPULAN

Algoritma ini mampu mendeteksi perbedaan antara dua citra yang dibandingkan dengan baik walaupun dengan data pelatihan yang kecil. Kinerja model diukur dengan nilai *contrastive loss* dan mencapai akurasi yang baik. Namun, terdapat beberapa keterbatasan dalam pengembangan dan penerapan model dalam kehidupan sehari-hari, seperti model membutuhkan pembelajaran berulang setiap menerima data baru dan model membutuhkan gambar yang jelas untuk mengurangi resiko gambar tidak terdeteksi. Sebagai kesimpulan, penelitian kami menunjukkan keefektifan jaringan Siam dalam pengenalan wajah dan keunggulannya dibandingkan metode konvensional. Hasilnya menunjukkan bahwa *Siamese Network* dapat diterapkan dalam berbagai tugas pengenalan wajah, termasuk sistem keamanan, manajemen sumber daya manusia, dan pengenalan identitas. Penelitian di masa depan dapat mengeksplorasi penggunaan arsitektur jaringan yang lebih kompleks dan kumpulan data yang lebih besar untuk lebih meningkatkan kinerja *Siamese Network* dalam pengenalan wajah

**REFERENSI**

1. Z. Ding, Y. Guo, L. Zhang, and Y. Fu, “Generative One-Shot Face Recognition,” Sep. 2019, [Online]. Available: <http://arxiv.org/abs/1910.04860>
2. S. Omatu, A. Chekima, J. Dargham, and E. Moug, “Face Recognition State-of-the-art, Enablers, Challenges and Solutions: A Review,” *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, pp. 96–105, Apr. 2020, doi: 10.30534/ijatcse/2020/1691.22020.
3. G. R. Koch, “Siamese Neural Networks for One-Shot Image Recognition,” 2015.
4. S. Z. Li and A. K. Jain, “Handbook of Face Recognition (The Second Edition) Springer-Verlag,” 2011.
5. A. R. Dewi and I. Septadepi, “Face Recognition System Using Eigenface Method based on Facial Component Region,” 2012.
6. Y. Zhang, “Speech Recognition Using Deep Learning Algorithms,” 2013.
7. K. Kowsari *et al.*, “Text Classification Algorithms: A Survey,” *Information (Switzerland)*, vol. 10, Apr. 2019, doi: 10.3390/info10040150.
8. S. Annareddy and S. Tammina, “A Comparative Study of Deep Learning Methods for Spam Detection,” in *2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, 2019, pp. 66–72. doi: 10.1109/I-SMAC47947.2019.9032627.
9. B. M. Lake, R. Salakhutdinov, J. Gross, and J. B. Tenenbaum, “One shot learning of simple visual concepts,” 2011. [Online]. Available: <http://web.mit.edu/brenden/www/charactervideos.html>.
10. W. Wang *et al.*, “Pyramid Vision Transformer: A Versatile Backbone for Dense Prediction without Convolutions,” Feb. 2021, [Online]. Available: <http://arxiv.org/abs/2102.12122>
11. C. Mingsung and L. Cai, *Research on the Application of Face Recognition System*. 2020. doi: 10.2991/assehr.k.200727.057.
12. R. Jafri and H. Arabnia, “A Survey of Face Recognition Techniques,” *JIPS*, vol. 5, pp. 41–68, Jun. 2009, doi: 10.3745/JIPS.2009.5.2.041.
13. L. Fei-Fei, R. Fergus, and P. Perona, “A Bayesian Approach to Unsupervised One-Shot Learning of Object Categories,” 2006. [Online]. Available: www.google.com
14. X. Chen, T. Labe, A. Milioto, T. Röhling, J. Behley, and C. Stachniss, “OverlapNet: a siamese network for computing LiDAR scan similarity with applications to loop closing and localization,” *Auton Robots*, vol. 46, no. 1, pp. 61–81, Jan. 2022, doi: 10.1007/s10514-021-09999-0.
15. H. Hindy *et al.*, “Leveraging siamese networks for one-shot intrusion detection model,” *J Intell Inf Syst*, 2022, doi: 10.1007/s10844-022-00747-z.
16. A. Sahoolzadeh, B. Heidari, and C. Dehghani, “A new face recognition method using PCA, LDA and neural network,” *International Journal of Computer Science and Engineering*, vol. 2, Jan. 2008.