

Pengenalan Sistem Isyarat Bahasa Indonesia (SIBI) Menggunakan Gradient-Convolutional Neural Network

DARMATASIA

Teknik Informatika Fakultas Sains dan Teknologi
Universitas Islam Negeri Alauddin Makassar
E-mail: darmatasia@uin-alauddin.ac.id

ABSTRAK

Penelitian ini bertujuan untuk melakukan pengenalan alfabet pada Sistem Isyarat Bahasa Indonesia (SIBI). Penelitian ini memiliki dua kontribusi utama, Pertama dilakukan pengumpulan dataset alfabet SIBI. Kedua, pengenalan alfabet SIBI menggunakan algoritma *Convolutional Neural Network* (CNN). Pada penelitian ini, citra masukan berupa alfabet bahasa isyarat pada lapisan *input* diberikan filter *gradient* agar bentuk objek menjadi lebih jelas. Hasil penelitian menunjukkan bahwa pemberian filter pada lapisan *input* dapat meningkatkan akurasi pengenalan yaitu sekitar 85%. Citra masukan yang tidak difilter hanya memperoleh akurasi sebesar 25%. Akurasi terbaik yang diperoleh yaitu 98% dengan meningkatkan jumlah iterasi. Metode yang diusulkan juga diuji menggunakan tiga *benchmark* dataset. Hasil pengujian menunjukkan bahwa metode yang diusulkan dapat meningkatkan akurasi pengenalan pada *benchmark* dataset yang memiliki *background* yang kompleks.

Kata Kunci: *Convolutional Neural Network*; *Gradient*; Sistem Isyarat Bahasa Indonesia

I. PENDAHULUAN

Bahasa Isyarat biasanya digunakan oleh penyandang tunarungu untuk berkomunikasi dengan orang lain. Bahasa isyarat merupakan gerak yang menggunakan tubuh atau fisik untuk menyampaikan informasi (Panwar, 2012).

Permasalahan yang sering terjadi dalam komunikasi menggunakan bahasa isyarat adalah ketika orang biasa berkomunikasi dengan penyandang tunarungu. Orang biasa seringkali sulit memahami bahasa isyarat yang digunakan oleh penyandang tunarungu. Masalah tersebut dapat mempengaruhi kehidupan dan hubungan interpersonal antara penyandang tunarungu dan masyarakat biasa. (R Hartanto, 2014).

Seiring dengan perkembangan teknologi, penggunaan aplikasi atau sistem penerjemah bahasa isyarat telah banyak dikembangkan. Hal tersebut dapat memudahkan komunikasi antara penyandang tunarungu dan orang biasa yang sulit memahami bahasa isyarat (R Hartanto, 2014).

Dalam penelitian (Rudy Hartanto et al., 2013) menggunakan algoritma SIFT untuk mengenali gestur tangan dalam Sistem Isyarat Bahasa Indonesia. Penelitian tersebut bertujuan untuk menerjemahkan gestur tangan ke dalam huruf alfabet. Pada penelitian selanjutnya, (R Hartanto, 2014) mengkombinasikan deteksi kulit dan ruang warna untuk segmentasi citra masukan, kemudian algoritma SURF digunakan untuk mengenali setiap gestur tangan dalam bahasa isyarat. Prototipe yang dikembangkan digunakan untuk mengenali gestur tangan dalam SIBI secara *real time*.

Banyaknya variasi gestur dalam bahasa isyarat menyebabkan sulitnya dilakukan pengenalan bahasa isyarat dengan tingkat akurasi yang tinggi. Berbagai pendekatan telah dilakukan oleh para peneliti, salah satunya yaitu pendekatan *deep learning*. Pendekatan *deep learning* mampu melakukan ekstraksi fitur secara otomatis pada sebuah citra masukan sehingga tidak perlu dilakukan ekstraksi fitur secara manual.

Dalam penelitian ini akan dilakukan pengenalan alfabet SIBI dengan menggunakan algoritma CNN. CNN merupakan salah satu arsitektur *deep learning*. CNN dapat digunakan untuk citra masukan yang memiliki *background* yang rumit dan variasi postur gestur tangan yang beragam. Arsitektur CNN dengan pendekatan *bottom-up* diusulkan oleh (Yamashita & Watasue, 2014) untuk pengenalan bahasa isyarat. Pendekatan tersebut terdiri dari dua arsitektur utama yaitu pertama citra masukan diekstraksi dari *background* yang kompleks kemudian selanjutnya dilakukan binerisasi citra sebagai masukan.

(Jie Huang et al., 2015) mengusulkan 3D-CNN yang terdiri dari lima tipe data masukan yaitu citra warna (R, G, dan B), *depth*, dan *skeleton* untuk pengenalan bahasa isyarat Microsoft Kinect. Hasil penelitian menunjukkan bahwa citra masukan *multi-channels* memberikan akurasi pengenalan yang lebih tinggi dibandingkan hanya menggunakan citra masukan berskala keabuan (*grayscale*).

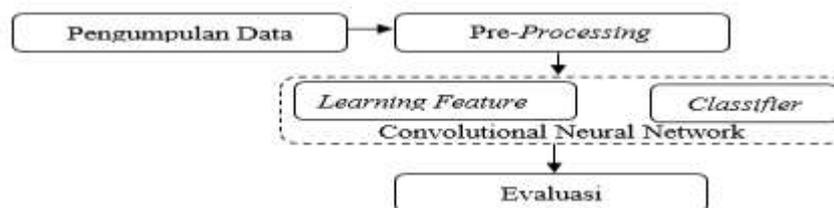
Modifikasi algoritma CNN terus dikembangkan untuk meningkatkan hasil pengenalan. Modifikasi dapat dilakukan pada arsitektur CNN, pemberian filter pada data masukan, kombinasi dengan metode ekstraksi fitur serta penggabungan dengan metode klasifikasi lain.

Modifikasi lapisan *input* pada CNN telah dilakukan dalam penelitian (Santoni et al., 2015) yaitu citra masukan diekstraksi menggunakan GLCM. Penelitian tersebut menunjukkan adanya peningkatan akurasi dibandingkan dengan menggunakan *input* data asli.

Umumnya citra masukan pada lapisan *input* CNN menggunakan citra skala keabuan (*grayscale*). Pada penelitian ini dilakukan modifikasi pada lapisan *input* CNN. Citra masukan berupa alfabet SIBI difilter terlebih dahulu. Hal tersebut bertujuan untuk mengatasi adanya perubahan cahaya pada citra masukan yang dapat mempengaruhi hasil pengenalan (Molchanov et al., 2015). Filter yang diberikan pada citra masukan adalah filter *gradient* yang dapat membuat tepi objek menjadi lebih tajam sehingga bentuk objek menjadi lebih jelas.

II. METODE PENELITIAN

Penelitian ini terdiri dari 4 tahapan yang dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

A. Pengumpulan Data

Pada penelitian ini dilakukan pengumpulan data dari video alfabet SIBI yang diperoleh melalui berbagai sumber kemudian dikonversi menjadi gambar. Hasil *capture* video kemudian di-*crop* pada bagian tangan untuk mengurangi dimensi gambar. Dataset yang dibuat terdiri dari 5 alfabet saja yaitu (A, B, C, D, dan E) dengan total 600 gambar.



Gambar 2. (a) Hasil *Capture Video* Alfabet SIBI; (b) Hasil *Crop Area Tangan*



Gambar 3. Alfabet SIBI (A-E)

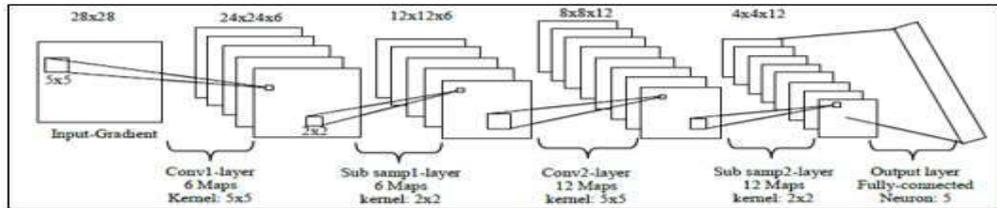
Gambar 2(a) adalah contoh hasil *capture* alfabet SIBI sedangkan Gambar 2(b) merupakan hasil perolehan pada area tangan. Gambar 3 merupakan contoh dataset yang telah dikumpulkan yang terdiri dari alfabet A, B, C, D, dan E dari kiri ke kanan.

B. Metode yang Diusulkan

Metode yang digunakan dalam penelitian ini adalah CNN. CNN pertama kali diperkenalkan oleh Yann Lecun pada tahun 1998. Secara umum, CNN memiliki 3 prinsip dasar yaitu *local receptive fields*, *shared weight*, dan *pooling* (Santoni et al., 2015). *Receptive fields* merupakan bagian yang ditentukan dari ruang spasial yang berisi unit yang memberi masukan ke sekumpulan unit pada lapisan yang sesuai. Setiap *local receptive fields* pada lapisan *input* akan terkoneksi dengan satu neuron pada lapisan tersembunyi (*hidden layer*). Setiap koneksi memiliki bobot dan bias yang akan dibagi (*shared weight*) pada *local receptive fields* yang sama pada lapisan yang sama. Ukuran *receptive fields* ditentukan oleh ukuran filter pada lapisan konvolusi.

Arsitektur CNN terdiri dari lapisan konvolusi, lapisan *sub-sampling*, dan *fully connected layer*. Pada layer konvolusi dapat diterapkan banyak filter konvolusi dimana setiap filter akan mengekstraksi satu fitur yang berbeda. Semakin banyak filter yang digunakan maka akan semakin banyak fitur yang dapat diperoleh. Lapisan *sub-sampling* atau dikenal juga dengan *pooling* diterapkan pada keluaran dari lapisan konvolusi. Hal tersebut bertujuan untuk menyederhanakan informasi keluaran pada lapisan konvolusi (Michael Nielsen, 2015). Lapisan terakhir dari CNN adalah *fully connected layer* seperti pada MLP yang berfungsi sebagai *classifier*.

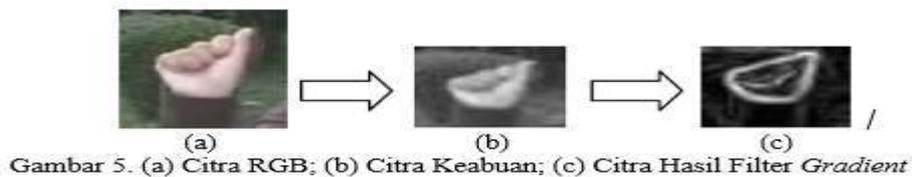
Arsitektur CNN yang diusulkan dalam penelitian ini dapat dilihat pada Gambar 4. Jaringan CNN terdiri dari 5 lapisan. 4 lapisan pertama merupakan dua pasang lapisan konvolusi dan *sub-sampling*. Lapisan kelima merupakan *fully connected layer* yang berfungsi sebagai *classifier*.



Gambar. 4 Arsitektur CNN

Input gambar berupa alfabet SIBI memiliki dimensi 28x28. Keluaran pada lapisan konvolusi memiliki 6 *feature map* dengan dimensi 24x24. Filter konvolusi yang digunakan berukuran 5x5 dengan nilai yang di-*generate* secara acak. Masing-masing lapisan konvolusi diikuti oleh lapisan *sub-sampling*. Keluaran dari lapisan *sub-sampling* pertama memiliki dimensi 12x12 dan pada lapisan *sub-sampling* kedua adalah 4x4. Fungsi aktivasi yang digunakan adalah sigmoid.

Citra masukan pada lapisan *input* CNN dilakukan *pre-processing* terlebih dahulu dengan menggunakan filter *gradient*. Filter *gradient* memberikan besarnya gradien data yang diperoleh dengan menghitung gradien arah (Gx) dan (Gy). Sumbu horizontal (x) menunjukkan arah peningkatan subskrip kolom. Sumbu vertikal (y) merupakan arah peningkatan subskrip baris. Besar dan arah gradien kemudian dihitung dari komponen ortogonal dari (Gx) dan (Gy).



Gambar 5. (a) Citra RGB; (b) Citra Keabuan; (c) Citra Hasil Filter Gradient

Berdasarkan Gambar 5, dapat dilihat bahwa pada citra yang telah difilter dengan menggunakan filter *gradient* memiliki *edge* yang lebih jelas pada bagian objek tangan dibandingkan dengan citra keabuan.

III. HASIL DAN PEMBAHASAN

Metode yang diusulkan diuji menggunakan dua skenario. Pertama, metode dievaluasi dengan menggunakan dua jenis citra masukan yang berbeda yaitu citra *grayscale* dan citra yang telah difilter menggunakan filter *gradient*. Hal ini bertujuan untuk melihat pengaruh penggunaan filter *gradient* terhadap hasil pengenalan alfabet SIBI. Akurasi pengenalan pada skenario pertama dapat dilihat pada Tabel. 1.

Tabel 1. Akurasi Pengenalan Alfabet SIBI

Jumlah Iterasi	Akurasi (%)	
	<i>Input-Grayscale</i>	<i>Input-Gradient</i>
50	25.00	85.00
100	81.00	97.50
200	97.00	98.00

Tabel 1 menunjukkan bahwa CNN yang menggunakan citra masukan *gradient* memberikan akurasi pengenalan yang lebih tinggi dibandingkan dengan citra *grayscale* meskipun dengan jumlah iterasi yang kecil. Akurasi pengenalan menggunakan metode yang diusulkan memperoleh tingkat akurasi yang cukup tinggi dengan hanya 50 iterasi saja sedangkan citra masukan *grayscale* membutuhkan jumlah iterasi dua kali lebih besar untuk memperoleh akurasi yang cukup baik.

Selain akurasi, metode yang diusulkan juga dievaluasi nilai presisi dan *recall*-nya. Hasil presisi dan *recall* dapat dilihat pada Tabel 2.

Tabel 2. Presisi dan *Recall* Pengenalan Alfabet SIBI Menggunakan CNN

Kelas	Presisi (%)		Recall (%)	
	<i>Input-Grayscale</i>	<i>Input-Gradient</i>	<i>Input-Grayscale</i>	<i>Input-Gradient</i>
A	96.00	100.00	96.00	98.00
B	95.24	95.00	100.00	95.00
C	94.87	94.87	100.00	100.00
D	100.00	100.00	94.87	100.00
E	100.00	100.00	94.12	97.06
Rata-rata	97.222	97.974	96.998	98.012

Tabel 2 menunjukkan bahwa rata-rata presisi dan recall citra masukan *gradient* memperoleh hasil yang lebih tinggi dibandingkan dengan citra masukan *grayscale*.

Skenario kedua pada penelitian ini dilakukan dengan mengevaluasi metode yang diusulkan menggunakan tiga *benchmark dataset* berbeda. Dataset yang

digunakan sebagai acuan yaitu *American Sign Language* (ASL) dengan pertimbangan ASL memiliki bentuk alfabet yang hampir sama dengan SIBI.

Dataset pertama diperoleh dari *Michal Gallery* (Kawulok et al., 2013) yang terdiri dari alfabet dan angka namun pada penelitian ini hanya digunakan 5 alfabet saja yaitu alfabet A, B, C, D, dan E. Contoh dataset Michal dapat dilihat pada Gambar 6.



Gambar 6. Dataset Michal Alfabet A, B, C, D, dan E (Kawulok et al., 2013)

Dataset kedua diperoleh dari (Massey, 2012) yang dibuat oleh Massey University dengan jumlah data sebanyak 2524 yang terdiri dari alfabet dan angka. Gambar alfabet telah disegmentasi dan memiliki latar belakang warna hitam. Dataset yang digunakan juga hanya alfabet yaitu A, B, C, D, dan E saja. Contoh dataset Masey dapat dilihat pada Gambar 7.



Gambar 7. Dataset Massey: Alfabet A, B, C, D, dan E (Massey, 2012)

Dataset ketiga diperoleh dari *Marcel Static Hand Posture Database* (Sebastian, 2012) yang terdiri dari 6 alfabet (A, B, C, *Point*, *Five*, dan V) yang diambil dari 10 orang berbeda.



Gambar 8. Dataset Marcel: A, B, C, *Point*, *Five*, dan V (Sebastian, 2012)

Metode yang diusulkan dievaluasi dengan menggunakan tiga *benchmark dataset* yang telah dijelaskan. Hasil evaluasi dapat dilihat pada Tabel 3.

Tabel 3. Akurasi Pengenalan pada *Benchmark Dataset* Menggunakan CNN

Dataset	Jumlah Iterasi	Akurasi (%)		Jumlah Data	Jumlah Kelas
		<i>Input-Grayscale</i>	<i>Input-Gradient</i>		
Michal	200	44.00	64.00	200	5
Massey-University	200	98.00	94.00	350	5
Marcel	200	83.75	90.25	1300	6

Berdasarkan Tabel 3, metode yang diusulkan memberikan akurasi yang lebih tinggi pada dataset Michal dan Marcel sedangkan pada dataset Massey, citra masukan berupa *grayscale* memberikan akurasi yang lebih tinggi. Dataset pada penelitian ini memiliki karakteristik yang sama dengan dataset Michal dan Marcel yaitu memiliki *background* sedangkan pada dataset Massey, objek telah disegmentasi.

Filter *gradient* tidak memberikan pengaruh yang signifikan pada objek yang telah disegmentasi sebab tujuan dari filter *gradient* hampir sama dengan segmentasi yaitu membuat tepi objek lebih tajam sehingga bentuk dari objek menjadi lebih jelas. Meskipun demikian, filter *gradient* bekerja dengan baik pada citra yang mengandung *background*.

Hasil penelitian ini juga dibandingkan dengan penelitian-penelitian sebelumnya yang relevan dengan pengenalan bahasa isyarat.

Tabel 4. Perbandingan Akurasi Pengenalan Dengan Penelitian Yang Relevan

Penulis	Dataset	Metode	Jumlah Kelas	Akurasi (%)
(Triesch & von der Malsburg, 2001)	Marcel (dibuat peneliti)	Constrained Generative Model (CGM)	6	93.20
(Lee & Park, 2008)	Tidak disebutkan	CNN-WFMM	6	80.00-97.50
(Nagi et al., 2011)	Jawad (dibuat peneliti)	Max-Pooling CNN (MPCNN)	6	96.00
(Rudy Hartanto et al., 2013)	ISL/SIBI (dibuat peneliti)	SIFT+k-means	24	62.60
(Jie Huang et al., 2015)	Microsoft Kinect	3D-CNN	25	94.20
Darmatasia	ISL/SIBI (dibuat peneliti)	Gradient-CNN	5	98.00

Berdasarkan Tabel 4, metode yang diusulkan memberikan akurasi yang cenderung lebih tinggi untuk pengenalan alfabet bahasa isyarat. Metode yang diusulkan juga memberikan akurasi yang cenderung lebih tinggi dibandingkan dengan penelitian lain yang juga menggunakan algoritma CNN.

IV. KESIMPULAN

Pada penelitian ini, metode CNN telah digunakan untuk mengenali alfabet pada Sistem Isyarat Bahasa Indonesia. Citra masukan diberikan filter *gradient* untuk

membuat objek yang mengandung *background* memiliki tepi yang lebih tajam sehingga bentuk dari objek tersebut menjadi lebih jelas. Berdasarkan hasil penelitian yang telah dilakukan menunjukkan bahwa metode yang diusulkan dapat meningkatkan akurasi pengenalan alfabet SIBI. Selain itu, untuk menguji performa metode yang diusulkan maka digunakan *benchmark* dataset dengan memberikan perlakuan yang sama dengan eksperimen yang telah dilakukan. Hasil penelitian menggunakan *benchmark* dataset menunjukkan bahwa metode yang diusulkan dapat meningkatkan akurasi pengenalan untuk citra masukan yang mengandung *background*.

Keuntungan yang dimiliki oleh metode yang diusulkan adalah ekstraksi fitur dan klasifikasi terintegrasi dalam satu kesatuan arsitektur sehingga tidak perlu dilakukan ekstraksi fitur secara terpisah. Selain itu, filter gradient dapat menghasilkan objek dengan bentuk yang lebih jelas sehingga tidak perlu dilakukan segmentasi citra.

Penelitian ini tentunya memiliki keterbatasan yaitu hanya mengenali lima alfabet SIBI saja. Penelitian selanjutnya, metode yang diusulkan dapat dievaluasi kembali dengan menggunakan 26 alfabet SIBI. Selain itu, pada penelitian ini dilakukan pengenalan alfabet SIBI secara statis sehingga pada penelitian selanjutnya, perlu dipertimbangkan pengenalan alfabet atau pun kosa kata SIBI secara *real time*.

DAFTAR PUSTAKA

- Hartanto, Rudy, Susanto, A., & Santosa, P. I. 2013. Preliminary design of static Indonesian sign language recognition system. *2013 International Conference on Information Technology and Electrical Engineering (ICITEE)*, 2, 187–192. <https://doi.org/10.1109/ICITEED.2013.6676236>
- Jie Huang, Wengang Zhou, Houqiang Li, & Weiping Li. 2015. Sign Language Recognition using 3D convolutional neural networks. *2015 IEEE International Conference on Multimedia and Expo (ICME)*, 1–6. <https://doi.org/10.1109/ICME.2015.7177428>
- Kawulok, M., Tomasz, G., Nalepa, J., & Knyc, M. 2013. *Database for hand gesture recognition*. <http://sun.aei.polsl.pl/~mkawulok/gestures/>
- Lee, J. S., & Park, J.-H. 2008. Dynamic hand gesture recognition using a CNN model with 3D receptive fields. *2008 International Conference on Neural Networks and Signal Processing*, 14–19. <https://doi.org/10.1109/ICNNSP.2008.4590300>

- Massey, U. 2012. *Gesture Dataset*. Massey University. http://www.massey.ac.nz/~albarcza/gesture_dataset2012.html
- Michael Nielsen. 2015. *Deep Learning*. <http://neuralnetworksanddeeplearning.com/chap6.html>
- Molchanov, P., Gupta, S., Kim, K., Kautz, J., & Clara, S. 2015. *Hand Gesture Recognition with 3D Convolutional Neural Networks*. 1–7.
- Nagi, J., Ducatelle, F., Caro, G. a. Di, Ciresan, D., Meier, U., Giusti, A., Nagi, F., Schmidhuber, J., & Gambardella, L. M. 2011. Max-pooling convolutional neural networks for vision-based hand gesture recognition. *2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, 342–347. <https://doi.org/10.1109/ICSIPA.2011.6144164>
- Panwar, M. 2012. Hand gesture recognition based on shape parameters. *2012 International Conference on Computing, Communication and Applications*, 1–6. <https://doi.org/10.1109/ICCCA.2012.6179213>
- Santoni, M. M., Sensuse, D. I., Arymurthy, A. M., & Fanany, M. I. 2015. Cattle Race Classification Using Gray Level Co-occurrence Matrix Convolutional Neural Networks. *Procedia Computer Science*, 59(October), 493–502. <https://doi.org/10.1016/j.procs.2015.07.525>
- Sebastian, M. 2012. *Hand Posture and Gesture Datasets*. <https://sites.google.com/site/autosignlan/source/image-data-set>
- Triesch, J., & von der Malsburg, C. 2001. A system for person-independent hand posture recognition against complex backgrounds. *Pattern Analysis and Machine Intelligence, IEEE Transactions On*, 23(12), 1449–1453. <https://doi.org/10.1109/34.977568>
- Yamashita, T., & Watasue, T. 2014. Hand posture recognition based on bottom-up structured deep convolutional neural network with curriculum learning. *Image Processing (ICIP), 2014 IEEE International Conference On*, 853–857. <https://doi.org/10.1109/ICIP.2014.7025171>