

Pengenalan Gestur Tangan Menggunakan Convolutional Neural Network Untuk Mengendalikan Karakter dalam Game Snake

James Jeremy Sijaya¹⁾ Yulia Ery Kurniawati²⁾

¹⁾ Informatika, Fakultas Industri Kreatif Institut Teknologi dan Bisnis Kalbis
Jalan Pulomas Selatan Kav. 22, Jakarta 13210
Email: jamesjeremysijaya@gmail.com

²⁾ Informatika, Fakultas Industri Kreatif Institut Teknologi dan Bisnis Kalbis
Jalan Pulomas Selatan Kav. 22, Jakarta 13210
Email: yulia.kurniawati@kalbis.ac.id

Abstract: *Human-Computer Interaction (HCI) is a field that includes interactions between humans and computers. One of the HCI interfaces that allow humans to interact with computers in a contactless manner is Hand Gesture Recognition (HGR). In this research, HGR will be applied as input to control the character in Snake game. The algorithm used for image classification is Convolutional Neural Network (CNN). Before, making a classification model, the image data first goes through the preprocessing stage. The preprocessing stage that is performed is setting ROI, grayscaling, smoothing, background subtraction, thresholding, resizing, and normalization. The image that has gone through the preprocessing stage is then used in making the CNN model. Based on the research results, the model gets an accuracy value of 98%, a precision value of 98.1%, and a recall value of 98%.*

Keywords: *classification, convolutional neural network, game, hand gesture recognition*

Abstrak: *Human-Computer Interaction (HCI) merupakan bidang yang mencakup interaksi antara manusia dengan komputer. Salah satu antarmuka HCI yang memungkinkan manusia berinteraksi dengan komputer secara nirsentuh adalah Hand Gesture Recognition (HGR). Dalam penelitian ini, HGR akan diaplikasikan sebagai inputan untuk mengendalikan karakter dalam game Snake. Algoritma yang digunakan untuk pengklasifikasian citra adalah Convolutional Neural Network (CNN). Sebelum, pembuatan model klasifikasi, data citra terlebih dahulu melalui tahap preprocessing. Tahap preprocessing yang dilakukan yakni menetapkan ROI, grayscaling, smoothing, background subtraction, thresholding, resizing, dan normalisasi. Citra yang telah melalui tahap preprocessing kemudian digunakan dalam pembuatan model CNN. Berdasarkan hasil penelitian, model mendapatkan nilai akurasi sebesar 98%, nilai presisi sebesar 98,1%, dan nilai recall sebesar 98%.*

Kata kunci: *convolutional neural network, game, klasifikasi, pengenalan gestur tangan*

I. PENDAHULUAN

Interaksi antara manusia dengan komputer atau *Human-Computer Interaction (HCI)* merupakan bidang yang mengkaji tentang interaksi antara manusia dan komputer. Penggunaan perangkat keras seperti *keyboard* dan *mouse* sering digunakan sebagai alat untuk interaksi antara manusia dan komputer. Interaksi dengan menggunakan perangkat keras

tersebut memerlukan kontak langsung antara manusia dan komputer.

Seiring perkembangan teknologi, *computer vision* memungkinkan interaksi antara manusia dan komputer secara nirsentuh. Metode tersebut tidak membutuhkan kontak langsung antara pengguna dengan peralatan input, melainkan komputer menangkap gerakan pengguna melalui kamera dan menginterpretasikan gerakan tersebut [1]. Salah satu pengaplikasian

interaksi manusia dan komputer secara nirsentuh yakni dengan menggunakan *Hand Gesture Recognition* (HGR).

Pengenalan gestur tangan atau yang dikenal sebagai *Hand Gesture Recognition* (HGR) telah banyak diaplikasikan dalam bidang yang berbeda seperti *game*, *virtual reality*, dan *Sign Language Recognition* (SLR) [2]. Metode pengklasifikasian gestur tangan dapat diterapkan dengan menggunakan algoritma *Machine Learning*.

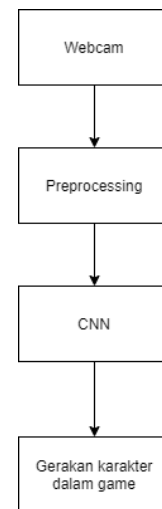
Penelitian ini akan mengklasifikasikan citra gestur tangan yang digunakan sebagai masukan untuk mengendalikan karakter dalam *game* Snake. Alasan *game* Snake dipilih yakni karena umumnya *game* Snake dimainkan dengan menggunakan *arrow key* pada *keyboard* sebagai penunjuk arah untuk menggerakkan karakter. Sedangkan gestur tangan juga dapat digunakan sebagai penunjuk arah yang dapat digunakan untuk menggerakkan karakter. Algoritma yang digunakan untuk pengklasifikasian citra adalah algoritma *Convolutional Neural Network* (CNN). Tahapan dimulai dari *preprocessing* citra gestur tangan yang diperoleh dari *webcam*. Tahap *preprocessing* dimulai dari penempatan *Region Of Interest* (ROI), *grayscale*, *smoothing*, *background subtraction*, *thresholding*, *resizing*, dan normalisasi. Setelah melalui tahap *preprocessing*, citra akan digunakan untuk pembuatan model menggunakan algoritma CNN. Model dari algoritma CNN kemudian dilakukan *testing* terlebih dahulu sebelum diuji coba ke dalam *game*. Model yang memiliki nilai akurasi, presisi, dan *recall* terbesar kemudian akan diaplikasikan kedalam *game*. Dengan memasukkan input berupa video dari *webcam*, program dapat menghasilkan keluaran berupa aksi yang dilakukan oleh karakter berdasarkan input dari pengguna.

II. METODE PENELITIAN

Penelitian ini bertujuan untuk membuat aplikasi yang mampu mengenali citra untuk mengontrol gerakan karakter dalam *game* Snake dengan menggunakan algoritma CNN. Data citra yang digunakan dalam penelitian ini

adalah citra tangan kanan dengan lima kategori yakni atas, bawah, kiri, kanan dan latar.

Alur kerja sistem dimulai dengan video yang direkam oleh *webcam* kemudian dibaca secara *frame per frame*, melalui tahap *preprocessing*, lalu masuk ke dalam model klasifikasi CNN. Model kemudian mengklasifikasikan *frame* atau citra yang telah melalui tahap *preprocessing* menjadi sebuah gerakan karakter dalam *game* Snake.



Gambar 1 Alur Kerja Sistem

A. Tinjauan Pustaka

1. Citra

Citra atau *image* merupakan istilah lain dari gambar. Citra adalah informasi yang berbentuk visual. Suatu citra dapat dilihat dari penangkapan sinar yang dipantulkan oleh objek. Ketika sumber cahaya menerangi suatu objek, objek memantulkan kembali sebagian dari cahaya tersebut. Pantulan tersebut ditangkap oleh alat pengindera optik, seperti mata manusia, kamera, dll. Bayangan dari objek tersebut akan terekam sesuai dengan intensitas dari pantulan cahaya. Saat mesin digital yang merekam pantulan cahaya, maka citra yang dihasilkan adalah citra digital [3].

Citra digital adalah representasi dari citra dua dimensi yang berisi kumpulan nilai digital yang disebut sebagai piksel. Piksel adalah elemen terkecil yang menyusun sebuah citra dan berisi tentang nilai yang mewakili kecerahan dari sebuah warna pada suatu titik tertentu [3].

2. Computer Vision

Computer vision adalah suatu proses transformasi data yang diperoleh melalui kamera berupa video maupun gambar menjadi sebuah keputusan ataupun sebuah presentasi yang baru, dimana hasil dari kegiatan transformasi tersebut memiliki kepentingan untuk mencapai suatu tujuan. Data yang dimasukkan ke dalam kegiatan transformasi tersebut memiliki beberapa informasi yang tersirat, contohnya sebuah gambar yang di dalamnya terdapat berbagai objek dan kita ingin mengetahui jumlah objek tersebut [4].

3. Preprocessing

Preprocessing merupakan langkah awal dalam mengolah suatu data. Data yang diolah tidak selalu dalam kondisi yang ideal untuk diproses. Oleh karenanya perlu melakukan tahap *preprocessing* untuk menghilangkan bagian-bagian yang dapat mengganggu hasil dari suatu proses data [5].

4. Convolutional Neural Network

Convolutional Neural Network (CNN) merupakan salah satu algoritma *deep learning* yang didesain untuk mengolah data bertopologi *grid* [6]. CNN termasuk dalam jenis *Deep Neural Network* yang telah teruji hasilnya dalam bidang pengenalan citra. Selain data citra, CNN juga dapat mengolah data *temporal*, *spatial*, dan *spatiotemporal*. CNN terdiri dari *convolutional layer*, *pooling layer*, dan *fully connected layer*.

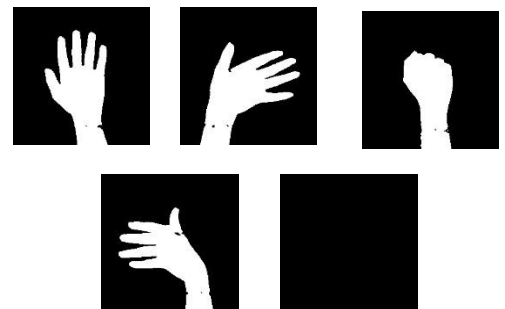
5. Game

Game atau yang dikenal dengan istilah gim adalah sebuah aktivitas rekreasi dengan tujuan bersenang-senang, mengisi waktu luang, atau berolahraga ringan. *Game* biasanya dilakukan sendiri atau bersama-sama. Di dalam sebuah *Game*, terdapat target-target yang perlu dicapai oleh para pemain [7]. Hal yang esensial dari sebuah *game* adalah adanya aturan yang dapat menuntun jalannya sebuah permainan. Dengan adanya peraturan, pemain dapat bereksplorasi dalam *game* tetapi tidak terlepas dari aturan yang telah dibuat [8].

Game Snake merupakan *Game* dimana pemain mengendalikan sebuah *snake* yang dapat tumbuh memanjang dalam sebuah ruang. *Snake* tersebut dapat bertambah panjang ketika *snake* tersebut memakan poin yang tersebar secara acak di dalam suatu ruang. *Game* akan berakhir ketika *snake* tersebut menabrakkan kepalanya ke bagian tubuhnya atau ketika *snake* tersebut menabrak dinding ruang.

B. Data Preparation

Dataset yang digunakan dalam penelitian dikumpulkan dengan cara yang berbeda-beda. Data latih dikumpulkan dengan cara melakukan *looping* untuk setiap *frame* atau citra yang ditangkap oleh *webcam*, citra tersebut kemudian melalui tahapan *preprocessing*, dan disimpan. Jadi, selama dalam keadaan *looping*, subjek menggerakkan gestur tangan secara terus menerus dalam kurun waktu tertentu untuk pengambilan data latih. Data validasi dikumpulkan dengan cara menyimpan citra yang dibaca oleh *webcam* dan telah melalui tahapan *preprocessing* secara satu per satu. Cara pengumpulan data uji sama seperti data validasi, namun data uji dikumpulkan dengan ukuran tangan yang berbeda.



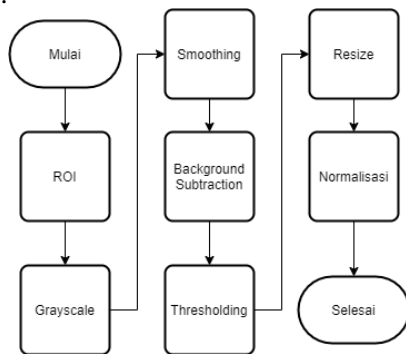
Gambar 2 Kategori gestur tangan (atas, kanan, bawah, kiri, dan latar)

C. Preprocessing

Tahap *preprocessing* dalam penelitian ini dimulai dengan menetapkan *Region of Interest* (ROI). ROI digunakan sebagai daerah yang difokuskan untuk merekam citra sehingga daerah lain diluar ROI diabaikan. Citra dalam ROI kemudian diubah menjadi citra keabuan. Tahap *smoothing* diterapkan dengan menggunakan metode *gaussian blur* untuk menghilangkan *noise* pada citra.

Tahap *background subtraction* (absdiff) dilakukan dengan cara menetapkan *frame* pertama yang telah melalui tahap *smoothing* sebagai *background* lalu membandingkannya dengan *frame* berikutnya. Hasil perbandingan antara *frame* awal (*background*) dan *frame-frame* seterusnya kemudian melalui tahap *thresholding* yakni dengan mengubah citra menjadi citra biner (hitam-putih) dengan nilai *threshold* 30. Citra kemudian diubah ukurannya menjadi 50x50 piksel. Setelah itu, citra disimpan dalam format .jpg dan dikelompokkan berdasarkan kategori (atas, bawah, kiri, kanan, dan latar). Citra yang disimpan nantinya akan digunakan untuk pelatihan dan evaluasi model.

Sebelum citra dimasukkan ke dalam model, citra yang dibaca akan melalui tahap normalisasi terlebih dahulu. Pada tahap normalisasi, nilai-nilai dari piksel citra akan dibagi dengan nilai 255, sehingga nilai dari piksel tersebut berada pada rentang nol sampai satu.



Gambar 3 Alur preprocessing

D. Arsitektur Convolutional Neural Network

Pengenalan gestur tangan dalam penelitian ini dapat dilakukan dengan membuat model klasifikasi dari algoritma CNN. CNN akan mengekstraksi fitur dan untuk klasifikasinya menggunakan neural network. Penelitian ini menggunakan 3 macam arsitektur CNN. Fungsi aktivasi yang digunakan yakni ReLu, namun untuk *output layer* menggunakan fungsi aktivasi *softmax*.

1. Arsitektur C1

Tabel 1 Arsitektur C1

Layer	Kernel	Stride	Padding
Input (citra biner 50x50 piksel)			
Convolution 32	3x3	1	VALID
Max Pooling	2x2	2	VALID
Convolution 64	3x3	1	SAME
Max Pooling	2x2	2	VALID
Convolution 128	3x3	1	VALID
Max Pooling	2x2	2	VALID
Flatten			
Fully Connected 1024			
Dropout 0,5			
Fully Connected 1024			
Fully Connected 5			

Arsitektur C1 terdiri dari *input layer*, tiga *convolution layer*, tiga *max pooling layer*, *flatten layer*, *dropout layer* dengan rasio 0,5, dan tiga *fully connected layer*. Ukuran *filter* pada *convolution layer* yakni 3x3 dengan jumlah 32 *output filters* pada layer pertama, 64 *output filters* pada layer kedua, dan 128 *output filters* untuk layer ketiga. Ukuran *filter* pada *max pooling layer* adalah 2x2. Jumlah *node* pada *fully connected layer* pertama, kedua, dan ketiga masing-masing adalah 1024, 1024, dan 5.

2. Arsitektur C2

Tabel 2 Arsitektur C2

Layer	Kernel	Stride	Padding
Input (citra biner 50x50 piksel)			
Convolution 32	3x3	1	VALID
Max Pooling	2x2	2	VALID
Convolution 64	3x3	1	VALID
Convolution 64	3x3	1	VALID
Max Pooling	2x2	2	VALID
Convolution 128	3x3	1	VALID
Convolution 128	3x3	1	VALID
Max Pooling	2x2	2	VALID
Flatten			
Fully Connected 1024			
Dropout 0,5			
Fully Connected 1024			
Fully Connected 5			

Arsitektur C2 terdiri dari *input layer*, lima *convolution layer*, tiga *max pooling layer*, *flatten layer*, *dropout layer* dengan rasio 0,5, dan tiga *fully connected layer*. Ukuran *filter* pada *convolution layer* yakni 3x3 dengan jumlah 32 *output filters* pada *layer* pertama, 64 *output filters* pada *layer* kedua dan ketiga, serta 128 *output filters* untuk *layer* keempat dan kelima. Ukuran *filter* pada *max pooling layer* adalah 2x2. Jumlah *node* pada *fully connected layer* pertama, kedua, dan ketiga masing-masing adalah 1024, 1024, dan 5.

3. Arsitektur C3

Tabel 3 Arsitektur C3

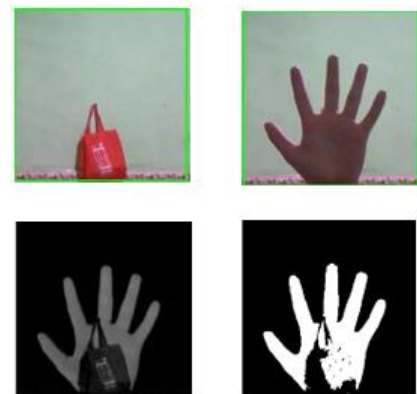
Layer	Kernel	Stride	Padding
Input (citra biner 50x50 piksel)			
Convolution 32	3x3	1	VALID
Max Pooling	2x2	2	VALID
Convolution 64	3x3	1	VALID
Max Pooling	2x2	2	VALID
Convolution 128	3x3	1	VALID
Convolution 128	3x3	1	VALID
Convolution 128	3x3	1	VALID
Flatten			
Fully Connected 1024			
Dropout 0,5			
Fully Connected 1024			
Fully Connected 5			

Arsitektur C3 terdiri dari *input layer*, lima *convolution layer*, dua *max pooling layer*, *flatten layer*, *dropout layer* dengan rasio 0,5, dan tiga *fully connected layer*. Ukuran *filter* pada *convolution layer* yakni 3x3 dengan jumlah 32 *output filters* pada *layer* pertama, 64 *output filters* pada *layer* kedua, serta 128 *output filters* untuk *layer* ketiga, keempat, dan kelima. Ukuran *filter* pada *max pooling layer* adalah 2x2. Jumlah *node* pada *fully connected layer* pertama, kedua, dan ketiga masing-masing adalah 1024, 1024, dan 5.

III. PEMBAHASAN

Hasil *preprocessing* citra dengan menggunakan latar polos menghasilkan citra biner dengan optimal. Namun ketika, terdapat

objek pada latar, lalu objek tangan menutupi objek pada latar, dapat menghasilkan citra biner yang kurang optimal. Hal tersebut dikarenakan, metode *absdiff* yang digunakan untuk melakukan *background subtraction* bekerja dengan cara mengurangi nilai piksel citra latar dengan citra yang dibandingkan. Hasilnya, ketika citra hasil *background subtraction* memiliki nilai piksel dibawah 30, nilai piksel akan diubah menjadi warna hitam. Hal tersebut yang membuat citra biner tidak optimal ketika terdapat objek pada *background*.



Gambar 4 Hasil preprocessing jika ada objek pada latar

Tabel 4 Hasil pengujian model

Konfigurasi	Akurasi	Presisi	Recall
C1	97,8	97,8	97,8
C2	96,9	96,9	96,9
C3	98	98,1	98

Berdasarkan hasil evaluasi yang dilakukan pada ketiga model, diketahui bahwa model C3 memiliki nilai akurasi, presisi, dan *recall* yang lebih tinggi dibandingkan model lainnya. Oleh karena itu model C3 dipilih menjadi model yang akan disinkronisasikan dengan *game Snake*.

Penerapan model klasifikasi pada *game Snake* berhasil dilakukan. Dengan menggunakan *webcam* lalu melakukan *preprocessing* pada citra dan kemudian memasukkan citra hasil *preprocessing* ke dalam model. Model klasifikasi dapat mengeluarkan *output* berupa kategori citra yang dapat digunakan sebagai arah gerak *snake*. Untuk kategori atas, arah gerak *snake* dapat bergerak ke arah atas, begitu pula untuk kategori kiri,

kanan, dan bawah. Sedangkan kategori latar tidak melakukan apapun.

IV. SIMPULAN

Simpulan yang dapat diambil dari penelitian ini adalah:

1. Menghasilkan dataset citra gestur tangan sebanyak 10000 citra dengan format JPG.
2. Berdasarkan hasil pengujian terhadap tiga arsitektur CNN yang digunakan dalam penelitian ini, arsitektur C3 merupakan arsitektur yang memberikan hasil paling baik. Arsitektur C3 yang terdiri dari *input layer*, lima *convolution layer* (32, 64, 128, 128, 128) dengan ukuran *filter* 3x3, dua *max pooling layer* dengan ukuran *filter* 2x2, *flatten layer*, *dropout layer* dengan rasio 0,5, dan tiga *fully connected layer* (1024, 1024, 5) dapat menghasilkan nilai akurasi dan *recall* sebesar 98% serta 98,1% untuk nilai presisi.
3. Metode *grayscale*, *smoothing*, *background subtraction* (absdiff), dan *thresholding* dapat menghasilkan citra objek biner secara efektif jika kondisi *background* tetap dan tidak ada objek pada *background*.

DAFTAR RUJUKAN

- [1] I. Komang and S. Buana, "Aplikasi Mendeteksi Gerakan Tangan untuk Bermain Game Pingpong dengan Teknik Pengolahan Citra Application to Detect Hand Movements to Play Ping Pong Game with Image Processing Technique," *Citec J.*, vol. 4, no. 1, pp. 1–8, 2016.
- [2] M. R. Islam, U. K. Mitu, R. A. Bhuiyan, and J. Shin, "Hand gesture feature extraction using deep convolutional neural network for recognizing American sign language," *2018 4th Int. Conf. Front. Signal Process. ICFSP 2018*, pp. 115–119, 2018, doi: 10.1109/ICFSP.2018.8552044.
- [3] M. Buana, "Modul Pengolahan Citra," p. 166, 2004.
- [4] R. A. Safitri, "Implementation of Viola and Jones methods for identifying fingers sign as a means of navigation music players application," Universitas Komputer Indonesia, 2015.
- [5] Samir, "What was Preprocessing," 2019. [Online]. Available: <https://hendroprasetyo.com/apa-itu-preprocessing/#.Xn3GvIgzaUk>. [Accessed: 27-Mar-2020].
- [6] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*.
- [7] R. Wardhani and M. H. Yaqin, "Game Dasar-Dasar

Hukum Islam Dalam Kitab Mabadi'ul Fiqh Jilid I," vol. 5, 2013.

- [8] H. Silvianita, "Pengertian Game Beserta Sejarah, Manfaat, serta Jenis-Jenis Game, Lengkap!," 2019. [Online]. Available: <https://www.nesabamedia.com/pengertian-game/>. [Accessed: 27-Mar-2020].