

**Jurnal Politeknik Caltex Riau**Terbit Online pada laman <https://jurnal.pcr.ac.id/index.php/jkt/>
| e- ISSN : 2460-5255 (Online) | p- ISSN : 2443-4159 (Print) |

Deteksi Tangan Otomatis Pada Video Percakapan Bahasa Isyarat Indonesia Menggunakan Metode *Deep Gated Recurrent Unit* (GRU)

Gulpi Qorik Oktagalu Pratamasunu¹, Fathorazi Nur Fajri² dan Puji Kurnia Sari³¹Universitas Nurul Jadid, email: pratamasunu@unuja.ac.id²Universitas Nurul Jadid, email: r4si.bint4ng@gmail.com³Universitas Nurul Jadid, email: pujikurniasari29@gmail.com

[1] Abstrak

Tunarungu adalah hilangnya pendengaran baik salah satu telinga maupun keduanya. Untuk melakukan komunikasi tunarungu menggunakan bahasa isyarat menggunakan kedua tangannya. Saat ini terdapat 2 sistem bahasa isyarat yaitu BISINDO dan SIBI. Bahasa isyarat adalah bahasa yang menggunakan gerakan tubuh berupa tangan dan bibir untuk menjelaskan sebuah arti. Namun masih banyak masyarakat yang belum paham dengan bahasa isyarat, oleh karena itu dilakukan sebuah penelitian dengan menggunakan metode *deep gated recurrent unit* (GRU) yang dapat membaca gerakan tangan pada video bahasa isyarat indonesia. Penelitian ini menggunakan 3 class yaitu ucapan hallo, terimakasih, dan sama sama. Tahapan yang dilakukan pada penelitian ini terdiri dari pengolahan video atau proses pre-processing pada video. Pengujian terhadap 45 video data training dan 36 video data testing. Pengujian yang dilakukan terhadap data testing menghasilkan nilai akurasi sebesar 88%.

Kata kunci: Bahasa isyarat, Tunarungu, Deep gated recurrent unit

[2] Abstract

Deafness is the loss of hearing in one or both ears. To communicate with the deaf, use sign language using both hands. Currently there are 2 sign language systems, namely BISINDO and SIBI. Sign language is a language that uses body movements in the form of hands and lips to explain a meaning. However, there are still many people who do not understand sign language, therefore a study was conducted using the deep gated recurrent unit (GRU) method which can read hand movements on Indonesian sign language videos. This study uses 3 classes, namely hello, thank you, and the same. The stages carried out in this research consist of video processing or video pre-processing. Testing of 45 video data training and 36 video data testing. Tests carried out on testing data resulted in an accuracy value of 88%.

Keywords: Sign language, Deaf, Deep gated recurrent unit

1. Pendahuluan

Bahasa adalah alat atau media untuk berkomunikasi antara manusia satu dengan yang lain [1]. Setiap negara pasti mempunyai bahasa masing-masing. Di Indonesia masyarakat menggunakan bahasa verbal untuk berkomunikasi sehari-hari. Bahasa verbal adalah bahasa yang dapat dilakukan melalui suara dan tulisan.

Namun, tidak semua orang dapat menggunakan bahasa verbal dengan sempurna. Seperti halnya kaum tunarungu dan tunawicara mereka tidak bisa melakukan komunikasi dengan baik. Tunarungu atau tuli adalah hilangnya kemampuan untuk mendengar dari salah satu atau kedua telinga. Bahasa isyarat adalah metode komunikasi yang tidak menggunakan suara, tetapi menggunakan gerakan tangan, tubuh dan bibir untuk menyampaikan sebuah informasi atau biasa dikenal dengan komunikasi *nonverbal* [2].

Di Indonesia mempunyai 2 bahasa isyarat yang sering digunakan, Bahasa Isyarat Indonesia (BISINDO) dan Sistem Isyarat Bahasa Indonesia (SIBI) [3]. SIBI secara resmi digunakan oleh semua SLB di bawah Kementerian Pendidikan dan Kebudayaan (Kemendikbud). Sedangkan BISINDO adalah bahasa yang digalakkan oleh Gerakan Kesejahteraan Tunarungu Indonesia (GERKATIN) dan dikembangkan oleh masyarakat tunarungu sendiri. Namun saat ini BISINDO masih kurang dikenal oleh masyarakat luas, karena kurangnya pemahaman manusia normal terhadap BISINDO maka dapat menghambat komunikasi antar penyandang tuna rungu dan tuna wicara dengan masyarakat [4]. Oleh karena itu, dibutuhkan suatu metode yang dapat membaca gerakan isyarat tangan untuk diterjemahkan menjadi sebuah teks.

Saat ini banyak sekali pemanfaatan computer vision diberbagai bidang seperti pada bidang perkebunan [5], transportasi [6]. Beberapa penelitian telah di kembangkan sebelumnya salah satunya ialah Pengenalan Abjad pada Video Bahasa Isyarat Indonesia (BISINDO) menggunakan Metode Ekstraksi Warna dan *Deep Learning* dengan Penghapusan Wajah, hasil akurasi yang diperoleh 92% dan telah berhasil dilakukan penghapusan wajah. Namun sistem kurang efektif karena metode yang digunakan hanya dapat membaca satu gambar saja berupa frame dan belum bisa membaca gerakan sehingga informasi yang diperoleh kurang banyak.

Oleh karena itu dibutuhkanlah sebuah metode yang dapat membaca seluruh video secara runtut sehingga informasi yang diperoleh bisa lebih akurat, maka metode yang akan dilakukan pada penelitian ini adalah metode *Deep Gated Recurrent Unit*. *Gated Recurrent Unit* (GRU) adalah struktur berulang yang dirancang dengan hati-hati dan membuat trade-off yang baik antara kinerja dan kecepatan yang berfungsi untuk membuat setiap recurrent unit agar dapat menangkap dependencies dalam skala waktu yang berbeda beda secara adaptif.

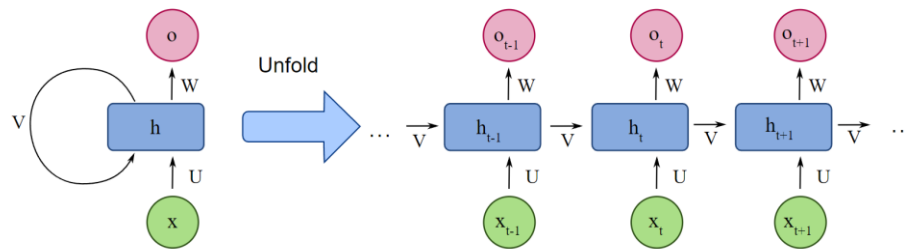
2. Tinjauan Pustaka

2.1 Bahasa Isyarat Indonesia (BISINDO)

Bahasa isyarat adalah bahasa yang universal karena bahasa isyarat dipakai oleh penyandang tunarungu ketika mereka berbicara atau melakukan komunikasi dengan orang normal atau dengan sesamanya [7]. Bahasa Isyarat Indonesia pertama kali dicetuskan oleh Dimiyati Hakim selaku ketua DPP PERTRI ia adalah seorang tunarungu, Dimiyati menciptakan bahasa isyarat indonesia ini, dikarenakan banyak terjadinya pertikaian dan polemik terkait penerapan dan penggunaan bahasa isyarat di Indonesia [8]. Penggunaan bahasa isyarat dalam pergaulan komunitas tunarungu harus memiliki 3 syarat utama diantaranya ialah kecepatan, keringkasan, dan pemahaman.

2.2 Recurrent Neural Network (RNN)

Recurrent Neural Network (RNN) adalah sebuah jaringan syaraf berulang, mengapa dikatakan demikian karena nilai *neuron* pada *hidden layer* sebelumnya akan diproses kembali sebagai data input. *Neuron* yang ada pada *hidden layer* akan disimpan ke dalam context layer lalu akan terus dilakukan update data hingga kondisi RNN terpenuhi [9]. Seperti halnya manusia, *Recurrent Neural Network* menyimpan informasi dari masa lalu dengan melakukan *looping* dalam arsitekturnya. Simulasi RNN dapat dilihat pada ilustrasi Gambar 2.

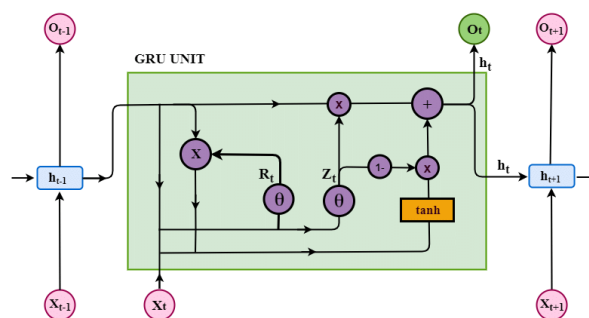


Gambar 1. Arsitektur RNN

2.3 Gated Recurrent Unit (GRU)

GRU adalah arsitektur yang diciptakan oleh Kyunghun Cho pada tahun 2014, GRU memberikan konvergensi yang lebih cepat dan hasilnya bisa dibandingkan dengan LSTM. Berdasarkan hasil penelitian R. Fu, 2016 dalam jurnal [10], menunjukkan bahwa metode GRU memiliki nilai kesalahan (error) yang lebih kecil dibandingkan metode LSTM.

Tujuan dari GRU adalah untuk membuat setiap *recurrent unit* agar bisa menangkap setiap hubungan (dependensi) dalam skala waktu yang berbeda-beda yang dilakukan secara adaptif. GRU mempunyai sebuah komponen pengatur alur informasi yang disebut dengan *gate* dan GRU disini mempunyai 2 *gate* yaitu, *reset gate* dan *update gate* [9]. Struktur dari GRU bisa dilihat pada Gambar 3. Struktur GRU di bawah ini, *r* melambangkan *reset gated*, *z* melambangkan update gate, sedangkan *h* dan \tilde{h} melambangkan *activation* dan *candidate activation*.



Gambar 2. Arsitektur GRU

3. Metode Penelitian



Gambar 3. Kerangka Penelitian

3.1 Pengumpulan Dataset

Pada penelitian ini dilakukan pengumpulan data yang diambil dari sebuah video Bahasa Isyarat Indonesia (BISINDO). Serta menggunakan video yang diambil dari relawan mahasiswa Universitas Nurul Jadid. Pada saat pengambilan video posisi tangan tidak boleh menutupi wajah, serta pengambilan video ini dilakukan di dalam ruangan dengan menggunakan baju dan kerudung yang berbeda – beda. Alat yang digunakan ialah smartphone dengan kualitas HD 1280 x 720.

3.2 Pre-Processing

Pre-processing yang dilakukan pada penelitian ini adalah proses dalam mengedit beberapa video BISINDO agar mendapatkan ukuran yang sama, dengan melakukan convert video dari bentuk *.MP4 ke bentuk *.AVI. Data video asli dipotong dari resolusi awal 1280 x 720 *pixel* menjadi resolusi 704 x 384 *pixel* proses memperkecil video ini bertujuan untuk mempermudah proses perhitungan di dalam sistem.

3.3 Modeling

Pada proses modeling dilakukan *combine model* antara RNN dan GRU. Dimana tujuan penggunaan GRU untuk membuat setiap *recurrent unit* agar bisa menangkap hubungan (dependensi) dalam skala waktu yang berbeda-beda dan secara adaptif [10]. Implementasi metode *Deep Gated Recurrent Unit* (GRU) ialah untuk mengetahui pose tangan. Mesin atau komputer akan dilatih dengan menggunakan beberapa video yang telah diambil. Selanjutnya video akan di inputkan untuk melatih mesin atau komputer dengan data yang sudah ada. Sehingga dapat memberikan hasil isyarat dari deteksi bahasa isyarat tangan dengan baik

3.4 Testing

Proses pengujian dilakukan dengan cara menguji model dengan melihat keakuratan model yang dihasilkan. Berikut rumus yang digunakan pada pengujian.

$$Akurasi = \frac{\sum DataBenar}{\sum DataUji} \times 100\% \tag{1}$$

4. Pembahasan

4.1 Pengumpulan Dataset

Pada penelitian ini data yang diambil adalah dengan membuat sebuah video berbahasa isyarat indonesia (BISINDO) berupa isyarat percakapan sehari-hari. Pengambilan video ini dilakukan dengan jarak kurang lebih 70 cm antara objek dan kamera dengan posisi kamera tetap harus lurus dengan objek. Pakaian yang digunakan disini bebas dan tangan tidak menutupi wajah saat pengambilan video berlangsung. Hasil video yang telah diambil ini nantinya akan dijadikan *data training* dan *data testing*.



Gambar 4. Hasil *dataset* yang telah terkumpul

Data video yang terkumpul berjumlah 81 video yang di dalamnya ada 3 gerakan bahasa isyarat sehari-hari. Video tersebut akan dibagi menjadi *data training* dan *data testing*, untuk *data training* menggunakan 45 video dan data testing menggunakan 36 video. Resolusi video adalah 1280×720 *pixel* dengan durasi video 39 menit dengan format *.mp4. Video gerakan bahasa isyarat indonesia ini belum bisa diolah ke dalam sistem karena ukuran video yang terlalu besar maka perlu dilakukan *compression* agar video dapat di proses oleh sistem dengan baik dan cepat.

4.2 Pre-Processing

Pada tahap *preprocessing* ini adalah memotong video dengan membuang hal-hal yang tidak dibutuhkan dalam video tersebut sehingga dapat menjadi sebuah video yang sempurna dengan penggunaan bahasa isyarat yang benar, serta merubah format video ke dalam bentuk *.AVI.



Gambar 5. Proses pemotongan video

Data video yang terkumpul berjumlah 81 video yang di dalamnya ada 3 gerakan bahasa isyarat sehari-hari. Video tersebut akan dibagi menjadi *data training* dan *data testing*, untuk *data training* menggunakan 45 video dan data testing menggunakan 36 video. Ukuran data video harus disamakan yaitu 704×384 *pixel* agar sistem dapat membaca lebih baik ketika diproses pada metode *Gated Recurrent Unit*.

4.3 Modeling

4.3.1 Import Tensorflow

Tahap awal yang harus dilakukan ialah memasukkan *library tensorflow* kedalam Google Colab. Keberadaan *tensorflow* ini digunakan untuk mempercepat performa pada *platform* Google Colab, serta dapat menangani detail dibalik layar dengan baik

4.3.2 Sliding frame

Sliding frame ini berfungsi untuk menslide atau menggeser setiap gerakan yang dilakukan oleh tangan. Dengan adanya proses *sliding* ini video akan benar benar diseleksi dengan baik untuk menghasilkan akurasi atau hasil yang bagus.

4.3.3 Inisialisasi *hyperparameter* model GRU

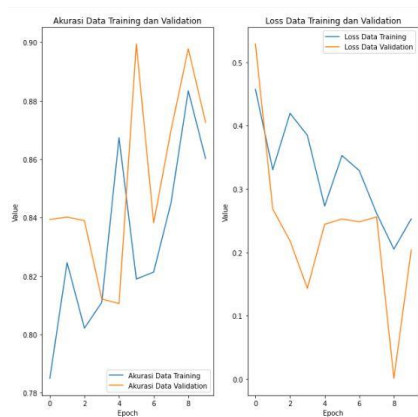
Hyperparameter merupakan salah satu faktor yang berpengaruh dalam mendapatkan performa yang lebih tinggi dalam melatih model *Gated Recurrent Unit*. Terdapat beberapa percobaan untuk mendapatkan *hyperparameter* yang maksimal nilai akurasi dan errornya. Pada model ini *Hyperparameter* yang di inisialisasikan ialah *batch_size* 8, *nb_frames* 5, *split_val* 0.2, *momentum* 0.2, dan memakai *epoch* 10.

4.3.4 Menampilkan video BISINDO

Setelah data diproses oleh sistem dengan dilakukan *sliding frame* generator maka data video akan ditampilkan sesuai dengan jumlah *frame* yang diinginkan. Pada Gambar 6 menunjukkan hasil *data training* yang telah dilakukan proses *sliding* dengan 5 *frame* setiap video.

4.3.5 Training data BISINDO

Proses *running* ini dilakukan selama 24 jam atau 1 hari. Proses *training* menggunakan *epoch* 10 yang modelnya akan disimpan dalam bentuk “.h5” untuk menjadi bekal pada tahapan *testing*. Gambar 7. Grafik hasil dari *epoch* 10 yang terdiri dari nilai *loss*, *accurasi*, *val_loss*, dan *val_accuracy*.



Gambar 6. Grafik hasil *epoch* 10

4.3.6 Akurasi

Nilai *accuracy* ini diperoleh dari hasil *epoch* yang dijalankan yaitu *epoch* 1 sampai *epoch* 10 untuk mencari nilai *accuracy* tertinggi. Hasil dari *epoch* 10 ini mendapat nilai *loss* sebesar 0,2034, nilai *accuracy* sebesar 0,9115, dan dengan waktu 794 second.

Tabel 1. Nilai akurasi *epoch* 10

<i>Epoch</i>	Nilai <i>Accuracy</i>	Nilai <i>Loss</i>
1	0,8269	0,3939
2	0,8157	0,4561
3	0,8373	0,3181
4	0,8566	0,2815
5	0,8954	0,2355

6	0,8898	0,2237
7	0,8682	0,3123
8	0,8770	0,2926
9	0,9115	0,2034
10	0,9099	0,2340

4.4 Testing

Uji coba yang dilakukan dengan menggunakan 36 video BISINDO yang terdiri dari tiga *class* yang mana setiap *class* terdapat 12 data video. Hasil dari tahap uji coba data *testing* sebanyak 36 video dan menggunakan *epoch* 10 yang telah dilakukan dalam tahap implementasi menghasilkan nilai akurasi sebesar 88%. Dengan jumlah data benar atau data yang dapat terdeteksi sebanyak 32 data video dan 4 data video tidak dapat terdeteksi oleh sistem.

Tabel 2. Hasil *testing dataset*

Banyak data uji coba	Data benar	Data salah	Nilai <i>accuracy</i>
36	34	4	88%

5. Kesimpulan dan Saran

5.1 Kesimpulan

Dari penelitian ini bisa ditarik kesimpulan bahwa deteksi tangan otomatis pada video percakapan bahasa isyarat indonesia (BISINDO) menggunakan metode *Deep Gated Recurrent Unit* (GRU) berhasil untuk diterapkan. Dengan melakukan pencarian akurasi terbaik pada *epoch* 10 dengan kombinasi pencarian model menggunakan *batch_size* 8, *nb_frames* 5, *split_val* 0.2, dan momentum 0.2. Hasil dari uji coba yang telah dilakukan ialah mendapat nilai akurasi sebesar 88% yang menunjukkan bahwa kemampuan deteksi tangan pada percakapan bahasa isyarat indonesia (BISINDO) sudah baik.

5.2 Saran

Saran dan masukan yang akan disampaikan dari penelitian ini untuk kedepannya adalah sebagai berikut:

1. Deteksi tangan otomatis pada video percakapan bahasa indonesia dengan menggunakan metode *Deep Gated Recurrent Unit* ini dapat dikembangkan lebih lanjut lagi sehingga bisa menghasilkan sebuah sistem aplikasi yang bagus sert siap pakai sehingga dapat membantu kaum tunarungu atau tunanetra dalam berkomunikasi sehari hari.
2. Melakukan penambahan data dan variasi pada video BISINDO sehingga sistem dapat membaca atau mendeteksi gerakan lebih baik lagi.
3. Menggunakan komputer atau processor yang lebih besar dan lebih baik serta menggunakan kapasitas memori yang lebih besar sehingga dapat menghasilkan akurasi yang lebih baik lagi.

Daftar Pustaka

- [1] R. I. Borman, "Klasifikasi Objek Kode Tangan Pada Pengenalan Alphabet Bahasa Isyarat Indonesia," *Seminar Nasional Informatika dan Aplikasinya (SNIA) 2017*, 2017.
- [2] N. N. Aslinda, "Pengenalan Isyarat Tangan Berupa Huruf pada Video Percakapan Bahasa Isyarat Indonesia (BISINDO) menggunakan Metode Ekstraksi Warna Kulit dan Deep Learning," *Skripsi Teknik Informatika*, 2019.
- [3] Nuryazid, "Pengembangan Aplikasi Kamus Bahasa Isyarat Indonesia (Bisindo) Dengan Mengintegrasikan Cloud Video Berbasis Android," 2016.
- [4] E. R. Fitriana, "Pengenalan Abjad pada Video Bahasa Isyarat Indonesia (BISINDO) menggunakan Metode Ekstraksi Warna dan Deep Learning dengan Penghapusan Wajah," *Skripsi Teknik Informatika*, 2020.
- [5] F. N. Fajri, N. Hamid and R. A. Pramunendar, "The recognition of mango varieties based on the leaves shape and texture using back propagation neural network method," in *2017 International Conference on Sustainable Information Engineering and Technology (SIET). IEEE*, 2017.
- [6] M. Hasanah , G. Q. O. Pratamasunu and R. E. Pawening, "Automatic Car Detection Using Haar Cascade Classifier and Convolutional Neural Network for Traffic Density Estimation," *Indonesian Journal of Artificial Intelligence and Data Mining*, vol. 4, no. 1, pp. 11-18, 2021.
- [7] M. D.K, "Bahasa Isyarat Indonesia Dikomunitas Gerakan Untuk Kesejahteraan Tunarungu Indonesia.," *Skripsi Komunitas Islam* , 2017.
- [8] R. Zahro, "Pengenalan Isyarat Tangan berupa Angka pada Video Percakapan Bahasa Isyarat Indonesia menggunakan Metode Ekstraksi Warna Kulit dan Deep Learning," *Skripsi Teknik Informatika*, 2019.
- [9] R. P. Wardana, "Penerapan Model Gated Recurrent Unit Untuk Peramalan Jumlah Penumpang Kereta Api di PT. KAI (Persero)," *Skripsi Fakultas Ilmu Sains dan Teknologi*, 2020.
- [10] R. A. saputra, "Prediksi Permintaan Kargo pada Cargo Service Center Tangerang City Menggunakan Metode Gated Recurrent Unit," 2020.