

RANCANG BANGUN PENERJEMAH BAHASA ISYARAT MENGGUNAKAN PENGOLAHAN CITRA DENGAN METODE *YOU ONLY LOOK ONCE (YOLO)*

Syahrul Hidayatullah¹, Ii Joko Endrasmono², M. Khoirul Hasin³

¹²³Program Studi Teknik Otomasi, Jurusan Teknik Kelistrikan Kapal

Politeknik Perkapalan Negeri Surabaya

Jalan Teknik Kimia, Kampus ITS Sukolilo, Surabaya 60111

E-mail : syahrulhidayatullah@student.ppns.ac.id

Abstrak

Komunikasi merupakan salah satu hal penting bagi setiap individu agar dapat menyampaikan pesan kepada orang lain dan bersosialisasi terhadap lingkungan masyarakat. Bahasa isyarat merupakan alat komunikasi bagi penyandang tuna rungu dan tuna wicara untuk menjalin hubungan, bersosialisasi, memahami satu sama lain, dan sebagainya. Namun tidak semua lawan bicara memahami tentang bahasa isyarat. Oleh karena itu, perantara atau alat bantu penerjemah bahasa isyarat sangatlah diperlukan yaitu dengan menggunakan pengolahan citra untuk mengklasifikasi gerakan Bahasa Isyarat yang dideteksi menggunakan *webcam* yang telah diprogram untuk mendeteksi gerakan Bahasa Isyarat secara *real-time* dengan menggunakan metode *You Only Look Once (YOLO)* yang dianggap metode yang bisa digunakan oleh peneliti dalam penyusunan Tugas Akhir ini. Hasil deteksi dan klasifikasi akan ditampilkan pada *interface* dengan *output* suara dan teks dari perangkat. Hasil deteksi dengan metode *You Only Look Once (YOLO)* memiliki beberapa variasi hasil. Dari 20 kata percobaan, terdapat 19 kata yang mendeteksi secara sempurna dan 1 kata yang mendapat persentase keberhasilan sebesar 90%. Proses evaluasi mendapatkan nilai yang bervariasi seperti *accuracy* sebesar 94%, *precision* sebesar 99,9%, dan *recall* sebesar 100%. Proses deteksi memerlukan tingkat pencahayaan sebesar 77,5 lux agar dapat mendapatkan hasil yang maksimal.

Kata Kunci: Bahasa Isyarat, Pengolahan Citra, *You Only Look Once, real-time*.

1. PENDAHULUAN

Komunikasi merupakan salah satu hal penting bagi setiap individu agar dapat menyampaikan pesan kepada orang lain dan bersosialisasi terhadap lingkungan masyarakat. Ada banyak cara yang dilakukan untuk berkomunikasi, diantaranya adalah dengan lisan, tulisan ataupun secara isyarat. Tanpa komunikasi, seseorang akan mengalami kesulitan untuk berinteraksi maupun menjalankan segala rutinitasnya, inilah yang menjadi ancaman bagi penderita tuna wicara dan tuna rungu, mereka tidak bisa berkomunikasi secara verbal, sehingga mereka harus menggunakan bahasa isyarat untuk bisa berkomunikasi dengan individu lainnya.

Pada umumnya, Bahasa isyarat digunakan sebagai komunikasi bagi para penyandang tuna rungu dan tuna wicara. Menurut Hallan dan Kauffman (1991) dan Hardman (1990) tuna rungu merupakan gangguan pada organ pendengaran sehingga mengakibatkan ketidakmampuan mendengar, mulai dari yang ringan hingga yang sangat berat yang

dikelompokkan ke dalam tuli (*deaf*) dan kurang dengar (*hard of hearing*). Sedangkan tuna wicara menurut Purwanto (1998) dalam buku *Ortopedagogik Umum* tuna wicara adalah apabila seseorang mengalami kelainan baik dalam pengucapan (artikulasi) Bahasa maupun suaranya dari bicara normal, sehingga menimbulkan kesulitan dalam berkomunikasi lisan dalam lingkungan. Bahasa isyarat merupakan alat komunikasi bagi para penyandang disabilitas.

Bahasa isyarat merupakan alat komunikasi bagi penyandang tuna rungu dan tuna wicara untuk menjalin hubungan, bersosialisasi, memahami satu sama lain, dan sebagainya. Menurut Pujiati (2019) Perbedaan Bahasa isyarat dengan Bahasa lisan terdapat pada alat produksi Bahasa dan alat penerima Bahasa. Bahasa lisan menggunakan alat ucapan manusia (*oral*) dan Bahasa diterima menggunakan indra pendengaran (*auditoris*), sedangkan bahasa isyarat menggunakan gestur dan diterima menggunakan indra penglihatan.

Bahasa isyarat yang digunakan di Indonesia

ada dua jenis, yaitu Sistem Isyarat Bahasa Indonesia (SIBI) dan Bahasa Isyarat Indonesia (Bisindo). SIBI adalah sistem bahasa yang dibakukan dan diresmikan oleh pemerintah Indonesia. Namun karena terdapat imbuhan dan struktur yang sepadan dengan bahasa lisan yang dianggap sulit oleh para penyandang disabilitas. Pada akhirnya organisasi Gerakatin (Gerakan untuk Kesejahteraan Tunarungu Indonesia) memperkenalkan BISINDO pada tahun 1960. Menurut Wijaya (2018) Bisindo menggunakan *visualgestural mode* yang memudahkan pemahaman informasi dan proses komunikasi bagi penyandang tunawicara dan tunarungu.

Menurut International Labour Organization (ILO), sejumlah 15% dari seluruh penduduk di dunia merupakan penyandang disabilitas. Menurut Hasil Survei Ekonomi Nasional (SUSENAS, 2012) menunjukkan jumlah penyandang disabilitas di Indonesia sebanyak 6.008.661 jiwa. Jumlah penyandang disabilitas rungu wicara adalah 472.855. Istilah Disabilitas merupakan kata bahasa indonesia yang berasal dari serapan kata Bahasa Inggris yaitu *Disability* yang berarti cacat atau ketidakmampuan. Permasalahan inilah yang menjadi jurang komunikasi yang memisahkan antara individu normal dengan individu penderita tuna wicara ataupun tuna rungu. Di Indonesia terdapat peraturan yang mengatur tentang hak untuk berkomunikasi bagi penyandang disabilitas sesuai dengan UU No. 8 Tahun 2016.

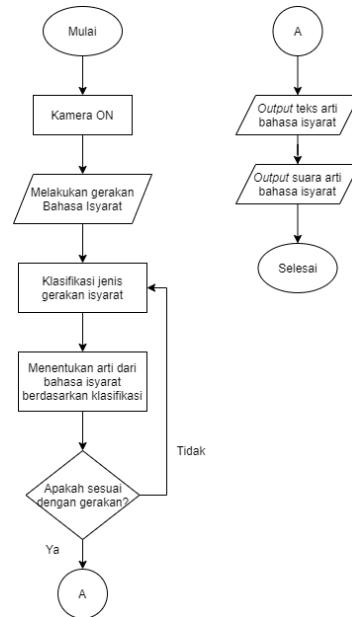
Namun, permasalahan yang sering terjadi di lingkungan sekitar, banyak orang tidak mengerti apa itu bahasa isyarat, sehingga sulit sekali untuk para penyandang disabilitas ini untuk berkomunikasi serta berinteraksi dengan individu lain, oleh karena itu peneliti dalam tugas akhir ini membuat judul “**Rancang Bangun Penerjemah Bahasa Isyarat menggunakan Pengolahan Citra dengan Metode You Only Look Once (YOLO)**” yang akan menjadi jembatan dari jurang yang telah memisahkan komunikasi bagi penyandang disabilitas. Harapannya tidak akan ada lagi kesenjangan komunikasi serta meningkatnya produktivitas para penyandang disabilitas yang bisa meningkatkan kesejahteraan individu masing masing.

2. METODOLOGI

2.1 Diagram Alur Penelitian

Diagram alur penelitian dari sistem ini secara sistematis merupakan langkah – langkah yang dilakukan dalam penelitian ini. Diagram alur penelitian dapat ditunjukkan

pada **Gambar 2.1**.

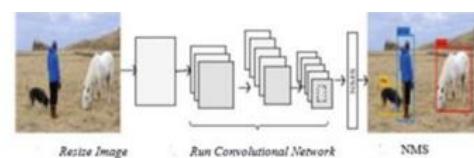


Gambar 2.1 Diagram Alur Penelitian

Pada **Gambar 2.1** dijelaskan bagaimana alur kerja sistem yang akan dibuat. Sistem bermula ketika kamera menyala dan mendekerti gerakan bahasa isyarat. Pada proses klasifikasi jenis gerakan isyarat memerlukan proses pengolahan citra dan menentukan arti dari bahasa isyarat tersebut berdasarkan klasifikasi yang sudah ada. Apabila gerakan bahasa isyarat tadi sesuai dengan klasifikasi maka akan muncul suara dan teks sebagai *output* yang memiliki makna sesuai dengan bahasa isyarat yang diperagakan.

2.2 You Only Look Once (YOLO)

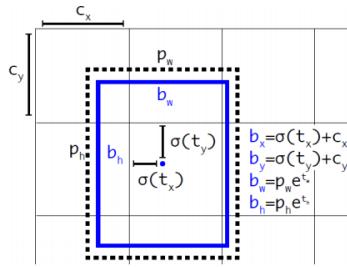
Proses deteksi sebuah gerakan bahasa isyarat menggunakan program yang sudah dibuat sesuai dengan *flowchart* pada **Gambar 2.1**, yang nantinya akan ditambahkan beberapa program untuk kebutuhan analisis Tugas Akhir. Proses pengenalan gerakan bahasa isyarat ini dilakukan dengan metode pendeketkan *You Only Look Once*. YOLO menggunakan pendeketkan jaringan saraf tiruan untuk mendeketkan objek pada sebuah citra (Rachmawati, 2020).



Gambar 2.2 Proses Deteksi YOLO

Pada **Gambar 2.2** dijelaskan bahwa

gambar *input* diubah menjadi ukuran 448 x 448. Lalu melalui proses *Single Convolutional Network* dan hasilnya nanti akan diperoleh berupa sebuah *frame* pembatas dan juga hasil keakurasi yang juga disebut nilai *confidence*. Setelah melalui proses tersebut, NMS (*Non-Max Suppression*) menjadi sebuah *output* yang digunakan untuk menghasilkan *threshold* deteksi yang dihasilkan oleh proses sebelumnya, probabilitas klasifikasi, dan koordinat *bounding box*.



Gambar 2.3 Bounding Box YOLO

Setiap *bounding box* terdapat 5 parameter prediksi yaitu x,y,w,h, dan c seperti **Gambar 3.16**. Koordinat (x,y) adalah pusat kotak relatif pada gambar dan *confidence* adalah *Intersection over Union* (IoU) antara *predicted box* dengan *ground-truth box*. Setiap *grid-cell* memprediksi probabilitas kelas C. Probabilitas itu nantinya akan dikondisikan berdasarkan *grid-cell* pada gambar atau objek, sehingga terdapat satu kelas probabilitas yang terdeteksi pada setiap *grid-cell*. Terdapat formula untuk mendapatkan nilai *confidence* dari **Gambar 2.2** yang dapat dilihat pada (3.1) dan (3.2) :

$$\Pr(\text{Class}_i | \text{Object}) \times \Pr(\text{Object}) \times \text{truthIoU}_{\text{pred}} \quad (3.1)$$

$$\Pr(\text{Class}_i) \times \text{truthIoU}_{\text{pred}} \quad (3.2)$$

2.3 Proses Evaluasi

Proses evaluasi dilakukan untuk mengetahui kecepatan proses deteksi sebuah gerakan bahasa isyarat berdasarkan klasifikasi gerakan tangan. Pada proses evaluasi diukur suatu tingkat akurasi deteksi gerakan, maka akan dihitung berapa banyak gerakan yang terdeteksi benar dari keseluruhan gerakan yang terdeteksi oleh sistem. Proses evaluasi ini melibatkan beberapa parameter yakni:

1) True Positive (TP)

True Positive merupakan objek sebenarnya yang terdeteksi benar oleh sistem

2) False Positive (FP)

False Positive merupakan *noise* yang terdeteksi sebagai objek

3) False Negative (FN)

False Negative merupakan objek

sebenarnya yang terdeteksi salah atau tidak terdeteksi oleh sistem

4) Accuracy

Accuracy merupakan hasil perhitungan dari tingkat akurasi *object detection* terhadap gerakan bahasa isyarat selama melakukan proses deteksi. Formula dari perhitungan *accuracy* dapat dilihat pada (3.3).

$$\text{Accuracy} = \frac{\text{True Positive (TP)}}{\text{Jumlah Keseluruhan Object Detection}} \quad (3.3)$$

5) Precision

Precision merupakan jumlah prediksi yang benar dibandingkan dengan keseluruhan hasil yang diprediksi oleh sistem. Rumus dari perhitungan *precision* dapat dilihat pada (3.4).

$$\text{Precision} = \frac{\text{True Positive}}{(\text{True Positive} + \text{False Positive})} \quad (3.4)$$

6) Recall

Recall merupakan jumlah prediksi yang benar dibandingkan dengan keseluruhan hasil yang sebenarnya. *Recall* dapat dihitung dengan rumus persamaan yang dapat dilihat pada (3.5).

$$\text{Recall} = \frac{\text{True Positive}}{(\text{True Positive} + \text{False Negative})} \quad (3.5)$$

3. HASIL DAN PEMBAHASAN

Pada penelitian ini, hasil pengujian dilakukan 2 kali, gerakan Bahasa Isyarat yang diperagakan dengan intensitas cahaya yang berbeda-beda dan pada Perbedaan bentuk tangan dan juga warna kulit. **Tabel 3.1** merupakan hasil perhitungan proses evaluasi dari metode yang digunakan, **Tabel 3.2** merupakan hasil pendekslan dengan tingkat intensitas cahaya yang berbeda dan **Tabel 3.3** merupakan hasil pendekslan dengan bentuk dan warna kulit yang berbeda.

Tabel 3.1 Hasil Perhitungan Proses Evaluasi Metode *You Only Look Once* (YOLO)

Load Model		Hasil
Halo	AP	100%
	TP	36
	FP	0
Kapan	AP	100%
	TP	43
	FP	0
Makan	AP	100%

Load Model		Hasil
	TP	46
	FP	0
Minum	AP	100%
	TP	50
	FP	0
Tidur	AP	100%
	TP	59
	FP	0
Terimakasih	AP	100%
	TP	79
	FP	0
Sama-sama	AP	100%
	TP	56
	FP	0
Kamu	AP	100%
	TP	62
	FP	0
Coba	AP	100%
	TP	55
	FP	0
Benar	AP	100%
	TP	53
	FP	1
Bertemu	AP	100%
	TP	53
	FP	0
Maaf	AP	100%
	TP	58
	FP	0
Motor	AP	100%
	TP	57
	FP	0
Mobil	AP	100%
	TP	55
	FP	0
Bis	AP	100%
	TP	51
	FP	0
Kereta	AP	100%
	TP	51
	FP	0
Sekarang	AP	100%

Load Model		Hasil
	TP	63
	FP	0
Toilet	AP	100%
	TP	49
	FP	0
Telefon	AP	100%
	TP	93
	FP	0
Semangat	AP	100%
	TP	62
	FP	0
FN		0
Waktu Pemrosesan (s)		1
Precision		1.00
Recall		1.00
F1-score		1.00
IoU		85.84%
mAP @0.5		100%

Berdasarkan Pada **Tabel 3.1** ditunjukkan bahwa nilai mAP tertinggi yang didapatkan adalah 100%. Pada deteksi 20 kelas yang merupakan bahasa isyarat, didapatkan nilai *True Positive* (TP) lebih besar daripada *False Positive* (FP) dimana pada seluruh kelas hanya terdapat 1 *False Positive* (FP) yaitu pada kelas “Benar”. Terdapat juga nilai *Precision* yaitu 1.00 dan juga nilai *Recall* yaitu 1.00.

Tabel 3.2 Hasil Pengujian dengan intensitas cahaya yang berbeda.

No.	Bahasa Isyarat	Lux	Keterangan
1.	Halo	0	Tidak Terdeteksi
2.	Halo	12.5	Terdeteksi
3.	Motor	1.67	Tedeteksi
4.	Terima kasih	2.5	Terdeteksi
5.	Bertemu	3.33	Terdeteksi
6.	Bis	4.17	Terdeteksi
7.	Toilet	9.17	Terdeteksi
8.	Tidur	10.0	Terdeteksi

No.	Bahasa Isyarat	Lux	Keterangan
9.	Mobil	14.17	Terdeteksi
10.	Maaf	77.5	Terdeteksi

Pada **Tabel 3.2** dapat dilihat hasil dari pengujian intensitas cahaya terhadap gerakan bahasa isyarat yang akan dideteksi. Ketika intensitas cahaya 0 lux maka tidak dapat mendeteksi gerakan apapun. Namun pada intensitas cahaya 1.67 lux dapat mendeteksi Bahasa Isyarat “Motor”. Agar pendekstian dapat maksimal, maka dibutuhkan intensitas cahaya yang lebih dari 77.5 lux.

Tabel 3.3 Hasil Pengujian dengan Bentuk Tangan dan Warna Kulit yang Berbeda.

No	Bahasa Isyarat	Bentuk Tangan	FPS	Akurasi
1.	Maaf	Gemuk, Kecil, Putih	7,52	100%
2.	Kapan	Kurus, Kecil, Coklat	6,81	97%
3.	Coba	Gemuk, Besar, Putih	6,81	100%
4.	Kereta	Gemuk, Kecil, Coklat	6,79	100%
5.	Mobil	Kurus, Besar, Coklat	6,67	98%

Tabel 3.3 merupakan hasil pengujian dengan bentuk tangan dan warna kulit yang berbeda. Menghasilkan nilai akurasi diatas 90% dengan fps tertinggi yaitu 7,44 dan rata-rata keseluruhan akurasi tertinggi adalah 94%.

4. PENUTUP

Bagian Penutup merupakan bagian akhir dari penelitian ini, sehingga pada bagian ini berisi kesimpulan dari pengujian yang telah dilakukan. Serta terdapat saran yang berguna sebagai acuan untuk penelitian selanjutnya.

4.1 Kesimpulan

Berdasarkan hasil pengujian yang telah dilakukan pada penelitian ini, maka dapat disimpulkan bahwa:

- 1) Sistem mampu mendeteksi bahasa isyarat melalui kamera dengan memanfaatkan metode YOLO, dan range FPS yang dihasilkan yaitu 7.52 – 8.01. Deteksi dengan menggunakan metode You Only Look Once (YOLO) memiliki hasil dari 20 gerakan Bahasa Isyarat hanya 1 gerakan bahasa isyarat “Benar” yang mengalami *error* dengan persentase keberhasilan 90%. Selain Bahasa Isyarat tersebut seluruh gerakan Bahasa Isyarat memiliki persentasi keberhasilan sebesar 100% dengan 10 kali coba deteksi tiap gerakan Bahasa Isyarat.
- 2) Penerapan metode YOLO pada sistem mampu mendeteksi bahasa isyarat jika intensitas cahaya lebih dari 1.67 lux, namun dengan intensitas cahaya lebih dari 77.5 membuat pendekstian bahasa isyarat menjadi lebih optimal. Hasil deteksi menggunakan 5 tangan yang berbeda, menghasilkan rata-rata akurasi tertinggi yaitu 94% pada kata “Kereta” dan rata-rata fps tertinggi yaitu 7,44 pada kata “Maaf” dan nilai persentase error keseluruhan adalah 0%.
- 3) *Output* dari sistem disajikan dalam bentuk teks dan suara, ketika mendekstibsebuah gerakan baru maka akan langsung memberikan *output* yang sesuai dengan gerakan yang tedeteksi. Jika tidak terdapat gerakan baru, maka hanya *output* berupa teks yang muncul, hal ini di atur agar tidak menimbulkan *spamming audio* dan mengakibatkan kebisingan.

4.2 Saran

Adapun saran yang dapat diberikan dari penelitian ini, antara lain:

- 1) Bahasa Isyarat “Benar” mendapatkan akurasi 90% disebabkan kemiripan gestur dengan bahasa isyarat “Toilet”. Disarankan penggunaan *webcam* dengan spesifikasi lebih tinggi dan menambahkan *dataset* untuk bahasa isyarat dengan kemiripan gestur agar dapat mengklasifikasikan dengan lebih optimal.
- 2) Menggunakan laptop dengan

- spesifikasi yang lebih tinggi agar dapat mencapai lebih dari 30 FPS.
- 3) Menggunakan peraga yang lebih banyak agar mendapatkan bentuk tangan yang lebih variatif untuk pembuatan *dataset* agar hasil yang dihasilkan lebih maksimal.
- 5. DAFTAR PUSTAKA**
- [1] Andrian, M. Y., Purwanto, D. and Mardiyanto, R. (2017) ‘Penerjemahan bahasa isyarat indonesia menggunakan kamera pada telepon genggam android’, *Jurnal Teknik ITS*, 6(1), p. 180.
 - [2] Hallan and Kauffman (1991) *Exceptional learners*, Allyn and Bacon. Boston.
 - [3] Hardman (1990) *Human Exceptionality*. third. Massachusetts: A Division of Simon & Schuster Inc.
 - [4] Hasibuan, I. H. A., Mulyana, A. and Brian, A. (2015) ‘Perancangan dan Implementasi Aplikasi Penerjemah Bahasa Isyarat Menjadi Suara Berbasis Kinect Menggunakan Metode Hidden Markov Model’, in *e-Proceeding of Engineering*, pp. 1–8.
 - [5] Hu, G. *et al.* (2019) ‘When Face Recognition Meets with Deep Learning: an Evaluation of Convolutional Neural Networks for Face Recognition’, in *Proceedings of the IEEE international conference on computer vision workshops*, pp. 142–150.
 - [6] Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012) ‘ImageNet Classification with Deep Convolutional Neural Networks.’ *Advances in Neural Information Processing Systems 25 (NIPS2012)*, in *NeurIPS Proceedings*, pp. 1–9.
 - [7] Maggiori, E. *et al.* (2016) ‘Convolutional neural networks for large-scale remote-sensing image classification’, *IEEE Transactions on Geoscience and Remote Sensing*, 55(2), pp. 645–657.
 - [8] Pigou, L. *et al.* (2015) ‘Sign Language Recognition Using Convolutional Neural Networks’, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8927(1), pp. 572–578. doi: 10.1007/978-3-319-16178-5.
 - [9] Pujiati, D. (2019) *Perbandingan Struktur Antara Sistem Isyarat Bahasa Indonesia (SIBI) Dengan Bahasa Isyarat Indonesia (BISINDO)*. Universitas Pendidikan Indonesia.
 - [10] Purwanto, H. (1998) *Ortopedagogik Umum*. Yogyakarta: IKIP Yogyakarta.
 - [11] Rachmawati, F. (2020) ‘Deteksi Jumlah Kendaraan di Jalur SSA Kota Bogor Menggunakan Algoritma Deep Learning YOLO’, in *PROSIDING LPPM UIKA BOGOR*.
 - [12] Redmon, J. and Farhadi, A. (2018) *Yolov3: An incremental improvement*, Cornell University. Available at: arx.