

## Pengembangan YOLO untuk Mengunci Sebuah Obyek Telapak Tangan Diantara Sejumlah Objek Seragam

Agustinus Rudatyo Himamunanto<sup>1\*</sup> Jean Pinter Son Zalukhu<sup>2</sup>  
 Informatika, Universitas Kristen Immanuel, Yogyakarta, Indonesia  
 \*e-mail *Corresponding Author*: rudatyo@ukrim.ac.id

### Abstract

*The hand is one part of the body that is often used in daily human activities. Nowadays, in line with the development of digital technology and computing technology, the role of hands has the potential to become even wider. Hands can potentially be used to model operational facilities or control input models for a device. Problems arise if more than one visual of a hand is captured by the camera. This has the potential to cause ambiguity due to the emergence of multiple hand gesture control input models. It is necessary to lock or tag the original hand control input model in such a way that the other hand can be ignored. In this research, the Google Media Pipe Hand (GMPH) framework is used to mark the hand area based on the input image and the YOLO framework will work to recognize the hand with a marker, which will lock the marked hand between the visuals of other hands. Based on test results involving 800 test data in the form of video data with visual images of one hand gesture or more than one hand gesture image, it is known that the results of the YOLO modification show success with an accuracy of 97.5%.*

*Keywords: Hand; Markers; Google Media Pipe Hand; YOLO Framework*

### Abstrak

Tangan merupakan salah satu anggota tubuh yang cukup sering dipergunakan dalam keseharian aktifitas manusia. Dewasa ini sesuai dengan perkembangan teknologi digital dan teknologi komputasi maka peran tangan berpotensi menjadi lebih luas lagi. Tangan berpotensi dipergunakan untuk model sarana operasional atau model input kendali terhadap suatu perangkat. Permasalahan muncul bila suatu visual tangan yang tertangkap oleh kamera berjumlah lebih dari satu. Hal ini berpotensi menimbulkan ambiguitas karena munculnya model input kendali gestur tangan ganda. Perlu dilakukan proses penguncian atau penandaan pada satu model input kendali tangan yang asli sedemikian rupa sehingga tangan yang lain dapat diabaikan. Pada penelitian ini, *framework Google Media Pipe Hand* (GMPH) dipergunakan untuk menandai area tangan berdasarkan citra input dan *framework YOLO* akan bekerja untuk mengenali tangan dengan marker, yang akan mengunci tangan bermarker diantara visual tangan yang lain. Berdasarkan hasil pengujian yang melibatkan 800 data uji berupa data video dengan visual citra satu gestur tangan maupun lebih dari satu citra gestur tangan, diketahui bahwa hasil modifikasi YOLO menunjukkan keberhasilan dengan akurasi 97,5%.

**Kata kunci:** *Tangan; Marker; Google Media Pipe Hand; Framework YOLO*

### 1. Pendahuluan

Tangan merupakan salah satu anggota tubuh yang cukup sering dipergunakan dalam keseharian aktifitas manusia [1]. Salah satu peran tangan bagi manusia adalah dipergunakan sebagai media komunikasi nonverbal, yaitu komunikasi tanpa kata-kata [2]. Tangan dalam hal ini akan berperan dalam menyatakan makna komunikasi melalui pola gestur-gesturnya.

Dewasa ini sesuai dengan perkembangan teknologi digital dan teknologi komputasi maka peran tangan berpotensi menjadi lebih luas lagi. Tangan berpotensi dipergunakan untuk sarana operasional langsung atau kendali langsung terhadap suatu perangkat. Semisal tangan yang dipergunakan sebagai: kendali perintah computer [3][4], kendali pesawat televisi [5], kendali level cahaya lampu [6], kendali gerak robot [7], kendali skala suara speaker [8], pengenalan bahasa isyarat [9], keamanan akses area [10] dan lain sebagainya.

Obyek tangan yang berperan sebagai kendali terlebih dahulu secara visual ditangkap menggunakan perangkat kamera. Visual tangan itu kemudian diproses dan dicoba dipahami

makna gesturnya untuk selanjutnya dilaksanakan perintah yang sudah berkolerasi dengan gesturnya [11]. Situasi unik muncul bila visual tangan yang tertangkap oleh kamera jumlahnya lebih dari satu. Hal ini berpotensi menimbulkan ambigu dan berpotensi membuat kendali perintah menjadi tidak akurat.

Penelitian ini mencoba mencari cara untuk mengunci hanya satu tangan sedemikian rupa sehingga tangan yang lain dapat diabaikan sekalipun obyek serupa atau obyek yang seragam jumlahnya cukup banyak. Hasil dari penelitian ini diharapkan memberikan kontribusi pada: upaya pengendalian operasional sistem yang berbasis pengenalan pola gestur tangan, upaya pengenalan bahasa isyarat baik pada personal berkebutuhan khusus maupun personal berusia lanjut dengan keterbatasan kemampuan verbal, upaya pengamanan akses melalui pengenalan kunci berbasis gestur tangan.

GMPH adalah *framework* populer yang bekerja untuk menemukan area tangan hingga pergelangan tangan [12]. YOLO adalah *framework* populer yang memiliki karakteristik kerja yaitu menandai obyek serupa atau obyek yang seragam yang muncul dalam satu frame [13]. YOLO perlu mendapat perlakuan khusus saat harus menandai atau mengunci hanya satu obyek diantara sejumlah obyek serupa atau sejumlah obyek yang seragam. Sesuai dengan peran dan pelatihannya, YOLO akan bekerja serta berupaya menemukan semua obyek yang memiliki kemiripan dengan data latihnya.

Pada sebagian kegiatan penelitian, YOLO dilatih dan dipergunakan untuk membantu menemukan suatu obyek khusus yang ingin diamati atau dikenai operasi secara lebih lanjut. Pada penelitian ini hasil kerja YOLO akan dibalik dalam upaya menemukan dan menandai hanya satu obyek kunci berupa satu gestur tangan diantara lebih dari satu visual obyek gestur tangan. Pertanyaan selanjutnya adalah: apakah modifikasi YOLO dapat dilakukan? Apakah modifikasi tersebut dapat mengunci atau menandai satu obyek visual tangan diantara sejumlah obyek visual yang seragam dalam satu frame? Seberapa jauh performa YOLO termodifikasi nanti dalam mengunci satu obyek visual gestur tangan?

## 2. Tinjauan Pustaka

Pada penelitian visi komputer bidang minat pengenalan gestur tangan, operasi deteksi gestur tangan [14] dan operasi segmentasi gestur tangan [15] merupakan operasi yang dipergunakan untuk menandai keberadaan suatu obyek visual tangan. Beberapa menunjukkan hasil deteksi yang cukup tepat bahkan mencapai akurasi diatas 95% [16] dan hasil segmentasi yang cukup bagus [7]. Namun terdapat beberapa keterlambatan proses terutama pada saat diperlukan penandaan atau penguncian visual obyek [17]. Fakta ini terkadang diperparah lagi dengan kondisi obyek yang secara visual bergerak yang menyebabkan keterlambatan proses penguncian menjadi jauh dari akurat [18].

Penelitian yang melibatkan pendekatan YOLO sebagai pemeran utama menunjukkan bahwa YOLO bekerja cukup bagus [19]. Namun begitu YOLO tetap dalam peran tugasnya untuk menemukan semua obyek serupa atau seragam yang muncul dari dalam frame yang sama [20].

Beberapa penelitian terkait pemrosesan data berupa visual gestur tangan yang melibatkan pendekatan YOLO pernah dilakukan. Tahun 2020, Shuai Yuan dkk [21] dalam artikel "Research Approach of Hand Gesture Recognition based on Improved YOLOV3 network and Bayes classifier", melibatkan YOLO serta metode PCA dan metode Bayes, memprediksi kategori isyarat tangan dengan mengurangi dimensi data dan meningkatkan akurasi klasifikasi. Penelitian ini mengolah data visual hanya satu gestur tangan. Abdullah Mujahid dkk [22] pada tahun 2021 dalam publikasinya "Real-Time Hand Gesture Recognition Based on Deep Learning" YOLOv3 Model" memaparkan keterlibatan YOLO dan Pascal VOC dalam mengekstraksi fitur dari tangan dan mengenali gerakan tangan dari model berbasis YOLOv3 yang menghasilkan akurasi, presisi, perolehan, dan skor F-1 masing-masing sebesar 97,68, 94,88, 98,66, dan 96,70% (sitasi).

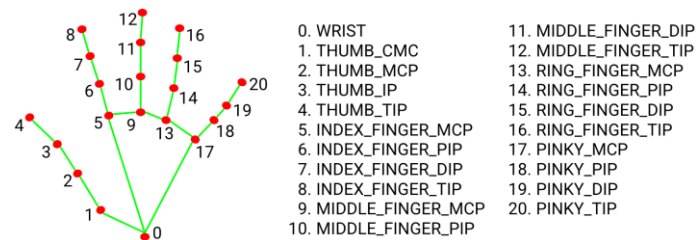
Pada tahun 2022, Tsui-Ping Chang dkk. [23] dalam artikelnya yang berjudul "Deep Learning Model for Dynamic Hand Gesture Recognition for Natural Human- Machine Interface on End Devices", melibatkan YOLO dan menyediakan metode (yaitu ModelOps) untuk menerapkan model terlatih pada perangkat akhir, dengan membangun arsitektur komputasi edge menggunakan Kubernetes. Penelitian ini memberikan pengembangan komponen pengenalan gerakan berdasarkan input visual satu tangan secara real-time (sitasi). Tahun 2023, Sugiarto Wibowo [24] dalam artikel "Hand Symbol Classification for Human-Computer

Interaction Using the Fifth Version of YOLO Object Detection”, dengan melibatkan YOLO meneliti peran gestur tangan untuk media interaksi komputer, namun hanya berdasarkan input visual satu tangan saja.

Pada penelitian ini akan dilakukan modifikasi pada YOLO sehingga diharapkan dapat menemukan metode yang membuat YOLO mampu mengunci satu obyek visual gestur tangan dan mengabaikan sejumlah obyek visual tangan lain yang diketemukan dalam satu frame yang sama.

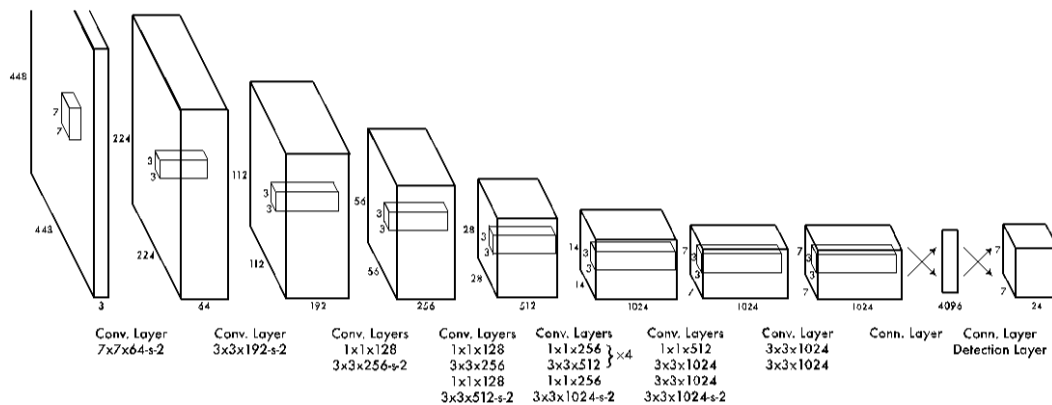
### 3. Usulan Metode

GMPH adalah framework yang dikembangkan oleh *Google* dan diperkenalkan pertama kali oleh Fan Zang dkk pada Fourth Workshop on Computer Vision for AR/VR tahun 2020. Frame ini bekerja menemukan area tangan serta pergelangan dilanjutkan dengan membangun konstruksi jaringan persendian tangan berdasarkan 21 titik konstruksi permukaan tangan [12].



Gambar 1. 21 titik penanda konstruksi tangan (diambil dari *Google MediaPipe Hands*)

Pada sisi lain, YOLO diperkenalkan pertama kali oleh Joseph Redmon [13] pada konferensi internasional visi komputer dan pengenalan pola pada tahun 2016 (CVPR 2016). Algoritma YOLO menggunakan *Convolution Neural Network* (CNN) untuk mendeteksi objek secara real-time. Seperti namanya, algoritma ini hanya membutuhkan satu propagasi maju melalui jaringan saraf untuk mendeteksi objek. Ini berarti bahwa prediksi di seluruh gambar dilakukan dalam satu kali algoritma. CNN digunakan untuk memprediksi berbagai probabilitas kelas dan kotak pembatas secara bersamaan. Garis besar cara kerja YOLO ditunjukkan dalam diagram pada Gambar 2.



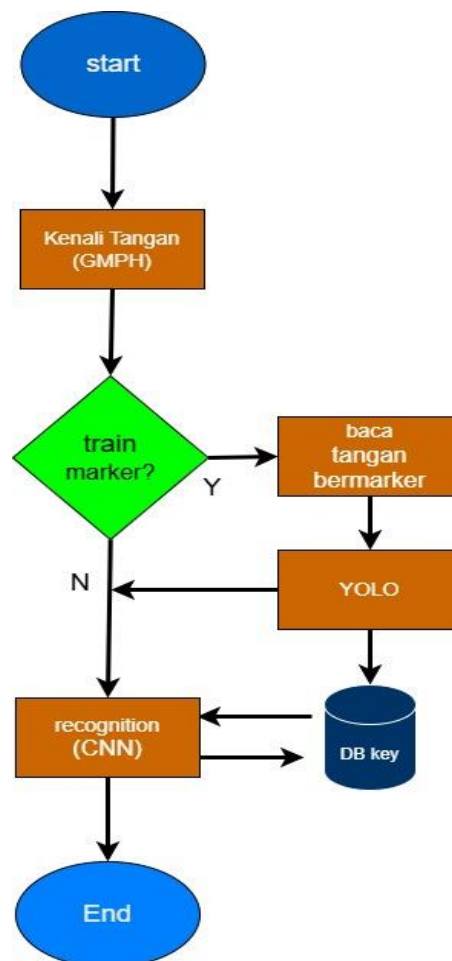
Gambar 2. Diagram alir proses deteksi obyek pada YOLO [25]

Algoritma YOLO menerapkan neural network pada sebuah gambar dan membagi input gambar menjadi sejumlah grid, lalu memprediksi bounding box serta probabilitas untuk masing-masing grid. Untuk mendapatkan bounding box, dilakukan proses konvolusi dari input gambar, sehingga hasil akhirnya akan diperoleh ukuran bounding box sebesar  $S \times S \times (B * 5 + C)$  dengan B adalah banyaknya bounding box dalam 1 grid, dan C adalah banyaknya class yang dapat dideteksi. Nilai B dikalikan dengan 5 karena sebuah bounding box memiliki 5 nilai yang perlu disimpan di antaranya: koordinat x, koordinat y, width (lebar), height (tinggi), confidence score (nilai probabilitas bounding box yang mendeteksi sebuah objek). Semua atribut pada bounding

box melalui proses normalisasi, sehingga akan menghasilkan nilai antara 0 dan 1. Koordinat x dan y akan dinormalisasi dengan menyesuaikan titik kiri atas dari grid yang bersangkutan. Width (lebar) dan height (tinggi) juga akan dinormalisasi menyesuaikan dengan ukuran gambar.

Berdasarkan karakteristik GHMP dan YOLO maka diharapkan area tangan dapat ditemukan dengan cepat dan obyek tangan bermarker dapat segera ditandai. Penelitian ini dilakukan juga dalam kerangka mempelajari potesi pemaduan framework GMPH-YOLO untuk potensi lain terkait implementasi berdasarkan kendali tangan. Karakteristik inilah yang selanjutnya dimodifikasi untuk mengarahkan YOLO pada upaya hanya menandai satu obyek saja.

Pada penelitian ini, untuk keperluan penguncian obyek, akan dicoba dilakukan pengembangan metode YOLO agar hanya satu obyek yang ditandai oleh YOLO diantara sejumlah obyek yang seragam. Gambaran diagram alir yang direncanakan menjadi pengembangan YOLO ditunjukkan pada Gambar 03.



Gambar 3. Diagram alir pengembangan GMPH-YOLO

Data input yang dipergunakan didalam penelitian ini berupa video gerakan gestur tangan dengan durasi yang cukup singkat yaitu antara 3 detik hingga 5 detik dalam satuan frame sebesar 25-30 fps. Data input video secara visual akan dipersiapkan atau dipilih dalam beberapa tingkat pencahayaan untuk menguji performa metode usulan yaitu YOLO termodifikasi. Beberapa video data input berisi visualisasi gerakan satu obyek gestur tangan dan selebihnya data input berisi visualisasi gerakan lebih dari satu obyek gestur tangan.

Adapun contoh beberapa video yang akan dipergunakan sebagai data input terbagi menjadi 2 kelompok yaitu kelompok video berisi gerakan satu tangan saja dan kelompok video berisi gerakan lebih dari satu tangan, seperti ditunjukkan dalam Gambar 04.



Gambar 4. Contoh input data visual 1 tangan dan banyak tangan

Total jumlah data input video gestur tangan yang akan disediakan adalah sebanyak 3000 data baik berupa dataset primer maupun dataset sekunder. Dataset sekunder yang akan dipergunakan diambil dari dataset gestur tangan HaGRID [26], dataset gestur tangan 20bn Jester Hand dataset [27] dan dataset IPN Hand Dataset [28].

HaGRID dataset adalah kepanjangan dari HAnd Gesture Recognition Image Dataset. Ukuran HaGRID adalah 716GB dan kumpulan data berisi 552.992 gambar FullHD (1920 x 1080) RGB yang dibagi menjadi 18 kelas gerakan. Selain itu, beberapa gambar memiliki kelas `no_gesture` jika ada tangan kedua yang bebas di dalam bingkai. Kelas tambahan ini berisi 123.589 sampel. Data dibagi menjadi 92% pelatihan, dan 8% pengujian ditetapkan berdasarkan `user_id` subjek, dengan 509.323 gambar untuk pelatihan dan 43.669 gambar untuk pengujian. 20bn-Jester Hand dataset memiliki jumlah data latih sebesar 71724 data dan data sampel validasi sebesar 9036 data. Ukuran gambar 100x100 dengan 36 frame. IPN Hand Dataset berisi lebih dari 4.000 contoh gerakan dan 800.000 bingkai dari 50 subjek, 13 kelas gerakan statis dan dinamis. Dataset lain yang juga dipergunakan untuk proses train adalah

Dataset sekunder adalah data input berupa video gerakan gestur tangan yang dibangun secara mandiri pada saat penelitian. Dataset primer sendiri akan disediakan hingga sejumlah 500 model data input berupa video gerakan tangan.

Proses analisis data hasil pengujian dalam penelitian ini akan dilakukan hanya apabila semua proses pengujian sudah dinyatakan selesai dan menghasilkan luaran informasi. Proses pengujian akan dilakukan berdasarkan metode *cross validation* dengan mengambil standar 3 *fold cross validation*. *Cross validation* adalah suatu metode tambahan dari teknik data mining yang bertujuan untuk memperoleh hasil akurasi yang maksimal. Metode ini sering juga disebut dengan *k-fold cross validation* dimana percobaan sebanyak *k* kali untuk satu model dengan parameter yang sama [29].

#### 4. Persiapan Awal dan Inisiasi

Metode yang digunakan dalam model pengunci tangan adalah teknik deteksi dan klasifikasi. Proses akan melibatkan *MediPipe Hands* untuk mendeteksi lokasi pergelangan tangan, lalu menggunakan *YOLOv8* untuk melakukan klasifikasi ada tidaknya marker pada pergelangan tersebut.

*MediPipe Hands* digunakan untuk mendeteksi landmark-landmark tangan dalam video. Setelah mendapatkan landmark, diambil posisi min dan max dari seluruh landmark pada tangan tersebut untuk membuat kotak bounding dengan koordinat  $x_1, y_1, x_2, y_2$  yang mencakup tangan dan pergelangan tangan.

```
x1_hand = int(min(hand_landmarks.landmark[i].x * frame.shape[1] for i in range(21)))
y1_hand = int(min(hand_landmarks.landmark[i].y * frame.shape[0] for i in range(21)))
x2_hand = int(max(hand_landmarks.landmark[i].x * frame.shape[1] for i in range(21)))
y2_hand = int(max(hand_landmarks.landmark[i].y * frame.shape[0] for i in range(21)))
```

Setelah mendapatkan posisi kotak bounding, dilakukan segmentasi untuk mendapatkan ROI tangan yang akan digunakan untuk klasifikasi. Pemotongan ROI ini dapat memanfaatkan fungsi bawaan `opencv` yang memungkinkan pemotongan gambar dengan memberikan koordinat  $x_1, y_1, x_2, y_2$  yang ingin dipotong dari frame asli. Skrip dan contoh gambar segmentasi dapat dilihat di bawah.

```
frame[y1_hand:y2_hand, x1_hand:x2_hand]
```

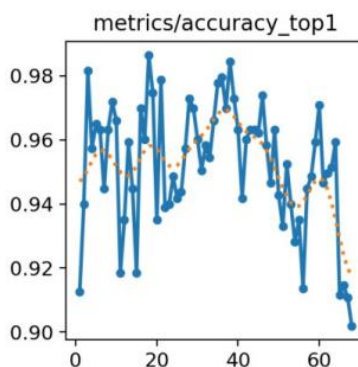
Berikut dibawah adalah contoh data hasil segmentasi yang dipergunakan untuk pelatihan.



Gambar 5. Contoh hasil segmentasi: marker putih, marker hitam dan tanpa marker

Terdapat 1000 data tanpa marker (tangan normal) dan 1000 data dengan marker (tangan dengan marker di pergelangan). Data tersebut kemudian dilabeli secara manual dengan dua kelas yang itu band, yaitu tangan dengan marker dan no\_band, untuk tangan tanpa marker. Seluruh data ini lalu dibagi dalam 70% data latih, 20% data validasi, 10% data uji (3 fold cross validation).

YOLO digunakan untuk melatih model klasifikasi guna membedakan antara tangan dengan dan tanpa marker di pergelangan. Data pelatihan yang telah dilabeli digunakan untuk melatih model. Training data menggunakan 65 epochs dan menghasilkan akurasi di atas 90%, seperti ditunjukkan pada Gambar 06. Berikut grafik hasil pengecekan akurasinya.



Gambar 6. Grafik pengecekan akurasi paska tahap pelatihan

Setelah model YOLO selesai, semua script disatukan di mana video diambil, setiap frame diproses dengan mendeteksi bagian tubuh pergelangan tangan menggunakan *MediaPipe Hands*, hasil pergelangan tangan dicrop, lalu menjadi inputan pada *YOLOv8* untuk diklasifikasi apakah memiliki marker atau tidak. Hanya tangan dengan marker yang akan diambil dan diproses untuk kebutuhan selanjutnya.

### 5. Hasil dan Pembahasan

Pengujian dilakukan dengan melibatkan proses train inisiasi telapak tangan kunci dengan penanda, penyimpanan kedalam basis data, dilanjutkan dengan train telapak tangan tanpa penanda. Pengujian dilakukan dengan melibatkan dataset primer dan sekunder. Pengujian dengan satu tangan dipergunakan untuk memastikan bahwa sistem tetap dapat memproses namun tidak menemukan obyek telapak tangan atau tidak menunjukkan penanda tangan. Pengujian selanjutnya dilakukan dengan satu tangan dengan marker yang dipergunakan untuk memastikan bahwa sistem tetap dapat memproses dan menemukan obyek telapak tangan. Pengujian selanjutnya dilakukan dengan banyak tangan namun salah satu obyek dengan marker yang dipergunakan untuk memastikan bahwa sistem tetap dapat memproses dan menemukan obyek telapak tangan. Contoh pengujian ditunjukkan pada Gambar 7.





Gambar 7. Input data visual untuk satu tangan tanpa marker, satu tangan dengan marker, banyak tangan dengan satu tangan ber-marker

Model yang dibentuk dengan menggabungkan GMPH untuk deteksi area tangan dan YOLO untuk klasifikasi bekerja dengan baik dalam menemukan dan mengunci tangan menggunakan marker dengan akurasi di atas 90% dalam beberapa kondisi seperti pencahayaan yang rendah dan bergerak dengan cepat. Adapun syarat yang diperlukan agar model dapat mengunci tangan dengan baik, yaitu tangan harus berada pada jarak 1 – 1.1 meter untuk mendapatkan hasil yang terbaik dan marker sebaiknya tidak terhalangi objek.

Berikut dibawah (Tabel 1.) adalah tabel hasil pengujian selengkapnya yang dilakukan sebanyak 800 kali dengan melibatkan 800 data beragam kondisi.

Pengujian	jumlah	Hasil benar
Satu tangan nonmarker	200	200
Satu Tangan ber-marker	200	200
Banyak tangan nonmarker	200	200
Banyak tangan + Satu tangan bermarker	200	180

Berdasarkan hasil pengujian yang melibatkan 800 data uji berupa data video dengan visual citra satu tangan maupun lebih dari satu citra tangan, diketahui bahwa hasil modifikasi YOLO menunjukkan keberhasilan dengan akurasi 97,5%.

Pada saat melakukan pengujian sistem, ditemukan juga beberapa catatan tambahan yaitu:

Situasi	Hasil
Ada tangan dengan marker	Model dapat menemukan tangan marker dengan akurasi 98%
Banyak tangan tanpa marker dimana terdapat satu tangan dengan marker	Terdapat dua hasil: <ul style="list-style-type: none"> <li>Model dapat menemukan tangan dengan akurasi 99% dengan catatan tangan yang menggunakan marker harus terlihat terlebih</li> </ul>

Situasi	Hasil
	dahulu oleh kamera <ul style="list-style-type: none"> <li>Model tidak dapat menemukan tangan sama sekali. Kedua hasil ini diakibatkan oleh sifat dari library MediaPipe yang hanya akan membaca dua tangan dan tidak lebih dalam satu frame. Sehingga jika kedua tangan yang terbaca pertama kali tidak terdapat marker, sehingga sekalipun tangan yang menggunakan marker muncul, deteksi dan pengunci tangan tidak akan aktif</li> </ul>
Tangan bergerak dengan cepat	Model masih dapat menemukan tangan yang diayun-ayun dengan cepat dengan akurasi 99%
Pencahayaan Normal (Pencahayaan Ruangan)	Model dapat menemukan dan mengunci tangan dengan akurasi 99%
Pencahayaan Rendah (Bukan Gelap Total)	Model dapat menemukan dan mengunci tangan dalam ruangan dengan keadaan pencahayaan yang rendah dengan akurasi 97%
Kondisi marker tertutup	Saat marker tertutup, model sudah tidak akan mengunci tangan. Namun, saat tangan bergerak dengan cukup cepat melewati objek penghalang, maka model masih dapat membaca tangan.
Jarak tengah jauh dari kamera	Jarak terbaik adalah 1 – 1.9 meter. Setelah di atas 1.9 meter hasil deteksi sudah tidak direkomendasi

YOLO secara teori bekerja untuk menemukan obyek yang dilatih sebarang banyaknya obyek tersebut dapat diketemukan oleh YOLO dalam satu frame citra inputnya [25]. Penelitian ini bukan hanya memanfaatkan YOLO untuk menemukan banyak obyek tangan [30] atau hanya menemukan satu tangan [31] namun lebih daripada itu berupaya mengembangkan YOLO untuk menjawab permasalahan yang muncul yaitu menandai/mengunci hanya satu tangan diantara banyak obyek tangan yang diketemukan. Hal ini terasa urgent mempertimbangkan pada implementasi kendali perangkat berbasis pengenalan tangan, satu tangan pengendali berpotensi terganggu dengan kemunculan obyek tangan yang lain.

Berdasarkan penelitian dan pengujian diketahui bahwa pengembangan GMPH-YOLO menunjukkan keberhasilan yang cukup baik dan signifikan, namun begitu menyisakan beberapa temuan yang muncul pada saat pengujian dilakukan seperti ditunjukkan pada Tabel 02. Bahkan muncul pertanyaan baru: apakah permasalahan ini dapat didekati dengan konsep YOLO bertingkat (*Stacked-YOLO*)? Bila jawaban ya bisa, lalu seberapa baik *GMPH-YOLO* dibandingkan dengan *Stacked YOLO*? Mengingat urgensinya permasalahan ini, tentu menjadi tantangan lebih lanjut yang menarik untuk diteliti.

### 6. Simpulan

Berdasarkan hasil pengujian yang melibatkan 800 data uji berupa data video dengan visual citra satu gestur tangan maupun lebih dari satu citra gestur tangan, diketahui bahwa hasil modifikasi GMPH-YOLO menunjukkan keberhasilan dengan akurasi 97,5%.

Penelitian ini menyisakan ruang permasalahan berikutnya mempertimbangkan bahwa obyek telapak tangan dapat dikenali dengan melibatkan penanda atau marker, maka penelitian selanjutnya akan diarahkan pada upaya pengenalan obyek telapak tangan utama namun tidak melibatkan penanda atau marker.

### Daftar Referensi

[1] A. P. Dhote and V. R. Parihar, "Gesture Recognition," no. Icici, pp. 512–516, 2017.  
 [2] N. Patel and S. J. He, "A Survey on Hand Gesture Recognition Techniques , Methods and Tools," vol. 6, no. 6, 2018.  
 [3] P. Xu, "A Real-time Hand Gesture Recognition and Human-Computer Interaction System," *Int. Conf. Adv. Comput. 2016 IEEE 6th Int. Conf. Adv. Comput. 2016*, pp. 1–8, 2016.  
 [4] M. C. Roh, D. Kang, S. Huh, and S. W. Lee, "A virtual mouse interface with a two-layered



- Bayesian network,” *Multimed. Tools Appl.*, vol. 76, no. 2, pp. 1615–1638, 2017, doi: 10.1007/s11042-015-3144-x.
- [5] U.V. Solanki, & N.H. Desai, "Hand gesture based remote control for home appliances: Handmote. In *2011 World Congress on Information and Communication Technologies*, IEEE, pp. 419-423, 2021.
- [6] A. Chaudhary and J. L. Raheja, "Light invariant real-time robust hand gesture recognition,” *Optik (Stuttg.)*, vol. 159, pp. 283–294, 2018, doi: 10.1016/j.ijleo.2017.11.158.
- [7] A. Chaudhary, *Robust Hand Gesture Recognition for Robotic Hand Control*. Singapore: Springer, 2918
- [8] W.P. Fang, An intelligent hand gesture extraction and recognition system for home care application. In *2012 sixth international conference on genetic and evolutionary computing*, IEEE, pp. 457-459, 2012.
- [9] C. N. Nyaga and R. D. Wario, "Sign Language Gesture Recognition through Computer Vision,” *2018 IST-Africa Week Conf.*, pp. 1-8, 2018.
- [10] J. Xu and X. Zhang, "A High-Security and Smart Interaction System Based on Hand Gesture Recognition for Internet of Things,” *Security and Communication Networks*, vol. 2018, no. 1, 2018.
- [11] J. Wang and S. Payandeh, "Hand Motion and Posture Recognition in a Network of Calibrated Cameras,” *Adv. Multimed.*, vol. 2017, pp. 1–25, 2017, doi: 10.1155/2017/2162078.
- [12] F. Zhang *et al.*, "MediaPipe Hands: On-device Real-time Hand Tracking,” 2020, [Online]. Available: <http://arxiv.org/abs/2006.10214>
- [13] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 779–788, 2016, doi: 10.1109/CVPR.2016.91.
- [14] T. Grzejszczak, M. Kawulok, and A. Galuszka, "Hand landmarks detection and localization in color images,” *Multimed. Tools Appl.*, vol. 75, no. 23, pp. 16363–16387, 2016, doi: 10.1007/s11042-015-2934-5.
- [15] J. Lin, X. Ruan, N. Yu, and J. Cai, "Multi-cue based moving hand segmentation for gesture recognition,” *Autom. Control Comput. Sci.*, vol. 51, no. 3, pp. 193–203, 2017, doi: 10.3103/s0146411617030063.
- [16] A. Tagliasacchi, M. Schröder, and A. Tkach, "Robust Articulated-ICP for Real-Time Hand Tracking,” *Eurographics (Wiley)*, vol. 34, no. 5, pp. 101-114, 2015, doi: 10.1111/cgf.12700.
- [17] A. I. Maqueda, C. R. Del-Blanco, F. Jaureguizar, and N. García, "Human-computer interaction based on visual hand-gesture recognition using volumetric spatiograms of local binary patterns,” *Comput. Vis. Image Underst.*, vol. 141, pp. 126–137, 2015, doi: 10.1016/j.cviu.2015.07.009.
- [18] W. H. Chun and H. Tobias, "Real-time Hand Interaction for Augmented Reality on Mobile Phones,” *Proceedings of the 2013 international conference on Intelligent user interfaces*, pp. 307–314, 2013.
- [19] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO Series in 2021,” pp. 1–7, 2021.
- [20] X. Long *et al.*, "PP-YOLO: An Effective and Efficient Implementation of Object Detector,” 2020.
- [21] B. Ye, S. Jin, B. Li, S. Yan, and D. Zhang, "applied sciences Dual Histogram Equalization Algorithm Based on Adaptive Image Correction,” *Applied Sciences*, vol. 13, no. 19, p.10649. 2023.
- [22] A. Mujahid *et al.*, "Real-time hand gesture recognition based on deep learning YOLOv3 model,” *Appl. Sci.*, vol. 11, no. 9, 2021, p. 4164. Doi: 10.3390/app11094164.
- [23] T.-P. Chang, H.-M. Chen, S.-Y. Chen, and W.-C. Lin, "Deep Learning Model for Dynamic Hand Gesture Recognition for Natural Human-Machine Interface on End Devices,” *Int. J. Inf. Syst. Model. Des.*, vol. 13, no. 10, pp. 1–23, 2022, doi: 10.4018/ijismd.306636.
- [24] S. Wibowo and I. Sugiarto, "Hand symbol classification for human-computer interaction using the fifth version of yolo object detection,” *CommIT J.*, vol. 17, no. 1, pp. 43–50, 2023, doi: 10.21512/commit.v17i1.8520.
- [25] A. Farhadi, & J. Redmon, Yolov3: An incremental improvement. In *Computer vision and pattern recognition*, Vol. 1804, pp. 1-6. Berlin/Heidelberg, Germany: Springer, 2018
- [26] A. Kapitanov, K. Kvanchiani, A. Nagaev, R. Kraynov, & A. Makhliarchuk, HaGRID--HAnd

- Gesture Recognition Image Dataset. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 4572-4581, 2024
- [27] J. Materzynska, G. Berger, I. Bax, and R. Memisevic, "The jester dataset: A large-scale video dataset of human gestures," *Proc. - 2019 Int. Conf. Comput. Vis. Work. ICCVW 2019*, pp. 2874–2882, 2019, doi: 10.1109/ICCVW.2019.00349.
- [28] R. Vezzani, "Proceedings of ICPR 2020 25th International Conference on Pattern Recognition : Milan, 10-15 January 2021," no. ii, pp. 4340–4347, 2021.
- [29] D. Berrar, "Cross-validation," *Encycl. Bioinforma. Comput. Biol. ABC Bioinforma.*, vol. 1–3, no. January 2018, pp. 542–545, 2018, doi: 10.1016/B978-0-12-809633-8.20349-X.
- [30] C. Dewi, H. Dwi Purnomo, B. Kristanto, D. Hartomo, S. Zaiton, and M. Hashim, "Utilizing the YOLOv8 Model for Accurate Hand Gesture Recognition with Complex Background," 2017, [Online]. Available: <http://dx.doi.org/10.1016/j.cviu.2017.00.000><http://dx.doi.org/10.1016/j.cviu.2017.00.000>[1077-3142/](http://dx.doi.org/10.1016/j.cviu.2017.00.000)
- [31] B. Alsharif, E. Alalwany, and M. Ilyas, "Transfer learning with YOLOv8 for real-time recognition system of American Sign Language Alphabet," *Franklin Open*, vol. 8, no. September, p. 100165, 2024, doi: 10.1016/j.fraope.2024.100165.