

Analisa Kemampuan Algoritma YOLOv8 Dalam Deteksi Objek Manusia Dengan Metode Modifikasi Arsitektur

Aris Setiyadi¹, Ema Utami², Dhani Ariatmanto³
^{1,2,3}Univeristas Amikom Yogyakarta, Indonesia
e-mail: arisyunior86@gmail.com

Abstract

Object detection is a skill that can be taught to a machine with the help of a camera sensor to capture a digital image. By using the YOLO algorithm we can teach machines to detect, for example, humans. Much research in object detection has been carried out previously using different algorithms and methods and also on different objects and images. In this research, a method was carried out to modify the architecture of YOLOv8 in the head section to be used to detect human objects in grayscale images. The training process was carried out 4 times using the default architecture, Model 1, 2 and 3 architecture. With the default model results, the mAP value was 76, Model 1 had an mAP value of 66, model 2 had an mAP value of 81 and model 3 produced an mAP value of 80. From the research carried out modifications The YOLOv8 architecture in the head section can influence the training results and produce a better model than the default architecture which only produces an mAP value of 76. The best results were obtained in model 2 with layers used of 40x40x512xW resulting in a model with an mAP value of up to 81.

Keywords: Deep learning, Yolov8, object detection, YOLO architecture

Abstrak

Deteksi objek merupakan kemampuan yang dapat diajarkan pada sebuah mesin dengan bantuan sensor kamera untuk menangkap sebuah citra digital. Dengan menggunakan Algoritma YOLO kita bisa mengajarkan mesin untuk mendeteksi misalkan manusia. Penelitian dalam deteksi objek sudah banyak dilakukan sebelumnya menggunakan algoritma dan metode yang berbeda dan juga pada objek dan citra yang berbeda. Pada penelitian ini dilakukan metode memodifikasi arsitektur pada YOLOv8 pada bagian head untuk digunakan medeteksi objek manusia dalam gambar yang berbentuk grayscale. Proses training dilakukan sebanyak 4 kali menggunakan arsitektur default, Arsitektur model 1, 2 dan 3. Dengan hasil model default nilai mAP 76, Model 1 nilai mAP 66, model 2 nilai mAP 81 dan model 3 menghasilkan nilai mAP 80. Dari penelitian yang dilakukan modifikasi arsitektur YOLOv8 pada bagian head dapat mempengaruhi hasil training dan menghasilkan model yang lebih baik dari arsitektur default yang hanya menghasilkan nilai mAP 76. Hail terbaik didapatkan pada model 2 dengan layer yang digunakan 40x40x512xW menghasilkan model dengan nilai nilai mAP mencapai 81.

Kata Kunci: Deep learning, Yolov8, Detection objek, Arsitektur YOLO

1. PENDAHULUAN

Salah satu kecerdasan manusia adalah kemampuan dalam mengenali suatu objek yang ada disekitarnya misalkan objek manusia. Manusia dapat mengenali sebuah objek, dengan menggunakan mata sebagai indra pengelihatian untuk menangkap sebuah citra objek yang kemudian akan direkam dan disimpan dalam memori otak. Perkembangan teknologi saat ini menunjang sebuah mesin untuk dapat belajar seperti manusia dalam mengenali sebuah objek. Mesin membutuhkan kecerdasan buatan untuk dapat mengenali dan mengklasifikasikan

sebuah objek citra digital. Deep learning merupakan bagian dalam ilmu AI (artificial intelligence) yang merupakan penerapan transformasi abstraksi nonlinier dan lanjutan kedalam basis data [1]. Salah satu algoritma yang sering digunakan dalam bidang deteksi objek adalah YOLO (You Only Look One) yang masuk pada kategori Algoritma CNN yang cara kerjanya meniru cara kerja otak manusia. Selain YOLO ada juga algoritma Single Shot Detector (SSD) RetinaFace and (Faster Recurrent Convolutional Neural Network) Faster R-CNN [2].

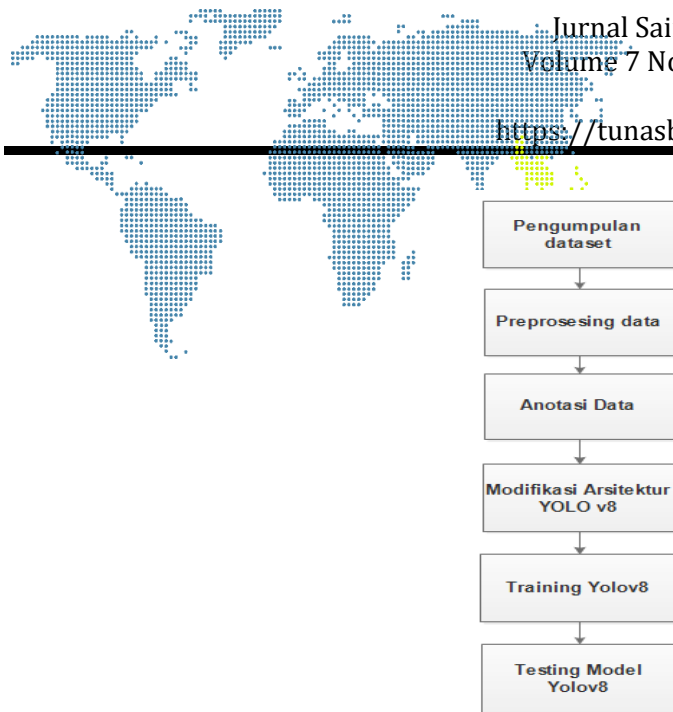
Penelitian dalam bidang detection objek manusia sudah banyak dilakukan dengan menggunakan berbagai metode yang ada. Chien-Yao Wang dkk melakukan penelitian deteksi objek menggunakan YOLOv7 dengan hasil akurasi mAP 56% dan kecepatan 30 fps [3]. Amelia dkk (2019) melakukan penelitian deteksi manusia menggunakan MobilNet untuk memproses segmentasi bentuk tubuh hasil yang diperoleh dalam penelitian untuk deteksi manusia mendapatkan akurasi 92,33% akurasi untuk bentuk tubuh mencapai 82% kemudian untuk keakurasian dalam menentukan ukuran nyata baju mencapai 96,06% dengan kondisi jarak pengambilan citra 2 meter, dalam penelitian ini tidak disebutkan berapa kecepatan deteksi objeknya [4]. Wahyu dkk (2021) melakukan penelitian deteksi manusia dengan membandingkan beberapa algoritma MobileNetv1, mobileNetv2 dan SSD inception V1 dengan optimasi jetson TX2 dan nano. dengan hasil MobileNetV2 memiliki akurasi tinggi dibandingkan yang lainnya mencapai PR 0.94 Recall 0.99 dan F1 0.98 [5]. Swastika, dkk (2019) melakukan penelitian deteksi manusia menggunakan YOLO dengan tingkat akurasi 71% pada video1 dan pada video berikutnya akurasi mencapai 73%, dalam penelitian ini tidak disebutkan berapa kecepatan yang didapatkan dalam deteksi objek manusia [6]. Khairunnas, dkk (2021) melakukan penelitian deteksi objek manusia menggunakan YOLOv4 dengan hasil nilai mAP 87,03%. [7]. Kristo dkk (2020) melakukan penelitian deteksi objek manusia dan bukan manusia pada citra sensor panas dengan menggunakan algoritma YOLOv3 dengan dataset berupa citra RGB dari COCO dan Citra sensor panas hasilnya dengan menggunakan citra sensor panas dapat mencapai nilai mAP 97,98% dibandingkan menggunakan dataset COCO RGB hanya mampu mencapai nilai AP 19% untuk kelas manusia [8]. Kurniawan, dkk (2022) melakukan penelitian deteksi objek manusia dengan membandingkan kemampuan YOLOv3 – YOLOv4 dengan optimasi TensorRT AGX dan NX, penelitian dilakukan pada citra RGB dengan hasil Optimasi AGX menghasilkan kecepatan yang lebih dibandingkan NX dan YOLOv4 menghasilkan akurasi yang lebih baik dibandingkan versi sebelumnya [9]. M. Musser (2020) menggunakan Bayesian statistics dan VGG mampu mendapat nilai akurasi 85% [10]. Harahap (2020) melakukan penelitian deteksi objek manusia menggunakan metode thinning berdasarkan nilai local maxima, hasil penelitian yang dilakukan mendapatkan nilai akurasi 80% pada deteksi wajah [11]. M Arfan dkk (2021) melakukan penelitian pengenalan aktivitas yang dilakukan oleh manusia dengan menggunakan convolution Neural network metode yang digunakan YOLO dan SORT tingkat akurasi yang didapatkan 74% pada gambar dan 80% pada video [12].

Berdasarkan uraian diatas telah banyak dilakukan penelitian deteksi objek manusia yang pada dilakukan pada gambar RGB maupun dalam bentuk gambar thermal dengan metode yang berbeda baik menggunakan optimasi pada algoritma maupun tidak dan hasil yang disampaikan dengan nilai akurasi yang baik. Bagaimana dengan kemampuan algoritma YOLOv8 untuk deteksi objek manusia dalam citra *grayscale* seperti halnya gambar yang dihasilkan mesin CCTV pada malam hari yang dengan bantuan sensor IR menghasilkan gambar *grayscale* dan juga bagaimana optimasi kemampuan algoritma YOLOv8 dari sisi akurasi apakah masih bisa dilakukan optimasi. Maka peneliti akan melakukan penelitian deteksi objek manusia dalam citra *grayscale* menggunakan algoritma YOLOv8 dengan modifikasi arsitektur YOLOv8 dan untuk meningkatkan akurasi deteksi objek.

2. METODOLOGI PENELITIAN

Metodologi penelitian yang digunakan dalam artikel jurnal ini adalah penelitian literatur. Penelitian literatur adalah jenis penelitian yang dilakukan dengan mengumpulkan dan menganalisis berbagai sumber literatur terkait topik penelitian, seperti artikel jurnal, buku, dan publikasi ilmiah lainnya (Ridwan dkk., 2021) [14]. Penelitian ini berfokus pada deteksi objek manusia pada citra *grayscale* dengan menggunakan algoritma You Only Look Once (YOLO) , YOLO merupakan algoritma yang masuk pada kategori deeplearning yang diusulkan oleh Josep Remond tahun 2015. Algoritma YOLO menggunakan Neural network dan membagi sebuah inputan citra dalam sebuah grid, lalu digunakan untuk peneganalan citra dan mendeteksi bounding box dan probabilitasnya. Pada penelitian ini menggunakan YOLO versi terbaru yaitu YOLOv8.

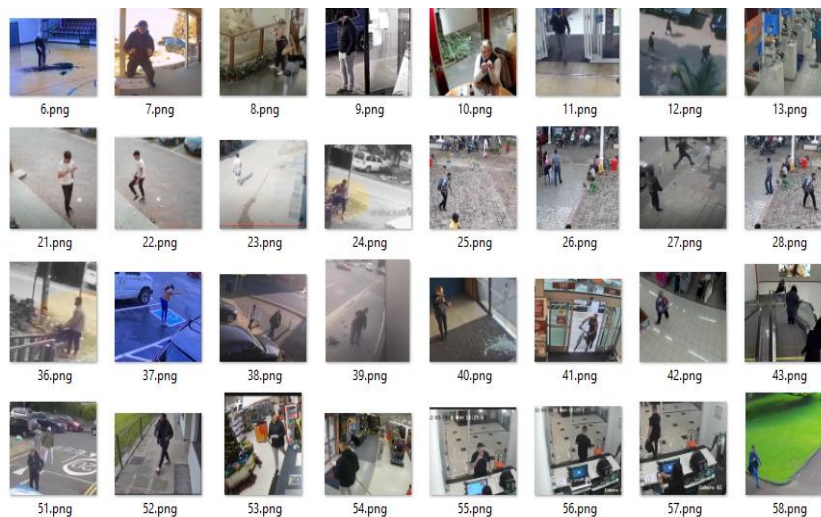
Algoritma YOLO berdasarkan ujicoba dapat beroperasi pada kecepatan 45 fps (frame per second) dengan menggunakan kartu grafis Titan X Dengan penelitian tersebut algoritma YOLO sangat baik digunakan pada proyek yang memiliki sifat realtime seperti pada deteksi objek pada image ataupun video. Seiring banyaknya penelitian pada bidang computer vision maka algoritma deteksi objek, YOLO berkembang dengan cepat dan yang terbaru YOLO v8. Perbedaan pada setiap versi YOLO terletak dari peningkatan akurasi dan kecepatan pendeteksian objek. Rangga, 2023 [15]. YOLO termasuk dalam metode Deep Learning memerlukan komputasi yang tinggi. Maka membutuhkan perangkat keras pendukung yang memiliki performa yang tinggi pula. Untuk penelitian ini menggunakan algoritma YOLOv8 dan untuk mengatasi kebutuhan komputasi pada penelitian ini akan menggunakan Google Colab. Berikut ini adalah tahapan penelitian meliputi pengumpulan dataset, anotasi untuk pelatihan dan validasi, pelatihan dan validasi model, serta pengujian model yang dapat dilihat pada Gambar.



Gambar 1. Metode Penelitian

1. Pengumpulan dataset

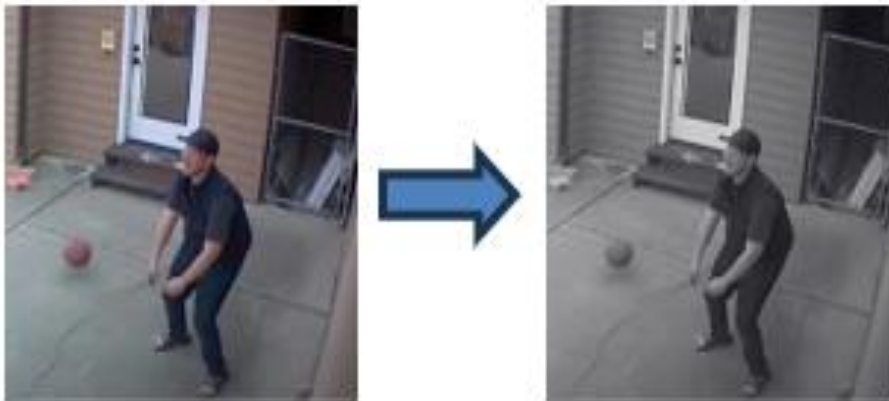
Dataset berupa citra yang menggambarkan objek manusia, pengumpulan dataset menggunakan data internal yang diambil dengan kamera digital maupun cctv, Selain dataset pribadi juga menggunakan dataset external yang merupakan dataset penelitian sebelumnya atau dengan mencari dataset dari sumber public.



Gambar 2. Citra objek manusia

2. Preprosesing data

Dalam penelitian ini menganalisa kemampuan YOLOv8 dalam deteksi objek manusia dalam citra *grayscale*, untuk keperluan *training* maka dibutuhkan dataset objek manusia dalam bentuk *grayscale*. *Preprosesing* data dilakukan dengan merubah dataset citra dalam bentuk RGB menjadi *Grayscale*.



Gambar 3. Merubah gambar RGB menjadi *Grayscale*

3. Anotasi data

Proses anotasi data merupakan proses bounding box dan label pada gambar. Proses anotasi ini dilakukan menggunakan aplikasi labeling yang tersedia dan bersifat open source. Hasil anotasi pada setiap gambar kemudian akan disimpan kedalam bentuk file .xml.

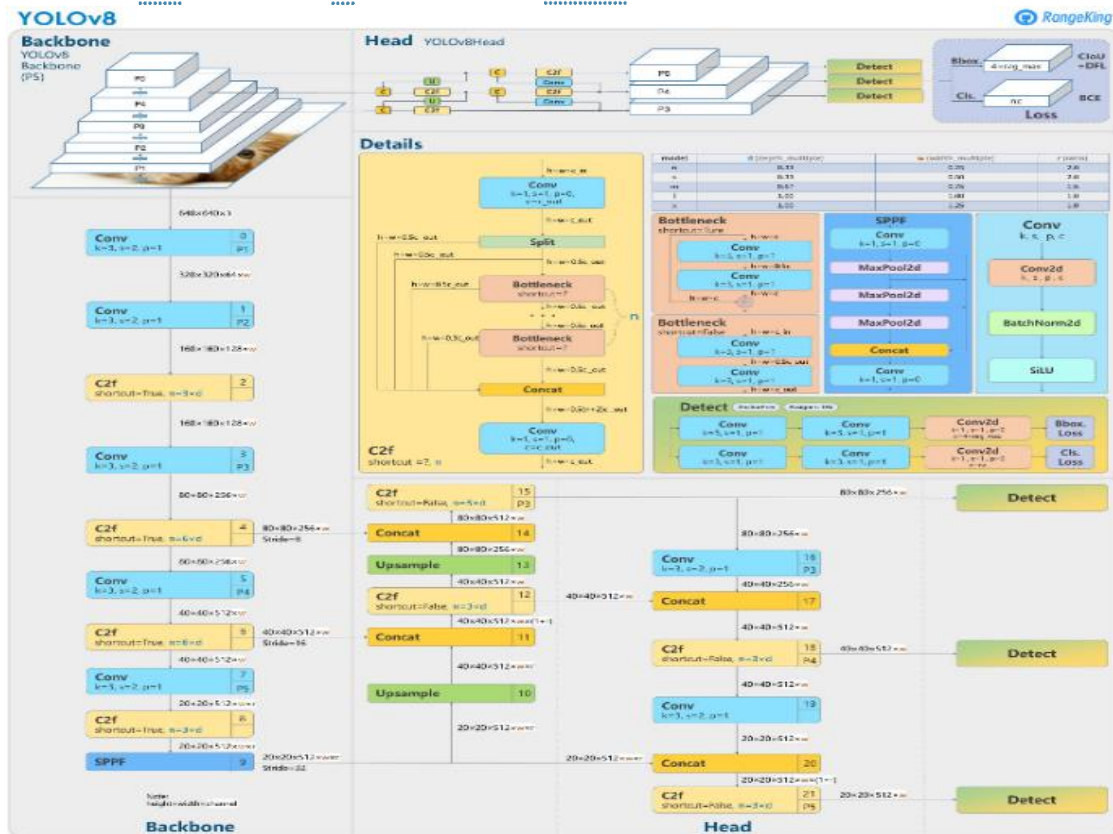


Gambar 4. Anotasi Data Image

4. Modifikasi Arsitektur YOLOv8

Pada penelitian ini dilakukan modifikasi pada arsitektur YOLOv8 bertujuan untuk dapat meningkatkan kemampuan YOLOv8 dalam deteksi objek manusia dalam gambar *grayscale*. YOLOv8 merupakan algoritma yang terdiri berbagai layer yang terbagi menjadi menjadi 3 bagian utama backbone, neck dan head. Dalam penelitian ini akan dilakukan modifikasi pada bagian head diharapkan dengan

arsitektur yang lebih sederhana dengan menyesuaikan ukuran objek deteksi dapat mempercepat kemampuan YOLOv8 dalam deteksi objek.



Gambar 5. YOLOv8 Architecture, visualisation made by GitHub user RangeKin

3. HASIL DAN PEMBAHASAN

3.1. Modifikasi Arsitektur YOLOv8

Pada arsitektur YOLOv8 dapat kita ketahui pada bagian head yang berfungsi untuk menghasilkan output dari proses deteksi objek dalam proses *training*. Pada arsitektur Head terdiri dari 3 layer dengan nilai rasio yang berbeda yang digunakan untuk deteksi objek, pada default arsitektur YOLOv8 3 layer yang ada digunakan secara bersamaan. Pada penelitian ini dilakukan modifikasi arsitektur dengan membagi menjadi 3 model arsitektur dengan setiap model menyisakan 1 layer saja.

Tabel 1. Pembagian Layer Head YOLOv8

No	Nama	Rasio layer head
1	Model 1	80 x 80 x 256 x W
2	Model 2	40 x 40 x 512 x W
3	Model 3	20 x 20 x 512 x W

3.2. Training dengan *custome dataset*

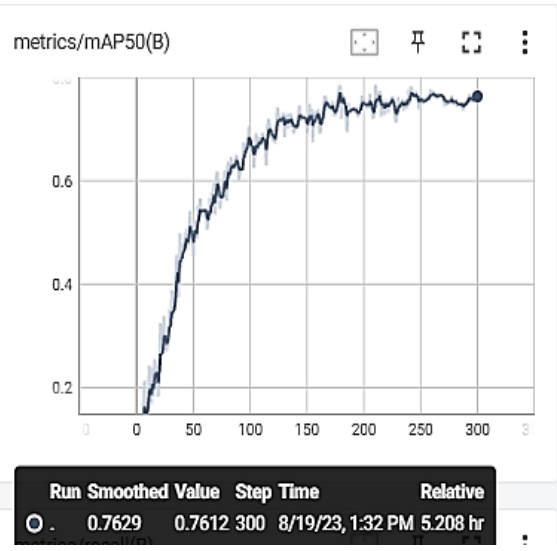
Dalam penelitian ini akan dirubah arsitektur yolov8 menjadi 3 model dengan cara memodifikasi pada bagian head YOLOv8 seperti namapak pada tabel.



Epoch	GPU_mem	box_loss	cls_loss	dfl_loss	Instances	Size
9/100	1.69G	2.835	2.142	2.006	8	640: 100% 220/220 [00:38<00:00, 5.78it/s]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% 14/14 [00:02<00:00, 5.82it/s]
	all	55	82	0.303	0.159	0.138 0.0508
Epoch	GPU_mem	box_loss	cls_loss	dfl_loss	Instances	Size
10/100	1.69G	2.85	2.114	1.965	8	640: 100% 220/220 [00:36<00:00, 6.07it/s]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% 14/14 [00:02<00:00, 5.98it/s]
	all	55	82	0.202	0.341	0.209 0.0667
Epoch	GPU_mem	box_loss	cls_loss	dfl_loss	Instances	Size
11/100	1.7G	2.828	2.183	1.976	8	640: 100% 220/220 [00:37<00:00, 5.94it/s]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% 14/14 [00:02<00:00, 5.93it/s]
	all	55	82	0.312	0.316	0.182 0.0591
Epoch	GPU_mem	box_loss	cls_loss	dfl_loss	Instances	Size
12/100	1.69G	2.684	2.1	1.928	6	640: 100% 220/220 [00:36<00:00, 6.02it/s]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% 14/14 [00:03<00:00, 4.39it/s]
	all	55	82	0.416	0.317	0.311 0.098
Epoch	GPU_mem	box_loss	cls_loss	dfl_loss	Instances	Size
13/100	1.69G	2.644	2.123	1.897	7	640: 100% 220/220 [00:40<00:00, 5.49it/s]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% 14/14 [00:03<00:00, 4.23it/s]
	all	55	82	0.0394	0.317	0.0431 0.0157
Epoch	GPU_mem	box_loss	cls_loss	dfl_loss	Instances	Size
14/100	1.69G	2.661	2.029	1.9	10	640: 100% 220/220 [00:35<00:00, 6.20it/s]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% 14/14 [00:02<00:00, 5.29it/s]
	all	55	82	0.469	0.28	0.287 0.105
Epoch	GPU_mem	box_loss	cls_loss	dfl_loss	Instances	Size
15/100	1.69G	2.663	2.156	1.891	8	640: 53% 117/220 [00:22<00:14, 7.21it/s]

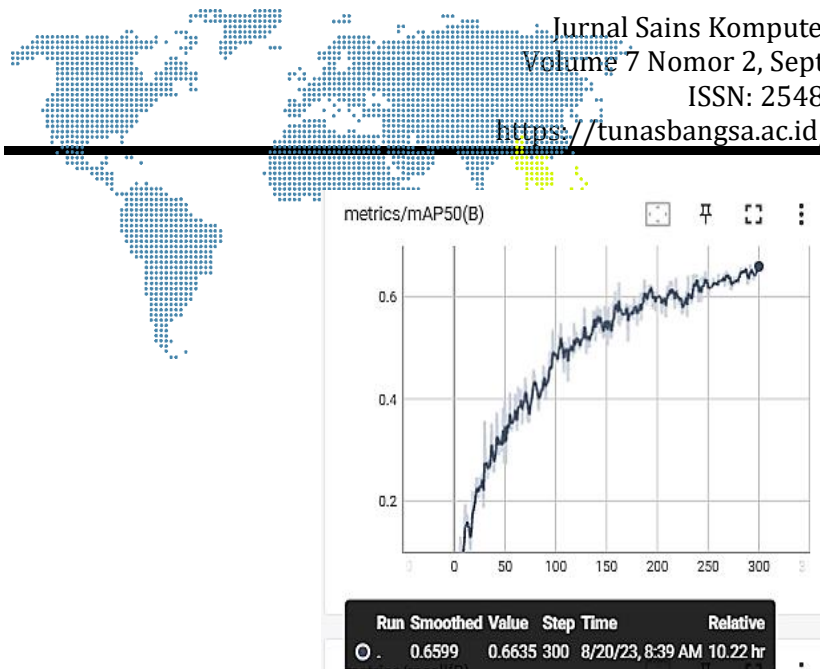
Gambar 6. Proses *training*

Dalam penelitian ini menggunakan custom dataset image yang menggambarkan objek manusia. Untuk menganalisa model yang didapatkan maka dilakukan *training* model dari awal menggunakan dataset yang sudah diseplit sebelumnya dengan pembagian *Training* 80% Test 10% dan validasi 10%. Proses *training* data dilakukan menggunakan komputasi Google colab pro dan dilakukan 4 kali proses *training*.



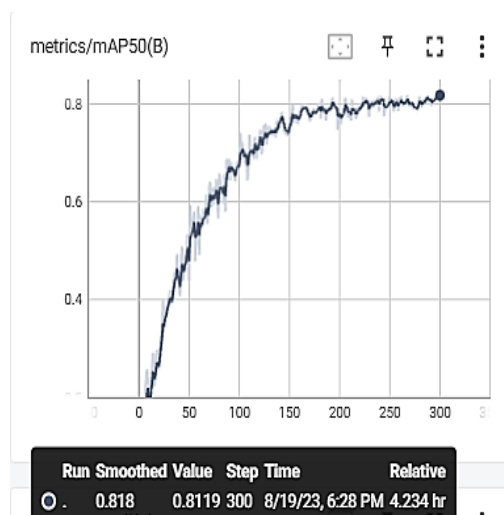
Gambar 7. nilai mAP pada model default

Hasil *training* pada arsitektur default YOLOV8 dengan menggunakan 3 layer pada bagian head. Hasil *training* yang dilakukan menghasilkan nilai mAP sebesar 76%.



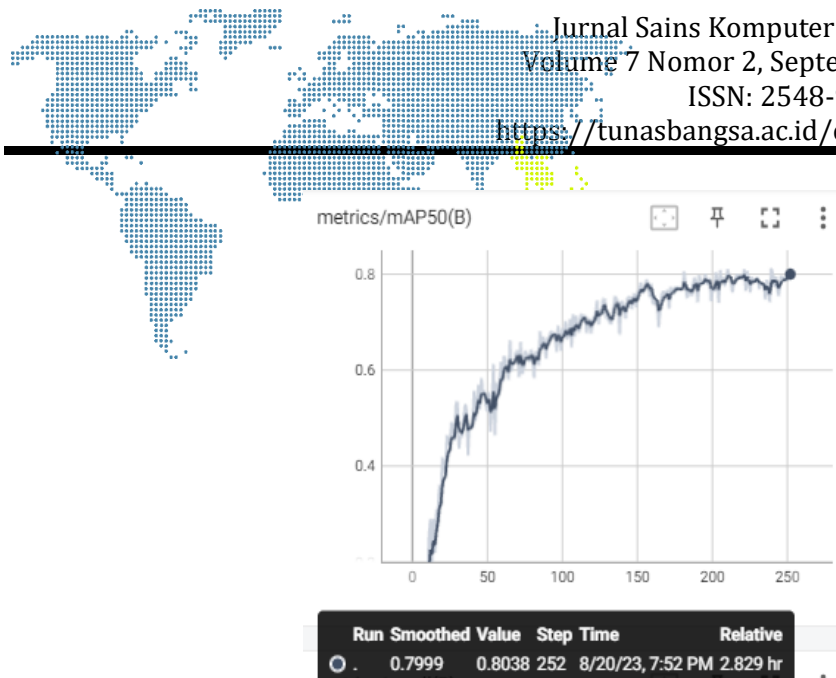
Gambar 8. nilai mAP pada model 1

Hasil *training* pada arsitektur modifikasi YOLOV8 dengan menggunakan 1 layer bagian head dengan nilai layer $80 \times 80 \times 256 \times w$. Hasil *training* yang dilakukan menghasilkan nilai mAP sebesar 66%.



Gambar 9. nilai mAP pada model 2

Hasil *training* pada arsitektur modifikasi YOLOV8 dengan menggunakan 1 layer bagian head dengan nilai layer $40 \times 40 \times 512 \times W$. Hasil *training* yang dilakukan menghasilkan nilai mAP sebesar 81%.



Gambar 10. nilai mAP pada model 3

Hasil *training* pada arsitektur modifikasi YOLOV8 dengan menggunakan 1 layer bagian head dengan nilai layer 20x20x512xW. Hasil *training* yang dilakukan menghasilkan nilai mAP sebesar 80%

3.3. Testing Model

Stelah proses *training* selesai akan menghasilkan model YOLOv8 berdasarkan arsitektur yang digunakan untuk mengenalkan pada objek manusia. Hasil testing menunjukkan model dapat mengenali objek manusia pada gambar *grayscale* dengan baik dengan mencapai nilai 80%.



Gambar 11. Testing model

3.4. Analisa Hasil Model

Training model dilakukan 4 kali menghasilkan permodelan berdasarkan modifikasi arsitektur

Tabel 2. Model YOLOv8

NO	Model	Epoch	Nilai mAP
1	Default	300	76
2	Model 1	300	66
3	Model 2	300	81
4	Model 3	300	80

Hasil *training* yang dilakukan sebanyak 4 kali dengan menggunakan cutome dataset menghasilkan model yang berbeda-beda . terlihat pada tabel diatas bahwa hasil model 2 dengan layer 40x40x512xW menghasilkan nilai mAP paling tinggi diantara model yang lainya dengan nilai mAP mencapai 81 %

4. SIMPULAN

Dari penelitian yang sudah dilakukan, dapat ditarik kesimpulan bahwa modifikasi arsitektur YOLOv8 dengan merubah pada head layer dapat mempengaruhi hasil *training*. Dari hasil *training* yang ada hasil modifikasi pada model 2 mendapatkan nilai yang paling tinggi dari model yang lainya mencapai mAP 81% ini lebih tinggi dari hasil model default YOLOv8 yang hanya menghasilkan nilai mAP 76%. Penelitian ini mencoba menganalisa kemampuan algoritma yolov8 dengan modifikasi arsitektur pada bagian Head dengan membuat 3 model arsitektur yang masing-masing model memiliki 1 layar deteksi saja. Untuk peneliti dengan tema yang sama bisa melakukan penelitian dengan menggunakan modifikasi bagian selain head misalkan pada bagian arsitektur backbone ataupun neck.

DAFTAR PUSTAKA

- [1] Ata Jahangir Moshayedi, Atanu Shuvam Roy, Amin Kolahdooz, Yang Shuxin (2022), " Deep Learning Application Pros and Cons Over Algorithm", EAI Endorsed Transactions on AI and Robotics, doi: 10.4108/10.4108/airo.v1i.19.
- [2] W. Zhu, H. Zhang, J. Eastwood, X. Qi, J. Jia, and Y. Cao, "Concrete crack detection using lightweight attention feature fusion single shot multibox detector," Knowl Based Syst, vol. 261, p. 110216, Feb. 2023.
- [3] Chien-Yao Wang1 , Alexey Bochkovskiy, and Hong-Yuan Mark Liao1, 2022 " YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors", arXiv:2207.02696v1.
- [4] Amelia, Dewi, I. A & Rafiqi, F. 2019. Deteksi Manusia Menggunakan Pre-Trained MobileNet untuk Segmentasi Citra Menentukan Bentuk Tubuh. Mindjournal.

- [5] Wahyu, Rahmantiar, W & Hernawan, A. 2021,. Real-Time Detection Using Deep Learning on Embedded platform: A Review. *Journal of Robotics and Control (JRC)* Volume 2.
- [6] Swastika, W., Nur, A.W., Kelana, O.H., 2019., Monitoring Ruangn Untuk Deteksi Manusia berbasis CNN Dengan Fitur Push Notification., *TEKNIKA*, Volume 8 Nomer 2.
- [7] Kharunnas, Mulyanto, E. & Zaini, A., 2021. Pembuatan Modul Deteksi Objek Manusia Menggunakan Metode YOLO untuk Mobile Robot., *Jurnal Teknik ITS., Volume 10 No. 1.*
- [8] Kristo, M., Kos, M.I & Pobar, M. 2020. Thermal Object Detection in Difficult Weather Conditions Using Yolo.*IEEE.*
- [9] Kurniawan, E. Adinanta, H. Suryadi. Sirenden, B.H. Ula, R.K. Pratomo, H Purwowbowo & Prakosa, J.A. 2022. Deep Neural Network-based physical distancing Monitoring System with TensorRT Optimization. *Internasional Journal Of Advances in Intelligent Informatics (ijain). Vol 8 No 2.*
- [10] M. Musser, 2020 “Detecting Autism Spectrum Disorder in Children With Computer Vision,”. <https://towardsdatascience.com>.
- [11] Harahap, M., Agustinaa, E.S., Lubis, M. M., Apriadi, & Anggar, A., 2020. Deteksi Objek Manusia Pada Image Dengan Metode Thinning Berdasarkan Local Maxima., *Jatisi Vol7 No 3.*
- [12] Arfan, M. Nurjalal, A., Somantri, M & Sudjadi., 2021. Pengenalan aktivitas Manusia pada Area Tambak Udang dengan Convolutional Neural Network. *Jurnal RESTI, Vol5 no 1.*
- [13] Sumit, S.S. Rambli, D.R.A., Mirjalili, S., Ejaz, M.M. & Miah, M.S.U. 2022. ReStinet: Improving the Perfomance of Tiny-YOLO-Based CNN Architecture for Applications in Human Detection. *MDPI.*
- [14] Muannif Ridwan, Suhar AM, Bahrul Ulum dan Fauzi Muhammad, 2021, “ The Importance Of Application Of Literature Review In Scientific Research” *jurnal masoho vol 2.*
- [15] Rangga Gelar Guntara, 2023 , “ Pemanfaatan Google Colab Untuk Aplikasi Pendeteksian Masker Wajah Menggunakan Algoritma Deep LearningYOLOv7” *Jtkis vol 5 no 1.*