

Human Detection menggunakan metode YOLO v4

Achmad Salim Aiman
Informatika
Telkom University
Bandung, Indonesia
salimaiman@student.telkomuniversity.ac.id

Gempur Bayu Aji
Informatika
Telkom University
Bandung, Indonesia
gempurbayu@student.telkomuniversity.ac.id

Windu Firmansyach
Informatika
Telkom University
Bandung, Indonesia
windufirmansyach@student.telkomuniversity.ac.id

Abstrak—Perkembangan teknologi sudah semakin pesat di berbagai bidang salah satunya adalah bidang keamanan. Salah satu teknologi yang digunakan adalah CCTV untuk mengawasi dan mengantisipasi terjadinya hal hal yang dapat mengganggu keamanan di suatu area. Sistem deteksi objek pun telah diterapkan untuk mempermudah manusia dalam mendapatkan informasi mengenai objek yang terdapat di suatu tempat. Sehingga pada penelitian ini kami membangun suatu sistem deteksi manusia menggunakan metode YOLO v4. Sistem yang dibangun dapat menghasilkan nilai mAP 92.95%.

Kata Kunci—Human Detection, YOLOv4, mean Average Precision

I. INTRODUCTION

Perkembangan teknologi di bidang kecerdasan buatan telah meningkat secara pesat khususnya di bidang *computer vision* yang diterapkan sering diterapkan pada sistem keamanan menggunakan kamera CCTV, kemudian di bidang medis yakni dengan kamera pendeteksi penyakit pada suatu organ tubuh. Selain itu juga *computer vision* diterapkan pada bidang kendaraan yakni *autonomous vehicle* dimana suatu kendaraan dapat menentukan dan menempuh perjalanan menuju tempat tujuannya tanpa harus dikendalikan oleh supir.

Sistem deteksi objek merupakan salah satu penerapan *computer vision* untuk mendeteksi suatu objek yang terdapat dalam gambar, maupun video. Sistem deteksi objek ini merupakan gabungan dari sistem klasifikasi dan *object localization*. Sehingga yang dilakukan ialah tidak hanya mengkategorikan suatu gambar atau video pada suatu kelas tertentu, namun harus mengetahui letak setiap objek yang terdapat di dalamnya sehingga perlu diterapkan beberapa metode seperti *face detection*, *pedestrian detection*, ataupun *skeleton detection* [2].

Sistem deteksi objek berbasis *deep learning* telah banyak dikembangkan oleh para peneliti. Metode deteksi objek terbagi menjadi dua jenis *two-stage* dan *one-stage* (satu tahap). Beberapa algoritma dua-tahap antara lain R-CNN, Fast R-CNN, Faster R-CNN, dan masih banyak algoritma lainnya. Beberapa algoritma *one-stage* (satu tahap) yakni YOLO, SSD, DetectNet, dan algoritma lainnya.

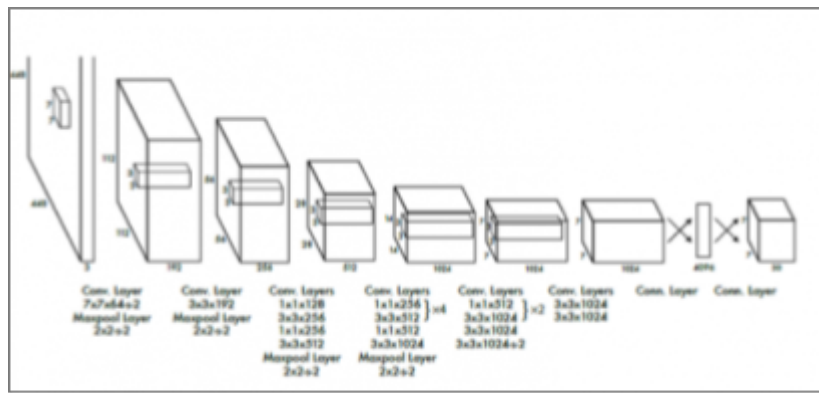
Pada penelitian kami melakukan deteksi objek manusia yang terdapat dalam sebuah gambar atau video dengan menggunakan algoritma *one-stage* yakni YOLO pada versi ke-4. Harapan dari penelitian ini yakni kami dapat membangun sebuah model yang mampu melakukan pendeteksian objek dalam waktu yang singkat dengan tingkat *confidence* yang tinggi.

Hal-hal yang ditampilkan pada jurnal ini dibagi kedalam beberapa sub-bagian. Pada sub-bagian II dipaparkan beberapa pekerjaan yang telah selesai yang berhubungan dengan penelitian kami (*related works*). Kemudian penjelasan mengenai dataset yang digunakan dibahas pada sub-bagian III. Dan pada sub-bagian IV dipaparkan pembahasan mengenai metode yang kami gunakan. Sub-bagian V dan VI merupakan inti dari penelitian dimana proses eksperimen dan hasil yang didapat beserta kesimpulan dari penelitian ini dibahas.

II. RELATED WORKS

Penelitian mengenai deteksi objek manusia telah banyak dilakukan, salah satunya yang dilakukan oleh Ivašić-Kos dkk pada penelitian [3] dimana mereka mendeteksi objek manusia dengan algoritma YOLOv3 pada gambar dan video yang diambil menggunakan kamera *thermal* yang diambil di sebuah hutan pada malam hari dengan berbagai kondisi cuaca. Penelitian tersebut menghasilkan model yang lebih baik untuk mendeteksi objek pada gambar atau video dari kamera *thermal* dibandingkan *pre-trained* model YOLO yang dilatih dengan dataset COCO.

Perbandingan algoritma YOLOv4, YOLOv3, dan Region-Based-CNN yang dilakukan pada penelitian oleh Niharika dkk [5]. Tiga algoritma tersebut diuji pada data citra yang didapat dari sistem *mobile eye-tracker*. *Mobile eye-tracker* ini adalah sebuah alat yang digunakan digunakan untuk melacak aktivitas belajar siswa di mana alat ini dapat merekam apa saja yang dilihat oleh siswa ketika belajar sehingga perlu dideteksi apa saja yang dilihat oleh siswa tersebut. Dari penelitian tersebut didapatkan hasil dimana YOLOv4 memiliki kecepatan dan akurasi yang paling tinggi dibandingkan YOLOv3 dan R-CNN.



Gambar 1. Arsitektur YOLO

III. DATA

Citra manusia yang kami butuhkan cukup mudah didapat dari berbagai sumber seperti *MS COCO* [7], *PASCAL VOC* [8], *Google Open Image* [9], dan *ImageNet* [10]. Namun pada tugas besar ini kami mencoba menggunakan dataset selain dataset diatas yang kami dapat dari *kaggle.com*¹. Dataset tersebut telah digunakan di penelitian yang dilakukan Karthika dkk [11]. Dataset ini mengandung 944 citra latih, 160 citra validasi, dan 235 citra test.

Dataset ini dibuat untuk mengatasi masalah *false positive*. Jika data hanya mengandung citra manusia maka akan mengakibatkan model melakukan *false positive prediction*. Oleh karena itu sebagai data yang digunakan untuk deteksi objek manusia, dataset ini tidak hanya mengandung citra manusia (*person*) tetapi juga mengandung citra objek yang menyerupai manusia (*person-like*) dengan tujuan mengatasi masalah tersebut. Kelas *person* merepresentasikan data citra manusia dengan berbagai posisi, sedangkan kelas *person-like* merepresentasikan data citra objek yang menyerupai manusia seperti patung, manekin, orang-orangan sawah, boneka, dan robot. Contoh data citra dapat dilihat pada Gambar 2 berikut.



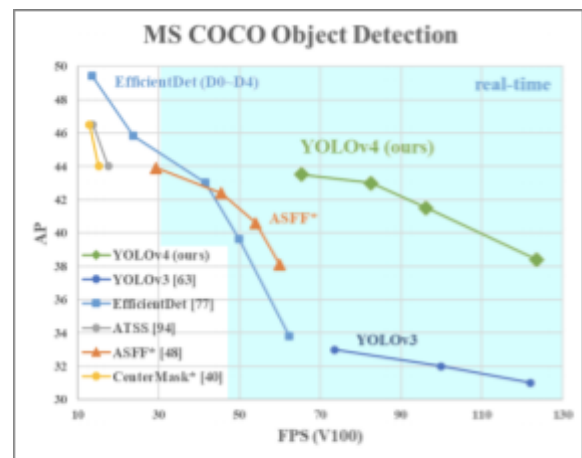
Gambar 2. Sample dataset

IV. YOU ONLY LOOK ONCE (YOLO)

Salah satu teknik dalam pendeteksian objek adalah dengan menggunakan algoritma YOLO (You Only Look Once). YOLO merupakan neural network pada sebuah citra, kemudian membagi citra menjadi daerah dan memprediksi *bounding box* serta probabilitas untuk masing-masing daerah. Probabilitas untuk setiap *bounding box* kemudian dihitung untuk mengklasifikasikan sebagai objek atau bukan. YOLO dapat melakukan sebuah pengenalan objek secara realtime dengan kecepatan 45 frame per second [2].

YOLO merupakan salah satu algoritma yang paling baik dalam hal deteksi objek. YOLO bekerja dengan melakukan satu CNN dan kemudian gambar akan dibagi menjadi *grid* atau kotak berukuran $S \times S$ kemudian *bounding box* beserta *confidence* dihitung pada setiap *grid* [3]. YOLO memiliki 24 layer konvolusi dan 2 lapisan *fully-connected* seperti dapat dilihat pada Gambar 1.

Setelah dilakukan pengujian dan pengembangan pada algoritma YOLO mulai dari YOLOv1, YOLO9000, YOLOv3 hingga YOLOv4 bahkan YOLOv5. Setiap pengembangan dilakukan tentunya karena terdapat kekurangan di setiap algoritma. YOLOv4 merupakan metode terbaik diantara tiga YOLO sebelumnya. Sebagaimana pada jurnal penelitiannya didapatkan hasil bahwa YOLOv4 menghasilkan akurasi dan kecepatan latih yang paling baik seperti dapat pada Gambar 2 [6].

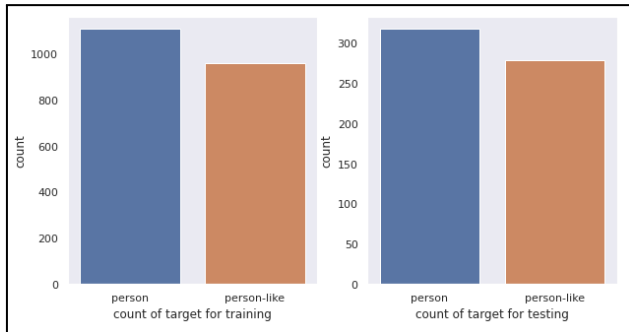


Gambar 3. Perbandingan performansi YOLOv4 dengan algoritma sebelumnya

¹ <https://www.kaggle.com/datasets/karthika95/pedestrian-detection>

V. EXPERIMENT

Tahap percobaan dilakukan dengan mempersiapkan dataset. Dataset terdiri dari data training, validation, dan testing. Namun kami hanya menggunakan data train dan tes untuk efisiensi waktu latih serta *computational cost* pada saat pelatihan. Data train mengandung 944 citra dan data test mengandung 235 citra, sehingga total data sebanyak 1179 citra. Jumlah data *person* dan *person-like* pada data train dan test dapat dilihat pada Gambar 4.



Gambar 4. Persebaran kelas *person* dan *person-like*

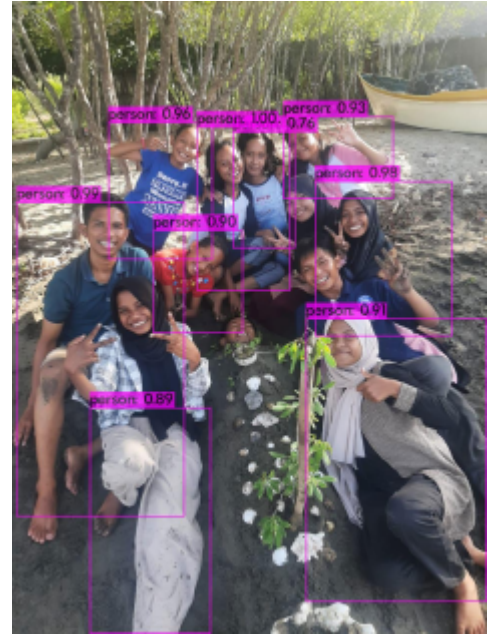
Data anotasi pada dataset masih dalam format PASCAL VOC (.xml). Oleh karena itu seluruh data anotasi perlu dikonversi ke dalam format YOLO (.txt). Setelah didapatkan anotasi dalam format YOLO maka anotasi dapat digunakan dalam proses training maupun sebagai *ground truth* pada proses testing dan evaluasi.

Proses training dilakukan dengan Google Colab menggunakan GPU dan didukung dengan CUDA versi 11.1 yang berlangsung selama kurang lebih 5 jam. Setelah itu dilakukan testing pada model dengan data citra baru. Model melakukan deteksi pada objek dengan baik, namun terkadang model mengalami kebingungan sehingga masih salah dalam membedakan objek manusia dan objek yang menyerupai manusia. Dapat dilihat pada Gambar 5 bahwa model salah mendeteksi objek yang seharusnya *person* tetapi dideteksi sebagai *person-like*.



Gambar 5. Kebingungan yang dialami model

Pada percobaan lain modal dapat melakukan deteksi dengan baik dan menghasilkan *confidence* yang cukup tinggi hingga mendekati 100%. Namun hasil tersebut menyimpulkan bahwa model mengalami *overfit*.



Gambar 6. Hasil deteksi dengan *confidence* tinggi.

Pada tahap evaluasi model kami menggunakan metode mAP yakni dengan menghitung nilai *precision* yang dihasilkan model. Didapatkan nilai mAP sebesar 92,95% dimana AP kelas *person* 91,26% dan AP kelas *person-like* sebesar 94.64%. Hasil *confusion* dapat dilihat pada Tabel 1.

Tabel 1. *Confusion Report* model YOLOv4

True Positive	543
False Positive	99
False Negative	47
Precision	85%
Recall	92%
F-1 Score	88%
Average IoU	67,69%

VI. CONCLUSION

Dari keseluruhan proses training data hingga menghasilkan model deteksi objek, dari hasil evaluasi dapat diambil kesimpulan bahwa model dapat melakukan deteksi dengan baik namun model terkadang masih mengalami kebingungan dalam proses deteksi. Hal tersebut disebabkan

dari jumlah data train pada data *person* dan *person-like* perlu ditambah serta melakukan lebih banyak metode ekstraksi fitur selain yang telah dipakai pada model YOLOv4.

Dari penelitian ini terdapat beberapa inovasi yang bisa dilakukan diantaranya dengan melakukan deteksi pada jenis kelamin pada suatu objek manusia, dan juga deteksi usia dari objek manusia tersebut.

REFERENCES

- [1] Zhao, Z. Q., Zheng, P., Xu, S. T., & Wu, X. (2019). Object detection with deep learning: A review. *IEEE transactions on neural networks and learning systems*, 30(11), 3212-3232..
- [2] Karlina, O. E., & Indarti, D. (2019). Pengenalan Objek Makanan Cepat Saji Pada Video Dan Real Time Webcam Menggunakan Metode You Look Only Once (Yolo). *Jurnal Ilmiah Informatika Komputer*, 24(3), 199–208. <https://doi.org/10.35760/ik.2019.v24i3.2362>
- [3] Ivašić-Kos, Marina, Mate Krišto, and Miran Pobar. "Human detection in thermal imaging using YOLO." *Proceedings of the 2019 5th International Conference on Computer and Technology Applications*. 2019.
- [4] Shinde, Shubham, Ashwin Kothari, and Vikram Gupta. "YOLO based human action recognition and localization." *Procedia computer science* 133 (2018): 831-838.
- [5] Kumari, Niharika, et al. "Mobile Eye-Tracking Data Analysis Using Object Detection via YOLO v4." *Sensors* 21.22 (2021): 7668.
- [6] Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao. "Yolov4: Optimal speed and accuracy of object detection." *arXiv preprint arXiv:2004.10934* (2020).
- [7] Lin, Tsung-Yi, et al. "Microsoft coco: Common objects in context." *European conference on computer vision*. Springer, Cham, 2014.
- [8] Vicente, Sara, et al. "Reconstructing pascal voc." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014.
- [9] Kuznetsova, Alina, et al. "The open images dataset v4." *International Journal of Computer Vision* 128.7 (2020): 1956-1981.
- [10] J. Deng, W. Dong, R. Socher, L. -J. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248-255, doi: 10.1109/CVPR.2009.5206848.
- [11] Karthika, N. J., and Saravanan Chandran. "Addressing the False Positives in Pedestrian Detection." *Electronic Systems and Intelligent Computing*. Springer, Singapore, 2020. 1083-1092.