

DOI: doi.org/10.21009/03.1201.FA04

DETEKSI OBJEK BAYANGAN KENDARAAN MENGUNAKAN FASTER R-CNN

Deeva Nabila^{1, a)}, Bambang Heru Iswanto^{1, b)}, Risnandar^{2,3, c)}

¹Program Studi Fisika, FMIPA, Universitas Negeri Jakarta, Jl. Rawamangun Muka No. 01, Rawamangun 13220, Indonesia

²Computer Vision & Image Processing Research Group-Pusat Riset Kecerdasan Artifisial dan Keamanan Siber-BRIN, Jl. Sangkuriang, Dago, Kecamatan Coblong, Kota Bandung, Jawa Barat 40135, Indonesia

³Intelligence Systems Research Group-Program Studi Informatika-Fakultas Informatika, Telkom University Jl. Telekomunikasi. 1, Terusan Buahbatu - Bojongsoang, Telkom University, Sukapura, Kec. Dayeuhkolot, Kabupaten Bandung, Jawa Barat, 40257, Indonesia

Email: ^{a)}deevanabila54@gmail.com, ^{b)}bhi@unj.ac.id, ^{c)}risnandar@brin.go.id

Abstrak

Objek bayangan memiliki intensitas dan bentuk yang beragam, yang dapat menimbulkan masalah dalam sistem visi kendaraan otonom. Bayangan yang ditimbulkan dari pohon, bangunan, dan objek lain di sekitar jalan dapat mempengaruhi kinerja sistem pengenalan dan pelacakan target. Maka itu, diperlukan suatu model pendeteksian bayangan untuk mengetahui lokasi bayangan agar dapat digunakan pada penelitian terkait eliminasi bayangan. Penelitian ini bertujuan untuk mengetahui tingkat akurasi model dengan variasi dataset yang kami berikan dan mendefinisikan masing-masing label objek *non-shadow* dan *shadow* merupakan metode yang digunakan untuk membedakan antara bayangan dan objeknya yang mirip. Pelatihan model dilakukan dengan *fine-tuning* Faster R-CNN pada kerangka kerja Pytorch menggunakan arsitektur ResNet50 sebagai rancangan dasar. Implementasi model untuk dapat mendeteksi bayangan diterapkan pada video perjalanan kendaraan otonom. Hasil penelitian menunjukkan bahwa dari kelima model yang dibuat, model P5 berhasil mendeteksi bayangan dengan rata-rata akurasi *F1-score* sebesar 46%.

Kata-kata kunci: Bayangan, Deteksi, Faster R-CNN, R-CNN, ResNet50, Pytorch.

Abstract

Shadow objects exhibit varying intensities and shapes, which can pose problems in autonomous vehicle vision systems. Shadows generated by trees, buildings, and other objects in the vicinity of the road can impact the performance of the recognition and tracking system. Thus, a shadow detection model is necessary to determine the location of shadows, which can be employed in studies related to shadow removal. This study aims to determine the accuracy level of the model with our given diverse dataset and defining distinct labels for non-shadow and shadow objects to differentiate between shadows and similar-looking objects. The model training was performed by fine-tuning Faster R-CNN on the PyTorch framework, utilizing ResNet50 as the backbone architecture. The implemented model aimed to detect shadows in videos of autonomous vehicle. The results indicated that out of the five models developed, P5 model successfully detected shadows with an average accuracy based on F1-score is 0.46%.

Keywords: Shadow, Detection, Faster R-CNN, R-CNN, ResNet50, Pytorch.

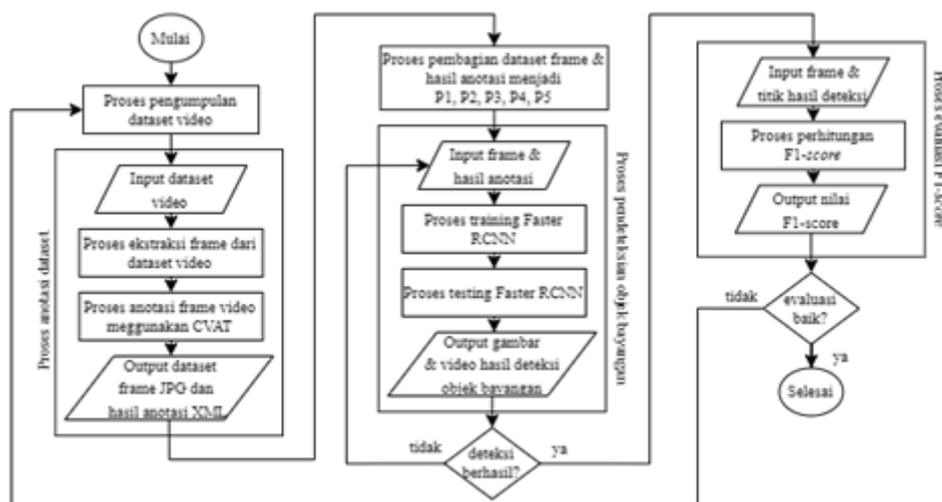
PENDAHULUAN

Perubahan iluminasi cahaya yang menimpa sebuah objek dapat membentuk bayangan dengan karakteristik yang berbeda dengan objek tersebut. Pada kondisi pencahayaan di luar ruangan, bayangan memiliki variasi bentuk dan ukuran yang beragam. Bahkan dapat menyebabkan distorsi bentuk dan ukuran objek asli karena dikelirukan sebagai bagian dari objek yang bergerak. Oleh karena itu, keberadaan bayangan dapat mempengaruhi kinerja aplikasi visi komputer. Misalnya, pada aplikasi dalam sistem perencanaan pergerakan kendaraan otonom yang memerlukan informasi lingkungan yang akurat [3]. Kesalahan pengenalan objek di sekitar kendaraan harus diminimalkan, termasuk pengenalan bayangan. Iluminasi yang ditimbulkan oleh bayangan dari objek seperti mobil, pohon, dan bangunan dapat mempengaruhi pembentukan markah jalan [13]. Dampak lain yang ditimbulkan oleh bayangan antara lain dalam aplikasi segmentasi objek, deteksi objek [12], dan pelacakan garis [2]. Oleh karena itu, penting mengetahui lokasi bayangan untuk menghasilkan aplikasi visi komputer yang kuat dan akurat.

Teknik pembelajaran mendalam menjadi teknik yang efektif untuk mendeteksi bayangan karena mampu mengatasi berbagai kasus iluminasi cahaya yang ekstrem sehingga fitur-fitur bayangan dapat diekstraksi secara mendalam [8]. Metode deteksi bayangan bergerak berbasis CNN oleh Kim dkk [4] dan Lee dkk [9] perlu melakukan proses pengurangan latar belakang sebagai masukan. Sayangnya, teknik pengurangan latar belakang memiliki banyak tantangan, kamufase, *bootstrap*, latar belakang dinamis, dan kebisingan video [10]. Faster RCNN merupakan salah satu perkembangan dari CNN yang mampu mendeteksi banyak objek dalam satu citra, sekaligus meningkatkan akurasi dengan performa yang lebih cepat [11]. Walaupun banyak digunakan untuk deteksi objek seperti kerusakan jalan [1], transportasi [5], hewan [7], penggunaannya untuk deteksi bayangan masih terbatas. Penelitian oleh Bao dkk [6], menggunakan Faster RCNN untuk deteksi bayangan pada video *Synthetic Aperture Radar* (SAR), menunjukkan akurasi yang lebih tinggi sebesar 9% dibandingkan dengan tingkat dasar eksperimental Faster R-CNN.

Penelitian ini menarik permasalahan pada bayangan objek yang ada di jalur kendaraan. Di mana pada kasus tertentu bayangan hampir menutupi sebagian jalur sedangkan objek bayangan tersebut ada di luar. Bahkan beberapa bayangan memiliki bentuk dan warna yang hampir mirip dengan objeknya. Tujuan penelitian ini adalah untuk mengetahui tingkat akurasi model dalam mendeteksi objek bayangan menggunakan variasi dataset yang kami berikan. Selain itu, kami juga mendefinisikan masing-masing label objek *non-shadow* dan *shadow* untuk membedakan antara bayangan dan objeknya yang mirip. Hal ini dilakukan untuk menganalisis kemampuan model dalam mendeteksi bayangan secara bertahap.

METODOLOGI



GAMBAR 1. Diagram alir penelitian

Pengumpulan Dataset

Dataset CDNet (Bungalows dan Highway) dan dataset pribadi digunakan untuk pelatihan model. Bungalows dan Highway merupakan dataset video rekam kendaraan beserta bayangannya pada keadaan kamera yang statis. Sedangkan dataset pribadi adalah rekam jalur kendaraan pada keadaan kamera yang dinamis. Total kelas label adalah tujuh yang diberi nama motorcycle, shadow_motorcycle, drainase, shadow_flag, car, shadow_car, dan shadow_tree.

Anotasi Dataset

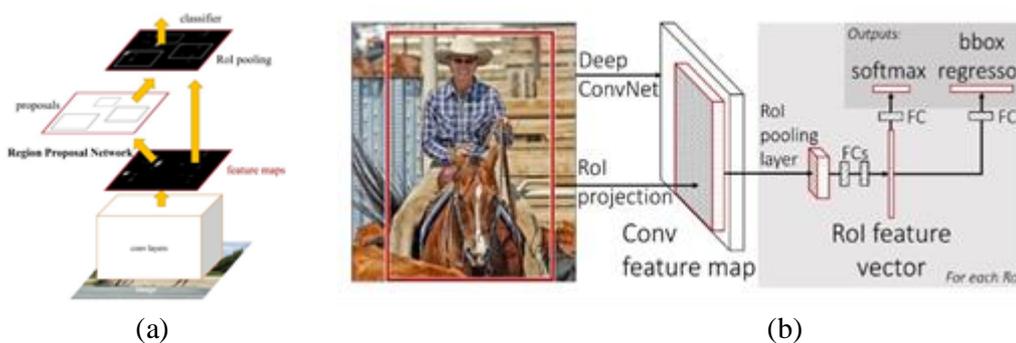
Anotasi dilakukan secara manual menggunakan aplikasi berbasis web CVAT (*Computer Vision Annotation Tools*) untuk menentukan lokasi setiap label yang kemudian disebut sebagai kotak *ground truth*. Hasil anotasi bayangan disimpan dalam format xml dengan total gambar adalah 586 dan jumlah masing-masing label adalah 150.

Pembagian Dataset

Dataset dibagi menjadi lima bagian yang disebut dengan P1, P2, P3, P4, dan P5. P1 merupakan dataset yang terdiri dari label motorcycle dan shadow_motorcycle, P2 terdiri dari label pada P1 dan label drainase, P3 terdiri dari label pada P2 dan label car dan shadow_car, P4 terdiri dari label pada P3 dan label shadow_flag, dan P5 adalah label pada P4 dan label shadow_tree. Pada proses pelatihan, untuk setiap dataset digunakan sebesar 70% untuk pelatihan, 20% untuk validasi, dan 10% untuk pengujian.

Pelatihan Model dengan Faster R-CNN

Faster RCNN merupakan model deteksi objek yang dikembangkan dari Fast RCNN dan menawarkan metode *Region Proposal Network* (RPN) sebagai pengganti algoritma *selective search* untuk menentukan proposal daerah. Gambar yang dimasukkan untuk pelatihan diekstraksi menggunakan arsitektur CNN dan menghasilkan peta fitur yang kemudian diarahkan ke RPN untuk menghasilkan proposal daerah dengan berbagai skala dan aspek rasio. Selanjutnya, proposal daerah diarahkan ke Fast RCNN sebagai tahap akhir dari pendeteksian objek Faster RCNN.



GAMBAR 2. Arsitektur Faster RCNN: (a) RPN dan (b) Fast RCNN [11]

Pengkodean dalam penelitian ini menggunakan platform komputasi berbasis awan gratis yang disediakan oleh Google, yakni Google Colaboratory. Bahasa pemrograman yang digunakan adalah python, dan beberapa library utama antara lain: Numpy, OpenCV, Matplotlib, Torch. Penelitian ini menggunakan teknik *fine-tuning* yang disediakan oleh kerangka kerja Pytorch, dengan rancangan dasar ResNet-50. Beberapa parameter yang digunakan diantaranya *learning rate* 0,001; *momentum* 0,9; *weight decay* 0,0001; *batch* 4; yang di latih sebanyak 50 epoch.

Evaluasi Menggunakan *Confusion Matrix*

Pengujian sistem deteksi dilakukan dengan menentukan nilai *confusion matrix*. Kami menggunakan nilai *F1-score* sebagai indeks akurasi karena memungkinkan untuk menemukan titik keseimbangan antara tingkat deteksi yang tinggi (*recall* tinggi) dan tingkat peringatan palsu yang rendah (*presisi* rendah) [6]. Dengan demikian, nilai *F1-score* dapat memberikan indikasi tentang seberapa baik sistem dapat mengenali bayangan yang benar sambil meminimalkan kesalahan dalam memberikan peringatan palsu. Dirumuskan:

$$F1 - score = 2 \times \left(\frac{Precision \times Recall}{Precision + Recall} \right)$$

atau

$$F1 - score = 2 \times \left(\frac{TP}{2TP + FP + FN} \right)$$

Dengan *True Positive* (TP) menunjukkan label hasil pendeteksian yang benar pada lokasi yang benar, *False Positive* (FP) menunjukkan adanya pendeteksian pada label yang benar namun di lokasi yang salah, dan *False Negative* (FN) menunjukkan hasil pendeteksian yang salah pada lokasi yang salah.

HASIL DAN PEMBAHASAN

Penelitian ini akan memfokuskan pembahasan pada label bayangan, yakni: *shadow_motorcycle*, *shadow_car*, *shadow_flag*, dan *shadow_tree*. Akurasi berdasarkan *F1-score* untuk setiap dataset ditunjukkan dalam tabel berikut ini:

TABEL 1. Nilai akurasi *F1-score* (%)

Label	P1	P2	P3	P4	P5
<i>shadow_motorcycle</i>	38%	50%	16%	16%	45%
<i>shadow_car</i>			26	26%	50%
<i>shadow_flag</i>				26%	49%
<i>shadow_tree</i>					40%
rata-rata	38%	50%	21%	23%	46%

Berdasarkan hasil matriks evaluasi di atas, dalam suatu keadaan, penambahan label yang berbeda dapat mengurangi tingkat akurasi. Namun, akurasi model dapat meningkat ketika adanya penambahan dan variasi label tertentu. Pada setiap dataset selalu terdapat label bayangan motor, di mana akurasi model pada kelas label ini mengalami penurunan pada dataset P3 dari 0,50% menjadi 0,16% dan kembali meningkat pada dataset P5 menjadi 0,45%. Selain itu, tidak ada perubahan akurasi kelas label pada dataset P3 ke P4 meskipun terdapat penambahan kelas label *shadow_flag*. Terlihat pada nilai akurasi deteksi bayangan motor dan mobil yang stabil pada dataset P3 dan P4, secara berurutan yakni 0,16% dan 0,26%. Sedangkan, rata-rata keseluruhan akurasi tiap kelas label paling tinggi ada pada dataset P5.

TABEL 2. Nilai TP, FP, dan FN per kelas

Label	TP					FP					FN				
	P1	P2	P3	P4	P5	P1	P2	P3	P4	P5	P1	P2	P3	P4	P5
<i>motorcycle</i>	15	18	18	18	19	7	14	34	34	8	0	9	50	50	16
<i>shadow_motorcycle</i>	6	11	10	10	12	20	21	57	57	24	1	1	46	46	5
<i>drainase</i>		12	12	12	12		1	23	23	14		8	67	87	38

Label	TP					FP					FN				
	P1	P2	P3	P4	P5	P1	P2	P3	P4	P5	P1	P2	P3	P4	P5
car			17	17	16			22	22	7			60	60	7
shadow_car			14	14	15			55	55	11			25	25	19
shadow_flag				11	10				10	6				54	15
shadow_tree					8					10					15

Mendefinisikan masing-masing label objek *non-shadow* dan *shadow* merupakan teknik yang digunakan dalam penelitian ini untuk membedakan antara bayangan dan objeknya yang mirip bayangan. Berdasarkan tabel di atas, nilai FN pada kelas label drainase yang cukup besar pada dataset P3, P4, dan P5 menunjukkan adanya kesalahan pendeteksian yang cukup banyak pada label tersebut. Hal ini dapat dikarenakan drainase yang memiliki warna dan bentuk yang mirip dengan bayangan bendera, dan bayangan mobil. Namun, memberikan nilai batas pendeteksian bayangan sebesar 0,6 menghasilkan model yang tepat memprediksi label benar pada *ground truth* yang sesuai.

Hasil implementasi pada video memberikan rata-rata *Frame per Second* (FPS) secara berurutan pada setiap model P1 hingga P5 adalah: 12.520, 12.487, 12.656, 12.829, dan 11.972. Model P5 memproses lebih sedikit frame per detik, menunjukkan bahwa kemampuan model mendeteksi dan melacak bayangan lebih lambat dibandingkan model lainnya. Meskipun berdasarkan tingkat akurasi, model P5 memiliki rata-rata akurasi yang paling tinggi.

SIMPULAN

Beberapa kesimpulan yang dapat ditarik berdasarkan hasil evaluasi dan analisa adalah:

1. Dengan memberikan variasi dataset, hasilnya menunjukkan bahwa semakin heterogen dataset yang digunakan maka semakin tinggi tingkat akurasi pendeteksian bayangan dengan akurasi *F1-score* tertinggi terdapat pada model P5 sebesar 46%.
2. Dengan teknik mendefinisikan masing-masing label objek *non-shadow* dan *shadow*, berhasil dilakukan namun masih kurang efektif pada bayangan dan objek dengan warna dan bentuk yang sangat mirip.

REFERENSI

- [1] A. Fajar *et al.*, "Identifikasi Kerusakan Jalan 34 dengan Metode Faster R- CNN Studi Kasus di Jalan Pakansari Bogor Jawa barat," *Smart Comp: Jurnalnya Orang Pintar Komputer*, vol. 11, no. 2, pp. 247-256, 2022, doi: 10.30591/smartcomp.v11i2.3509.
- [2] A. Yoneyama *et al.*, "Moving cast shadow elimination for robust vehicle extraction based on 2D joint vehicle/shadow models," *Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance*, Miami, USA, pp. 229-236, 2003, doi: 10.1109/AVSS.2003.1217926.
- [3] B. Chougula *et al.*, "Road segmentation for autonomous vehicle: A review," *3rd International Conference on Intelligent Sustainable Systems (ICISS)*, Thoothukudi, India, pp. 362-365, 2020, doi: 10.1109/ICISS49785.2020.9316090.
- [4] D. Kim, M. Arsalan, K. Park, "Convolutional Neural Network-Based Shadow Detection in Images Using Visible Light Camera Sensor," *Sensors*, vol. 18, no. 4, p. 960, 2018, doi: 10.3390/s18040960.
- [5] F. Charli *et al.*, "Implementasi Metode Faster Region Convolutional Neural Network (Faster R-CNN) Untuk Pengenalan Jenis Burung Lovebird," *Journal of Information Technology Ampera*, vol. 1, no. 3, pp. 185-197, 2020.
- [6] J. Bao *et al.*, "ShadowDeNet: A Moving Target Shadow Detection Network for Video SAR," *Remote Sensing*, vol. 14, no. 2, p. 320, 2022, doi: 10.3390/rs14020320.

- [7] J. Pardede, H. Hardiansah, "Deteksi Objek Kereta Api menggunakan Metode Faster R-CNN dengan Arsitektur VGG 16," *MIND Journal*, vol. 7, no. 1, pp. 21-36, 2022, doi: 10.26760/mindjournal.v7i1.21-36.
- [8] L. Qu *et al.*, "Evaluation Of Shadow Features," *IET Computer Vision*, vol. 12, no. 1, p. 95-103, 2017, <https://doi.org/10.1049/iet-cvi.2017.0159>.
- [9] Lee *et al.*, "Moving Shadow Detection from Background Image and Deep Learning," in: Huang, F., Sugimoto, A. (eds) *Image and Video Technology – PSIVT 2015 Workshops, PSIVT 2015*, Lecture Notes in Computer Science, vol. 9555, 2016.
- [10] S. Brutzer, B. Hoferlin, G. Heidemann, "Evaluation of background subtraction techniques for video surveillance," in *CVPR*, Providence, RI, pp. 1937-1944, 2011.
- [11] S. Ren *et al.*, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149, 2017, doi: 10.1109/TPAMI.2016.2577031.
- [12] Sarda *et al.*, "Object Detection for Autonomous Driving using YOLO [You Only Look Once] algorithm," in *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, pp. 1370-1374, 2021.
- [13] T. M. Hoang *et al.*, "Road Lane Detection Robust to Shadows Based on a Fuzzy System Using a Visible Light Camera Sensor," *Sensors*, vol. 17, no. 11, p. 2475, 2017, doi: 10.3390/s17112475.