

UJIAN TENGAH SEMESTER

IMAGE PROCESSING

Semester Genap T.A. 2020/2021

Dosen Pengampu : Dr. INDAH SOESANTI, S.T., M.T.



Object Detection dengan Deep Learning dan Penerapannya pada Self-Driving Car

Oleh

Nama: Salman Rahwidean Janotama

NIM: 18/428633/TK/47135

Fakultas Teknik
Universitas Gadjah Mada
2021

INTISARI

Deep Learning menjadi salah satu teknologi yang terus dikembangkan. Penggunaannya untuk Self-Driving Car menjadi salah satu alasan pengembangan Deep Learning khususnya Object Detection semakin pesat. Untuk mewujudkan SAE Level 5 dimana self-driving car sudah tidak ada campur tangan manusia sama sekali beragam metode object detection dikembangkan seperti R-CNN Family (R-CNN, Fast R-CNN, dan Faster R-CNN), SSD, dan YOLO. Metode yang digunakan adalah dengan studi literature dan juga eksperimen dengan 50 citra gambar per class dengan 40 citra untuk training dan 10 citra untuk testing. Penerapan Faster R-CNN dapat meningkatkan kecepatan komputasi dan mempersingkat waktu komputasi ditambah dengan Inception V2 sebagai feature extractornya. Open Image Dataset V6 digunakan untuk ujicoba dengan mengambil subset class Person, Car, dan Motorcycle. Dari hasil training didapatkan sebuah model dengan metode Faster R-CNN dengan Inception V2 yang dapat mendeteksi object pada citra gambar maupun video dengan nilai learning rate 0.0002.

Kata kunci: Faster R-CNN, Open Image, Self-Driving Car

DAFTAR ISI

INTISARI.....	2
DAFTAR ISI.....	3
PENDAHULUAN.....	4
A. Latar Belakang.....	4
B. Tujuan	5
KAJIAN PUSTAKA.....	6
A. Faster R-CNN.....	6
B. Inception V2.....	7
C. Open Image Dataset V6	7
METODOLOGI.....	8
A. Alat dan Bahan	8
B. Langkah Uji.....	8
C. Metode.....	8
HASIL DAN ANALISIS	9
A. Data Acquisition dan Preprocessing	9
B. Training dan Model	9
KESIMPULAN	10
A. Simpulan	10
B. Saran	10
DAFTAR PUSTAKA	11

PENDAHULUAN

A. Latar Belakang

Perkembangan *artificial intelligence* (AI), *machine learning* (ML), dan *deep learning* (DL) sudah semakin luas. Berbagai penerapan dari teknik ini sudah dilakukan, salah satu diantaranya adalah *self-driving car*. Konsep dari *self-driving car* itu sendiri sebenarnya sudah ada selama 80 tahun, diperkenalkan pertama kali pada acara *World's Fair* tahun 1939 di New York, Amerika Serikat oleh Futurama milik General Motor (GM). Saat ini perusahaan-perusahaan besar seperti Google, Tesla Motors, dan General Motors berlomba-lomba untuk menciptakan *self-driving car* terbaik (Gupta, Anpalagan, Guan, & Khwaja, 2021). Pada prinsipnya terdapat enam tingkatan dari kendaraan otomatis (*vehicular automation*) yang didefinisikan oleh *Society of Automotive Engineers* (SAE). Mulai dari SAE Level 0 hingga SAE Level 3 yang merupakan kolaborasi antara AI dan manusia hingga SAE Level 4 dan SAE Level 5 yang sepenuhnya otomatis dan bergantung pada AI.

- 1) SAE Level 0 (L0), tanpa otomatisasi sama sekali.
- 2) SAE Level 1 (L1), *Driving assistance* seperti *automatic braking* dan *cruise control*.
- 3) SAE Level 2 (L2), *Partial automation* seperti *hands-free driving* dan *driving assistance* yang lebih mutakhir lagi seperti *automatic parking*, *steering assist*, dan *emergency braking*.
- 4) SAE Level 3 (L3), *Conditional automation*, secara umum kendaraan dapat memiliki kontrol penuh pada kemudi dalam beberapa kondisi tertentu.
- 5) SAE Level 4 (L4), *Highly automated*, hampir sepenuhnya otomatis dan tidak ada interaksi dengan manusia, hanya saja pengemudi tetap memiliki kendali jika diperlukan
- 6) SAE Level 5 (L5) *Fully autonomous* dalam seluruh keadaan, waktu, dan lokasi. Pada tingkat ini tidak ada interaksi manusia sama sekali.

Salah satu teknologi yang memungkinkan kendaraan untuk bisa dijalankan secara otomatis adalah dengan penerapan object recognition. Pada dasarnya object recognition dibagi menjadi beberapa segmen, image identification (apakah ini bangunan?), image classification (gambar apakah ini?), object detection (dimanakah pejalan kaki?), scene understanding (adegan apakah ini?), dan specific object recognition (apakah lampu merah itu menandakan 'berhenti'?) (Fujiyoshi, Hirakawa, & Yamashita, 2019). Obejct recognition yang digunakan untuk *self-driving car* umumnya adalahh object detection, dengan tujuan untuk mengetahui objek apakah itu dan dimana letak object tersebut. *Self-driving car* perlu mengetahui kondisi sekitar dengan mengetahui object apa yang ada didepannya, apakah motor, mobil, sepeda, bus, dan pejalan kaki. *Self-driving car* juga harus mengetahui letak pasti dari object tersebut apakah tepat didepan mobil, disamping, atau dibelakang. Banyak metode object detection yang diterapkan pada *self-driving car* seperti SSD, R-CNN family, dan juga YOLO. Beberapa parameter yang harus diperhatikan oleh *self-driving car* antara lain adalah jumlah class yang dapat

diuji dan seberapa tinggi akurasi serta level of confident nya. Semakin banyak class yang dapat dideteksi maka kendaraan akan semakin aman. Dengan adanya perkembangan pada Deep Learning khususnya object detection, SAE Level 5 bukan lagi menjadi sebuah konsep saja melainkan dapat dipraktekan secara nyata dan lebih lagi bisa mulai diproduksi secara masal.

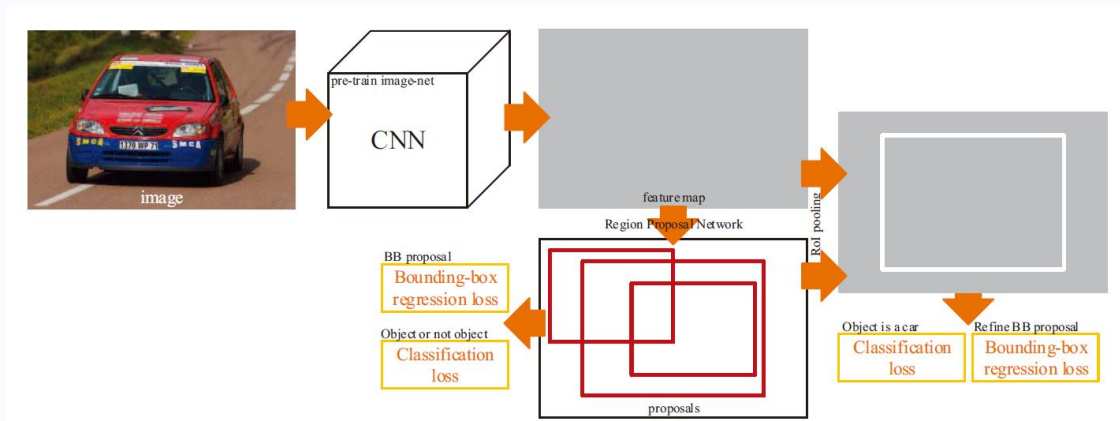
B. Tujuan

Bekaitan dengan uraian latar belakang diatas, tujuan yang ingin dicapai didalan karya ini adalah melakukan percobaan Deep Learning khususnya Object Detection dengan metode Faster R-CNN dan penerapannya pada Self-driving car.

KAJIAN PUSTAKA

A. Faster R-CNN

Faster R-CNN merupakan salah satu dari varian dari metode R-CNN disamping model R-CNN dan Fast R-CNN. Faster R-CNN memiliki dua module utama, yaitu deep fully convolutional network dan Fast R-CNN. Pengembangan yang dilakukan pada varian ini adalah dengan mengganti metode selective search algorithm yang relative lambat dengan fast neural network, atau lebih dikenal dengan istilah Region Proposal Network (RPN) (Cai, Li, Zhongzhao, Zhao, & LU, 2018).



Gambar 1. Struktur dari Faster R-CNN

Komputasi dari Faster R-CNN untuk object detection jauh lebih cepat dibandingkan dengan metode lain di keluarga R-CNN. Seperti pada Gambar 2 test time per image (with proposals) yang dibutuhkan oleh Faster R-CNN hanya 0.2 detik, jauh dibandingkan dengan Fast R-CNN selama 2 detik dan R-CNN selama 50 detik dengan nilai mAP pada VOC 2007 sebesar 66.9 (Dixit, Chadaga, Savalgimath, Rakhith, & R, 2019).

	R-CNN	Fast R-CNN	Faster R-CNN
Test time per image (with proposals)	50 seconds	2 seconds	0.2 seconds
(Speedup)	1x	25x	250x
mAP (VOC 2007)	66.0	66.9	66.9

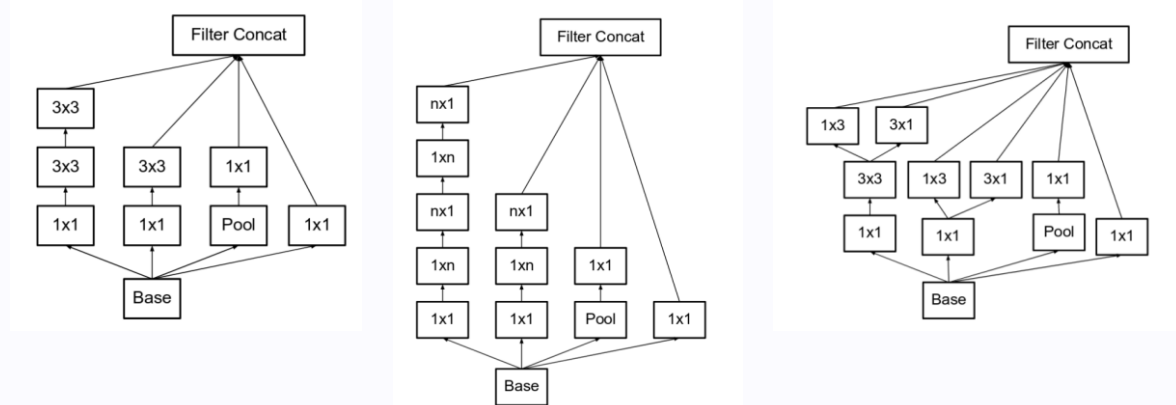
Gambar 2. Perbandingan komputasi R-CNN Family

Region proposals didapat dengan menggeser sebuah window diatas convolutional feature map output dari sebuah convolutional layer yang dipilih didalam network pada posisi yang berbeda ($n \times n$ grid).

Pada setiap pergeseran lokasi window, bounding box proposals yang banyak dibuat dengan menggunakan sejumlah k anchor boxes (Kalliomaki, 2019).

B. Inception V2

Inception V2 adalah pengembangan dari Inception V1 dengan menyelesaikan beberapa masalah yang ditemui pada Inception V1 Architecture. Pada Inception V2 Architecture, 5×5 convolution diganti dengan dua buah 3×3 convolution yang menyebabkan berkurangnya waktu komputasi dan meningkatkan kecepatan komputasi karena 5×5 convolution 2.78 kali lebih lama dibandingkan dengan 3×3 convolution. Pada architecture ini juga merubah $n \times n$ factorization menjadi $1 \times n$ dan $n \times 1$ factorization sehingga mempermudah proses komputasi yang awalnya kompleks. Untuk mengatasi permasalahan representational bottleneck pada Inception V1, feature banks dari modul di perbesar bukan diperdalam sehingga dapat mencegah loss information yang terjadi bisa dibuat semakin dalam (pawangfg, 2020).



Gambar 3. Inception V2

C. Open Image Dataset V6

Open image adalah dataset dengan lebih dari sembilan juta gambar yang telah di anotasi dengan image-level labels, object bounding boxes, object segmentation maske, visual relationships, dan localized narrative. Total terdapat 16 juta bounding boxes untuk 600 object classes pada 1,9 juta gambar yang membuat open image menjadi dataset dengan object location annotation terbesar didunia. Box yang ada secara manual dibuat oleh para anotator profesional untuk menjaga akurasi dan konsistensi. Gambar yang disediakan sangat beragam dan mengandung beragam scenario yang kompleks dengan beberapa object (rerata 8.3 per gambar). Open images juga menyeduk visual relationship annotations yang dapat menandakan hubungan antar object seperti “perempuan yang bermain gitar” dan “minuman di atas meja” serta object properties seperti “meja dari kayu” dan “perempuan yang sedang loncat” dengan total 3,3 juta anotasi dari 1.466 relationship triplets yang berbeda.

METODOLOGI

A. Alat dan Bahan

Pada percobaan object detection dengan deep learning ini digunakan alat dan bahan antara lain:

Alat dan Bahan	Fungsi
Open Image Dataset V6	Dataset untuk training model
Google Colab	Software untuk data acquisition, preprocessing, training dan testing
Google Drive	Cloud storage untuk menyimpan file proyek

B. Langkah Uji

Dalam proses pengujian ini Langkah yang dilakukan mengikuti pada AI Project Cycle, yaitu:

1) Problem Scoping

Pada tahap ini masalah yang diangkat adalah tentang bagaimana memberikan gambaran pada kendaraan otomatis untuk bisa membedakan object disekitarnya, object yang diambil untuk percobaan ini adalah Person, Car, dan Motorcycle.

2) Data Acquisition

Pada tahap ini class object yang ingin di deteksi diambil dari dataset Open Image Dataset V6 untuk subset Person, Car, dan Motorcycle.

3) Data Exploration

Pada tahap ini data yang sudah didapat diformat ulang sebelum siap untuk dimasukan kedalam proses training

4) Modeling

Pada tahap ini digunakan model pretrained dari tensorflow untuk metode Faster R-CNN dengan backbone Inception V2 dan dataset Open Image Dataset V6

5) Deployment

Setelah proses training model yang sudah jadi di tes dengan dua buah citra masukan berupa gambar dengan berbagai kualitas serta video dengan durasi mulai dari 8 detik, 20 detik, dan 3 menit.

C. Metode

Metode yang digunakan pada percobaan ini adalah dengan studi literatur untuk mengetahui bagaimana kerja dari Faster R-CNN dengan Inception V2 dan eksperimen untuk membuat secara langsung model dari Deep Learning ini.

HASIL DAN ANALISIS

A. Data Acquisition dan Preprocessing

Proses akuisisi data dilakukan dengan menggunakan tools dari Github untuk mendownload subset class Person, Car, dan Motorcycle. Citra gambar dan label yang didownload dari Open Image Dataset berjumlah 50 gambar per class nya, sehingga total ada 150 gambar yang digunakan. Gambar ini nantinya dibagi menjadi 40 untuk training dan 10 untuk testing per classnya. Setelah seluruh data didownload dan di split, maka dilakukan proses untuk merubah format dari label yang digunakan. Hasil dari Open Image label dengan ekstensi *.txt dirubah menjadi *.xml dengan tools dari Github yang kemudian dibuat *.csv dari seluruh label yang ada. Setelah data siap maka proses akan dilanjutkan pada proses training.

B. Training dan Model

Pada proses ini digunakan pretrained model dari tensorflow dengan metode Faster R-CNN dan backbone Inception V2. Tensorflow yang digunakan adalah versi 1.15.0 dengan accelerator GPU pada Google Colab. Proses training berlangsung selama kurang lebih satu setengah jam hingga step 30.656 hingga nilai loss dibawah 0.05 dan proses training dihentikan. Setelah proses training selesai maka didapat output file berupa graph dengan ekstensi *.pb yang berisi trained model dengan learning rate 0.0002. Dari TensorBoard didapatkan beberapa grafik terlampir. Hasil dari model kemudian di test dengan citra gambar sebanyak tiga kali dan dengan video sebanyak 3 kali dengan variasi gambar berbagai resolusi dan video berbagai durasi, resolusi, dan fps. Hasil dari citra gambar dapat dilihat pada Gambar 4, sedangkan untuk video terlampir.



Gambar 4. Hasil tes model dengan citra gambar

KESIMPULAN

A. Simpulan

Berdasarkan percobaan tersebut didapat beberapa kesimpulan, diantaranya adalah metode Faster R-CNN dengan Inception V2 menunjukkan hasil yang memuaskan untuk citra gambar dan video dengan variasi berbagai kualitas hanya dengan proses training data yang sedikit serta durasi yang tidak lama dan sangat berpotensi untuk dikembangkan sebagai dasar dari self-driving car.

B. Saran

Pada implementasi yang lebih jauh banyak hal yang bisa di tingkatkan dari model ini, seperti jumlah class object yang lebih beragam, jumlah data yang lebih besar lagi dan proses training yang lebih lama. Sehingga model dapat diterapkan pada Self-driving car.

DAFTAR PUSTAKA

- Cai, W., Li, J., Zhongzhao, X., Zhao, T., & LU, K. (2018). Street Object Detection Based on Faster R-CNN. *Chinese Control Conference*, (p. 9500). Wuhan.
- Dixit, K. G., Chadaga, M. G., Savalgimath, S. S., Rakhith, G. R., & R, N. K. (2019). Evaluation and Evolution of Object Detection Techniques YOLO and R-CNN. *International Journal of Recent Technology and Engineering (IJRTE)*, 824.
- Fujiyoshi, H., Hirakawa, T., & Yamashita, T. (2019). Deep learning-based image recognition for autonomous driving. *IATSS Reasearch*.
- Gupta, A., Anpalagan, A., Guan, L., & Khwaja, A. S. (2021). Deep learning for object detection and scene perception in self-driving cars: Survey, challenges, and open issues. *Array*.
- Kalliomaki, R. (2019). *Real-time object detection for autonomous vehicles using deep learning*. Uppsala: Uppsala Universitet.
- pawangfg. (2020, 7 16). *Inception V2 and Inception V3 - Inception Network versions*. Retrieved from GeeksforGeeks.org: <https://www.geeksforgeeks.org/inception-v2-and-v3-inception-network-versions/>