

Modifikasi Arsitektur Single Shoot Multibox Detector Untuk Deteksi Penggunaan Masker

Bayu Krisna Mukti¹

Program Studi Teknik Elektro, Fakultas Teknik dan Informatika, Univeritas PGRI Semarang
 Jl. Sidodadi Timur No.24-Dr.Cipto, 50232, Semarang
 Email: bayukrisnam21@gmail.com

Abstrak— berakhirnya pandemi covid-19 berimbas pada monitoring penggunaan masker yang semakin diabaikan, meskipun memakai masker dapat mengurangi resiko penularan. Memanfaatkan Artificial Intelligence untuk melakukan object detection menjadi salah satu solusi agar monitoring tetap dapat dilaksanakan, salah satunya dengan metode Single Shoot Multibox Detector (SSD) [1]. Arsitektur SSD memiliki 3 layer utama yaitu, base network, extra convolutional feature layers, dan convolutional predictor layer. Convolutional predictor layer menghasilkan 8732 deteksi untuk tiap kelas, dari hasil kombinasi base network dan extra convolutional feature layer dengan default boxes. Pada penelitian ini, arsitektur SSD dilakukan modifikasi (SSD v2) pada convolutional predictor layer yang hanya mendapat input dari extra convolutional feature layers, sehingga deteksi yang dihasilkan berkurang menjadi 790 deteksi. Dari hasil training dan testing dalam melakukan deteksi masker, SSD v2 mempunyai mAP sebesar 87,23%, sedangkan SSD mempunyai mAP sebesar 92,79%. Akan tetapi, SSD v2 memiliki nilai loss yang lebih kecil yaitu sebesar 1,03, sedangkan SSD memiliki nilai loss 1,39.

Kata kunci: Masker; Artificial Intelligence; Object detection; SSD

I. PENDAHULUAN

Berakhirnya pandemi covid-19 membuat masyarakat semakin lalai untuk tetap memakai masker guna mengurangi resiko penularan, apalagi dengan tidak adanya pengawasan yang dilakukan. Memanfaatkan artificial intelligence untuk melakukan deteksi objek menjadi salah satu metode untuk meningkatkan pengawasan tersebut. dalam object detection terdapat salah satu metode, yaitu single shoot multibox detector (SSD) [1]. yang berbasis convolutional neural network (CNN).

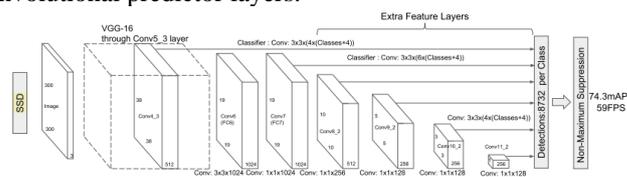
Arsitektur SSD memiliki 3 layers utama, yaitu base network, extra convolutional feature layers, dan convolutional predictor layer. Base network SSD menggunakan arsitektur VGG16 [2] sebagai feature extractor, kemudian extra convolutional feature layers merupakan layer tambahan yang mempunyai fungsi yang sama seperti base network, dan convolutional predictor layer yang digunakan untuk melakukan deteksi dengan teknik konvolusi.

Pada dasarnya, input convolutional predictor layer SSD berasal dari output 2 layer base network dan 4 extra convolutional feature layers. dalam penelitian [c arsitektur SSD juga dilakukan modifikasi bagian convolutional predictor layers dengan mengurangi sampai meniadakan input dari extra convolutional feature layer, dan menghasilkan nilai mAP yang beragam. Pada penelitian ini, penulis akan mencoba memodifikasi bagian output layer, dengan cara meniadakan output dari 2 layer base network, dan hanya menggunakan output dari 4 extra convolutional feature layers, lalu membandingkan hasilnya dengan arsitektur SSD yang asli dalam melakukan deteksi masker.

II. STUDI PUSTAKA

A. Single Shoot Multibox Detector

Single Shot Multibox Detector (SSD) adalah sebuah metode dalam object detection pada gambar yang berbasis regresi atau klasifikasi. Terdapat 3 layer utama dalam SSD, yaitu base network, extra convolutional feature layers, dan convolutional predictor layers.



Gambar 1. Arsitektur SSD [1]

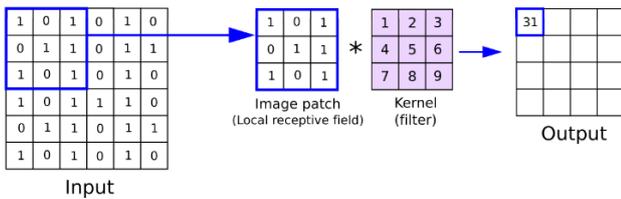
Pada Convolutional predictor layers output dari base network dan extra convolutional feature layer dihubungkan dengan default boxes yang memiliki skala, aspect ratio, dan jumlah yang berbeda-beda. Dengan filter konvolusi, setiap layer yang menjadi input dari convolutional predictor layer dapat menghasilkan sekumpulan prediksi yang kemudian dicocokkan dengan ground truth boxes. SSD memiliki Total 8732 deteksi untuk setiap kelas, yang kemudian dilakukan proses non-maximum suppression sehingga menghasilkan 200 prediksi teratas.

B. Convolutional Neural Network (CNN)

CNN merupakan sebuah system pengolahan informasi citra yang menggunakan teknik konvolusi. Arsitektur CNN terdiri dari beberapa layer, yaitu convolutional layer, pooling layer, dan fully connected layer.

1. Convolutional layer

Convolutional layer menjadi bagian paling penting dalam CNN. dalam layer ini dilakukan proses konvolusi, dimana kumpulan kernel atau filter yang diterapkan di seluruh area input data [3].



Gambar 2. Proses konvolusi [4]

2. Pooling layer

Ide utama dalam proses pooling di dalam pooling layer adalah meringkas informasi yang didapat dari convolutional layer [5]. Terdapat dua teknik pooling yang biasanya digunakan dalam CNN, yaitu average pooling dan max pooling. Average pooling mencari rata-rata nilai dari sebuah matrik hasil convolutional layer menggunakan sebuah filter, sedangkan max pooling mencari nilai tertinggi dari matrik tersebut.

3. Fully connected layer

Fully connected layer merupakan lapisan paling ujung dari arsitektur CNN, dimana semua neuron saling terhubung. Sebelum memasuki fully connected layer terlebih dahulu dilakukan proses flatten, hingga menghasilkan output yang kemudian diolah dengan fungsi aktivasi seperti softmax atau sigmoid untuk menjadi hasil akhir dari proses CNN (misalnya klasifikasi)

III. METODE/DESAIN

Tujuan dari penelitian ini adalah untuk mencari perbandingan akurasi arsitektur SSD [1] dengan arsitektur SSD yang telah dimodifikasi (SSD v2) pada bagian convolutional predictor layer, dalam melakukan deteksi masker. Secara garis besar proses yang terjadi dalam penelitian ini adalah sebagai berikut:

A. Preprocessing dataset

Dalam penelitian ini terdapat dua kelas yang harus dideteksi model SSD, yaitu citra object orang menggunakan masker dan tidak, oleh karena itu, Dataset yang digunakan dalam penelitian ini berupa 2000 citra orang menggunakan masker dan tidak menggunakan masker yang didapat dari google image. proses preprocessing dataset meliputi :

4. Labelling

Dataset citra yang telah dikumpulkan dilakukan proses pelabelan dengan memberi kotak pembatas terhadap object yang terdapat pada image sesuai kelas 'masked' atau 'un masked'. dalam proses training SSD kotak pembatas tersebut akan digunakan sebagai bounding boxes.

5. Dataset Splitting

Dataset yang telah melalui proses pelabelan dipisah secara acak untuk proses trainig dan testing dengan dengan perbandingan 80% data training dan 20% data testing.

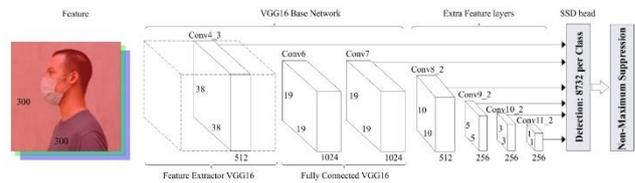
6. Transform

Transform merupakan proses augmentasi data, dengan tujuan untuk membuat dataset lebih beragam. Proses

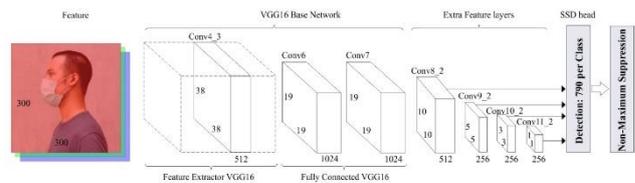
augmentasi data yang dilakukan meliputi random crop, expand, flip, resize, dan photometric distort.

B. Perancangan model

Kedua model yang dirancang dalam penelitian ini memiliki perbedaan pada convolutional predictor layer. Pada gambar 1, convolutional predictor layer mendapat 2 input dari base network dan 4 inputan dari extra convolutional feature layer, sedangkan dalam penelitian ini convolutional predictor layer hanya mendapat 2 input dari base network saja. Berikut merupakan 2 arsitektur model SSD yang digunakan dalam penelitian ini:



Gambar 3. Arsitektur SSD



Gambar 4. Arsitektur SSD v2

C. Training dan testing

Proses training dan testing dilakukan secara bersamaan menggunakan sebuah parameter early stop. Dalam proses training, weight akan terus dilakukan update menggunakan metode backpropagation [6] dan optimizer [7], sedangkan dalam proses testing weight yang didapat pada proses training tiap epoch akan langsung digunakan untuk proses testing. Dari proses testing inilah didapat nilai akurasi dari model.

D. Analisa hasil training dan testing

Setelah dilakukan proses training dan testing, selanjutnya dilakukan Analisa perbandingan akurasi berdasarkan hasil testing dan implementasi dari masing-masing model. Dalam penelitian ini, nilai akurasi model didapat menggunakan matrik mean average precision (mAP) [8].

IV. HASIL DAN PEMBAHASAN

Penelitian ini dilakukan untuk membandingkan 2 model SSD dalam melakukan deteksi masker. parameter-parameter yang digunakan mulai dari batch size, optimizer, learning rate, weight decay, momentum, sampai nilai early stop untuk kedua model sama, yang menjadi perbedaan hanyalah bagian convolutional predictor layer. Berikut merupakan nilai berbagai parameter yang digunakan dalam penelitian ini:

Tabel 1. Parameter training

Parameter	Nilai / Jenis
Batch Size	32
Optimizer	SGD
Learning Rate	$10^{-3}, 10^{-4}, 10^{-5}$
Early Stop	10, 15, 20
Weight Decay	5×10^{-4}
Momentum	0,9

Setiap input convolutional predictor layer dihubungkan dengan default boxes dengan jumlah yang berbeda-beda. Oleh karena itu, jumlah deteksi yang dihasilkan dari masing-masing input akan berbeda-beda, sesuai dengan ukuran dimensi dan jumlah default box-nya. Berikut merupakan jumlah deteksi yang didapat dari masing-masing layers yang digunakan sebagai input untuk convolutional predictor layers:

Tabel 2. Jumlah deteksi tiap layer

Layers Name	Default Boxes	Dimensi Output	Jumlah Deteksi
Conv4_3	4	38 x 38	5776
Conv7	6	19 x 19	2166
Conv8_2	6	10 x 10	600
Conv9_2	6	5 x 5	150
Conv10_2	4	3 x 3	36
Conv11_2	4	1 x 1	4

A. Training dan Testing

Proses training dan testing dilakukan secara bersamaan tanpa menggunakan Batasan epoch. Proses training dan testing akan berhenti apabila telah mencapai nilai early stop. Berikut merupakan hasil dari proses training dan testing dari masing-masing model:

Tabel 3. Hasil training dan testing

Model Arsitektur	Jumlah Deteksi	Jumlah Epoch	Loss	Mean Average Precision
SSD	8732	128	1,39	0,9279
SSD v2	790	193	1,03	0,8723

Seperti yang ditunjukkan dalam gambar 1, convolutional predictor layer mendapatkan input dari 6 layer, dimana 2 layer dari base network dan 4 layer dari extra feature layers sehingga menghasilkan 8732 deteksi, sedangkan pada gambar 2 model hasil modifikasi hanya mendapat 4 input saja dari extra feature layers, sehingga hanya menghasilkan 790 deteksi.

Menggunakan nilai early stop yang sama, SSD v2 membutuhkan jumlah epoch yang lebih banyak untuk mencapai akurasi terbaiknya, yaitu sebanyak 193 epoch. Sedangkan SSD hanya membutuhkan 128 epoch untuk mencapai akurasi terbaiknya. Meskipun memiliki jumlah epoch yang lebih banyak, berdasarkan hasil evaluating model SSD v2 memiliki akurasi yang lebih rendah yaitu sebesar 87,23% dibanding dengan SSD yang memiliki akurasi sebesar 92,79% dengan jumlah epoch yang lebih sedikit. Meskipun demikian, SSD v2 memiliki loss yang lebih sedikit dibanding dengan SSD, yang mana SSD v2 mempunyai loss terbaik sebesar 1,03, sedangkan SSD memiliki loss terbaik sebesar 1,39.

Nilai loss dari masing-masing model didapat dari gabungan localization loss dan confidence loss. Localization loss adalah nilai loss yang didapat dari ketidakcocokan antara ground truth box dengan kotak prediksi, dimana model menggunakan Smooth L1 loss dalam prosesnya.

Sedangkan confidence loss merupakan loss yang dihasilkan pada saat melakukan prediksi class menggunakan Softmax loss

B. Implementasi Model

Setelah dilakukan proses training dan testing, selanjutnya dilakukan implementasi model menggunakan 10 data baru. Berikut merupakan salah satu contoh hasil dari implementasi masing-masing model:



Gambar 5. Implementasi SSD



Gambar 6. Implementasi SSD v2

Sama seperti training dan testing, akurasi pada saat implementasi model juga diperoleh dengan matrik mAP. Untuk itu terlebih dahulu dilakukan perhitungan nilai AP_c untuk masing-masing kelas c dengan mencari rata-rata dari 11 titik interpolate P_r , dengan P sebagai nilai precision dan r merupakan 11 titik recall dengan rentang 0 sampai 1.

$$AP_c = \frac{(P_r(0) + \dots + P_r(1.0))}{11} \tag{1}$$

Setelah didapat nilai AP_c untuk tiap kelas, maka didapat nilai mAP model dengan mencari rata-rata AP_c untuk semua kelas .

$$mAP = \frac{AP_{c1} + \dots + AP_{cn}}{n \text{ class}} \tag{2}$$

Berdasarkan persamaan (1) dan (2), diperoleh hasil dari implementasi untuk masing-masing model sebagai berikut:

Tabel 4. Hasil implementasi model

Model Arsitektur	Average Precision		Mean Average Precision
	Not Masked	Masked	
SSD	0,56	0,86	0,71
SSD v2	0,73	0,73	0,73

Dengan demikian akurasi model SSD pada saat implementasi model sebesar 71%, sedangkan SSD v2 mempunyai nilai akurasi sebesar 73%.

V. KESIMPULAN

Dari hasil training dan testing, diperoleh kesimpulan bahwa model SSD[1] memiliki akurasi yang lebih baik dalam melakukan deteksi masker, yaitu sebesar 92,79%. Sedangkan SSD v2 hasil modifikasi mempunyai akurasi 87,23%.

Akan tetapi, ketika dilakukan implementasi model, SSD v2 memiliki akurasi lebih baik yaitu 73%, sedangkan SSD[1] memiliki akurasi 71%. Meskipun demikian, SSD[1] memiliki akurasi yang lebih baik dalam mendeteksi masked class daripada SSD v2.

REFERENSI

[1] W. Liu dkk., "SSD: Single shot multibox detector," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes*

- Bioinformatics*), vol. 9905 LNCS, hal. 21–37, 2016.
- [2] F. O. R. L. Arge dan C. I. Mage, “VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION,” hal. 1–14, 2015.
- [3] I. Vasilev, D. Slater, G. Spacagna, P. Roelants, dan V. Zocca, Python Deep Learning. 2019.
- [4] A. H. Reynolds, “Convolutional Neural Network (CNNs),” <https://anhreynolds.com>, 2019. [Daring]. Tersedia pada: <https://anhreynolds.com/blogs/cnn.html>. [Diakses: 08-Jun-2023].
- [5] J. W. G. Putra, “Artificial Neural Network,” 2020.
- [6] Hecht-Nielsen, “Theory of the backpropagation neural network,” in International 1989 Joint Conference on Neural Networks, 1989, hal. 593–605 vol.1.
- [7] S. Ruder, “An overview of gradient descent optimization,” hal. 1–14, 2016.
- [8] J. Hui, “mAP (mean Average Precision) for Object Detection,” <https://jonathan-hui.medium.com/>, 2018. [Daring]. Tersedia pada: <https://jonathan-hui.medium.com/map-mean-average-precision-for-object-detection-45c121a31173>. [Diakses: 13-Jun-2023].