Modifikasi Default-Boxes Pada Model SSD Untuk Meningkatkan Keakuratan Deteksi

Muhammad Arfina Afwani[1], Ema Utami[2], Eko Pramono[3]
[1,2,3]Magister Teknik Informatika Universitas AMIKOM Yogyakarta
Jl. Ring Road Utara, Condong Catur, Depok, Sleman, Yogyakarta 55281
Email : arfina.afwani@gmail.com[1], ema.u@amikom.ac.id[2], eko.p@amikom.ac.id[3],

Abstract

Modified Default-boxes In SSD Model To Increase Accuracy Detection is a research to improve the detection accuracy of the Single Shot Multibox Detector method. In this study the default-boxes are modified with an additional aspect ratio of 2:3 and 3:2 . The dataset used for this research is PASCAL VOC 2007. The model used to classify objects is Single Shot Multibox Detector. The layers on the convolutional network are created by defining them in the form of a protocol buffer. To perform the training and testing process, the caffe framework is used. The conclusion of this research is that adding default-boxes can improve the overall accuracy. But in certain classes the accuracy actually decreased.

***Keywords** : SSD, CNN, Neural Network, Object Detection*

Abstrak

Modifikasi Default-boxes Pada Model SSD Untuk Meningkatkan Keakuratan Deteksi adalah penelitian untuk meningkatkan keakuratan pendeteksian metode Single Shot Multibox Detector. Pada penelitian ini default-boxes dimodifikasi dengan tambahan aspek rasio 2:3 dan 3:2.Dataset yang digunakan untuk penelitian ini adalah PASCAL VOC 2007. Model yang digunakan untuk mengklasifikasikan objek adalah Single Shot Multibox Detector. Layer pada convolutional network dibuat dengan mendefinisikannya dalam bentuk protocol buffer. Untuk melakukan proses training dan testing, digunakan framework caffe. Kesimpulan yang didapatkan dari penelitian ini adalah penambahan default-boxes dapat meningkatkan keakuratan secara keseluruhan. Namun pada kelas-kelas tertentu keakuratan justru mengalami penurunan.

***Kata kunci:** SSD, CNN, neural network, pendeteksian objek*

1. PENDAHULUAN

Convolutional Neural Network (CNN) telah menjadi metode defacto untuk penyelesaian permasalahan pendeteksian objek sejak [1] membuktikan keakuratan pendeteksian yang cukup signifikan dibandingkan metode lain. Sejak saat itu penelitian CNN untuk mendeteksi objek berkembang pesat.

Penelitian mutakhir pendeteksian objek dengan variasi pendekatan *bounding boxes*, sampel ulang piksel atau *feature* untuk masing-masing *box* dan pengaplikasian *classifier* berkualitas tinggi telah menduduki peringkat teratas pada PASCAL VOC[2],

COCO[3], dan ILSVRC[4].

SSD[5] merupakan salah satu metode yang menerapkan *bounding boxes* untuk memperkirakan lokalisasi objek yang dideteksi. Keakuratan SSD lebih baik dibandingkan dengan YOLO dan Faster R-CNN.

Pada SSD perkiraan lokalisasi diterapkan dengan default-box. Penelitian ini mencoba menambahkan default-boxes menjadi 6 pada setiap feature map. Dari penelitian ini ditemukan bahwa perubahan defatul-boxes dapat meningkatkan keakuratan.

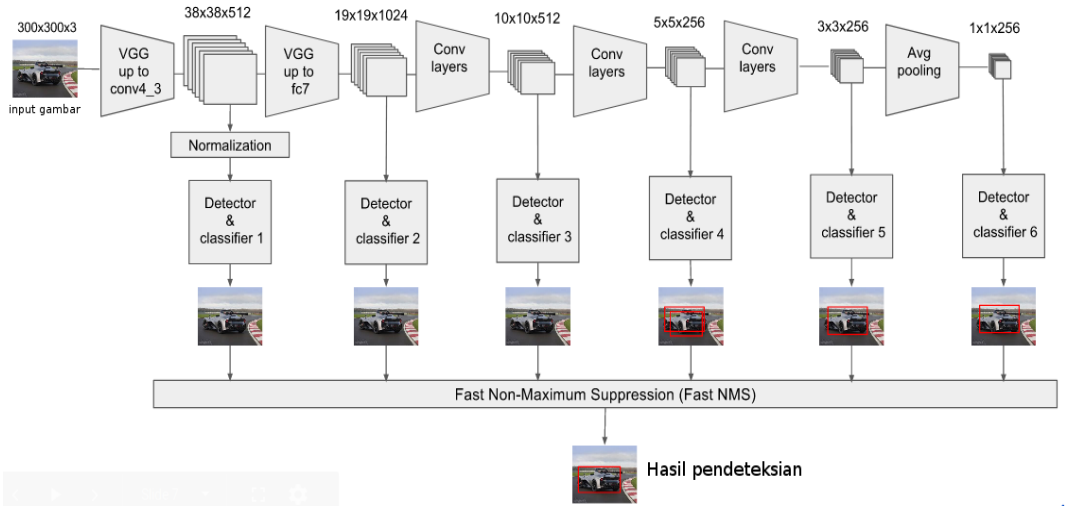
2. METODE PENELITIAN

Pada penelitian ini pengambilan data dilakukan dengan metode observasi, dokumentasi dan studi pustaka. Observasi yaitu pengambilan data dengan mengamati lingkungan penelitian dalam hal ini salah satunya adalah obyek data. Hasil dari observasi yaitu didapatkannya *sample* data baik benda mati maupun binatang. Dokumentasi yaitu pengambilan data dengan mengambil data-data *sample* salah satunya adalah foto atau gambar *sample* yang dibutuhkan dalam penelitian. Studi Pustaka merupakan pengambilan data dengan mencari referensi-referensi ilmiah sebagai acuan dan untuk memperkuat penelitian.

3. HASIL DAN PEMBAHASAN

1. *Network*

Untuk menerapkan SSD, maka diperlukan pembuatan *network*. Pada penelitian ini, penulis menerapkan tool yang sama yang digunakan oleh SSD yaitu caffe. Untuk mempersiapkan caffe, perlu dibuat database dalam bentu lmdb dan pembuatan file prototxt. Network yang dibuat dapat dilihat pada Gambar 1.



Gambar 1. *Network* SSD

Secara struktur, *network* yang dibuat pada penelitian ini sama dengan, *network* SSD, namun pada layer prior box penulis mengubah dan menambahkan default-boxes untuk setiap *feature map* yang ada.

Jumlah *default-box* untuk penelitian ini dapat dilihat pada Tabel 1 dibawah ini:

Tabel 1. Tabel Dimensi Kolom, Balok, dan Dinding

Layer	Bottom	w	h	Jumlah Box	min	max
conv4_3_norm_mbox_priorbox	conv4_3_norm, data	38	38	8.664	30	60
fc7_mbox_priorbox	fc7, data	19	19	2.166	60	111
conv6_2_mbox_priorbox	conv6_2, data	10	10	600	111	162
conv7_2_mbox_priorbox	conv7_2, data	5	5	150	162	213
conv8_2_mbox_priorbox	conv8_2, data	5	5	150	213	264
conv9_2_mbox_priorbox	conv9_2, data	5	5	150	264	315
Total				11.880		

2. Testing

Default Boxes

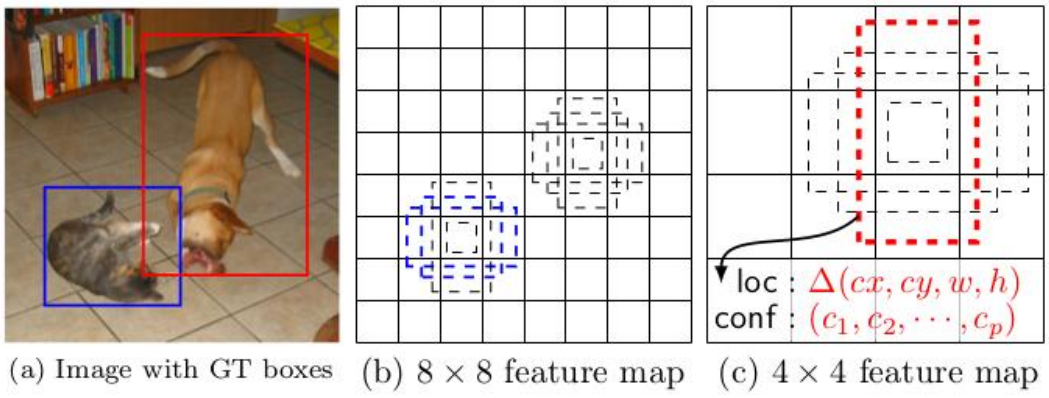
Default-box diasosiasikan untuk setiap *feature map cell* sehingga posisi relatif antara *default-box* dengan *cell* yang bersangkutan tetap. Gambar 2 merupakan ilustrasi pembuatan *default-box* dengan jumlah *default-box* 3 (1, 1:2, 2:1) untuk *feature map* 5x5. Setiap *cell* pada *feature map* tersebut akan dibuat *default-boxes* sehingga akan menghasilkan total 5x5x3 = 75 *default-boxes* untuk *feature map* tersebut.

Untuk setiap *cell* akan diprediksi skor kelas *c* dan 4 offset relatif terhadap bentuk awal *default-box*. Dengan jumlah filter *n*, jumlah *default-box k* dan ukuran *feature map w x h*, maka metode ini akan menghasilkan *(n + 4)k x w x h* hasil deteksi untuk setiap *feature map*. Offset yang dihitung adalah posisi horizontal *dx,* posisi vertikal *dy,* lebar *dw* dan tinggi *dh*.

Matching strategy

Indeks Jaccard yang juga disebut dengan *intersection over union* (IoU) digunakan untuk memisahkan *default boxes* yang cocok dengan *ground-truth box* dari yang tidak cocok.

Untuk setiap *default-box* yang memiliki IoU > 0.5 maka akan diberi label positif sedangkan yang lain dianggap negatif. Sebagai contoh, pada Gambar 2 terdapat 3 *default-boxes* positif 2 *default-boxes* positif pada kucing dan 1 *default-box* positif pada anjing.



Gambar 2. *Default box* pada SSD

Training objective

Training SSD merupakan turunan dari training *Multibox* namun dengan kemampuan untuk menangani beberapa kategori sekaligus. Misalkan $x_{ij}^{p} = \{1,0\}$ merupakan indikator untuk *matching* pada posisi *i* pada *default-box* dan pada posisi *j* pada *ground-truth box* pada kategori *p*, maka dalam strategi matching kita bisa mendapatkan $\sum_i x_{ij}^{p} \geq 1$. Secara keseluruhan fungsi *objective loss* merupakan jumlah dengan pemberat dari *localization loss (loc)* dan *confidence loss (conf)* sebagai berikut:

$$L(x,c,l,g) = \frac{1}{N}(L_{conf}(x,c) + \alpha L_{loc}(x,l,g)) \qquad (1)$$

dimana N adalah jumlah *default-box* yang cocok. Jika N=0, maka *loss* di tentukan dengan 0. *Localization loss* adalah *Smooth L1 loss* yang merupakan *loss* dari parameter box yang diprediksi (*l*) dan *ground-truth box* (*g*). Kemudian dilakukan regresi untuk offest pada titik tengah *(cx, cy)* dari *bounding-box* dengan lebar *(w)* dan tinggi *(h)*.

$$\begin{gathered} L_{loc}(x,l,g) = \sum_{i \in Pos}^{N} \sum_{m \in \{cx,cy,w,h\}} x_{ij}^{k} \mathrm{smooth}_{\mathrm{L1}}(l_i^m - \hat{g}_j^m) \\ \hat{g}_j^{cx} = (g_j^{cx} - d_i^{cx})/d_i^w \qquad \hat{g}_j^{cy} = (g_j^{cy} - d_i^{cy})/d_i^h \\ \hat{g}_j^{w} = \log\left(\frac{g_j^w}{d_i^w}\right) \qquad \hat{g}_j^{h} = \log\left(\frac{g_j^h}{d_i^h}\right) \end{gathered} \qquad (2)$$

Loss tingkat kepercayaan adalah *softmax loss* untuk beberapa kelas tingkat kepercayaan.

$$L_{conf}(x,c) = -\sum_{i \in Pos}^{N} x_{ij}^{p} log(\hat{c}_i^p) - \sum_{i \in Neg} log(\hat{c}_i^0) \quad \text{where} \quad \hat{c}_i^p = \frac{\exp(c_i^p)}{\sum_p \exp(c_i^p)} \qquad (3)$$

dan pemberat *α* dibuat 1 untuk validasi silang.

Training network dilakukan dengan menggunakan *stochastic gradient desent*. Untuk setiap sampel training dengan *ground truth g* dan *output network* (c,l) , maka x^* yang sesuai dihitung dengan meminimalkan loss:

$$x^* = arg \ \ min \ \ F(x,c,l,g)$$

sedemikian hingga

$$x_{i,j} \in \{0,1\}, \ \sum_i x_{ij} = 1$$

dan parameter kemudian disesuaikan mengikuti gradien yang dievaluasi pada x^* yang ditemukan.

Adapun contoh pengetikkan tabel dapat dilihat pada Tabel 2 di bawah ini:

3. Hasil Testing

Hasil testing dapat dilihat pada tabel 2 dibawah ini:

Tabel 2. Hasil tes perkelas

No	Kelas	mAP
1	cat	0,9221845756
2	train	0,8961131215
3	bus	0,8611806983
4	dog	0,8587043082
5	horse	0,8436180096
6	diningtable	0,8261594841
7	sofa	0,8209965659
8	motorbike	0,8179529437
9	sheep	0,8026550145
10	aeroplane	0,7965221585
11	car	0,7926940506
12	bicycle	0,7789231043
13	tvmonitor	0,7638968086
14	bird	0,7535152622
15	cow	0,743660099
16	person	0,7224381317
17	chair	0,6704108613
18	boat	0,6345923465
19	pottedplant	0,625741696
20	bottle	0,6016949112
mAP		0,7766827076

Dari Tabel 2 terlihat keakuratan tertinggi didapatkan oleh kelas cat sedangkan bottle memiliki keakuratan yang paling rendah. Keakuratan secara keseluruhan yang didapatkan dari hasil tes adalah 0,7766827076 atau 77,67%. Jika dibandingkan dengan

hasil penelitian lain, maka datanya dapat dilihat pada Tabel 3

Tabel 3. Perbandingan dengan penelitian lain

Method	Fast	Fast	Faster	Faster	Faster	SSD300	SSD300	SSD300	SSD512	SSD512	SSD512	This
data	07	07+12	07	07+12	07+12+COCO	07	07+12	07+12+COCO	07	07+12	07+12+COCO	07+12
mAP	66.9	70.0	69.9	73.2	78.8	68.0	74.3	79.6	71.6	76.8	81.6	77.7
aeroplane	74.5	77.0	70.0	76.5	84.3	73.4	75.5	80.9	75.1	82.4	86.6	79.7
bicycle	78.3	78.1	80.6	79.0	82.0	77.5	80.2	86.3	81.4	84.7	88.3	77.9
bird	69.2	69.3	70.1	70.9	77.7	64.1	72.3	79.0	69.8	78.4	82.4	75.4
boat	53.2	59.4	57.3	65.5	68.9	59.0	66.3	76.2	60.8	73.8	76.0	63.5
bottle	36.6	38.3	49.9	52.1	65.7	38.9	47.6	57.6	46.3	53.2	66.3	60.2
bus	77.3	81.6	78.2	83.1	88.1	75.2	83.0	87.3	82.6	86.2	88.6	86.1
car	78.2	78.6	80.4	84.7	88.4	80.8	84.2	88.2	84.7	87.5	88.9	79.3
cat	82.0	86.7	82.0	86.4	88.9	78.5	86.1	88.6	84.1	86.0	89.1	92.2
chair	40.7	42.8	52.2	52.0	63.6	46.0	54.7	60.5	48.5	57.8	65.1	67.0
cow	72.7	78.8	75.3	81.9	86.3	67.8	78.3	85.4	75.0	83.1	88.4	74.4
diningtable	67.9	68.9	67.2	65.7	70.8	69.2	73.9	76.7	67.4	70.2	73.6	82.6
dog	79.6	84.7	80.3	84.8	85.9	76.6	84.5	87.5	82.3	84.9	86.5	85.9
horse	79.2	82.0	79.8	84.6	87.6	82.1	85.3	89.2	83.9	85.2	88.9	84.4
motorbike	73.0	76.6	75.0	77.5	80.1	77.0	82.6	84.5	79.4	83.9	85.3	81.8
person	69.0	69.9	76.3	76.7	82.3	72.5	76.2	81.4	76.6	79.7	84.6	72.2
pottedplant	30.1	31.8	39.1	38.8	53.6	41.2	48.6	55.0	44.9	50.3	59.1	62.6
sheep	65.4	70.1	68.3	73.6	80.4	64.2	73.9	81.9	69.9	77.9	85.0	80.3
sofa	70.2	74.8	67.3	73.9	75.8	69.1	76.0	81.5	69.1	73.9	80.4	82.1
train	75.8	80.4	81.1	83.0	86.6	78.0	83.4	85.9	78.1	82.5	87.4	89.6
tvmonitor	65.8	70.4	67.6	72.6	78.9	68.5	74.0	78.9	71.8	75.3	81.2	76.4

Pada tabel 3 diatas dapat dibandingkan penelitian ini dengan SSD 300 dengan dataset yang sama yaitu PASCAL VOC 2007-20012. Dari perbandingan tersebut terlihat ada kelas yang memiliki kenaikan keakuratan namun ada pula kelas yang memiliki penurunan keakuratan.

4. Kesimpulan

Penambahan default-boxes pada setiap feature-map dapat meningkatkan keakuratan untuk kelas-kelas tertentu. Namun kelas-kelas tertentu justru mengalami penurunan keakuratan. Kesimpulan yang didapatkan dari penelitian ini adalah ukuran dan aspek rasio dari default boxes mempengaruhi keakuratan berdasarkan kelas dan feature map.

Daftar Pustaka

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in Neural Information Processing Systems 25 (F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, eds.), pp. 1097–1105, Curran Associates, Inc., 2012.

[2] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes challenge: A retrospective," International Journal of Computer Vision , vol. 111, pp. 98–136, Jan. 2015

[3] T. Lin, M. Maire, S. J. Belongie, L. D. Bourdev, R. B. Girshick, J. Hays, P. Perona, D. Ramanan, P. Doll´ar, and C. L. Zitnick, "Microsoft COCO: common objects in context," CoRR , vol. Abs/1405.0312, 2014

[4] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," International Journal of Computer Vision (IJCV) , vol. 115, no. 3, pp. 211–252, 2015.

[5] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. E. Reed, C. Fu, and A. C. Berg, "SSD: single shot multibox detector," CoRR , vol. Abs/1512.02325, 2015.

[6] J. Uijlings, K. van de Sande, T. Gevers, and A. Smeulders, "Selective search for object recognition," International Journal of Computer Vision, 2013