JURNAL TEKNOINFO
Volume 17, Nomor 1, Januari 2023, Page 66-73
ISSN: 1693-0010(Print), ISSN: 2615-224X(Online)
Available online at https://ejurnal.teknokrat.ac.id/index.php/teknoinfo/index

DETEKSI WAJAH MANUSIA BERBASIS *ONE STAGE DETECTOR* MENGGUNAKAN METODE *YOU ONLY LOOK ONCE* (YOLO)

Muhammad Yusqi Alfan Thoriq[1)], Kurniawan Eka Permana[2)], Indah Agustien Siradjuddin[3)]

[1,2,3]Teknik Informatika/Informatika, Universitas Trunojoyo Madura
[1,2,3]Jl. Raya Telang Kamal, Madura
Email: [1]yusqialfann@gmail.com, [2]kurniawan@trunojoyo.ac.id, [3]indah.siradjuddin@trunojoyo.ac.id

Abstract

The purpose of face detection is to find the location of the face in an image; hence it can be utilized for further applications such as face recognition, finding specific faces in a video, and others. We present Algorithm You Only Look Once based on the one-stage detector approach. This algorithm divided an image into grids of a specific size, where each grid or cell represents the candidate location of the target object. The implemented architecture of the Convolutional Neural Network of this algorithm is to classify the candidate object within each grid into a face or non-face class and justify the face's location. We used the trainable VGG-16 model for the convolutional layers and trained the fully connected layers with the appropriate target label. The experiments are conducted using the WIDER Face dataset with various face objects in each image. As a result, we achieved the highest precision, recall, and f1-score are 0.253, 0.247, and 0.25

Keyword: *Face Detection, Convolutional Neural Network, One Stage Detector, You Only Look Once, Visual Geometry Group-16.*

Abstrak

Deteksi wajah bertujuan untuk memperoleh lokasi wajah pada suatu citra, sehingga dengan lokasi wajah ini dapat dimanfaatkan untuk beberapa aplikasi seperti pengenalan wajah pada suatu citra yang sudah terdeteksi, pencarian wajah tertentu dari suatu data video. Pendekatan *one stage detector* dengan algoritma *You Only Look Once* (YOLO) digunakan pada penelitian ini, dimana data citra dibagi menjadi *grid* dengan ukuran terentu. Setiap *grid* ini merupakan representasi lokasi kandidat wajah yang terdapat pada citra. Untuk proses klasifikasi wajah dan non wajah serta perbaikan lokasi wajah di setiap *grid* ini menggunakan arsitektur *Convolutional Neural Network* (CNN), dimana pada penelitian ini menggunakan model yang sudah dilatih yaitu *Visual Geometry Group*-16 (VGG-16) pada lapisan konvolusi, sedangkan pada lapisan *Fully Connected* (FC) dilakukan proses pelatihan dengan data target yang sudah ditentukan. Ujicoba dilakukan pada WIDER *Face dataset* yang memiliki variasi jumlah wajah di dalam setiap citranya. Hasil ujicoba yang telah dilakukan mendapatkan nilai akurasi *Precision* sebesar 0.253, nilai akurasi *Recall* sebesar 0.247 dan nilai akurasi *F1-Score* sebesar 0.25

Kata Kunci: Deteksi Wajah, *Convolutional Neural Network, One Stage Detector, You Only Look Once, Visual Geometry Group-16.*

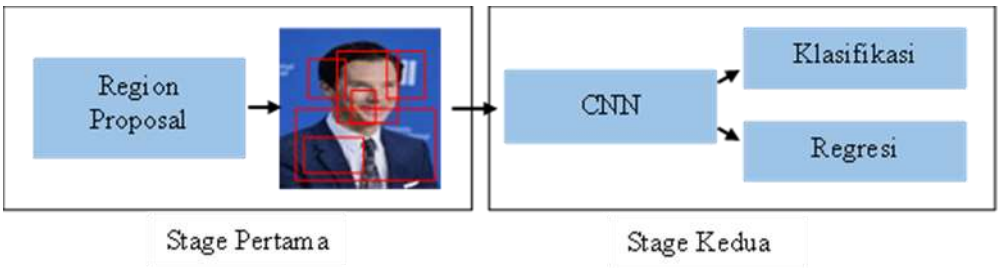
1. PENDAHULUAN

Deteksi wajah merupakan sebuah proses yang digunakan untuk mencari keberadaan wajah pada suatu citra. Deteksi wajah dalam awal penemuan masih relatif sederhana pada kategori dan latar belakang[1]. Seiring berkembangnya teknologi dalam bidang kecerdasan buatan deteksi wajah jika ditambahkan metode pengenalan dapat dimanfaatkan untuk informasi dasar dalam pengenalan ekspresi wajah dan sistem absensi.

Arsitektur yang digunakan adalah *Convolutional Neural Network* (CNN). CNN mempunyai kelebihan salah satunya adalah memiliki kemampuan *feature learning* [2]–[4]

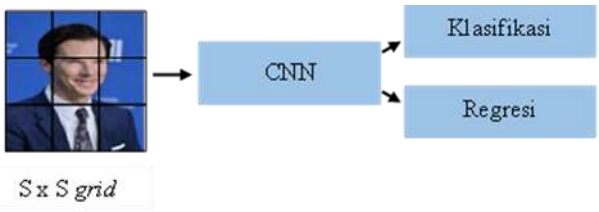
Terdapat dua pendekatan dalam melakukan deteksi wajah yaitu *multi stage detector* dan *one stage detector*. *Multi stage detector* terdiri dari dua tahapan yaitu proses *region proposal* dan proses klasifikasi. Terdapat berbagai metode untuk mendapatkan *region proposal* diantaranya menggunakan *sliding window* [5], [6], *Region Proposal Network* (RPN) [7] dan *selective search* [8]. Beberapa metode yang menggunakan konsep *multi stage detector* adalah *Region-based Convolutional Neural Network* (R-CNN) [9], *Fast R-CNN* [10], dan *Faster R-CNN* [7]. Ilustrasi proses *multi stage detector* dapat dilihat pada Gambar 1

JURNAL TEKNOINFO
Volume 17, Nomor 1, Januari 2023, Page 66-73
ISSN: 1693-0010(Print), ISSN: 2615-224X(Online)
Available online at https://ejurnal.teknokrat.ac.id/index.php/teknoinfo/index



Gambar 1. *Multi Stage Detector*

One stage detector menggabungkan proses *region proposal* dan proses klasifikasi dalam satu *network*. Salah satu metode yang menerapkan konsep *one stage detector* adalah *You Only Look Once* (YOLO). Berikut adalah ilustrasi *one stage detector* yang ditampilkan pada Gambar 2



Gambar 2. *One Stage Detector*

Pada penelitian ini digunakan pendekatan *one stage detector*, yaitu algoritma *You Only Look Once* (YOLO).

2. METODE PENELITIAN

Penelitian ini menggunakan metode *You Only Look Once* (YOLO). YOLO merupakan salah satu algoritma *one stage detector* untuk deteksi objek. Pada YOLO tidak diperlukan pemilihan *region*. Setiap citra pada YOLO akan dibentuk *grid* dengan ukuran *S* x *S,* dimana setiap *grid* pada citra akan memprediksi *bounding box B* dan Nilai probabilitas kelas *C*. Terdapat lima prediksi dalam *bounding box B* yaitu *confidence score* (*p*), *x, y, w,* dan *h*. Nilai *confidence score* menyatakan ada atau tidaknya objek pada *grid* tertentu. Nilai *x* dan *y* menyatakan titik pusat dari suatu objek, *w* dan *h* merupakan lebar dan tinggi *bounding box* dari objek [11]. Hasil ukuran data target dalam bentuk *grid* dapat dihitung dengan Persamaan sebagai berikut

$$S\ x\ S\ x\ (B * 5 + C) \tag{1}$$

Dimana *S* ukuran *grid*, *B* dalah jumlah *bounding box* dan *C* adalah probabilitas kelas

Terdapat tiga tahapan utama pada YOLO, yaitu pembentukan *bounding box*, pembentukan model dan prediksi lokasi objek.

Pembentukan *bounding box* pada YOLO dapat dilakukan dengan konversi koordinat anotasi ke koordinat *grid*. Ada empat tahapan dalam melakukan konversi koordinat anotasi ke koordniat *grid* yaitu

1. *Resize* citra yang bertujuan untuk menyamakan ukuran citra. Dalam proses *resize* citra koordinat anotasi harus disesusaikan dengan ukuran citra yang sudah diubah ukurannya melalui Persamaan berikut

$$X = \frac{w\ resize}{w\ awal} \tag{2}$$

$$Y = \frac{h\ resize}{h\ awal} \tag{3}$$

$$Xmin = xmin * X \tag{4}$$

$$Ymin = ymin * Y \tag{5}$$

$$Xmax = xmax * X \tag{6}$$

JURNAL TEKNOINFO
Volume 17, Nomor 1, Januari 2023, Page 66-73
ISSN: 1693-0010(Print), ISSN: 2615-224X(Online)
Available online at https://ejurnal.teknokrat.ac.id/index.php/teknoinfo/index

$$Ymax = ymax * Y \tag{7}$$

2. Perhitungan titik tengah koordinat anotasi dapat dihitung melalui Persamaan berikut

$$Xc = \frac{Xmin + Xmax}{2} \tag{8}$$

$$Yc = \frac{Ymin + Ymax}{2} \tag{9}$$

3. Penentuan lokasi koordinat *grid* dapat dihitung melalui Persamaan berikut

$$Cx = \left|\frac{Xc}{ukuran\ grid}\right| \tag{10}$$

$$Cy = \left|\frac{Yc}{ukuran\ grid}\right| \tag{11}$$

4. Konversi koordinat anotasi ke koordinat *grid* tahap terakhir ini dapat dilakukan melalui Persamaan berikut

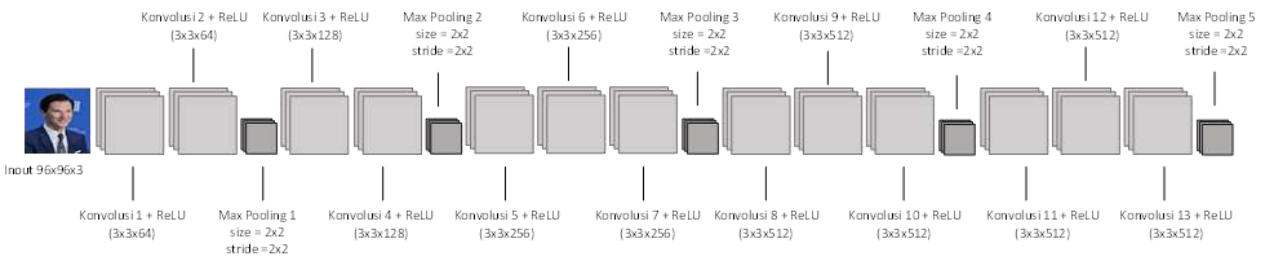
$$Xg = \frac{\left(Xc - (Cx * ukuran\ grid)\right)}{(ukuran\ grid - 1)} \tag{12}$$

$$Yg = \frac{(Yc - (Cy * ukuran\ grid))}{(ukuran\ grid - 1)} \tag{13}$$

$$Wg = \frac{((Xmax - Xmin))}{ukuran\ citra} \tag{14}$$

$$Hg = \frac{((Ymax - Ymin))}{ukuran\ citra} \tag{15}$$

Pembentukan model pada YOLO menggunakan *Convolutional Neural Network* (CNN). Pada penelitian ini model yang dibangun menggunakan *convolution layer* yang sudah dilatih yaitu *Visual Geometry Group*-16 (VGG-16). Berikut adalah arsitektur VGG-16 yang ditampilkan pada Gambar 3



Gambar 3. Arsitektur VGG-16

Prediksi lokasi pada YOLO memiliki dua tahapan yang pertama adalah memfilter *confidence score* yang memiliki nilai diatas *threshold*. Proses kedua adalah konversi koordinat *grid* ke koordinat anotasi yang dapat dilakukan melalui Persamaan berikut

JURNAL TEKNOINFO
Volume 17, Nomor 1, Januari 2023, Page 66-73
ISSN: 1693-0010(Print), ISSN: 2615-224X(Online)
Available online at https://ejurnal.teknokrat.ac.id/index.php/teknoinfo/index

$$Xc = \big(Xg * (ukuran\ grid - 1)\big) + (Cx * ukuran\ grid) \quad (16)$$

$$Yc = \big(Yg * (ukuran\ grid - 1)\big) + (Cy * ukuran\ grid) \quad (17)$$

$$W = Wg * ukuran\ citra \quad (18)$$

$$H = Hg * ukuran\ citra \quad (19)$$

$$Xmin = Xc - (\frac{W}{2}) \quad (20)$$

$$Ymin = Yc - (\frac{H}{2}) \quad (21)$$

$$Xmax = Xc + (\frac{W}{2}) \quad (22)$$

$$Ymax = Yc + (\frac{H}{2}) \quad (23)$$

3. HASIL DAN PEMBAHASAN

Dataset yang digunakan pada penelitian ini adalah WIDER *Face dataset* [12]. Jumlah *dataset* citra yang digunakan adalah 2400 citra dengan pembagian 1920 citra *training* dan 480 citra *testing*.

Terdapat tiga skenario percobaan pada penelitian ini diantaranya perbandingan ukuran *grid,* perbandingan ukuran citra, dan perbandingan jumlah *epoch.*

Perbandingan ukuran *grid* dilakukan dengan membandingkan penggunaan *grid* 3 x 3 dan 7 x 7 dengan masing-masing ukuran tiap *grid* adalah 32. Sehingga ukuran citra yang digunakan untuk *grid* 3 x 3 adalah 96 x 96 piksel dan ukuran *grid* 7 x 7 adalah 224 x 224 piksel.

Perbandingan ukuran citra dilakukan dengan membandingkan ukuran citra 96 x 96 piksel dan 222 x 222 piksel, *grid* yang digunakan adalah hasil dari *grid* yang paling baik dari skenario percobaan sebelumnya.

Perbandingan jumlah *epoch* dilakukan dengan membandingkan penggunaan jumlah *epoch* 50, 100 dan 150, ukuran *grid* dan citra yang digunakan adaah hasil yang paling baik dari skenario percobaan sebelumnya.

Hasil uji coba dari ketiga skenario yang sudah dilakukan diperoleh hasil dan analisa sebagai berikut

1. Perbandingan ukuran *grid* yang dilakukan menghasilkan bahwa ukuran *grid* 3 x 3 lebih baik dari ukuran *grid* 7 x 7 pada penelitian ini. Hal tersebut dapat dilihat dari hasil akurasi yang telah dihitung seperti pada Tabel 1 berikut

Tabel 1. Hasil Akurasi Perbandingan Ukuran *Grid*

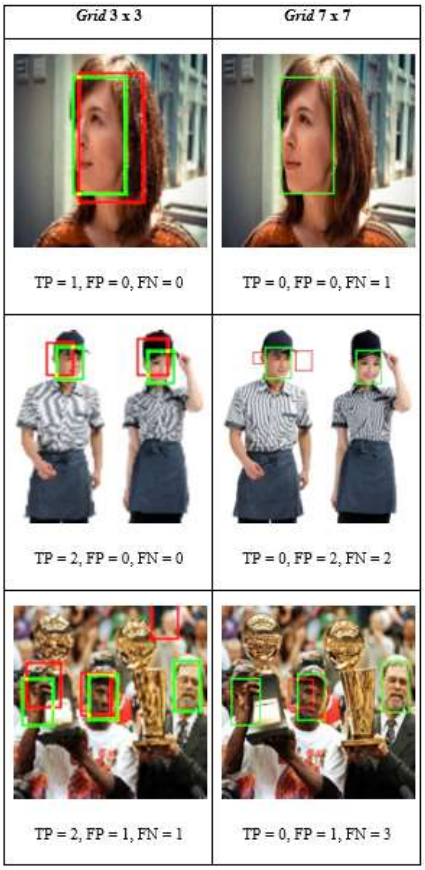
Ukuran *Grid*	Ukuran Citra	*Precision*	*Recall*	*F1-Score*
3 x 3	96 x 96	0.130	0.128	0.129
7 x 7	224 x 224	0.12	0.072	0.090

Dari hasil akurasi yang sudah didapat bisa digambarkan grafik akurasi perbandingan penggunaan ukuran *grid* seperti pada Gambar 4

JURNAL TEKNOINFO
Volume 17, Nomor 1, Januari 2023, Page 66-73
ISSN: 1693-0010(Print), ISSN: 2615-224X(Online)
Available online at https://ejurnal.teknokrat.ac.id/index.php/teknoinfo/index

Gambar 4. Grafik Akurasi Perbandingan Ukuran *Grid*

Hasil citra deteksi yang dihasilkan melalui perbandingan ukuran *grid* 3 x 3 dengan 7 x 7 dapat dilihat pada Gambar 5



Gambar 5. Hasil Citra Deteksi Perbandingan Ukuran *Grid*

Dari hasil skenario perbandingan *grid* yang telah dilakukan penggunaan ukuran *grid* 3 x 3 lebih baik dari *grid* 7 x 7. Hal tersebut disebabkan karena ketidakseimbangan jumlah data target positif dan target negatif yang dihasilkan pada saat pembuatan *bounding box.*

Pembuatan bounding box untuk data target berbentuk *grid* dengan ukuran 3 x 3 x 5, pada tiap *grid* hanya bisa diisi oleh satu objek wajah. Sehingga jika citra hanya memiliki 1 objek wajah maka akan terdapat 1 data target postif dan 8 data target negatif yang dihasilkan. Jika ukuran *grid* 7 x 7 maka akan terdapat 1 data target positif dan 48 data target

JURNAL TEKNOINFO
Volume 17, Nomor 1, Januari 2023, Page 66-73
ISSN: 1693-0010(Print), ISSN: 2615-224X(Online)
Available online at https://ejurnal.teknokrat.ac.id/index.php/teknoinfo/index

negatif. Hal tersebut yang menyebabkan penggunaan *grid* 3 x 3 lebih baik dari *grid* 7 x 7 pada penelitian ini.

2. Perbandingan ukuran citra yang dilakukan menghasilkan bahwa ukuran citra 222 x 222 pikel lebih baik dari 96 x 96 piksel pada *grid* yang sama yaitu ukuran *grid* 3 x 3. Hal tersebut dapat dilihat dari hasil akurasi yang telah dihitung seperti pada Tabel 2 berikut

Tabel 2. Hasil Akurasi Perbandingan Ukuran Citra

Ukuran *Grid*	Ukuran Citra	*Precision*	*Recall*	*F1-Score*
3 x 3	96 x 96	0.130	0.128	0.129
3 x 3	222 x 222	0.250	0.238	0.244

Dari hasil akurasi yang sudah didapat bisa digambarkan grafik akurasi perbandingan penggunaan ukuran citra sepeti pada Gambar 6

Gambar 6. Grafik Akurasi Perbandingan Ukuran Citra

Hasil citra deteksi yang dihasilkan melalui perbandingan ukuran citra 96 x 96 piksel dengan 222 x 222 piksel dapat dilihat pada Gambar 7

Gambar 7. Hasil Citra Deteksi Perbandingan Ukuran Citra

JURNAL TEKNOINFO
Volume 17, Nomor 1, Januari 2023, Page 66-73
ISSN: 1693-0010(Print), ISSN: 2615-224X(Online)
Available online at https://ejurnal.teknokrat.ac.id/index.php/teknoinfo/index

Dari hasil skenario perbandingan ukuran citra yang telah dilakukan penggunaan ukuran citra 222 x 222 piksel lebih baik dari ukuran citra 96 x 96 piksel . Hal tersebut disebabkan karena semakin kecil ukuran citra maka semakin banyka informasi yang dihilangkan pada saat pelatihan.

3. Perbandingan jumlah *epoch* menghasilkan bahwa *epoch* 150 adalah yang terbaik. Hal tersebut dapat dilihat dari hasil akurasi yang telah dihitung seperti pada Tabel 3 berikut

Tabel 3. Hasil Akurasi Perbandingan Jumlah *Epoch*

Epoch	*Precision*	*Recall*	*F1-Score*
50	0.225	0.207	0.215
100	0.250	0.238	0.244
150	0.253	0.247	0.250

Dari hasil akurasi yang sudah didapat bisa digambarkan grafik akurasi perbandingan penggunaan jumlah *epoch* sepeti pada Gambar 8

Gambar 8. Grafik Akurasi Perbandingan Jumlah *Epoch*

Hasil citra deteksi yang dihasilkan melalui perbandingan jumlah *epoch* 50, 100 dan 150 dapat dilihat pada Gambar 9

Gambar 9. Hasil Citra Deteksi Perbandingan Jumlah *Epoch*

JURNAL TEKNOINFO
Volume 17, Nomor 1, Januari 2023, Page 66-73
ISSN: 1693-0010(Print), ISSN: 2615-224X(Online)
Available online at https://ejurnal.teknokrat.ac.id/index.php/teknoinfo/index

Dari hasil skenario perbandingan jumlah *epoch* yang telah dilakukan menunjukkan bahwa semakin tinggi jumlah *epoch* yang digunakan akan semakin tinggi hasil akurasi yang didapatkan meskipun kenaikan akurasi tidak signifikan. Waktu komputasi yang dibutuhkan juga akan semakin lama dengan jumlah *epoch* yang tinggi.

4. KESIMPULAN

Setelah dilakukan tiga skenario pengujian didapatkan kesimpulan bahwa deteksi wajah manusia menggunakan *You Only Look Once* dengan *dataset* WIDER *Face* berjumlah 1920 citra *training* dan 480 citra *testing* didapatkan hasil terbaik adalah pada ukuran *grid* 3 x 3, ukuran citra 222 x 222 piksel dan jumlah *epoch* yang digunakan adalah 150. Hasil akurasi yang didapatkan adalah nilai *Precision* sebesar 0.253, *Recall* sebesar 0.247, dan *F1-Score* sebesar 0.25. Nilai akurasi dapat dipengaruhi oleh penggunaan model *Convolutional Neural Network* (CNN) yang dibangun dan keseimbangan data target positif dan negatif yang digunakan pada saat membangun model untuk deteksi wajah.

DAFTAR PUSTAKA

[1] J. Lu, S. Tang, J. Wang, H. Zhu, and Y. Wang, "A Review on Object Detection Based on Deep Convolutional Neural Networks for Autonomous Driving," *Proc. 31st Chinese Control Decis. Conf. CCDC 2019*, pp. 5301–5308, 2019, doi: 10.1109/CCDC.2019.8832398.

[2] I. A. Siradjuddin, A. Sakinah, and M. K. Sophan, "Combination of feature engineering and feature learning approaches for classification on visual complexity images," *Int. J. Innov. Comput. Inf. Control*, vol. 17, no. 3, pp. 991–1005, 2021, doi: 10.24507/ijicic.17.03.991.

[3] J. Fan, W. Xu, Y. Wu, and Y. Gong, "Human tracking using convolutional neural networks," *IEEE Trans. Neural Networks*, vol. 21, no. 10, pp. 1610–1623, 2010, doi: 10.1109/TNN.2010.2066286.

[4] H. Khalajzadeh, M. Mansouri, and M. Teshnehlab, "Face recognition using convolutional neural network and simple logistic classifier," *Adv. Intell. Syst. Comput.*, vol. 223, pp. 197–207, 2014, doi: 10.1007/978-3-319-00930-8_18.

[5] H. Li, Z. Lin, X. Shen, J. Brandt, and G. Hua, "A convolutional neural network cascade for face detection," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June, pp. 5325–5334, 2015, doi: 10.1109/CVPR.2015.7299170.

[6] M. M. Cheng, Y. Liu, W. Y. Lin, Z. Zhang, P. L. Rosin, and P. H. S. Torr, "BING: Binarized normed gradients for objectness estimation at 300fps," *Comput. Vis. Media*, vol. 5, no. 1, pp. 3–20, 2019, doi: 10.1007/s41095-018-0120-1.

[7] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, 2017, doi: 10.1109/TPAMI.2016.2577031.

[8] J. R. R. Uijlings, K. E. A. Van De Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," *Int. J. Comput. Vis.*, vol. 104, no. 2, pp. 154–171, 2013, doi: 10.1007/s11263-013-0620-5.

[9] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 580–587, 2014, doi: 10.1109/CVPR.2014.81.

[10] R. Girshick, "Fast R-CNN," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2015 Inter, pp. 1440–1448, 2015, doi: 10.1109/ICCV.2015.169.

[11] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 779–788, 2016, doi: 10.1109/CVPR.2016.91.

[12] S. Yang, P. Luo, C. C. Loy, and X. Tang, "WIDER FACE: A face detection benchmark," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 5525–5533, 2016, doi: 10.1109/CVPR.2016.596.