

Perbandingan Metode ResNet, YoloV3, dan TinyYoloV3 pada Deteksi Citra dengan Pemrograman Python

Agung Slamet Riyadi¹; Ire Puspa Wardhani²; Maria Sri Wulandari³; Susi Widayati⁴

¹ Jurusan Teknik Informatika, Fakultas Teknologi Industri, Universitas Gunadarma

¹ Jl. Margonda Raya No. 100, Depok 16424, Jawa Barat

² Program Studi Teknologi Informasi, STMIK Jakarta STI&K

^{3,4} Program Studi Sistem Informasi, STMIK Jakarta STI&K

^{2,3,4} Jl. BRI Radio Dalam No.17 Keb. Baru Jakarta Selatan 12140

¹ agungsr357rvr@gmail.com

² irepuspa@gmail.com

³ mswuland89@gmail.com

⁴ widayatisusi@gmail.com

ABSTRACT

Image processing is an image processing technique that transforms the input image into another image so that the output has a better quality than the input image quality. Detection of objects in digital image processing is a process used to determine the existence of certain objects in a digital image. The detection process can be carried out by various methods which generally read the features of all objects in the input image. There are several methods for detecting these objects, some of which are ResNet, YOLOv3 and TinyYOLOv3 methods. In this study, the differences between the three methods will be carried out. The purpose of the trial in this study is to find out more about the results of object detection in the image with a good level of accuracy. And this study can prove that the results of object detection using the YOLOv3 model, the accuracy results are higher than the ResNet model and the Tiny YOLO model.

Keywords: Deep Learning, ResNet, YoloV3, TinyYoloV3, Image Detection

ABSTRAK

Pengolahan citra (image processing) adalah teknik mengolah citra yang mentransformasikan citra masukan menjadi citra lain agar keluaran memiliki kualitas yang lebih baik dibandingkan kualitas citra masukan. Deteksi objek dalam digital image processing adalah suatu proses yang digunakan untuk menentukan keberadaan objek tertentu di dalam suatu citra digital. Proses deteksi tersebut dapat dilakukan dengan berbagai macam metode yang umumnya melakukan pembacaan fitur-fitur dari seluruh objek pada citra input. Pada pendeteksian objek tersebut terdapat beberapa metode beberapa diantaranya yaitu metode ResNet, YOLOv3 dan TinyYOLOv3. Dalam penelitian ini akan dilakukan perbandingan ketiga metode tersebut. Tujuan dari uji coba dalam penelitian ini adalah mengetahui lebih jauh hasil deteksi objek pada citra dengan tingkat keakuratan yang baik. Dan penelitian ini dapat membuktikan bahwa hasil pendeteksian objek dengan menggunakan model YOLOv3, hasil keakuratannya lebih tinggi dibandingkan dengan model ResNet dan model Tiny YOLO.

Kata Kunci: Deep Learning, ResNet, YoloV3, TinyYoloV3, Deteksi Citra

1. PENDAHULUAN

Teknologi pengolahan citra digital (*Digital Image Processing*) berkembang demikian pesat, dengan banyaknya kemudahan dalam pengimplementasian yang diperoleh pada kehidupan manusia, banyak aplikasi yang menggunakan teknologi ini pada berbagai bidang. Pengolahan citra (*image processing*) adalah teknik mengolah citra yang mentransformasikan citra masukan menjadi citra lain agar keluaran memiliki kualitas yang lebih baik dibandingkan kualitas citra masukan. Pengolahan citra sangat bermanfaat, diantaranya adalah untuk meningkatkan kualitas citra, menghilangkan cacat pada citra, mengidentifikasi objek, penggabungan dengan bagian citra yang lain.

Deteksi objek dalam *digital image processing* adalah suatu proses yang digunakan untuk mendeteksi dan menentukan keberadaan objek tertentu di dalam sebuah citra digital. Proses deteksi dilakukan dengan berbagai macam cara dan metode dengan pembacaan fitur-fitur yang ada dari seluruh objek pada citra yang ada sebagai input data. Fitur-fitur tersebut kemudian dibandingkan dengan fitur dari objek referensi yang telah ada sebelumnya. Hasil perbandingan antara fitur input dan fitur referensi kemudian digunakan untuk menentukan hasil yang sesuai yaitu apakah objek yang terdeteksi dapat sebagai template yang dimaksud atau tidak. Sistem pendeteksian objek diharapkan mampu menemukan objek di dunia nyata dengan memanfaatkan model objek yang telah diketahui sebelumnya. Permasalahan yang ada pada pendeteksian objek adalah permasalahan pelabelan yang berdasarkan pada model objek yang ingin diketahui. Citra yang berisi satu atau lebih objek dan satu set label yang sesuai dengan satu set model yang telah dikenal oleh sistem, sistem diharapkan dapat menetapkan label yang benar pada citra. Pendeteksian objek ini tidak dapat diselesaikan sampai citra dapat tersegmentasi dengan baik serta tanpa deteksi parsial, proses segmentasi tidak dapat diterapkan dan deteksi yang dimaksud merujuk pada banyak kemampuan visual yang dapat diidentifikasi, dikategorisasikan dan diskriminasi. Dalam penelitian ini, memberikan hasil gambaran tentang perbandingan berbagai teknik dan pendekatan yang digunakan untuk mendeteksi objek dalam citra.

Husin et al, 2012 meneliti tentang berbagai teknik tentang cara mendeteksi mangga dari pohon mangga. Teknik-teknik tersebut adalah pengolahan warna yang digunakan sebagai penyaringan utama untuk menghilangkan warna atau objek yang tidak berhubungan pada citra. Selain itu juga digunakan deteksi bentuk dimana akan menggunakan deteksi tepi yaitu Circular Hough Transform (CHT) [5].

Torrent et al, 2013 melakukan penelitian dan mengusulkan kerangka kerja untuk secara bersamaan melakukan OD dan segmentasi pada objek dengan sifat yang berbeda, yang didasarkan pada prosedur boosting yang secara otomatis memutuskan sesuai dengan properti objek apakah lebih baik memberikan bobot lebih pada proses deteksi atau segmentasi untuk meningkatkan kedua hasil. Pendekatan mereka memungkinkan informasi untuk dilintasi dari deteksi ke segmentasi dan sebaliknya. Waktu tugas ini dapat meningkat jika pada awalnya objek yang terdeteksi bukan yang menarik [6].

Malagon-Borja dan Fuentes, 2009 melakukan penelitian dan telah menyajikan sistem deteksi objek yang bekerja tanpa menggunakan data asumsi pengetahuan dan pembelajaran sebelumnya tentang citra tersebut. Sistem mereka bekerja pada tahap pertama pengklasifikasi dan memeriksa setiap lokasi pada citra dengan skala berbeda. Tahap kedua sistem menghilangkan deteksi palsu berdasarkan heuristik. Pengklasifikasian objek berdasarkan pada Principal Component Analysis (PCA) yaitu mengompresi secara optimal jenis citra yang digunakan dengan menghitung komponen utama. Jadi pengklasifikasian dengan PCA dilakukan secara terpisah dari contoh positif dan dari contoh negatif; pada saat pengklasifikasian dengan pola baru, kedua set membandingkan rekonstruksi, menetapkan contoh ke

kelas dengan tingkat kesalahan rekonstruksi yang terkecil. Waktu akan menjadi masalah dengan pengklasifikasi tersebut ketika memeriksa setiap lokasi dalam citra [7].

Menurut Everingham M et Al, 2015 Dataset yang digunakan dalam Tantangan Pascal VOC terdiri dari dua subset: dataset trainval, yang selanjutnya diklasifikasikan ke dalam set terpisah untuk pelatihan dan validasi; dan kumpulan data uji. Semua citra yang terkandung sepenuhnya dijelaskan dengan bantuan kotak pembatas untuk semua contoh objek berikut untuk tantangan klasifikasi dan deteksi [8]. Seiring dengan pernyataan ini, atribut yang digunakan pada pendeteksian seperti sudut pandang, pemotongan, sulit, konsisten, akurat dan lengkap ditentukan, beberapa di antaranya ditambahkan dalam penelitian selanjutnya [9].

2. METODE DAN RANCANGAN PENELITIAN

Pada prinsipnya dasar sebuah sistem pendeteksian sebuah objek dapat diilustrasikan dengan skema penelitian yang terlihat pada Gambar 1 yaitu proses deteksi objek pada citra. Input dasar ke sistem deteksi objek dapat berupa gambar atau adegan dalam video. Tujuan sistem pendeteksian ini adalah untuk mendeteksi objek yang ada dalam citra atau adegan lain yang dibutuhkan sistem dalam mengkategorikan berbagai objek ke dalam kelas objek masing-masing dengan beberapa metode beberapa diantaranya yaitu metode ResNet, YOLOv3 dan TinyYOLOv3.



Gambar 1. Skema Penelitian

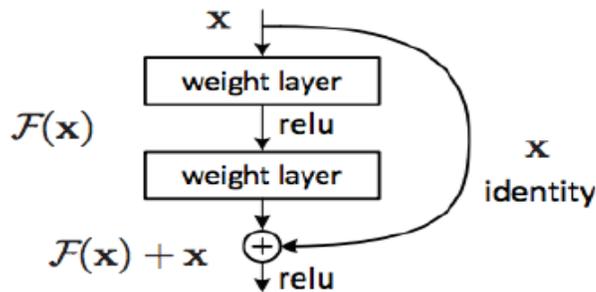
3. ANALISIS DAN PEMBAHASAN

Deteksi objek adalah proses komputer dengan mengidentifikasi keberadaan, lokasi, dan jenis satu atau lebih objek citra diberikan. Beberapa metode yang digunakan untuk pengenalan objek (misalnya di mana objek tersebut berada), lokalisasi objek (misalnya berapa luas objek tersebut), dan klasifikasi objek (misalnya jenis objek yang dimaksud).

3.1. Model ResNet (Residual Network)

Menurut Kaiming. He, X. Zhang et al Residual neural network atau yang biasa disebut sebagai Resnet adalah salah satu jenis arsitektur yang cukup populer. Resnet memiliki berbagai macam jenis arsitektur, mulai dari 18, 34, 50, 101, sampai 152 layer [1]. Arsitektur ini cukup revolusioner pada saat itu karena arsitektur ini menjadi state of the art pada saat itu tidak hanya dalam klasifikasi, namun dalam semua kategori termasuk object detection, dan semantic segmentation. Arsitektur CNN yang memiliki kedalaman tinggi adalah salah satu hal penting dalam membangun model CNN yang memiliki performa yang baik, namun model CNN yang memiliki kedalaman yang tinggi juga memiliki masalah, yaitu

vanishing gradient problem, yaitu suatu keadaan dimana hasil gradien yang dipelajari oleh model, tidak dapat mencapai layer pertama karena mengalami perkalian berkali-kali sehingga layer pertama tidak menerima gradien apa-apa, atau secara singkatnya, hal ini menyebabkan suatu CNN tidak dapat belajar dari error yang telah dikalkulasi [2]. Hal yang diusung oleh Kaiming He et al. pada saat itu adalah dengan menggunakan sesuatu yang bernama residual block, blok ini adalah blok yang ada pada tiap lapis arsitektur CNN Resnet dan menjadi fundamental dari arsitektur tersebut, gambaran dari blok ini dapat dilihat pada Gambar 2.

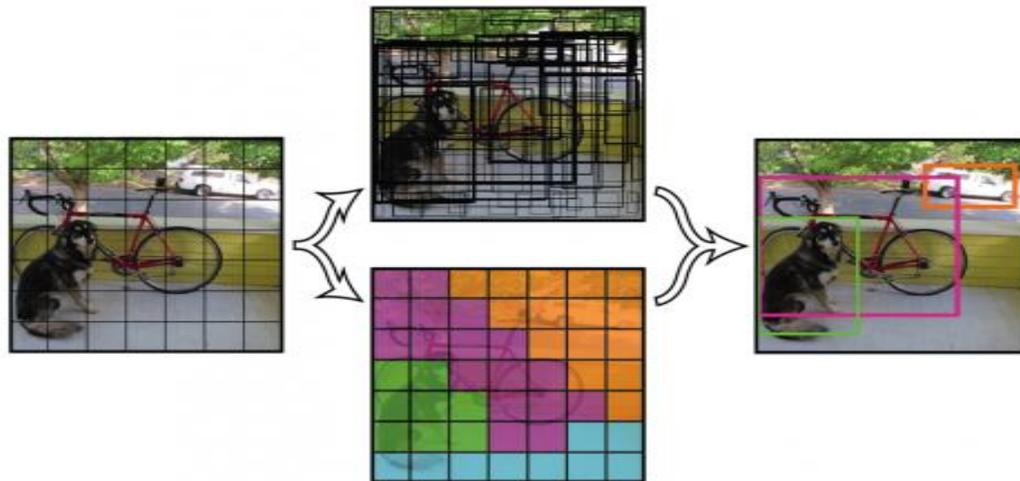


Gambar 2. Blok residual fondasi arsitektur ResNet.

3.2. Model YOLOv3

Dalam beberapa tahun terakhir, teknik pembelajaran pendeteksian objek menghasilkan hasil yang baik. Salah satu bentuk model pendeteksian yang terkenal adalah *You Only Look Once* (YOLO), merupakan keluarga *Convolutional Neural Networks* (CNN) yang mencapai hasil optimal dengan model ujung-ke-ujung tunggal yang dapat melakukan deteksi objek secara real-time. YOLOv3 merupakan salah satu *object detector* dengan tingkat akurasi yang relatif baik dan mampu mendeteksi objek secara real-time. YOLOv3 menggunakan metode anchor-based detection dimana metode ini menggunakan kotak jangkar (*anchor box*). Metode ini memberikan keuntungan yang relatif signifikan dibandingkan metode lama, yaitu *sliding window detection*. *Sliding window detection* sangatlah boros dalam proses komputasi karena jaringan syaraf buatan akan melakukan proses komputasi dari setiap pergeseran window pada citra sebagai detector dalam satu frame citra. Jika menggunakan metode anchor-based detector, maka proses komputasi akan sangat signifikan menurun karena anchor box akan di-generate dalam sekali proses untuk setiap frame.

Anchor-based detection pada YOLOv3 memilih *anchor box* yang tepat sebagai area objek yang dideteksi berdasarkan nilai confidence score dan IoU (Intersection over Union) yang memenuhi nilai threshold yang sudah ditentukan. Dengan metode ini YOLOv3 memperoleh *Mean Average Precision* pada COCO dataset sebesar 55,3% dengan kecepatan deteksi sebesar 35 Frame per Second menggunakan GPU Pascal Titan X. Dengan kemampuan ini, YOLOv3 memiliki potensi yang relatif baik untuk digunakan sebagai real-time face detection.



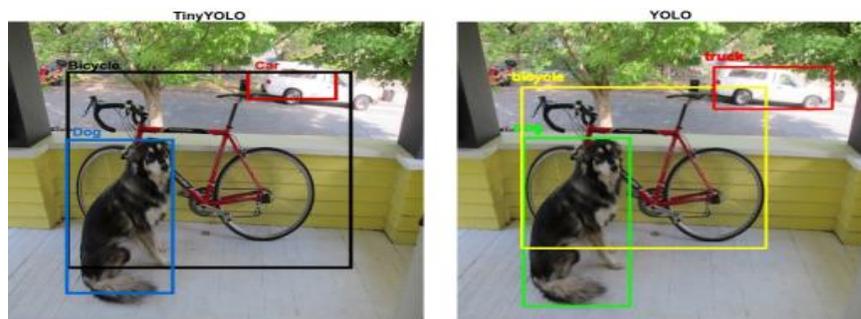
Gambar 3. YOLO: Real Time Object Detection (Sumber: Github pjreddie)

Jaringan YOLO didasarkan pada pembagian sistematis dari sebuah citra yang diberikan ke dalam bentuk grid. Grid terdiri dari 3 jenis. YOLO menggunakan batas-batas kotak pembatas dan jangkar untuk analisis citra. Kotak-kotak ini diakui sebagai hasil deteksi meskipun ribuan data diabaikan karena skor probabilitasnya yang rendah dan diperlakukan sebagai nilai positif palsu. Kotak-kotak ini adalah manifestasi dari pemecahan citra ketat menjadi kisi-kisi sel [10,11,12].

3.3. Model TinyYOLOv3

TinyYOLO (juga disebut tiny Darknet) adalah versi ringan dari jaringan saraf dalam pendeteksian objek waktu nyata YOLO (You Only Look Once). TinyYOLO lebih ringan dan lebih cepat dari YOLO sementara juga mengungguli akurasi model lampu lainnya. Tabel berikut menyajikan perbandingan antara YOLO, Alexnet, SqueezeNet, dan TinyYOLO. Jaringan TinyYOLOv3 dapat digunakan untuk pengenalan dan klasifikasi objek. Model ini dilatih dengan kumpulan data Coco dan dapat mendeteksi hingga 80 kelas.

TinyYOLOv3 membagi citra menjadi sel grid 13x13 dan 26x26. Setiap sel grid memiliki 3 kotak jangkar dan setiap kotak jangkar memiliki skor objek, 20 skor kelas, dan 4 koordinat kotak pembatas. Kode `tiny_yolo_v3.py` membaca jumlah kelas melalui argumen `-labels`.



Gambar 4. Perbedaan Hasil TinyYolo dan YOLO

Dari gambar 4 terlihat pendeteksian dengan model Tiny Yolo objek mobil tidak masuk dalam klasifikasi objek yang dimaksud, namun pendeteksian menggunakan Yolo objek mobil terlihat termasuk dalam objek yang terdeteksi dengan tingkat keakuratan yang tinggi.

4. HASIL UJICоба

Hasil dari penelitian ini adalah sebuah aplikasi yang dapat menentukan keberadaan objek tertentu di dalam suatu citra digital dengan menggunakan 3 model yang berbeda, yaitu ResNet, YOLOv3, dan TinyYOLOv3.

Pada awal ujicoba dilakukan penginstalan library 'ImageAI'. Fungsi dari library tersebut adalah untuk memungkinkan kita melakukan *image prediction*. Selain *image prediction*, ImageAI juga dapat melakukan object detection, video detection dan object tracking. Berikut adalah beberapa package yang digunakan :

1. Object Detection dari library ImageAI.
2. Lalu import urllib yang digunakan untuk mengambil request url dari link yang akan digunakan dan package IPython untuk menampilkan tampilan yang telah dideteksi.
3. Setelah mengimport semua package yang diperlukan, dengan input citra yang akan digunakan menggunakan proses deteksi dengan program python pada tiga model yang berbeda dan sampel citra yang digunakan :



Gambar 5. Citra Gambar Asli

Dan sampel Citra data set training menggunakan data set yang diambil dari kumpulan dataset <https://www.kaggle.com/karthika95/pedestrian-detection> .

Berikut ini merupakan ujicoba pendeteksian pada sampel gambar dengan 5 objek menggunakan tiga metode yang dapat dilihat hasilnya :

- a. Hasil ujicoba pendeteksian pada objek dengan menggunakan Model YOLOv3 dapat dilihat pada gambar 6 berikut ini :



Gambar 6. Model Model YOLOv3

Pada gambar 3 terlihat hasil pendeteksian dengan model YOLOv3 terlihat jelas 5 objek dengan keakuratan tinggi, yaitu 1 objek Bus dan 4 objek manusia.

b. Hasil ujicoba pendeteksian pada objek dengan menggunakan model ResNet dapat dilihat pada gambar 7 berikut ini :



Gambar 7. Hasil Deteksi ResNet

Pada gambar 4 terlihat hasil pendeteksian dengan model ResNet terlihat jelas 5 objek dengan keakuratan tinggi, yaitu 1 objek Bus dan 4 objek manusia, namun untuk objek yang bergerak memiliki tingkat keakuratan yang tinggi.

c. Hasil ujicoba pendeteksian pada objek dengan menggunakan model Tiny YOLOv3 dapat dilihat pada gambar 8 berikut ini :



Gambar 8. Hasil Deteksi Tiny YOLOv3

Pada gambar 5 terlihat hasil pendeteksian dengan model YOLOv3 terlihat jelas hanya 4 objek dengan keakuratan tinggi, yaitu objek manusia dan objek bus tidak terdeteksi, namun untuk objek yang tidak bergerak memiliki tingkat keakuratan yang rendah.

Dengan melihat ketiga hasil deteksi tersebut diatas dapat dilihat perbandingan prosentase keakuratan masing-masing model pada tabel 1 berikut ini :

Tabel 1. Perbandingan Model pada Deteksi Objek Citra

Model	Objek Bus (%)	Objek Person 1 (%)	Objek Person 2 (%)	Objek Person 3 (%)	Objek Person 4 (%)
YOLOv3 Type Result	99.182790517807	99.95113611221313	99.98750686645508	87.32494711875916	97.5606918334961
Retina Type Result	97.67121076583862	94.84061002731323	94.30742263793945	89.49382305145264	58.50703716278076
TiniYOL Ov3 Type Result	Tidak terdeteksi	98.48272204399109	99.94239211082458	61.96569800376892	89.44951295852661

Sistem Pendeteksian dengan menggunakan model Tiny YoloV3, jenis objeknya lebih spesifik dikenali, sedangkan dengan model resNet objek yang ada digambar tidak terdeteksi berdasarkan jenis, namun berdasarkan semua objek yang terlihat.T

4. KESIMPULAN

Berdasarkan pembahasan dan analisa hasil percobaan diatas dapat disimpulkan bahwa percobaan ketiga model dalam deteksi gambar dengan 5 objek di dalamnya sebagai berikut:

1. Berdasarkan hasil model YOLOv3 yang didapatkan tingkat keakuratan dengan rata - rata 96.79%.

2. Dilanjutkan dengan hasil dari penggunaan model ResNet yang didapatkan tingkat keakuratan dengan rata - rata 86.96%.
3. Terakhir hasil dari penggunaan model TinyYOLOv3 yang didapatkan tingkat keakuratan dengan rata - rata 65.96%.

Hasil yang didapatkan yang mendekati tingkat keakuratan terbaik adalah menggunakan model YOLOv3, dikarenakan model tersebut dapat mendeteksi setiap objek yang ada didalam gambar. Sedangkan untuk penggunaan model TinyYOLOv3 tidak mendapatkan hasil yang maksimal karena model tersebut hanya mampu mendeteksi empat objek dengan tingkat akurasi yang rendah. Untuk mengimplementasikan penggunaan deteksi gambar yang terbaik sebaiknya digunakan model YOLOv3 dengan menggunakan *shortcut connections* yang dapat memberikan performa yang lebih baik dengan objek yang lebih kecil.

DAFTAR PUSTAKA

- [1] Kaiming. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. arXiv:1512.03385 [cs], Dec 2015.
- [2] Y. Hu, A. E. G. Huber, J. Anumula, and S. Liu. Overcoming the vanishing gradient problem in plain recurrent networks. CoRR, abs/1801.06105, 2018.
- [3] Kamal, Amro. YOLO, YOLOv2 and YOLOv3: All You want to know. Url : <https://amrokamal-47691.medium.com/yolo-yolov2-and-yolov3-all-you-want-to-know-7e3e92dc4899>.
- [4] RidgeRun. 7 Desember 2020. GstInference with TinyYoloV3 architecture. https://developer.ridgerun.com/wiki/index.php?title=GstInference/Supported_architectures/Tiny_YoloV3.
- [5] Hussin, R., Juhari, M.R., Kang, N.W., Ismail, R.C. and Kamarudin, A. (2012) ‘Digital image processing techniques for object detection from complex background image’, in International Symposium on Robotics and Intelligent Sensors, Vol. 41, pp.340–344.
- [6] Torrent, A., Lladó, X., Freixenet, J. and Torralba, A. (2013) ‘A boosting approach for the simultaneous detection and segmentation of generic objects’, Journal of Pattern Recognition Letters, Vol. 34, No. 13, pp.1490–1498.
- [7] Malagon-Borja, L. and Fuentes, O. (2009) ‘Object detection using image reconstruction with PCA’, Journal of Image and Vision Computing, Vol. 27, Nos. 1–2, pp.2–9.
- [8] Everingham M, Eslami SA, Van Gool L, Williams CK, Winn J, Zisserman A. The pascal visual object classes challenge: a retrospective. Int J Computer Vision. 2015;111(1):98–136.
- [9] Everingham M, Van Gool L, Williams CK, Winn J, Zisserman A. The pascal visual object classes (voc) challenge. Int J Computer Vision. 2010;88(2):303–38.
- [10] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Rabinovich A. Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2015, p. 1–9.
- [11] Zhao L, Li S. Object detection algorithm based on improved YOLOv3. Electronics. 2020;9(3):537.

- [12] Syed NR. A PyTorch implementation of YOLOv3 for real time object detection (Part 1). [Internet] [Updated Jun 30 2020]. <https://nrsyed.com/2020/04/28/a-pytorch-implementation-of-yolov3-for-real-time-object-detection-part-1/>. Accessed 02 Feb 2021.