

Departemen Teknik Komputer - FTEIC  
Institut Teknologi Sepuluh Nopember

**EC184701 - PRA TUGAS AKHIR 2 SKS**

Nama Mahasiswa : Agung Wicaksono  
Nomor Pokok : 07211740000002  
Semester : Ganjil 2020/2021  
Dosen Pembimbing : 1. Dr. Eko Mulyanto Yuniarno, S.T., M.T.  
2. Prof. Dr. Ir. Mauridhi Hery Purnomo, M.Eng.  
Judul Tugas Akhir : **Deteksi Pejalan Kaki pada Zebracross  
Untuk Peringatan Dini Pengendara Mobil  
Menggunakan Mask R-CNN**  
Uraian Tugas Akhir :

Dewasa ini, fitur keselamatan pada kendaraan roda empat atau mobil sudah sangat berkembang pesat. Hal tersebut terbukti dengan banyaknya produsen mobil yang menerapkan teknologi *seat belt*, *air bag*, *adaptive cruise control*, *electronic stability control*, *autonomous emergency braking*, *blind spot monitoring* dan lain sebagainya. Namun, fitur yang sudah disebutkan diatas dinilai masih kurang ramah bagi pejalan kaki. Terbukti menurut data dari WHO, terdapat 270.000 pejalan kaki meninggal dunia setiap tahun atau sekitar 22% dari seluruh korban meninggal akibat kecelakaan di jalan. Berawal dari permasalahan tersebut, penulis akan melakukan penelitian mengenai pendeteksian pejalan kaki pada *zebracross* untuk peringatan dini pengendara mobil sebagai topik tugas akhir. Pada tugas akhir ini, terdapat 2 objek yang akan dideteksi yaitu pejalan kaki dan *zebracross* dengan menggunakan metode Mask R-CNN. Hasil yang diharapkan dari tugas akhir kali ini adalah terdapat model yang memiliki akurasi yang tinggi dari dataset yang tersedia yaitu Caltech Pedestrian Dataset.

Dosen Pembimbing I

Dosen Pembimbing II

Prof. Dr. Ir. Mauridhi Hery Purnomo, M.Eng.  
NIP. 195809161986011001

Dr. Eko Mulyanto Yuniarno, S.T., M.T.  
NIP. 196806011995121009

Mengetahui,  
Departemen Teknik Komputer FTEIC - ITS  
Kepala,

Dr. Supeno Mardi Susiki Nugroho, ST., MT.  
NIP. 197003131995121001

# Deteksi Pejalan Kaki pada Zebracross untuk Peringatan Dini Pengendara Mobil Menggunakan Mask R-CNN

## I PENDAHULUAN

Mobil merupakan salah satu jenis kendaraan bermotor yang banyak terdapat di Indonesia. Pada tahun 2018 Badan Pusat Statistik mencatat terdapat 16.440.987 mobil penumpang yang berada di Indonesia. Dengan bertambahnya jumlah mobil di Indonesia dari tahun ke tahun, meningkatkan juga jumlah kecelakaan mobil. Fitur keselamatan dan keamanan pada mobil sangat penting bagi para pengendara dan penumpang, sehingga para produsen mobil berusaha meningkatkan teknologi keselamatan dan keamanan pada mobil buatannya. Sebagai contoh beberapa fitur keselamatan dan keamanan yang terdapat pada mobil antara lain, adaptive cruise control, hill strat assist, blind spot monitoring, electronic stability control dan lain sebagainya.

Menurut data dari WHO, terdapat 270.000 pejalan kaki meninggal dunia setiap tahun atau sekitar 22% dari seluruh korban meninggal akibat kecelakaan di jalan. Melihat kegiatan para pejalan kaki yang jarang berada di badan jalan, angka tersebut tentu cukup tinggi. Para pejalan kaki hanya menggunakan badan jalan ketika hendak menyebrang jalan lewat zebracross. Kelalaian dari pejalan kaki maupun pengendara mobil merupakan faktor utama mengapa angka kematian pejalan kaki cukup tinggi. Salah satu contoh kelalaian pejalan kaki adalah pada saat menyebrang jalan tidak memperhatikan kendaraan yang akan lewat dan atau melihat rambu serta lampu lalu lintas. Di sisi pengendara mobil, kelelahan, kurangnya fokus saat berkendara dan tidak memperhatikan rambu maupun marka dapat berakibat fatal baik kepada pejalan kaki dan pengendara lain.

Teknologi artificial intelligent sudah banyak disematkan pada mobil pada masa kini, dibuktikan dengan adanya teknologi adaptive cruise control, hill start assist dan lain sebagainya. Artificial intelligent khususnya deep learning tentu dapat digunakan untuk deteksi pejalan kaki di zebracross guna mengurangi jumlah korban akibat kecelakaan. Deteksi pejalan kaki dapat digabungkan dengan buzzer dan atau LED sebagai komponen output untuk mengingatkan kepada pengendara bahwa ada pejalan kaki yang sedang menyebrangi jalan serta mengembalikan fokus untuk berkendara.

Berdasarkan penjelasan yang telah dipaparkan, dapat dirumuskan beberapa rumusan masalah sebagai berikut:

1. Cukup tingginya angka kematian pejalan kaki akibat kecelakaan lalu lintas.
2. Belum adanya deteksi pejalan kaki di zebracross untuk peringatan dini kepada pengendara mobil.

Penelitian mengenai deteksi pejalan kaki di jalan raya sudah ada sebelumnya seperti Real-Time Pedestrian Detection With Deep Network Cascades (Anelia Angelova et al.) [1], dimana pada penelitian ini dihasilkan average miss rate sebesar 26,2% yang berjalan secara *real time*. Selanjutnya terdapat penelitian yang berjudul Pedestrian Detection: The Elephant In The Room (Irtiza Hasan et al.). [2]. Pada penelitian ini membandingkan antara detektor objek secara umum yaitu *Cascade R-CNN* dengan detektor objek khusus pejalan kaki pada dataset yang berbeda. Pada penelitian yang berjudul Fast Vehicle and Pedestrian Detection Using Improved Mask R-CNN Vincent Vanhoucke (Chenchen Xu et al.). [3] dijelaskan mengenai deteksi terhadap kendaraan yang bergerak dan pejalan kaki. Penggunaan Restnet-86 untuk menggantikan Restnet-101 sebagai *backbone* membuat deteksi lebih akurat pada dataset yang terdapat pada MS COCO. Pada MASK R-CNN for Pedestrian Crosswalk Detection and Instance Segmentation (Mon Arjay Malbog). [4] deteksi hanya dilakukan pada *pedestrian crosswalk* atau *zebracross* saja. *Taining data* dilakukan dengan Mask R-CNN untuk deteksi objek serta Restnet-101 sebagai *backbone*. Selanjutnya pada A Pedestrian Detection and Tracking System Based on Video Processing Technology (Chen, Yuanyuan et al.) [5]. dijelaskan mengenai deteksi pada pejalan kaki, pelacakan penghitungan serta peringatan resiko pada sebuah segemen video.

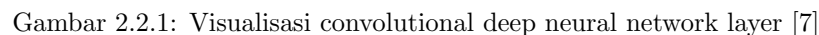
Penelitian yang telah ada dibandingkan dengan penelitian ini, penelitian sebelumnya berfokus pada satu objek saja yaitu pejalan kaki dan lebih mengutamakan tingkat akurasi yang tinggi. Selain itu, kamera ditempatkan pada posisi yang tetap atau tidak bisa dipindah-pindahkan.

Adapun tujuan dari penelitian Tugas Akhir ini adalah mendeteksi adanya pejalan kaki di zebracross untuk peringatan dini kepada pengendara mobil guna mengurangi angka kematian pejalan kaki akibat kecelakaan lalu lintas.

## 2.1 Deep Learning

## 2.2 Convolutional Neural Network

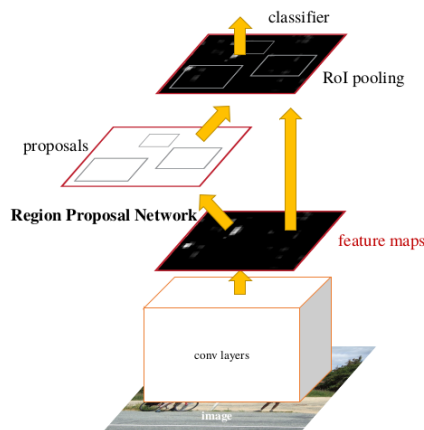
Asumsikan *input* dari *neural network* adalah sebuah gambar berwarna dengan ukuran 32x32 pixel dan kedalaman sebesar 3 (RGB). Jika *Input layer* disambungkan dengan hanya satu neuron (misalkan *hidden layer* pada *Multi-Layer perceptron*) maka diperlukan bobot koneksi sebesar 32x32x3. Jika kita ingin menambahkan neuron lagi pada *hidden layer*, maka dibutuhkan bobot koneksi sebesar 32x32x3 lagi, atau sebesar 32x32x3x2 parameter. Terdapat metode untuk mengurangi bobot koneksi, yaitu dengan menggunakan region lokal dari gambar. Gambar dengan ukuran 32x32x3 dapat diubah menjadi ukuran 5x5x3 dengan konvolusi. Dengan neuron sejumlah 32x32 pada *hidden layer*, maka jumlah koneksi akan berkurang drastis dari *input layer* 32x32x3 dengan 32x32 neuron menjadi 5x5x3 dengan 32x32 neuron. Gambar 2.2.1 merupakan visualisasi dari konvolusi pada CNN seperti yang sudah dijelaskan.



3

## 2.3 Faster R-CNN

Perkembangan detektor objek pada masa kini sejalan dengan perkembangan algoritma dari proposal region untuk menghipotesis letak dari suatu objek. Kemajuan dari detektor objek seperti pada arsitektur SPPNet dan Fast R-CNN telah mengurangi waktu yang diperlukan untuk proses pendeteksian, dengan mengekspos komputasi proposal region sebagai sebuah *bottleneck*. Pada Faster R-CNN diperkenalkan *Region Proposal Network* (RPN), yang dapat membagikan *full-image convolutional features* dengan jaringan deteksi, dimana hal tersebut tidak membutuhkan lagi proposal region. RPN adalah sebuah *fully convolutional network* yang secara simultan dapat memprediksi batas dari sebuah objek sekaligus skor objektivitas pada setiap posisi. RPN dilatih secara *end-to-end* sehingga dapat menghasilkan proposal region yang mempunyai kualitas tinggi dari yang dihasilkan oleh Fast R-CNN untuk deteksi. Faster R-CNN sebenarnya adalah penggabungan dari RPN dengan Fast-RCNN menjadi satu jaringan untuk membagikan hasil dari fitur konvolusi masing-masing.

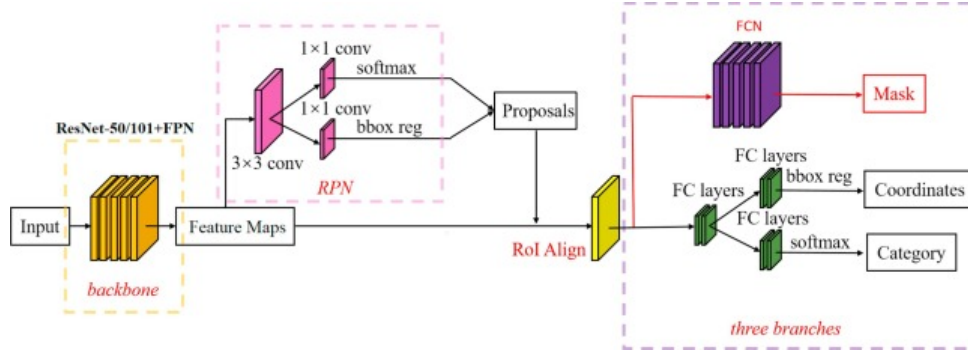


Gambar 2.3.1: Arsitektur Faster R-CNN [8]

Sistem deteksi objek pada Faster R-CNN mempunyai dua modul yang dijalankan, yaitu *deep convolutional neural network* yang akan mengusulkan beberapa region serta Fast R-CNN yang akan menggunakan region yang telah diusulkan. Keseluruhan sistem adalah sebuah jaringan yang terpadu seperti yang dimodelkan pada gambar 2.3.1. Dengan menggunakan terminologi populer terbaru, yaitu mekanisme '*attention*' modul RPN akan memberitahu pada module Fast R-CNN bagian mana yang seharusnya dilihat dan diproses. RPN akan mengambil sebuah gambar dengan ukuran berapapun sebagai input dan output satu set segiempat proposal objek. Untuk menghasilkan proposal region, dilakukan pergeseran jaringan kecil melalui *convolutional feature map output* oleh layer terakhir dari konvolusi yang dibagikan. Jaringan kecil ini mengambil input sebesar  $n \times n$  *spatial window* dari masukan *convolutional feature map*. Setiap *sliding window* akan dipetakan menjadi fitur dimensi yang lebih rendah. Fitur tersebut akan dibagi menjadi 2 *fully-connected layer* yaitu *box-regression layer* dan *box-classification layer*. [8]

## 2.4 Mask R-CNN

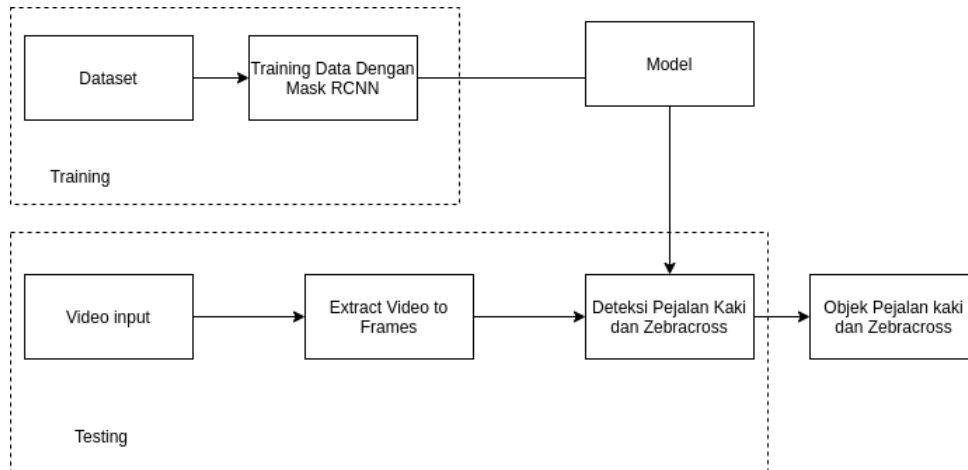
Mask R-CNN adalah sebuah framework untuk segmentasi instance objek. Pendekatan metode ini dilakukan dengan mendeteksi objek dalam gambar sekaligus menghasilkan mask segmentasi dengan kualitas yang tinggi untuk sebuah instance. Mask R-CNN merupakan perluasan dari Faster R-CNN dengan menambahkan branch untuk memprediksi mask objek secara paralel terhadap branch yang sudah ada untuk pengenalan bounding box. Mask R-CNN adalah metode yang sederhana dalam proses training dengan menambahkan sedikit overhead ke Faster R-CNN dan berjalan pada 5 fps. Selain itu, misalnya memungkinkan kita untuk memperkirakan pose manusia dalam kerangka yang sama.



Gambar 2.1.3: Arsitektur Mask R-CNN. [9]

Mask R-CNN mengadopsi dua langkah prosedur yang sama dengan faster R-CNN yaitu proses pengidentifikasian (RPN) pada langkah pertama, prediksi kelas dan *box offset* secara paralel pada langkah kedua serta juga dihasilkan *binary mask* pada setiap RoI. Sistem seperti itu tentu sangat kontras dengan sistem yang sudah ada sebelumnya, dimana klasifikasi bergantung pada *mask prediction*. Selama proses *training*, *multi-task loss* didefinisikan sebagai berikut pada setiap RoI  $L = L_{cls} + L_{box} + L_{mask}$ . Klasifikasi dan *bouding loss* merupakan *loss* indentifikasi seperti yang sudah dijelaskan di atas. Sedangkan *mask loss* memungkinkan jaringan dapat menghasilkan mask pada setiap kelas tanpa adanya persaingan antar kelas. Oleh sebab itu diperlukan cabang klasifikasi khusus untuk memprediksi label kelas yang akan digunakan pada pemilihan output *mask*. Hal tersebut tentu berbeda dari biasanya ketika menggunakan FCN pada segmentasi semantic yang biasanya menggunakan per-pixel *softmax* dan *multinomial cross-entropy loss*. [10]

### III METODOLOGI



Gambar 3.1: Blok diagram metodologi.

#### 3.1 Alat dan Material

Pada tugas akhir kali ini penulis menggunakan dataset dari Caltech Pedestrian Dataset. Caltech Pedestrian Dataset sendiri adalah dataset yang terdiri dari sekitar 10 jam video 640x480 30Hz yang diambil dari kendaraan yang mengemudi melalui lalu lintas di lingkungan perkotaan. Sekitar 250.000 frame (dalam 137 segmen dengan panjang kira-kira satu menit) dengan total 350.000 box bounding dan 2300 pejalan kaki diberi keterangan. Namun bounding box dari dataset tersebut hanya melabeli pedestrian saja, belum ada label untuk zebracross seperti yang penulis akan lakukan. Sehingga penulis perlu mengambil file berjenis gambar saja tanpa file json bounding box objek. Selanjutnya penulis akan melabeli objek berupa pedestrian dan zebracroos pada gambar yang diperoleh dari dataset tersebut. Sedangkan untuk pembagian rasio data training sebesar 70% dan data testing sebesar 30%. Gambar 3.1.1 merupakan sebagian contoh penampakan dari Caltech Pedestrian Dataset. Dataset tersebut akan diolah

lewat layanan cloud Google berupa Google Colab. Spesifikasi komputer yang digunakan oleh layanan Google Colab adalah Intel Xeon Processor @2.3 GHz, GPU Tesla K80 12 GB GDDR5 VRAM, dan RAM sebesar 16 GB. Sedangkan untuk spesifikasi dari komputer penulis adalah AMD FX-7500 Processor @2.1 GHz, GPU AMD Radeon R7 2GB GDDR3 VRAM dan RAM 8 GB. [11]



Gambar 3.1.1: Beberapa contoh gambar dari Caltech Pedestrian Dataset.

### 3.2 Training

Pada tahap selanjutnya dilakukan training data menggunakan salah satu metode deep learning yaitu Mask R-CNN. Mask R-CNN merupakan ekstensi dari Faster R-CNN dengan menambahkan cabang baru yaitu Fully Convolutional Network. Data anotasi dan gambar hasil preprocessing akan diolah menggunakan algoritma Faster R-CNN hingga menghasilkan kelas dan koordinat dari objek. Peran dari Fully Connected Network pada Mask R-CNN adalah untuk menghasilkan masking dari objek yang dideteksi. Pada proses kali ini dihasilkan *weight* yang akan digunakan untuk pendeteksian objek berupa pedestrian dan zebra-cross.

### 3.3 Testing

Pada tahap testing, model yang sudah dihasilkan dari proses training akan dilakukan percobaan mengenai keakuratan hasil pendeteksian objek berupa pedestrian dan zebra-cross. Dari hasil testing dapat dilihat tingkat akurasi model training yang telah kita lakukan dengan menggunakan algoritma loss function. Apabila akurasi masih dirasa kurang yaitu dibawah 70%, maka akan dilakukan training kembali dengan beberapa perubahan sehingga hasil yang akan didapatkan selanjutnya memiliki tingkat akurasi yang lebih baik.

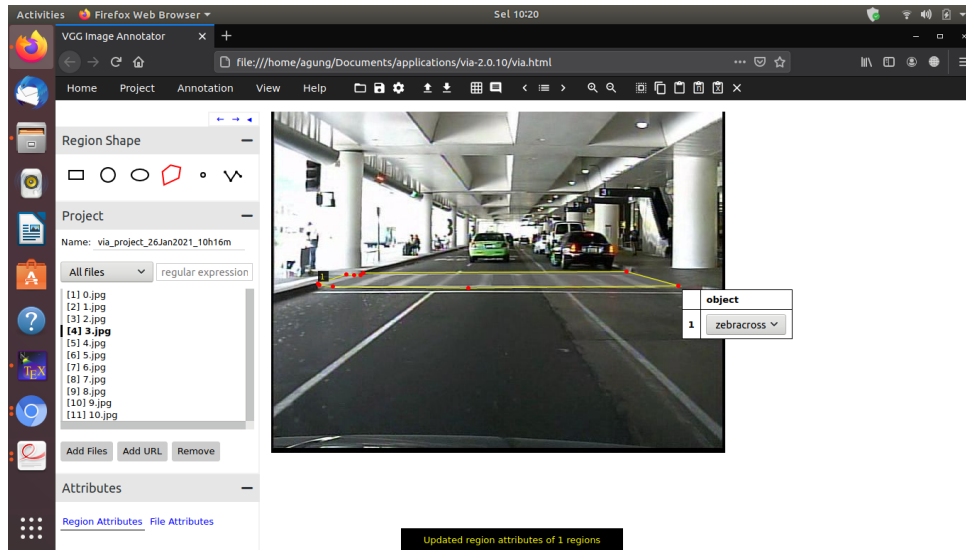
## IV HASIL YANG DIHARAPKAN

### 4.1 Hasil Yang Diharapkan

Pada tugas akhir ini, diharapkan dapat menghasilkan model yang memiliki akurasi yang tinggi untuk mendeteksi pejalan kaki dan zebra-cross untuk peringatan dini kepada pengendara mobil. Progres yang sudah dilakukan penulis sampai sekarang adalah menyiapkan dataset untuk training data Mask R-CNN. Pada deteksi pedestrian, sudah bisa dilakukan dengan menggunakan dataset dari MS COCO. Dikarenakan dataset yang digunakan pada tugas akhir penulis adalah Caltech Pedestrian Dataset maka perlu dilakukan preprocessing terhadap dataset tersebut.

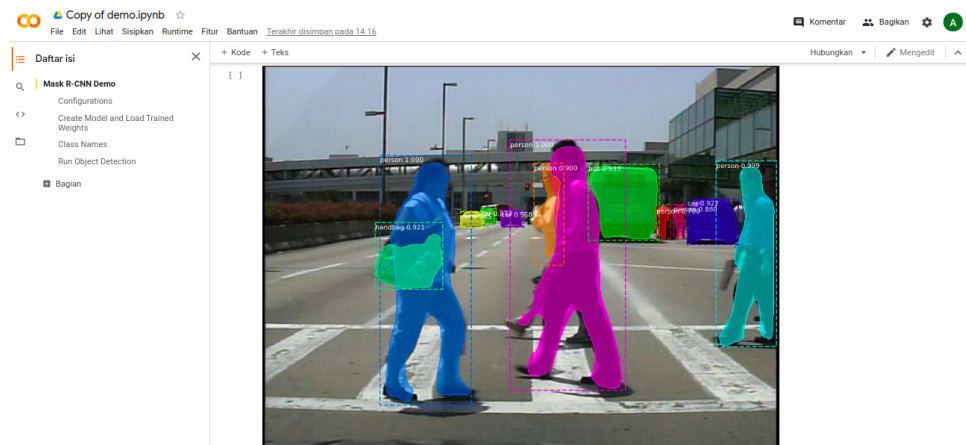
### 4.2 Hasil Pendahuluan

Pada pengerjaan awal, penulis telah memilih beberapa gambar yang akan dijadikan dataset yang berasal dari Caltech Pedestrian Dataset. Penulis juga melakukan labeling untuk objek yang akan di deteksi yaitu pejalan kaki dan *zebra-cross* dengan menggunakan *VGG Image Annotator* seperti tertampil pada gambar 4.2.1.



Gambar 4.2.1: Labeling gambar dengan VGG Image Anotator.

Selain itu penulis juga sudah mencoba deteksi pejalan kaki dengan menggunakan *pubic surce code* [12] dengan menggunakan dataset dari MS COCO seperti tertampil pada gambar 4.2.2.



Gambar 4.2.2: Hasil deteksi pejalan kaki dengan Mask R-CNN.

## V RENCANA KERJA

No.	Kegiatan	Minggu															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	Studi literatur																
2	Konfigurasi label dan membagi data <i>training</i> & <i>test</i>																
3	Proses <i>preprocessing</i>																
4	Proses <i>training</i>																
5	Proses <i>testing</i>																
6	Dokumentasi dan pembuatan laporan																

## VI DAFTAR PUSTAKA

### References

- [1] A. Angelova, A. Krizhevsky, V. Vanhoucke, A. Ogale, and D. Ferguson, “Real-time pedestrian detection with deep network cascades,” 2015.
- [2] I. Hasan, S. Liao, J. Li, S. Akram, and L. Shao, “Pedestrian detection: The elephant in the room,” 03 2020.
- [3] C. Xu, G. Wang, S. Yan, J. Yu, B. Zhang, S. Dai, Y. Li, and L. Xu, “Fast vehicle and pedestrian detection using improved mask r-cnn,” *Mathematical Problems in Engineering*, vol. 2020, 2020.
- [4] M. A. Malbog, “Mask r-cnn for pedestrian crosswalk detection and instance segmentation,” pp. 1–5, 2019.
- [5] Y. Chen, S. Guo, B. Zhang, and K.-L. Du, “A pedestrian detection and tracking system based on video processing technology,” *Proceedings - 2013 4th Global Congress on Intelligent Systems, GCIS 2013*, pp. 69–73, 12 2013.
- [6] N. Rusk, “Deep learning,” *Nature Methods*, vol. 13, no. 1, pp. 35–35, 2016.
- [7] S. Albawi, T. A. Mohammed, and S. Al-Zawi, “Understanding of a convolutional neural network,” in *2017 International Conference on Engineering and Technology (ICET)*, pp. 1–6, 2017.
- [8] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” *arXiv preprint arXiv:1506.01497*, 2015.
- [9] Y. Yu, K. Zhang, L. Yang, and D. Zhang, “Fruit detection for strawberry harvesting robot in non-structural environment based on mask-rcnn,” *Computers and Electronics in Agriculture*, vol. 163, p. 104846, 2019.
- [10] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask r-cnn,” pp. 2961–2969, 2017.
- [11] C. Xu, G. Wang, S. Yan, J. Yu, B. Zhang, S. Dai, Y. Li, and L. Xu, “Fast vehicle and pedestrian detection using improved mask r-cnn,” *Mathematical Problems in Engineering*, vol. 2020, 2020.
- [12] I. Matterport, “Mask r-cnn for object detection and segmentation.” [https://github.com/matterport/Mask\\_RCNN#mask-r-cnn-for-object-detection-and-segmentation](https://github.com/matterport/Mask_RCNN#mask-r-cnn-for-object-detection-and-segmentation), 2019. Accessed : 2020-11-4.