

## PEMBUATAN PENDETEKSI OBYEK DENGAN METODE YOU ONLY LOOK ONCE (YOLO) UNTUK AUTOMATED TELLER MACHINE (ATM)

ABI RACHMAN WASRIL, M. SHIDDIQ GHOZALI, DAN M. BANU MUSTAFA  
Program Studi Informatika, Fakultas Ilmu Rekayasa  
Universitas Paramadina

*Perkembangan Teknologi Informasi dan Komunikasi begitu pesat di zaman sekarang ini. Diikuti pula dengan perkembangan di bidang Artificial Intelligence (AI) atau Kecerdasan Buatan. Di Indonesia sendiri masih belum begitu populer dikalangan masyarakat akan tetapi perusahaan-perusahaan IT berlomba-lomba menciptakan inovasi dibidang Kecerdasan Buatan dan penerapan Kecerdasan Buatan disegala aspek kehidupan. Contoh kasus di Automated Teller Machine (ATM), seringkali terjadi kejahatan di ATM seperti pengintaian nomor pin, skimming, lebanese loop dan kejahatan lainnya. Walaupun di ATM sudah terdapat CCTV akan tetapi penjahat menggunakan alat bantu untuk menutupi wajahnya seperti helm, topi, masker dan kacamata hitam. Biasanya didepan pintu masuk ATM terpampang larangan untuk tidak menggunakan helm, topi, masker dan kacamata hitam serta tidak membawa rokok. Akan tetapi larangan itu masih tetap ada yang melanggar, dikarenakan tidak ada tindak lanjut ketika seseorang menggunakan benda-benda yang dilarang dibawa kedalam ATM. Oleh karena itu penulis membuat sistem pendeteksi obyek di bidang Kecerdasan Buatan untuk mendeteksi benda-benda yang dilarang digunakan ketika berada di ATM. Salah satu metode yang digunakan untuk menciptakan Object Detection yaitu You Only Look Once (YOLO). Implementasi ide ini tersedia pada DARKNET (open source neural network). Cara kerja YOLO yaitu dengan melihat seluruh gambar sekali, kemudian melewati jaringan saraf sekali langsung mendeteksi object yang ada. Oleh karena itu disebut You Only Look Once (YOLO). Pada penelitian ini, penulis membuat sistem yang masih dalam bentuk pengembangan, sehingga menjalankannya masih menggunakan command prompt.*

**Keywords : Automated Teller Machine (ATM), Kecerdasan Buatan, Pendeteksi Obyek, You Only Look Once (YOLO)**

### PENDAHULUAN

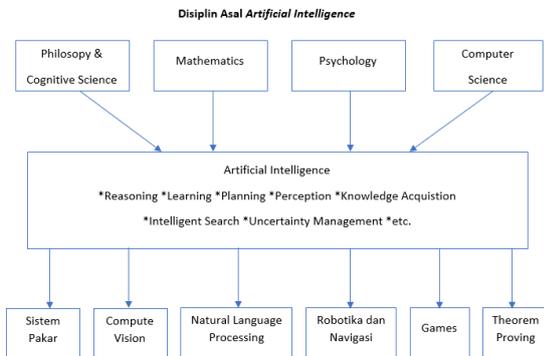
*Artificial Intelligence (AI) atau Kecerdasan Buatan adalah ilmu yang dikembangkan dengan kombinasi banyak subyek. Dalam istilah sederhana, kecerdasan buatan adalah memberi mesin dengan kecerdasan manusia, mensimulasikan pemikiran manusia sehingga membantu orang memecahkan masalah dan untuk mewujudkan aplikasi yang lebih canggih seperti produksi dan kehidupan manusia yang dibantu computer. Kecerdasan buatan, cabang ilmu komputer, dianggap sebagai salah satu dari tiga teknologi yang paling maju (Rekayasa genetika, ilmu Nano dan kecerdasan Buatan) pada abad ke-21 [1]. Kecerdasan buatan (*Artificial Intelligence*) memodifikasi atau mengadaptasi computer dari tindakan manusia (termasuk prediksi atau control robot), se-*

hingga bisa menjadi lebih akurat [2]. Kecerdasan buatan telah mencapai perkembangan yang sangat cepat di 30 tahun terakhir dan telah banyak digunakan di banyak negara bidang akademik [1]. Supaya computer dapat bertindak atau serupa dengan manusia, maka komputer harus diberi bekal pengetahuan, dan mempunyai kemampuan untuk menalar. Penerapan AI beraneka ragam. Tujuan dari sistem kecerdasan buatan dapat dibagi dalam 4 kategori :

1. Sistem yang dapat berfikir seperti manusia (Bellman, 1978)
2. Sistem yang dapat berfikir secara rasional (Winston, 1992)
3. Sistem yang dapat beraksi seperti manusia (Rich and Knight, 1991)

4. Sistem yang dapat bereaksi secara rasional (Nilsson, 1998)

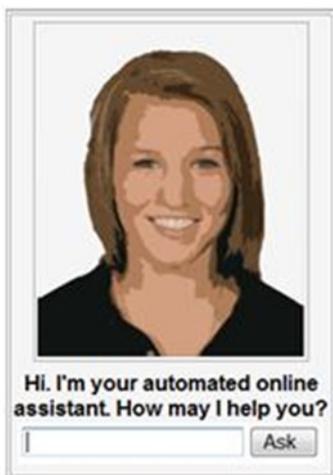
Secara garis besar bidang ilmu yang dipelajari dalam bidang AI bisa dilihat pada gambar di bawah ini :



Gambar 1. Domain Area Kecerdasan Buatan [3]

1. Natural Language Processing (NLP)

NLP mempelajari bagaimana bahasa alami itu diolah sedemikian rupa hingga user dapat berkomunikasi dengan komputer. Konsentrasi ilmu ini adalah interaksi antara komputer dengan bahasa natural yang digunakan manusia, yakni bagaimana komputer melakukan ekstraksi informasi dari *input* yang berupa *natural language* dan atau menghasilkan output yang juga berupa *natural language*, misalnya pada *system automated online assistant* seperti pada gambar ... dan deteksi *email spam* yang cerdas [3].

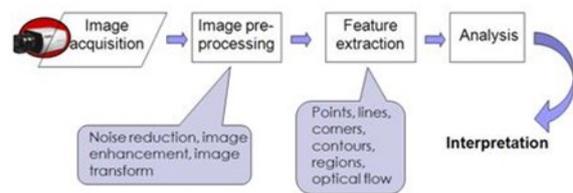


Gambar 2. Penerapan NLP pada *automated online assistant* berbasis web [3]

2. Computer Vision

Cabang ilmu ini erat kaitannya dengan pembangunan arti atau makna dari *image* ke obyek secara fisik. Yang dibutuhkan didalamnya adalah

metode – metode untuk memperoleh, melakukan proses, menganalisa dan memahami *image*. Apabila cabang ilmu ini dikombinasikan dengan Kecerdasan Buatan secara umum akan mampu menghasilkan sebuah *visual intelligence system* [3]. Akuisi dan pemrosesan informasi berupa *vision* dapat ditampilkan pada gambar berikut :



Gambar 3. Model persepsi visual pada komputer vision [3]



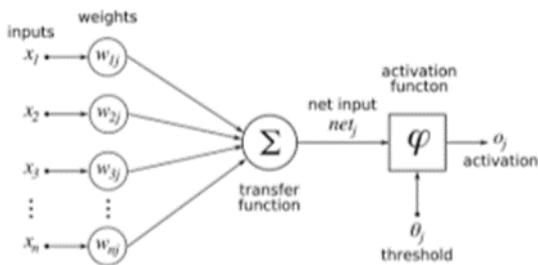
Gambar 4. Contoh penerapan komputer vision untuk identifikasi wajah [3]

TINJAUAN PUSTAKA

1. Artificial Neural Network (ANN)

*Artificial Neural Network* (ANN) atau Jaringan Syaraf Tiruan merupakan sebuah teknik atau pendekatan pengolahan informasi yang terinspirasi oleh cara kerja sistem saraf biologis, khususnya pada sel otak manusia dalam memproses informasi. Elemen kunci dari teknik ini adalah struktur sistem pengolahan informasi yang bersifat unik dan beragam untuk tiap aplikasi. *Neural Network* terdiri dari sejumlah besar elemen pemrosesan informasi (neuron) yang saling terhubung dan bekerja bersama-sama untuk menyelesaikan sebuah masalah tertentu, yang pada umumnya adalah masalah klasifikasi ataupun prediksi Komunikasi. Cara kerja *Neural Network* dapat dianalogikan sebagaimana halnya manusia belajar dengan menggunakan contoh atau yang disebut sebagai *supervised learning*. Sebuah *Neural Network* dikonfigurasi untuk aplikasi tertentu, seperti

pengenalan pola atau klasifikasi data, dan kemudian disempurnakan melalui proses pembelajaran (training). Proses belajar yang terjadi dalam sistem biologis melibatkan penyesuaian koneksi sinaptik yang ada antara neuron, dalam halnya pada *Neural Network* penyesuaian koneksi sinaptik antar *neuron* dilakukan dengan menyesuaikan nilai bobot yang ada pada tiap konektivitas baik dari *input*, *neuron* maupun *output* [4].



Gambar 5. *Artificial Neural Network* [4]

*Neural Network* memproses informasi berdasarkan cara kerja otak manusia. Dalam hal ini *Neural Network* terdiri dari sejumlah besar elemen pemrosesan yang saling terhubung dan bekerja secara paralel untuk memecahkan suatu masalah tertentu. Di sisi lain, komputer konvensional menggunakan pendekatan kognitif untuk memecahkan masalah. Dimana cara pemecahan masalah haruslah sudah diketahui sebelumnya untuk kemudian dibuat menjadi beberapa instruksi kecil yang terstruktur. Instruksi ini kemudian dikonversi menjadi program komputer dan kemudian ke dalam kode mesin yang dapat dijalankan oleh komputer [4].

**2. Convolutional Neural Network (CNN)**

*Convolutional Neural Network* (CNN) adalah salah satu jenis *Neural Network* (NN) yang biasa digunakan pada data image. CNN bisa digunakan untuk mendeteksi dan mengenali object pada sebuah image. Secara garis besar CNN tidak jauh beda dengan *Neural Network* biasanya. CNN terdiri dari neuron yang memiliki weight, bias dan activation function. Yang membedakan antara CNN dengan NN ialah arsitektur dari CNN dibagi menjadi dua bagian besar, yaitu *Feature Extraction Layer* dan *Fully Connected Layer* [5].

**a. Feature Extraction Layer**

Proses yang terjadi pada bagian ini adalah melakukan “*encoding*” dari sebuah image menjadi features yang berupa angka-angka yang merepresentasikan image tersebut (*Feature Extraction*). *Feature Extraction Layer* terdiri dari dua

bagian yaitu *Convolutional Layer* dan *Pooling* [5].

- **Convolutional Layer**

*Convolutional Layer* terdiri dari neuron yang tersusun sedemikian rupa sehingga membentuk sebuah filter dengan panjang dan tinggi (pixels). Sebagai contoh, layer pertama pada feature extraction layer biasanya adalah convolution layer dengan ukuran 5x5x3. Panjang 5 pixels, tinggi 5 pixels dan tebal atau jumlah 3 buah sesuai dengan channel dari image tersebut [5].

- **Pooling Layer**

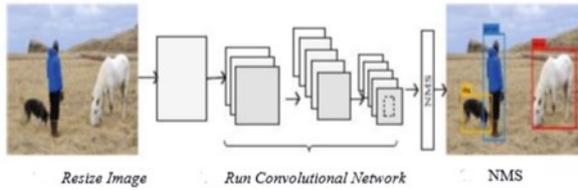
*Pooling Layer* biasanya berada setelah *Convolutional Layer*. Pada prinsipnya pooling layer terdiri dari sebuah filter dengan ukuran dan stride tertentu yang akan bergeser pada seluruh area feature map. Pooling yang biasa digunakan adalah Max Pooling dan Average Pooling. Sebagai contoh jika kita menggunakan Max Pooling 2x2 dengan stride 2, maka pada setiap pergeseran filter, nilai maximum pada area 2x2 pixel tersebut yang akan dipilih, sedangkan Average Pooling akan memilih nilai rata-ratanya [5].

**b. Fully Connected Layer**

Activation map yang dihasilkan dari feature extraction layer masih berbentuk multidimensional array, sehingga harus melakukan reshape activation map menjadi sebuah vector agar bisa digunakan sebagai input dari fully connected layer. Layer ini memiliki hidden layer, activation function, output layer, dan loss function. Layer ini adalah layer yang biasanya digunakan dalam penerapan multi layer perceptron dan bertujuan untuk melakukan transformasi pada dimensi data agar data dapat diklasifikasikan secara linear. Setiap neuron pada convolution layer perlu ditransformasi menjadi data satu dimensi terlebih dahulu sebelum dapat dimasukkan ke dalam sebuah fully connected layer. Karena hal tersebut menyebabkan data kehilangan informasi spasialnya dan tidak reversibel, sedangkan fully connected layer hanya dapat diimplementasikan di akhir jaringan [6].

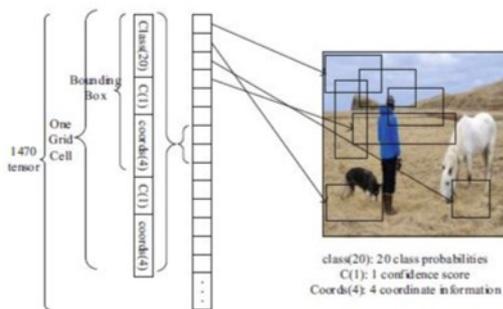
**c. You Only Look Once (YOLO)**

YOLO ditunjukkan pada gambar 14. Pertama – tama, gambar input diubah menjadi 448 x 448. Kemudian menjalankan single convolutional network pada gambar, kotak pembatas serta skor keyakinannya akan diperoleh. Pada akhirnya, NMS (non-max suppression) digunakan untuk meng-threshold deteksi yang dihasilkan oleh kepercayaan model, probabilitas kelas final dan koordinat bounding box diperoleh seperti yang ditunjukkan pada bagian paling kanan dari gambar berikut:



Gambar 6. Proses Deteksi YOLO [7]

YOLO membagi gambar input ke  $7 \times 7$  grid. Setiap sel grid sesuai dengan 2 kotak pembatas. Setiap sel grid memprediksi satu set probabilitas kelas (termasuk 20 kelas) terlepas dari jumlah kotak. Dan setiap kotak pembatas terdiri dari 5 prediksi:  $x, y, w, h$ , dan keyakinan. Koordinat  $(x, y)$  mewakili pusat kotak relatif terhadap batas – batas sel grid.  $(w, h)$  mewakili lebar dan tinggi kotak relatif terhadap keseluruhan gambar. Dengan demikian, setiap grid memiliki prediksi  $(4 + 1) \times 2 + 20 = 30$  dan seluruh gambar memiliki  $7 \times 7 \times 30 = 1470$  prediksi. YOLOs CNN mencakup 25 convolutional layers, 4 pooling layers, 1 dropout layer, 1 fully connected layer and 1 lapisan deteksi. Dimensi dari lapisan yang terhubung sepenuhnya adalah 1470, yang menunjukkan posisi dan keyakinan kotak prediksi yang terikat dan probabilitas kelasnya. Untuk CNN yang diinisialisasi secara acak, output dari lapisan yang terhubung sepenuhnya juga acak pada awal proses pelatihan, yaitu ada kesenjangan besar antara kotak prediksi dan kotak ground truth di awal. Sketsa yang ditunjukkan pada gambar 16, mengungkapkan pemetaan antara lapisan yang terhubung penuh dan kotak prediksi pada gambar. Diperlukan banyak iterasi untuk mendekati kotak ground truth, menghasilkan kecepatan konvergensi yang lambat. Disisi lain, YOLO mengusulkan kotak batas yang jauh lebih sedikit, hanya  $7 \times 7 \times 2 = 98$  per gambar, jadi kesalahan lokalikasinya besar [7].



Gambar 7. Pemetaan antara layer yang terhubung penuh dan kotak prediksi YOLO [7]

Abi R Wasril, M. Shiddiq Ghozali, dan M. Banu Mustafa

#### d. Automated Teller Machine (ATM)

Automated Teller Machine (ATM) adalah komputerisasi perangkat telekomunikasi yang menyediakan pelanggan lembaga keuangan dengan akses ke transaksi keuangan di ruang publik. Dengan menggunakan ATM, pelanggan bisa akses rekening bank mereka untuk penarikan tunai dan cek saldo akun mereka. ATM pertama dipasang di Amerika Serikat pada tahun 1969. ATM memiliki pengaruh positif pada pertumbuhan mata uang nominal, tetapi efek ini tidak sangat kuat. Di antara semua layanan ATM bank dipertimbangkan sebagai layanan yang lebih menguntungkan karena menarik jumlah pelanggan non-bank. Struktur ATM terdiri atas komponen utama seperti CPU, kartu chip magnetik, PIN pembayaran, keamanan *crypto processor*, tombol fungsi, dan ruang besi. Sejak diperkenalkan, pada tahun 1967, pelaku telah merencanakan cara untuk mencoba mencuri uang. Karena ATM menghilangkan kebutuhan untuk keterlibatan manusia sepanjang waktu dan cenderung ditempatkan di tempat yang membuat mereka lebih banyak rentan diserang, mereka sering menjadi sasaran yang menarik pelaku. Kejahatan ATM tidak terbatas pada pencurian uang tunai di ATM. Banyak serangan ATM berusaha mendapatkan informasi pribadi konsumen, seperti nomor kartu dan nomor identifikasi pribadi (PIN) mereka. Menurut perkiraan oleh Retail Banking Research, disana lebih dari 2,2 juta ATM dikerahkan di seluruh dunia. Ini adalah angka yang diperkirakan akan melebihi 3 juta pada tahun 2016. Sebagai jumlah ATM yang digunakan meningkat, begitu juga frekuensi dan kecanggihan ancaman keamanan, membuat perkembangan pencegahan penipuan mengukur prioritas utama untuk keuangan lembaga (FI) dan produsen ATM [8].

#### METODE

Metode yang digunakan penulis pada penelitian ini yaitu menggunakan *You Only Look Once* (YOLO) yang merupakan salah satu metode untuk melakukan deteksi objek. Implementasi ide ini tersedia pada DARKNET (*open source neural network*). Cara kerja YOLO yaitu dengan melihat seluruh gambar sekali, kemudian melewati jaringan saraf sekali langsung mendeteksi obyek yang ada. Oleh karena itu disebut YOLO (*You Only Look Once*). Untuk melakukan deteksi obyek selain YOLO juga terdapat metode lain seperti Faster RCNN dan SSD.

**HASIL DAN PEMBAHASAN**

**1. Pengumpulan Data Set**

Pengumpulan dataset merupakan langkah awal sebelum YOLO diajarkan untuk mengenal benda – benda yang harus dideteksi. Adapun pengambilan dataset dapat dilakukan baik dengan membuat *video record* terlebih dahulu kemudian dipotong – potong menjadi iamge ataupun dengan mengunduh gambar – gambar dari internet. Berikut cara penulis untuk mengumpulkan dataset :

**a. Pengambilan sumber data video**

Gambar dibawah ini merupakan *flow* dari pengambilan sumber data berupa video yang akan digunakan sebagai langkah awal untuk mentraining YOLO sebelum melakukan *labelling images*. Dengan melakukan pengambilan sumber data video dalam aplikasi *video management tools*.

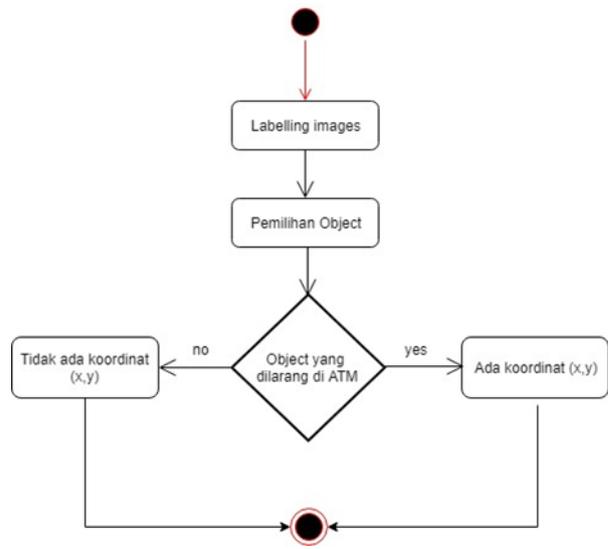


Gambar 8. Flow pengambilan sumber data video

**b. Labelling images**

Gambar dibawah merupakan *flow* dari *labelling images* yang digunakan untuk mengumpulkan data set training YOLO. Dalam hal ini nantinya, hasil dari *labelling* berupa koordinat (x, y). Jika gambar itu terdeteksi adanya obyek – obyek yang telah ditentukan seperti helm, topi, masker, rokok dan kaca mata hitam, maka akan menghasilkan koordinat x dan y, akan tetapi jika tidak ada obyek yang diten-

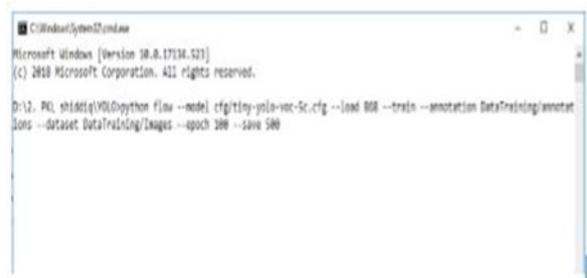
tukan maka tidak menghasilkan koordinat x dan y. Begitu cara YOLO mengenal obyek.



Gambar 9. Flow Labelling Images

**2. Training YOLO**

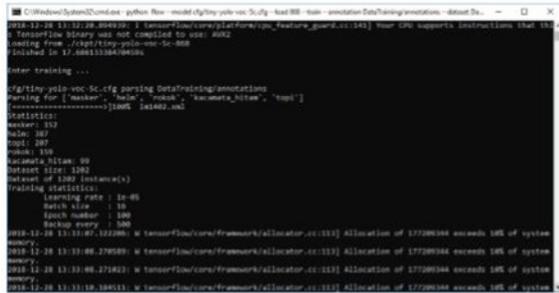
Data set yang telah ditentukan berupa koordinat – koordinat, yang dihasilkan dari *labelling images* kemudian di training agar YOLO dapat mengenal obyek – obyek yang telah ditentukan.



Gambar 10. Cara menjalankan training YOLO

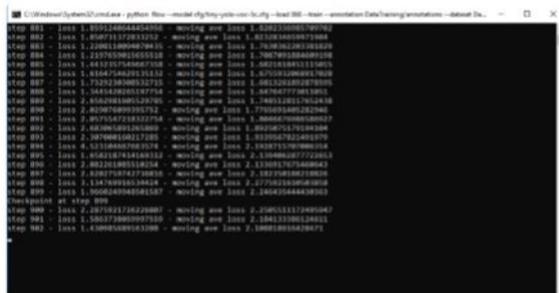
Dalam gambar ini load atau mulai trainingnya pada *check point* 868 dikarenakan penulis sudah melakukan training sebelumnya. Agar dataset yang sudah ditraining tidak hilang atau rusak. Akan tetapi jika ingin mengulang training dari awal maka perintahnya seperti berikut :

```
python flow --model cfg/tiny-yolo-voc-5c.cfg --load bin/tiny-yolo-voc-weights--train-annotation DataTraining/annotations--dataset DataTraining/Images--epoch 100 --save 500
```



Gambar 11. Keterangan jumlah dataset yang akan di training

Setelah menjalankan perintah training kemudian akan muncul data seperti gambar diatas. Terdapat jumlah dataset yang akan ditraining dan juga terdapat jumlah masing – masing obyek yang akan dikenali oleh YOLO.



Gambar 12. Check point 899 YOLO

Gambar ini menjelaskan pada check point 899 YOLO sudah mentraining dengan nave loss atau tingkat kesalahan untuk mendeteksi yaitu 2,246.

### 3. Running YOLO

Setelah dirasa cukup untuk training maka lakukan percobaan merenung YOLO untuk mengetahui hasilnya. Sebelumnya menjalankan dalam file VideoCapture1.py cari load dan diganti dengan checkpoint hasil training terakhir.

Ketika tidak ada obyek yang dimaksud seperti helm, topi, masker, kacamata hitam dan rokok maka YOLO akan menampilkan aman seperti gambar diatas. Akan tetapi jika ada obyek yang terdeteksi dan akan mengeluarkan suara maaf mohon lepas + nama obyek + Anda. Seperti gambar dibawah ini.



Gambar 13. Helm Terdeteksi

### KESIMPULAN

Berdasarkan rangkaian penelitian yang dijalankan, penulis mendapati beberapa poin kesimpulan dari penulisan ini. Berikut adalah poin – poin tersebut :

1. Penulis melakukan pembuatan *object detection* menggunakan YOLO dengan melakukan riset pada *source code* yang terbuka.
2. Aplikasi ini masih dalam bentuk pengembangan, sehingga harus menjalankannya dengan menggunakan *command prompt*.

### DAFTAR PUSTAKA

- [1] Z. X, N. Y and J. H, "Modified Object Detection Method on YOLO," *Second CCF Chinese Conference* , 2017 .
- [2] W. Budiharto, "School of Computer Science," Binus University, 06 06 2012. [Online]. Available: <http://socs.binus.ac.id/2012/06/06/mengenal-kecerdasan-buatan-kini-dan-akan-datang/>. [Accessed 27 December 2018].
- [3] A. Yanuar , "Universitas Gajah Mada Menara Ilmu Maching Learning," 25 06 2018. [Online]. Available: <http://machinelearning.mipa.ugm.ac.id/2018/06/25/fully-connected-layer-cnn-dan-implementasinya/>. [Accessed 03 January 2019 ].

- [4] H. Li, "On The Impact Artificial," *Information Recording Materials*, p. 18, 2018.
- [5] S. Sena, "A Medium Cooperation," Medium.com, 13 11 2017 . [Online]. Available: <https://medium.com/@samuelsena/pengenalan-deep-learning-part-7-convolutional-neural-network-cnn-b003b477dc94>. [Accessed 03 01 2019].
- [6] M. H. H. Sayed and Z. Riesjafari , "ATM Management prediction using Artificial Intelligence techniques," *A Suvery*, p. 4, 2017.
- [7] B. Sujith , "Crime Detection and Avoidnce in ATM: A New Framework," *International Journal of Computer Science and Information Technologies India*, 2014.
- [8] H. D. Widiputra , "Dosen.perbanas.id," Perbanas Institute, 12 10 2016. [Online]. Available: <https://dosen.perbanas.id/artificial-neural-network/>. [Accessed 03 01 2019].

