

# BAB I

## Pendahuluan

### 1.1. Latar Belakang

*Computer vision* atau dalam Bahasa Indonesia disebut sebagai visi komputer merupakan bidang keilmuan yang mempelajari tentang bagaimana komputer dapat memodelkan kerja sistem penglihatan manusia. Dalam sudut pandang ilmu teknik, *computer vision* mempunyai tujuan utama untuk membangun sistem yang mampu melakukan tugas yang sistem penglihatan manusia dapat lakukan atau bahkan dapat melebihi sistem penglihatan manusia (T. Huang, 1996). Pada masa sekarang ini *computer vision* sudah semakin berkembang sukses dan lebih praktikal dengan semakin banyak konferensi dan penelitian aktif, serta perusahaan besar yang turut berkontribusi dalam bidang *computer vision* (Yuille, 2012). Salah satu topik yang populer dan terus berkembang dibidang *computer vision* dalam beberapa tahun terakhir adalah pengenalan objek. Beberapa kontes penelitian mengenai pengenalan objek seperti PASCAL Visual Object Classes Challenge (PASCAL VOC), ImageNet Large Scale Visual Recognition Challenge (ILSVRC), dan Microsoft Common Object in Context Detection Challenge (COCO) yang diadakan tiap tahun menandakan bahwa topik mengenai pengenalan objek masih terus berkembang hingga saat ini. Seiring dengan perkembangannya, tingkat kesulitan yang dihadapi dalam topik pengenalan objek pun semakin kompleks. Salah satunya adalah jumlah data untuk proses *data learning* yang semakin banyak dan beragam, sebagai contoh pada *dataset* terbesar yang disediakan dalam ILSVRC mempunyai jumlah data sebanyak 14 juta citra digital dan 22 ribu kategori objek. Besarnya data dan proses komputasi yang rumit membuat diperlukannya metode yang tepat untuk dapat menangani permasalahan pengenalan objek. Salah satu yang paling populer untuk hal tersebut adalah dengan penggunaan metode *deep learning*.

*Deep learning* merupakan suatu kelas dari teknik *machine learning* yang memanfaatkan banyak lapisan atau *layer* dalam pemrosesan informasi untuk ekstraksi dan transformasi, serta untuk analisis pola dan klasifikasi terutama untuk memahami data seperti citra digital, suara, dan teks (Deng & Yu, 2014). Beberapa area pengembangan aplikasi berbasis *deep learning* antara lain adalah dalam bidang *computer vision*, pengenalan suara, pemrosesan bahasa alami, dan area lain yang berhubungan (Goodfellow, Bengio, & Courville, 2016). Tiga alasan penting yang mendasari populernya *deep learning* saat ini adalah karena meningkatnya kemampuan perangkat keras untuk pemrosesan komputasi, bertambah besarnya ukuran data yang digunakan untuk melatih jaringan, dan semakin berkembangnya penelitian terkait *machine learning* dan pemrosesan data. Kemampuan *deep learning* untuk mengolah data yang besar dan kompleks seperti data citra digital membuat teknik ini sangat sering digunakan dalam bidang *computer vision* terutama pengenalan objek. Penggunaan *deep learning* pada teknik pengenalan objek modern dimulai pada tahun 2012 dimana *Convolutional Neural Network* (CNN) pertama kali digunakan oleh Alex Krizhevsky dengan model jaringan AlexNet yang berhasil menjuarai kompetisi pengenalan objek ILSVRC 2012. Sejak saat itu CNN menjadi salah satu metode penting yang digunakan dalam metode pengenalan objek yang terus berkembang.

*Convolutional Neural Network* (CNN) merupakan salah satu kelas dari *deep neural network*. Seperti *neural network* pada umumnya, CNN terdiri dari banyak neuron yang mempunyai bobot dan bias yang dapat dilatih. Setiap neuron pada CNN dapat menerima beberapa *input*, melakukan operasi komputasi, menghasilkan *output*, dan terhubung ke neuron-neuron lainnya. Hal yang menjadi ciri khusus dari CNN adalah *input* awal pada jaringannya secara eksplisit selalu diasumsikan sebagai citra digital, sehingga parameter pada arsitektur jaringannya dapat lebih mudah disesuaikan dan lebih efisien (F.-F. Li, Karpathy, & Johnson, 2016). Terdapat beberapa metode berbasis CNN dalam ranah pengenalan objek yang saat ini banyak digunakan, salah satunya yang sempat menjadi *state-of-the-art* adalah *Region-based Convolutional*

*Neural Network* (R-CNN). R-CNN mulai dikembangkan pada tahun 2014 sebagai metode pengenalan objek yang pertama kali mengkombinasikan algoritma *Region Proposal Network* (RPN) untuk menemukan region atau lokasi objek berada dalam suatu citra digital, dan metode CNN sebagai jaringan untuk melakukan klasifikasi atau pengenalan objek dari region yang dihasilkan oleh RPN (Girshick, Donahue, Darrell, & Malik, 2014). Seiring berjalannya waktu metode R-CNN ini terus dikembangkan untuk meningkatkan performa baik kecepatan maupun akurasi dalam pengenalan objek, salah satunya adalah dengan dikembangkannya *Faster R-CNN* yang mengoptimasi penggunaan jaringan konvolusional untuk sekaligus melakukan proses *object proposal* sehingga dapat mengurangi terjadinya *bottleneck* pada deteksi objek (Ren, He, Girshick, & Sun, 2015).

Metode R-CNN ini merupakan metode yang fleksibel dan dapat diimplementasikan dengan berbagai jenis model jaringan CNN, maupun dataset yang berbeda. Implementasi yang sudah dilakukan sebelumnya oleh Girshick (2014) maupun Ren (2015) menggunakan model jaringan ZF dan VGG16 yang merupakan salah satu model jaringan yang ada pada saat implementasi tersebut dilakukan. Akan tetapi saat ini telah muncul beberapa model jaringan baru seperti GoogLeNet, ResNet, Inception, MobileNet, dan model jaringan lainnya, sehingga metode R-CNN ini belum teruji pada model jaringan yang baru. Kemudian dalam segi data, implementasi dari metode R-CNN yang ada sebelumnya menggunakan dataset VOC 2007, dimana saat ini telah ada beberapa *dataset* yang lebih baru seperti VOC 2012 dan Microsoft COCO. Berdasarkan faktor-faktor tersebut, terdapat ruang untuk melakukan implementasi kembali metode R-CNN dengan pembaharuan pada beberapa hal seperti model jaringan maupun *dataset* yang lebih baru.

Model jaringan yang akan digunakan untuk implementasi algoritma R-CNN dalam penelitian ini adalah VGG, ResNet, dan MobileNet. Model jaringan VGG dipilih karena sudah terbukti mempunyai performa yang baik serta sebagai pembanding dengan implementasi sebelumnya, kemudian ResNet dan MobileNet dipilih untuk melihat bagaimana performa metode R-

CNN pada model jaringan terbaru dan model jaringan yang dioptimalkan untuk perangkat *mobile*. Dalam segi data, VOC 2012 akan digunakan sebagai *dataset* pada implementasi ini karena ukurannya yang tidak terlalu besar dan mendukung metode R-CNN. Dengan penggunaan model jaringan dan *dataset* terbaru tersebut diharapkan performa seperti akurasi dan kecepatan pada metode R-CNN yang dihasilkan dapat meningkat.

Dalam implementasinya, sistem berbasis *deep learning* seperti R-CNN akan melibatkan ukuran jaringan yang besar dan *layer-layer* berisi fungsi komputasi dengan jumlah yang banyak. Hal tersebut membuat implementasi sistem pengenalan objek berbasis R-CNN ini akan sulit jika dilakukan secara manual. Dengan dibutuhkannya kemampuan komputasi yang tinggi, saat ini banyak model implementasi *deep learning* yang menggunakan *General Purpose Graphics Processing Unit* (GPGPU) dengan memanfaatkan kemampuan komputasi paralelnya untuk proses komputasi yang lebih cepat. Dengan adanya model pemrograman seperti CUDA, sistem paralelisasi yang masif, dan *bandwith memory* yang tinggi menjadikan penggunaan GPU sebagai platform yang ideal dalam pemrosesan *deep learning* (Ciresan, et al., 2010). Akan tetapi penulisan kode program yang efisien untuk dieksekusi di GPU tetaplah menjadi hal yang tidak mudah karena arsitekturnya yang berbeda dengan program yang dieksekusi di CPU.

Dengan tidak mudahnya pembangunan program berbasis *deep learning* serta diperlukannya model pemrograman yang dapat memanfaatkan GPU secara maksimal maka diperlukanlah sebuah *library* perangkat lunak yang dapat melakukan beragam operasi komputasi, untuk kemudian didefinisikan dalam bentuk model kerangka kerja atau *framework* yang dapat mempermudah proses tersebut (Goodfellow et al., 2016). Saat penelitian ini dilakukan, telah ada beberapa *framework* untuk *deep learning* seperti Theano, TensorFlow, Torch, Microsoft Cognitive Toolkit, Caffe, dan lain sebagainya yang juga menyediakan fitur serupa. Semua kerangka kerja tersebut mempunyai tujuan umum untuk memproses komputasi yang intensif, jumlah data yang besar, dan melibatkan proses iterasi yang cukup banyak. Beberapa *framework deep*

*learning* yang ada saat ini juga telah menyematkan dukungan komputasi paralel berbasis GPU. Pada penelitian ini, *framework* Tensorflow akan digunakan untuk mendukung implementasi pengenalan objek dengan algoritma R-CNN yang akan dibuat. Tensorflow dipilih dikarenakan dukungan fitur-fitur yang cukup lengkap, dokumentasi yang baik, serta Tensorflow merupakan *library* untuk *data science* paling populer dengan jumlah pengembang dan dukungan komunitas yang besar (M. Li, 2017).

Fokus dari penelitian ini adalah untuk melakukan implementasi pengenalan objek pada citra digital dengan metode *Region-based Convolutional Neural Network* dengan melakukan pembaharuan pada model jaringan dan *dataset* yang lebih baru dan memanfaatkan *framework* pemrograman *deep learning* untuk efektifitas dan kemudahan penulisan kode. Penelitian ini dilakukan untuk melihat bagaimana performa metode R-CNN dengan model jaringan dan *dataset* terbaru dalam melakukan pengenalan objek pada citra digital sebagai salah satu topik populer dalam bidang *computer vision*. Beberapa indikator pengukuran yang sesuai akan digunakan untuk mengetahui performa dari penelitian ini. Hasil dari penelitian ini nantinya dapat digunakan sebagai referensi dalam pemilihan metode maupun alat yang dapat digunakan dalam *deep learning* terutama dalam tugas pengenalan objek dan *computer vision*.

## **1.2. Rumusan Masalah**

Berdasarkan latar belakang permasalahan, maka dapat diidentifikasi permasalahan yang akan dibahas adalah sebagai berikut.

1. Mengembangkan pengenalan objek pada citra digital dengan algoritma R-CNN dengan model jaringan dan *dataset* terbaru.
2. Mengetahui performa dari pengenalan objek pada citra digital dengan algoritma R-CNN dengan model jaringan dan *dataset* terbaru.

### **1.3. Tujuan Penelitian**

Penelitian ini mempunyai tujuan sebagai berikut.

1. Mampu melakukan pengembangan pengenalan objek pada citra digital dengan algoritma R-CNN menggunakan model jaringan dan *dataset* terbaru.
2. Performa dari pengenalan objek pada citra digital dengan algoritma R-CNN dengan model jaringan dan *dataset* terbaru yang dihasilkan dapat setara atau lebih baik dari penelitian sebelumnya.

### **1.4. Batasan Masalah**

Berdasarkan permasalahan diatas, maka penelitian ini akan dibatasi pada hal-hal sebagai berikut:

1. Indikator yang akan digunakan dalam pengukuran performa dari pengenalan objek pada citra digital dengan algoritma R-CNN ini adalah lama proses pelatihan, kecepatan deteksi, ukuran bobot hasil pelatihan, dan akurasi dari hasil deteksi objek.

### **1.5. Metodologi Penelitian**

Adapun metode yang digunakan adalah sebagai berikut.

#### **1. Studi Pustaka**

Pada bagian langkah studi pustaka ini dilakukan untuk mencari sumber pustaka yang berkaitan dengan aplikasi yang akan dibuat. Pada penelitian ini dilakukan pencarian sumber pustaka untuk pengenalan objek berbasis R-CNN. Pada studi pustaka ini dapat membantu untuk membuat dan memberikan langkah-langkah pembantu selanjutnya serta memberikan teori yang ada. Dengan metode penelitian ini dapat didapatkan data yang merupakan hasil dari sumber pustaka yang sudah tertulis sebelumnya.

## 2. Analisis Algoritma

Pada tahapan analisis algoritma dilakukan untuk dapat memahami langkah-langkah proses cara kerja sistem pengenalan objek yang akan dibuat dengan metode yang telah ditentukan. Analisis dilakukan dengan melakukan studi literatur dari teori-teori yang didapatkan dari hasil studi pustaka. Tahapan ini dapat menghasilkan alur dan gambaran umum tentang fungsi-fungsi program yang akan dibuat, langkah proses kerjanya, dan arsitektur sistem yang akan dibuat.

## 3. Perancangan Program

Tahap perancangan program ini digunakan untuk merancang program berdasarkan fungsi-fungsi program, langkah proses kerja, dan arsitektur sistem yang telah dihasilkan pada tahap analisis algoritma. Tahap perancangan ini menghasilkan *flowchart* program dan *pseudocode* dari program yang akan dibuat.

## 4. Implementasi

Tahap implementasi dilakukan untuk mengembangkan rancangan yang telah dilakukan oleh tahapan sebelumnya. Hasilnya adalah sebuah sistem atau program yang sesuai dengan tujuan penelitian.

## 5. Pengujian

Tahap pengujian adalah tahap untuk menguji sistem yang telah dibuat pada tahap implementasi. Pengujian dilakukan untuk mengevaluasi indikator-indikator sesuai dengan tujuan penelitian dan dapat menghasilkan sebuah sistem yang sesuai dengan yang telah diharapkan. Pengujian dilakukan berdasarkan parameter yang telah ditentukan, dan hasil akhir pengujian akan dijadikan evaluasi untuk penelitian selanjutnya.

## 6. Penyusunan laporan

Tahap ini adalah tahap akhir pada penelitian ini dengan menuliskan hal-hal terkait seperti pendahuluan, perancangan dan hasil dari penelitian.

### 1.6. Alat dan Bahan

Pada penelitian ini akan digunakan bahasa pemrograman utama yaitu Python versi 3.6 yang berjalan pada sistem operasi berbasis Linux yaitu Ubuntu 16.04 (Xenial Xerus). Perangkat keras untuk melakukan penelitian ini yaitu.

1. *Personal computer* (PC) untuk proses pembangunan dan percobaan sistem dengan *processor* Intel i7-3770K, *memory* 16 GB, dan GPU Nvidia GTX 1080 8 GB .
2. *Cloud computing* untuk proses pelatihan dan *testing* akhir dengan *processor* Intel Xeon (8 vCPU) 2.50 GHz, *memory* 40 GB, dan GPU Nvidia Tesla P100 16 GB dengan 3584 CUDA *cores* dan 32 GB/s PCIe *bandwith*.

Dalam penelitian ini juga menggunakan beberapa library untuk bahasa pemrograman Python. Library yang digunakan dalam penelitian ini yaitu.

1. Tensorflow 1.8.0 untuk *framework deeplearning*.
2. Numpy 1.14 untuk pemrosesan *array* dan matriks.
3. OpenCV 3.4.1 untuk pemrosesan citra digital.
4. Matplotlib 2.0.2 untuk visualisasi hasil deteksi.
5. SciPy 1.1.0 untuk membantu perhitungan beberapa fungsi komputasi.
6. Cython untuk memproses beberapa modul eksternal kedalam bahasa Python.
7. Pyyaml 3.12 untuk membaca file konfigurasi dengan format yaml.

*Dataset* yang digunakan adalah PASCAL VOC 2007 sebagai *dataset* pembanding serta VOC 2012 sebagai *dataset* utama.

## **1.7. Sistematika Penulisan Laporan**

Laporan tugas akhir ini disusun dengan sistematika sebagai berikut.

### **BAB I : Pendahuluan**

Pada bab ini berisi tentang latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, alat dan bahan, metodologi penelitian, dan sistematika penulisan laporan yang terkait dengan proses penelitian ini.

### **BAB II : Tinjauan Pustaka**

Bab ini berisi tentang penjelasan singkat dan ringkasan mengenai penelitian terdahulu yang berhubungan atau memiliki kesamaan dengan permasalahan yang akan dibahas oleh penulis di dalam tugas akhir ini.

### **BAB III : Landasan Teori**

Pada bab ini berisi dasar-dasar teori yang melandasi serta mendukung dalam implementasi aplikasi dan proses yang berhubungan dengan pembangunan aplikasi.

### **BAB IV : Analisis dan Implementasi Algoritma**

Bab ini berisi penjelasan uraian analisis algoritma dan perancangan sistem yang akan dibuat.

### **BAB V : Hasil dan Pembahasan**

Bab ini berisi penjelasan mengenai hasil dari penelitian ini yang berupa data-data hasil pengujian dan pembahasannya.

### **BAB VI : Kesimpulan dan Saran**

Pada bab ini berisi kesimpulan mengenai penelitian yang telah dibuat beserta saran-saran yang berguna bagi pengembangan lebih lanjut.