

BAB II

TINJAUAN PUSTAKA

Banyak pengembangan sistem yang mendalami *Computer Vision* seperti *Face Recognition*, *Image Recognition* maupun pengenalan pola tertentu. Sistem seperti ini terbukti secara kongkrit menjadi sebuah fungsionalitas yang dapat mempermudah pekerjaan di banyak bidang. Implementasi *Deep Learning* pada permasalahan tersebut sangatlah tepat dan efektif. Penggunaan teknik *Deep Learning* dengan metode *Convolutional Neural Network* pertama kali berhasil diaplikasikan oleh Yann LeCun pada tahun 1998 (LeCun, Bottou, Bengio, & Haffner, 1998). Pada penelitian ini Yann LeCun mengemukakan metode CNN untuk mengenal tulisan tangan untuk keperluan pembacaan dokumen. Hasil yang didapat dari penelitian tersebut menunjukkan akurasi yang cukup tinggi hingga mencapai test error hanya sebesar 1,7%.

Penerapan metode *Convolutional Neural Network* dapat dikembangkan dari sisi arsitektur dan banyaknya lapisan yang digunakan pada jaringan. Penggunaan arsitektur yang benar akan sangat baik untuk klasifikasi citra dalam berbagai macam kategori. Contohnya seperti pada dataset ImageNet yang memiliki sebanyak 1000 kategori. Pada tahun 2012 teknik *Deep Learning* dengan metode CNN dipopulerkan dengan asritektur AlexNet yang diuji dengan dataset ImageNet (Krizhevsky, Sutskever, & Hinton, 2012). Arsitektur yang dibuat oleh Alex Krizhevsky menunjukan hasil yang sangat signifikan pada testing set dengan test error sebesar 17%. Hasil tersebut dinilai sudah sangat luar biasa karena citra pada dataset yang digunakan sangatlah kompleks dan banyak.

Kedalaman jaringan merupakan komponen penting untuk kinerja yang baik untuk pengenalan citra. Lebih dalam sebuah arsitektur jaringan maka akan lebih banyak lapisan yang digunakan. Jaringan yang dibuat pada penelitian oleh Karen Simonyan dan Andrew Zisserman menggunakan lapisan konvolusi sebanyak 16 hingga 19 lapisan dan jaringan tersebut menunjukkan performa yang sangat baik dengan akurasi yang lebih besar dibanding dengan arsitektur yang sudah dikemukakan sebelumnya (Simonyan & Zisserman, 2015).

Model-model jaringan yang sudah dibuat pada riset terbaru belakangan ini memang sudah terbukti sangat ampuh dalam klasifikasi citra. Model-model tersebut sudah menjadi standar dan patokan dalam membuat model CNN yang baik serta dapat digunakan pada sistem maupun aplikasi yang menggunakan fitur *Image Recognition*. Tentunya model yang sudah ada tersebut masih dapat dikembangkan lagi untuk hasil yang lebih baik, seperti penelitian yang dilakukan oleh Matthew D. Zeiler dan Rob Fergus pada tahun 2013 (Fergus & Zeiler, 2014). Pada penelitian tersebut model yang dibuat merupakan perbaikan dari model AlexNet dengan mengubah arsitektur *hyperparameter*, terutama dengan memperluas ukuran lapisan konvolusi di tengah susunan model dan membuat ukuran filter pada lapisan pertama lebih kecil. Penelitian mereka juga menunjukkan model yang baik untuk banyak dataset lain ketika klasifikasi *Softmax* dilatih ulang dengan data yang baru. Hal ini sangat meyakinkan mengingat hasil yang diberikan mengalahkan hasil riset sebelumnya untuk dataset Caltech-101 dan Caltech-256.

Penggunaan parameter yang baik antar lapisan dalam model CNN dapat mengurangi bobot komputasi saat pelatihan. Lebih sedikit parameter pada model yang digunakan maka akan lebih cepat proses pelatihan serta. Walaupun parameter pada model sangat sedikit, bukan berarti hasil yang didapat tidak lebih baik dibanding parameter yang banyak, bahkan bisa lebih baik. Pada tahun 2014 Christian Szegedy membuat terobosan baru dengan model GoogLeNet yang terbukti memperkecil penggunaan parameter (Szegedy, 2014). Kontribusi utamanya adalah pengembangan model pendahulunya yang secara dramatis mengurangi jumlah parameter di jaringan, yakni hanya sebesar 4 juta parameter, dibandingkan dengan AlexNet dengan 60 juta parameter. Selain itu, penelitian yang dilakukan oleh Christian Szegedy ini menggunakan lapisan *Average Pooling* daripada menggunakan lapisan *Fully Connected* di bagian akhir susunan model, sehingga menghilangkan sejumlah besar parameter yang tampaknya tidak penting. Ciri utama arsitektur model ini adalah pemanfaatan sumber daya komputasi yang baik di dalam jaringan. Hal ini dicapai dengan membuat desain yang memungkinkan untuk meningkatkan kedalaman jaringan

sementara harus menjaga penggunaan komputasi. Untuk mengoptimalkan kualitas dari arsitektur model, GoogLeNet menggunakan susunan 22 lapisan dan dinilai dari konteks klasifikasi serta pendeteksian yang dihasilkan maka dapat dikatakan bahwa model jaringan seperti ini sangatlah efektif.

Berbicara dari sisi fungsionalitas serta kemampuan suatu model jaringan, teknik *Deep Learning* sendiri terdapat model *state of the art* yang sudah diakui oleh banyak peneliti. Saat ini model *state of the art* terbaru adalah *Residual Network*. Model ini diajukan oleh Kaiming He pada tahun 2015 (He, Zhang, Ren, & Sun, 2015). Model ini memiliki fitur koneksi khusus dan penggunaan *Batch Normalization*. Arsitektur ini juga tidak memakai lapisan *Fully Connected* pada akhir susunan jaringan. ResNets saat ini merupakan model Jaringan CNN yang canggih dan merupakan pilihan standar untuk penelitian dengan dataset yang sangat besar. Model ini juga menyediakan fungsi komprehensif yang menunjukkan bahwa sisa dari lapisan di jaringan lebih mudah dioptimalkan, dan bisa mendapatkan akurasi yang lebih tinggi. Pada dataset ImageNet model ini memiliki jaringan sisa dengan kedalaman hingga 152 lapisan, 8 kali lebih banyak dari pada jaringan VGG namun masih memiliki kompleksitas yang lebih rendah. Hasil dari model ini mencapai tingkat kesalahan sebesar 3,57% pada dataset ImageNet.

Pada penelitian ini, model CNN yang digunakan tidak memerlukan banyak lapisan guna menghemat waktu pembelajaran. Kesederhanaan arsitektur CNN yang digunakan sudah cukup untuk kasus klasifikasi citra dengan dataset yang dapat terbilang kecil, terlebih kategori untuk klasifikasi hanya sedikit. Seditkitnya jumlah data yang akan dipelajari dapat menghasilkan akurasi yang kurang memuaskan, untuk itu ada beberapa hal yang harus dilakukan demi mendapat hasil yang cukup baik. Penggunaan regularisasi *Dropout* yang agresif akan sangat membantu model untuk mengurangi *Overfitting* sehingga data yang baru dapat diprediksi dengan benar. Satu hal lagi yang perlu digunakan ialah proses *Data Augmentation*. Hal ini dapat memperbanyak jumlah data yang dipelajari dengan mengubah beberapa sudut dari citra seperti rotasi, perubahan warna maupun pergeseran citra. Hasil pengujian pada penelitian ini juga

dibandingkan berdasarkan jumlah *epochs* yang digunakan dan setelah itu jumlah *epochs* yang optimal akan diambil dilihat dari sisi kecepatan dan ketepatan model dalam mempelajari data.

Tabel 2.1. Tabel Perbandingan Pustaka

Penulis	Dataset	Jumlah Lapisan	Tahun
Yann LeCun	MNIST	5	1998
Alex Krizhevsky	ImageNet	10	2012
Matthew D. Zeiler & Rob Fergus	Caltech	15	2013
Christian Szegedy	ImageNet	22	2014
Karen Simonyan & Andrew Zisserman	ImageNet	16 & 19	2015
Kaiming He	ImageNet	152	2015
Kefin Pudi Danukusumo	Candi	9	2017