

Technical Report UTS - UNHAS

Perbandingan GoogleNet, ResNet, dan DenseNet

Siti Nisrina Nabila P.S.Y, Septi Intan Amalia, Annisa Putri Aulia
Information Systems
Universitas Hasanuddin

CONTENTS

I	Introduction	1
II	CNN Architectures	1
II-A	GoogleNet	1
II-B	ResNet	1
II-C	DenseNet	2
III	RESEARCH METHODOLOGY	2
III-A	Dataset	2
III-B	Augmentasi data	2
IV	Result and Conclusion	2
	References	2

Technical Report UTS - UNHAS

Perbandingan GoogleNet, ResNet, dan DenseNet

Abstract—Ada banyak arsitektur yang berbeda telah diusulkan selama beberapa tahun terakhir. Beberapa yang paling berpengaruh, dan masih relevan hingga saat ini, adalah sebagai berikut: arsitektur GoogleNet/Inception (pemenang ILSVRC 2014), ResNet (pemenang ILSVRC 2015), dan DenseNet (pemenang makalah terbaik CVPR 2017). Semuanya adalah model terancang saat diusulkan, dan ide inti dari jaringan ini adalah fondasi untuk sebagian besar arsitektur terancang saat ini. Oleh karena itu, penting untuk memahami arsitektur ini secara detail dan mempelajari cara mengimplementasikannya.

I. INTRODUCTION

Convolutional Neural Network atau CNN merupakan salah satu jenis algoritme *Deep Learning neural network* yang biasa digunakan pada data image. CNN bisa digunakan untuk mendeteksi dan mengenali object pada sebuah image. Arsitektur CNN terbilang mirip dengan pola koneksi neuron atau sel saraf dalam otak manusia. CNN terinspirasi dari Visual Cortex, yaitu bagian pada otak yang bertugas untuk memroses informasi dalam bentuk visual.

Secara garis besarnya, CNN memanfaatkan proses dengan mengerjakan sebuah kernel konvolusi (filter) berukuran tertentu ke sebuah gambar, komputer mendapatkan informasi representatif baru dari hasil perkalian bagian gambar tersebut dengan filter yang digunakan. Kelebihan CNN adalah algoritma yang dapat mengenali informasi tersembunyi dari suatu objek baik itu gambar, suara, teks dan sebagainya walaupun objek tersebut diposisikan dimana saja dalam input. Terdapat beberapa model arsitektur dari CNN salah satunya adalah GoogleNet, ResNet, dan DenseNet. Pada *technical report* ini akan dilakukan perbandingan performa dari ketiga model tersebut dengan menggunakan dataset CIFAR-10.

II. CNN ARCHITECTURES

A. GoogleNet

Arsitektur GoogLeNet disajikan dalam ImageNet Large-Scale Visual Recognition Challenge 2014(ILSVRC14) untuk menyelesaikan tugas visi komputer seperti klasifikasi gambar dan deteksi objek. GoogLeNet digunakan untuk tugas visi komputer lainnya seperti face detection and recognition, adversarial training, dll. Blok Inception menerapkan empat blok konvolusi secara terpisah pada peta fitur yang sama: konvolusi 1x1, 3x3, dan 5x5, dan operasi kumpulan maks. Ini memungkinkan jaringan untuk melihat data yang sama dengan bidang reseptif yang berbeda. Tentu saja, mempelajari konvolusi 5x5 saja secara teoritis akan lebih kuat. Namun, ini tidak hanya lebih banyak komputasi dan memori, tetapi juga cenderung lebih mudah digunakan.

Blok awal keseluruhan terlihat seperti pada figure 1 (fig.credit - Szegedy et al.): Konvolusi 1x1 tambahan sebelum konvolusi 3x3 dan 5x5 digunakan untuk reduksi dimensi. Karena konvolusi 5x5 25 kali lebih mahal daripada konvolusi 1x1, sehingga dapat menghemat banyak komputasi dan parameter dengan mengurangi dimensi sebelum konvolusi besar. GoogleNet asli dirancang untuk ukuran gambar ImageNet (224x224 piksel) dan memiliki hampir 7 juta parameter. Jumlah saluran untuk pengurangan dimensi dan keluaran per filter (1x1, 3x3, 5x5, dan pooling maks) perlu ditentukan secara manual dan dapat diubah jika tertarik. Intuisi umumnya adalah memiliki filter terbanyak untuk konvolusi 3x3, karena filter tersebut cukup kuat untuk mempertimbangkan konteks sementara membutuhkan hampir sepertiga dari parameter konvolusi 5x5.

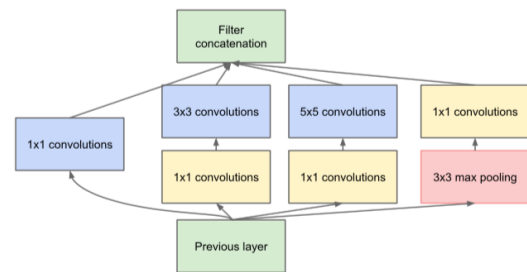


Fig. 1. Modul awal dengan pengurangan dimensi

B. ResNet

ResNet telah menjadi dasar untuk jaringan saraf dengan lebih dari 1.000 lapisan. Terlepas dari kesederhanaannya, ide koneksi residual sangat efektif karena mendukung propagasi gradien yang stabil melalui jaringan. Blok ResNet asli menerapkan fungsi aktivasi non-linear, biasanya ReLU, setelah koneksi lewati.

Sebaliknya, blok ResNet pra-aktivasi menerapkan non-linearitas di awal F. Keduanya memiliki kelebihan dan kekurangan masing-masing. Namun, untuk jaringan yang sangat dalam, pra-aktivasi ResNet telah menunjukkan kinerja yang lebih baik karena aliran gradien dijamin memiliki matriks identitas seperti yang dihitung di atas, dan tidak dirugikan oleh aktivasi non-linier yang diterapkan.

Implementasi asli menggunakan pemetaan identitas dengan langkah 2 dan menambahkan dimensi fitur tambahan dengan 0. Namun, implementasi yang lebih umum adalah menggunakan konvolusi 1x1 dengan langkah 2 karena memungkinkan kita

untuk mengubah dimensi fitur sekaligus efisien dalam parameter dan biaya komputasi. Arsitektur ResNet keseluruhan terdiri dari penumpukan beberapa blok ResNet, di mana beberapa di antaranya melakukan downsampling input. Ketika berbicara tentang blok ResNet di seluruh jaringan, maka pengelompokannya berdasarkan bentuk keluaran yang sama. Oleh karena itu, jika ResNet memiliki [3,3,3] blok, itu berarti bahwa memiliki 3 kali kelompok 3 blok ResNet, di mana subsampling terjadi di blok keempat dan ketujuh. ResNet dengan [3,3,3] blok di CIFAR10 divisualisasikan pada fig.2

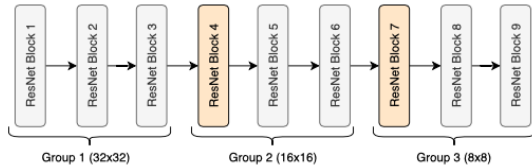


Fig. 2. ResNet dengan blok [3,3,3] pada CIFAR10

Tiga kelompok beroperasi pada resolusi 32x32, 16x16 dan 8x8 masing-masing. Blok berwarna oranye menunjukkan blok ResNet dengan downsampling. Notasi yang sama digunakan oleh banyak implementasi lain seperti di perpustakaan torchvision dari PyTorch.

C. DenseNet

DenseNet adalah arsitektur lain untuk mengaktifkan jaringan saraf yang sangat dalam dan mengambil perspektif yang sedikit berbeda pada koneksi residual. Alih-alih memodelkan perbedaan antar lapisan, DenseNet menganggap koneksi residual sebagai cara yang memungkinkan untuk menggunakan kembali fitur di seluruh lapisan, menghilangkan kebutuhan apa pun untuk mempelajari peta fitur yang berlebihan. Jika kita masuk lebih dalam ke jaringan, model mempelajari fitur abstrak untuk mengenali pola. Namun, beberapa pola kompleks terdiri dari kombinasi fitur abstrak (misalnya tangan, wajah, dan lain-lain), dan fitur tingkat rendah (misalnya tepi, warna dasar, dan lain-lain)

Untuk menemukan fitur tingkat rendah ini di lapisan dalam, CNN standar harus belajar menyalin peta fitur tersebut, yang membuang banyak kompleksitas parameter. DenseNet menyediakan cara yang efisien untuk menggunakan kembali fitur dengan membuat setiap konvolusi bergantung pada semua fitur input sebelumnya, tetapi hanya menambahkan sedikit filter ke dalamnya.

III. RESEARCH METHODOLOGY

A. Dataset

CIFAR-10 Merupakan dataset berisi kumpulan gambar yang dibagi menjadi 10 kelas yang masing-masing kelas terdiri dari 60000 gambar. 10 kelas dari dataset tersebut terdiri dari *airplane, automobile, cat, deer, dog, frog, horse, ship, truck*.

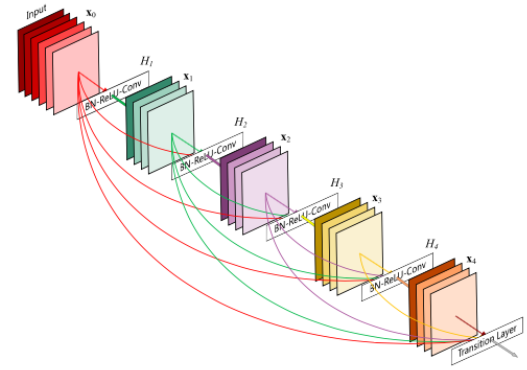


Fig. 3. Ilustrasi DenseNet

B. Augmentasi data

Augmentasi Data merupakan sebuah teknik memanipulasi sebuah data tanpa kehilangan inti atau esensi dari data tersebut. Ini mengurangi risiko over fitting dan dapat membantu CNN untuk menggeneralisasi lebih baik. Pada tahap augmentasi data, akan dilakukan 2 proses yakni proses data augmentasi dilakukan dengan melakukan *flip* pada gambar secara horizontal, dan melakukan *resize* pada gambar.

IV. RESULT AND CONCLUSION

Pertama-tama, kami melihat bahwa semua model berkinerja cukup baik. Model sederhana seperti yang telah Anda implementasikan dalam praktik mencapai kinerja yang jauh lebih rendah, selain jumlah parameter yang lebih rendah juga dikaitkan dengan pilihan desain arsitektur. GoogleNet adalah model untuk mendapatkan kinerja terendah pada set validasi dan pengujian, meskipun sangat dekat dengan DenseNet. Pencarian hyperparameter yang tepat untuk semua ukuran saluran di GoogleNet kemungkinan akan meningkatkan akurasi model ke tingkat yang sama, tetapi ini juga mahal mengingat sejumlah besar hyperparameter.

Model	Val Accuracy	Test Accuracy	Num Parameters
GoogleNet	90.40%	89.70%	260,650
ResNet	91.84%	91.06%	272,378
ResNetPreAct	91.80%	91.07%	272,250
DenseNet	90.72%	90.23%	239,146

Fig. 4. Perbandingan

ResNet mengungguli DenseNet dan GoogleNet lebih dari 1 persen pada set validasi, sementara ada perbedaan kecil antara kedua versi, asli dan pra-aktivasi. Kita dapat menyimpulkan bahwa untuk jaringan dangkal, tempat fungsi aktivasi tampaknya tidak terlalu penting.

REFERENCES

- [1] Wahyu setiawan "PERBANDINGAN ARSITEKTUR CONVOLUTIONAL-NEURAL NETWORK UNTUK KLASIFIKASI FUNDUS," in *Deep*

Residual Learning for Image Recognition, <https://arxiv.org/abs/1512.03385>. [Accessed: 01 May. 2022].

- [2] He.Kaiming,Zhang Xiangyu “Computer Vision and Pattern Recognition,” in *Deep Residual Learning for Image Recognition*, <https://arxiv.org/abs/1512.03385>. [Accessed: 01 May. 2022].
- [3] He.Kaiming,Zhang Xiangyu “Computer Vision and Pattern Recognition,” in *Deep Residual Learning for Image Recognition*, <https://arxiv.org/abs/1512.03385>. [Accessed: 01 May. 2022].