

# Technical Report UTS UNHAS

## CNN Inception,ResNet,Dan DesNet

Aso Ahmad Amin Rais,Muhammad Yusuf Syam,Bayu Ajid  
Departemen Ilmu Komputer  
Universitas Hasanuddin

### CONTENTS

<b>I</b>	<b>Introduction</b>	<b>1</b>
<b>II</b>	<b>CNN Architectures</b>	<b>1</b>
II-A	GoogleNet . . . . .	1
II-B	ResNet . . . . .	1
II-C	DenseNet . . . . .	2
<b>III</b>	<b>Research Methodology</b>	<b>2</b>
III-A	DataSet . . . . .	2
III-B	Augmentasi Data . . . . .	2
<b>IV</b>	<b>Result And Conclusion</b>	<b>2</b>
	<b>References</b>	<b>3</b>

# Technical Report UTS UNHAS

## CNN Inception, ResNet, Dan DesNet

**Abstract**—Dalam technical report ini, kita akan mengimplementasikan, membahas, dan membandingkan varian arsitektur CNN modern. Ada banyak arsitektur yang berbeda telah diusulkan selama beberapa tahun terakhir. Beberapa yang paling berdampak, dan masih relevan hingga saat ini seperti GoogleNet / Arsitektur Inception, Pemenang ILSVRC 2014, ResNet, Pemenang ILSVRC 2015 dan DenseNet, Penghargaan best paper CVPR 2017. Ketiga arsitektur tersebut menjadi state-of-the-art saat diusulkan dan menjadi fondasi dari sebagian besar arsitektur-arsitektur CNN saat ini. Oleh karena itu, penting untuk memahami arsitektur ini secara detail dan mempelajari cara mengimplementasikannya.

### I. INTRODUCTION

Convolutional Neural Network (CNN) adalah salah satu jenis neural network yang biasa digunakan pada data image. CNN bisa digunakan untuk mendeteksi dan mengenali object pada sebuah image. CNN adalah sebuah teknik yang terinspirasi dari cara mamalia — manusia, menghasilkan persepsi visual. Secara garis besar Convolutional Neural Network (CNN) tidak jauh beda dengan neural network biasanya. CNN terdiri dari neuron yang memiliki weight, bias dan activation function. Convolutional layer juga terdiri dari neuron yang tersusun sedemikian rupa sehingga membentuk sebuah filter dengan panjang dan tinggi (pixels).

CNN memanfaatkan proses konvolusi dengan menggerakkan 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

GoogleNet, yang diusulkan pada tahun 2014, memenangkan Tantangan ImageNet karena penggunaan modul Inception.

Blok Inception menerapkan empat blok konvolusi secara terpisah pada satu blok yang sama: konvolusi 1x1, 3x3, dan 5x5, dan operasi max poolin. Blok Inception keseluruhan terlihat seperti di bawah ini (kredit angka - Szegedy et al.):

Konvolusi 1x1 tambahan sebelum konvolusi 3x3 dan 5x5 digunakan untuk reduksi dimensi. Ini sangat penting karena

feature map dari semua cabang digabungkan nantinya, dan kita tidak ingin ada ledakan ukuran fitur. Karena komputasi konvolusi 5x5 lebih mahal 25 kali daripada konvolusi 1x1, kita dapat menghemat banyak komputasi dan parameter dengan mengurangi dimensi sebelum konvolusi besar selanjutnya. Arsitektur GoogleNet terdiri dari beberapa blok Inception dengan max pooling sesekali untuk mengurangi tinggi dan lebar feature map. GoogleNet asli dirancang untuk ukuran gambar ImageNet (224x224 piksel) dan memiliki hampir 7 juta parameter. Saat kita melakukan pelatihan dengan data CIFAR10 dengan ukuran gambar 32x32, kita tidak memerlukan arsitektur yang berat. Jumlah channel untuk reduksi dimensi dan keluaran per filter (1x1, 3x3, 5x5, dan max pooling) perlu ditentukan secara manual dan dapat diubah. Umumnya filter terbanyak untuk konvolusi adalah 3x3, karena filter tersebut cukup baik mempertimbangkan parameter yang dibutuhkan hanya sekitar sepertiga dari parameter konvolusi 5x5.

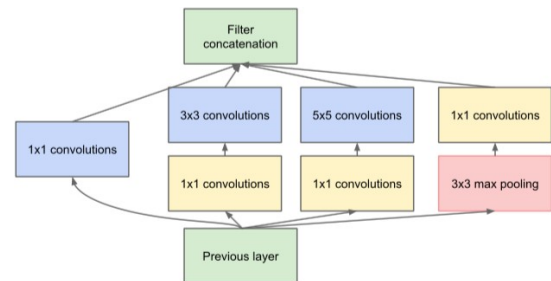


Fig. 1. Modul awal pengurangan dimensi

#### B. ResNet

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

Sebaliknya, blok ResNet pra-aktivasi menerapkan non-linearitas di awal  $F$ . Keduanya memiliki kelebihan dan kekurangan masing-masing. Namun, untuk neural network 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 padanya. Implementasi asli menggunakan pemetaan identitas dengan stride 2 dan menambahkan dimensi fitur tambahan dengan 0. Namun, implementasi yang lebih umum adalah menggunakan

konvolusi 1x1 dengan stride 2 karena memungkinkan kita untuk mengubah dimensi fitur dan tetap 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

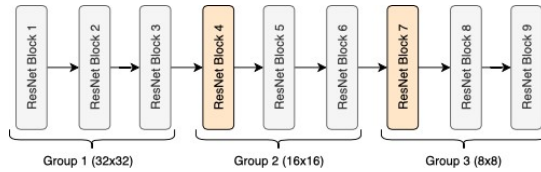


Fig. 2. blok [3,3,3] pada CIFAR10

### C. DenseNet

DenseNet adalah arsitektur lain untuk jaringan saraf yang sangat dalam dan mengambil perspektif yang sedikit berbeda tentang koneksi residual. Alih-alih memodelkan perbedaan antar layer, DenseNet menganggap koneksi residual sebagai cara yang mungkin untuk menggunakan kembali fitur di seluruh layer, menghilangkan kebutuhan apa pun untuk mempelajari feature map 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, dll.), dan fitur low-level (misalnya tepi, warna dasar, dll.). Untuk menemukan fitur low-level ini di layer yang dalam, CNN standar harus belajar menyalin feature map 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.

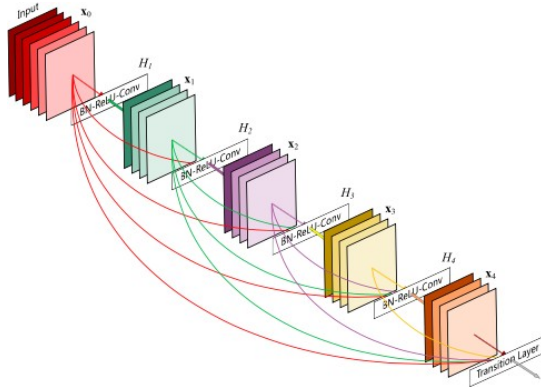


Fig. 3. contoh DensNet

Layer terakhir, yang disebut Transition layer, bertanggung jawab untuk mengurangi dimensi feature map dalam tinggi,

lebar, dan ukuran channel. Meskipun secara teknis merusak identitas backpropagation, hanya ada beberapa di jaringan sehingga tidak banyak mempengaruhi aliran gradien.

This may be a modified version of your proposal depending on previously carried out research or any feedback received.

## III. RESEARCH METHODOLOGY

### A. DataSet

CIFAR-10 adalah kumpulan data image yang berisi 6000 gambar tiap kelas dan terbagi menjadi 10 kelas berbeda, airplane, automobile, cat, deer, dog, frog, horse, ship, truck.

### B. Augmentasi Data

Augmentasi data adalah strategi yang memungkinkan praktisi untuk secara signifikan meningkatkan keragaman data yang tersedia untuk model pelatihan, tanpa benar-benar mengumpulkan data baru. Teknik augmentasi data seperti cropping, padding, dan flipping horizontal umumnya digunakan untuk melatih jaringan neural besar.

## IV. RESULT AND CONCLUSION

Pertama-tama, kita melihat bahwa semua model berkinerja cukup baik.

GoogleNet merupakan model dengan performa terendah pada set validasi dan pengujian, meskipun sangat dekat dengan DenseNet. Pencarian hyperparameter yang tepat untuk semua ukuran channel di GoogleNet kemungkinan akan meningkatkan akurasi model ke tingkat yang sama, tetapi ini juga mempunyai komputasi yang mahal mengingat sejumlah besar hyperparameter.

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,

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 Hasil

Menariknya, DenseNet mengungguli ResNet asli pada setup tetapi berada di belakang ResNet Pra-Aktivasi. Model terbaik, Jaringan Dual Path sebenarnya merupakan kombinasi dari ResNet dan DenseNet yang menunjukkan bahwa keduanya menawarkan keuntungan yang berbeda.

## REFERENCES

- [1] QOLBIYATUL LINA “Apa itu Convolutional Neural Network?,” Available: <https://medium.com/@16611110/apa-itu-convolutional-neural-network-836f70b193a4>. [Accessed: Apr. 30, 2022].
- [2] Joseph Sanjaya dan Mewati Ayub “Augmentasi Data Pengenalan Citra Mobil Menggunakan Pendekatan Random Crop, Rotate, dan Mixup,” Available: <https://journal.maranatha.edu/index.php/jutisi/article/view/2688>. [Accessed: Apr. 30, 2022].
- [3] Siddhart Das dan Mewati Ayub “CNN Architectures: LeNet, AlexNet, VGG, GoogLeNet, ResNet and more” Available: <https://medium.com/analytics-vidhya>. [Accessed: Apr. 30, 2022].