Review: FractalNet (Image Classification)

An Ultra Deep Network Without Using Residuals

SH Tsang [Follow]

Dec 15, 2018 · 4 min read

In 2015, after the invention of ResNet, with numerous champion won, there are plenty of researchers working on how to improve the ResNet, such as Pre-Activation ResNet, RiR, RoR, Stochastic Depth, and WRN. In this story, conversely, a non-residual-network approach, **FractalNet** is shortly reviewed. When VGGNet is starting to degrade when it goes from 16 layers (VGG-16) to 19 layers (VGG-19), **FractalNet can go up to 40 layers or even 80 layers**. And it is published in **2017 ICLR** with more than **100 citations**. (SH Tsang @ Medium)
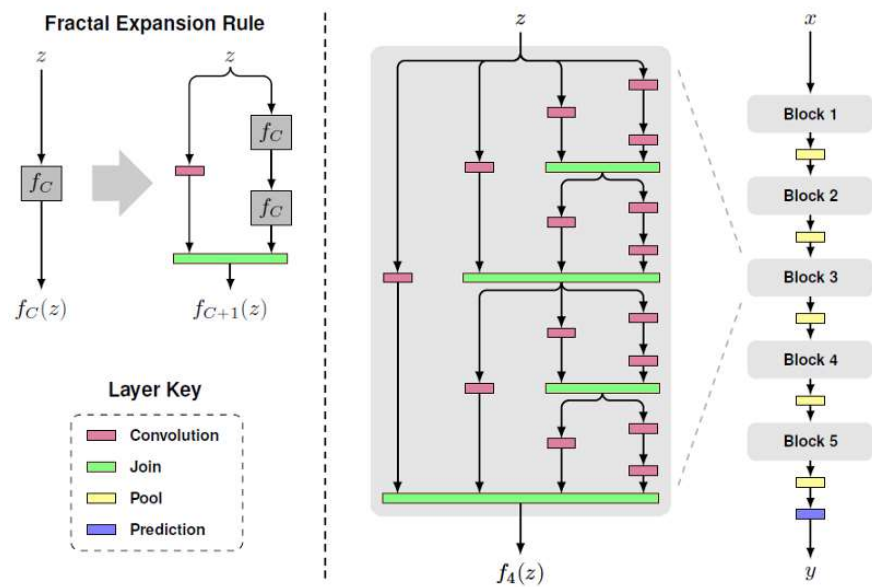
· · ·

What Are Covered

1. **FractalNet Architecture**

2. **Drop-Path as Regularization**

3. **Ablation Study**

4. **Results**

· · ·

1. FractalNet Architecture

Fractal Architecture: A Simple Fractal Expansion (Left), Recursively Stacking of Fractal Expansion as One Block (Middle), 5 Blocks Cascaded as FractalNet (Right)

For the base case, f1(z) is the convolutional layer:

$$f_1(z) = \text{conv}(z)$$

After that, the recursive fractals are:

$$f_{C+1}(z) = [(f_C \circ f_C)(z)] \oplus [\text{conv}(z)]$$
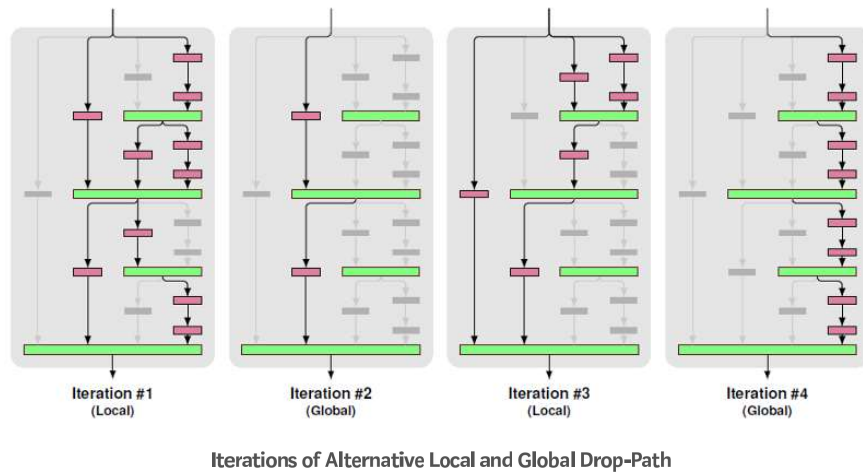
where C is the number of columns as in the middle of the figure. The number of convolutional layers at the deepest path within a block will have $2^{(C-1)}$. In this case, **C=4**, thereby, the number of convolutional layers is $2^3=8$ layers.

For the **join layer** (green), **element-wise mean** is computed. It is not concatenation or addition.

With 5 blocks (**B=5**) cascaded as FractalNet at the right of the figure, then the number of convolutional layers at the deepest path within the whole network is $B×2^{(C-1)}$, i.e. $5×2^3=$**40 layers**.

In between 2 blocks, 2×2 max pooling is done to reduce the size of feature maps. Batch Norm and ReLU are used after each convolution.

. . .

2. Drop-Path as Regularization

Iteration #1 (Local)   Iteration #2 (Global)   Iteration #3 (Local)   Iteration #4 (Global)

Iterations of Alternative Local and Global Drop-Path

There are local and global drop-paths:

- **Local**: A join **drops each input with fixed probability**, but it is sure that there is at least one survives.

- **Global**: **A single path is selected** for the entire network.

By dropping path randomly, it is just like noise is injected to the input, and finally **regularization effect** is achieved to **reduce overfitting.**

. . .

3. Ablation Study

3.1. Depth

Cols.	Depth	Params.	Error (%)
1	5	0.3M	37.32
2	10	0.8M	30.71
3	20	2.1M	27.69
4	40	4.8M	27.38
5	80	10.2M	26.46
6	160	21.1M	27.38

Number of Depths on CIFAR-100++

The network with depth of 80 yields the best performance. The network with depth of 160 has overfitting problem.

3.2. Effectiveness of Fractal Structure

Model	Depth	Train Loss	Error (%)
Plain	5	0.786	36.62
Plain	10	0.159	32.47
Plain	20	0.037	31.31
Plain	**40**	**0.580**	**38.84**
Fractal Col #1	5	0.677	37.23
Fractal Col #2	10	0.141	32.85
Fractal Col #3	20	0.029	31.31
Fractal Col #4	**40**	**0.016**	**31.75**
Fractal Full	**40**	**0.015**	**27.40**

Effectiveness of Fractal Structure on CIFAR-100++

If plain network with depth of 40 is used, overfitting occurs, while fractal structure can still obtain good performance, which shows FractalNet may be less prone to overfitting.

3.3. Student-Teacher Information Flow
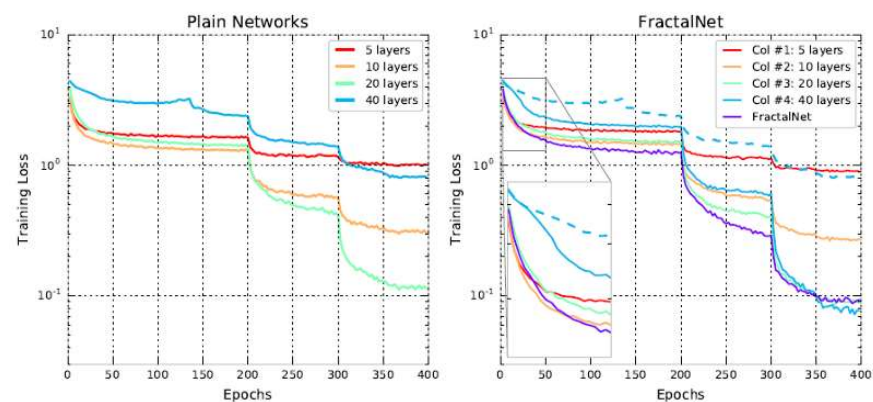
Training Loss Against Epoches for Plain Network (Left) and FractalNet (Right)

For plain network with 40 layers, the speed of convergence is very slow.

But for the col #4 of FractalNet which has the same 40 layers, the speed of convergence is much faster. Authors claim that there are help from other columns, i.e. Student-Teacher Information Flow.

· · ·

4. Results

4.1. CIFAR-10 (C10), CIFAR-100 (C100), SVHN (Street VIew House Number)

Method	C100	C100+	C100++	C10	C10+	C10++	SVHN
Network in Network (Lin et al., 2013)	35.68	-	-	10.41	8.81	-	2.35
Generalized Pooling (Lee et al., 2016)	32.37	-	-	7.62	6.05	-	1.69
Recurrent CNN (Liang & Hu, 2015)	31.75	-	-	8.69	7.09	-	1.77
Multi-scale (Liao & Carneiro, 2015)	27.56	-	-	6.87	-	-	1.76
FitNet Romero et al. (2015)	-	35.04	-	-	8.39	-	2.42
Deeply Supervised (Lee et al., 2014)	-	34.57	-	9.69	7.97	-	1.92
All-CNN (Springenberg et al., 2014)	-	33.71	-	9.08	7.25	4.41	-
Highway Net (Srivastava et al., 2015)	-	32.39	-	-	7.72	-	-
ELU (Clevert et al., 2016)	-	24.28	-	-	6.55	-	-
Scalable BO (Snoek et al., 2015)	-	-	27.04	-	-	6.37	1.77
Fractional Max-Pool (Graham, 2014)	-	-	26.32	-	-	3.47	-
FitResNet (Mishkin & Matas, 2016)	-	27.66	-	-	5.84	-	-
ResNet (He et al., 2016a)	-	-	-	-	6.61	-	-
ResNet by (Huang et al., 2016b)	44.76	27.22	-	13.63	6.41	-	2.01
Stochastic Depth (Huang et al., 2016b)	37.80	24.58	-	11.66	5.23	-	1.75
Identity Mapping (He et al., 2016b)	-	22.68	-	-	4.69	-	-
ResNet in ResNet (Targ et al., 2016)	-	22.90	-	-	5.01	-	-
Wide (Zagoruyko & Komodakis, 2016)	-	20.50	-	-	4.17	-	-
DenseNet-BC (Huang et al., 2016a)[1]	19.64	17.60	-	5.19	3.62	-	1.74
FractalNet (20 layers, 38.6M params)	35.34	23.30	22.85	10.18	5.22	5.11	2.01
+ drop-path + dropout	28.20	23.73	23.36	7.33	4.60	4.59	1.87
↳ deepest column alone	29.05	24.32	23.60	7.27	4.68	4.63	1.89
FractalNet (40 layers, 22.9M params)[2]	-	22.49	21.49	-	5.24	5.21	-

CIFAR-10, CIFAR-100, SVHN Results, (+: Data augmentation, ++: Heavy data augmentation)

- **Without any data augmentation,** FractalNet obtain the lowest test error except the DenseNet-BC. And **FractalNet outperforms ResNet by largin especially for C10 and C100.**

- **With data augmentation, FractalNet can still outperforms almost all ResNet variants except the wide one (WRN)** for C10+ and C100+.

- But with heavy data augmentation (C10++ and C100++), FractalNet still cannot have the best result when comparing with those in C10+ or C100+.

- And DenseNet-BC, which was the concurrent work at that moment, outperforms FractalNet for all tasks. But in this paper, the main purpose is to compare with ResNet, i.e. non-residual vs residual.

4.2. ImageNet

Method	Top-1 (%)	Top-5 (%)
VGG-16	28.07	9.33
ResNet-34 C	24.19	7.40
FractalNet-34	24.12	7.39

ImageNet Results

FractalNet-34 has a little better results comparing with ResNet-34 C.

When comparing with non-residual-network, FractalNet-34 outperforms VGG-16 by large margin.

.   .   .

Reference

[2017 ICLR] [FractalNet]
FractalNet: Ultra-Deep Neural Networks without Residuals

My Related Reviews on Image Classification

[LeNet] [AlexNet] [ZFNet] [VGGNet] [SPPNet] [PReLU-Net] [DeepImage] [GoogLeNet / Inception-v1] [BN-Inception / Inception-v2] [Inception-v3] [Inception-v4] [Xception] [MobileNetV1] [ResNet] [Pre-Activation ResNet] [RiR] [RoR] [Stochastic Depth] [WRN] [Trimps-Soushen] [PolyNet] [ResNeXt] [DenseNet]

Gain Access to Expert Views

First Name

Email

Give me access!

I agree to leave Medium and submit this information, which will be collected and used according to Upscribe's privacy policy.