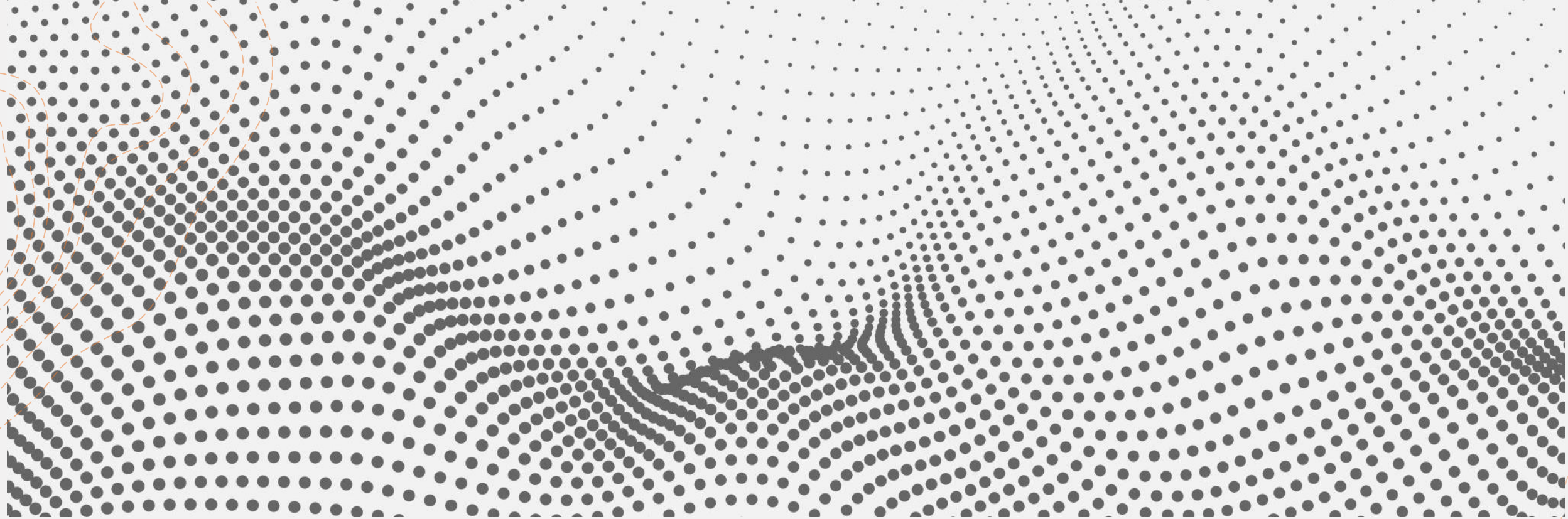


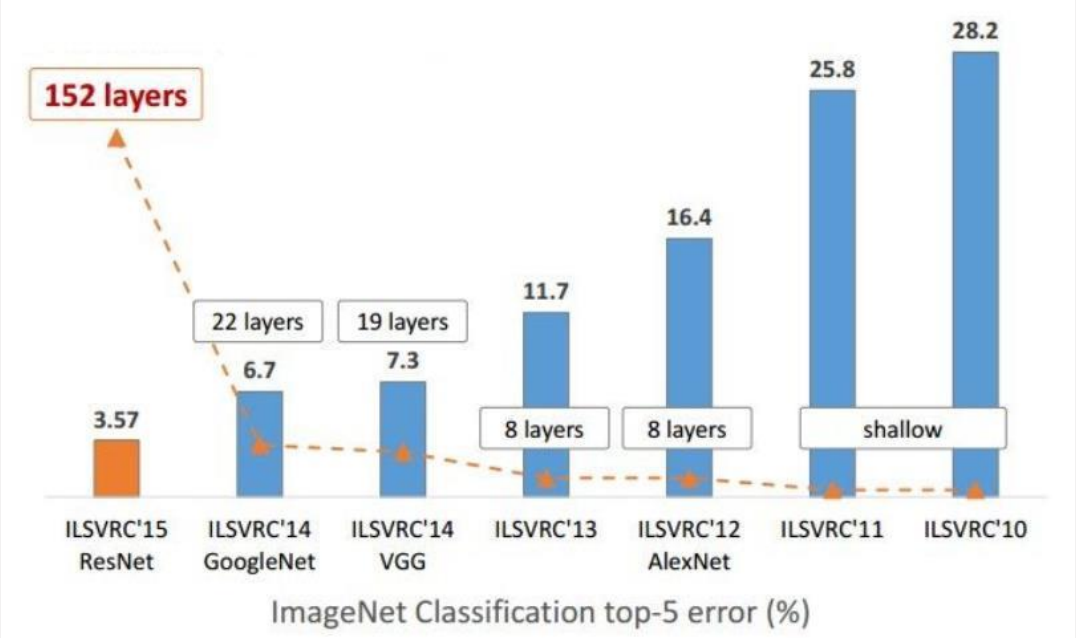
DenseNets

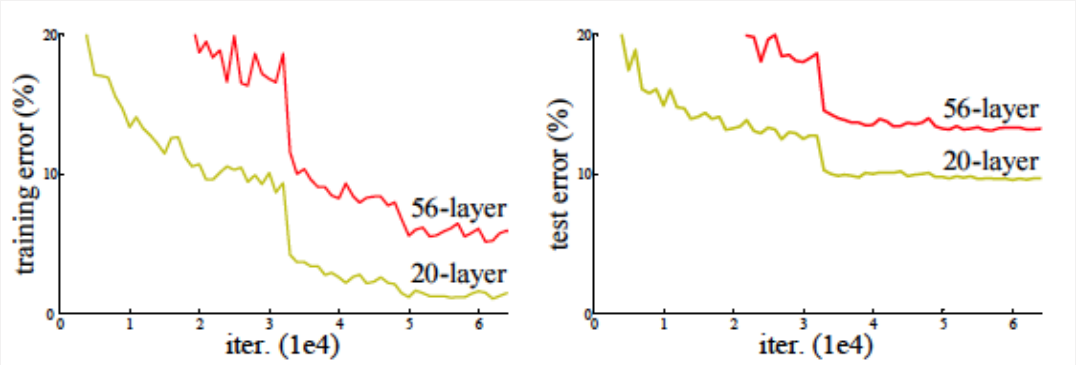
분석 - Classification

16기 분석 이재원

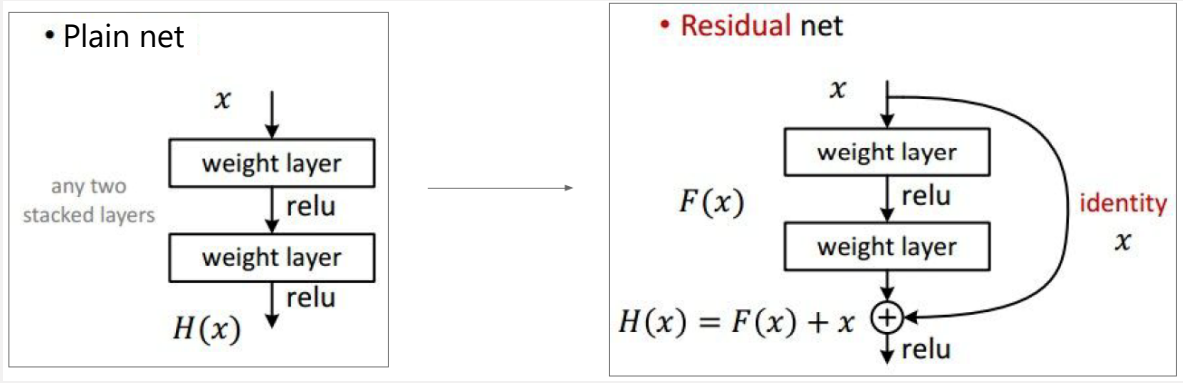
Introduction

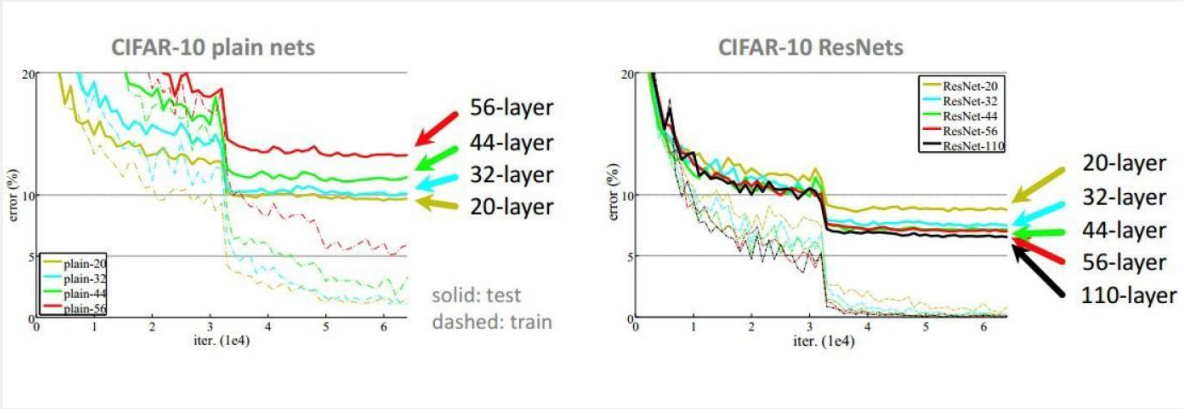
Problem statement: depth





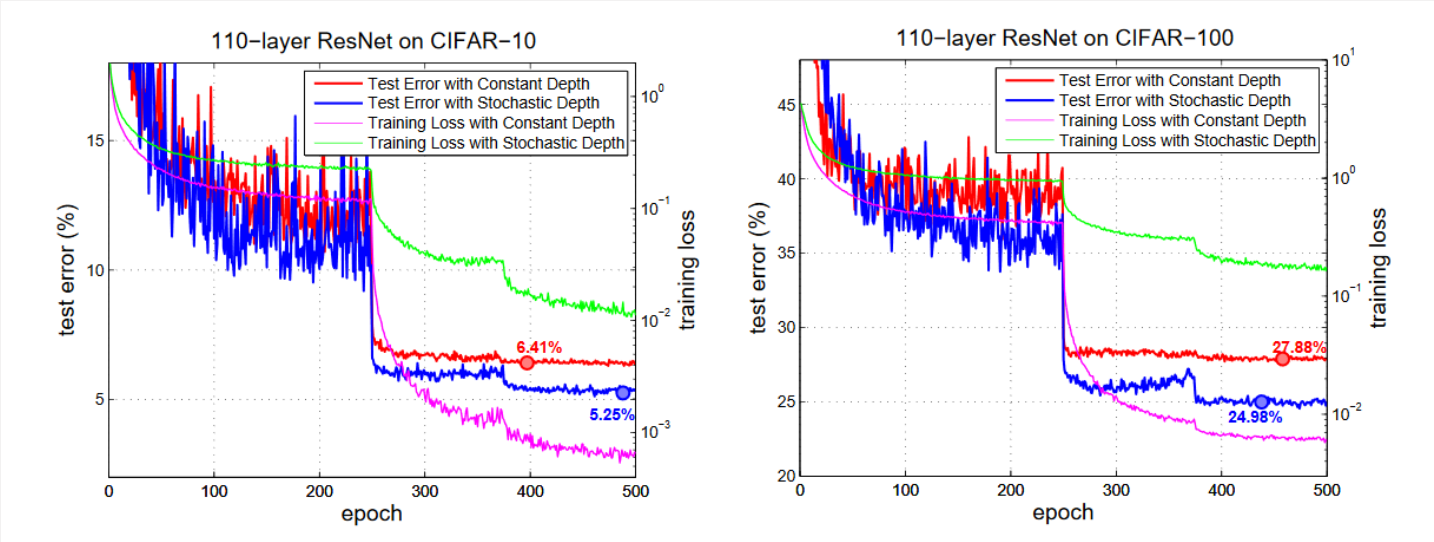
depth: ResNet (Skip Connection)

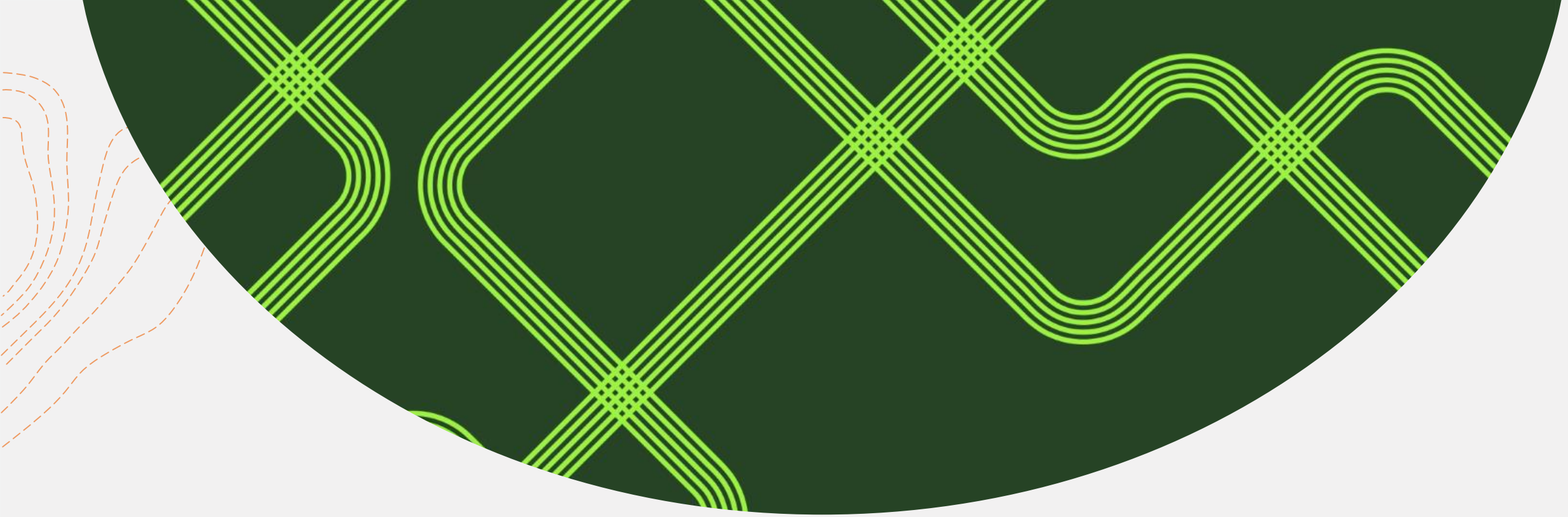




depth: Stochastic Depth

+testing(deep) + training(shallow)

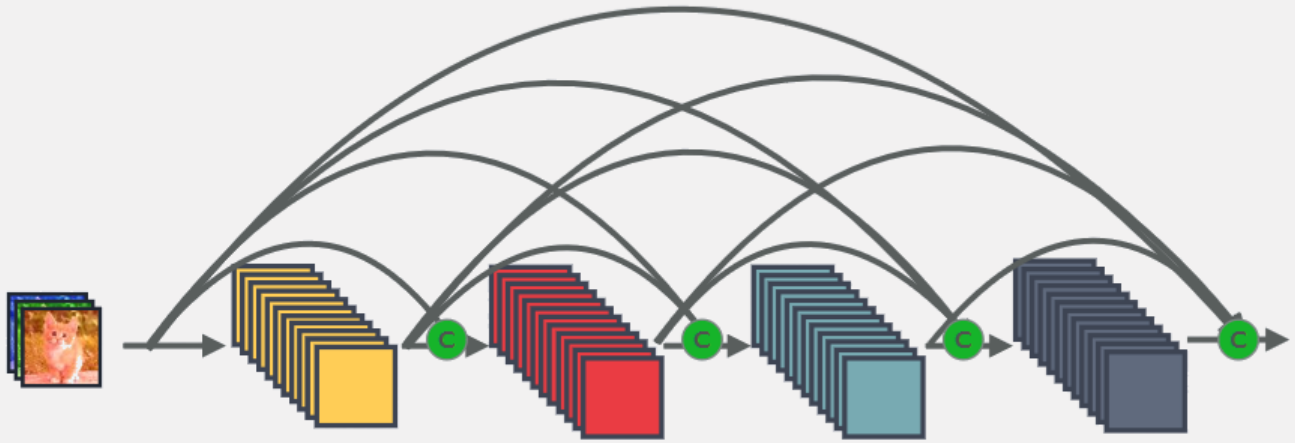




Model Structure

Dense Connectivity: *dense* + *concat*

+dense: 모든 layer를 서로 연결!



Concatenation!

- **Traditional Convolutional feed-forward networks :**

$$x_l = H_l\left(x_{l-1}\right)$$

- **ResNets :**

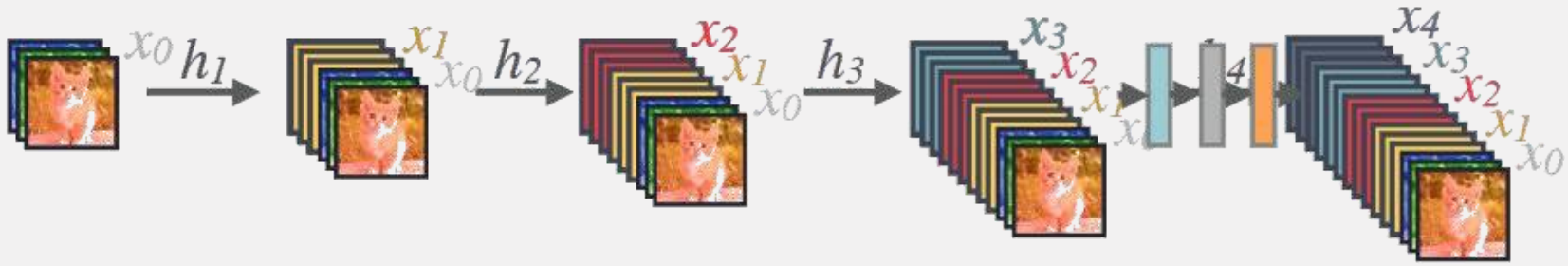
$$x_l = H_l\left(x_{l-1}\right) + x_{l-1}$$

- **DenseNets :**

$$x_l = H_l\left([x_0, x_1, \ldots, x_{l-1}]\right)$$

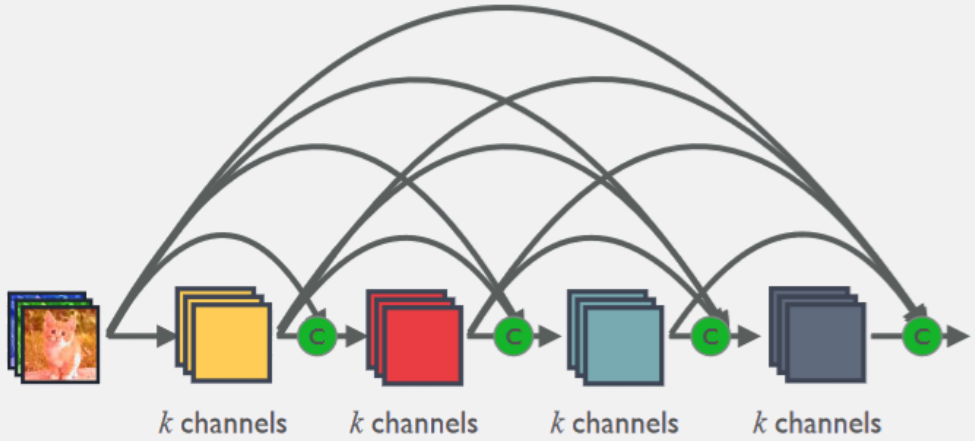
Where $[x_0, x_1, \ldots, x_{l-1}]$ refers to the concatenation of the feature-maps produced in layers 0….. l-1.

Concatenation!



+Parameter 개수는?

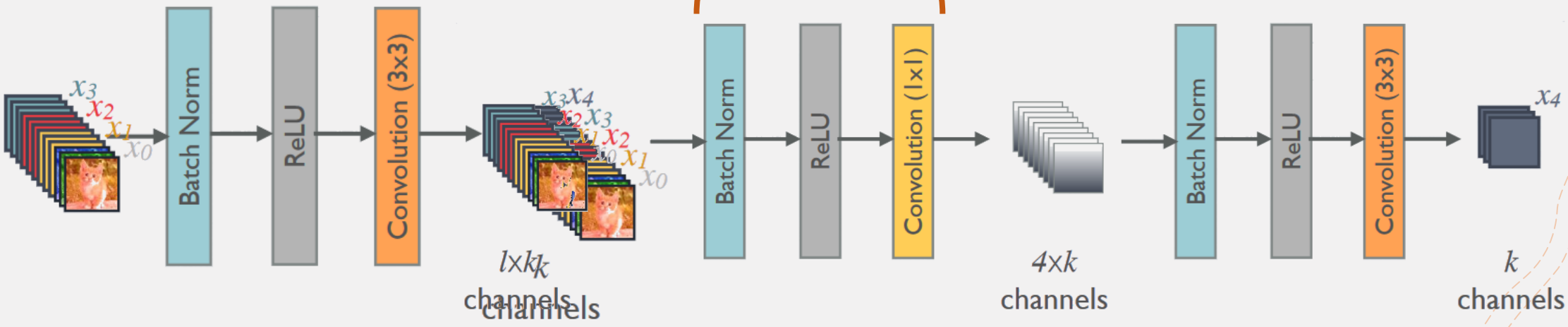
Growth Rate (k)



k : Growth Rate

Model Structure: Composite layer

+ Bottleneck layer

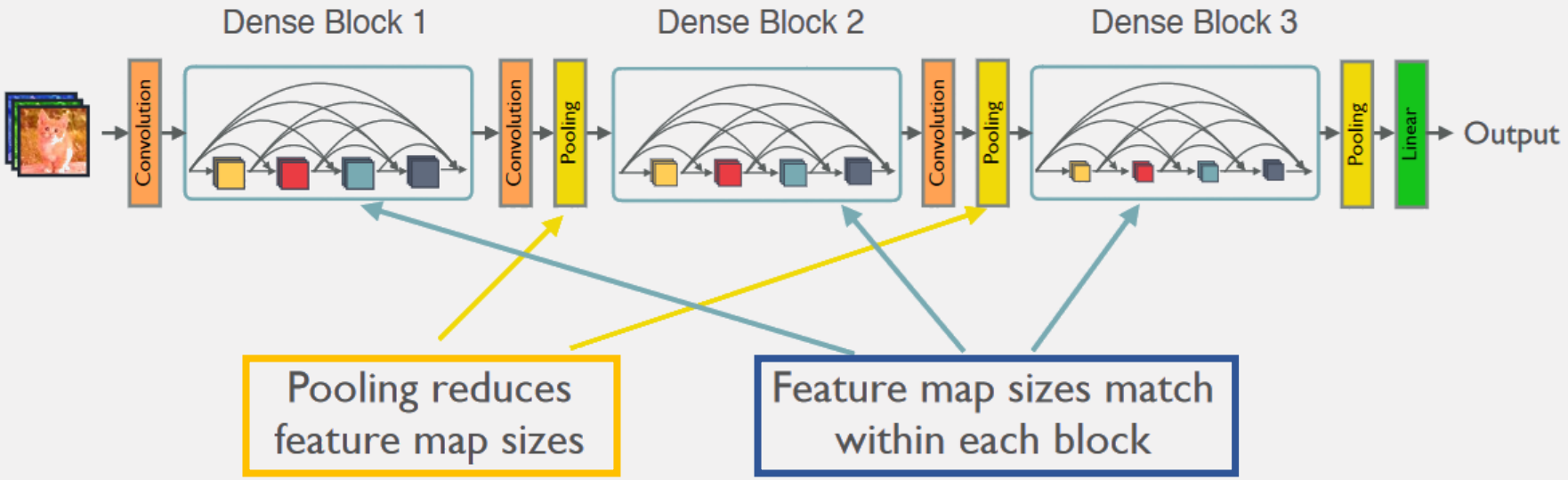


$x_5 = h_5([x_0, \ldots, x_4])$

Higher parameter and computational efficiency

"DenseNet-B"

Overall structure



Transition Layer: compression

- Reduce # of feature maps(channels) at transition layers
- 실험에서는 0.5로 진행
- “DenseNet – C”
- Bottleneck layer와 compression: “DenseNet – BC”

Overall Structure: implementation

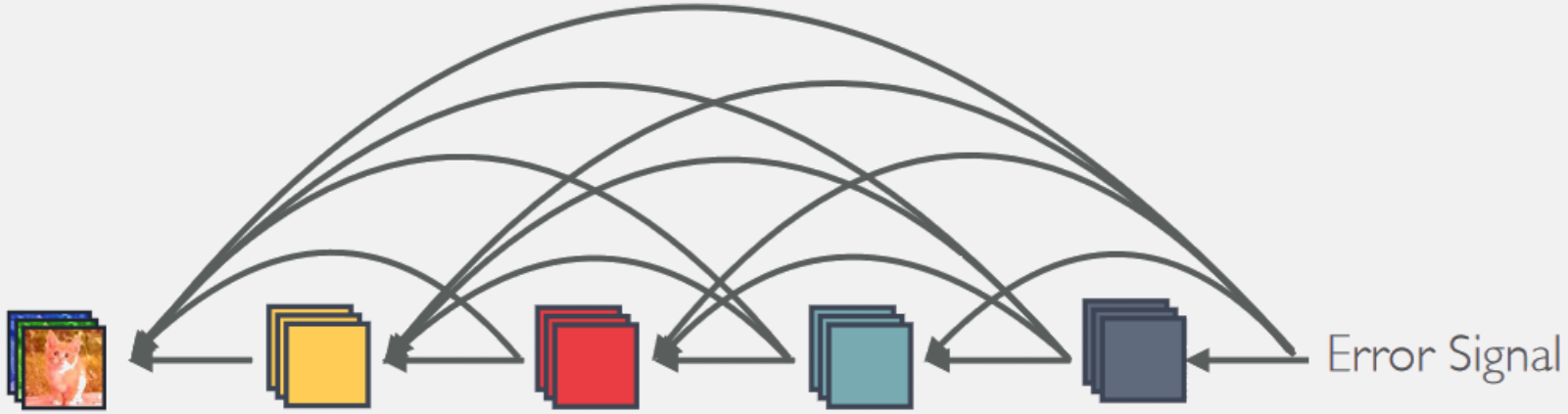
Bottleneck + Composite layer (BN + ReLU + Conv)

Layers	Output Size	DenseNet-121($k = 32$)	DenseNet-169($k = 32$)	DenseNet-201($k = 32$)	DenseNet-161($k = 48$)
Convolution	112 × 112	7 × 7 conv, stride 2			
Pooling	56 × 56	3 × 3 max pool, stride 2			
Dense Block (1)	56 × 56	[1 × 1 conv; 3 × 3 conv] × 6	[1 × 1 conv; 3 × 3 conv] × 6	[1 × 1 conv; 3 × 3 conv] × 6	[1 × 1 conv; 3 × 3 conv] × 6
Transition Layer (1)	56 × 56	1 × 1 conv			
	28 × 28	2 × 2 average pool, stride 2			
Dense Block (2)	28 × 28	[1 × 1 conv; 3 × 3 conv] × 12	[1 × 1 conv; 3 × 3 conv] × 12	[1 × 1 conv; 3 × 3 conv] × 12	[1 × 1 conv; 3 × 3 conv] × 12
Transition Layer (2)	28 × 28	1 × 1 conv			
	14 × 14	2 × 2 average pool, stride 2			
Dense Block (3)	14 × 14	[1 × 1 conv; 3 × 3 conv] × 24	[1 × 1 conv; 3 × 3 conv] × 32	[1 × 1 conv; 3 × 3 conv] × 48	[1 × 1 conv; 3 × 3 conv] × 36
Transition Layer (3)	14 × 14	1 × 1 conv			
	7 × 7	2 × 2 average pool, stride 2			
Dense Block (4)	7 × 7	[1 × 1 conv; 3 × 3 conv] × 16	[1 × 1 conv; 3 × 3 conv] × 32	[1 × 1 conv; 3 × 3 conv] × 32	[1 × 1 conv; 3 × 3 conv] × 24
Classification Layer	1 × 1	7 × 7 global average pool			
		1000D fully-connected, softmax			

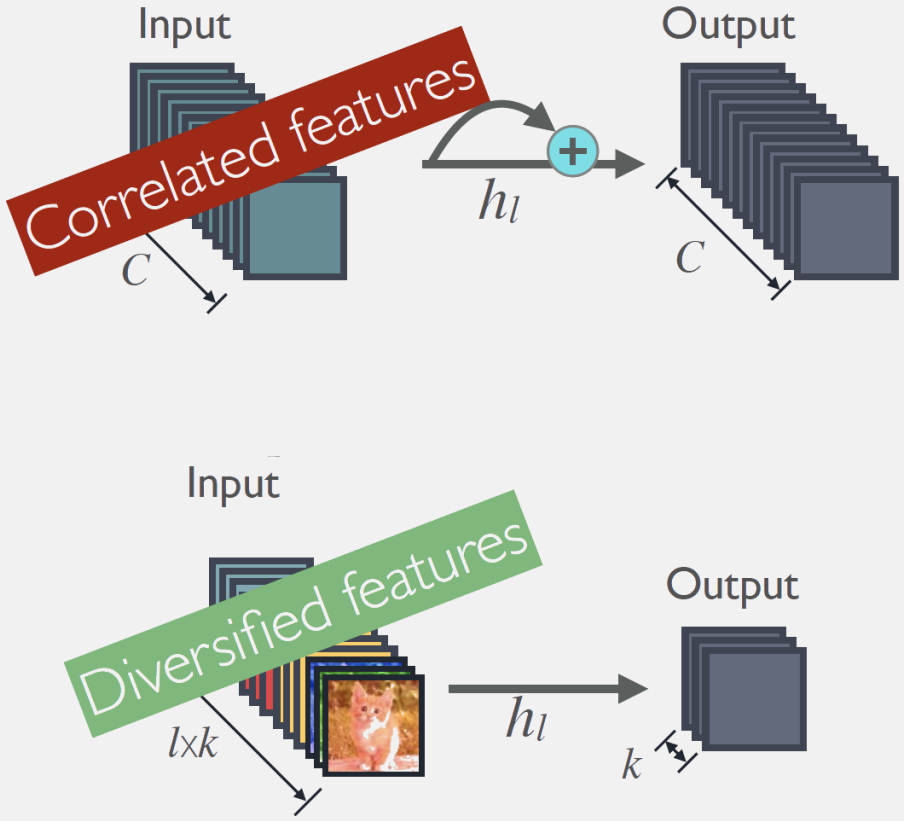
Advantages

1. Strong gradient flow

+ Backprop: ”deep supervision” with single classifier
+ Regularization



2. Parameter & computational efficiency



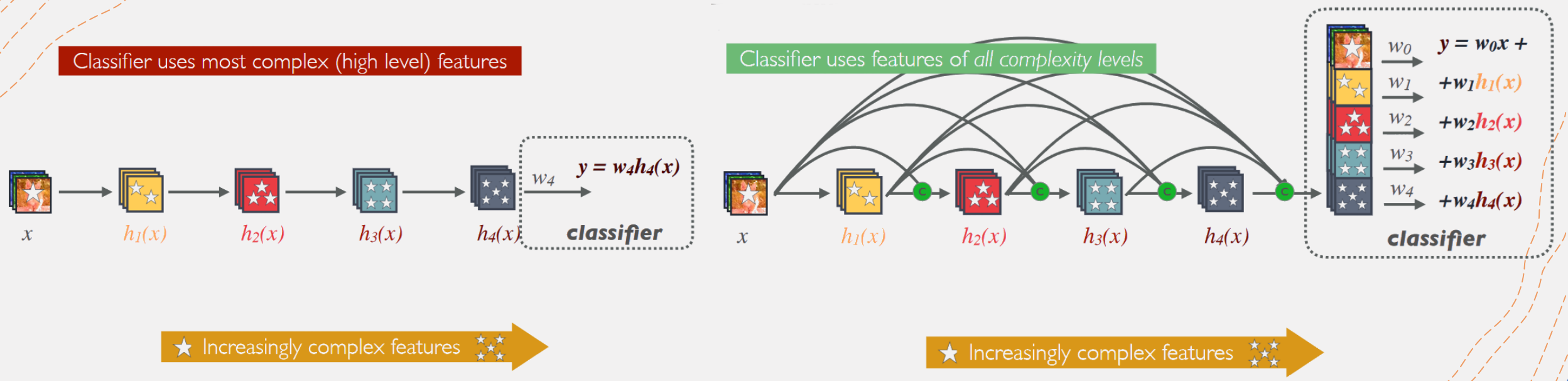
$O(C \times C)$

$k \ll C$

$O(l \times k \times k)$

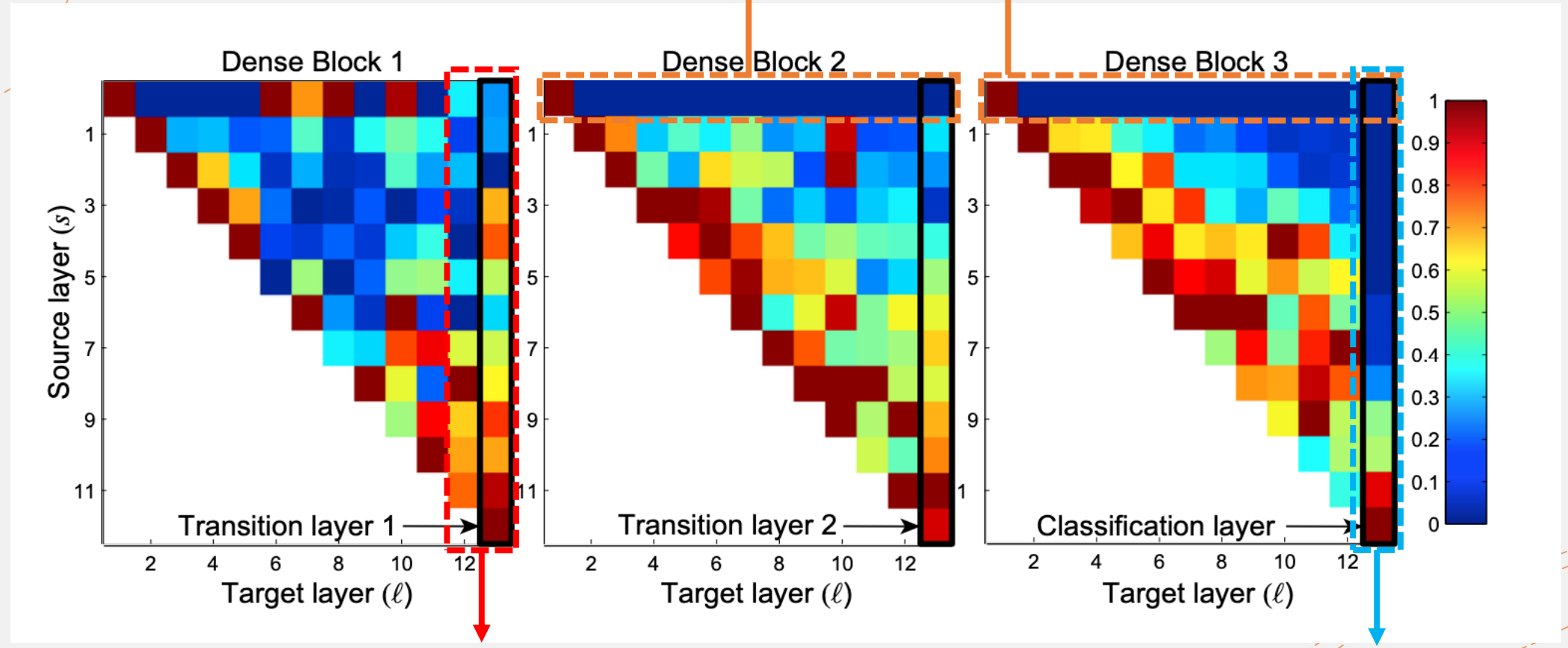
3. Low-level features

+ Feature Reuse



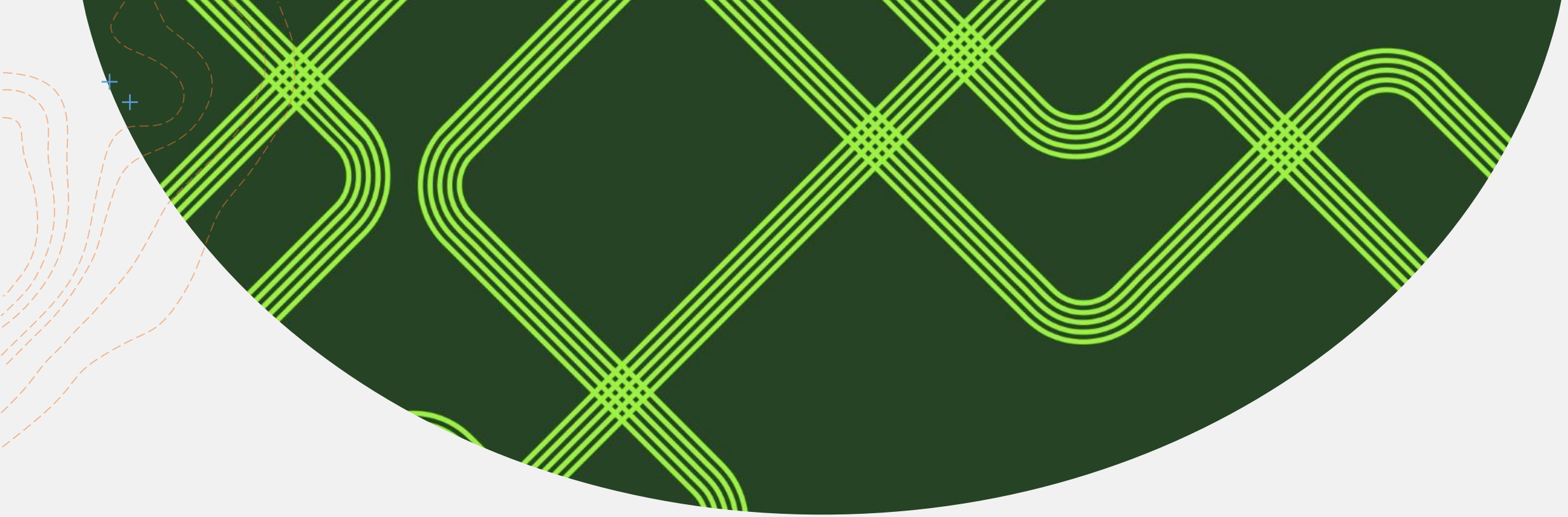
Low-level features

Redundant features



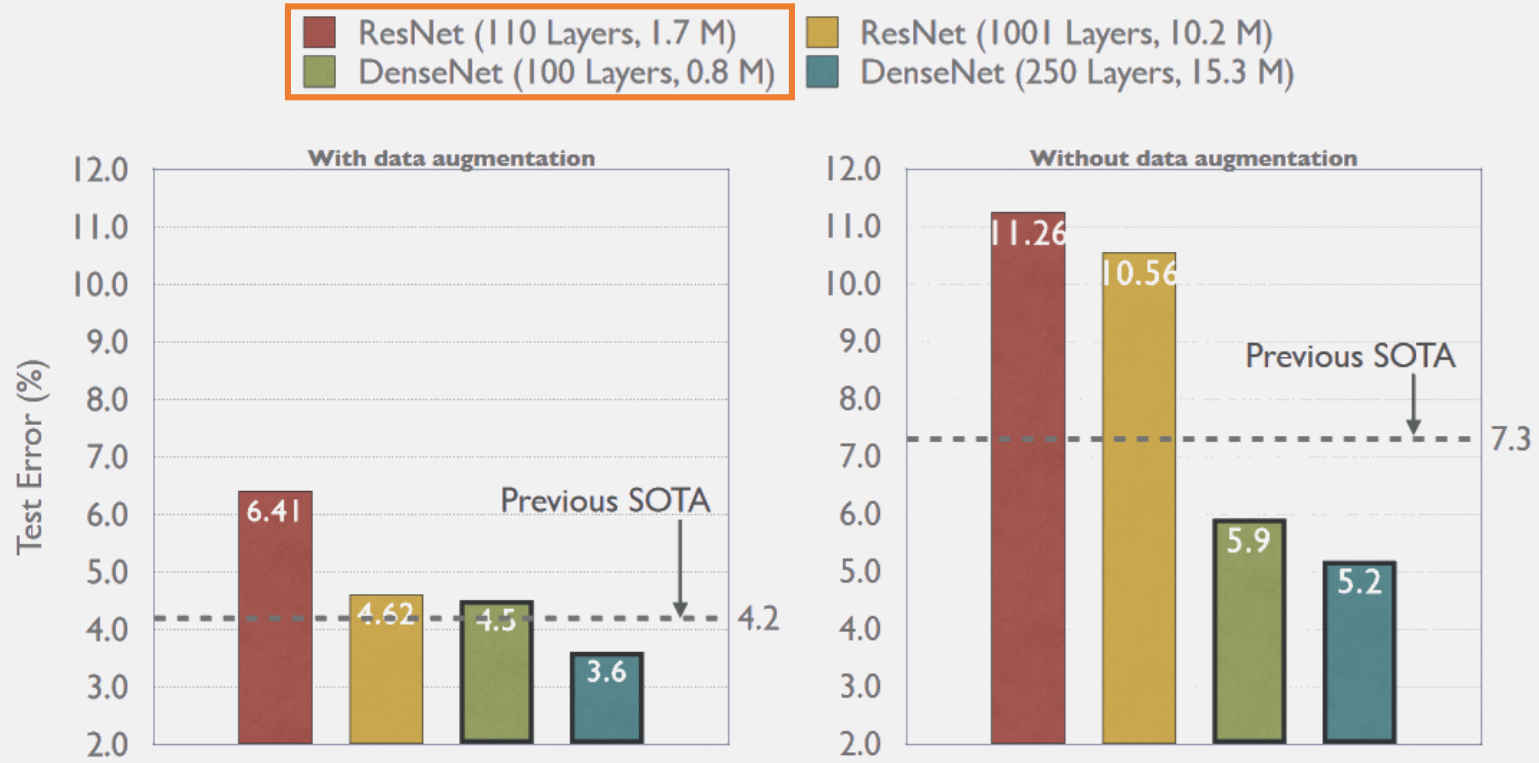
utilizing low-level features

high-level features



Results

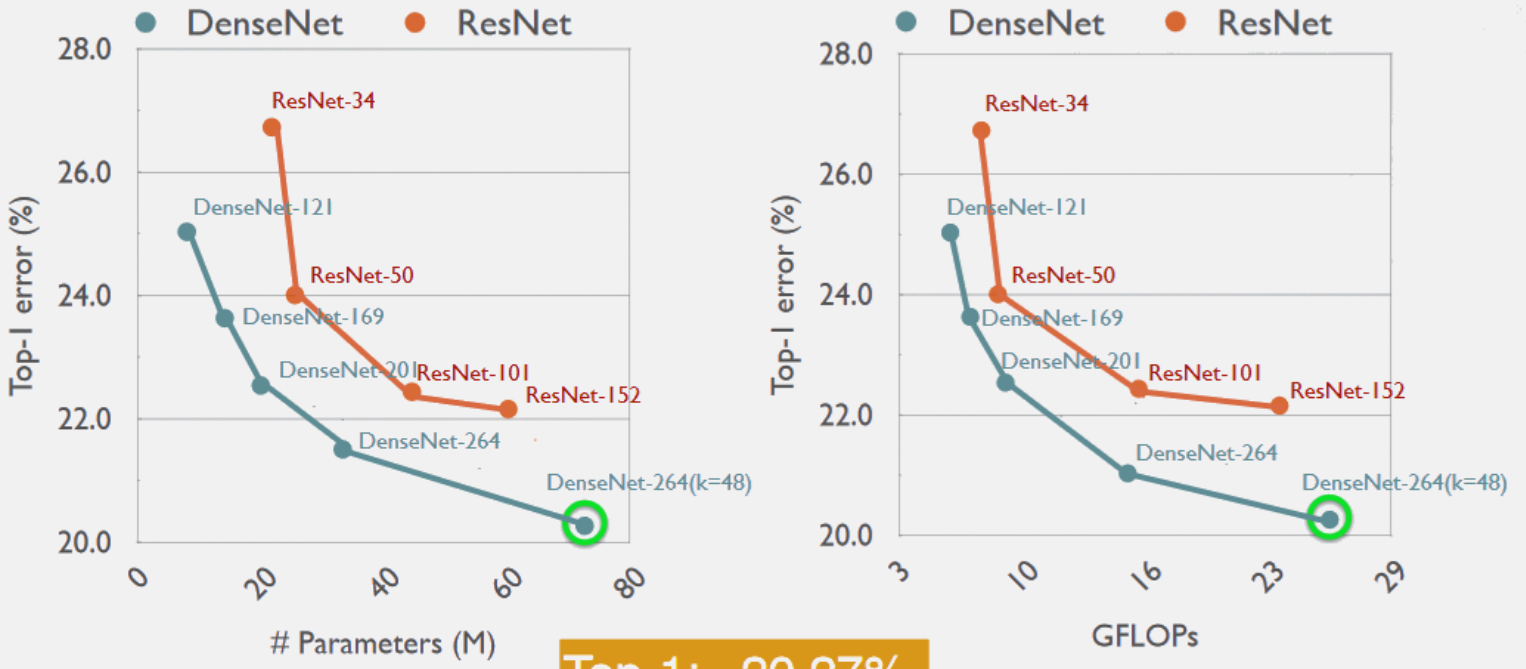
CIFAR-10


ResNet (110 Layers, 1.7 M)
DenseNet (100 Layers, 0.8 M)
ResNet (1001 Layers, 10.2 M)
DenseNet (250 Layers, 15.3 M)
With data augmentation
Test Error (%)
12.0
11.0
10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
6.41
4.62
4.5
3.6
Previous SOTA
4.2
Without data augmentation
11.26
10.56
5.9
5.2
Previous SOTA
7.3

Error rates & parameters

Method	Depth	Params	C10	C10+	C100	C100+	SVHN
Network in Network [22]	-	-	10.41	8.81	35.68	-	2.35
All-CNN [32]	-	-	9.08	7.25	-	33.71	-
Deeply Supervised Net [20]	-	-	9.69	7.97	-	34.57	1.92
Highway Network [34]	-	-	-	7.72	-	32.39	-
FractalNet [17]	21	38.6M	10.18	5.22	35.34	23.30	2.01
with Dropout/Drop-path	21	38.6M	7.33	4.60	28.20	23.73	1.87
ResNet [11]	110	1.7M	-	6.61	-	-	-
ResNet (reported by [13])	110	1.7M	13.63	6.41	44.74	27.22	2.01
ResNet with Stochastic Depth [13]	110	1.7M	11.66	5.23	37.80	24.58	1.75
	1202	10.2M	-	4.91	-	-	-
Wide ResNet [42]	16	11.0M	-	4.81	-	22.07	-
	28	36.5M	-	4.17	-	20.50	-
with Dropout	16	2.7M	-	-	-	-	1.64
ResNet (pre-activation) [12]	164	1.7M	11.26*	5.46	35.58*	24.33	-
	1001	10.2M	10.56*	4.62	33.47*	22.71	-
DenseNet ($k = 12$)	40	1.0M	**7.00**	5.24	**27.55**	24.42	1.79
DenseNet ($k = 12$)	100	7.0M	**5.77**	**4.10**	**23.79**	**20.20**	1.67
DenseNet ($k = 24$)	100	27.2M	**5.83**	**3.74**	**23.42**	**19.25**	**1.59**
DenseNet-BC ($k = 12$)	100	0.8M	**5.92**	4.51	**24.15**	22.27	1.76
DenseNet-BC ($k = 24$)	250	15.3M	**5.19**	**3.62**	**19.64**	**17.60**	1.74
DenseNet-BC ($k = 40$)	190	25.6M	-	**3.46**	-	**17.18**	-

ImageNet



Top-1: 20.27%
Top-5: 5.17%

감사합니다!

Reference:

"Densely Connected Convolutional Networks." arXiv:1608.06993