

The title of paper

Densely Connected Convolutional Networks

<https://arxiv.org/pdf/1608.06993.pdf>

One sentence described about the paper

Summary

By Noritsugu Yamada
2019/01/23



Densely Connected Convolutional Networks

Gao Huang*
Cornell University
gh349@cornell.edu

Zhuang Liu*
Tsinghua University
liuzhuang13@mails.tsinghua.edu.cn

Laurens van der Maaten
Facebook AI Research
lvdmaaten@fb.com

Kilian Q. Weinberger
Cornell University
kqw4@cornell.edu

- 1, Summary
- 2, What is ~ ...??
- 3, Experiments, conclusion, and discussion

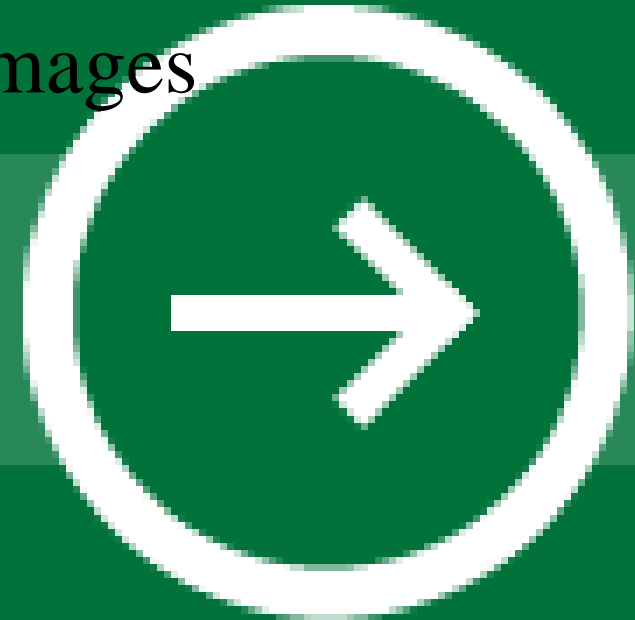
Auter

<https://arxiv.org/pdf/1608.06993.pdf>

Method of improving the quality of synthetic images

Summary

By Noritsugu Yamada
2019/01/23



Conclusion:

DenseNetは勾配消失問題に対して有効な方法を示したILSVRC2015 winner ResNetを改良し、DenseNetを導入することで画像分類精度を向上し、パラメータの減少を実現した。

What is this thesis for?

CNNにおける層が深いモデルの成功例であるResNetを改良したもの

How to verified whether it is valid?

CIFAR. , ImageNet. ,SVHN.という有名なデータセットにおけるResNetとの精度の比較

Where is an important point compared to previous researches?

DenseNetはResNetよりも画像分類の精度を大幅に向上しパラメータも減少させた

Is there discussions?

すいません。わかりません

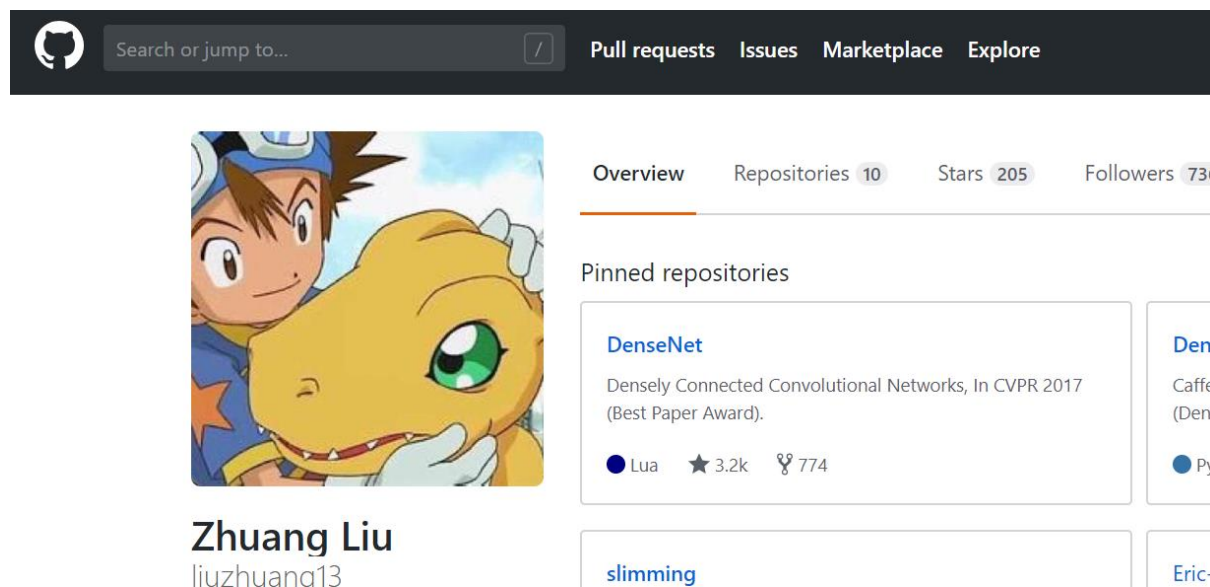
Where are the key points of technology and method?

ResNetの入力と学習の**合計**（正しくは求めたい関数-入力の差分を学習する）からDense Blockの**積み重ねの連結**に置き換えた

Which reserches should I read next?

Identity mappings in deep residual networks. In ECCV, 2016

著者のGithubがデジモンで親近感がわいた
太一の声優さん亡くなってしまった

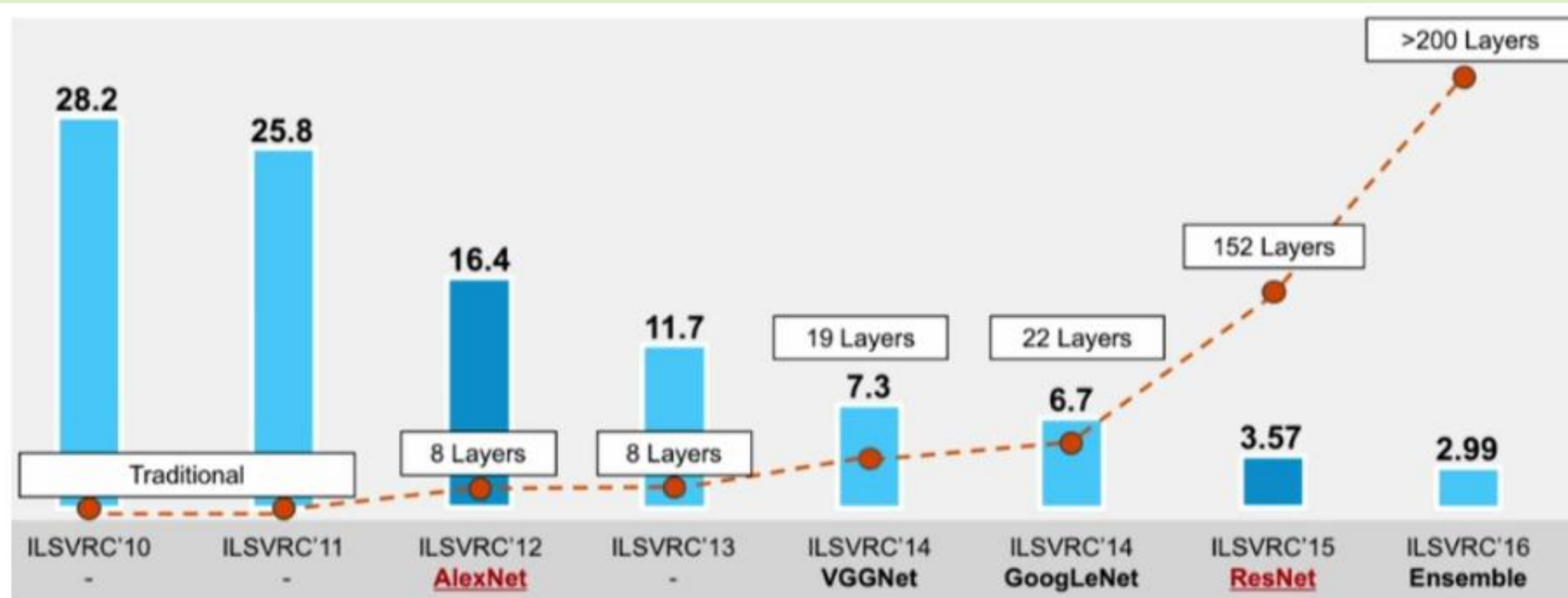


What is

DenseNet....??

What is ~...??:

Figure ResNet



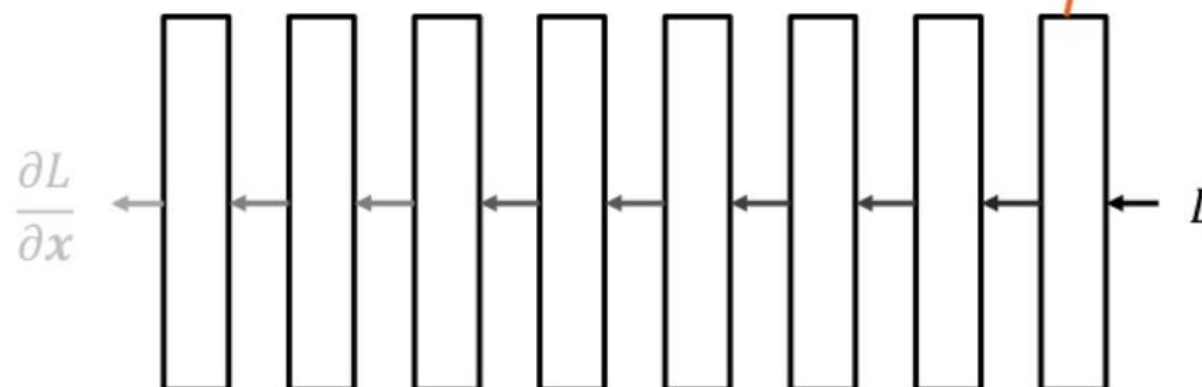
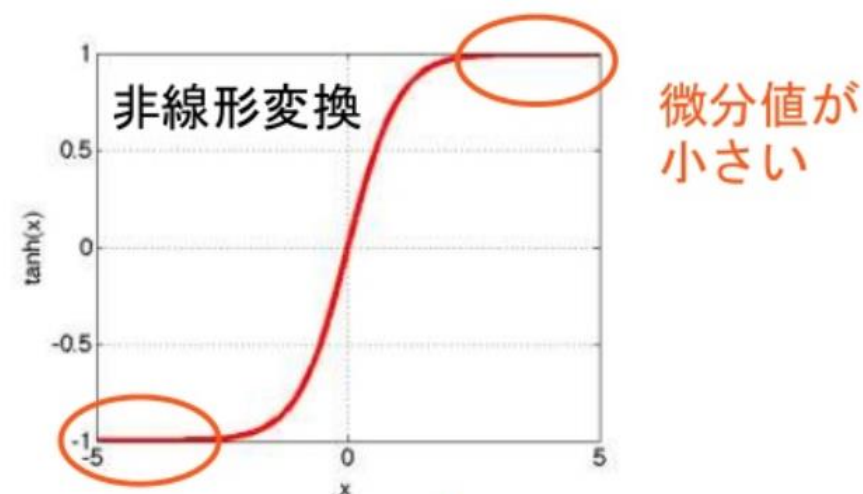
2014年までは層が深くても
22層だった

What is ~...??:

Figure ResNet

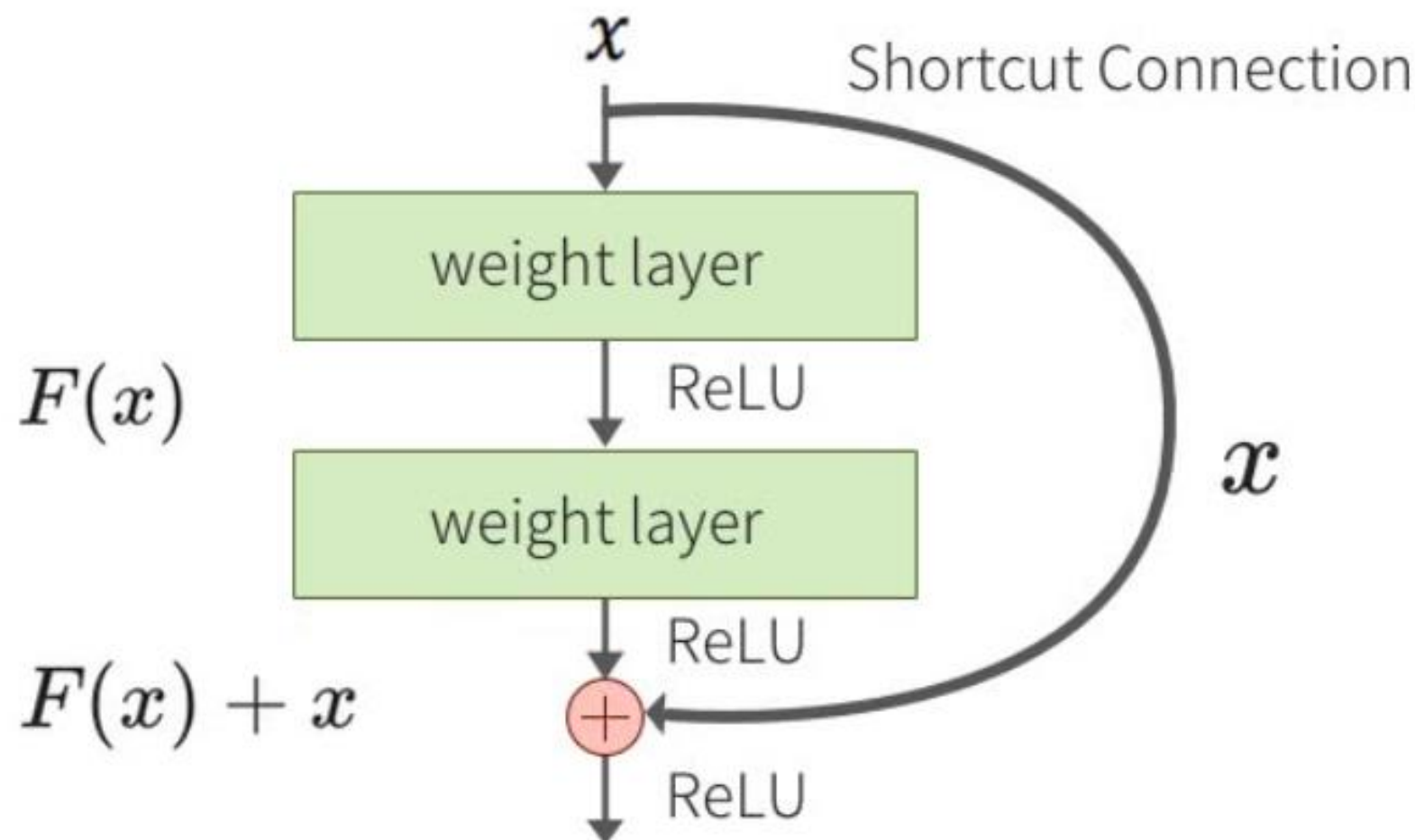
勾配消失問題 (Vanishing Gradient)

- チェインルールで微分を計算していくと、小さな値の掛け合わせで微分が0に近づく問題
- 深層ネットワークの学習の難しさの主要な要因
- LSTMやResNetなどで解決



What is $\sim\sim\ldots??$:

Figure ResNet

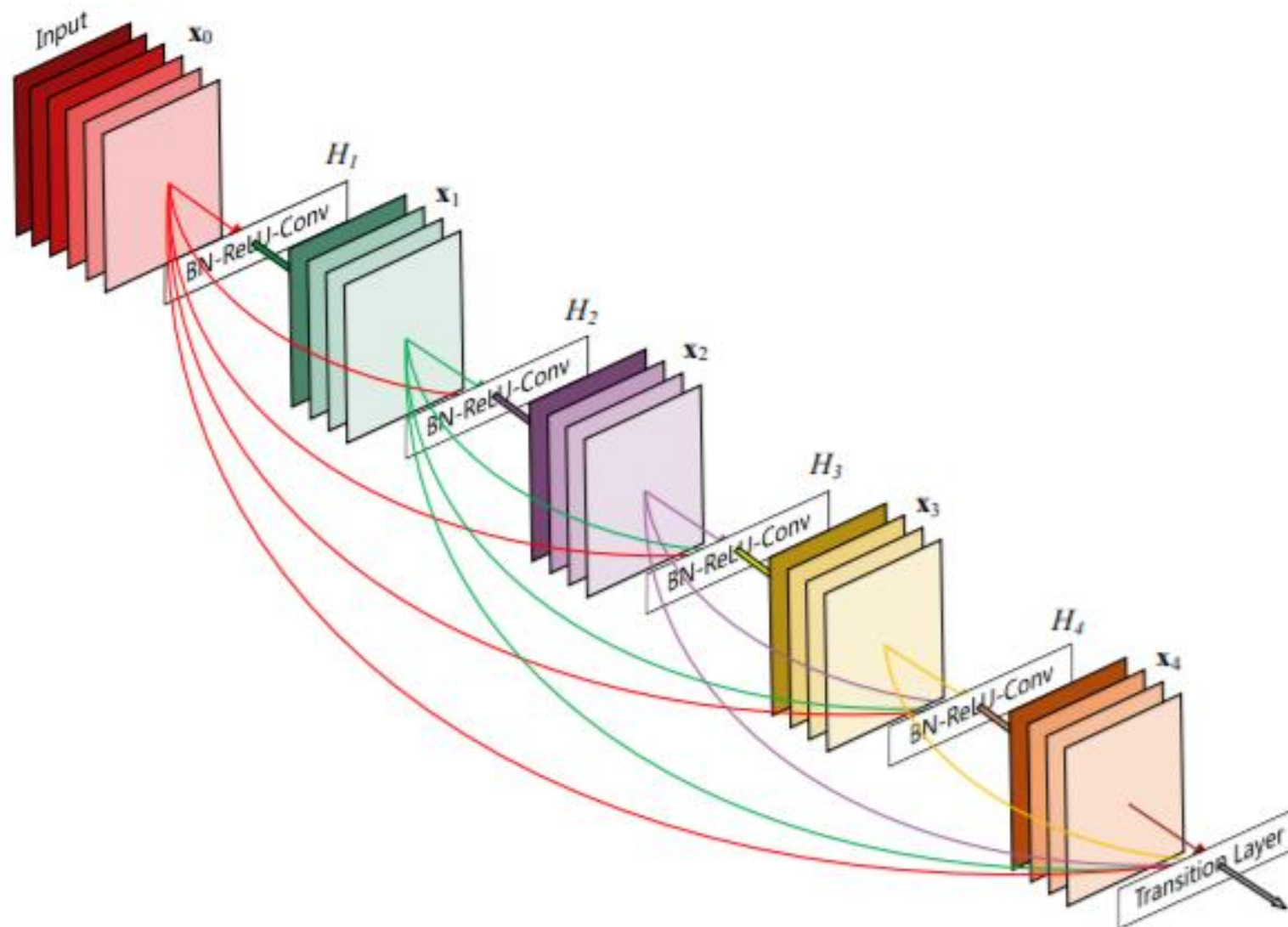


求める関数 $H(x)$ とすると
 $H(x)$ から入力 x を引いた **残差**
 $f(x)$ を CNN で学習する

出力で入力値を加算すること
 で勾配が小さくならず勾配消
 失問題に有効性を示した

What is $\sim\sim\ldots??$:

Figure 5-layer dense block

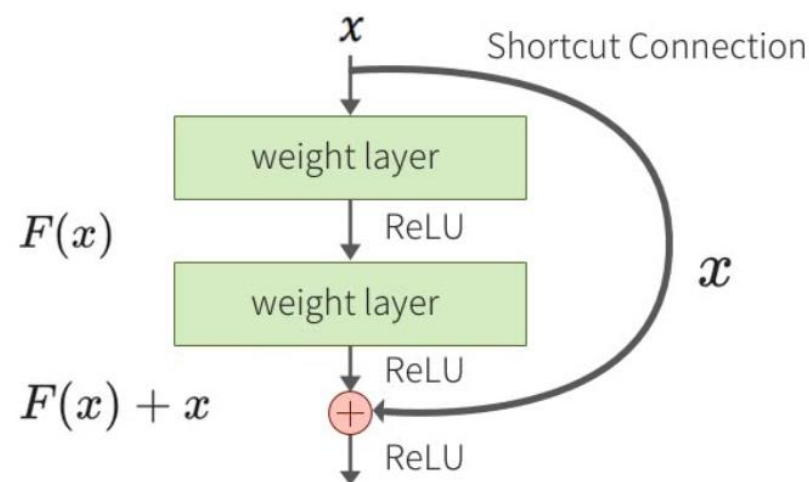


5層のDenseNet 構造
畳み込みを繰り返す密な
ブロック=Denseblock

浅い層で学習したものが後の層のすべての入力になる

What is ~...??:

Figure DenseNet and ResNet



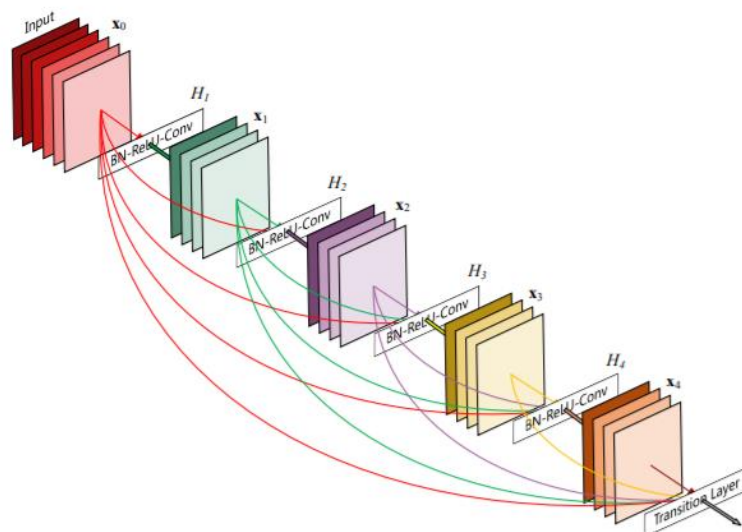
ResNets

ResNets は出力に恒等関数を使った非線形変換を迂回するスキップ接続を追加する

$$\mathbf{x}_\ell = H_\ell(\mathbf{x}_{\ell-1}) + \mathbf{x}_{\ell-1}.$$

DenseNet

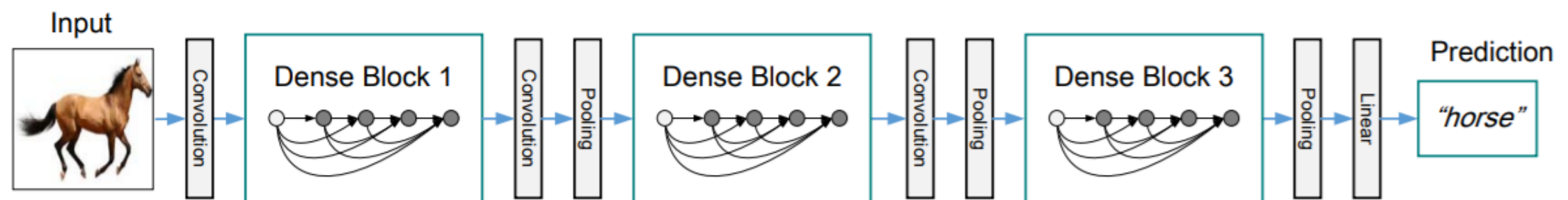
DenseNetは任意の層からそれ以降のすべての層への直接接続を導入
レイヤ間の情報の流れをさらに改善



$$\mathbf{x}_\ell = H_\ell([\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{\ell-1}]),$$

What is ~...??:

Figure TransitionLayer



隣接Denseblockの間の層はTransitionLayerと呼ばれ、
バッチ正規化層と 1×1 の畳み込み層と
 2×2 のaverageプーリング層で構成される

DenseNet-B

ボトルネック層 (BN-ReLU-Conv (1×1) -BN-ReLU-Conv (3×3)) を有するバージョン

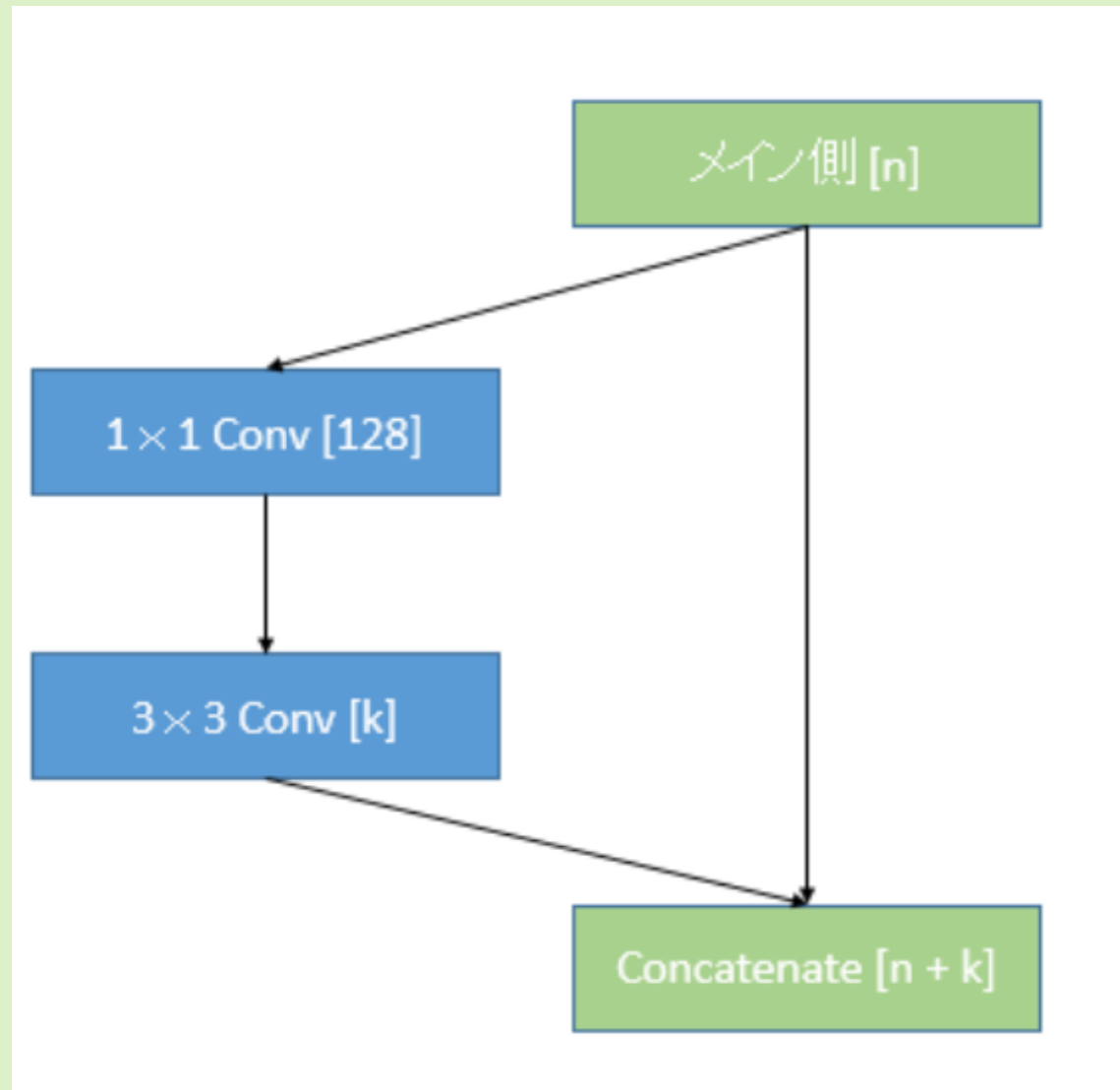
DenseNet-C

ボトルネックを使用せず圧縮係数を $0 < \theta < 1$ としてモデルを小型化するDenseNet (実験では $\theta = 0.5$)
 $\theta = 1$ なら圧縮しない

DenseNet-BCボトルネックと圧縮係数が $0 < \theta < 1$ の場合のモデル

What is ~~...??:

Figure 成長率k (Growth rate)



kは成長率といい
どれくらいメインに加えるか
設定するハイパーパラメータ

ℓ 層のfeaturemapsは
 $k_0 + k \times (\ell - 1)$

What is ~...??:

Figure DenseNet Architecture

Layers	Output Size	DenseNet-121	DenseNet-169	DenseNet-201	DenseNet-264
Convolution	112×112	7×7 conv, stride 2			
Pooling	56×56	3×3 max pool, stride 2			
Dense Block (1)	56×56	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$
Transition Layer (1)	56×56	1×1 conv			
	28×28	2×2 average pool, stride 2			
Dense Block (2)	28×28	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$
Transition Layer (2)	28×28	1×1 conv			
	14×14	2×2 average pool, stride 2			
Dense Block (3)	14×14	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 24$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 64$
Transition Layer (3)	14×14	1×1 conv			
	7×7	2×2 average pool, stride 2			
Dense Block (4)	7×7	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 16$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$
Classification Layer	1×1	7×7 global average pool			
		1000D fully-connected, softmax			

ImageNet用のDenseNetアーキテクチャ. すべてのネットワークの成長率は $k = 32$
 表に示されている各「conv」層は、シーケンスBN-RelU-Convに対応

What is ~...??:

Figure : Error rates (%) on CIFAR and SVHN datasets

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	-
	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	-

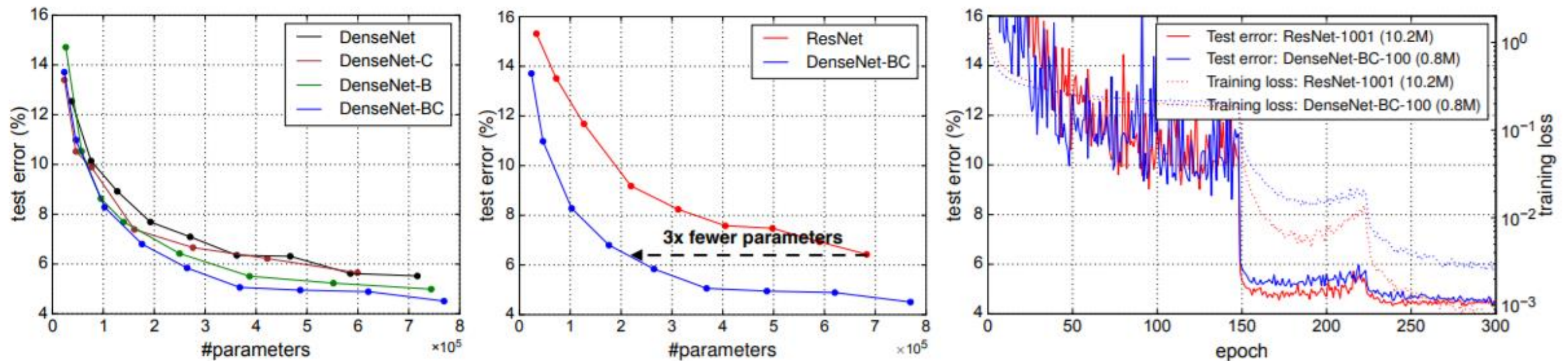
CIFARおよびSVHNデータセットのエラー率 (%)

全体で最良の結果は青文字

DenseNetは、ResNetよりも少ないパラメータを使用しながら低いエラー率を達成
データ増やさずに他のモデルよりDenseNetのパフォーマンスは大幅に向上

What is $\sim\sim\ldots??$:

Figure : :Comparison of the parameter efficiency on C10+ between DenseNet variations



左：DenseNetの種類ごとのC10+のパラメータ効率の比較

真ん中：DenseNet-BCと（preactivation）ResNetの間のパラメータ効率の比較

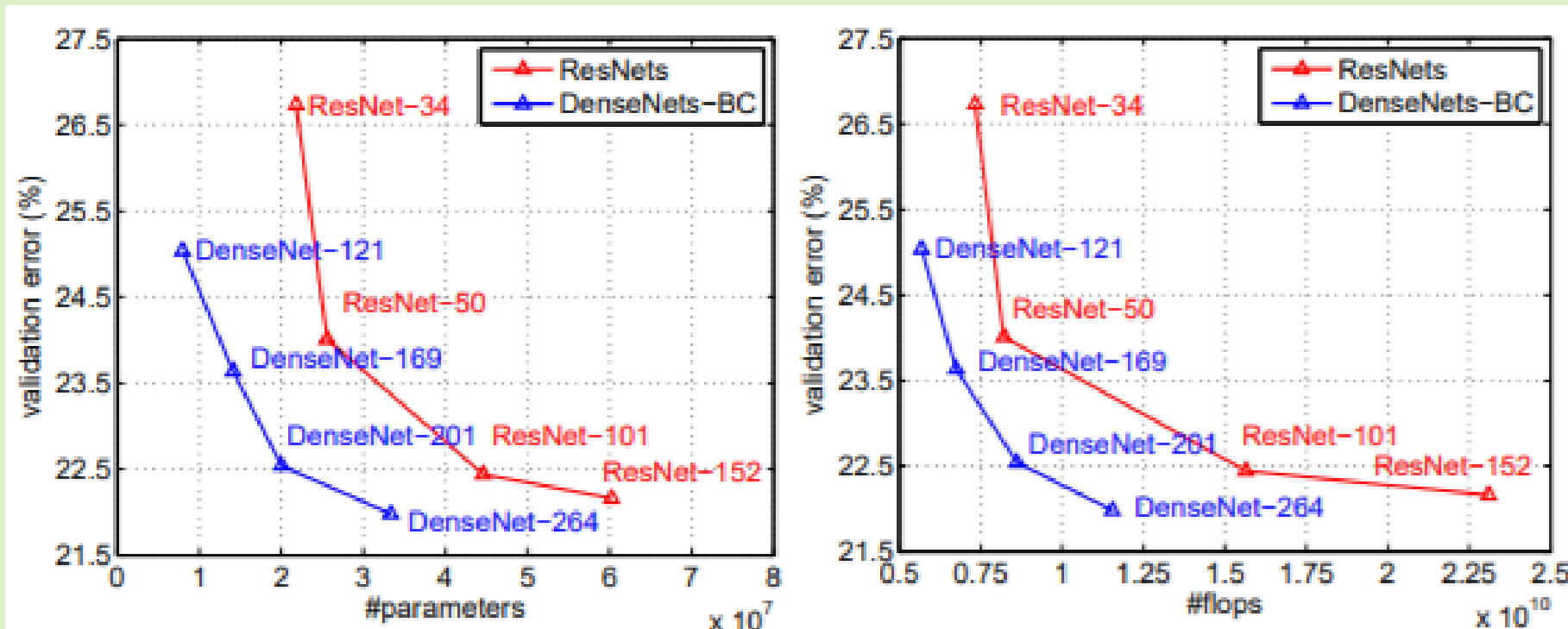
DenseNet-BCは同等の精度を達成するためにResNetとの約1/3のパラメータが必要

右：10M以上のパラメータを持つ1001層の（preactivation）ResNetと、

わずか0.8Mのパラメータを持つ100層のDenseNetのトレーニングとテスト曲線が同等

What is ~...??:

Figure :: Comparison of the DenseNets and ResNets top-1 error rates



Model	top-1	top-5
DenseNet-121	25.02 / 23.61	7.71 / 6.66
DenseNet-169	23.80 / 22.08	6.85 / 5.92
DenseNet-201	22.58 / 21.46	6.34 / 5.54
DenseNet-264	22.15 / 20.80	6.12 / 5.29

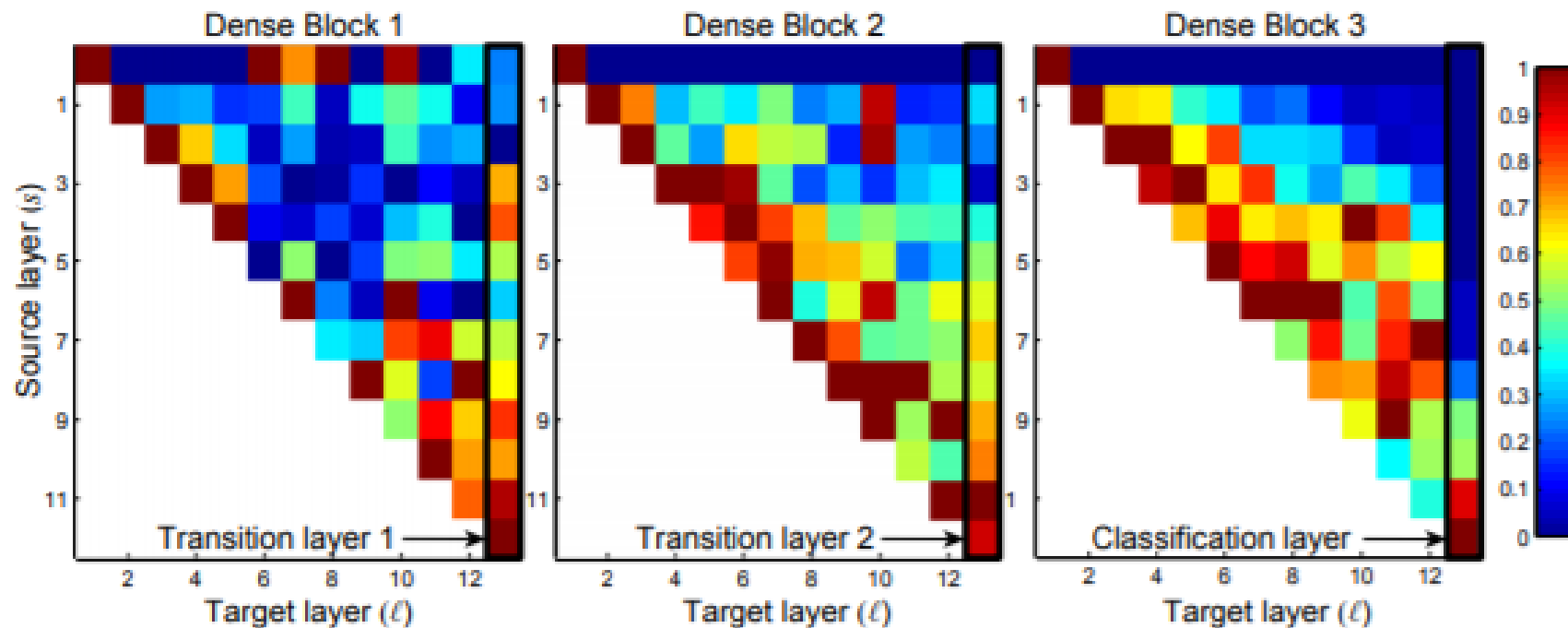
パラメーター数（左）とFLOP（右・損失）の関数としてのDenseNetsとResNetsのTop1の検証エラー

DenseNetが最先端のResNetと同等の性能を達成するために必要なパラメータと計算が大幅に少なくなる

DenseNetの層ごとのtop1とtop5のerror率

What is ~...??:

Figure :: The average absolute filter weights of convolutional layers in a trained DenseNet



1 すべてのレイヤーが同じブロック内の多数の入力に重みをかける
これは、早い層によって抽出された特徴が、同じDenseblock全体の深い層によって直接使用されることを示している

2 transitionlayerの重みもまた、前のDenseblock内のすべての層にわたって重みを分散させ、情報がDenseNetの最初の層から最後の層まで流れることを示している

3 第2および第3のDenseblock内の層はtransitionlayerの出力（三角形の最上列）に一貫して最小の重みを割り当てる (transitionlayerは多くの冗長な特徴を出力することを示している)

これは、DenseNet-BCがよいスコアとなる結果と一致する

4 Dense Block3に示されている最終的な分類層(Classificationlayer)もまたdenseblock全体にわたって重みを使用しているが、最終的な(図の下側の)特徴マップに集中している

Conclusion:

Problem>>Farmer Work>>

勾配消失問題に対して有効な方法を示したILSVRC2015 winner ResNetを改良したDenseNetはパラメータが減って精度も上がった

What is this thesis for?

CNNにおける層が深いモデルの成功例であるResNetを改良したもの

How to verified whether it is valid?

CIFAR. , ImageNet. ,SVHN.という有名なデータセットにおけるResNetとの精度の比較

Where is an important point compared to previous researches?

DenseNetはResNetよりも画像分類の精度を大幅に向上しパラメータも減少させた

Is there discussions?

すいません。わかりません

Where are the key points of technology and method?

ResNetの入力と学習の**合計**（正しくは求めたい関数-入力の差分を学習する）からDense Blockの**積み重ねの連結**に置き換えた

Which reserches should I read next?

Identity mappings in deep residual networks. In ECCV, 2016