

**深層学習と
その画像認識への応用に関する
最近の研究動向**

岡谷貴之

東北大学情報科学研究科, JST CREST

目次

- **最近までの状況：画像認識の革命**
- **いろいろなディープネット**
- **ディープネットの学習**
 - **確率勾配降下法とその発展**
 - **重みの圧縮**
 - **転移学習**
- **ディープネットの理解**
 - **多層の意義**
 - **実験的解析**
 - **生物との対応**
- **まとめ**

目次

- **最近までの状況：画像認識の革命**
- **いろいろなディープネット**
- **ディープネットの学習**
 - **確率勾配降下法とその発展**
 - **重みの圧縮**
 - **転移学習**
- **ディープネットの理解**
 - **多層の意義**
 - **実験的解析**
 - **生物との対応**
- **まとめ**

「物体カテゴリ認識」



'lion'



⇒ **'lion'**



⇒ **'table lamp'**



⇒ **'acoustic guitar'**



⇒ **'Blenheim spaniel'**



⇒ **'electric guitar'**



⇒ **'Japanese spaniel'**



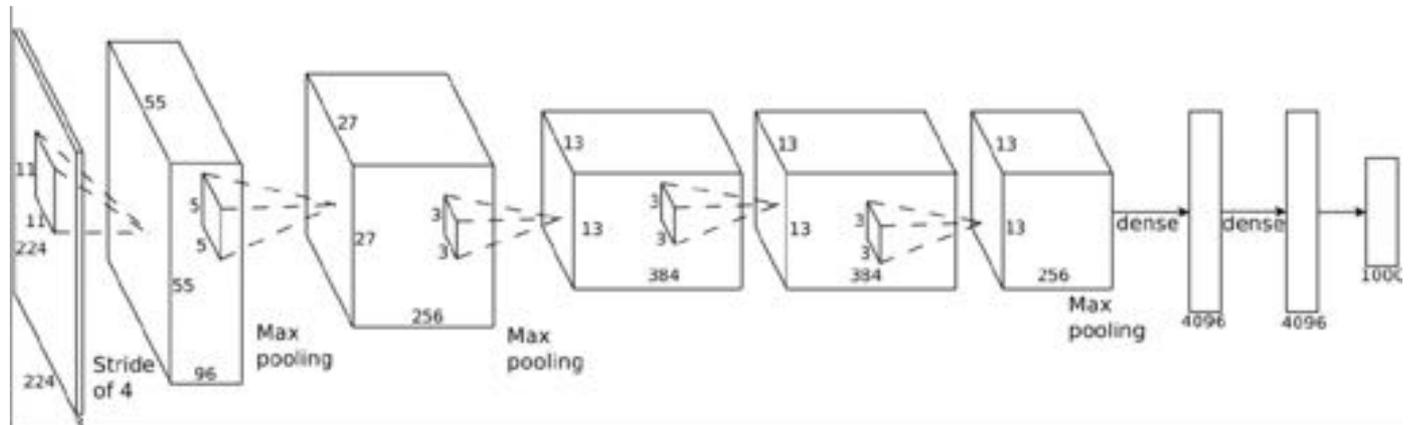
⇒ **'chambered nautilus'**



⇒ **'crane'**

CNN：畳込みニューラルネット

- **Convolutional Neural Network**
 - 畳込み層とプーリング層の交互反復＋全結合層を持つフィードフォワードネット
 - 重みをランダムに初期化し，教師あり学習



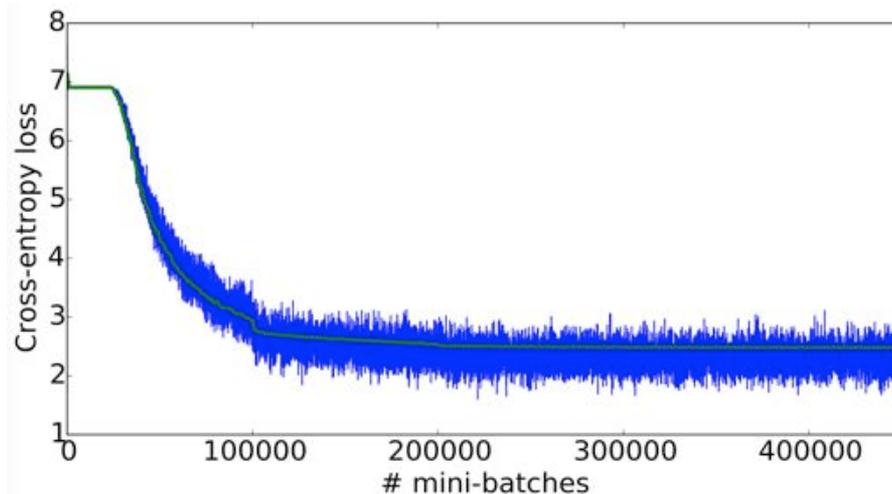
ILSVRC12のCNN [Krizhevsky+12]

“Alexnet”

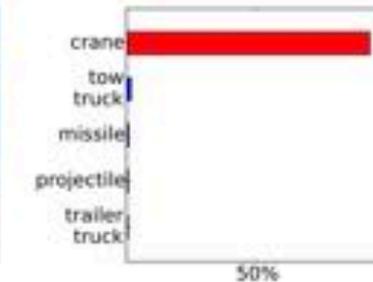
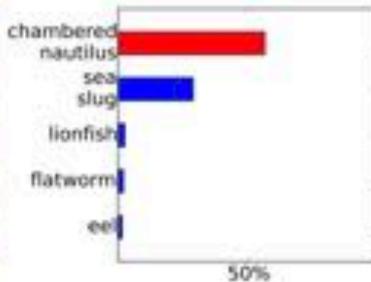
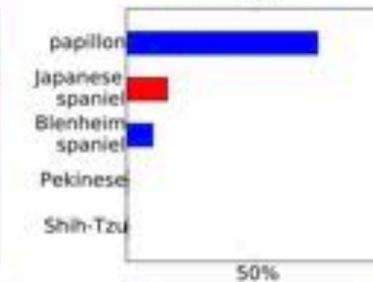
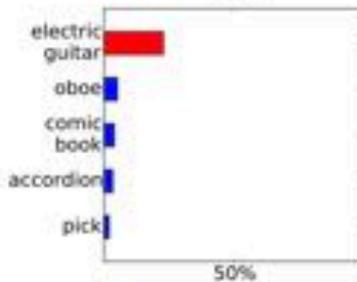
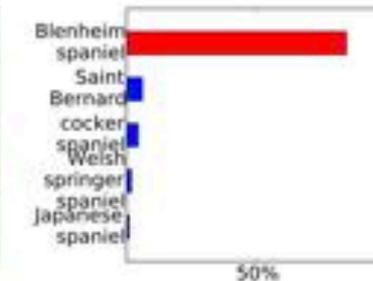
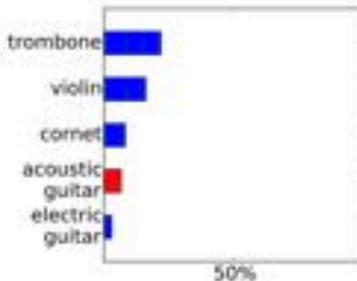
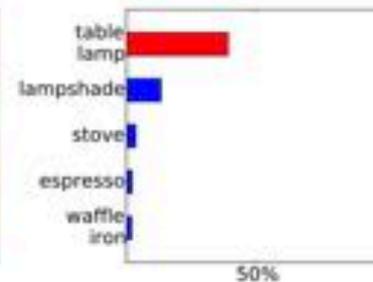
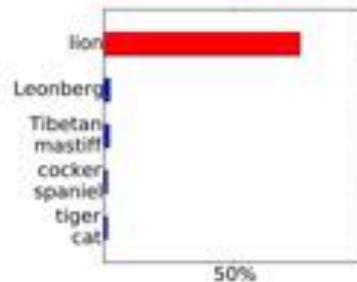
大規模な学習

- **100万枚を超える学習サンプル (画像・クラスラベル)**
 - 1000の物体カテゴリ・1000枚/カテゴリ
- **最新のGPUを用いて数日～数週間かけて学習**
- **人を超える認識精度**
(He+, Delving deep into rectifier, 2015)

The screenshot shows the ImageNet website interface. At the top, there is a search bar with the text "IMAGENET" and a search button. Below the search bar, there are navigation links for "Home", "About", "Explore", and "Download". The main content area displays the search results for "Car, auto, automobile, machine, motorcar". It includes a description: "A motor vehicle with four wheels; usually propelled by an internal combustion engine; *he needs a car to get to work*". There are also statistics: "1307 pictures", "92.48% Popularity Percentile", and "Wordnet E5". The interface features a "Treemap Visualization" and "Images of the Synset" section, which shows a grid of small images categorized by sub-classes like "Racer", "Ambulance", "Cab", "Jeep", "Minivan", "Convertible", "Modes", "Hardtop", "Touring", "Stanley", "Beach", "Horseless", "Subcompact", "Compact", "Pick", "Electric", "Sport", "Minicar", "Hatchback", "Hot", "Sports", "Coupe", "Limousine", "Sedan", "Gas", "Roadster", "Cruiser", "Used-car", and "Limousine".



実例

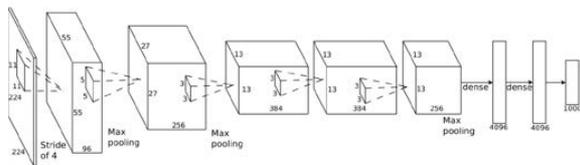


2012-2015の発展

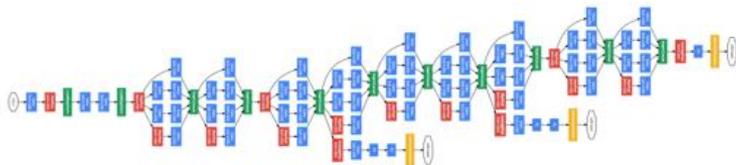
表1 代表的なモデルのパラメータ数および演算回数、畳込み層および全結合層での合計と総計。

モデル		Alexnet	VGGNet	GoogLeNet	ResNet
畳込み	層	5	13	21	151
	重み	380 万	0.15 億	580 万	-
	演算	10.8 億	153 億	15 億	113 億
全結合	層	3	3	1	1
	重み	0.59 億	1.24 億	100 万	200 万
	演算	0.59 億	1.24 億	100 万	200 万
合計	重み	0.62 億	1.38 億	680 万	-
	演算	11.4 億	155 億	15 億	113 億

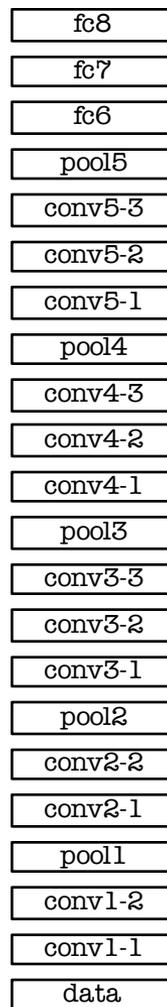
AlexNet [Krizhevsky+12]



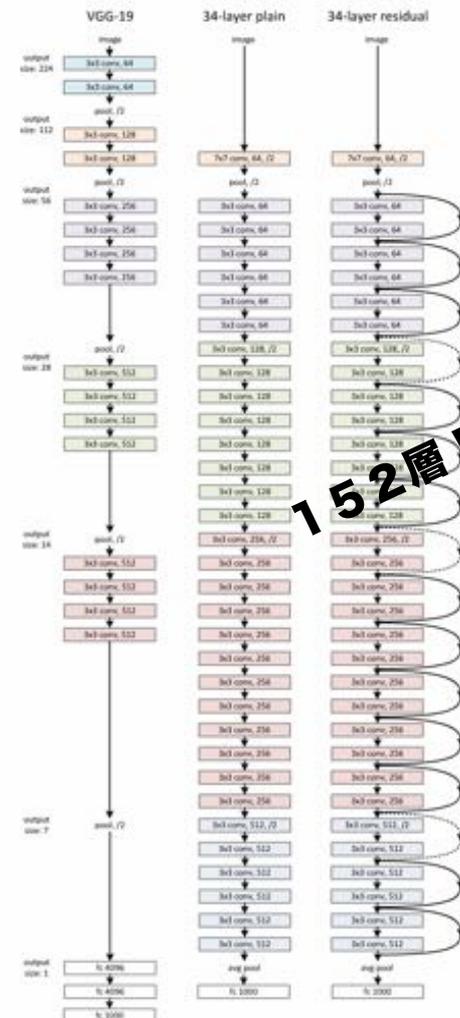
GoogLeNet [Szegedy+14]



VGGNet [Simonyan+14]



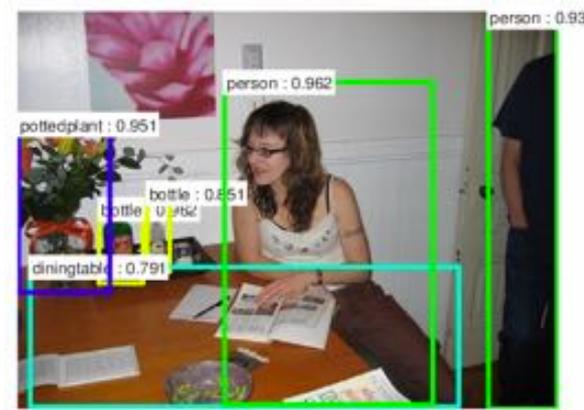
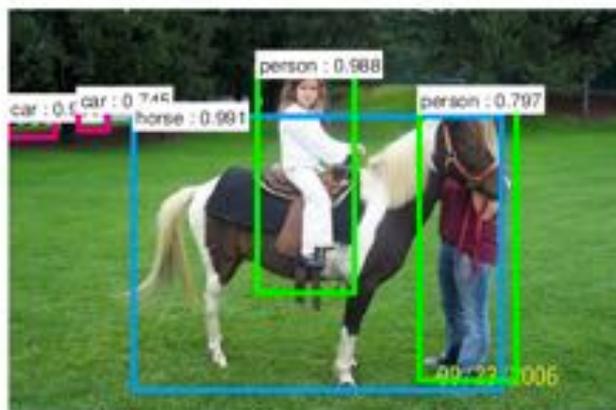
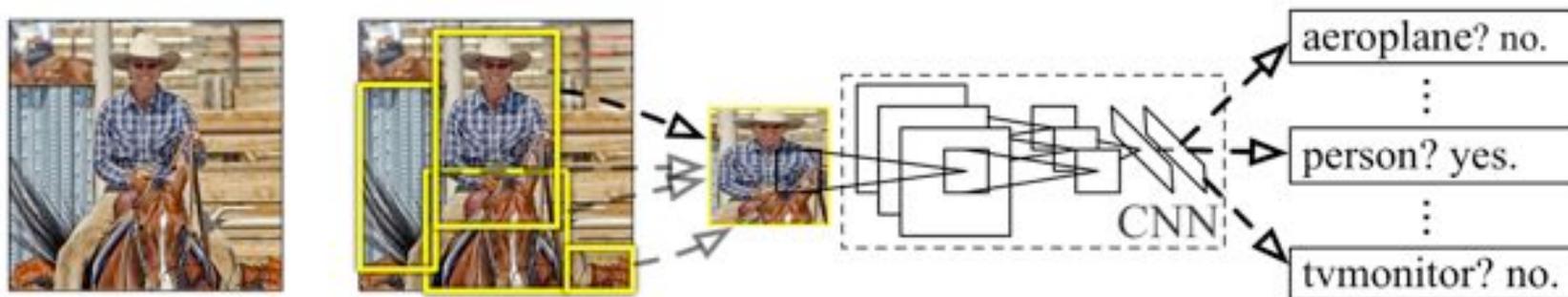
ResNet [He+15]



物体検出 (R-CNN)

Girshick et al., Rich feature hierarchies for accurate object detection and semantic segmentation, CVPR2014

- 物体の種類と位置を認識：物体認識のCNNを流用



Faster RCNN [Ren+15]

ポーズ認識

Toshev, Szegedy, DeepPose: Human Pose Estimation via Deep Neural Networks, CVPR2014

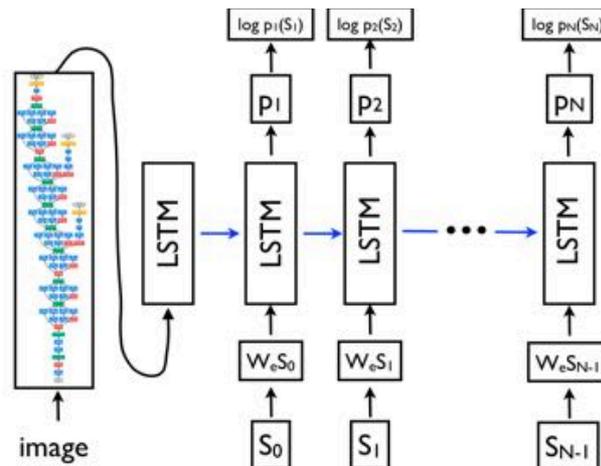
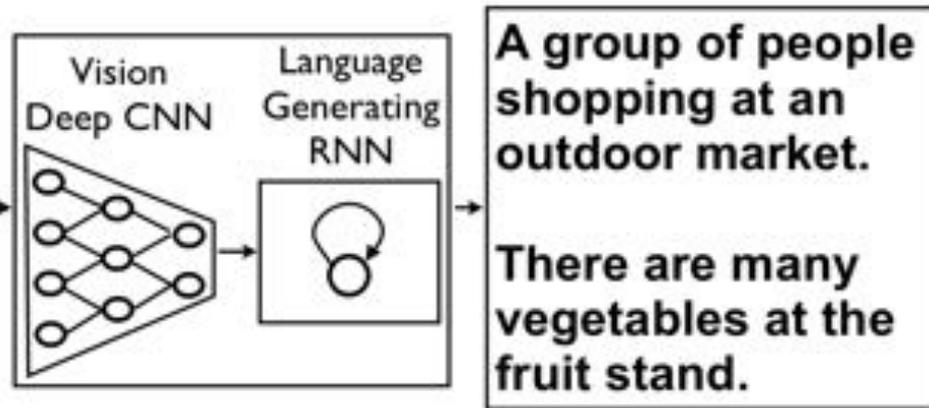


- 回帰問題として定式化
(画像から骨格の姿勢を予測)
- 物体認識のCNNを流用



画像記述

Vinyals+, Show and Tell: A Neural Image Caption Generator, CVPR2015



画像記述

Vinyals+, Show and Tell: A Neural Image Caption Generator, CVPR2015

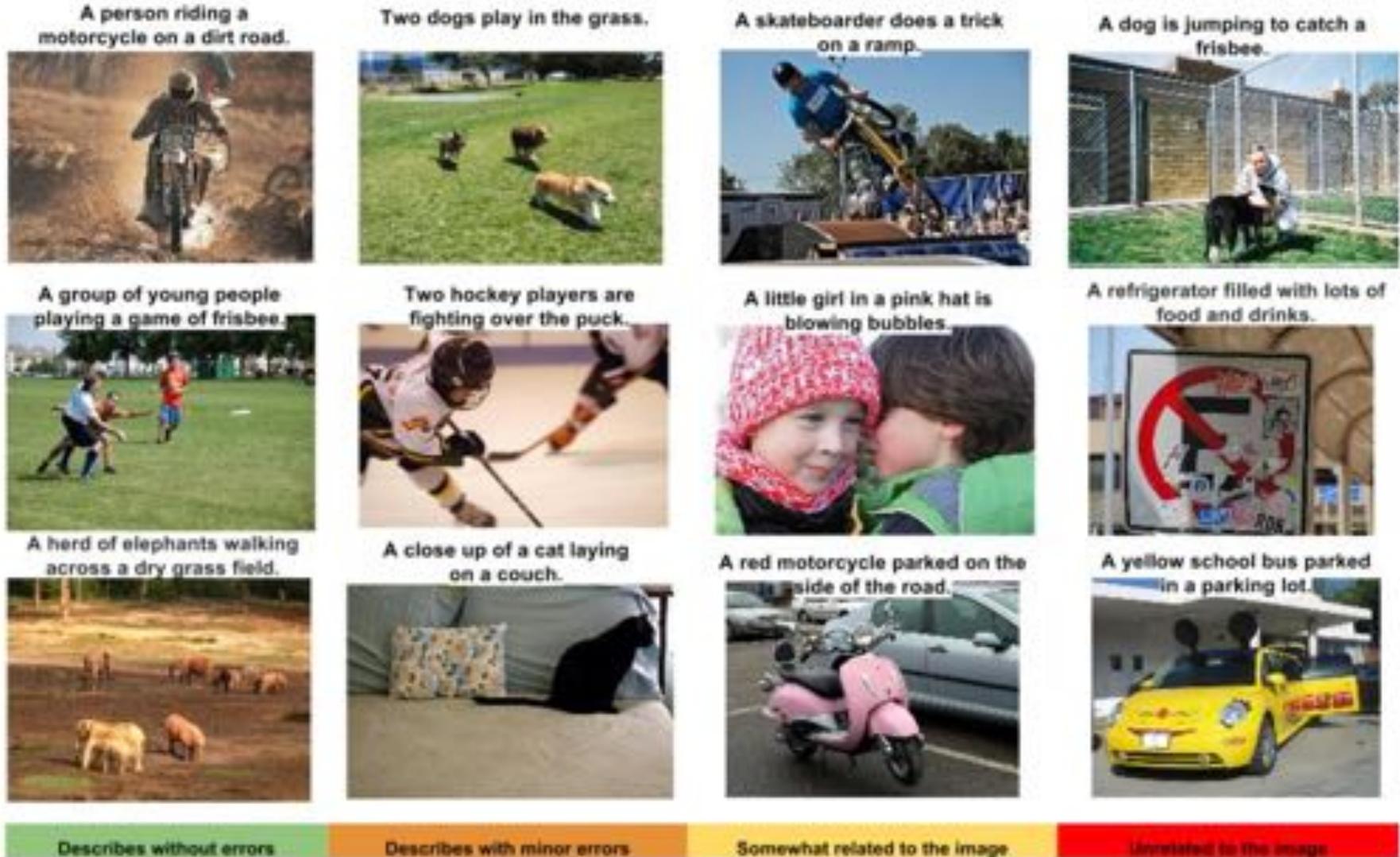
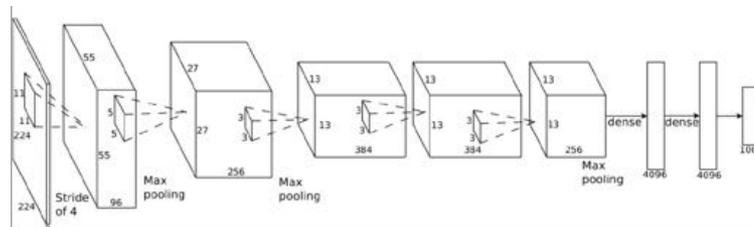
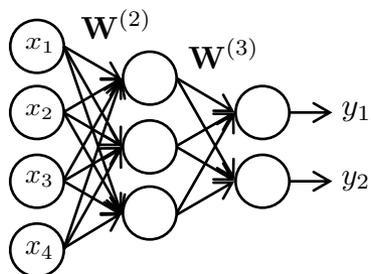


Figure 5. A selection of evaluation results, grouped by human rating.

目次

- **最近までの状況：画像認識の革命**
- **いろいろなディープネット**
- **ディープネットの学習**
 - **確率勾配降下法とその発展**
 - **重みの圧縮**
 - **転移学習**
- **ディープネットの理解**
 - **多層の意義**
 - **実験的解析**
 - **生物との対応**
- **まとめ**

様々なニューラルネットワーク



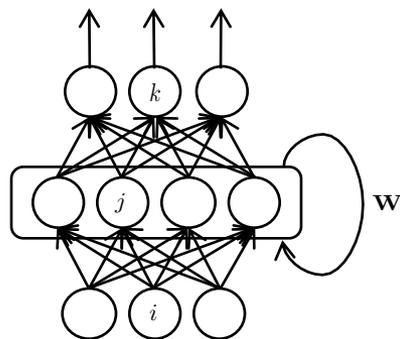
①②フィードフォワードネットワーク

④たたみこみネットワーク(CNN)

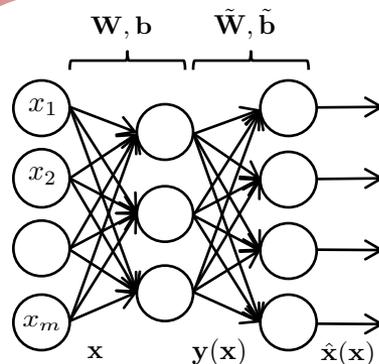
教師あり

静的

動的



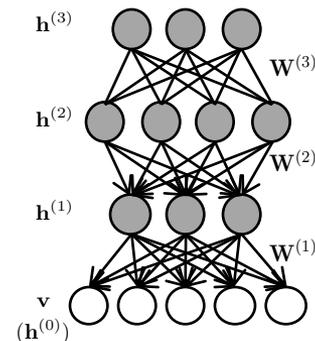
⑤リカレントネットワーク (RNN)



③オートエンコーダ (自己符号化器)

決定論的

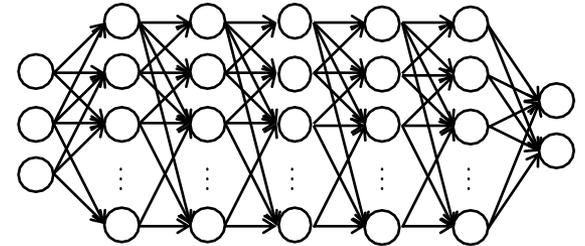
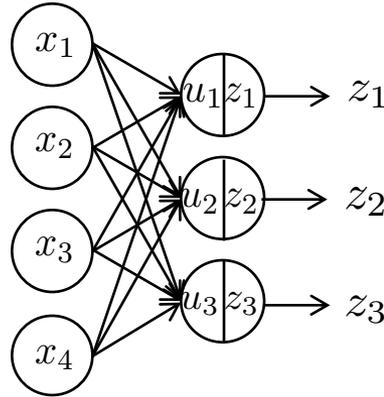
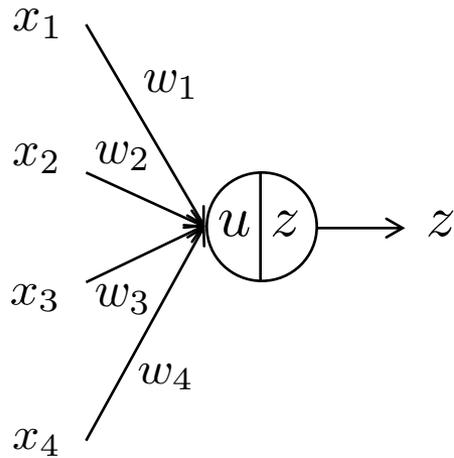
教師なし



⑥ボルツマンマシン

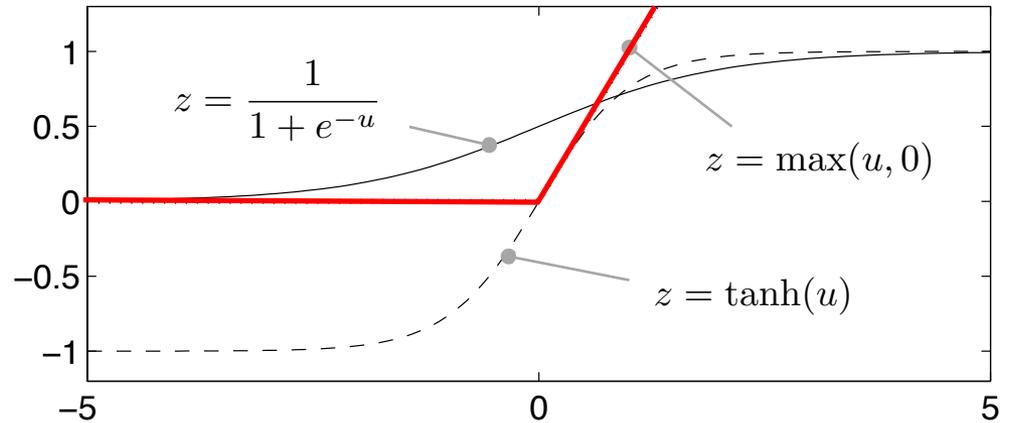
確率的

フィードフォワードネットワーク



$$u_j = \sum_{i=1}^I w_{ji} x_i + b_j$$

$$z_j = f(u_j)$$



CNN：畳込みニューラルネット

Convolutional Neural Network

- **Neocognitronにルーツ** [Fukushima80]
- **LeNet：手書き文字認識への応用で成功** [LeCun+89]
 - Backpropagation Applied to Handwritten Zip Code Recognition, 1989
- **神経科学の知見が基礎**
 - Hubel-Wiesel の単純細胞・複雑細胞
 - 局所受容野 (local receptive field)

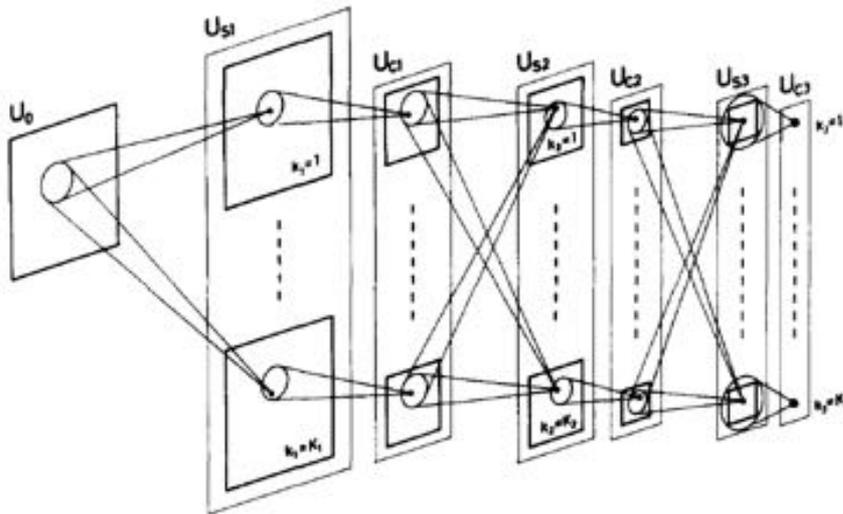


Fig 4 Schematic diagram illustrating the interconnections between layers in the neocognitron

[Fukushima+83]

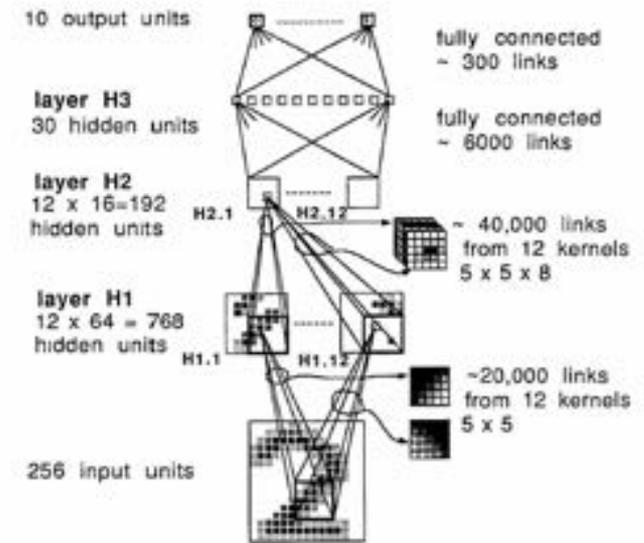
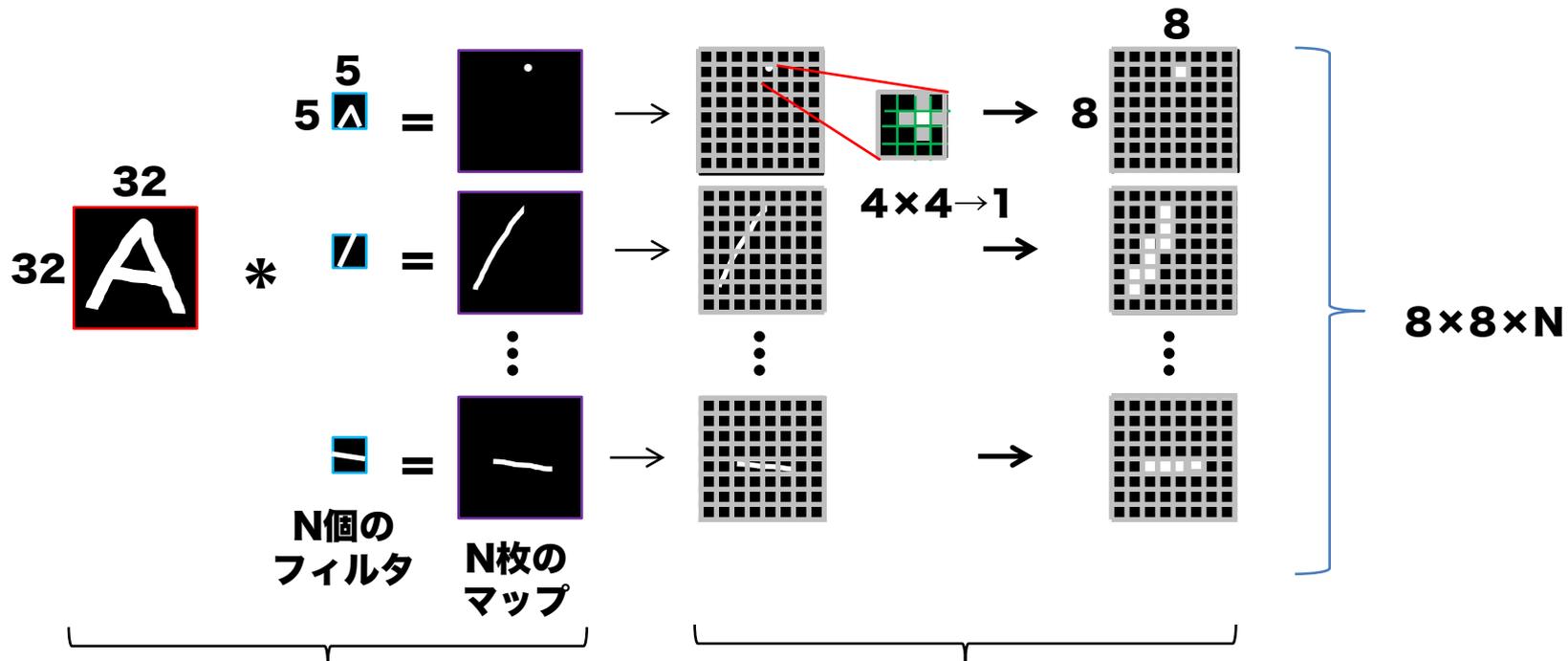


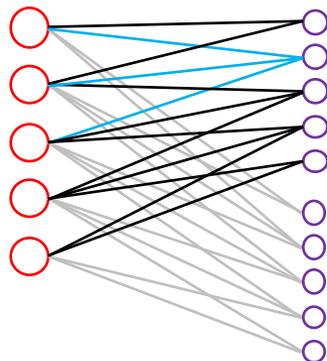
Figure 3 Log mean squared error (MSE) (top) and raw error rate (bottom) versus number of training passes

[LeCun+89]

CNNの二つの演算：畳込みとプーリング

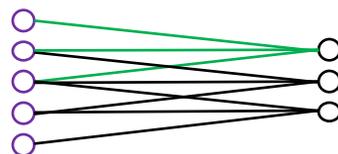


畳込み (層)



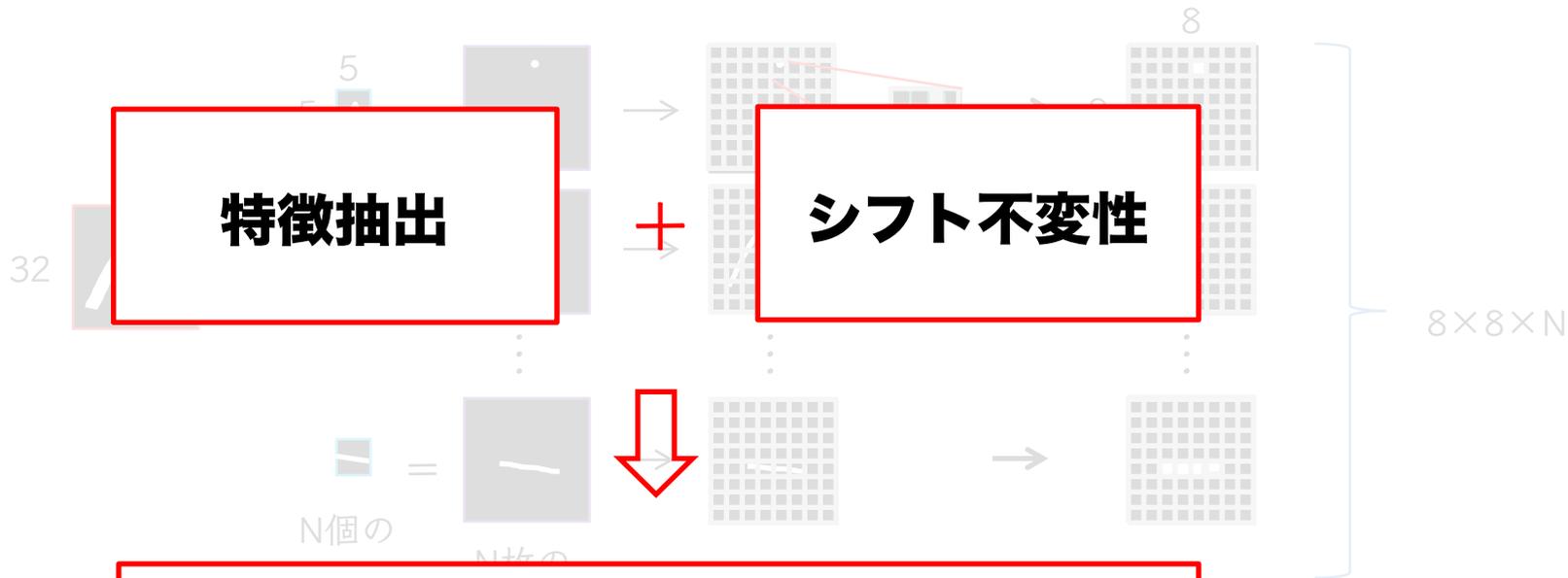
- 共有重み
- 疎結合

プーリング (層)



- 固定配線 (重み)
- 疎結合

二つの演算：畳込みとプーリング



微小変位に対する不変性を獲得

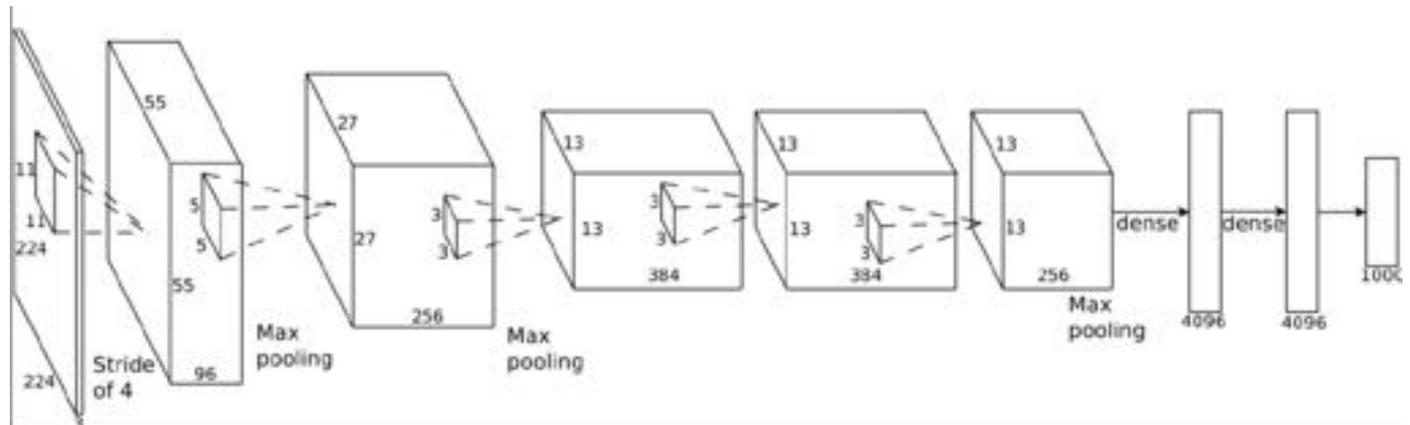


- 固定配線 (重み)
- 疎結合



CNN：畳込みニューラルネット

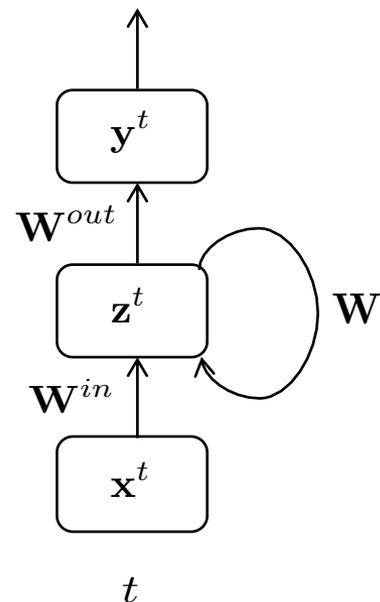
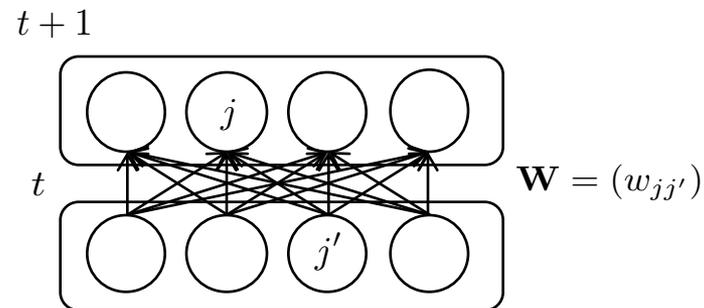
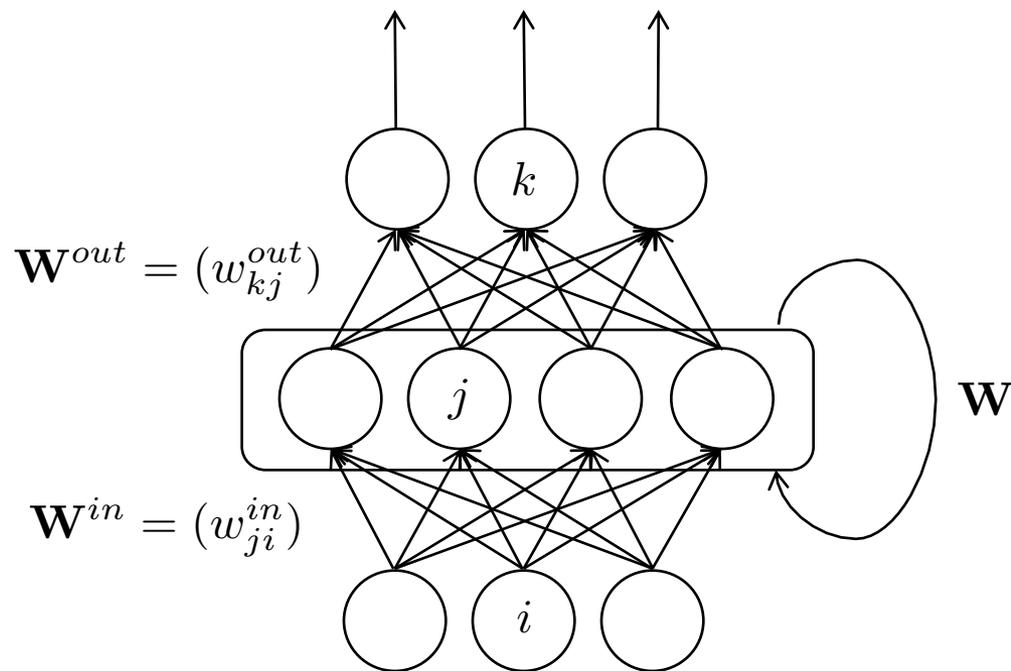
- **Convolutional Neural Network**
 - 畳込み層とプーリング層の交互反復＋全結合層を持つフィードフォワードネット
 - 重みをランダムに初期化し，教師あり学習



ILSVRC12のCNN [Krizhevsky+12]

“Alexnet”

リカレントニューラルネット (RNN)



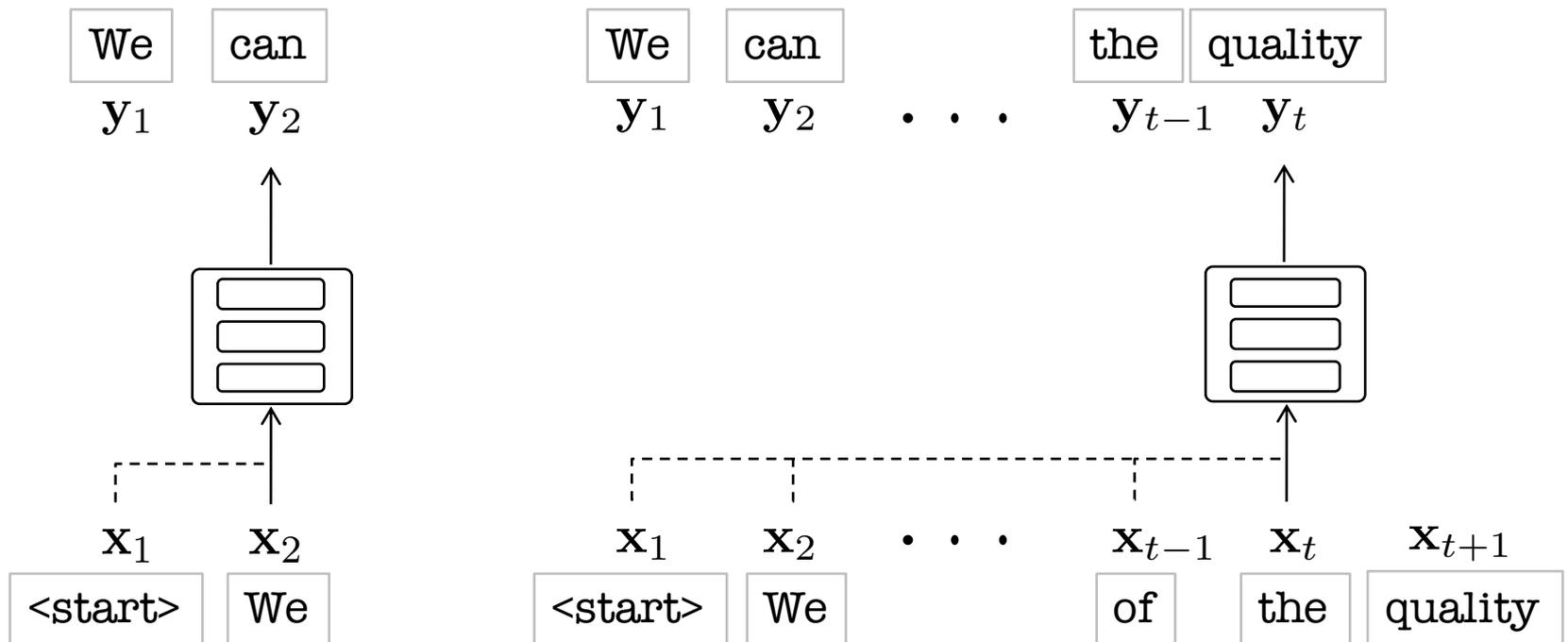
系列データへの適用

例：言語モデル

次の単語を予測

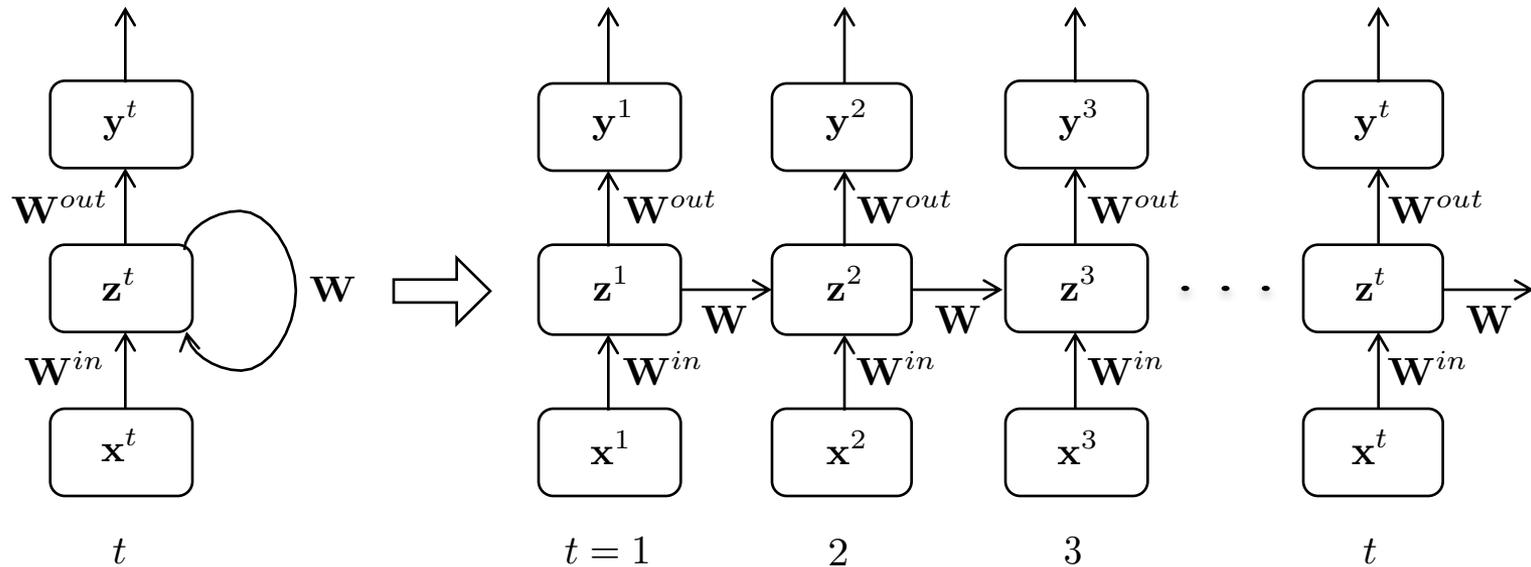


We can get an idea of the quality of the learned feature vectors by displaying them in a 2-D map.



RNN=ディープフィードフォワードネット

- 時間方向に展開 → フィードフォワードネットになる

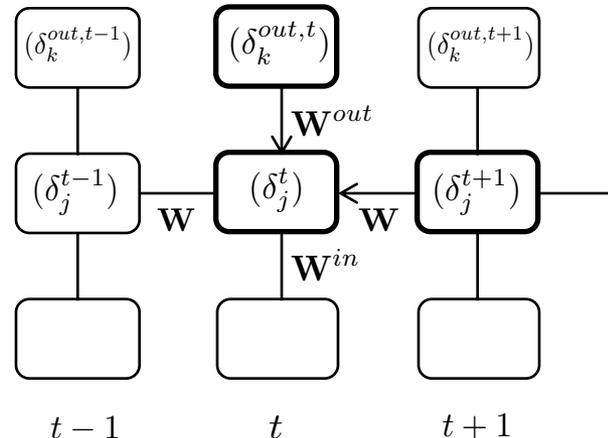


BPが使える：

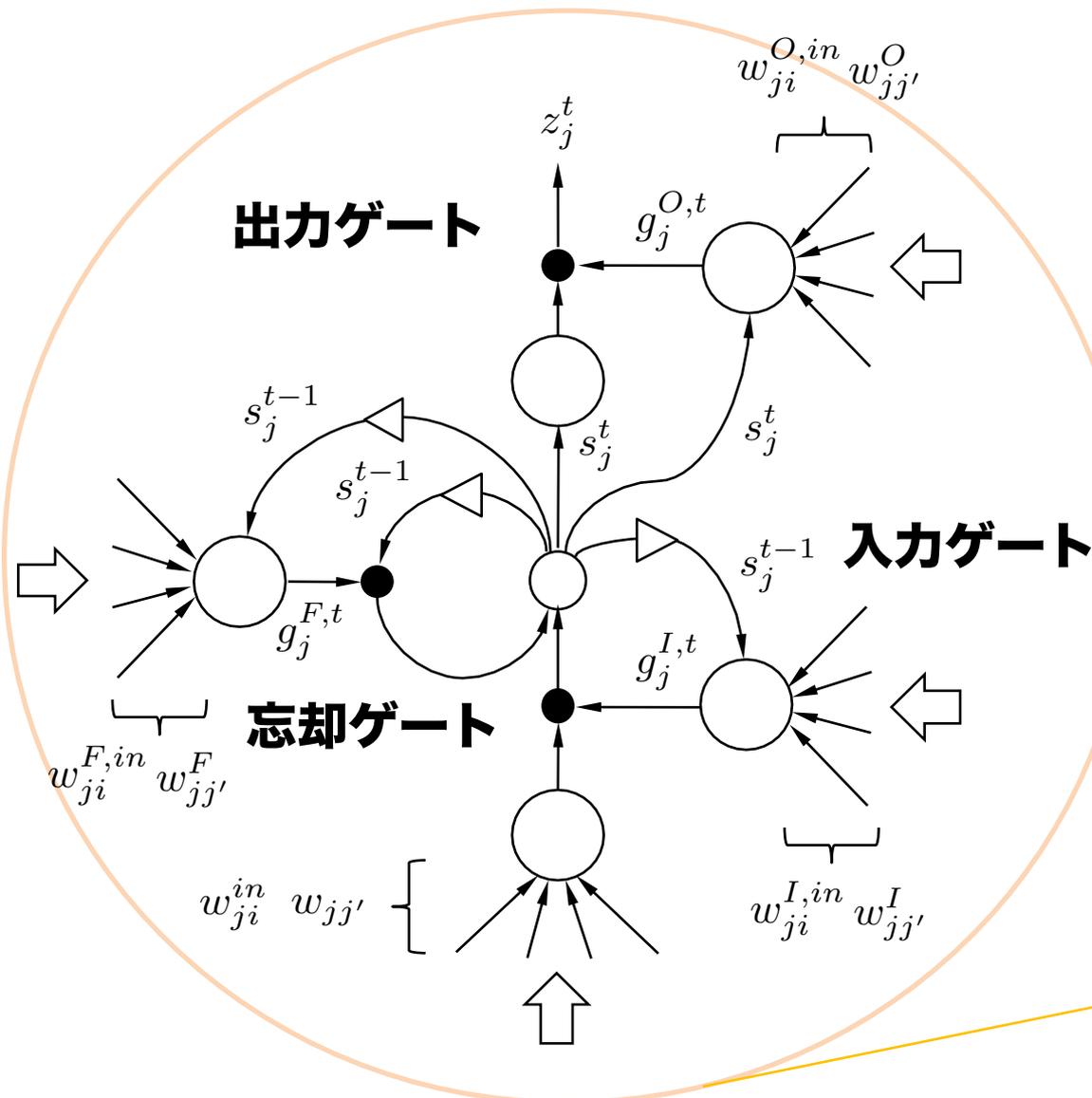
実はディープネット

→ **勾配消失問題との戦い**

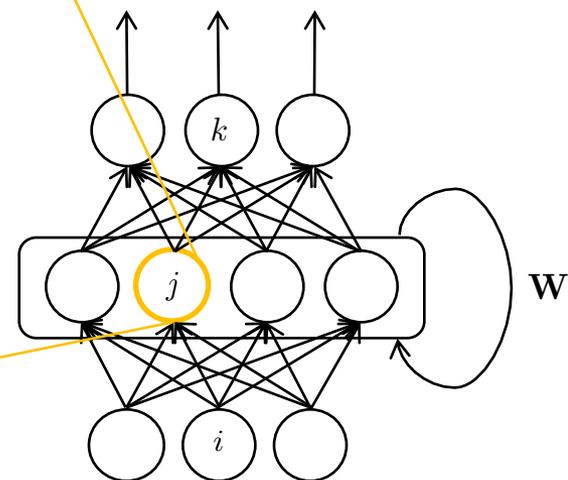
→ **高々10ステップ程度の履歴しか学習できない**



LSTM : Long Short-Term Memory

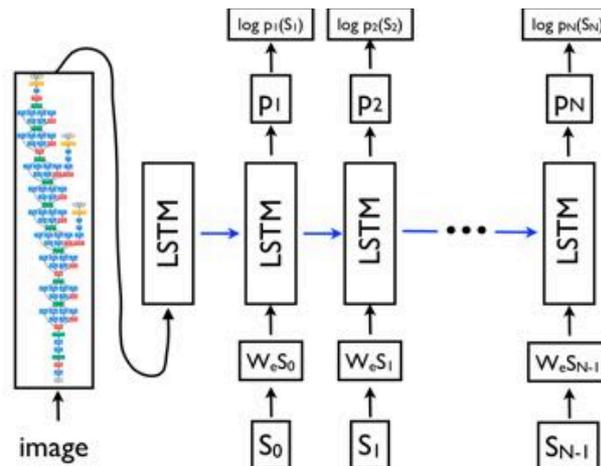
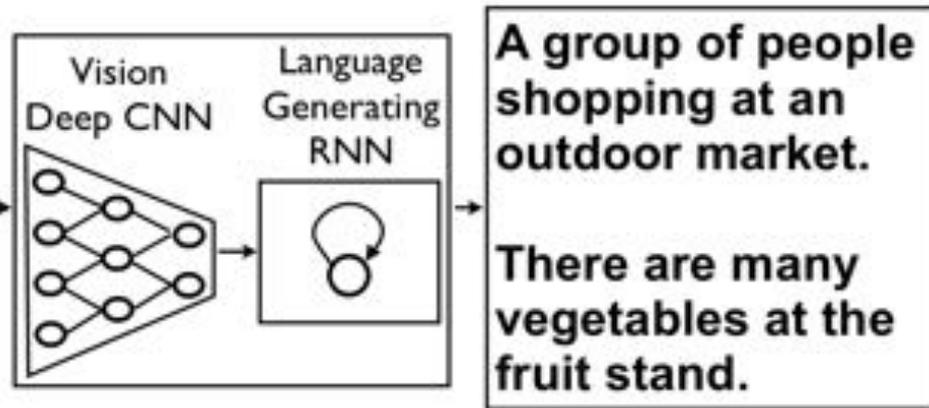


- 3つのゲートにより高度な動作
- より長い履歴を捉える
- BPで勾配は計算可 (ややこしい!)



画像記述

Vinyals+, Show and Tell: A Neural Image Caption Generator, CVPR2015

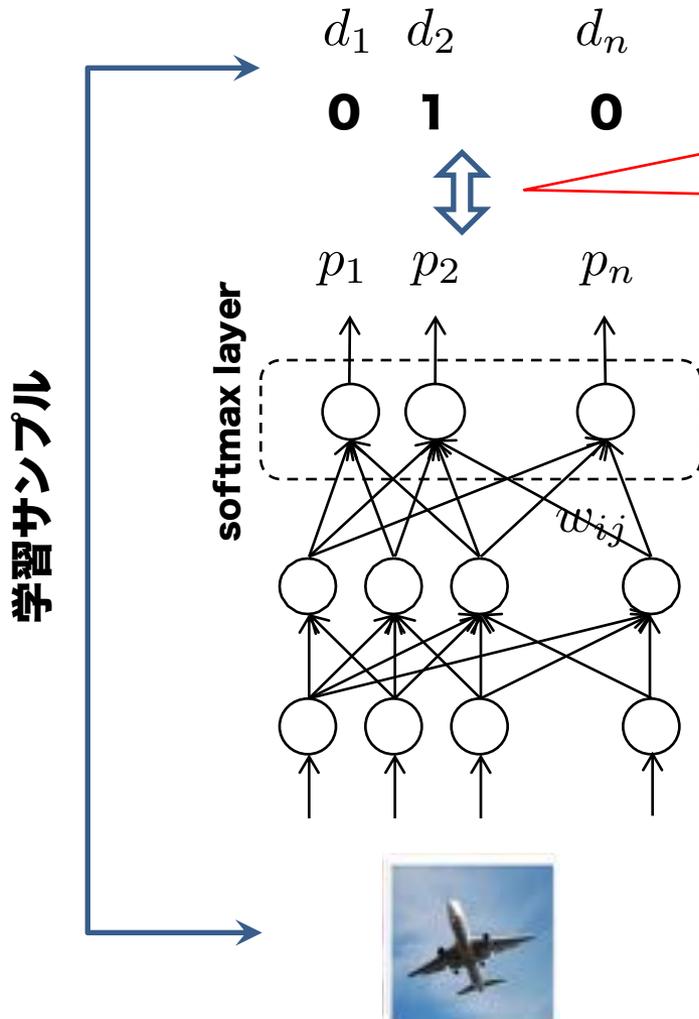


目次

- **最近までの状況：画像認識の革命**
- **いろいろなディープネット**
- **ディープネットの学習**
 - **確率勾配降下法とその発展**
 - **重みの圧縮**
 - **転移学習**
- **ディープネットの理解**
 - **多層の意義**
 - **実験的解析**
 - **生物との対応**
- **まとめ**

クラス分類の教師あり学習

- 出力の誤差が小さくなるように重み $\{w_{ij}\}$, $\{b_j\}$ を調節



$$C = - \sum_j^n d_j \log p_j$$

交差エントロピー

$$\Delta w_{ij} = \epsilon \frac{\partial C}{\partial w_{ij}}$$

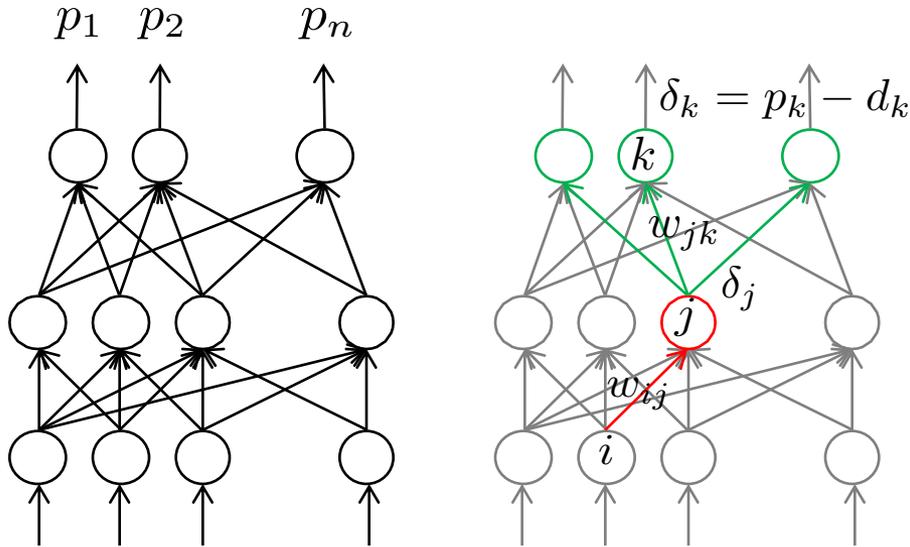
$$+ \alpha \Delta w'_{ij} - \epsilon \lambda w_{ij}$$

モメンタム Weight-decay

勾配降下法 (GD: Gradient Descent)

誤差逆伝播と確率的勾配降下

- 誤差逆伝播法



勾配の連鎖計算

$$\frac{\partial C}{\partial w_{ij}} = \delta_j y_i$$

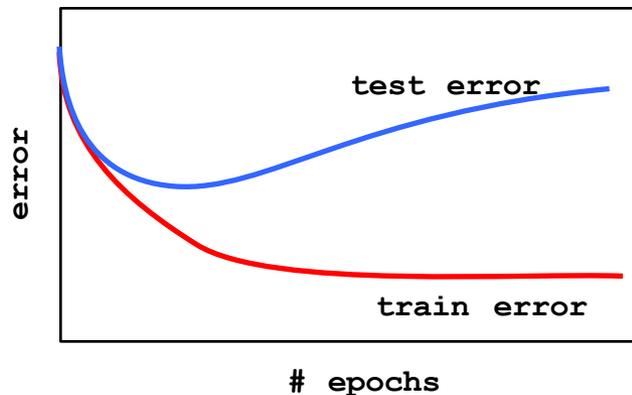
$$\delta_j = f'(x_j) \sum_k \delta_k w_{jk}$$

- 確率的勾配降下法 (SGD: stochastic gradient descent)

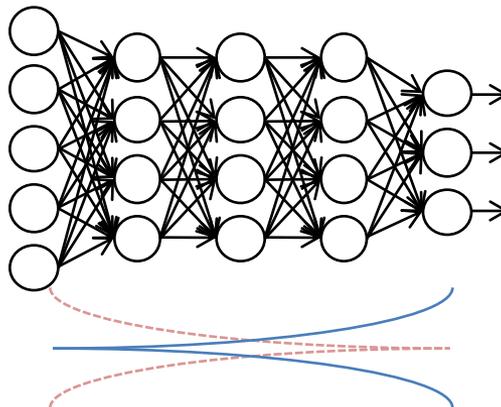
- ミニバッチ：100個程度のサンプルごとに重みを更新

ディープネットの学習の困難さ

- 多層になると過適合（過学習）・ローカルミニマム



- 勾配消失（vanishing gradient）問題
 - デルタが急速に小さく，あるいは大きくなり制御できない



確率的勾配降下法の改良

- 標準的方法

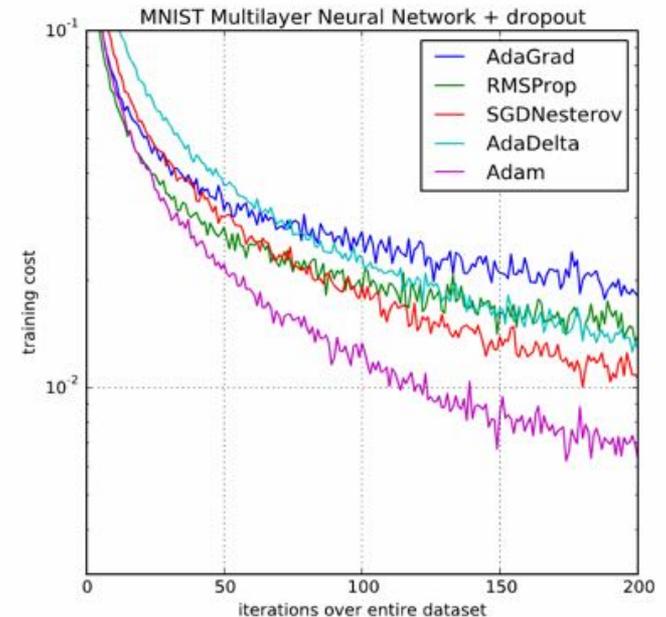
$$\theta_t \leftarrow \theta_{t-1} - \alpha g_t$$

- AdaGrad**
(adaptive gradient) [Duchi+2011]

$$\theta_t = \theta_{t-1} - \alpha g_t / \sqrt{\sum_{i=1}^t g_i^2}$$

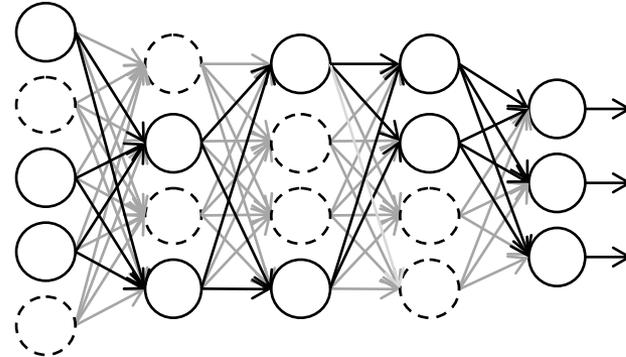
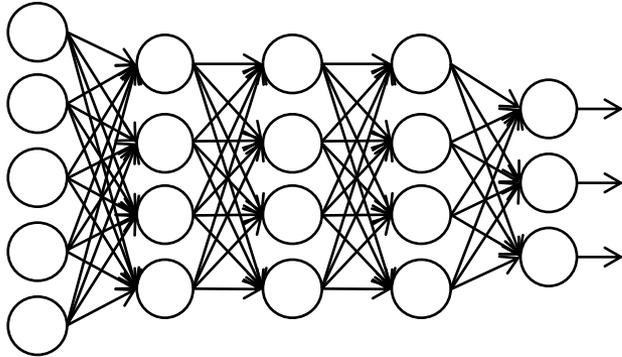
- ADAM**
(adaptive moment) [Kingma-Ba2015]

$$g_t \leftarrow \nabla_{\theta} f_t(\theta_{t-1}) \text{ (Get gradients w.r.t. stochastic objective at timestep } t)$$
$$m_t \leftarrow \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t \text{ (Update biased first moment estimate)}$$
$$v_t \leftarrow \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2 \text{ (Update biased second raw moment estimate)}$$
$$\hat{m}_t \leftarrow m_t / (1 - \beta_1^t) \text{ (Compute bias-corrected first moment estimate)}$$
$$\hat{v}_t \leftarrow v_t / (1 - \beta_2^t) \text{ (Compute bias-corrected second raw moment estimate)}$$
$$\theta_t \leftarrow \theta_{t-1} - \alpha \cdot \hat{m}_t / (\sqrt{\hat{v}_t} + \epsilon) \text{ (Update parameters)}$$

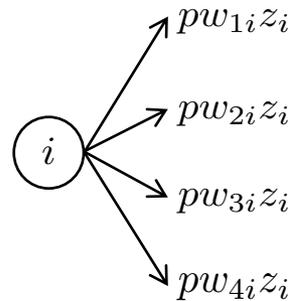
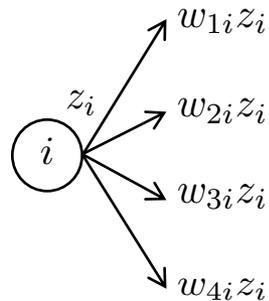


dropout [Hinton+12]

- 学習時一定の確率(p)で中間層のユニットを「無効化」



- 予測時：ユニット出力を p 倍
 - 学習時の補償



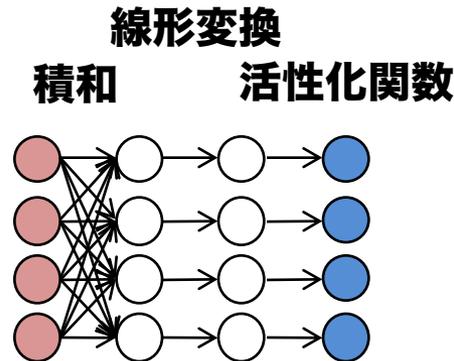
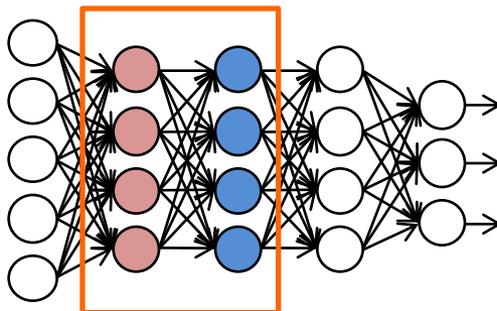
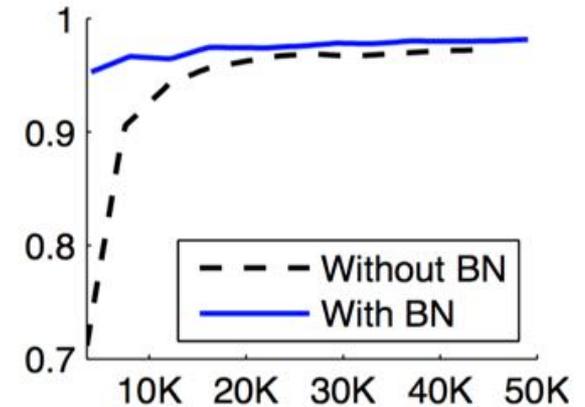
バッチ正規化

- ミニバッチ単位で各層出力の平均と分散を揃える

$$\hat{x}_i \leftarrow \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}$$

$$y_i \leftarrow \gamma \hat{x}_i + \beta \equiv \text{BN}_{\gamma, \beta}(x_i)$$

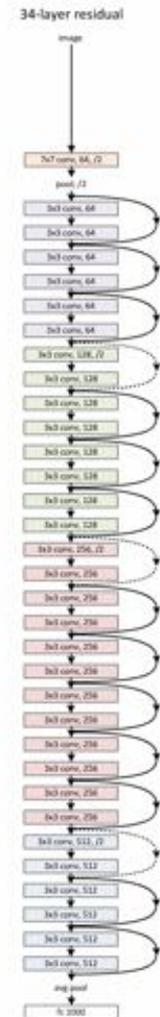
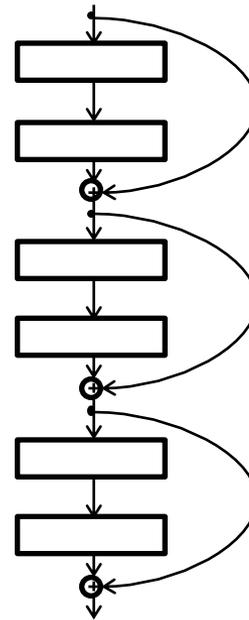
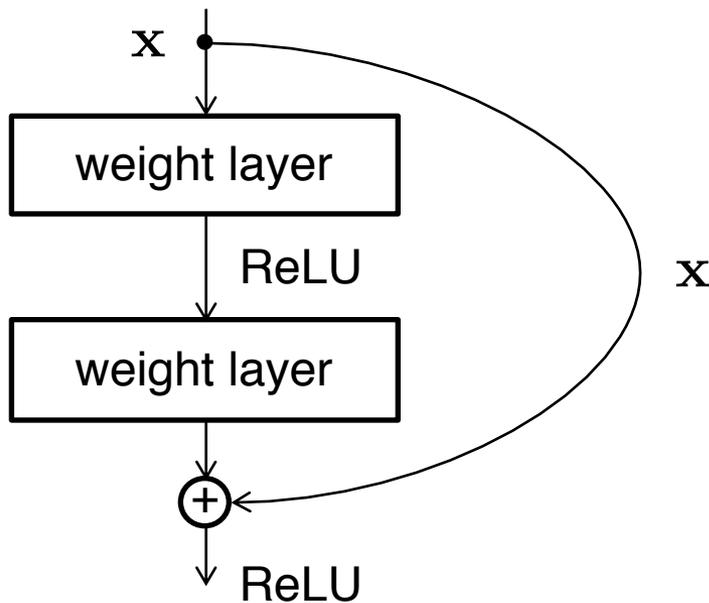
- 収束性能（速度と到達点）が向上
- 線形変換層の挿入で表現可



Deep residual learning

Deep residual learning for image recognition [He+2015]

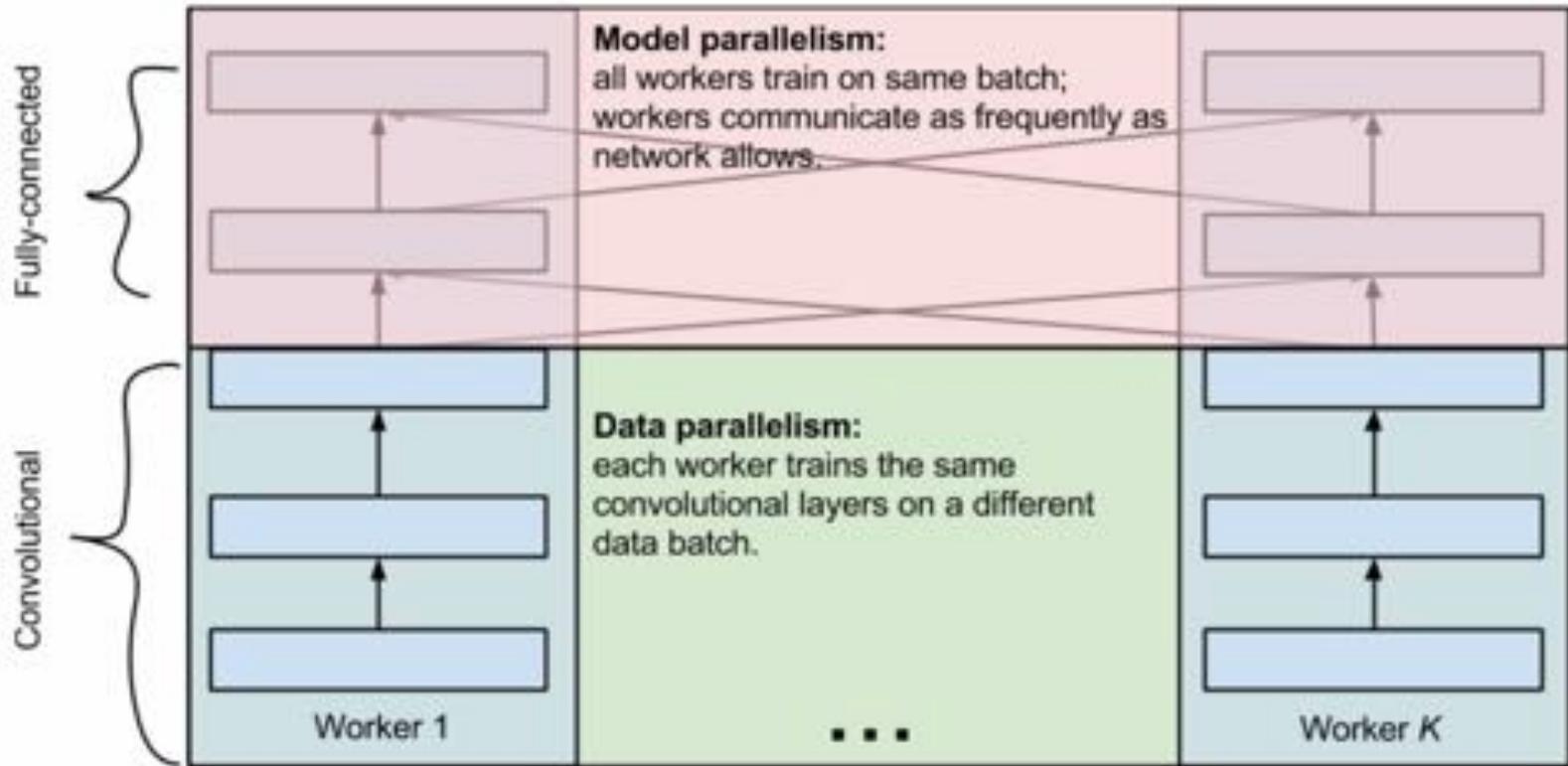
- **100層を超えるネットの学習が可能に**
– **ILSVRC2015チャンピオン**



学習の並列計算

Krizhevsky, One weird trick for parallelizing convolutional neural networks, 2014

- データ並列とモデル並列のミックス
 - データ並列：異なるminibatchを計算
 - モデル並列：同じminibatch・計算を分割



目次

- **最近までの状況：画像認識の革命**
- **いろいろなディープネット**
- **ディープネットの学習**
 - **確率勾配降下法とその発展**
 - **重みの圧縮**
 - **転移学習**
- **ディープネットの理解**
 - **多層の意義**
 - **実験的解析**
 - **生物との対応**
- **まとめ**

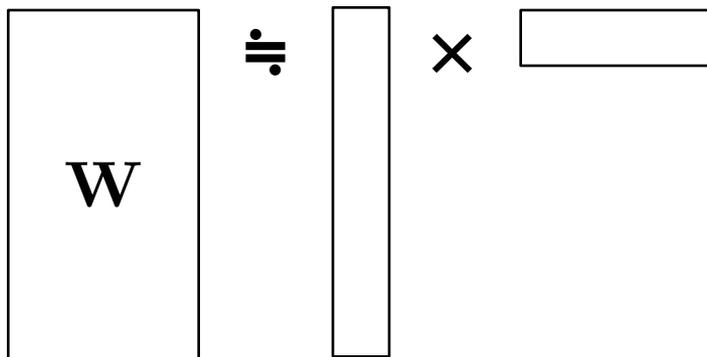
重みの冗長性と圧縮

Denil+, Predicting Parameters in Deep Learning, 2013

- **学習で得られる重みは冗長**
 - 一部から残りを予測可能；一部だけ学習すれば済む



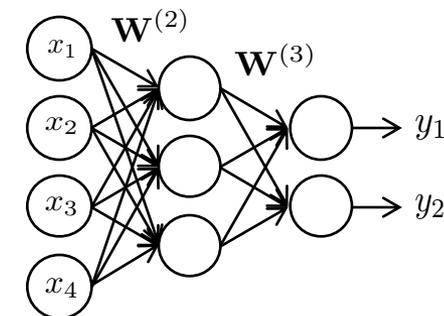
- **学習後のネットの圧縮法が様々に検討されている**
 - [Denton+14] [Ba-Caruaana13] [Collins-Kohli14]



$$\mathbf{u}^{(l+1)} = \mathbf{W}^{(l+1)}\mathbf{z}^{(l)} + \mathbf{b}^{(l+1)}$$

$$\mathbf{z}^{(l+1)} = \mathbf{f}(\mathbf{u}^{(l+1)})$$

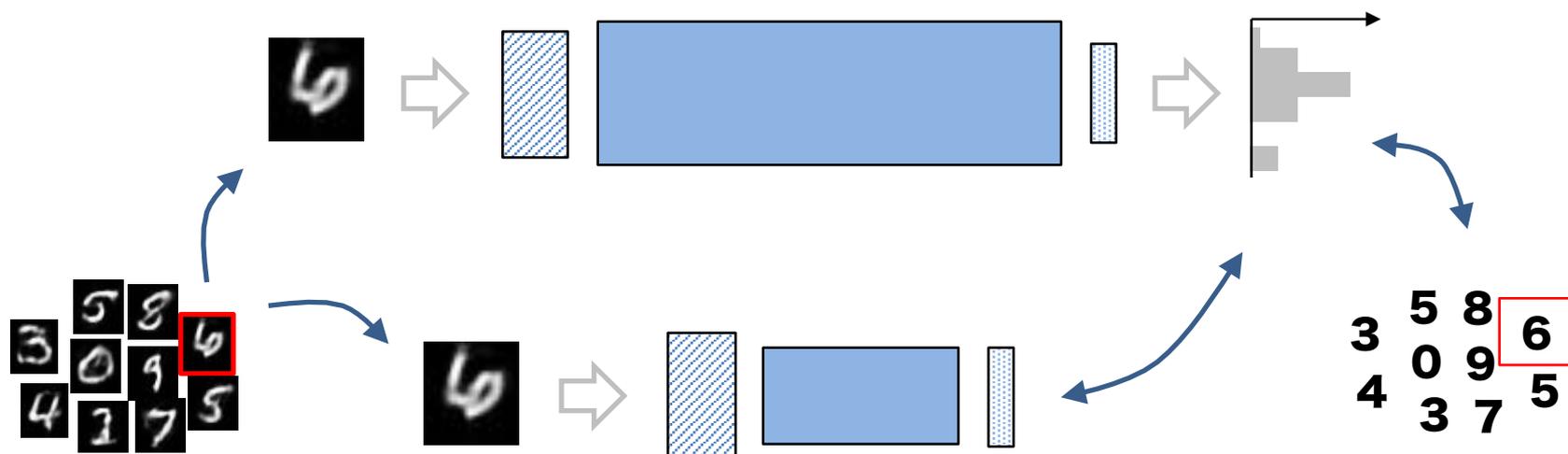
$l = 1 \quad 2 \quad 3$



"Knowledge Distillation"

Hinton+, Distilling the Knowledge in a Neural Network, 2015

- 目標を1-of-k符号ではなく、学習済みの大きいネットの確率（を鈍らせたもの）に



- MNISTの場合

- そのままだと誤り率1.46%の小さいネットを誤り率0.67%の大きなネットの知識を転送 ⇒ 誤り率0.74に減少
- '3'を故意に省いて学習 ⇒ 上の方法でやると1.09%の誤り率（'3'の誤り率は1.4%しかない）

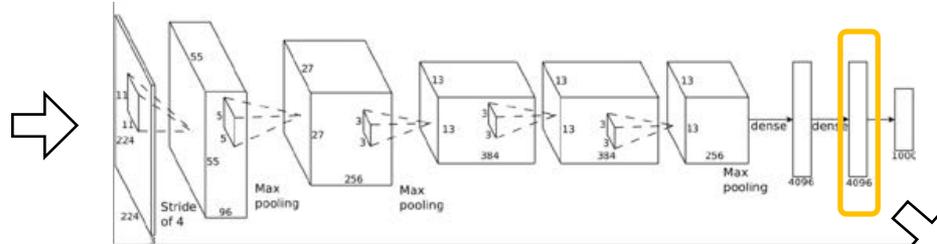
目次

- **最近までの状況：画像認識の革命**
- **いろいろなディープネット**
- **ディープネットの学習**
 - **確率勾配降下法とその発展**
 - **重みの圧縮**
 - **転移学習**
- **ディープネットの理解**
 - **多層の意義**
 - **実験的解析**
 - **生物との対応**
- **まとめ**

転移学習

DeCAF [Denahue+13], CNN features off-the-shelf [Razavian+14]

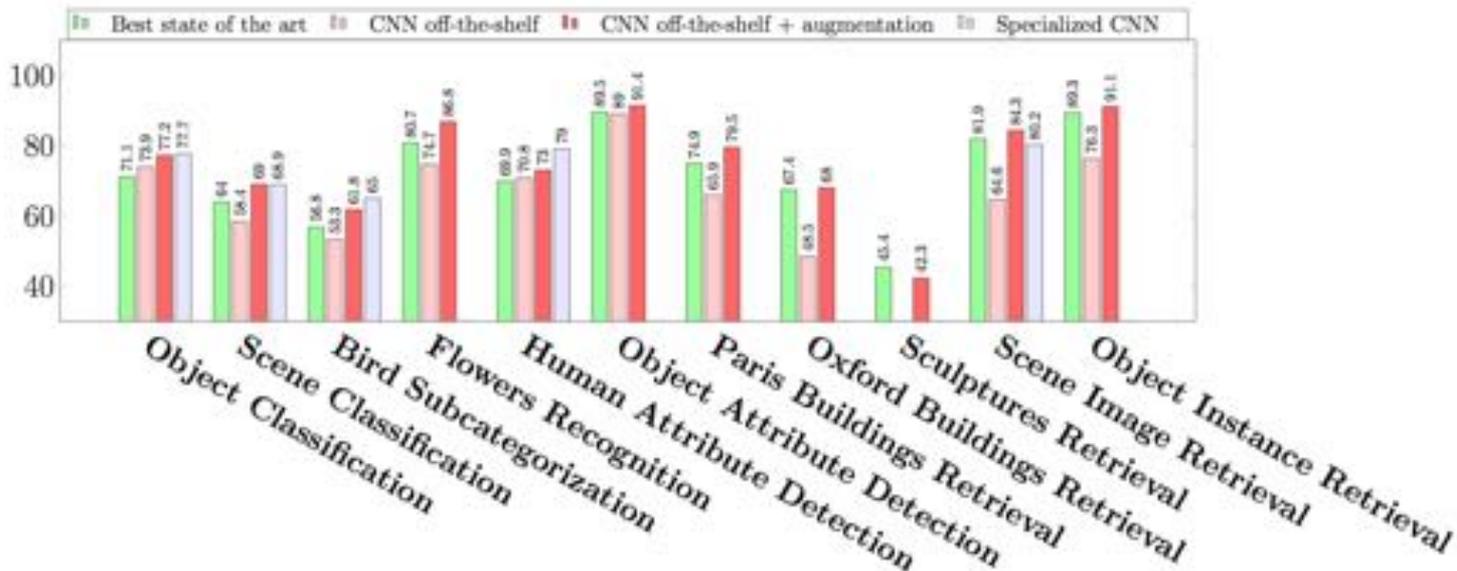
- 物体認識を学習したCNNが獲得した特徴は万能



'kitchen'

SVM etc.

Feature

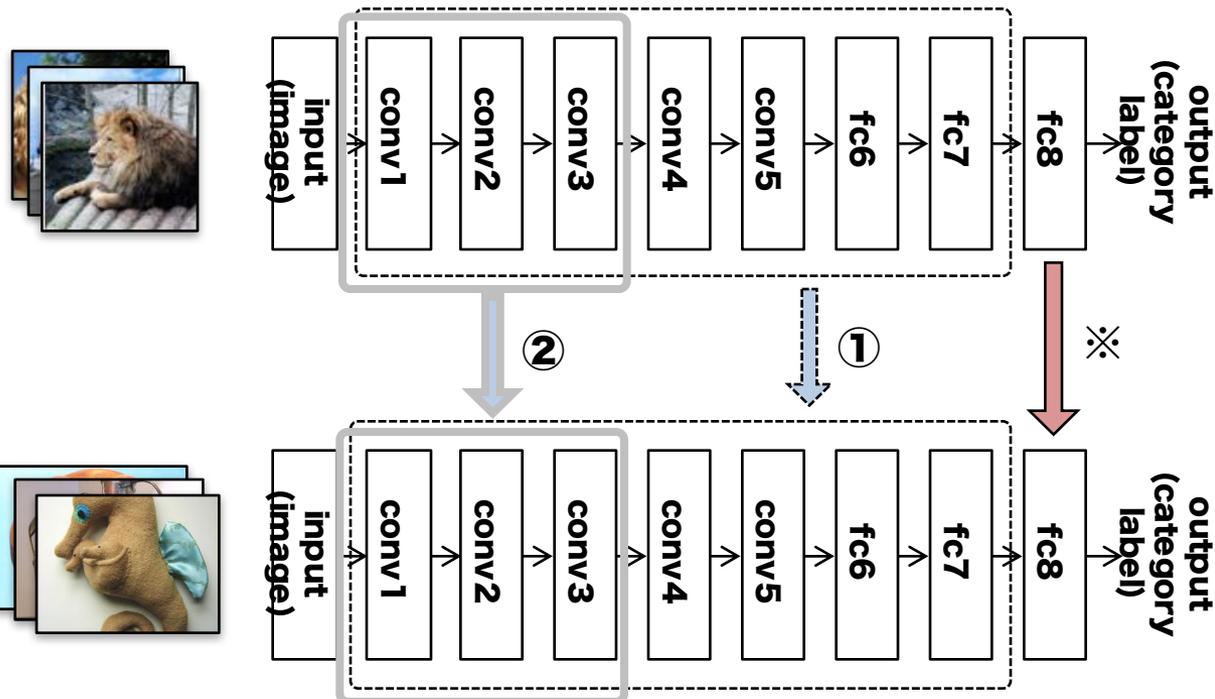


From Razavian+, CNN Features off-the-shelf: an Astounding Baseline for Recognition, 2014

転移学習・fine-tuning

- 1つのタスクを学習したネットの重みを別タスクに転用
 - 下位層では基礎的特徴が学習 ⇒ 何にでも流用される
 - 限られた上位層のみの再学習 ⇒ 小規模データで十分

物体カテゴリ認識

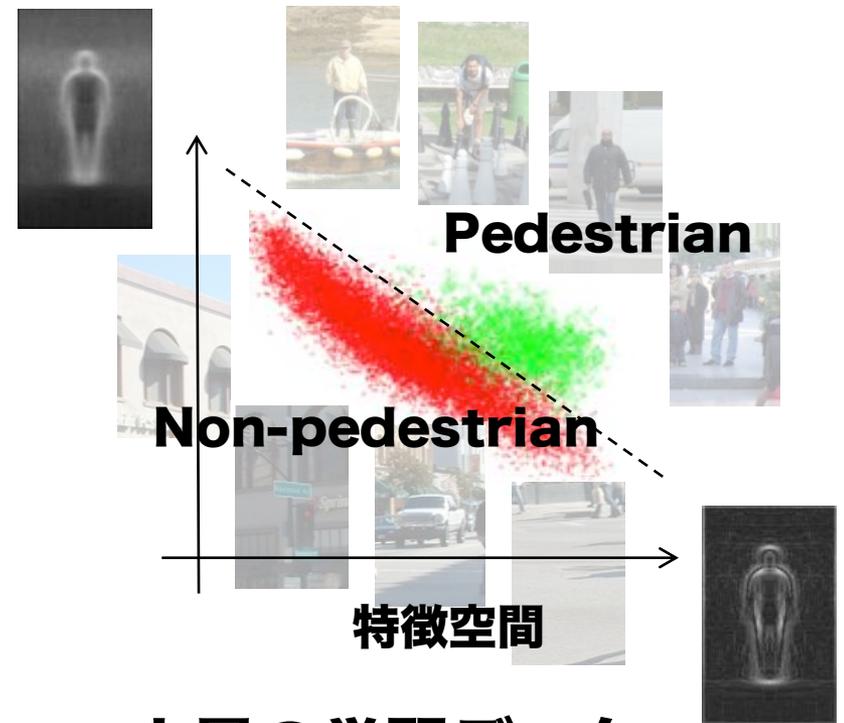
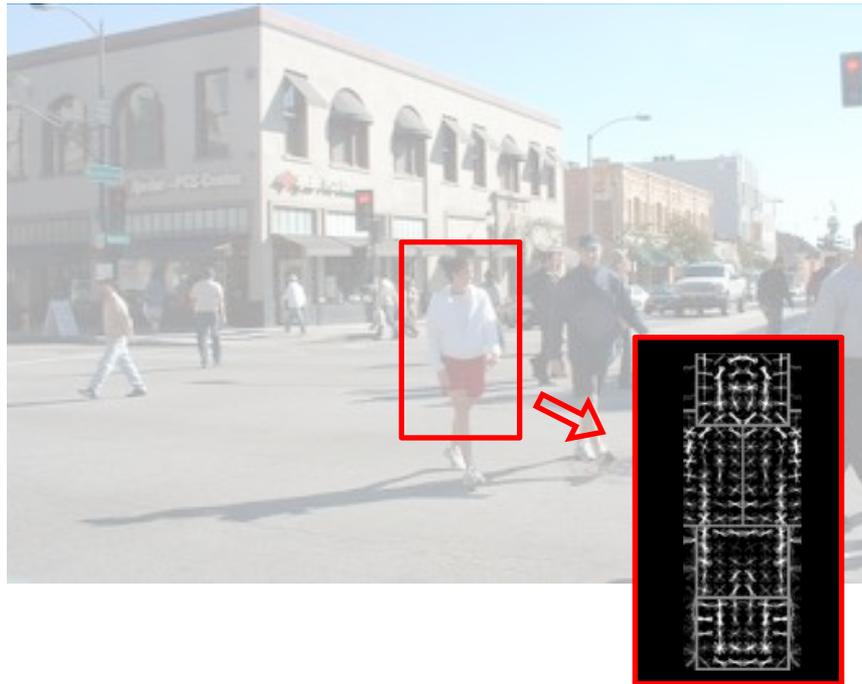


別のタスク

目次

- **最近までの状況：画像認識の革命**
- **いろいろなディープネット**
- **ディープネットの学習**
 - **確率勾配降下法とその発展**
 - **重みの圧縮**
 - **転移学習**
- **ディープネットの理解**
 - **多層の意義**
 - **実験的解析**
 - **生物との対応**
- **まとめ**

画像認識のプロセス



大量の学習データ
を用いて機械学習

画像認識技術の現状

实用レベル

研究途上

1970

1980

1990

2000

2010

???

バーコード



印刷文字

CHAPTER 1
Eigenfaces, eigenscripts, and stability

1.1 Introduction

In this and all the following chapters, we assume color but mean the usual in the chapter with exception of how they arise, unless otherwise stated in an application.

1.1.1 Change of basis and stability. Every invertible matrix is a change-of-basis matrix, and every change-of-basis matrix is invertible (see Section 0.10). Thus, if B is a given basis of a vector space V , if T is a given linear transformation on V , and if $\{u_i\}$ is the B -basis representation of T , the set of all possible basis representations of T is $\{u_i\} \cdot U$, where U is a basis of V .

1.1.2 This is the set of all matrices that are similar to the given matrix A . Similarity is an equivalence relation, and different bases represent the same linear transformation.

1.1.3 Our model requires stable features in order to ensure recognition performance. At first, those properties that are invariant to the underlying linear transformations — and that are an eigenbasis of the linear algebra, i.e., a basis of the null space of the matrix — are the only ones that are stable under any linear transformation.

1.1.4 The notion of stability is a key concept in this chapter.

画像から何を取り出すか？ (どんな特徴)

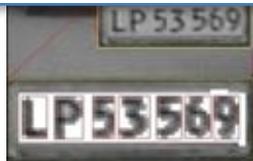
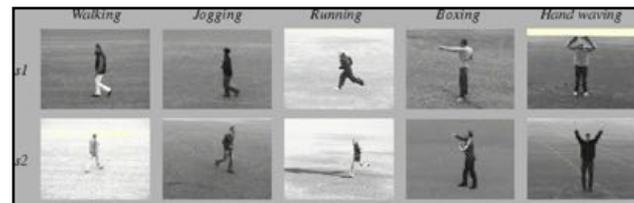
一般物体



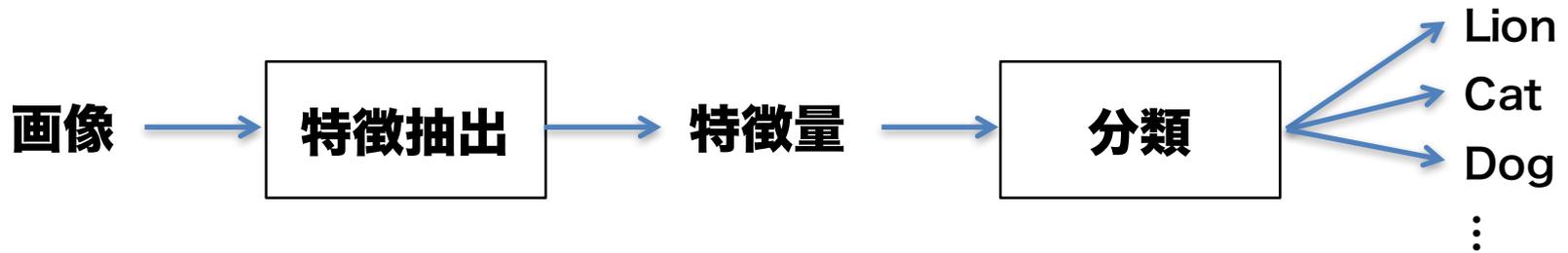
顔



人の行動



問題の難しさ



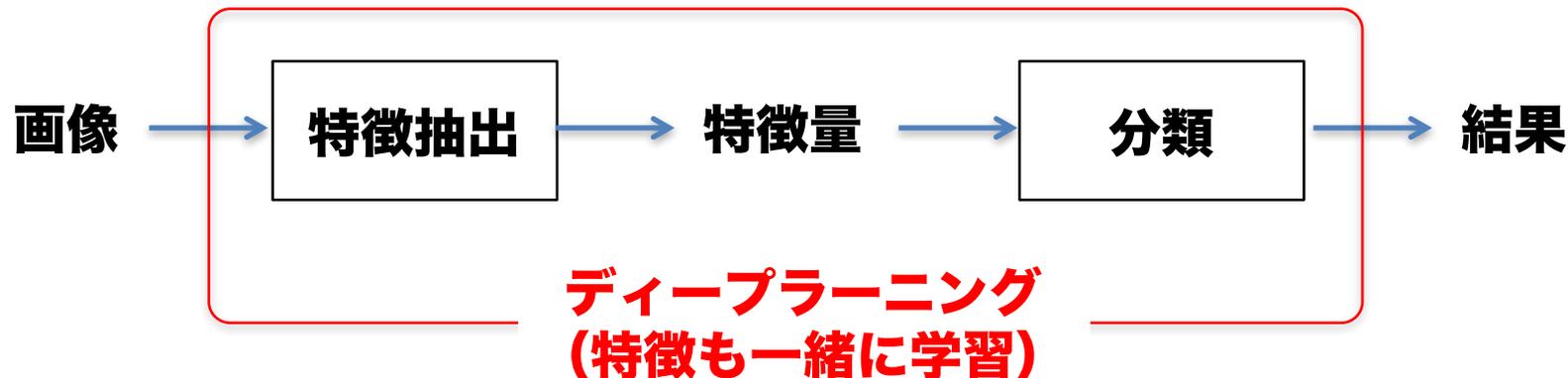
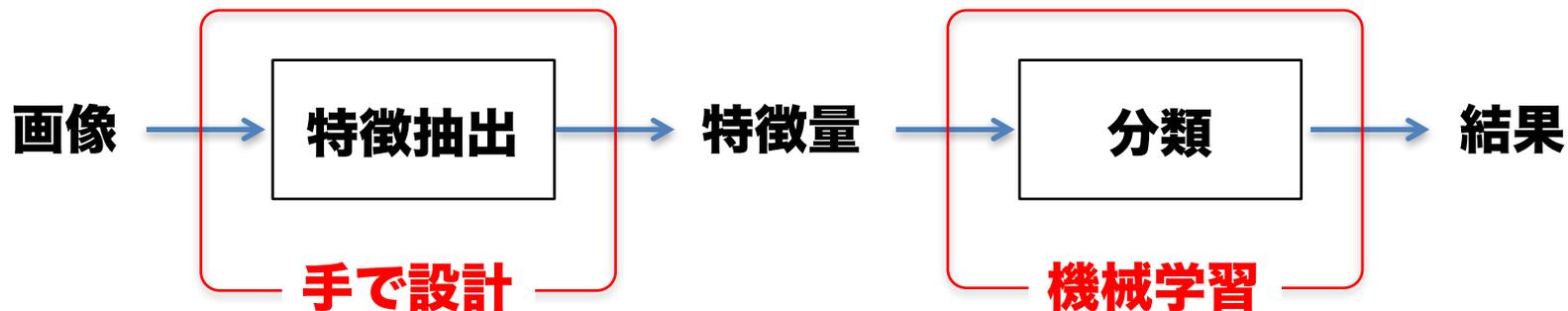
問題の難しさ

- 特徴に求められる性質
 - 不変性：カテゴリ内変動に不変（≒鈍感）
 - 識別力：僅かな違いにも敏感



“Television set”

ディープラーニング：特徴の設計から特徴の学習へ

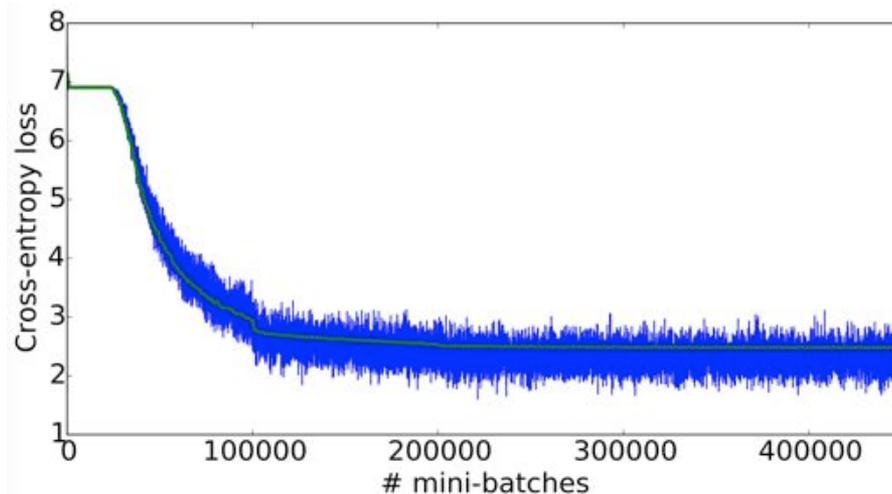


大規模な学習

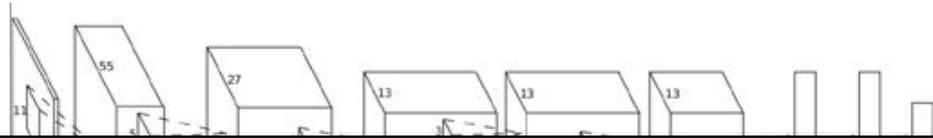
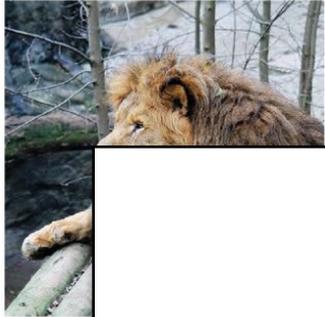
- **ImageNet** : 約2万synsetの計1400万枚
 - ILSVRC:1000の物体カテゴリ・1000枚/カテゴリ
- **GPUを用いて数日～数週間かけて学習**
- **物体カテゴリ認識などで人を超える認識精度**

(He+, Delving deep into rectifier, 2015)

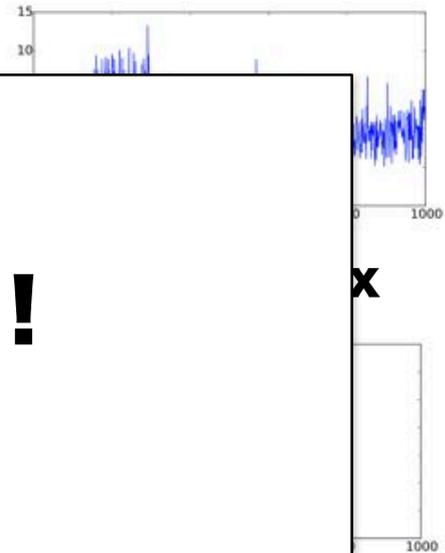
The screenshot shows the ImageNet website interface. At the top, there is a search bar with the text "IMAGENET" and a search button. Below the search bar, there are navigation links for "Home", "About", "Explore", and "Download". The main content area displays the search results for "Car, auto, automobile, machine, motorcar". It includes a description: "A motor vehicle with four wheels; usually propelled by an internal combustion engine; *he needs a car to get to work*". There are also statistics: "1307 pictures", "92.48% Popularity Percentile", and "Wordnet IDs". A treemap visualization is shown, displaying a grid of small images categorized by type. On the left side, there is a list of synsets with their respective counts in brackets, such as "ImageNet 2011 Fall Release (32326)", "plant, flora, plant life (4486)", "geological formation, formation (1)", "natural object (1112)", "sport, athletics (1176)", "artifact, artefact (10504)", "instrumentality, instrumentation device (2760)", "implement (726)", "container (744)", "wheeled vehicle (229)", "baby buggy, baby car, bicycle, bike, wheel, cy, boneshaker (0)", "car, railcar, railway car, handcart, pushcart, ca, horse-drawn vehicle (2)", "motor scooter, scoote, rolling stock (0)", "scooter (0)", "self-propelled vehicle (", "personnel carrier (C", "reconnaissance veh", "weapons carrier (0)", "armored vehicle, an, carrier (0)", "forklift (0)", "locomotive, engine, motor vehicle, auto", "amphibian, amsp", "bloodmobile (0)".



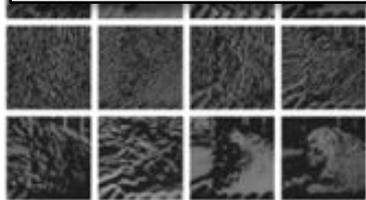
入力画像に対するレスポンス



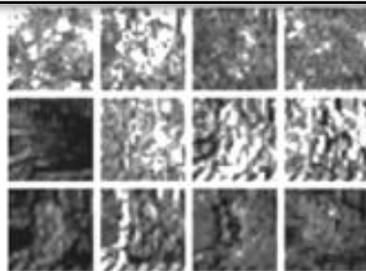
出力層



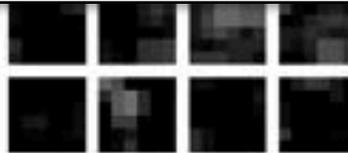
何を計算しているのか不明！
(解釈不可能)



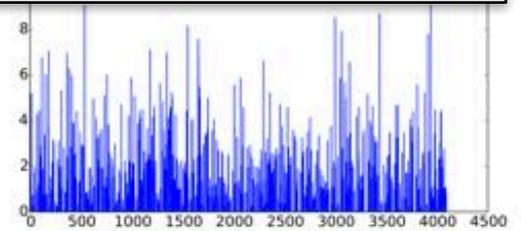
初期畳込み層



プーリング層

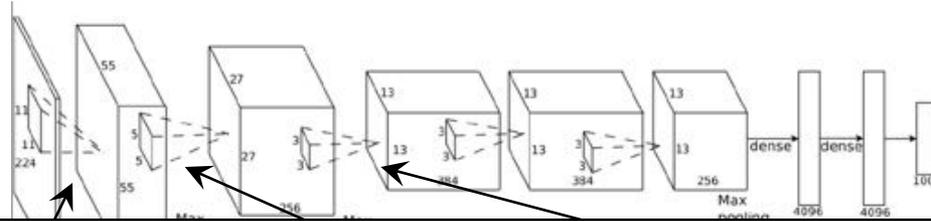


最終畳込み層



全結合層

学習された特徴



**何を学習したのか不明！
(解釈不可能)**

第1畳込み層フィルタ

第2畳込み層フィルタ

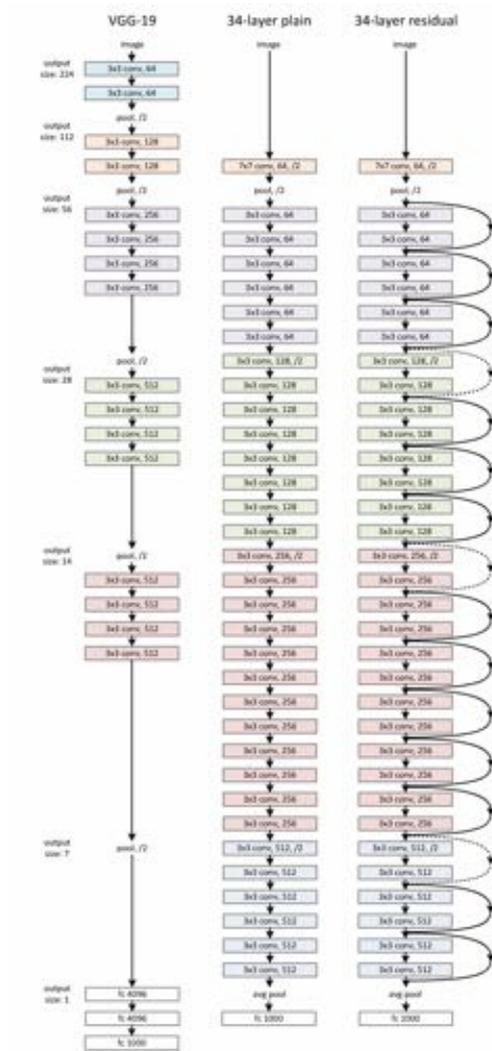
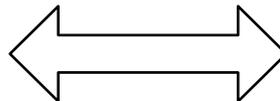
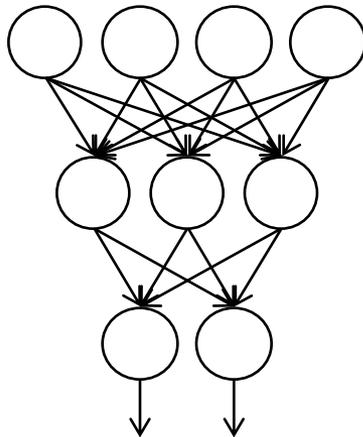
第3畳込み層フィルタ

目次

- **最近までの状況：画像認識の革命**
- **いろいろなディープネット**
- **ディープネットの学習**
 - **確率勾配降下法とその発展**
 - **重みの圧縮**
 - **転移学習**
- **ディープネットの理解**
 - **多層の意義**
 - **実験的解析**
 - **生物との対応**
- **まとめ**

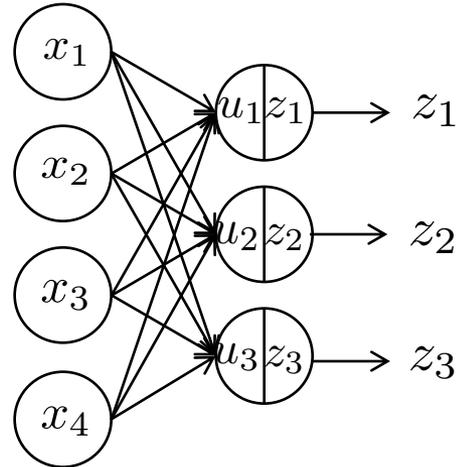
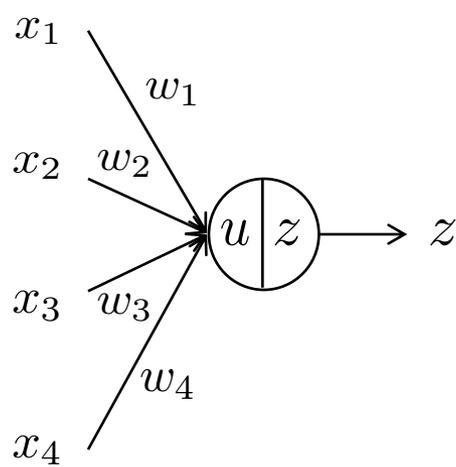
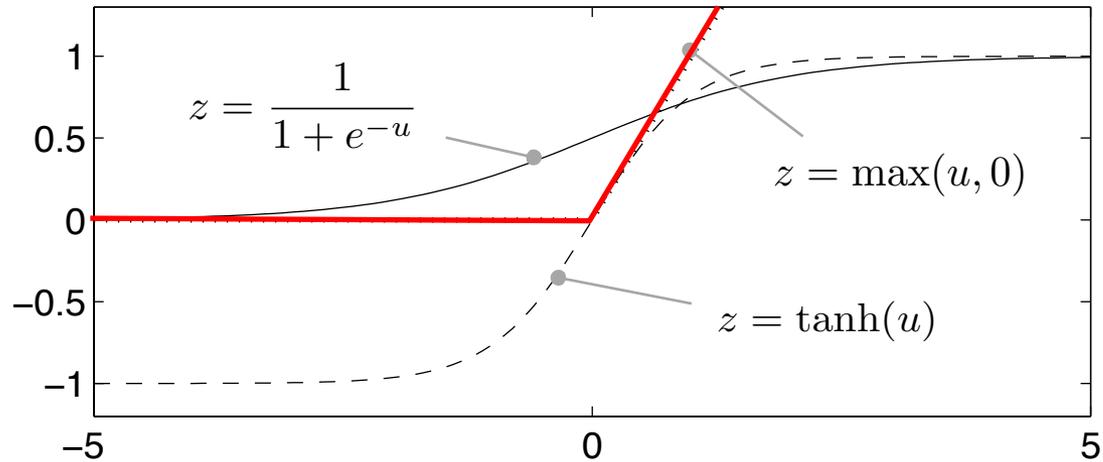
多層であることの意味

- **2層ネットでも中間層のユニット数次第でどんな関数も表現可能 (universal approximation theory) [Hornik+89, Cybenko89]**
- **152層のResNetがILSVRC2015でトップ [He+2015]**



Deep residual learning for image recognition [He+2015]

Rectifier Net



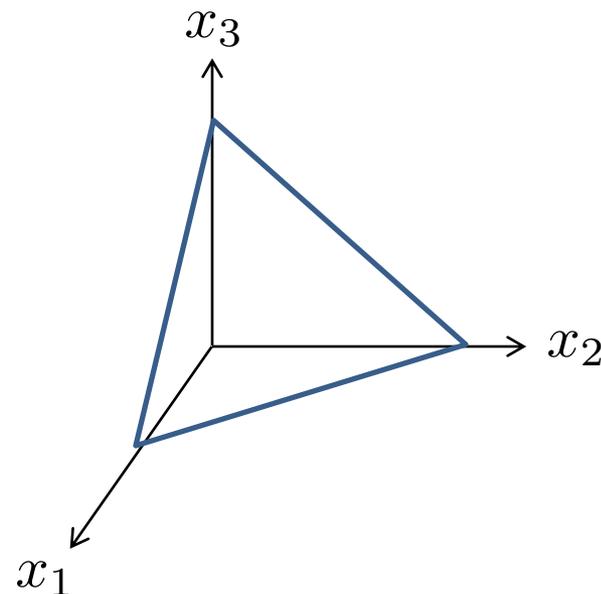
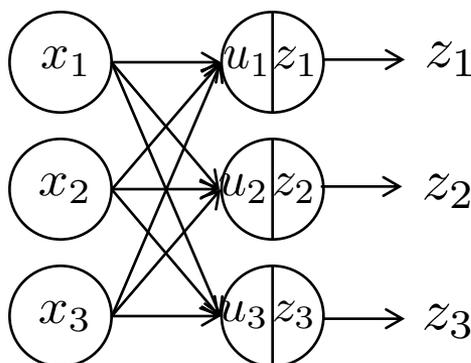
Rectifier Netの線形領域数

Montufar+, On the Number of Linear Regions of Deep Neural Networks, 2014

- Rectifier ネットは入力空間を超平面分割

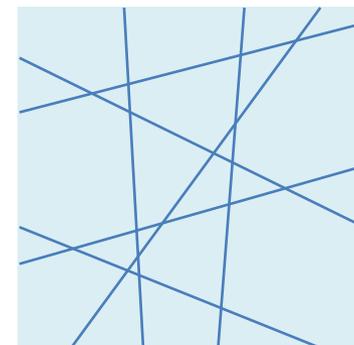
$$z_i = \max(\mathbf{w}_i^\top \mathbf{x} + b_i, 0)$$

$$\mathbf{x} = [x_1, x_2, x_3]$$



- n_0 次空間内の n_1 個の超平面による空間分割数は高々

$$\sum_{j=0}^{n_0} \binom{n_1}{j}$$

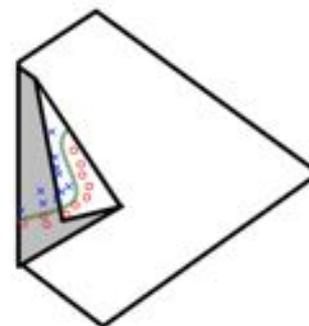
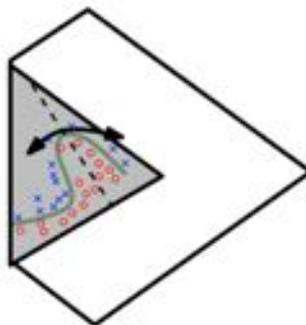
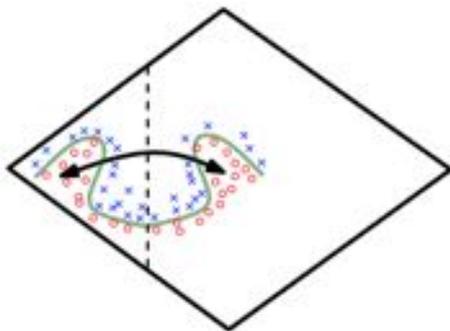
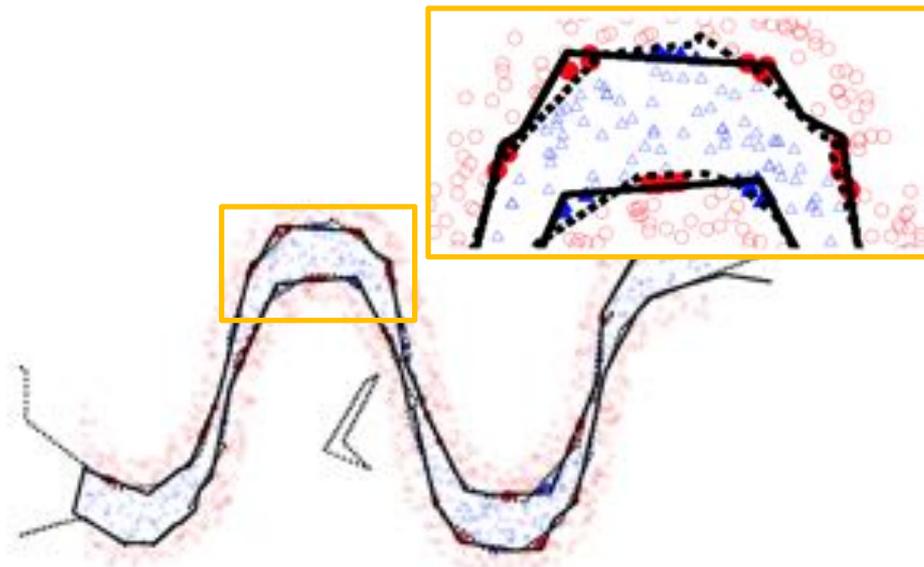
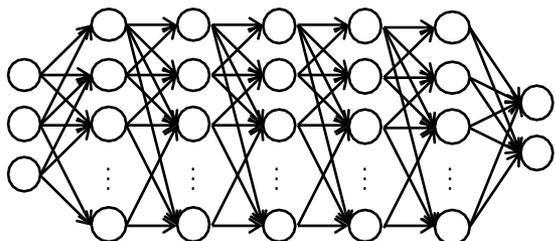


Rectifier Netの線形領域数

Montufar+, On the Number of Linear Regions of Deep Neural Networks, 2014

L層ネットの入力空間の線形領域数：

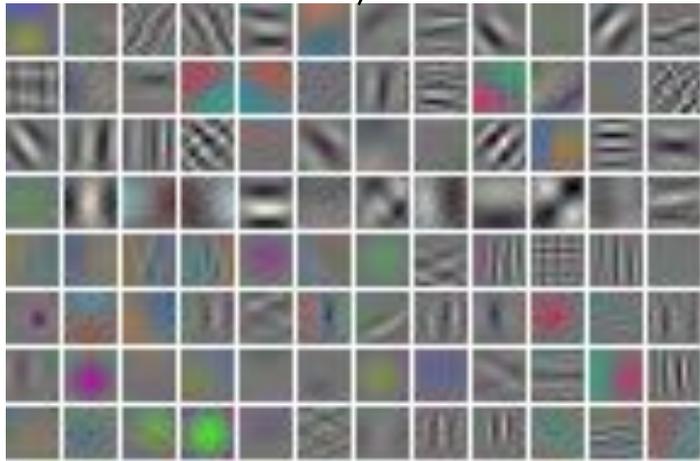
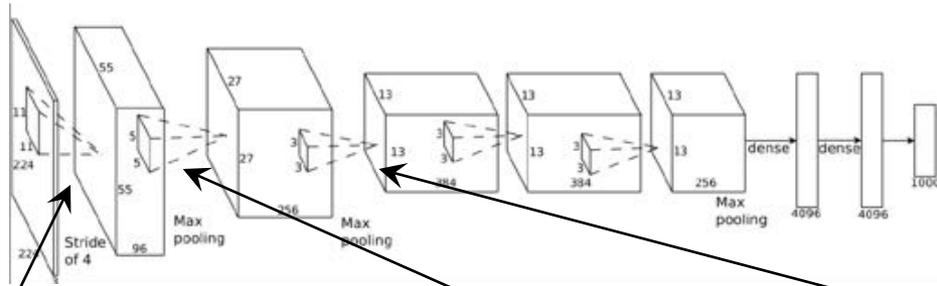
$$\left(\prod_{i=1}^{L-1} \left\lfloor \frac{n_i}{n_0} \right\rfloor \right) \sum_{j=0}^{n_0} \binom{n_L}{j}$$



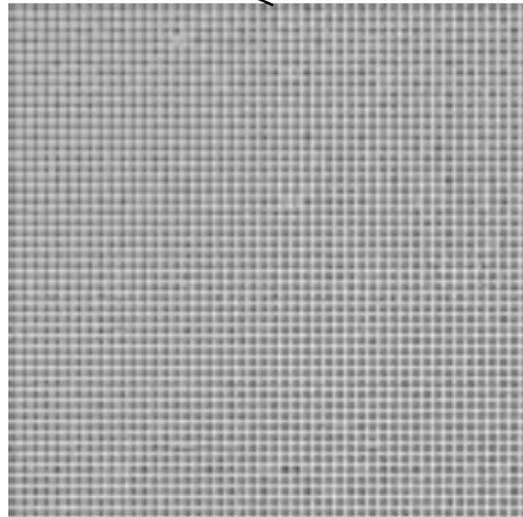
目次

- **最近までの状況：画像認識の革命**
- **いろいろなディープネット**
- **ディープネットの学習**
 - **確率勾配降下法とその発展**
 - **重みの圧縮**
 - **転移学習**
- **ディープネットの理解**
 - **多層の意義**
 - **実験的解析**
 - **生物との対応**
- **まとめ**

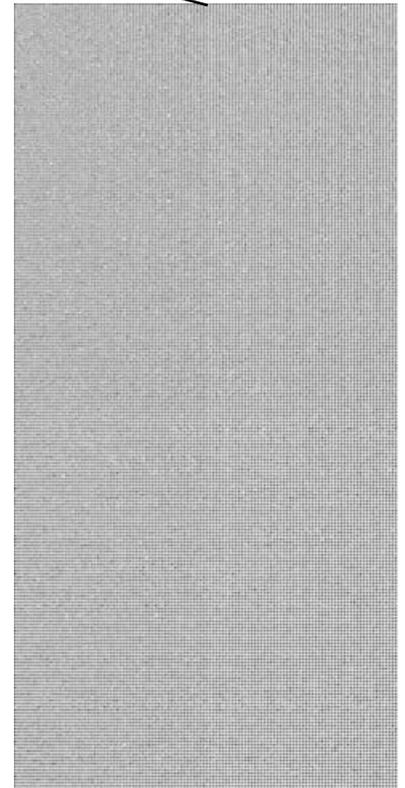
学習された特徴



第1畳込み層フィルタ



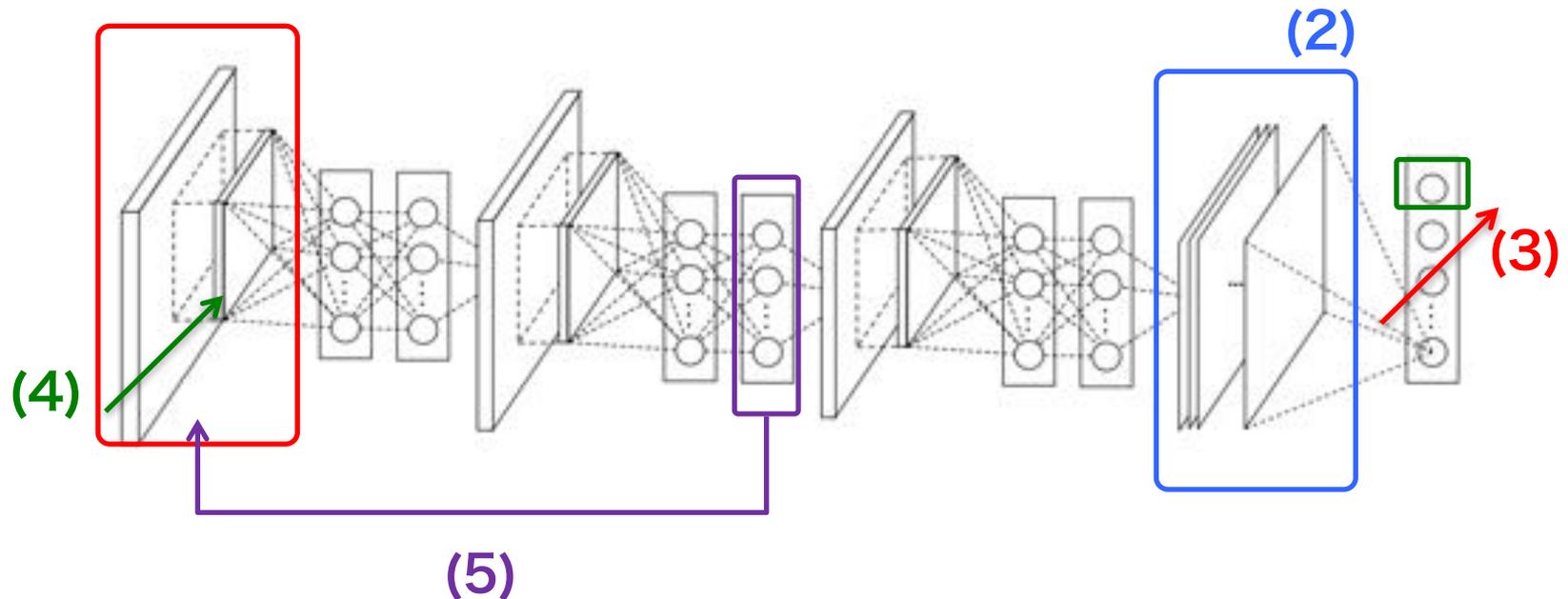
第2畳込み層フィルタ



第3畳込み層フィルタ

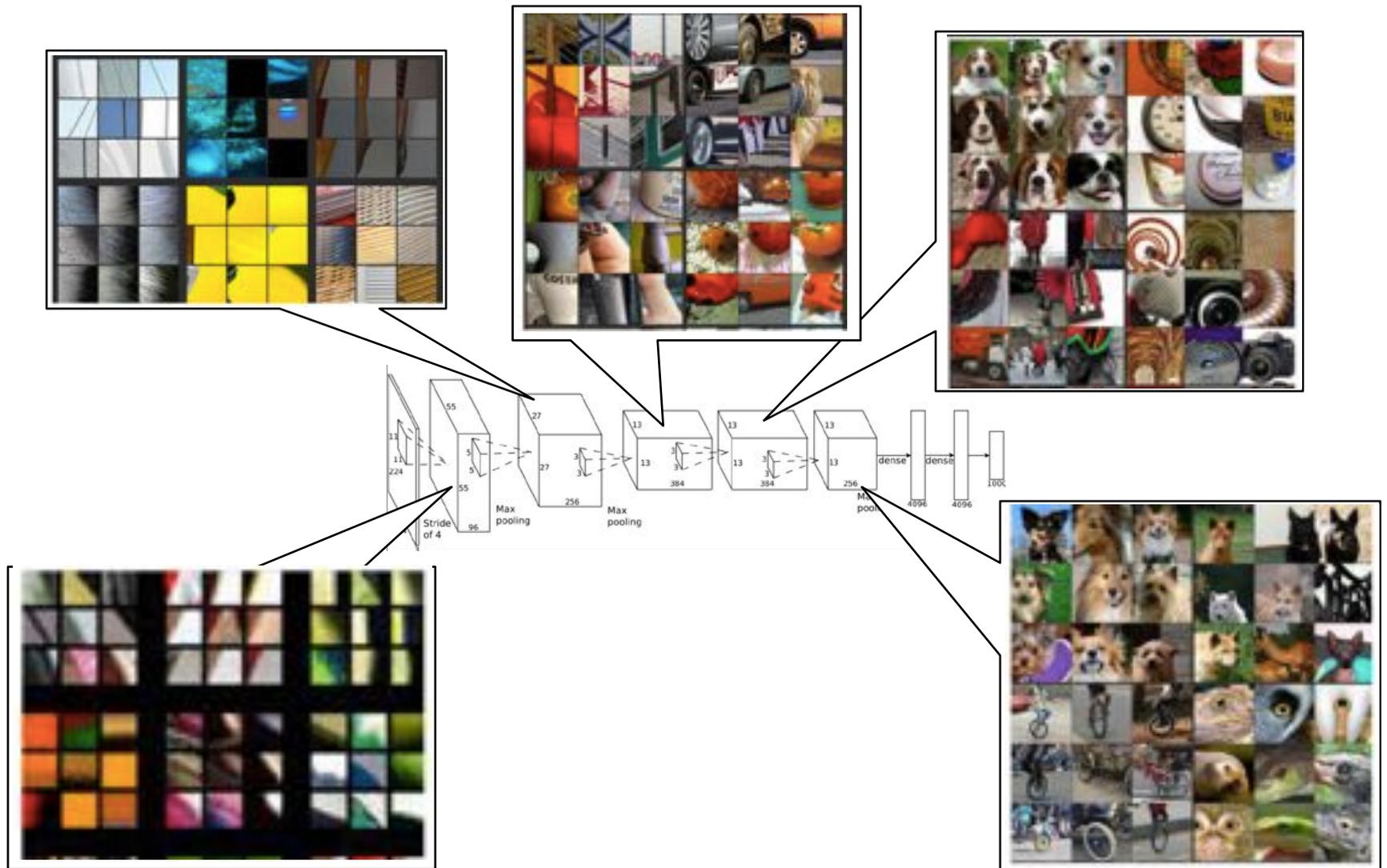
CNNの「可視化」

1. 各層各ユニットを最も活性化する画像 [Zeiler+14]
2. GAP(global average pooling)層の出力 [Lin+14]
3. カテゴリ尤度を最大化する最適入力 [Simonyan+14]
4. カテゴリ尤度のヤコビ行列（入力に関する微分） [Simonyan+14]
5. 中間層から入力を逆算 [Mahendran+15]



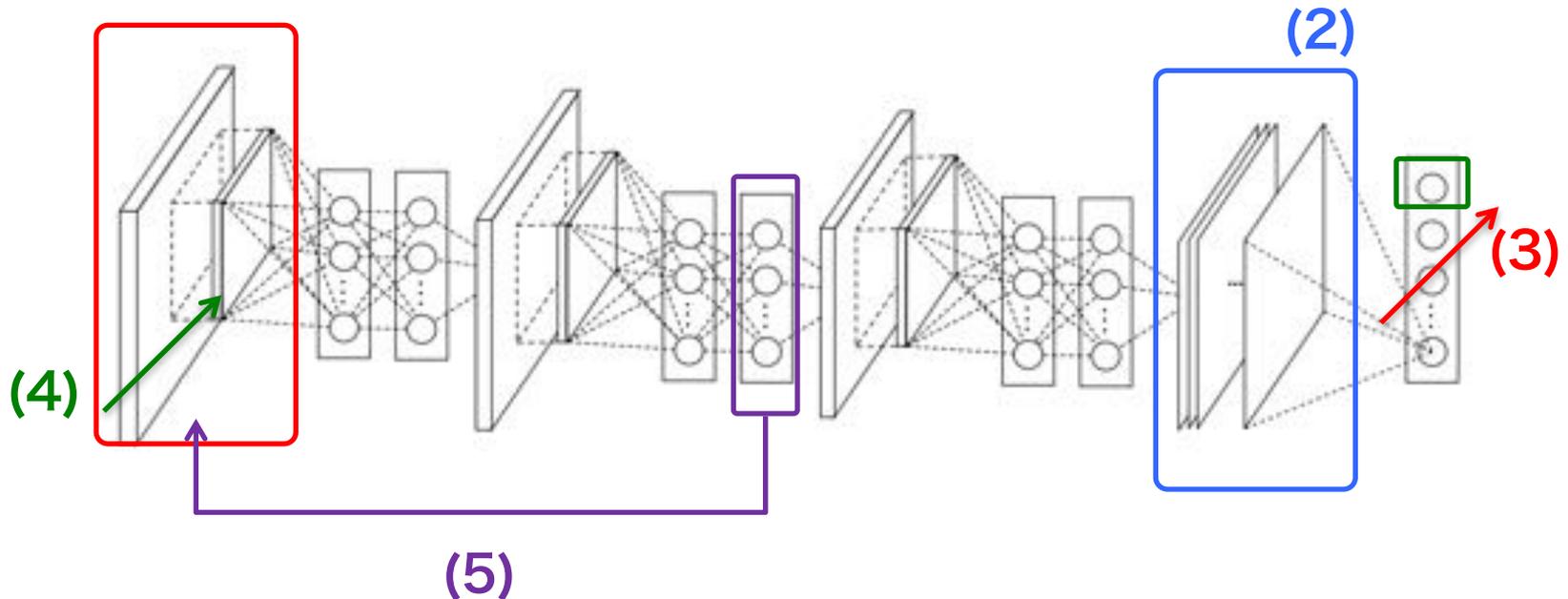
各層各ユニットを活性化する入力画像

Zeiler, Fergus, Visualizing and understanding convolutional networks, ECCV2014



CNNの「可視化」

1. 各層各ユニットを最も活性化する画像[Zeiler+14]
2. GAP(global average pooling)層の出力[Lin+14]
3. カテゴリ尤度を最大化する最適入力 [Simonyan+14]
4. カテゴリ尤度のヤコビ行列（入力に関する微分） [Simonyan+14]
5. 中間層から入力を逆算 [Mahendran+15]



CNNの「最適入力」

Simonyan+, Deep Inside Convolutional Networks: Visualising..., 2013



dumbbell



cup



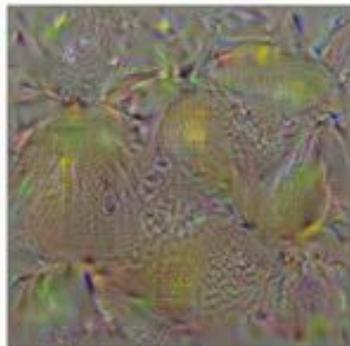
dalmatian



goose



bell pepper



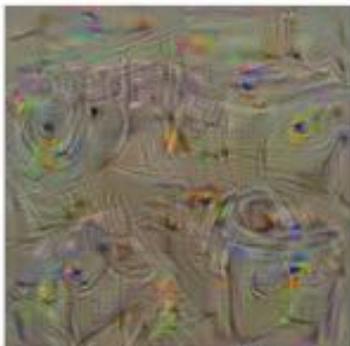
lemon



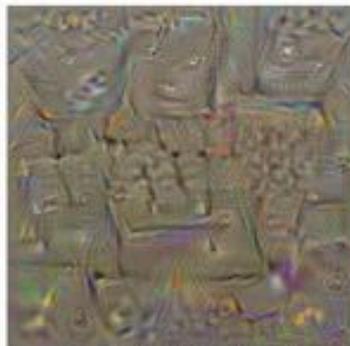
husky



ostrich



washing machine



computer keyboard



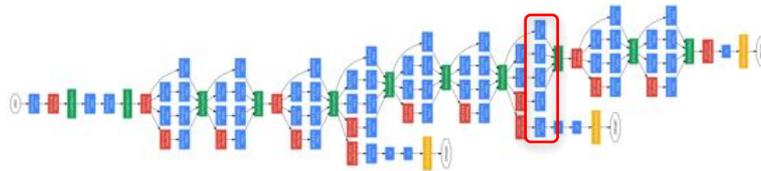
kit fox



limousine

CNNの「最適入力」

Google blogpost, Deep dream, 2015



"Admiral Dog!"



"The Pig-Snail"



"The Camel-Bird"

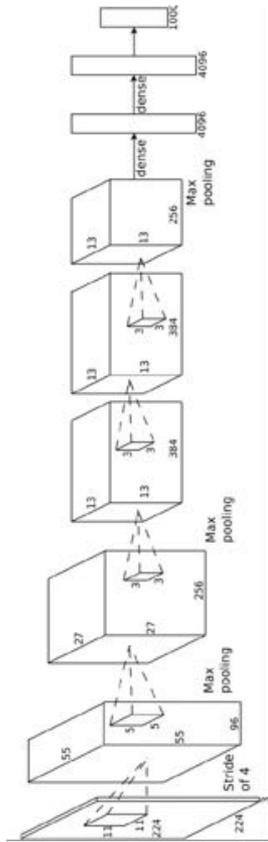


"The Dog-Fish"

画風を再現した風景画像

Gatys+, A Neural Algorithm of Artistic Style, 2015

- スタイルの層内チャンネル間の相関を入力の前層出力と一緒に再現



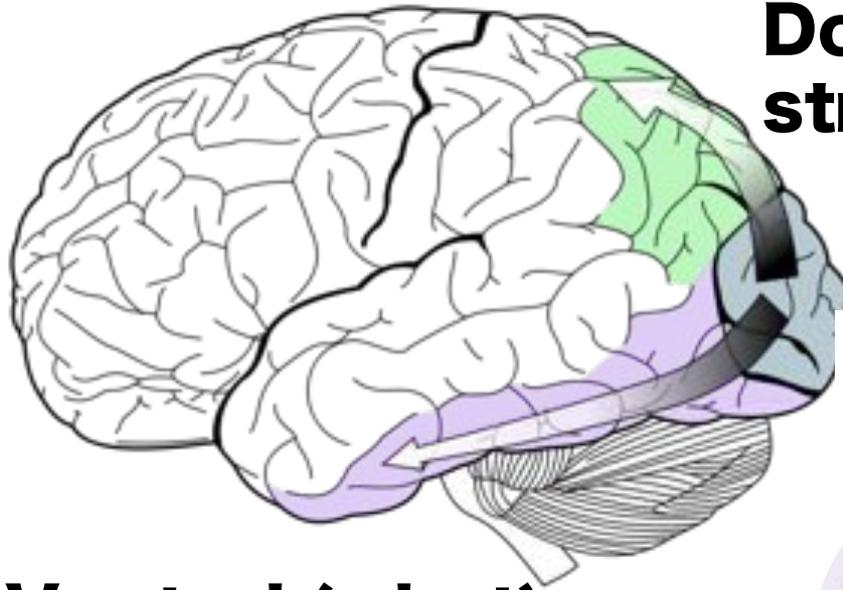
目次

- **最近までの状況：画像認識の革命**
- **いろいろなディープネット**
- **ディープネットの学習**
 - **確率勾配降下法とその発展**
 - **重みの圧縮**
 - **転移学習**
- **ディープネットの理解**
 - **多層の意義**
 - **実験的解析**
 - **生物との対応**
- **まとめ**

視覚皮質の2つの経路

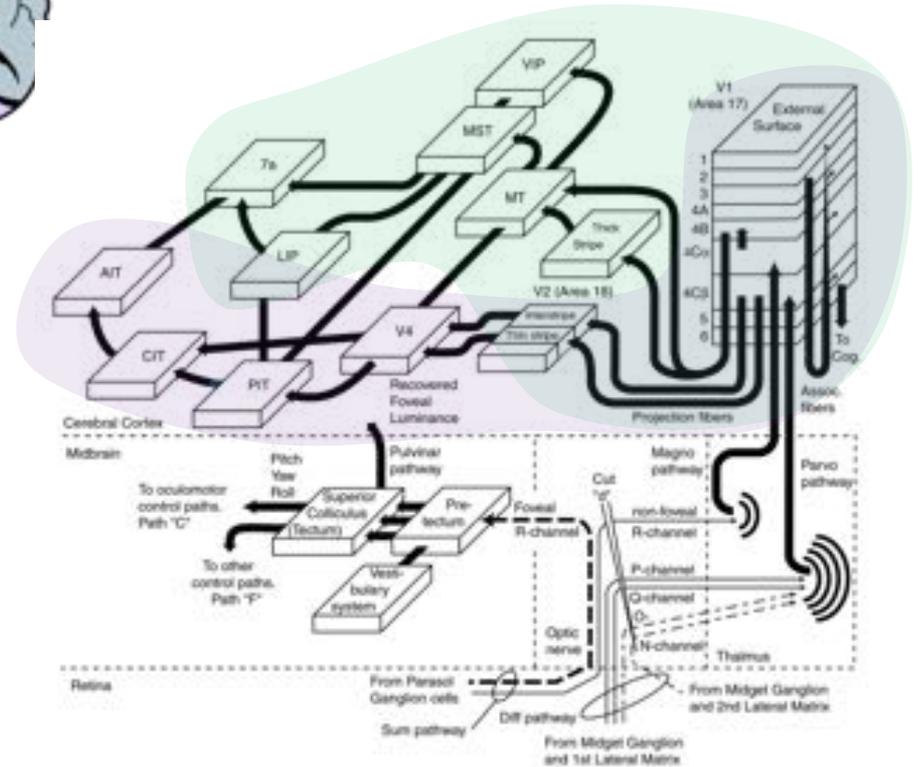
Dorsal (where) stream

(背側視覚皮質路)



Ventral (what) stream

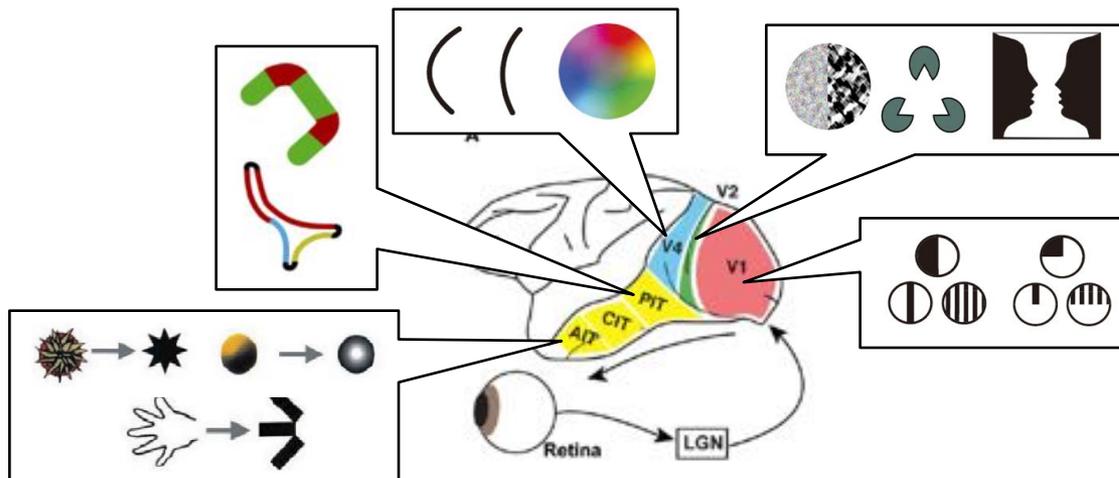
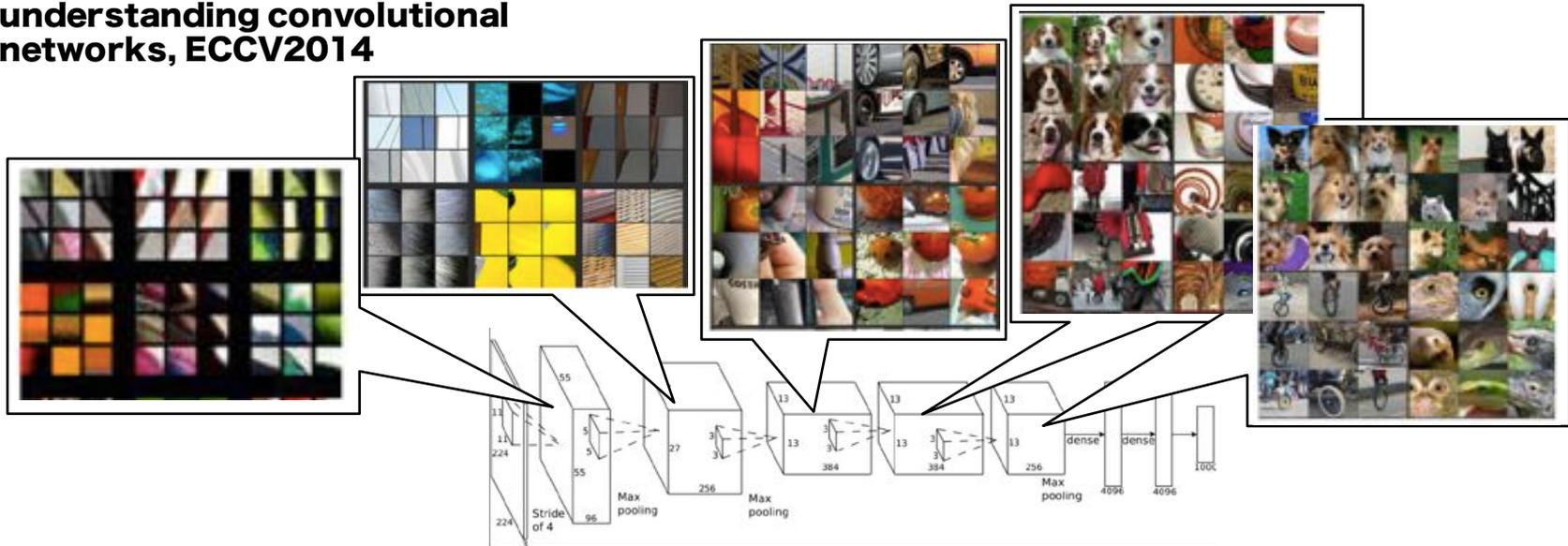
(腹側視覚皮質路)





CNN & ventral stream

Zeiler, Fergus, Visualizing and understanding convolutional networks, ECCV2014

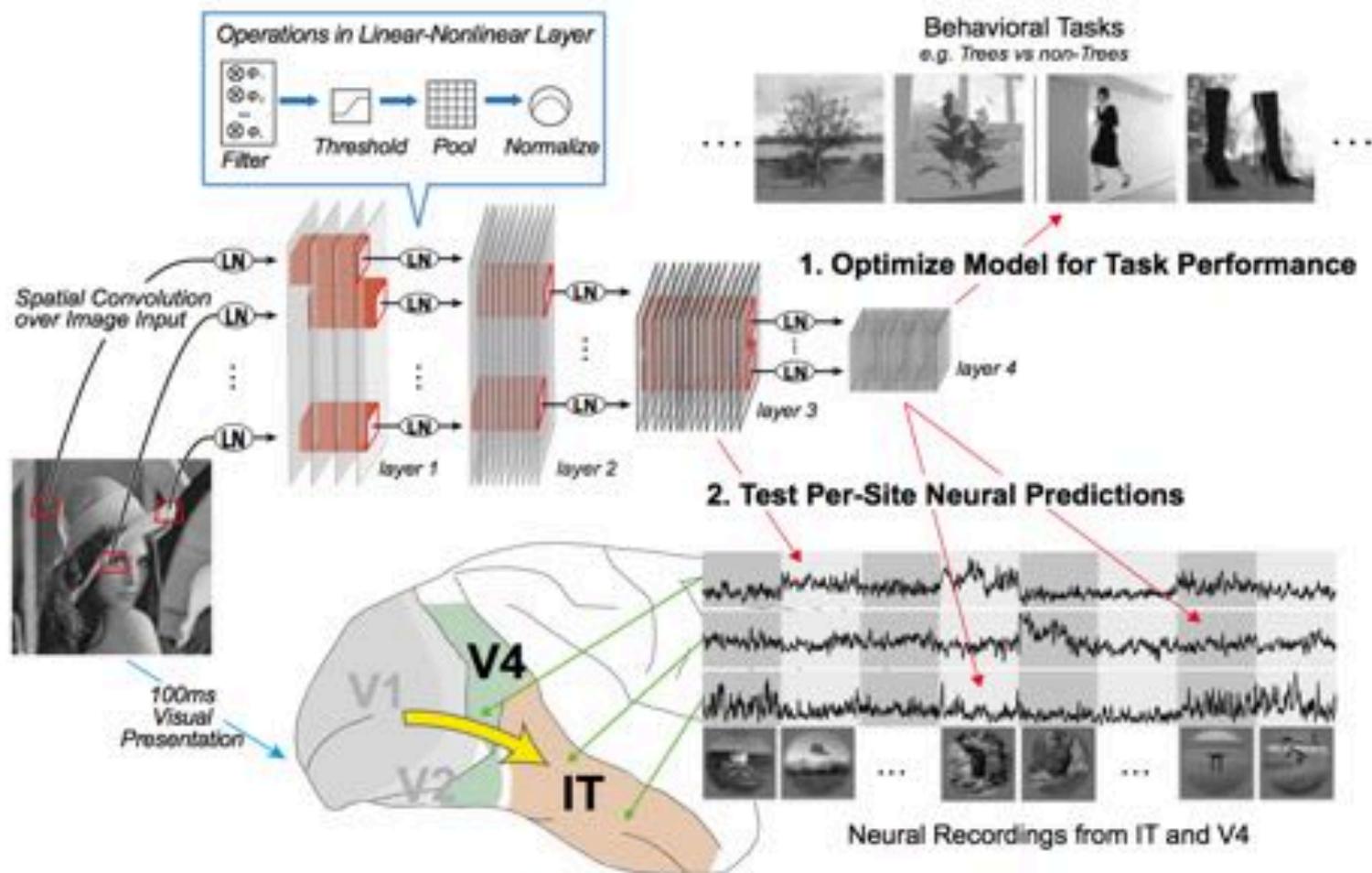


Kruger et al., Deep Hierarchies in the Primate Visual Cortex: What Can We Learn for Computer Vision?, PAMI2013

CNN = a good model of ventral stream

Yamins, Performance-optimized hierarchical models predict neural responses in higher visual cortex, 2014

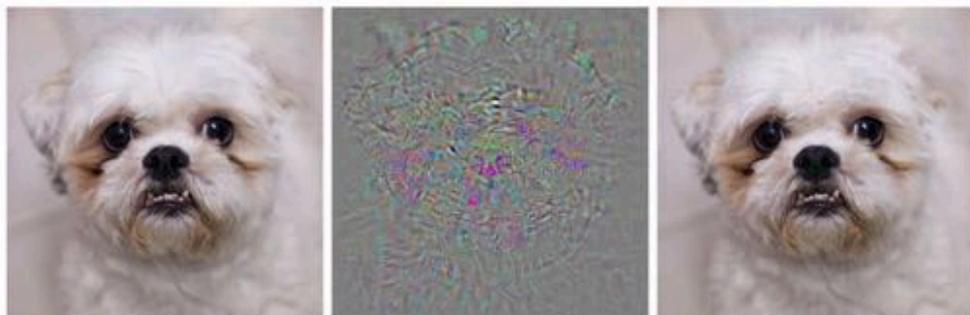
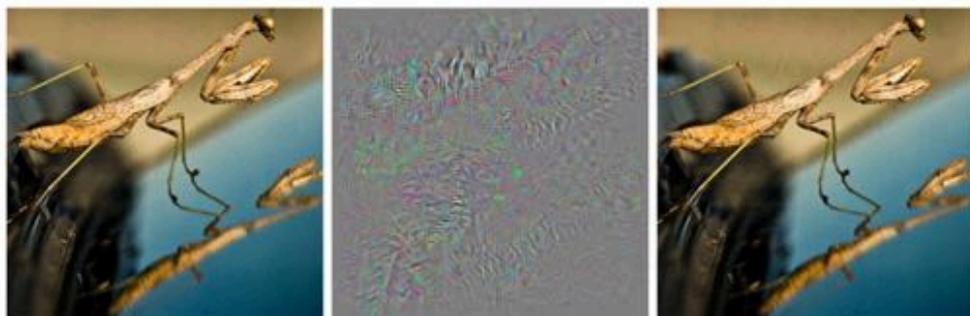
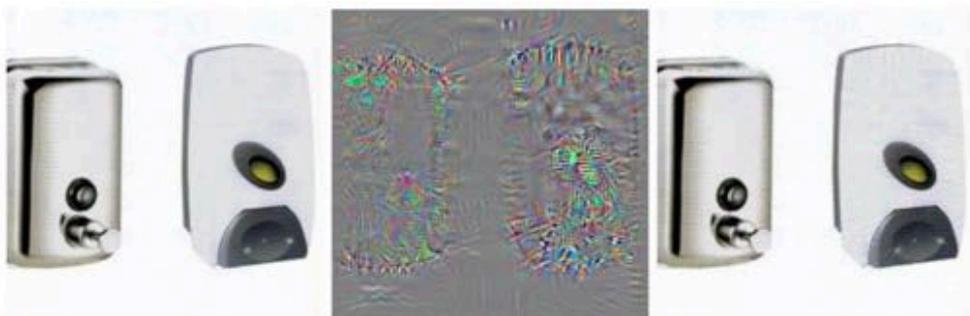
- 多層CNNとサルの高次視覚野の神経活動に高い類似性



CNNを騙す

Szegedy+, Intriguing properties of neural networks, 2014

- 人には同じに見えるが、CNNには全然違って見える



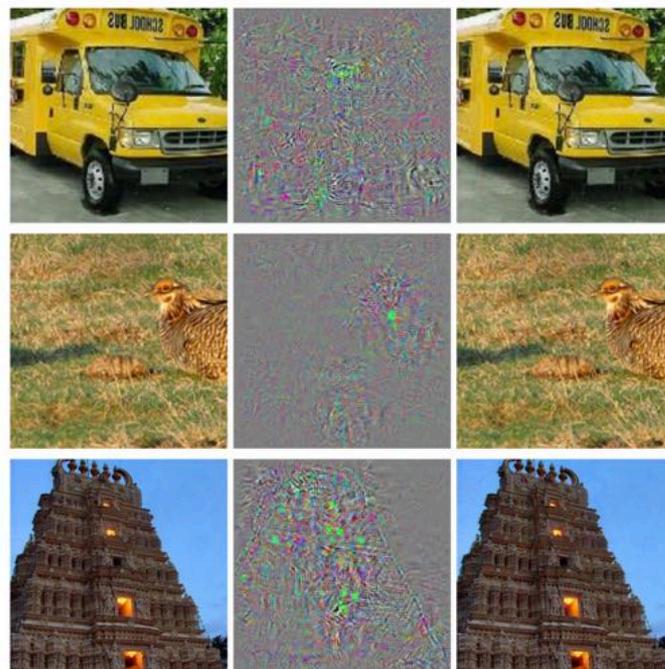
Correctly
recognized

Additive
noise

Recognized
as "Ostrich"

- Minimize $\|r\|_2$ subject to:

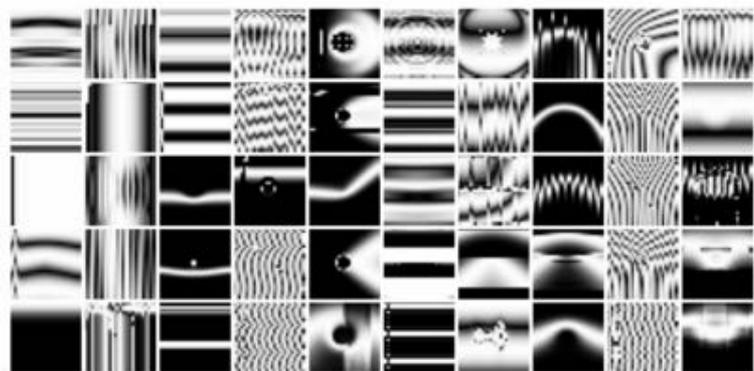
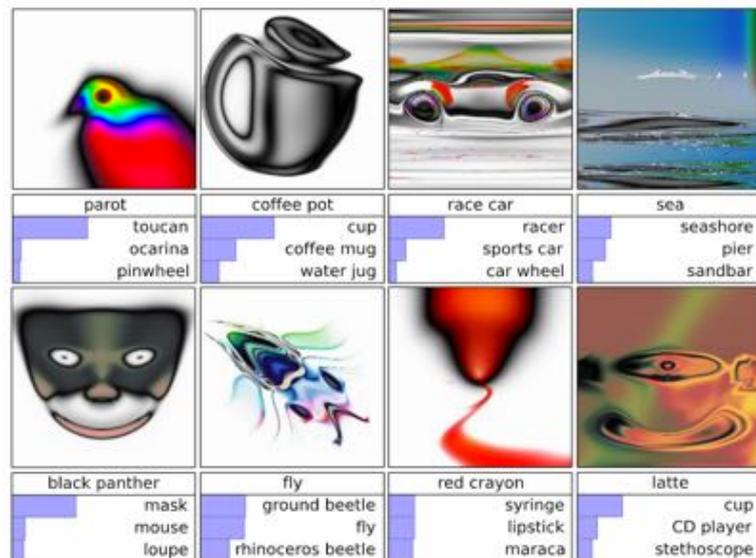
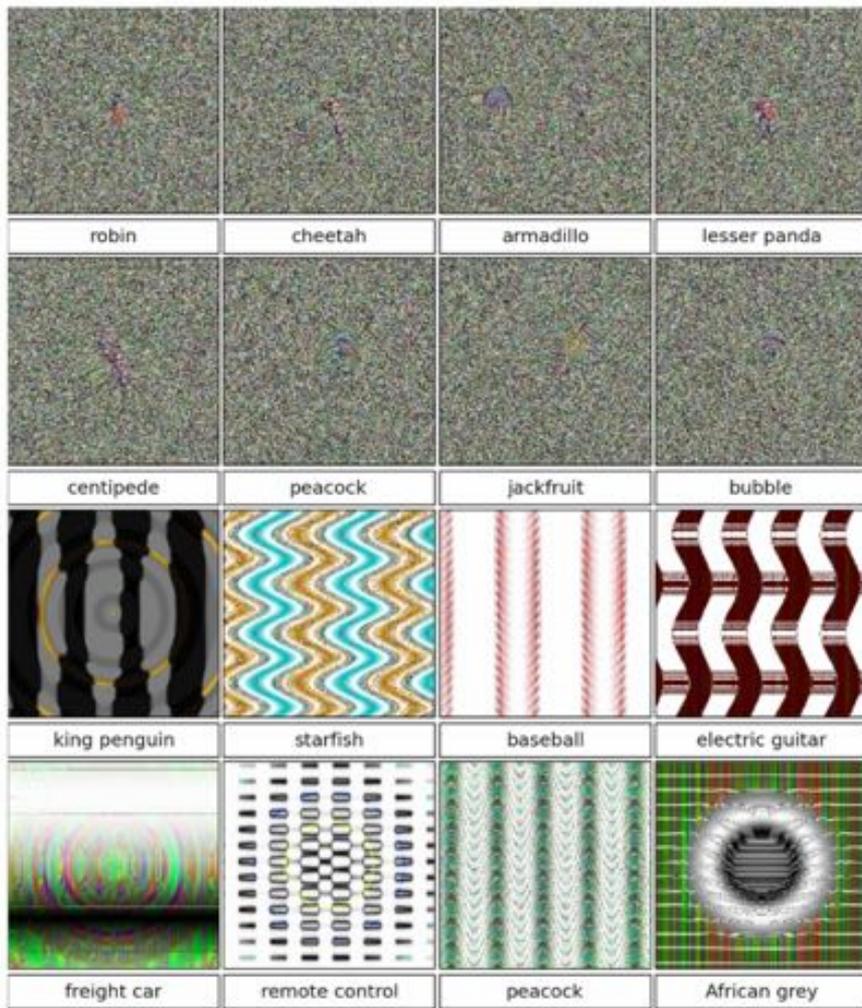
1. $f(x + r) = l$
2. $x + r \in [0, 1]^m$



CNNを騙す

Nguyen+, Deep Neural Networks are Easily Fooled..., 2014

- 人にはそうは見えないがCNNにはそう見える



目次

- **最近までの状況：画像認識の革命**
- **いろいろなディープネット**
- **ディープネットの学習**
 - **確率勾配降下法とその発展**
 - **重みの圧縮**
 - **転移学習**
- **ディープネットの理解**
 - **多層の意義**
 - **実験的解析**
 - **生物との対応**
- **まとめ**

まとめ

- **画像認識のいくつかのタスクで極めて高い性能**
 - 認識精度で人を上回る
 - できないタスクもまだ沢山ある
- **GPU 1 台で数日～数週間の学習**
- **学習方法の進展によってより多層に**
- **高い性能がどこから来るのかはよく分からない**
 - 視覚皮質の研究との関係
- **その他の分野：音声認識・自然言語処理**