企画セッション2「ディープラーニング」

- 趣旨：応用3分野におけるDeep Learning（深層学習）の研究の現状

- 画像：岡谷貴之（東北大学）
  - 「画像認識分野でのディープラーニングの研究動向」
- 音声：久保陽太郎（NTTコミュニケーション科学基礎研究所）
  - 「音声認識分野における深層学習技術の研究動向」
- 自然言語処理：渡邉陽太郎（東北大学）
  - 「自然言語処理におけるディープラーニングの現状」

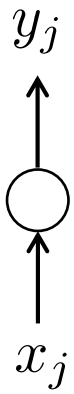
画像認識分野での
ディープラーニングの研究動向

岡谷 貴之
東北大学

NNの基本要素

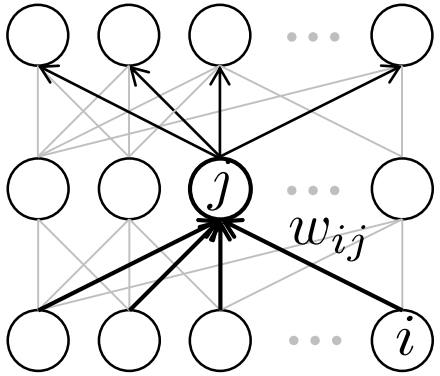
Neural Network

非線形入出力のユニット　　ネットワーク



$$y_j = f(x_j)$$

$$x_j = b_j + \sum_i y_i w_{ij}$$

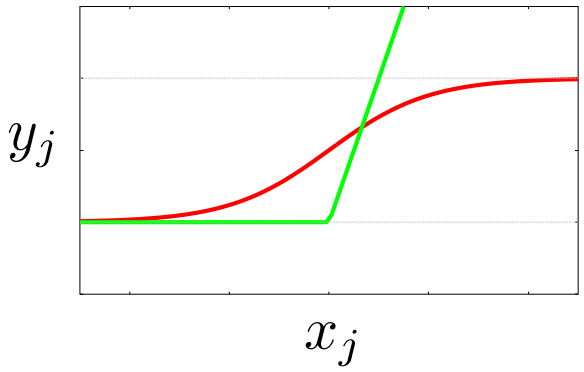


活性化（activation）関数

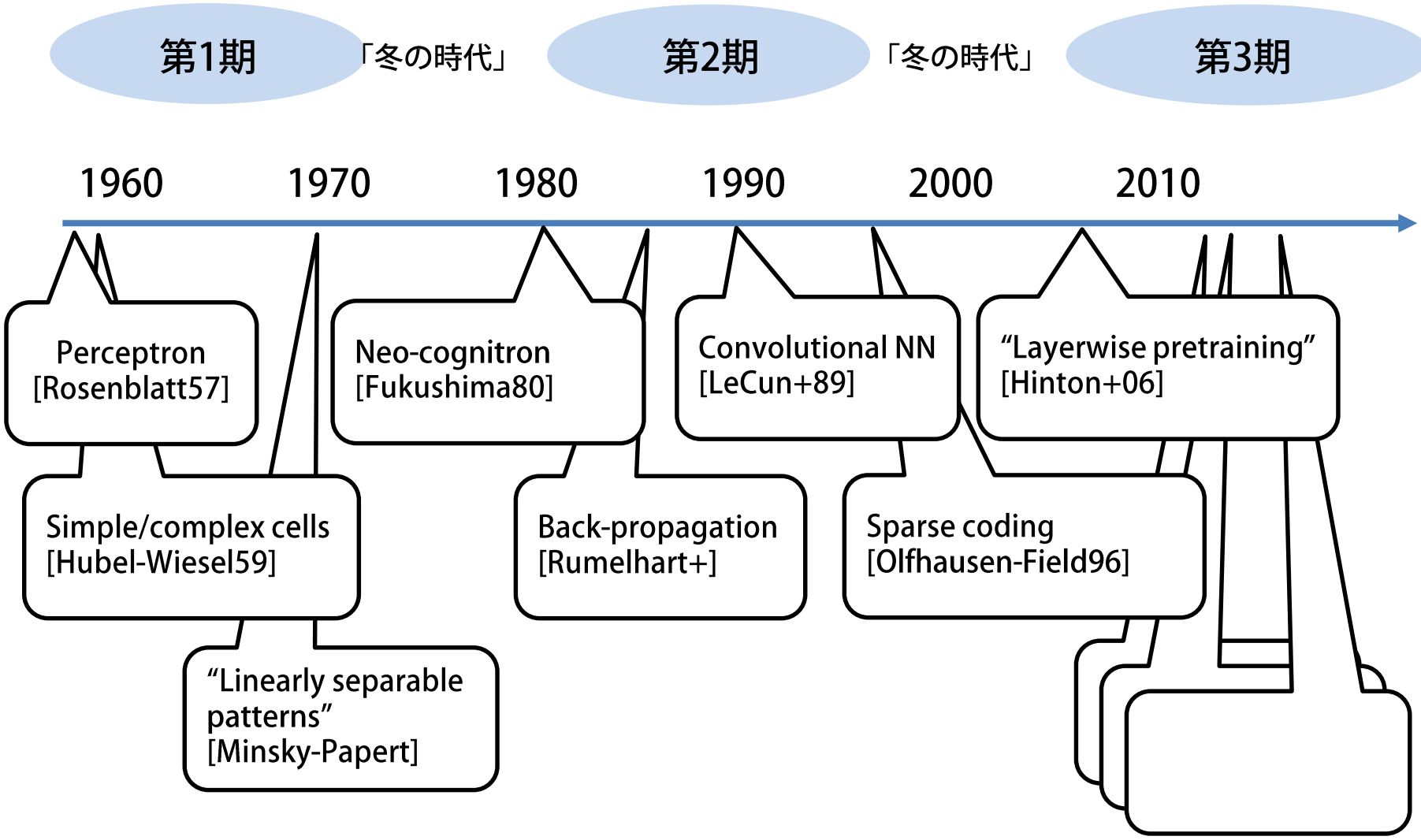
ロジスティックシグモイド　$f(x_j) = \dfrac{1}{1+e^{-x_j}}$

双曲線正接関数　$f(x_j) = \tanh(x_j)$

Rectified linear（ReLU）　$f(x_j) = \max(x_j, 0)$



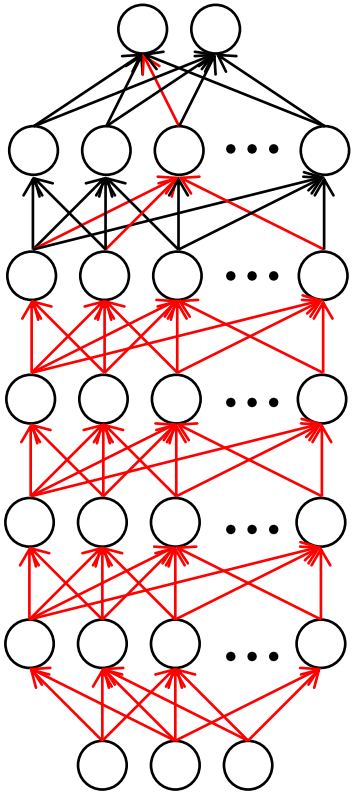
NN研究の歴史


第1期
「冬の時代」
第2期
「冬の時代」
第3期
1960
1970
1980
1990
2000
2010
Perceptron
[Rosenblatt57]
Simple/complex cells
[Hubel-Wiesel59]
“Linearly separable
patterns”
[Minsky-Papert]
Neo-cognitron
[Fukushima80]
Back-propagation
[Rumelhart+]
Convolutional NN
[LeCun+89]
Sparse coding
[Olfhausen-Field96]
“Layerwise pretraining”
[Hinton+06]

DNNの過学習とその克服

Deep Neural Network

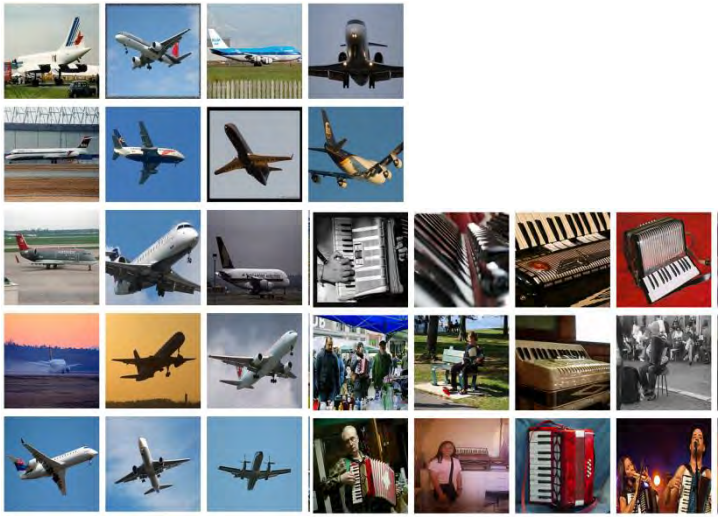
- DNN → 過学習が起こる
  - 訓練誤差は小さいが汎化誤差は大
- 現象
  - 下層まで情報が伝わらない；勾配が拡散；訓練データが上位層のみで表現可能
  - 全結合型のNNで顕著
- 過学習の克服
  1. 学習最適化の工夫
  2. ネットの構造の工夫
  3. データを増やす



DNN成功例：一般物体認識

Krizhevsky et al., ImageNet Classification with Deep Convolutional Neural Networks, NIPS2012

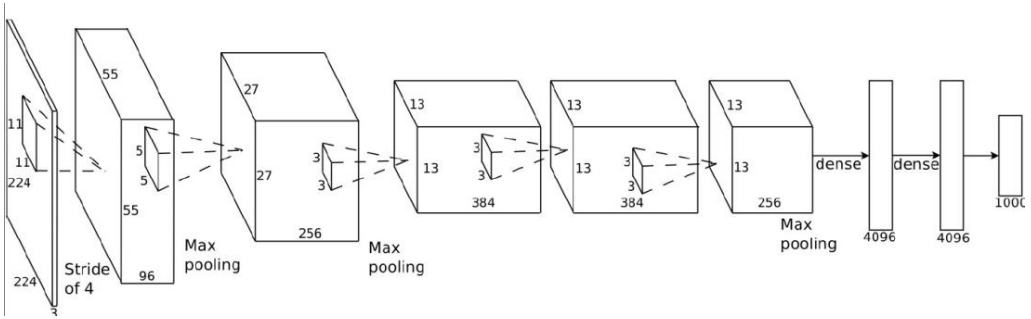
- IMAGENET Large Scale Visual Recognition Challenge 2012
  - 1000カテゴリ・カテゴリあたり約1000枚の訓練画像
  - たたみこみニューラルネット；rectified linear unit；drop-out



“airliner”

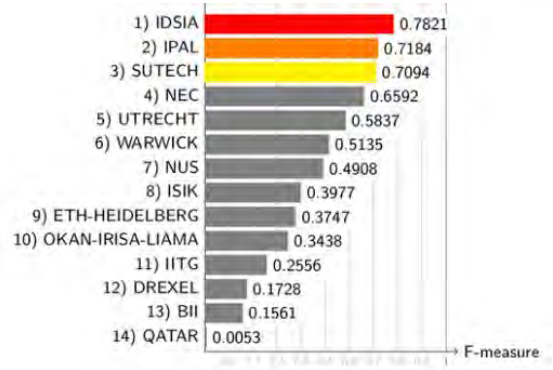
“accordion …”

	Team name	Error (5 guesses)
1	SuperVision	0.15315
2	ISI	0.26172
3	OXFORD_VGG	0.26979
4	XRCE/INRIA	0.27058
5	University of Amsterdam	0.29576
6	LEAR-XRCE	0.34464

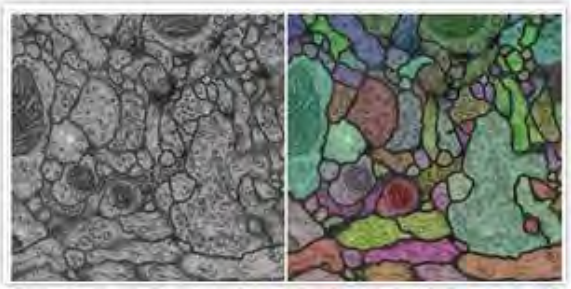


DNN成功例：文字・画像認識（Schmidhuberのグループ@IDSIA）

- コンテスト1位
  - IJCNN 2011 Traffic Sign Recognition Competition; 1st (0.56%), 2nd (1.16%, Humans), 3rd (1.69%), 4th (3.86%)
  - ICPR 2012 Contest on “Mitosis Detection in Breast Cancer Histological Images”
  - ISBI 2012 challenge on segmentation of neuronal structures
  - ICDAR 2011 Offline Chinese Handwriting Competition
  - ICDAR2009
    - Arabic Connected Handwriting Competition
    - French Connected Handwriting Competition



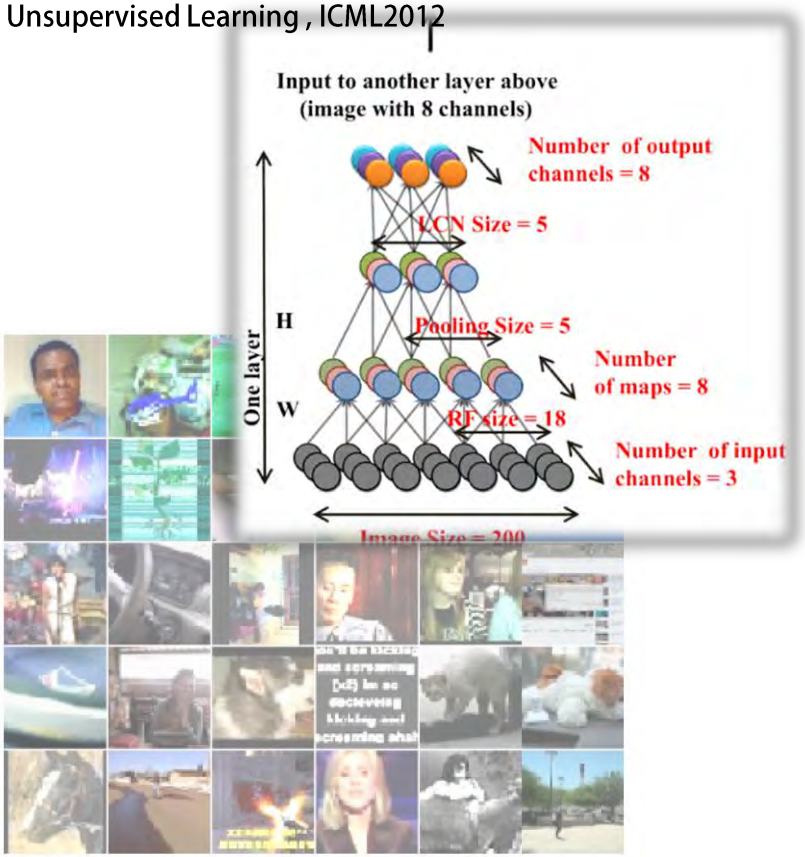
- 認識率最高位（2012年11月時点）
  - NORB object recognition benchmark
  - CIFAR image classification benchmark
  - MNIST handwritten digits benchmark; “human-competitive result”

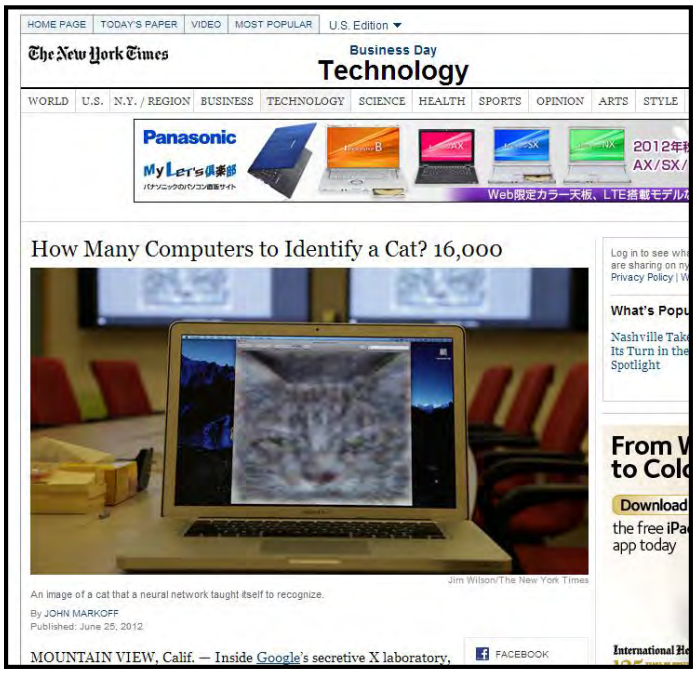


DNN成功例：画像特徴の無教師学習（「グーグルの猫」）

Le et al., Building High-level Features Using Large Scale Unsupervised Learning , ICML2012

- 9層NNを使った無教師学習
  - パラメータ数10億個！
  - 16コアPC1000台のPCクラスタ×3日間
  - YouTubeの画像1000万枚
- 「おばあさん細胞」の生成を確認





The New York Times (2012/6/25)

顔

人の体

概要

- ディープラーニング＝画像認識のパラダイム変化？
- 中核的方法
  - 教師なし：（スパース）オートエンコーダ
  - 教師あり：たたみこみネット
- たたみこみネットの周辺
- その他事例
- まとめ

概要

- ディープラーニング＝画像認識のパラダイム変化？
- 中核的方法
  - 教師なし：（スパース）オートエンコーダ
  - 教師あり：たたみこみネット
- たたみこみネットの周辺
- その他事例
- まとめ

画像認識技術の現状

実用レベル

研究途上

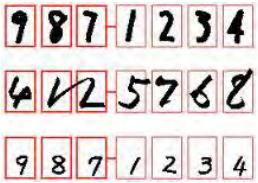
1970 1980 1990 2000 2010 ???

バーコード

印刷文字

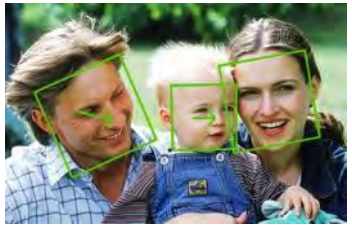
指紋

手書き文字



ナンバープレート

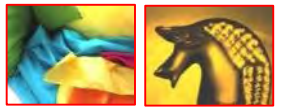
顔



道路標識

一般物体

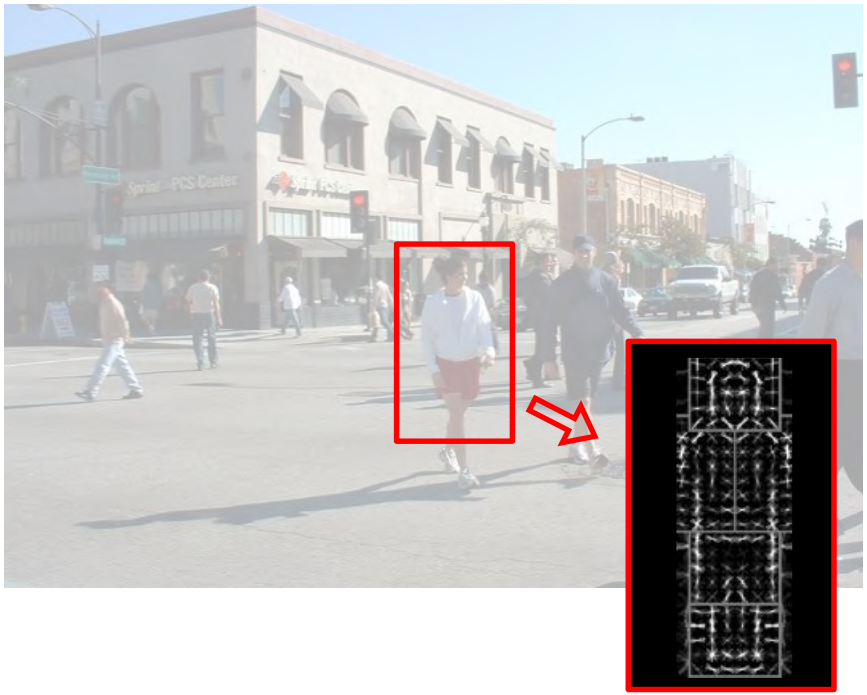
材質・質感

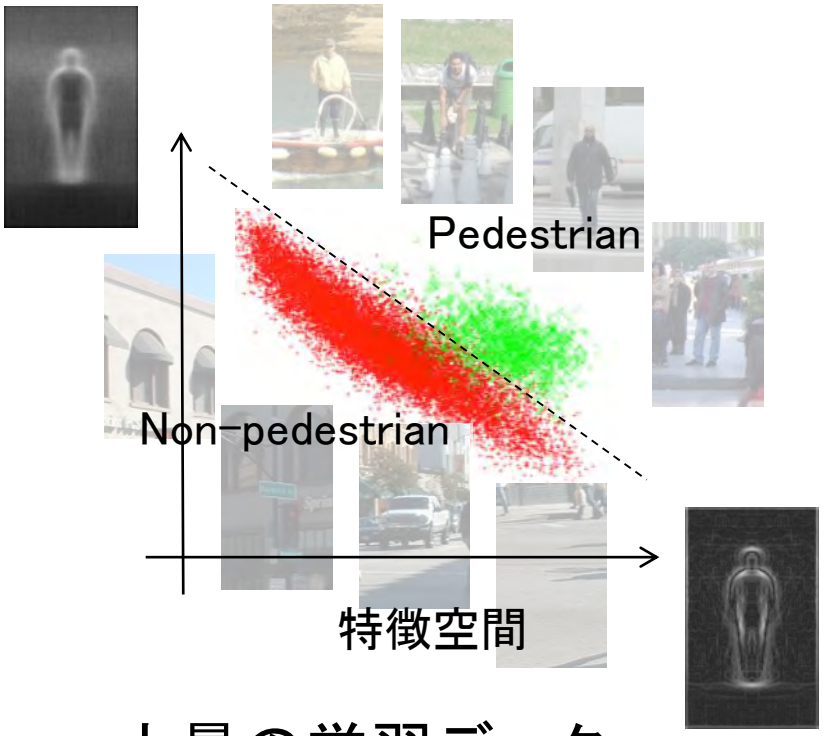


人の行動

画像認識のプロセス

画像 → 特徴抽出 → 特徴量 → 分類 → 結果





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

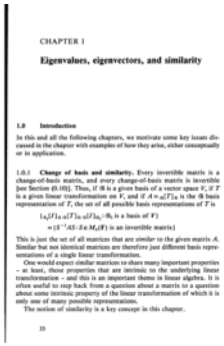
実用レベル

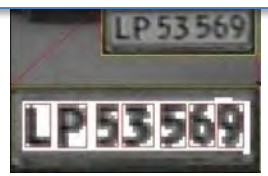
研究途上

1970 1980 1990 2000 2010 ???

バーコード

印刷文字





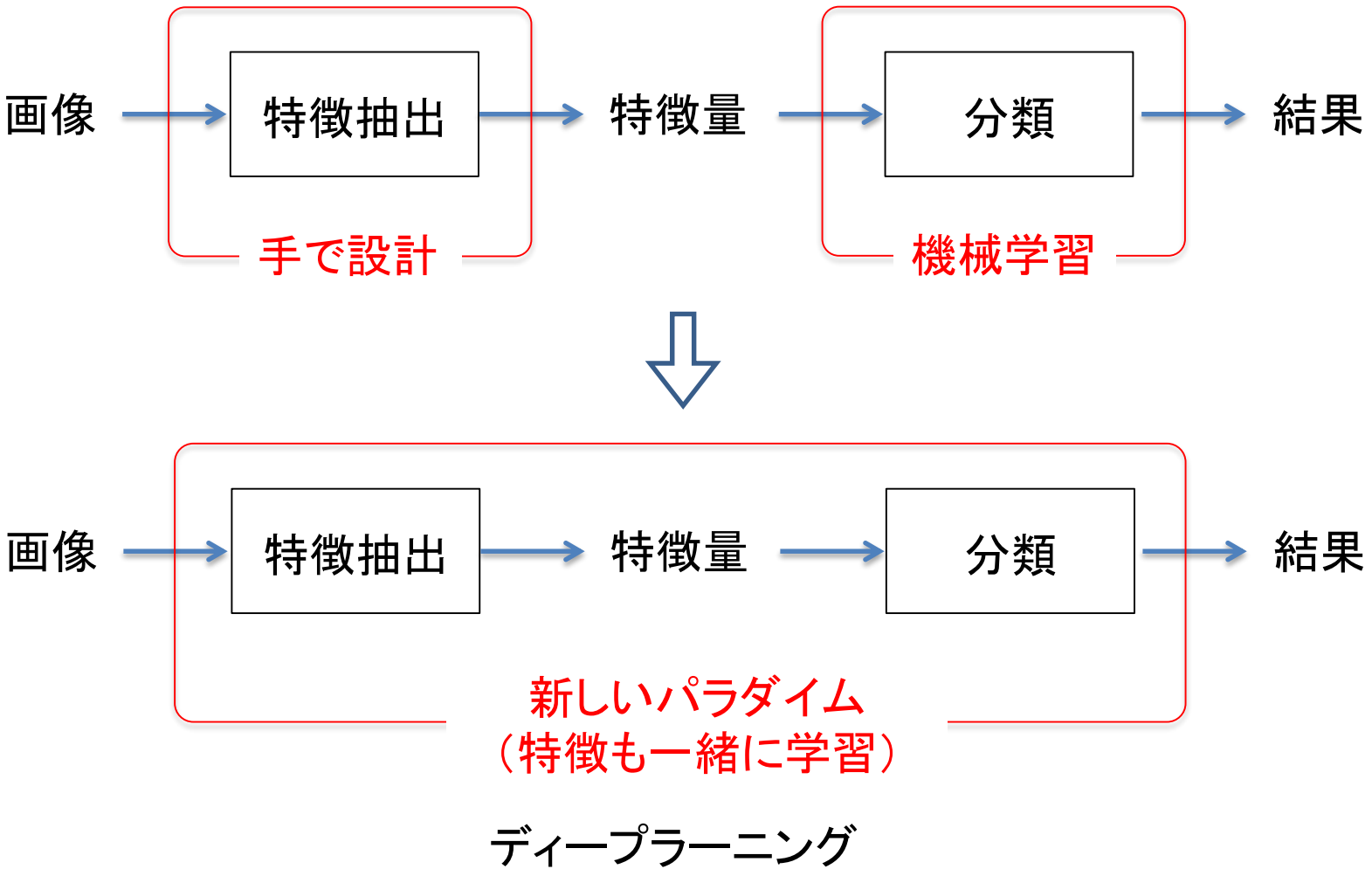
一般物体

感

人の行動

画像から何を取り出すか？
（どんな特徴）

パラダイムの変化：特徴の設計から特徴の学習へ

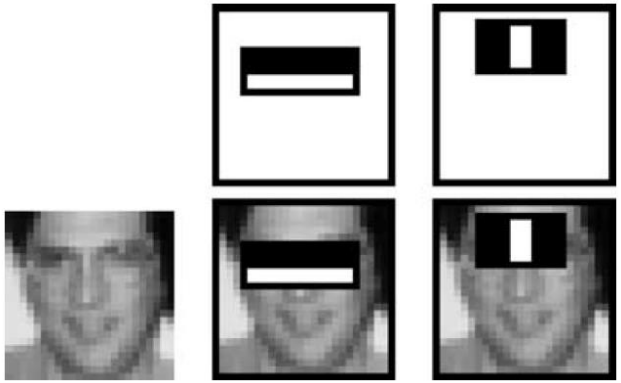


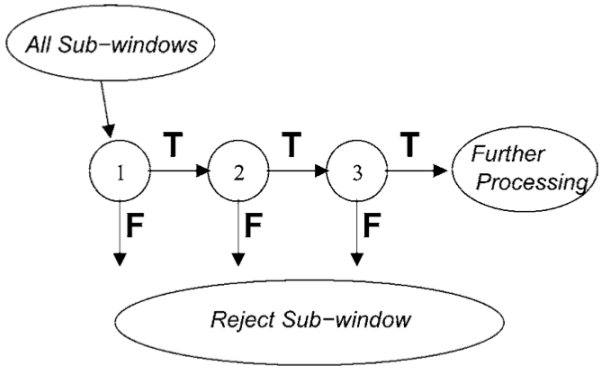
ディープラーニング

特徴選択

ブースティング：顔画像

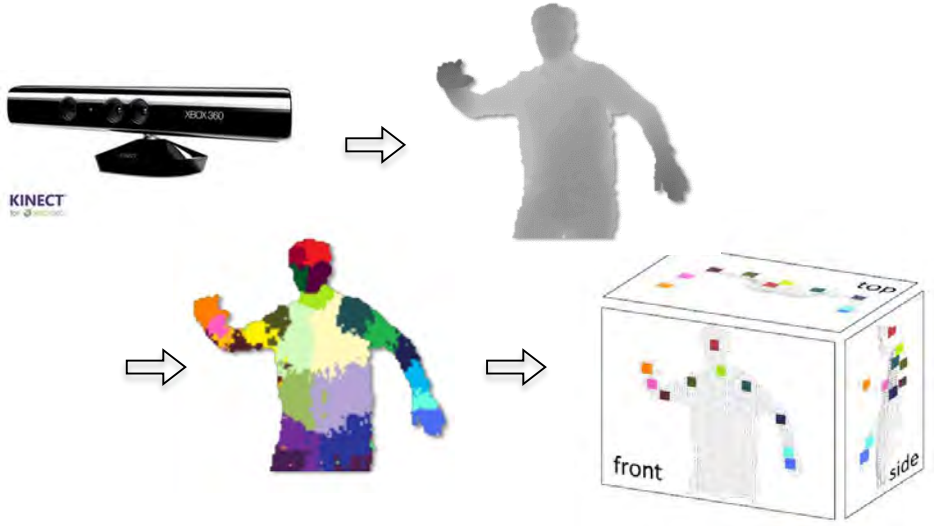
[Viola-Jones04]

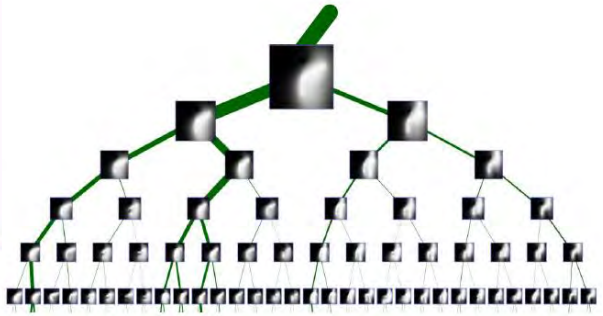




RandomForest：モーションキャプチャ

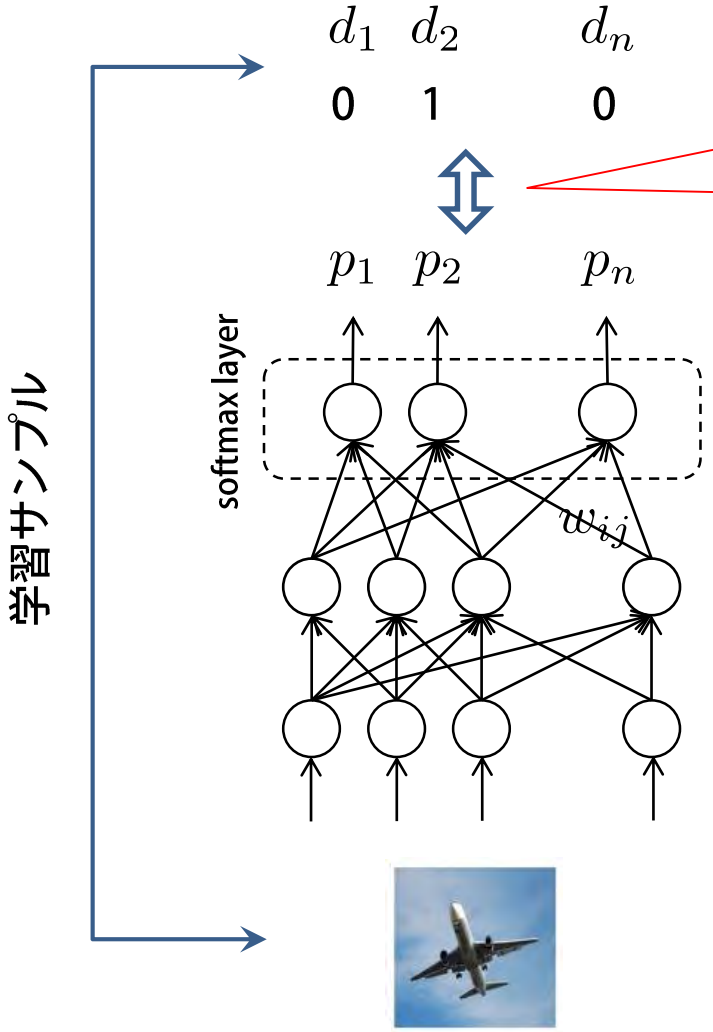
[Shotton+11]





画像（物体カテゴリー）認識の場合

- 出力の誤差が小さくなるように重み $\{w_{ij}\}$ 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

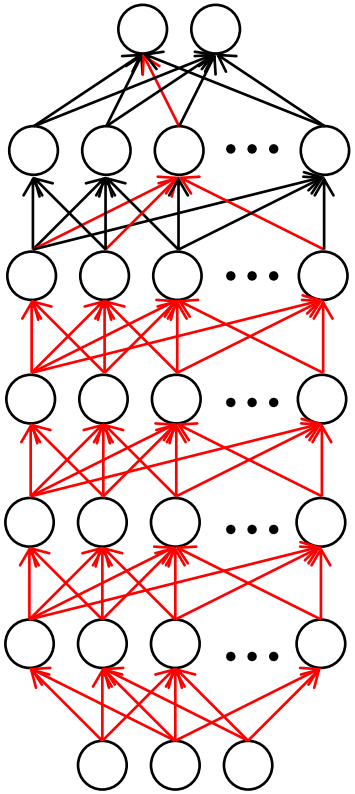
勾配降下法（GD:Gradient Descent）

概要

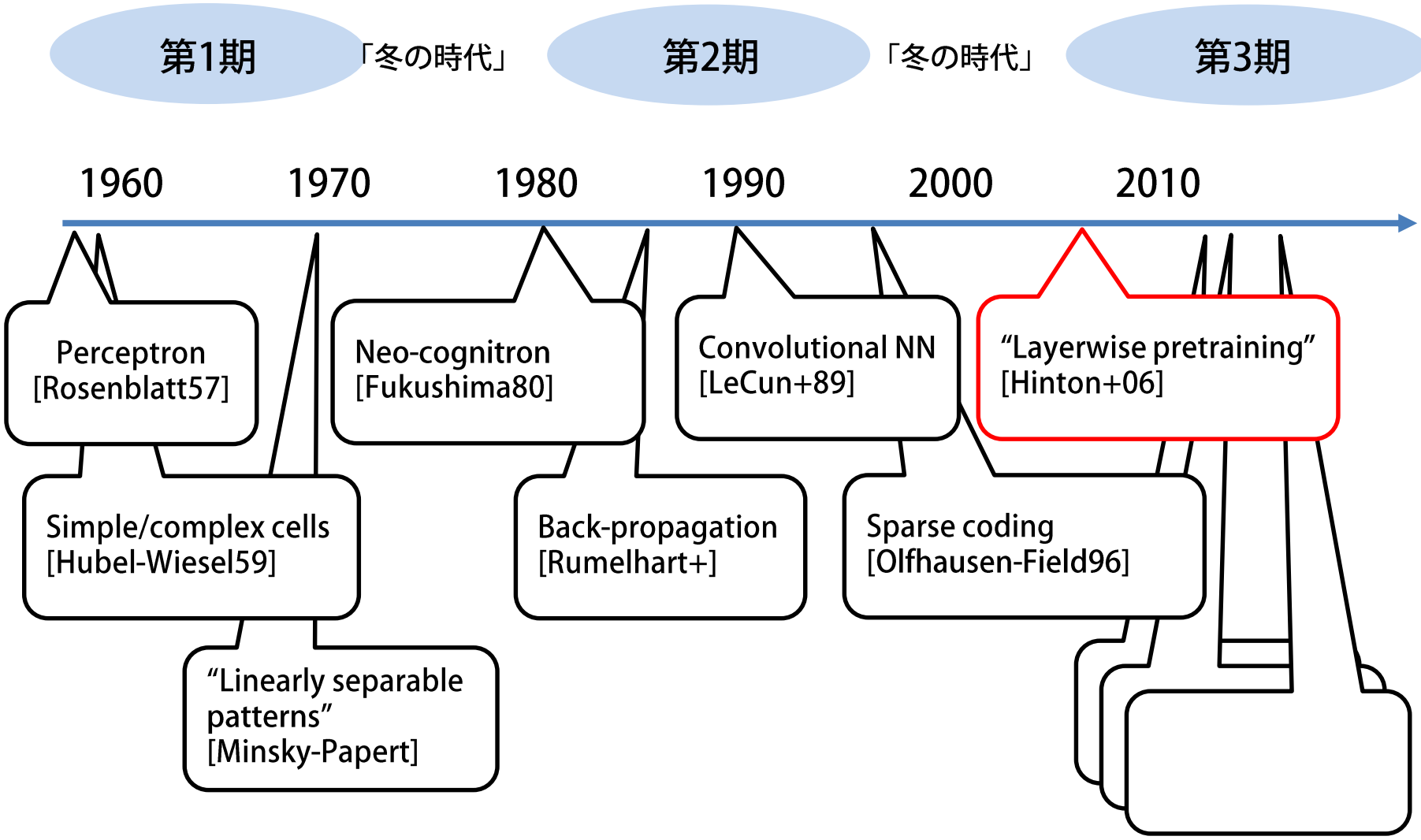
- ディープラーニング＝画像認識のパラダイム変化？
- 中核的方法
  - 教師なし：（スパース）オートエンコーダ
  - 教師あり：たたみこみネット
- たたみこみネットの周辺
- その他事例
- まとめ

多層ネットの過学習とその克服

- 多層ネット → 過学習が起こる
  - 訓練誤差は小さいが汎化誤差は大
- 現象
  - 下層まで情報が伝わらない；勾配が拡散；訓練データが上位層のみで表現可能
  - 全結合型のNNで顕著
- 過学習の克服
  1. 学習最適化の工夫
  2. ネットの構造の工夫
  3. データを増やす

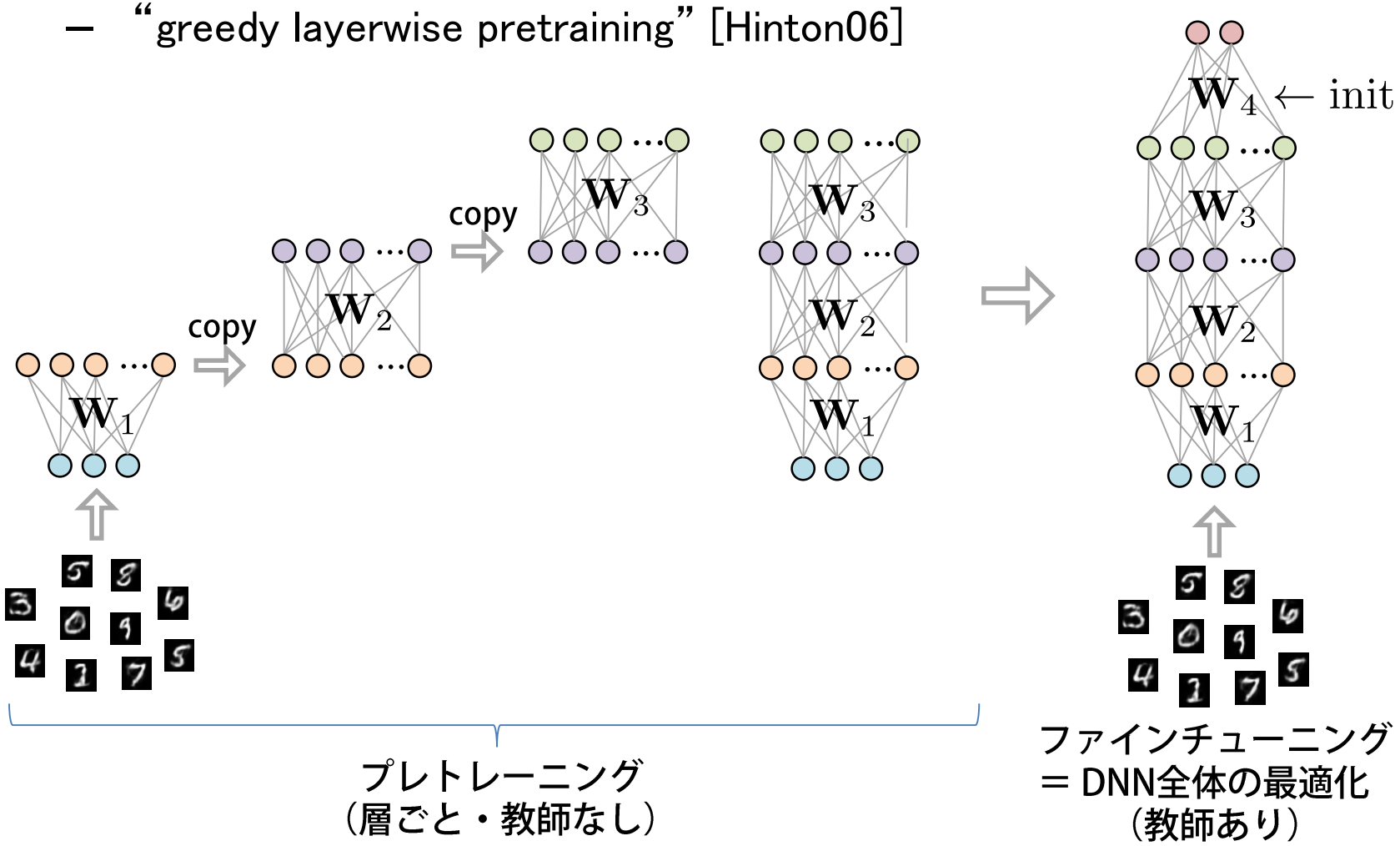


NN研究の歴史


第1期
「冬の時代」
第2期
「冬の時代」
第3期
1960
1970
1980
1990
2000
2010
Perceptron
[Rosenblatt57]
Simple/complex cells
[Hubel-Wiesel59]
“Linearly separable
patterns”
[Minsky-Papert]
Neo-cognitron
[Fukushima80]
Back-propagation
[Rumelhart+]
Convolutional NN
[LeCun+89]
Sparse coding
[Olfhausen-Field96]
“Layerwise pretraining”
[Hinton+06]

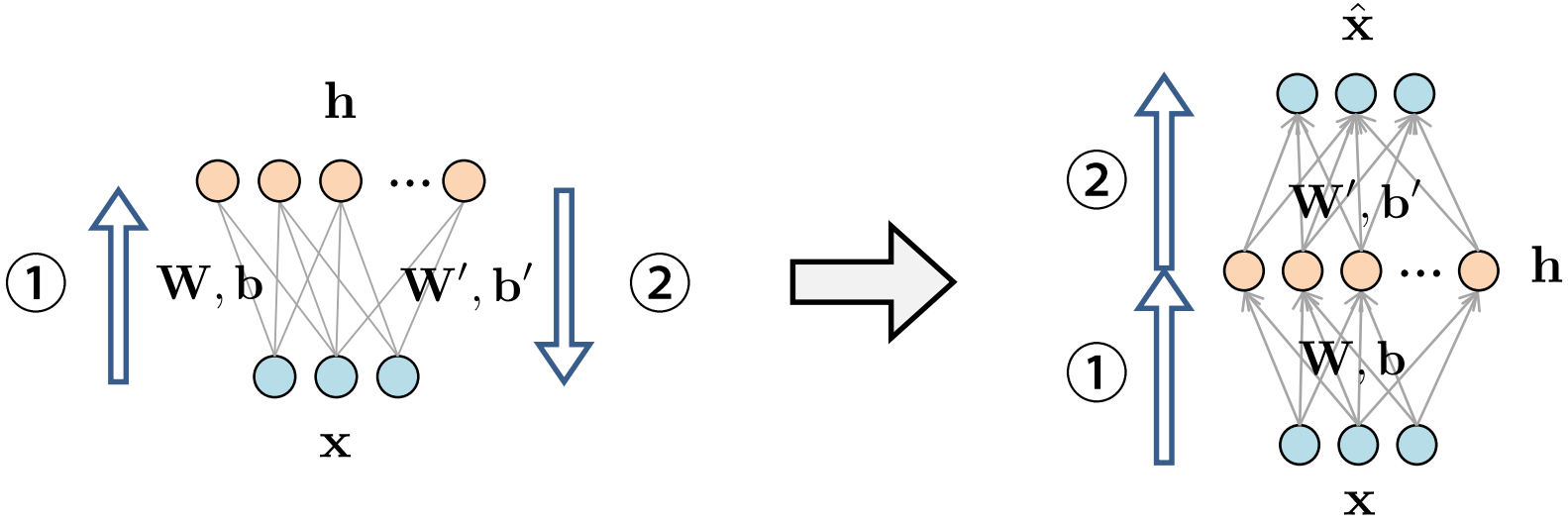
DNNのプレトレーニング

- 層ごとにオートエンコーダを学習 → 過学習を克服
  - “greedy layerwise pretraining” [Hinton06]



スパースオートエンコーダ

- 入力サンプルをよく再現するように
  - BPで or ボルツマンマシンとして学習
  - 中間層がスパースに活性化するように正則化を行う



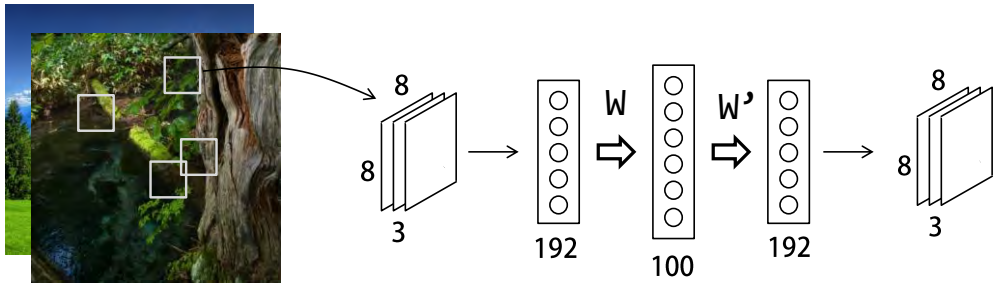
① $\mathbf{h}(\mathbf{x}_i) = f(\mathbf{W}\mathbf{x}_i + \mathbf{b})$

② $\hat{\mathbf{x}}(\mathbf{x}_i) = f'(\mathbf{W}'\mathbf{h}(\mathbf{x}_i) + \mathbf{b}')$

$$\min_{\theta} \sum_i \|\mathbf{x}_i - \hat{\mathbf{x}}(\mathbf{x}_i)\|^2 + g(\theta)$$

$$\theta = [\mathbf{W}, \mathbf{b}, \mathbf{W}', \mathbf{b}']$$

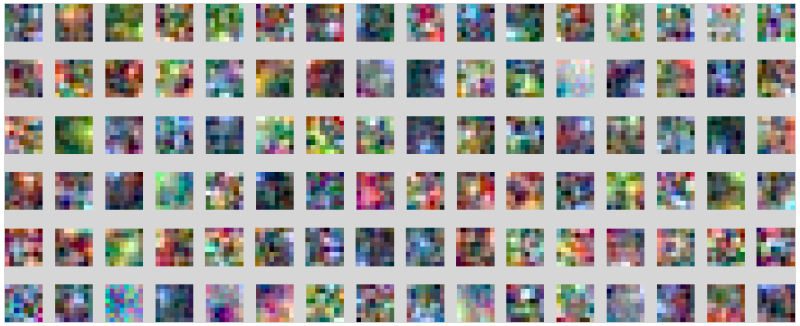
局所画像特徴の学習とスパース正則化の重要性



スパース正則化あり

スパース正則化なし

特徴
（W）



再現

オリジナル
（入力パッチ）

物体認識
正答率
（10カテゴリ）

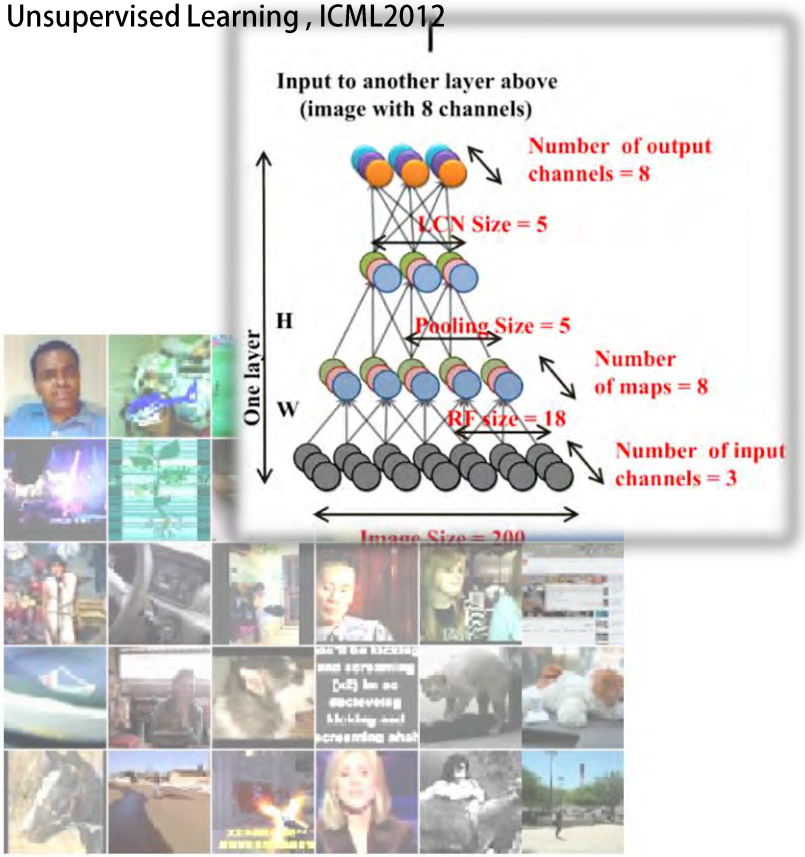
60%

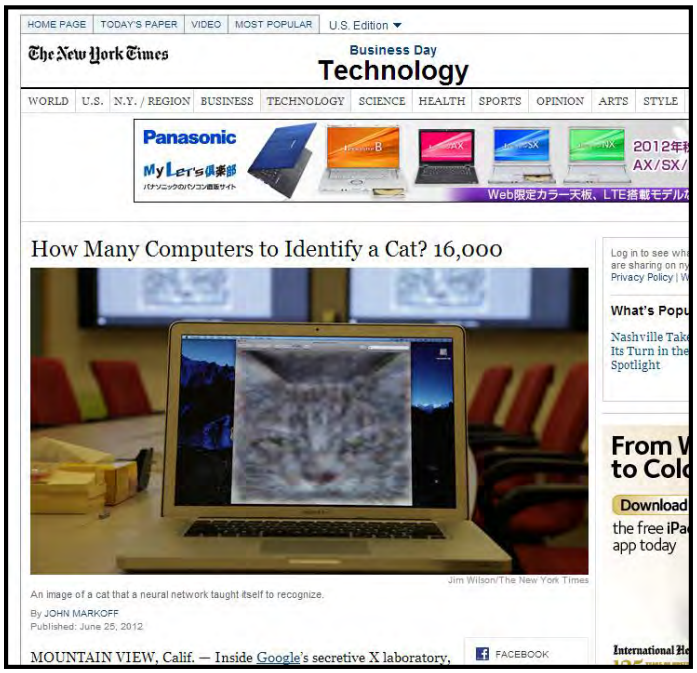
45%

DNN成功例：画像特徴の無教師学習（「グーグルの猫」）

Le et al., Building High-level Features Using Large Scale Unsupervised Learning , ICML2012

- 9層NNを使った無教師学習
  - パラメータ数10億個！
  - 16コアPC1000台のPCクラスタ×3日間
  - YouTubeの画像1000万枚
- 「おばあさん細胞」の生成を確認





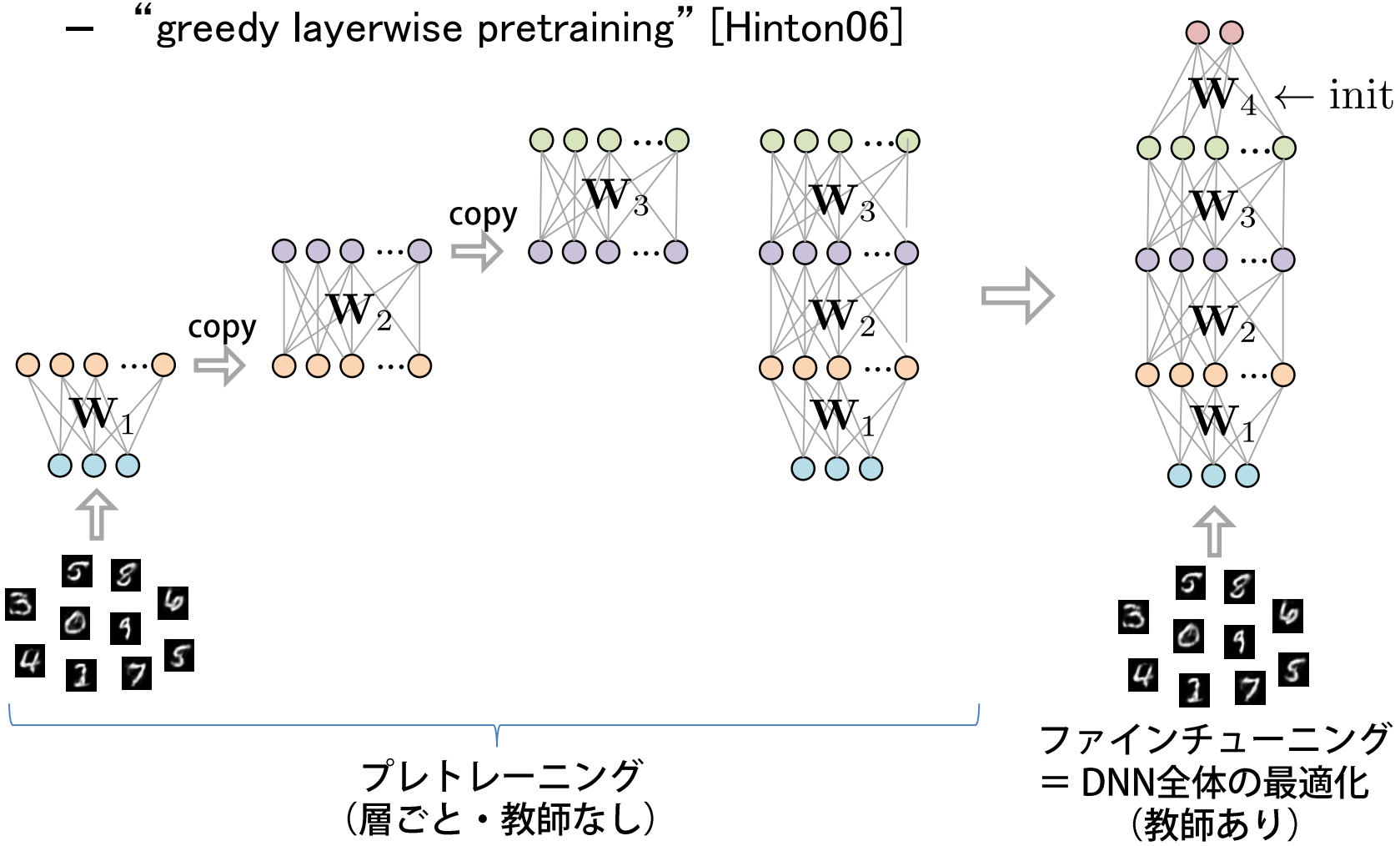
The New York Times (2012/6/25)

顔

人の体

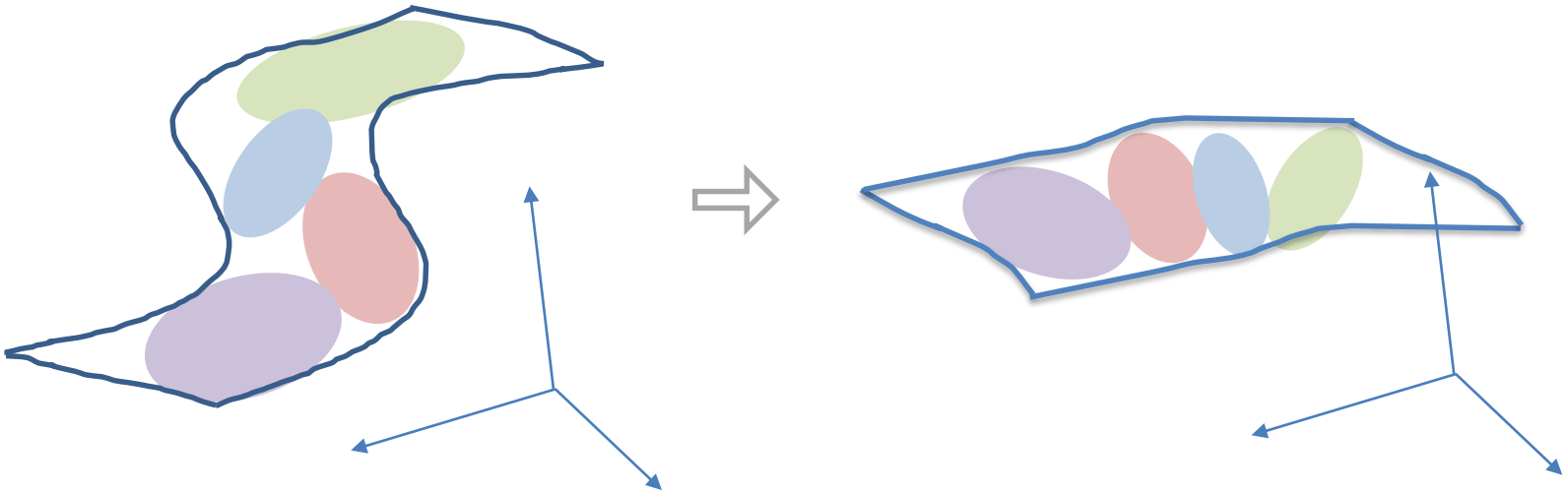
DNNのプレトレーニング

- 層ごとにオートエンコーダを学習 → 過学習を克服
  - “greedy layerwise pretraining” [Hinton06]



プレトレーニングはなぜ有効か

- データを正確に表現する特徴は識別にも役立つ
  - すべてが有効なわけではないとしても
- よい初期値を与えると同時に一種の正則化として機能 [Larochelle+09]
  - 正則化＝汎化や分布した内部表現を促進

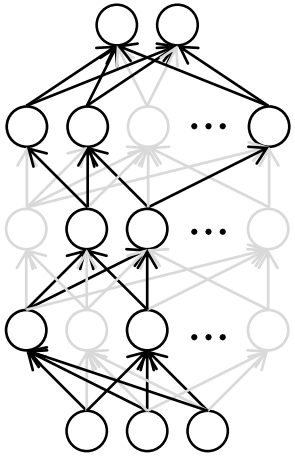


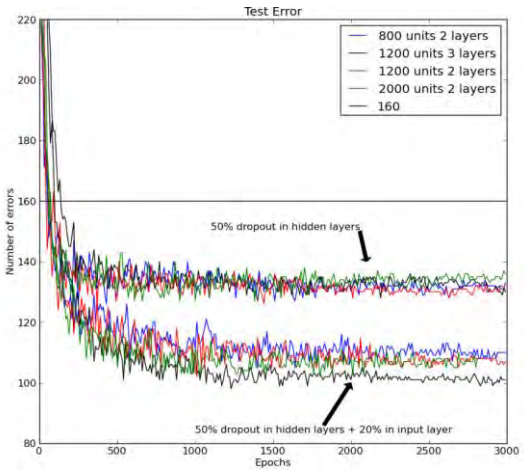
Larochelle et al., Exploring Strategies for Training Deep Neural Networks, JMLR, 2009
Hinton et al., Deep Neural Networks for Acoustic Modeling in Speech Recognition, IEEE SP magazine, Nov. 2012

第3の方法（全結合NNの学習）

ドロップアウト [Hinton+12]

- ランダムに隠れユニットを省いて学習

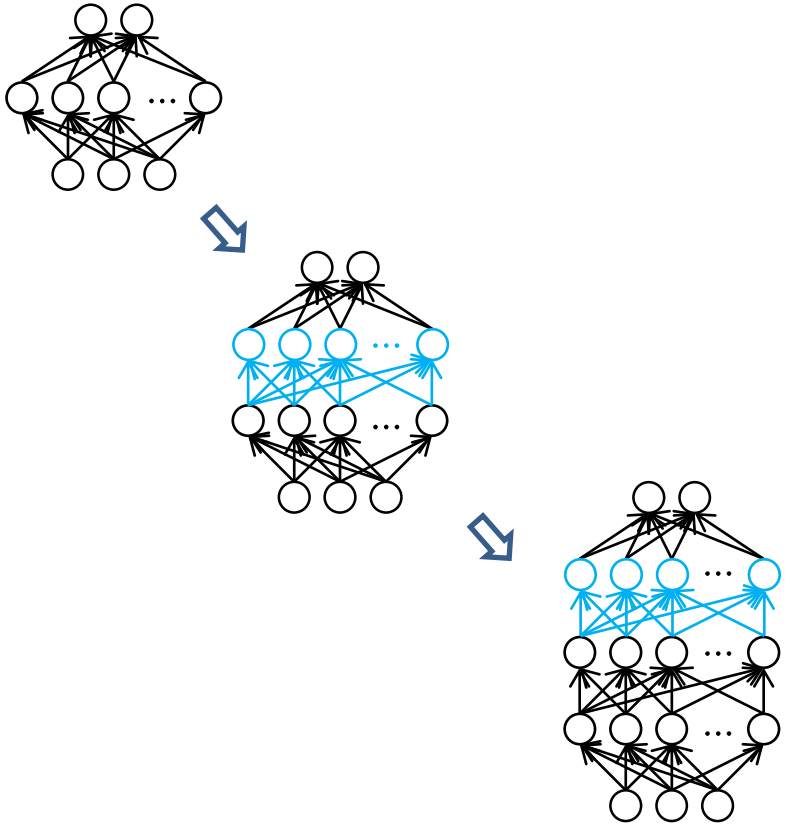




MNISTでの結果

識別的プレトレーニング [Seide+11]

- 浅い方から深いネットへ，教師あり学習を反復



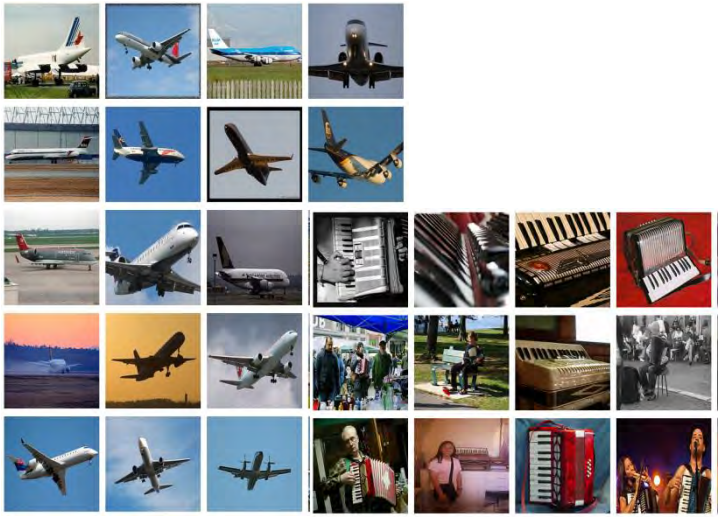
概要

- ディープラーニング＝画像認識のパラダイム変化？
- 中核的方法
  - 教師なし：（スパース）オートエンコーダ
  - 教師あり：たたみこみネット
- たたみこみネットの周辺
- その他事例
- まとめ

DNN成功例：一般物体認識

Krizhevsky et al., ImageNet Classification with Deep Convolutional Neural Networks, NIPS2012

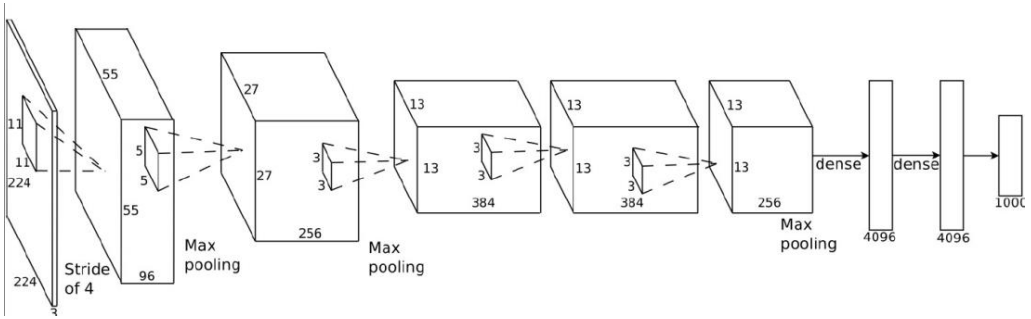
- IMAGENET Large Scale Visual Recognition Challenge 2012
  - 1000カテゴリ・カテゴリあたり約1000枚の訓練画像
  - たたみこみニューラルネット；rectified linear unit；drop-out



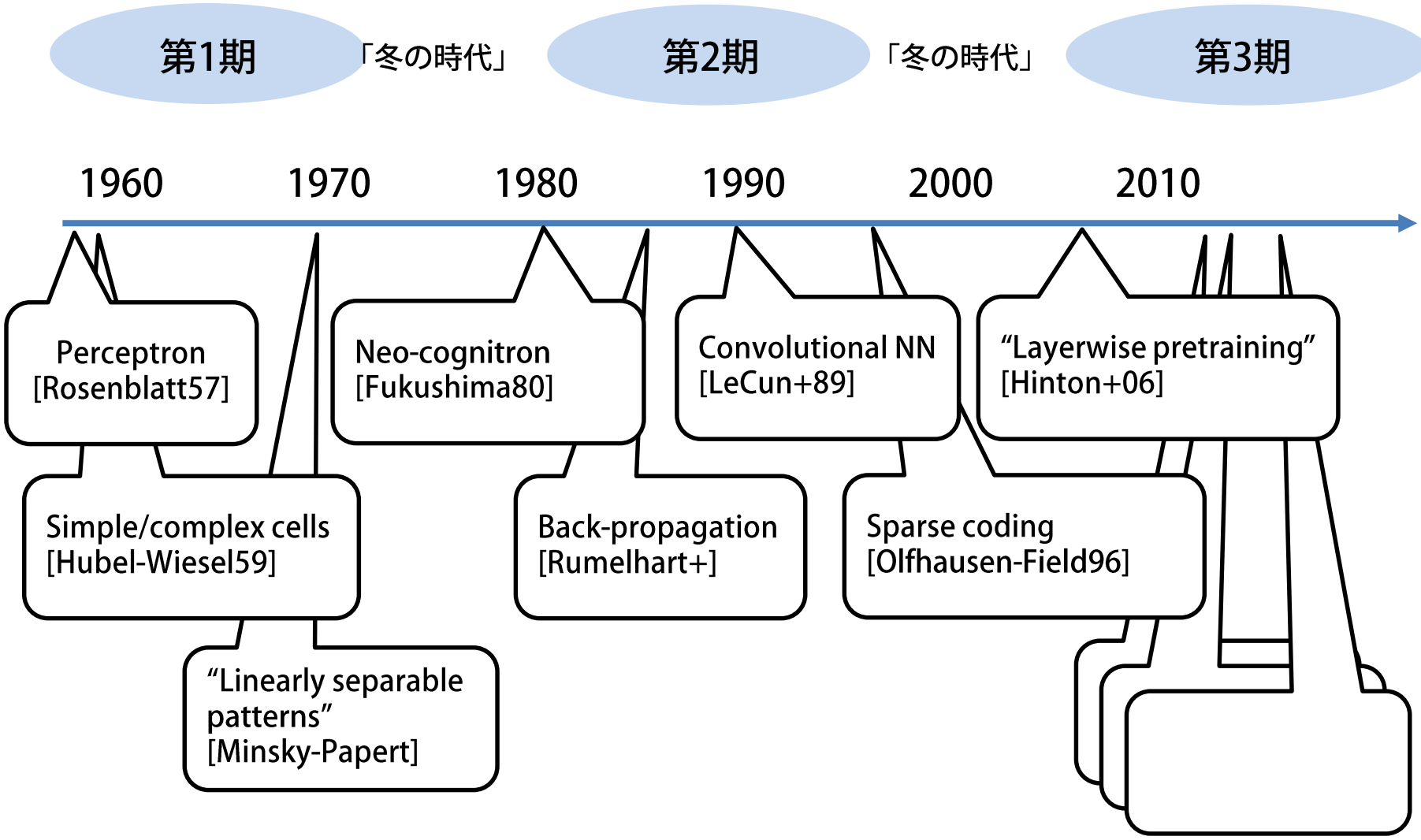
“airliner”

“accordion …”

	Team name	Error (5 guesses)
1	SuperVision	0.15315
2	ISI	0.26172
3	OXFORD_VGG	0.26979
4	XRCE/INRIA	0.27058
5	University of Amsterdam	0.29576
6	LEAR-XRCE	0.34464



NN研究の歴史


第1期
「冬の時代」
第2期
「冬の時代」
第3期
1960
1970
1980
1990
2000
2010
Perceptron
[Rosenblatt57]
Simple/complex cells
[Hubel-Wiesel59]
“Linearly separable
patterns”
[Minsky-Papert]
Neo-cognitron
[Fukushima80]
Back-propagation
[Rumelhart+]
Convolutional NN
[LeCun+89]
Sparse coding
[Olfhausen-Field96]
“Layerwise pretraining”
[Hinton+06]

たたみこみニューラルネット
Convolutional Neural Network (CNN)

- Neocognitronにルーツ [Fukushima80]
- Backpropによる教師有学習と手書き文字認識への応用 [LeCun+89]
  - Backpropagation Applied to Handwritten Zip Code Recognition, *Neural Computation*, 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)

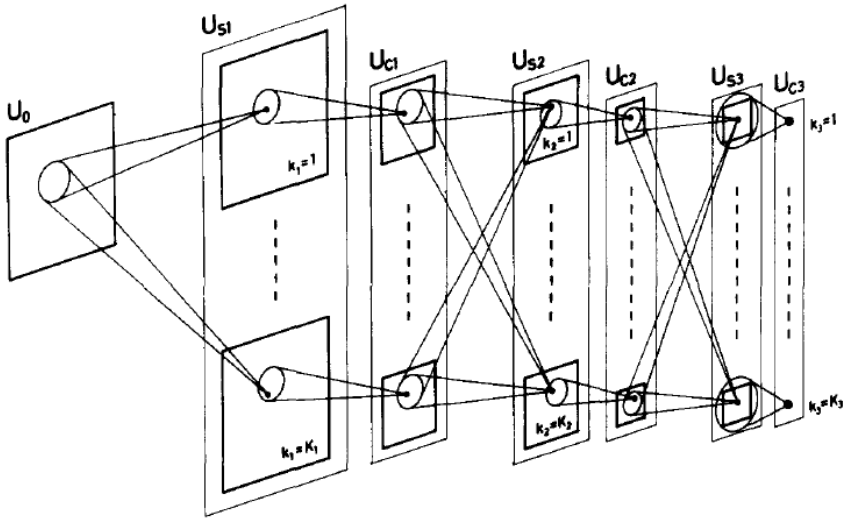


Fig 4 Schematic diagram illustrating the interconnections between layers in the neocognitron

[Fukushima+83]

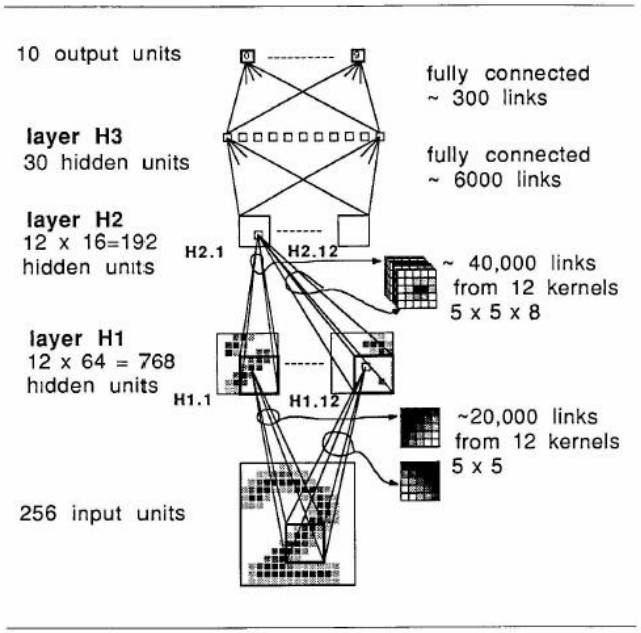
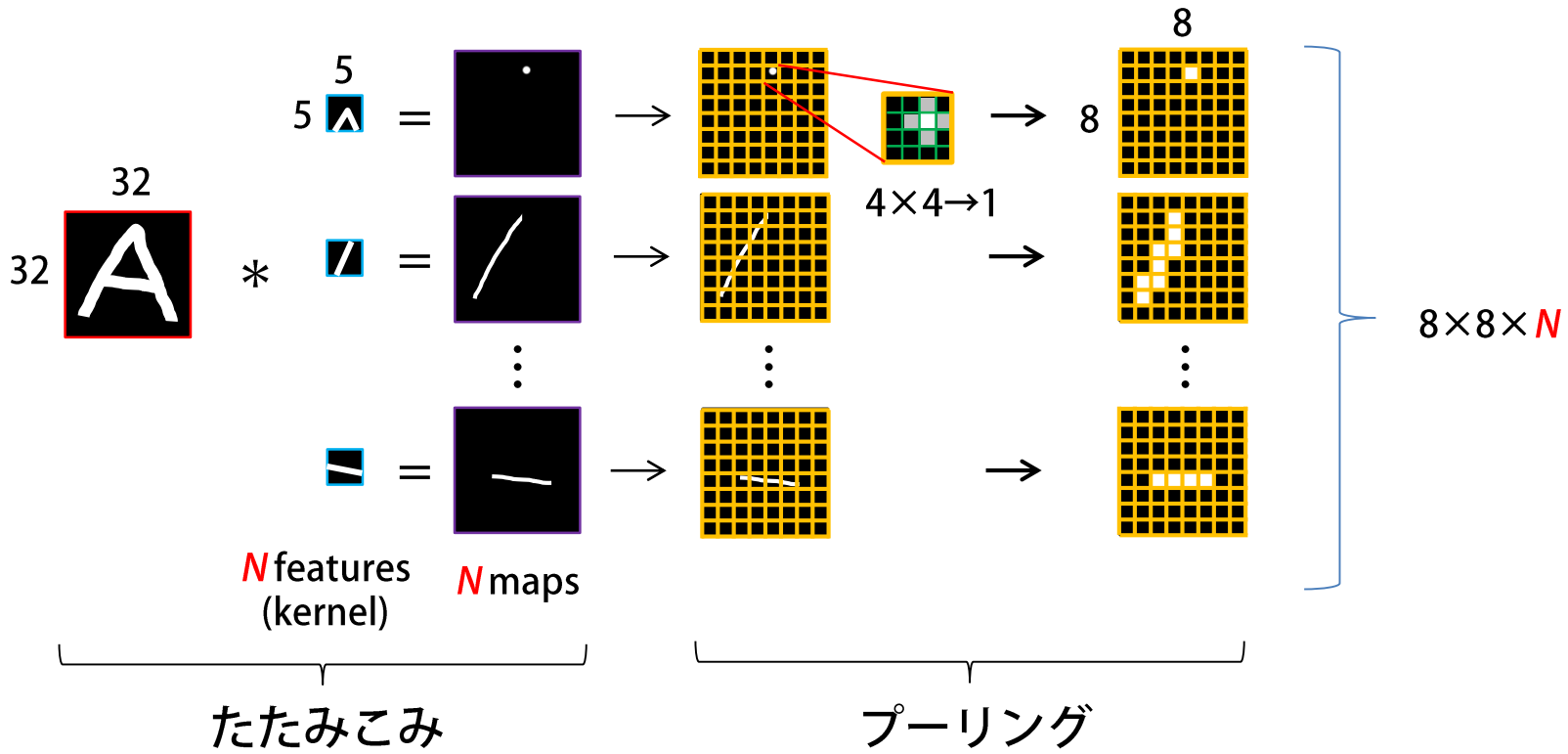
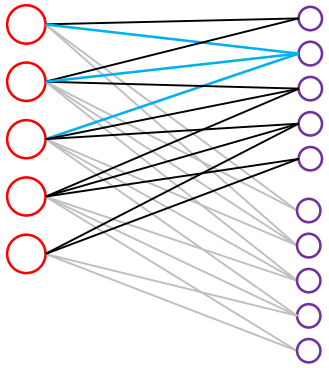


Figure 3 Log mean squared error (MSE) (top) and raw error rate (bottom) versus number of training passes

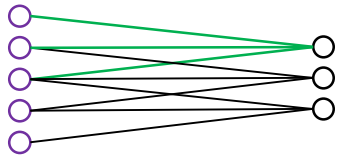
[LeCun+89]

たたみこみニューラルネット



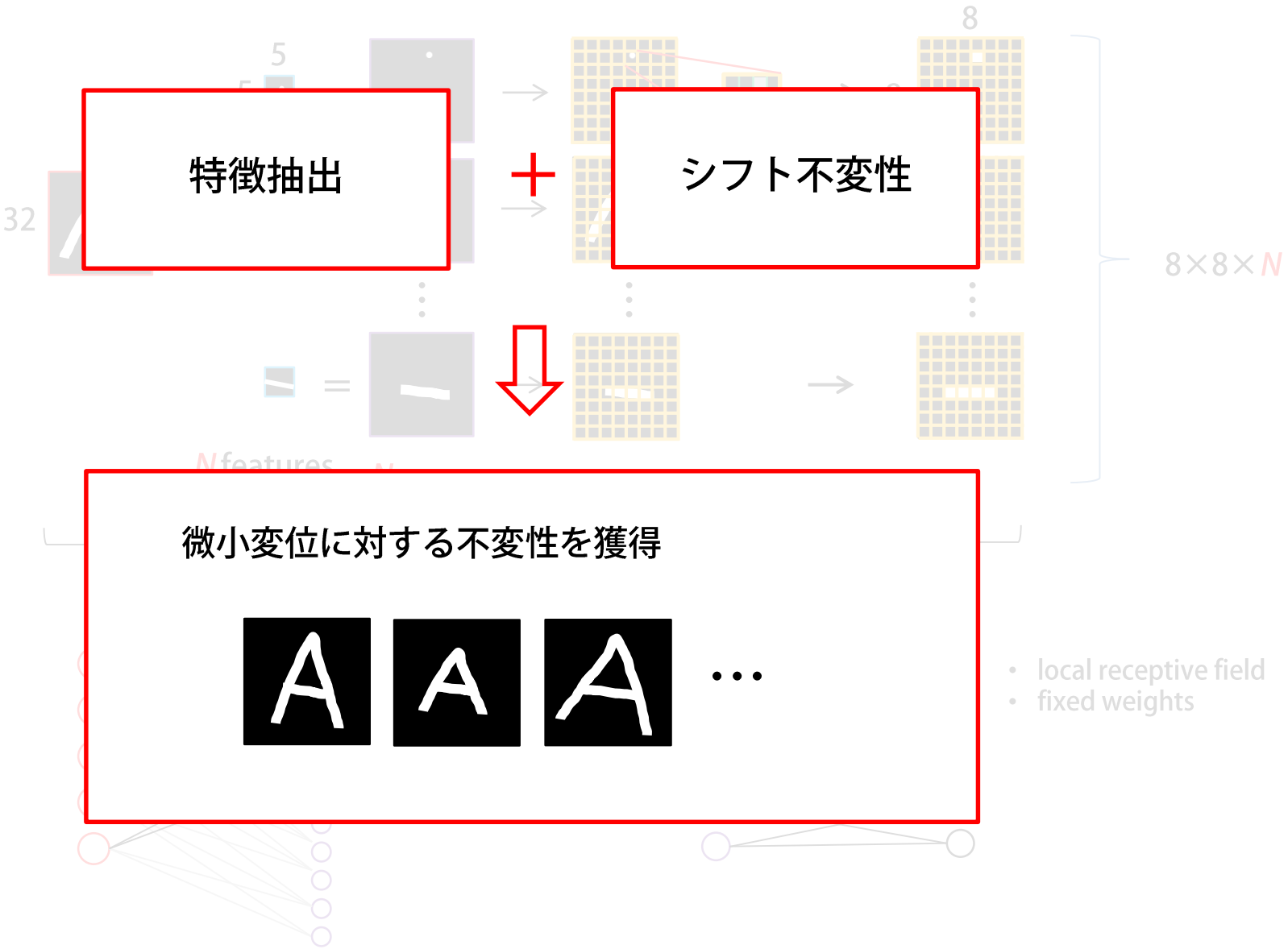


- local receptive field
- tied weights



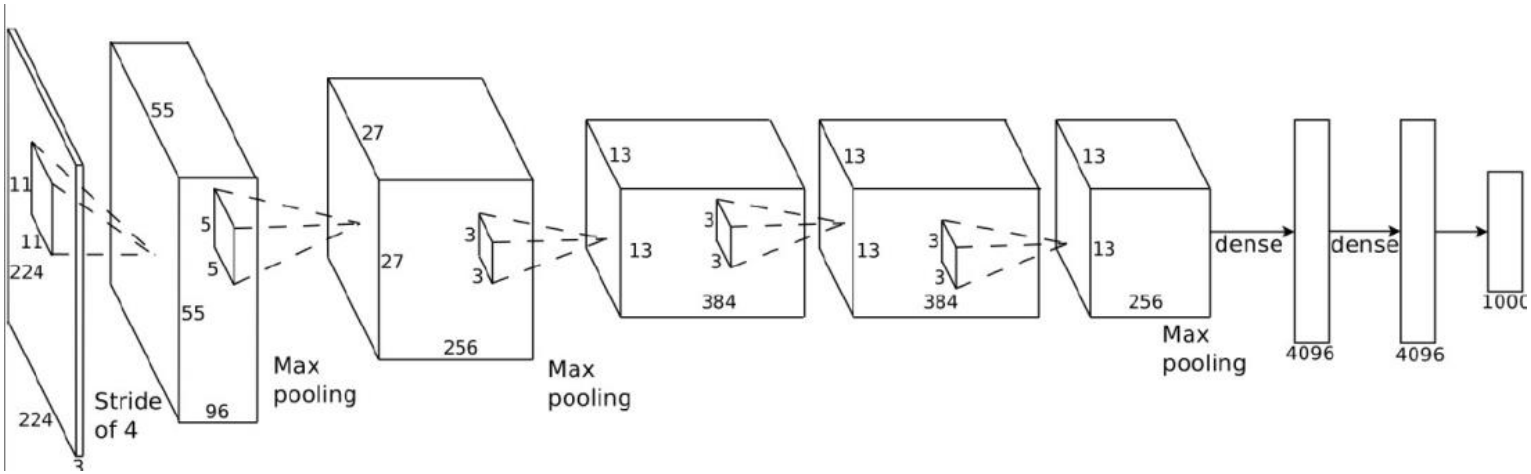
- local receptive field
- fixed weights

たたみこみニューラルネット



たたみこみニューラルネット

- たたみこみ+ プーリングを繰り返す（＝ Deep CNN）ことで，多様な変形に対する不変性を獲得
- フィルタと上位の全結合層を勾配降下法（BP）で学習
  - 全重み（フィルタ）をランダムに初期化

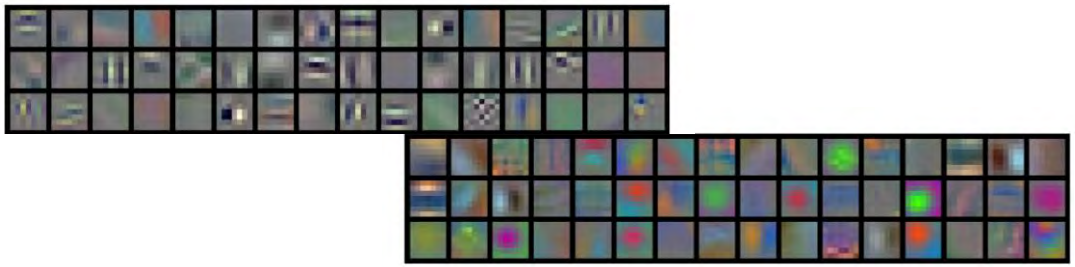


ILSVRC12のCNN [Krizhevsky+12]

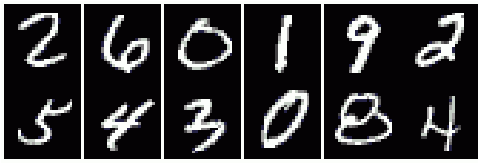
たたみこみネットが学習した特徴

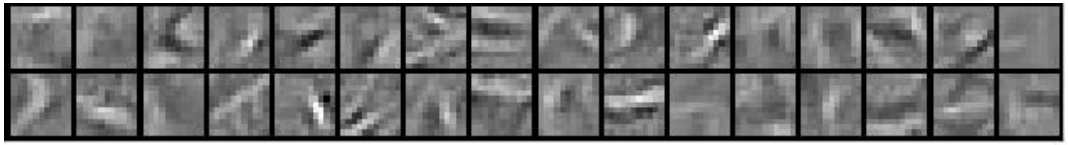
Object category (ImageNet)

第１層で学習された特徴（フィルタ）

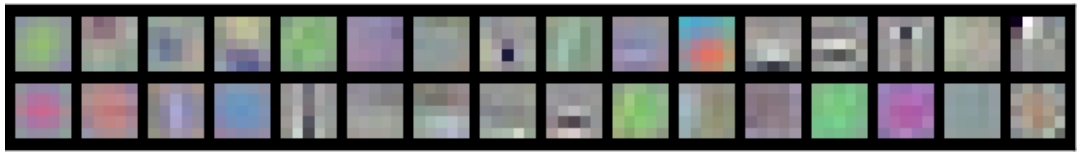


Handwritten digit （MNIST）

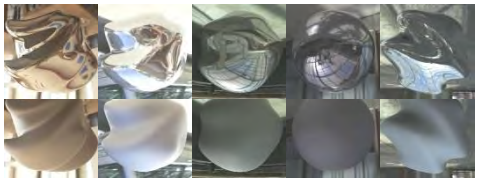


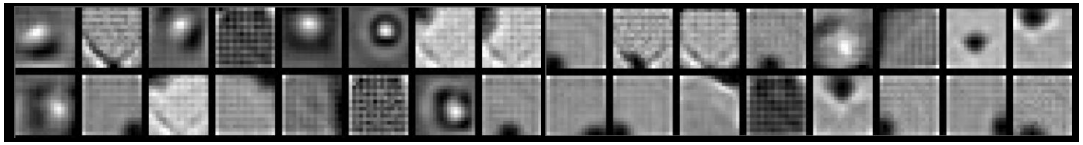


Material （FMD）

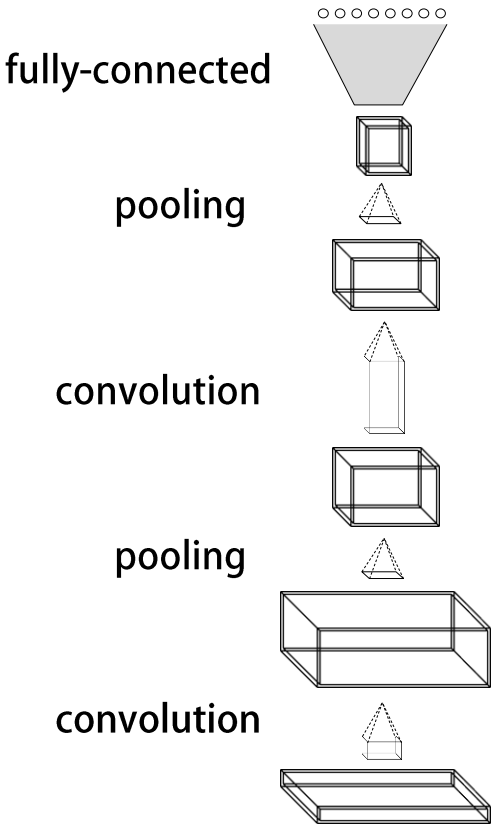


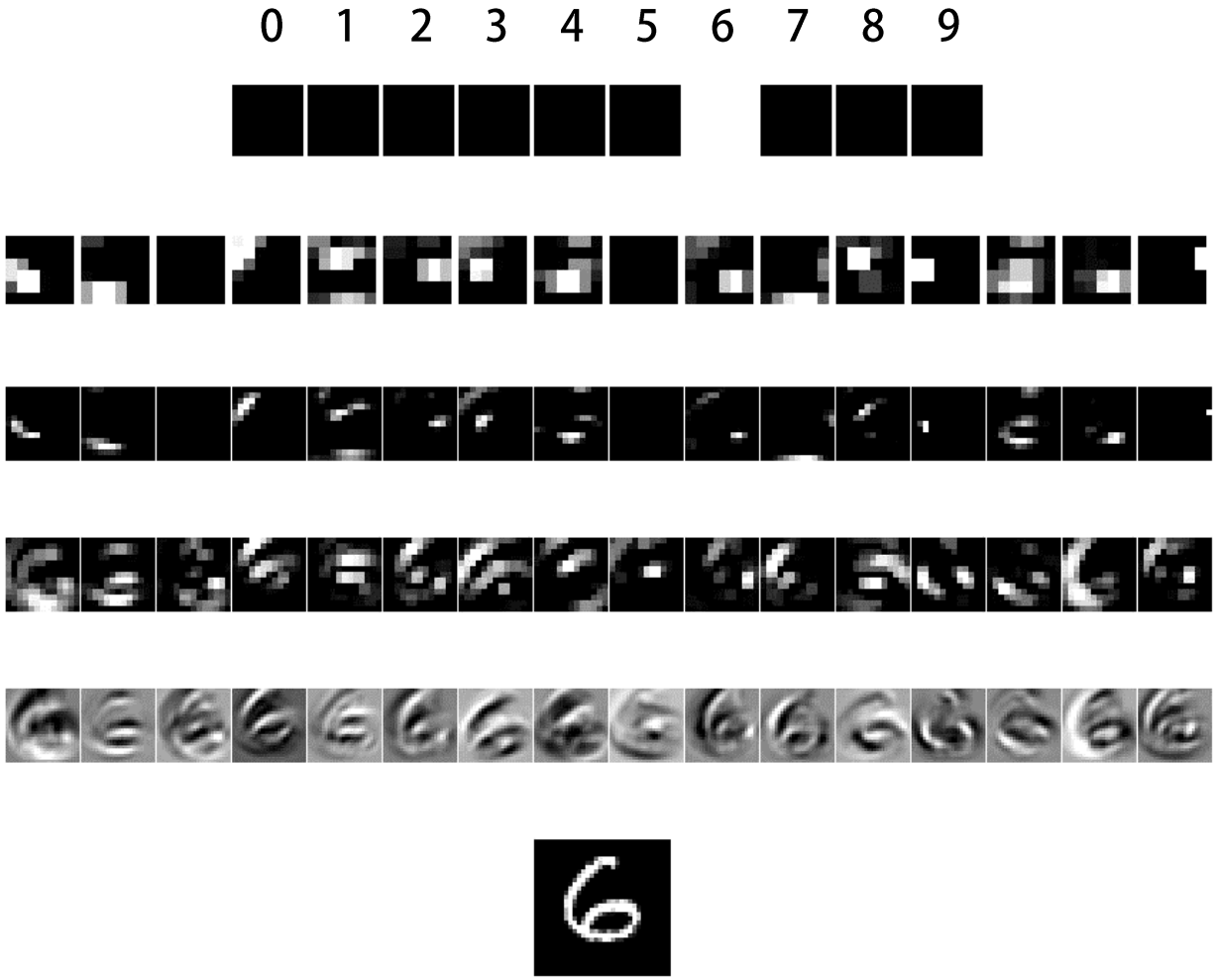
Glossiness



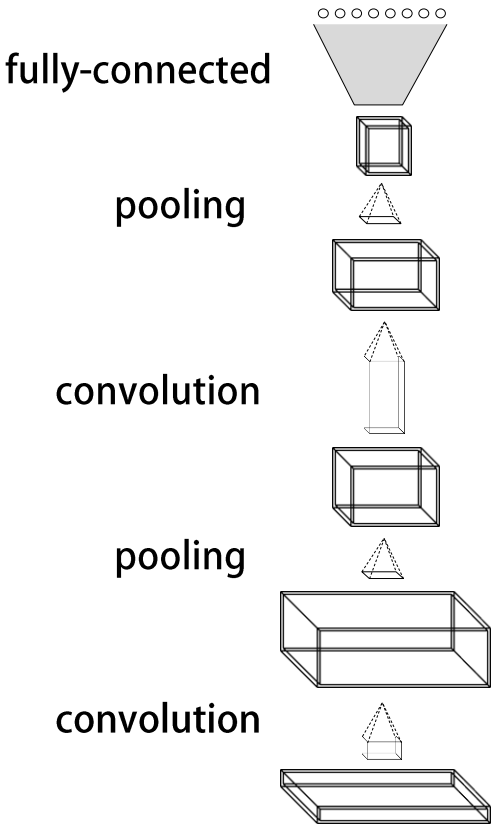


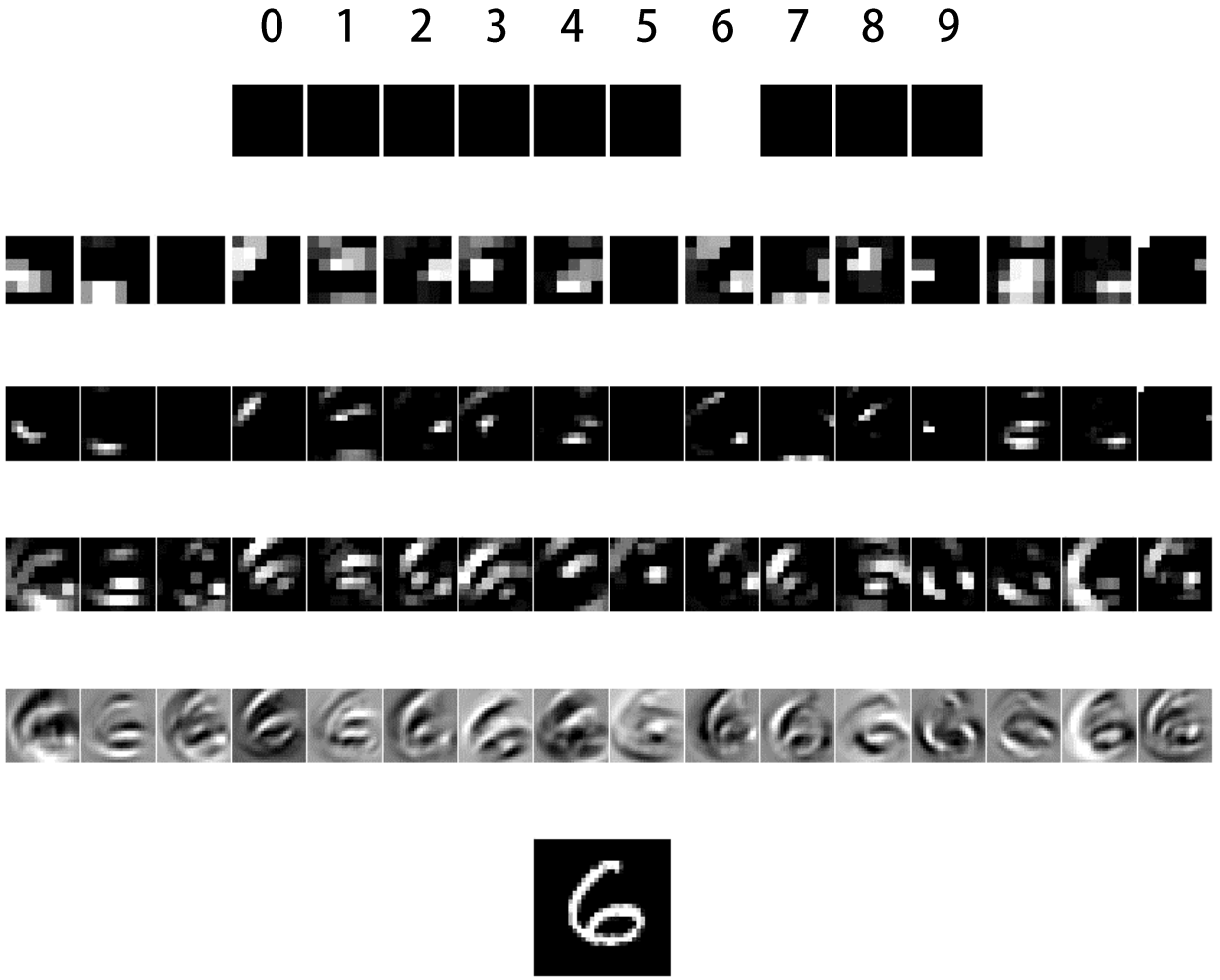
たたみこみネットの振舞い





たたみこみネットの振舞い





概要

- ディープラーニング＝画像認識のパラダイム変化？
- 中核的方法
  - 教師なし：（スパース）オートエンコーダ
  - 教師あり：たたみこみネット
- たたみこみネットの周辺
- その他事例
- まとめ

脳の視覚情報処理

DiCarlo, Zoccolan, Rust, How does the brain solve visual object recognition?, Neuron, 2012

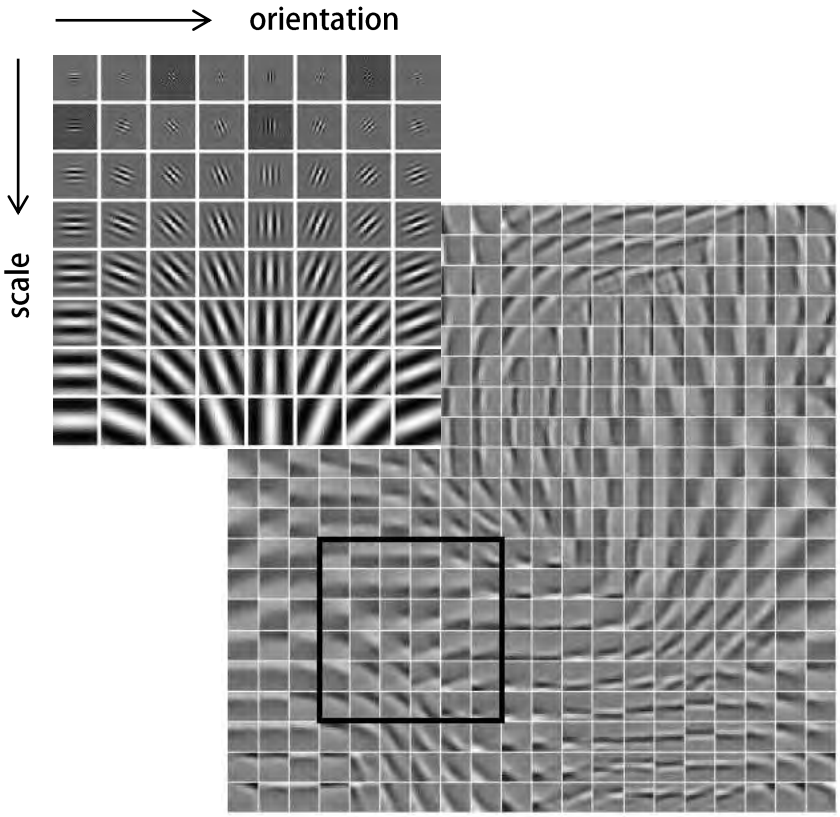
- 視覚野（腹側皮質視覚路 ）の構造
  - フィードフォワードで伝播
  - 階層性：単純な特徴抽出 → 複雑なものへ

（著作権への配慮から図を削除 ）

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

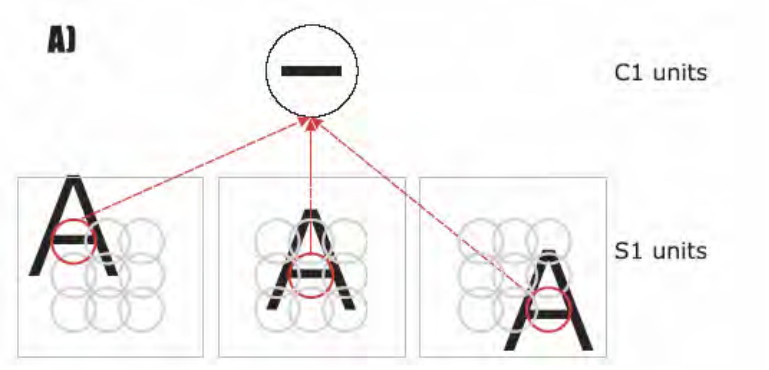
V1と単純細胞・複雑細胞

- ガボールウェーブレット
  - 位置 / 向き / スケール
  - Topographic map



Kavukcuoglu, Ranzato, Fergus, LeCun, Learning Invariant Features through Topographic Filter Maps, CVPR09

- Simple cells/complex cells [Huber-Wiesel59]



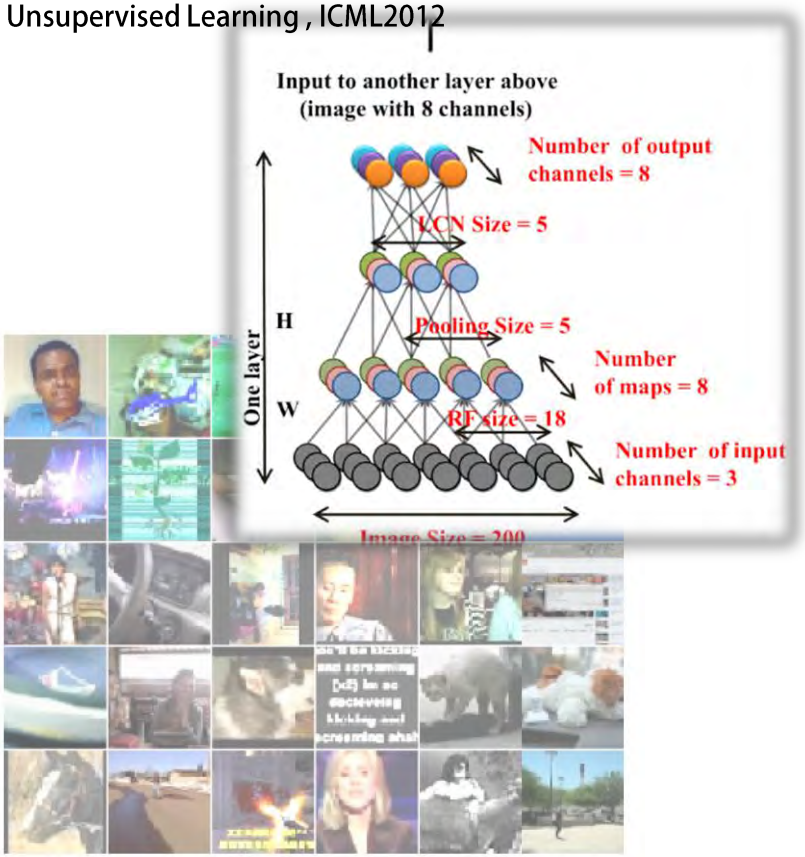
Serre et al, Object Recognition with Features Inspired by Visual Cortex, CVPR05

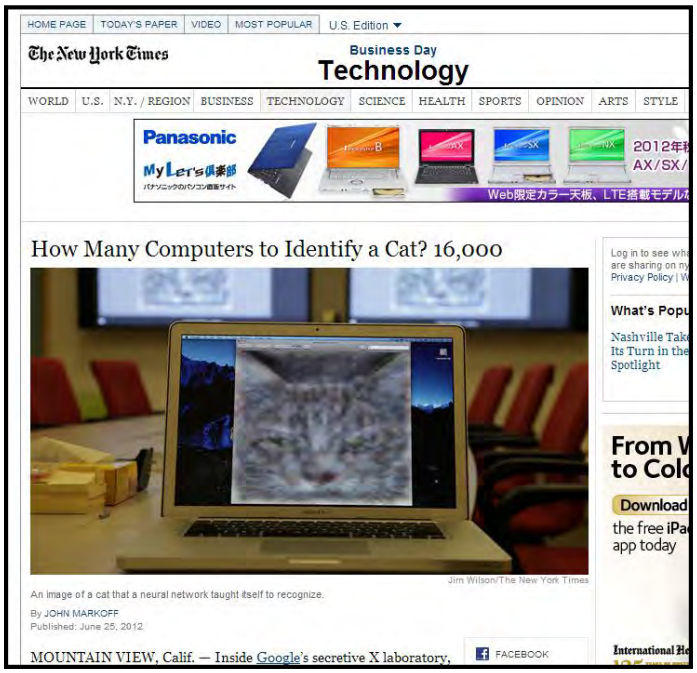
  - Slow feature analysis [Berkes-Wiskott05]
  - Gabor quadrature pair [Jones - Palmer87]

DNN成功例：画像特徴の無教師学習（「グーグルの猫」）

Le et al., Building High-level Features Using Large Scale Unsupervised Learning , ICML2012

- 9層NNを使った無教師学習
  - パラメータ数10億個！
  - 16コアPC1000台のPCクラスタ×3日間
  - YouTubeの画像1000万枚
- 「おばあさん細胞」の生成を確認





The New York Times (2012/6/25)

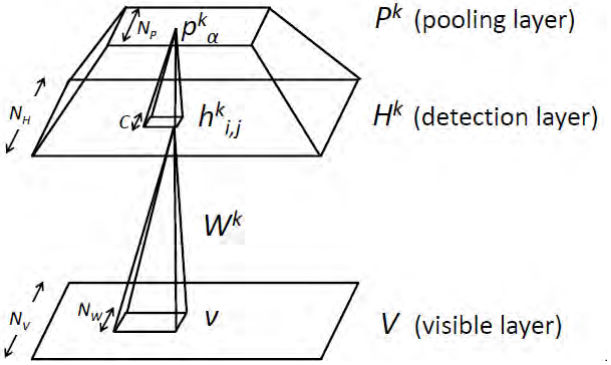
顔

人の体

Convolutional DBN

Lee et al., Convolutional Deep Belief Networks for Scalable Unsupervised Learning of Hierarchical Representations, ICML09

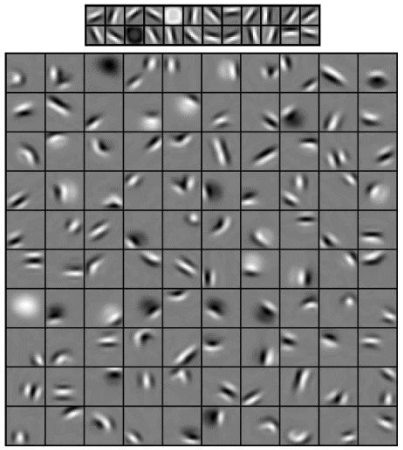
- たたみこみとプーリングを取り入れたDBN



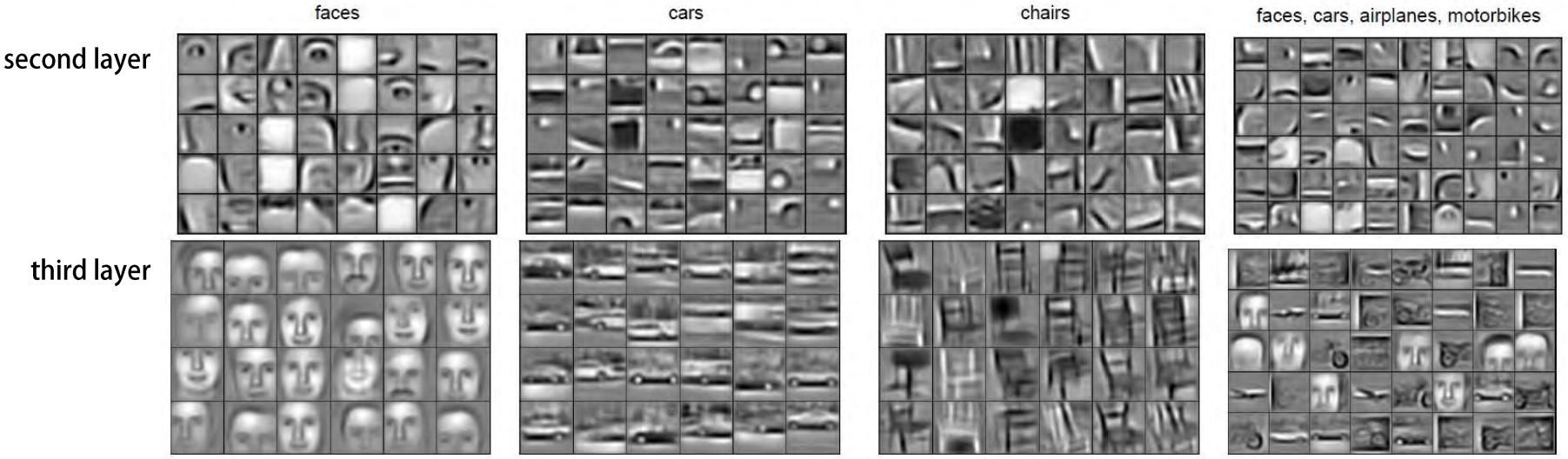
Learned bases for natural scenes:

first layer bases

second layer bases



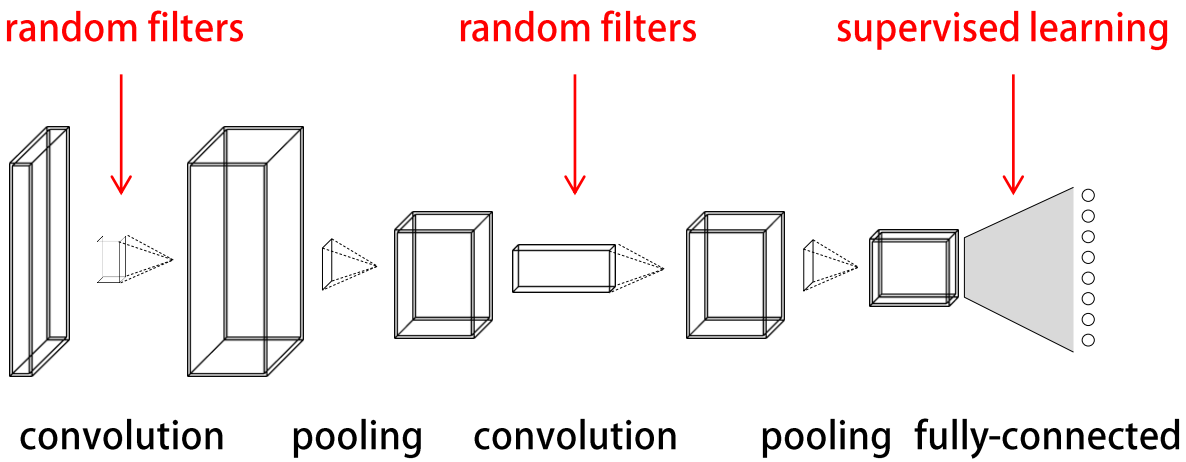
Unsupervised learning of object parts (three layers):



ランダムフィルタ：アーキテクチャの重要性

Jarrett et al., What is the best multi-stage architecture for object recognition? ICCV09

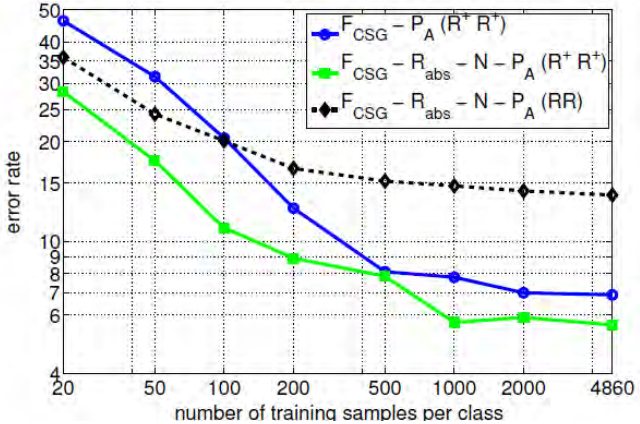
- フィルタをランダムとし，最上位の fully-connected 層のみ学習



Caltech-101

アーキテクチャ	ランダムフィルタ	フィルタも学習
2層，絶対値プーリング	62.9%	64.7%
2層，平均プーリング	19.6%	31.0%
1層，絶対値プーリング	53.3%	54.8%

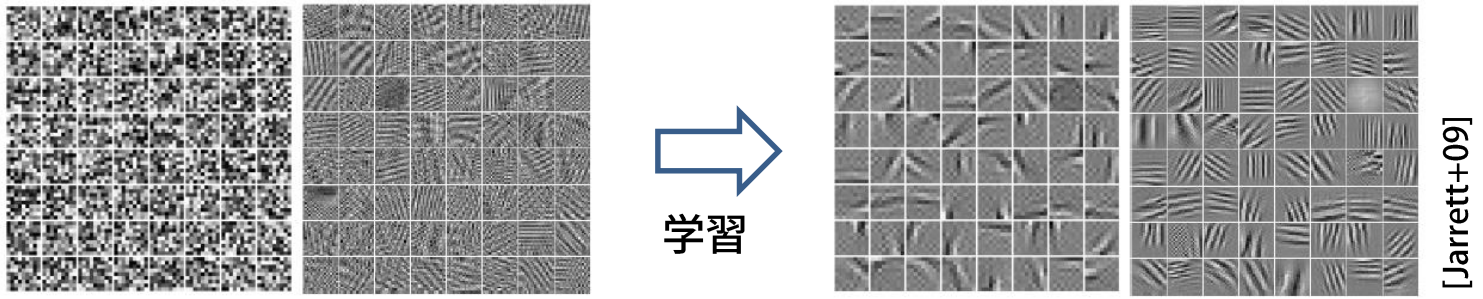
NORB



ランダムフィルタ:アーキテクチャの重要性

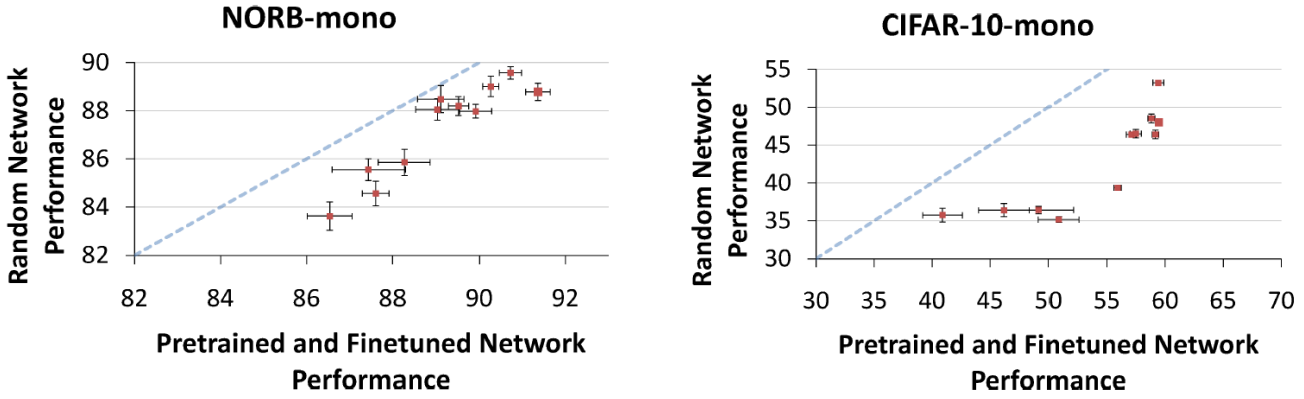
Saxe et al., On random weights and unsupervised feature learning, ICML2010

- 学習アルゴリズムよりもアーキテクチャがずっと大事
- プーリング層のユニットが最も活性化する最適入力:
  - 理論的説明 [Saxe+10]



ランダムフィルタ　最適入力　学習後フィルタ　最適入力

- アーキテクチャの性能予測をランダムフィルタで [Saxe+10]
  - アーキテクチャ探索時間を節約可

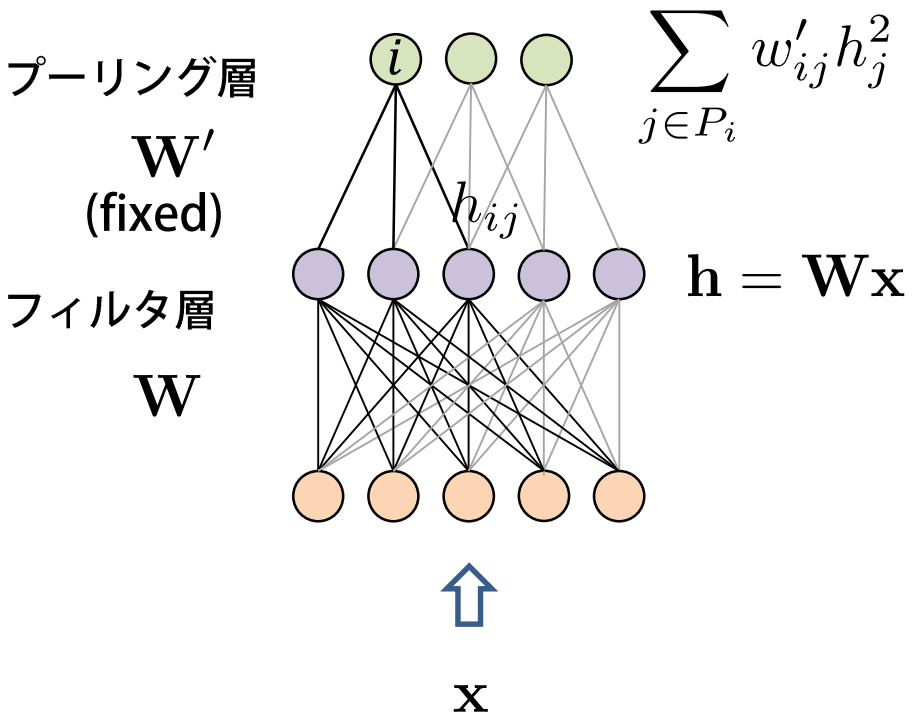


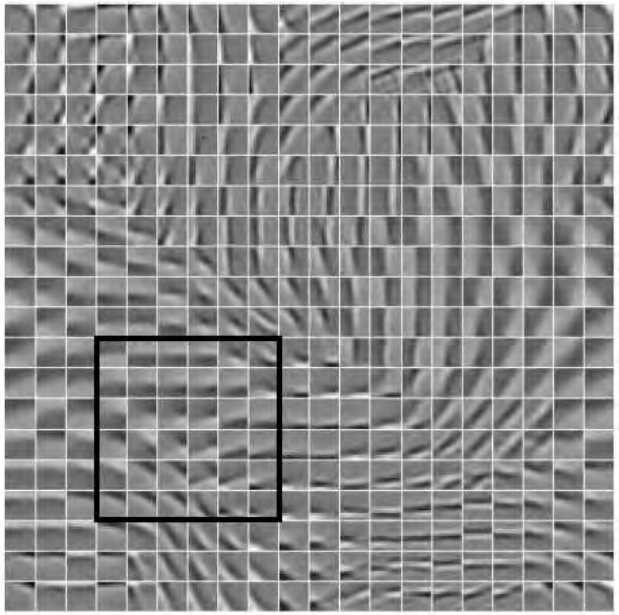
Topographic ICA

Hyvarinen and Hoyer, A two-layer sparse coding model learns simple and complex cell…, Vision Research, 2001

- 隣接ユニットの特徴（フィルタ）が類似するように→高度な不変性

$$\min_{\mathbf{W}} \sum_i \sqrt{\sum_{j \in P_i} w'_{ij} h_j^2} \quad \text{s.t.} \quad \mathbf{W}^\top \mathbf{W} = \mathbf{I}$$





[Kavukcuoglu09]

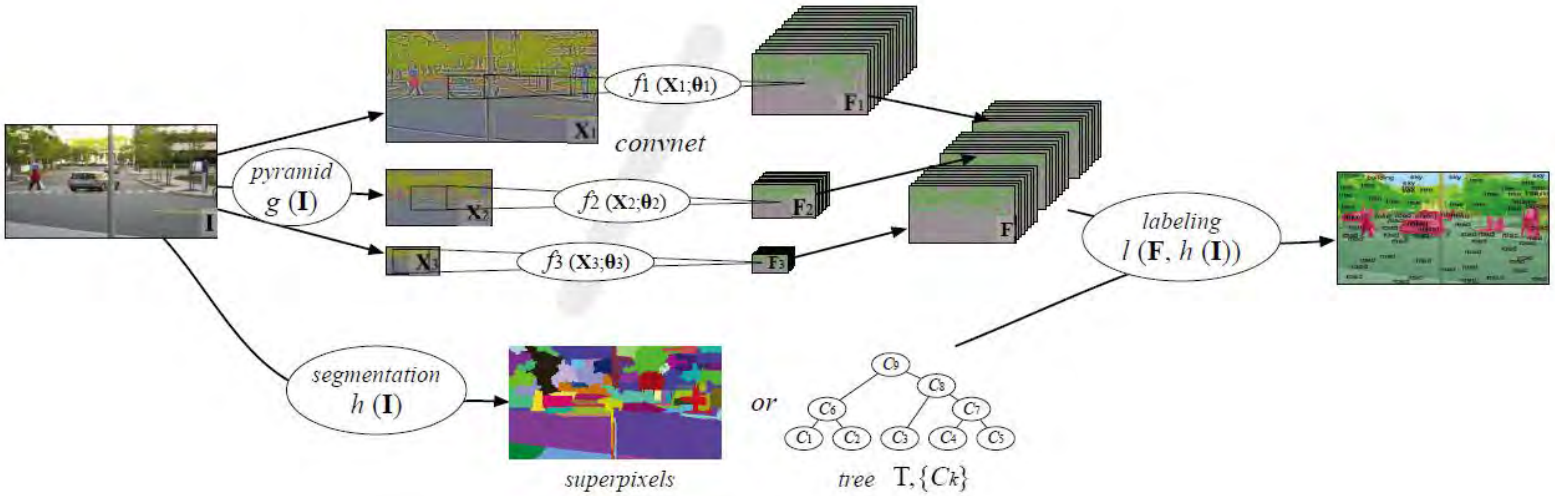
概要

- ディープラーニング＝画像認識のパラダイム変化？
- 中核的方法
  - 教師なし：（スパース）オートエンコーダ
  - 教師あり：たたみこみネット
- たたみこみネットの周辺
- その他事例
- まとめ

Scene Labeling

Farabet et al., Learning Hierarchical Features for Scene Labeling, IEEE PAMI, 2012

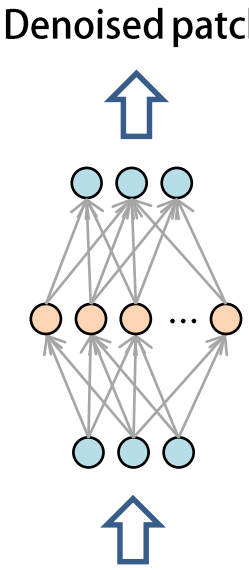
- 画素ごとにラベルを出力するCNNを教師あり学習
  - 上位層の出力を特徴量に空間方向の一致性を確保



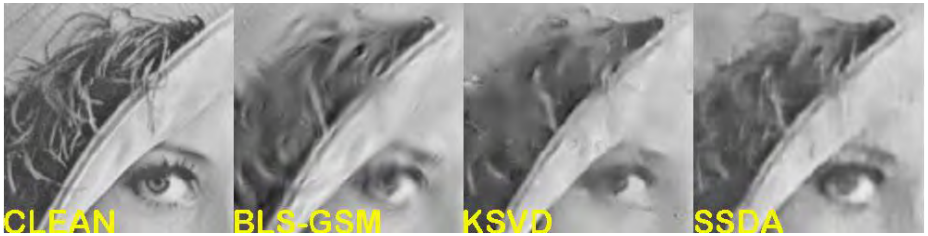
Denoising and inpainting

Xie et al., Image Denoising and Inpainting with Deep Neural Networks, NIPS, 2012

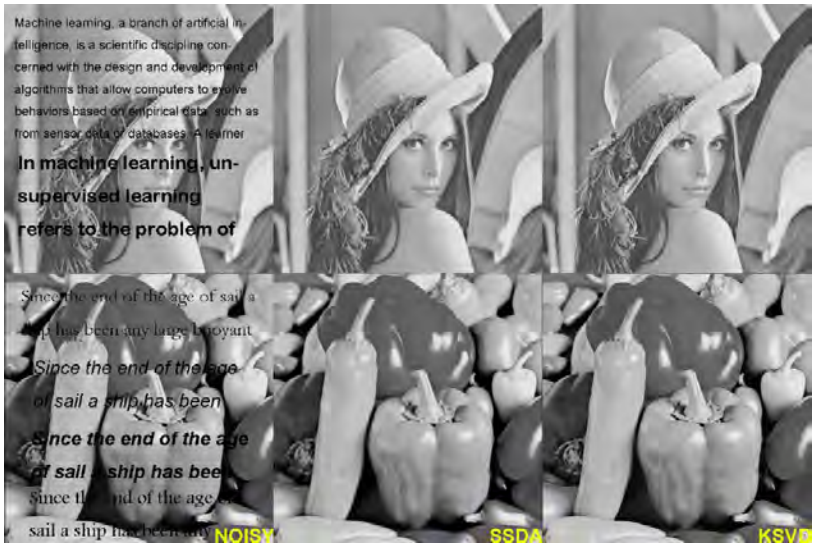
- デノイジング・オートエンコーダを用いたノイズ除去；Denoising性能は，PSNR評価では従来手法と同等だが，見た目で勝る
  - Inpaintingの問題をデノイジングとみなす → blind inpainting が可能に



多層の
スパースデノイジングオートエンコーダ
を応用



ノイズ除去結果



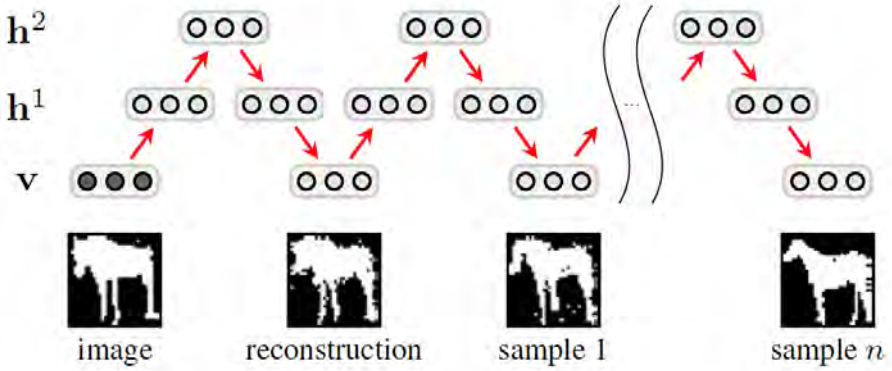
Blind inpainting 結果（KSVDはnon-blind）

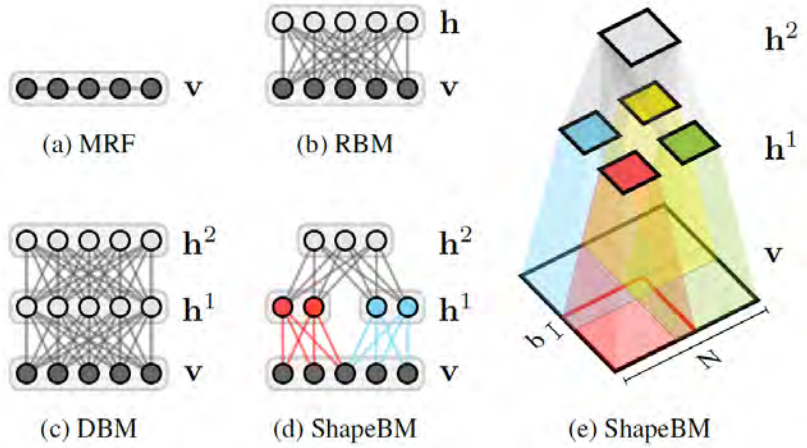
Shape Boltzmann machine

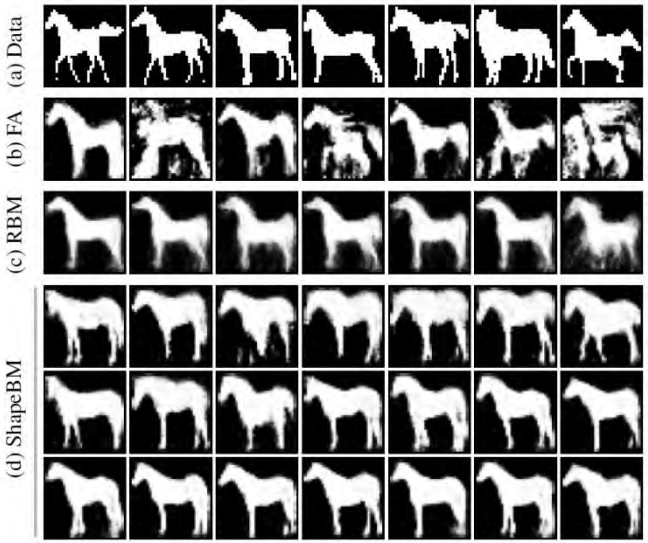
Ali Eslami et al., The Shape Boltzmann Machine: a Strong Model of Object Shape, CVPR12

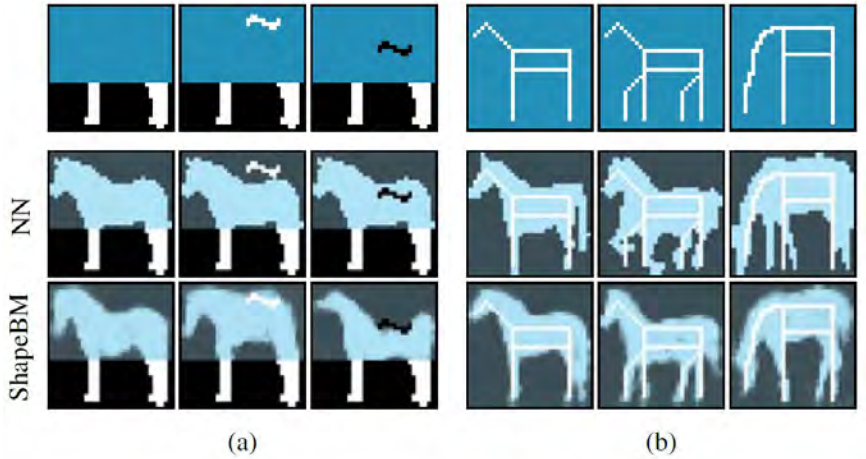
- 形を学習するDBM：realismとgeneralizationの達成

[video]



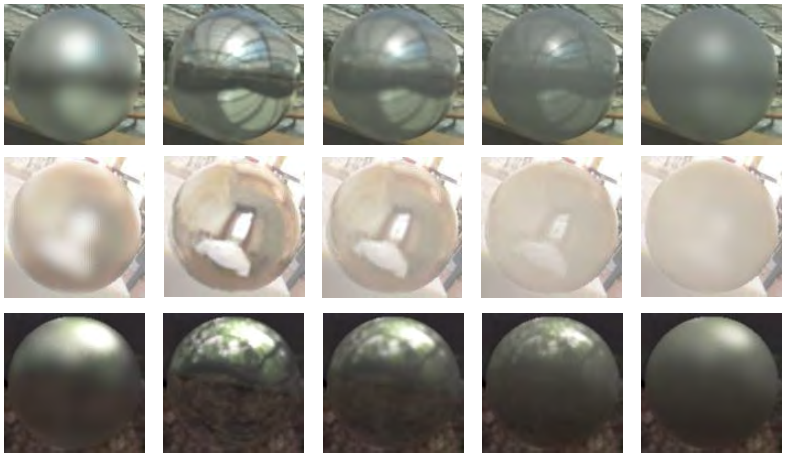




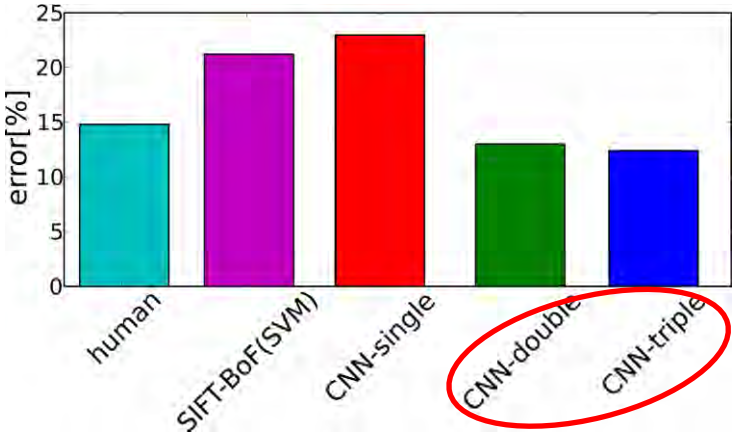


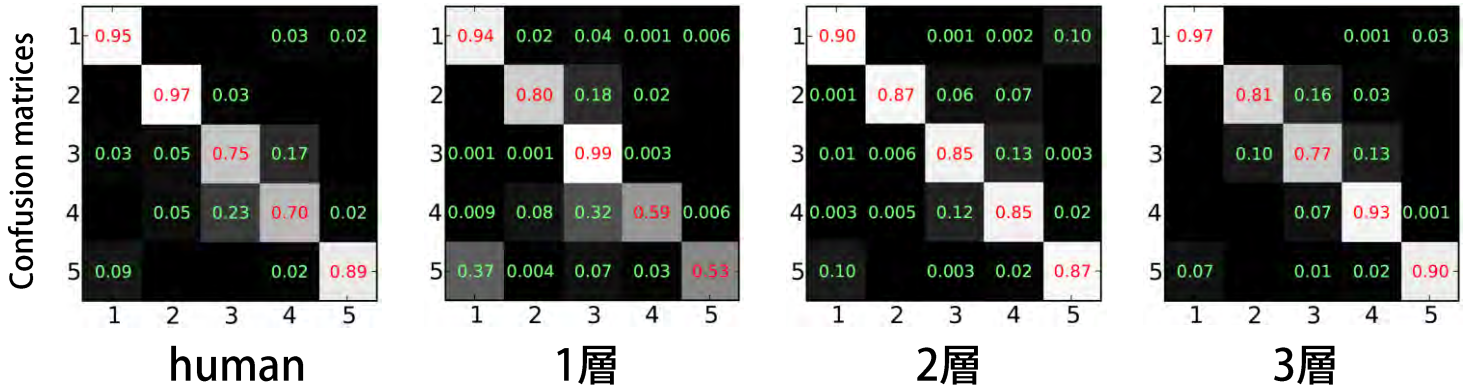
質感の認識

- 多層のCNNが高い性能
  - 人を上回る



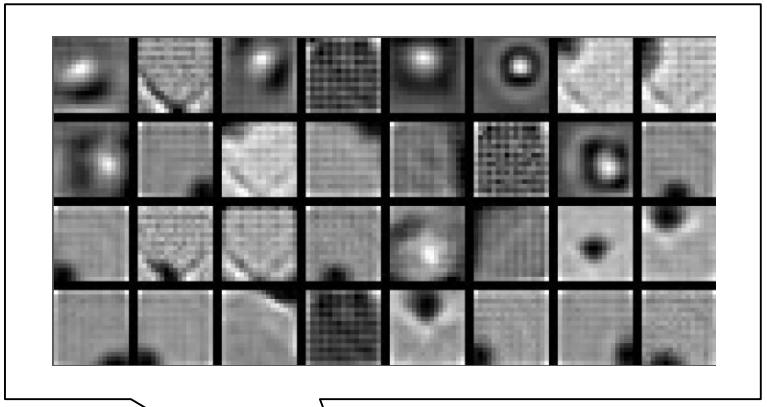
誤認識率（2000テスト中）

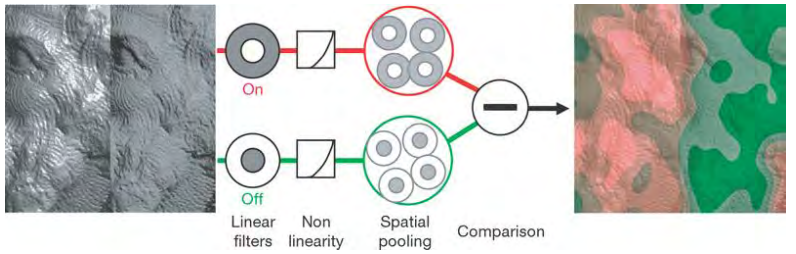




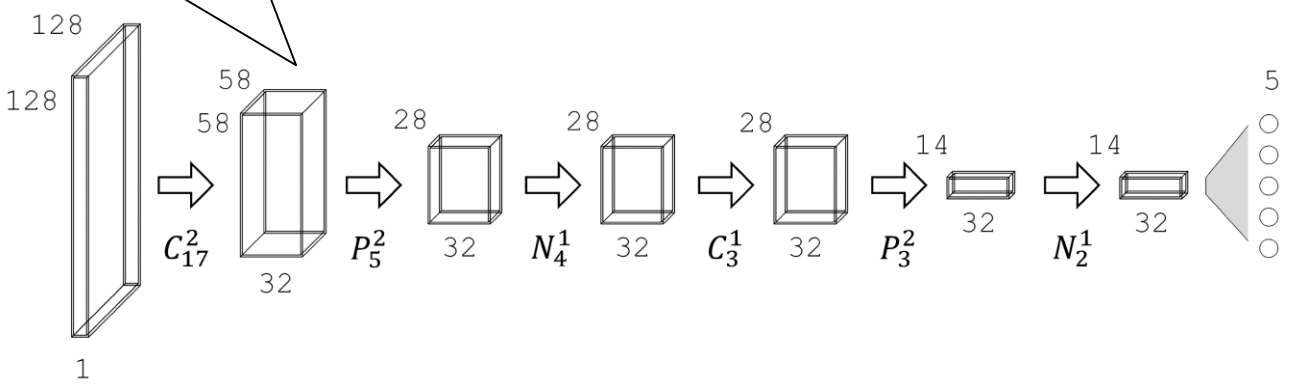
質感の認識

- 回転対称なフィルタを学習（on-center & off-center）





Motoyoshi et al. *Nature* 2007



まとめ：ディープラーニングの各技術の関係

DNNの学習をうまくやる方法：

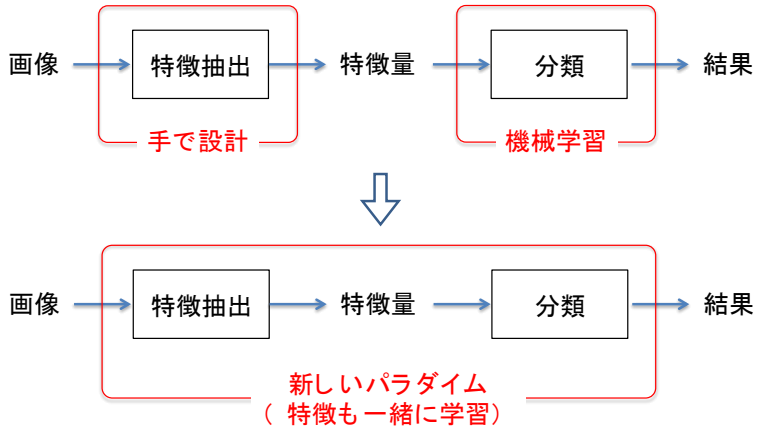
1. 計算方法の工夫
2. ネットの構造の工夫
3. データを増やす

教師あり	ドロップアウト [Hinton+12] 識別的プレトレーニング [Seide+11]	たたみこみネット [Fukushima80,LeCun+89, Krizhevsky+12]
教師なし	ディープAE プレトレーニング [Hinton+06]	再構成型TICA [Le+12-「グーグル猫」] たたみこみビリーフネット [Lee+09-CDBN]
	全結合	たたみこみ＋プーリング構造

まとめ

長所

- 圧倒的な性能
  - まだまだ向上しそう
- 特徴そのものの学習
  - パラダイム変化
  - （内部）表現の学習



短所

- 昔と変わらぬ使いづらさ
  - 学習パラメータ・ネット構造のチューニング
- エンジニアリング勝負
  - ネット・学習データの大規模化 → 大規模並列計算

$$\Delta w_{ij} = \epsilon \frac{\partial C}{\partial w_{ij}} + \alpha \Delta w'_{ij} - \epsilon \lambda w_{ij}$$