

Deep Learning (深層学習) 入門

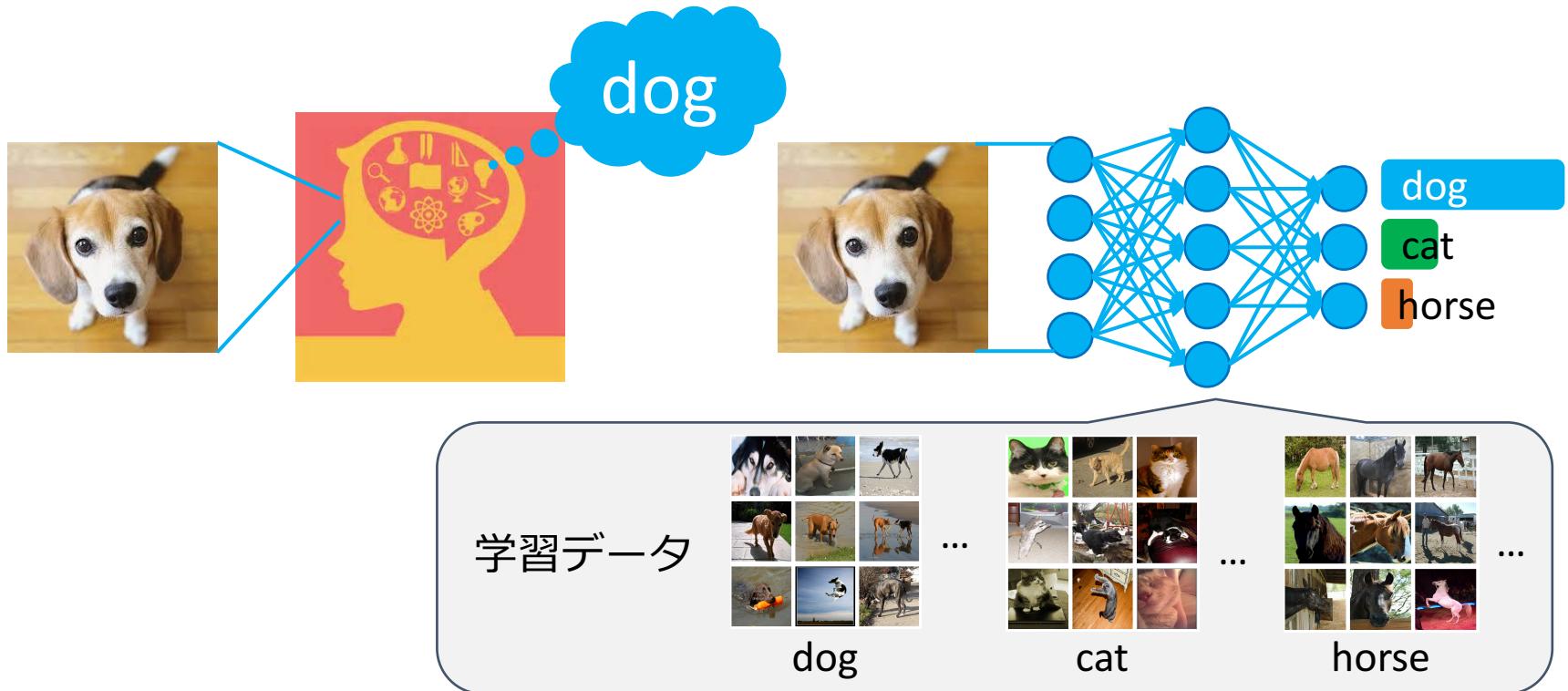
ビジネス科学研究科
経営システム科学専攻

目次

- Deep Learning とは
- ニューラルネットの歴史と基本
- 様々なニューラルネットワーク
- Deep Learning の実践
- 最新の研究

Deep Learning (深層学習)とは

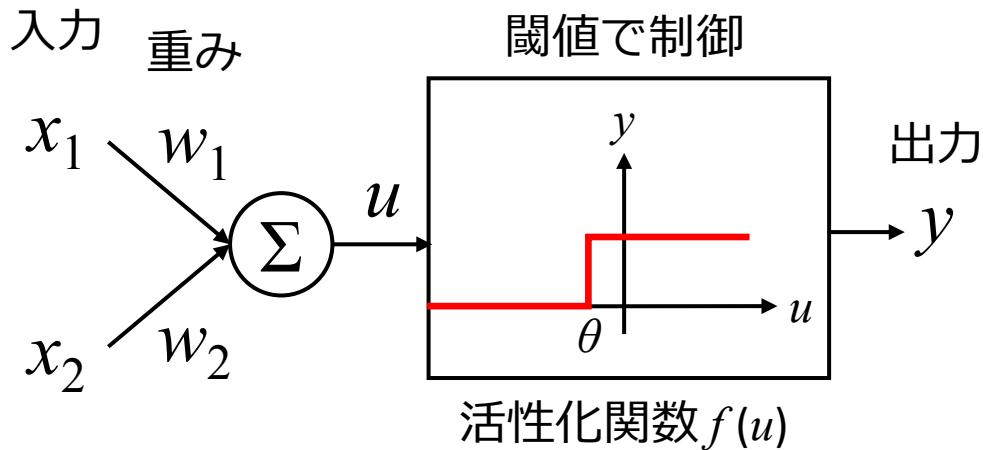
- ニューラルネットワークを用いた機械学習手法
 - 脳の神経細胞(ニューロン)の働きを模した
 - 機械学習とは、データを学習し、パラメータを獲得



ニューラルネットの歴史

ニューロンのモデル化 [McCulloch&Pitts,43]

- McCullochとPitts の素子モデル
 - 神経細胞を模した最初のニューロンモデル
 - 入力を重み付きで重ね合わせ,閾値で制御
 - パラメタ $\{w_1, w_2, \theta\}$ の組み合わせで,論理素子を表現
 - パラメタは手計算で決める



$$y = f(u - \theta) = f \left(\sum_{i=1}^2 w_i x_i - \theta \right)$$

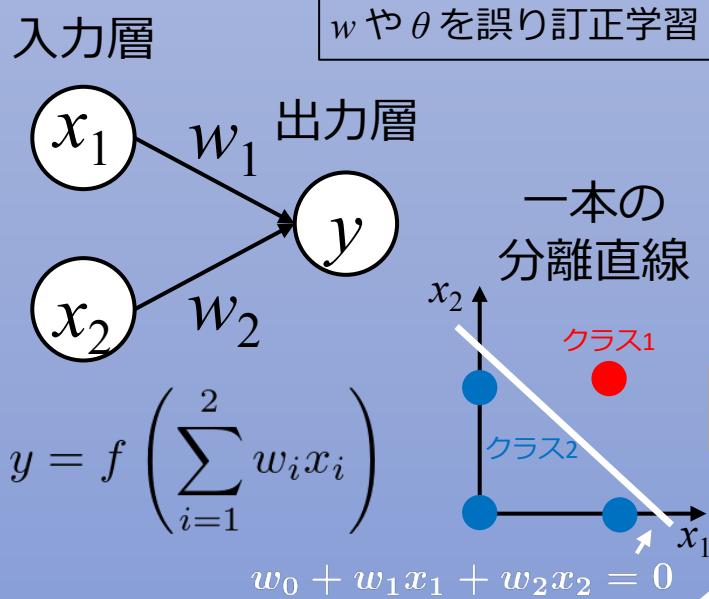
w_1	w_2	θ	論理素子
1	1	1.5	AND
1	1	0.5	OR
-1	-1	-1.5	NAND

パーセプトロン

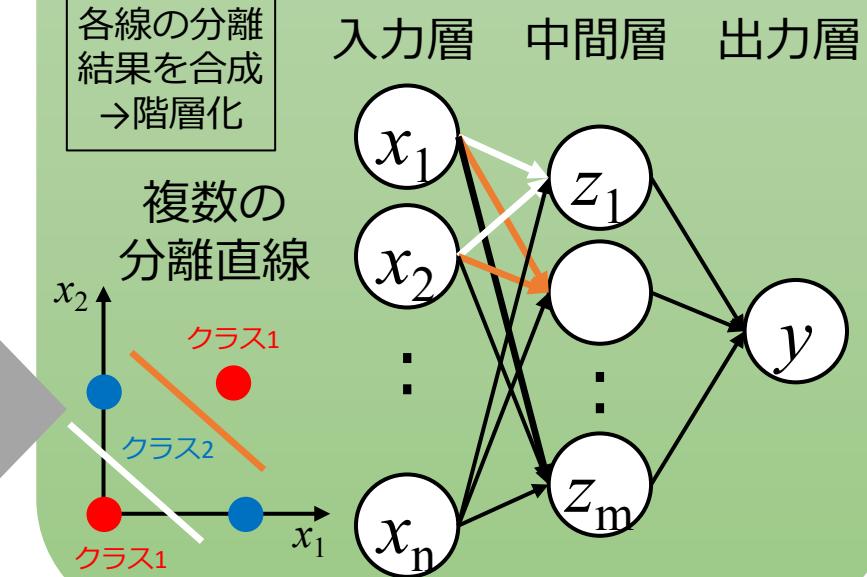
[Rosenblatt,57] → 単純パーセプトロン
[Minsky&Papert 68] → 線形分離不可能

- 学習サンプルからパラメタを自動的に決定
 - 単純パーセプトロン → 一本の直線で境界分離(線形分離)
 - 多層パーセプトロン → 単純パーセプトロンの限界を克服

単純パーセプトロン



多層パーセプトロン



冬の時代 1970～1980

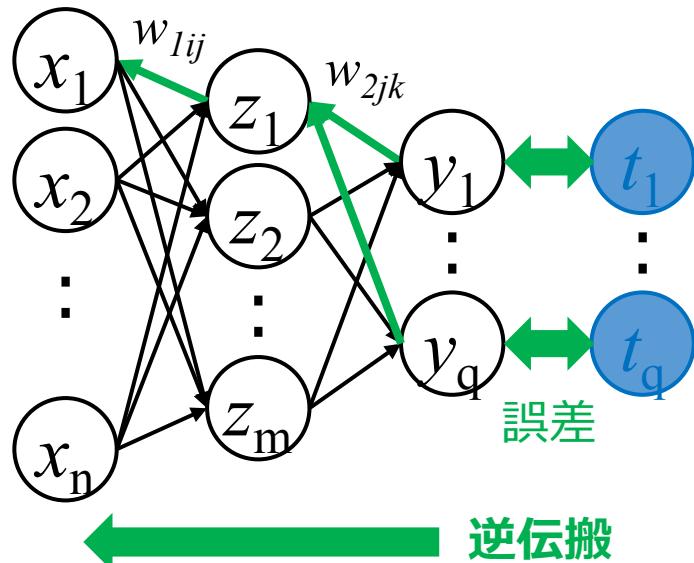
- ただし、階層的構造の学習方法が未解決

第1期	1940～	• McCullochとPittsが形式ニューロンモデルを発表 [McCulloch-Pitts,43]
	1950～	• Rosenblattがパーセプトロンを発表 [Rosenblatt,57]
	1960～	• MinskyとPapertが単純パーセプトロンの線形分離不可能問題への限界を指摘 [Minsky-Papert,69]
冬	1970～	冬の時代 (階層的構造の学習方法が未解決)
第2期	1980～	• Fukushimaらがネオコグニトロンを提案 [Fukushima,80] • Rumelhartらが誤差逆伝播法を提案 [Rumelhart+,86] • LeCunらが畳み込みニューラルネット Conv.net を提案 [LeCun,89]
	1990～	冬の時代 (学習時間や過学習に課題、一方でSVMが流行)
第3期	2000～	• Hintonらが事前学習とオートエンコーダを導入した多層NNを提案 [Hinton+,06]
	2010～	• Seideらが音声認識のベンチマークで圧勝 [Seide+,11] • KrizhevskyらがReLUを提案し画像認識コンペで圧勝 [Krizhevsky,12]

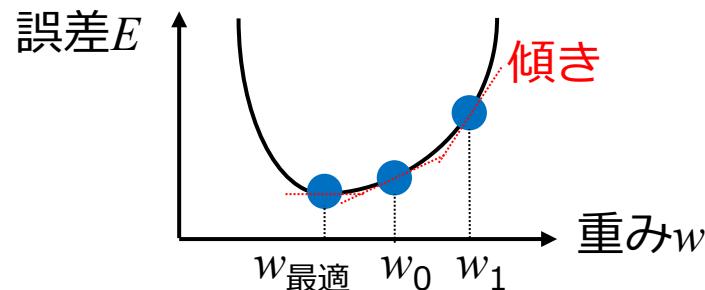
誤差逆伝搬法 [Rumelhart+,86]

- 2層以上の多層ニューラルネットの学習法
 - 誤差が小さくなるように全体の重みを調整
 - 入力層から逆に伝播させ,各層の重みの勾配を計算

入力層 中間層 出力層 正解



勾配降下法



連鎖律を使って勾配を計算

$$\frac{\partial E}{\partial w_{1ij}} = \sum_{k=1}^q \left(\frac{\partial E}{\partial y_k} \frac{\partial y_k}{\partial u_{2k}} \frac{\partial u_{2k}}{\partial w_{1ij}} \right)$$

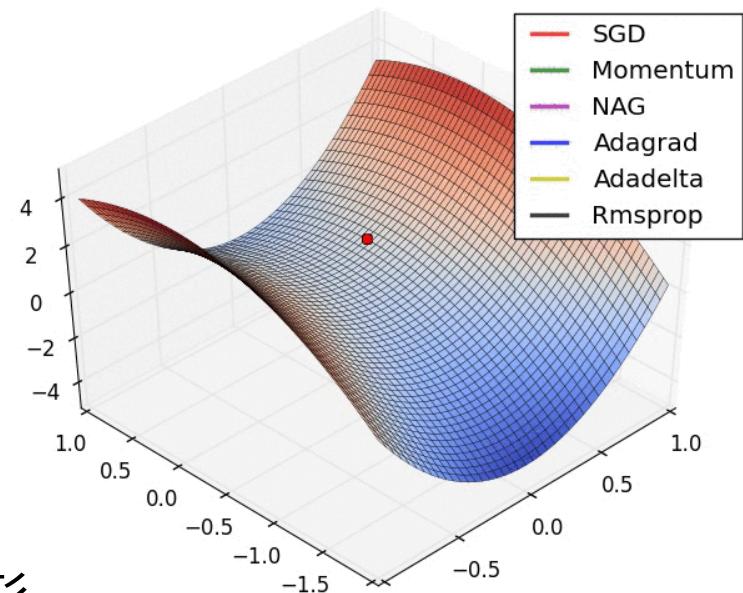
冬の時代 1990～2000

- ただし、多層では学習時間や過学習に課題

第1期	1940～	• McCullochとPittsが形式ニューロンモデルを発表 [McCulloch-Pitts,43]
	1950～	• Rosenblattがパーセプトロンを発表 [Rosenblatt,57]
	1960～	• MinskyとPapertが単純パーセプトロンの線形分離不可能問題への限界を指摘 [Minsky-Papert,69]
冬	1970～	冬の時代 (階層的構造の学習方法が未解決)
第2期	1980～	• Fukushimaらがネオコグニトロンを提案 [Fukushima,80]
		• Rumelhartらが誤差逆伝播法を提案 [Rumelhart+,86]
		• LeCunらが畳み込みニューラルネット Conv.net を提案 [LeCun,89]
冬	1990～	冬の時代 (学習時間や過学習に課題、一方でSVMが流行)
第3期	2000～	• Hintonらが事前学習とオートエンコーダを導入した多層NNを提案 [Hinton+,06]
	2010～	• Seideらが音声認識のベンチマークで圧勝 [Seide+,11] • KrizhevskyらがReLUを提案し画像認識コンペで圧勝 [Krizhevsky,12]

最適化方法の改善

- AdaGrad [Duchi,11]
 - まれなパラメータに対してはより大きな更新
 - 頻出のパラメータに対してはより小さな更新
- RMSProp
 - AdaGrad の変形
- AdaDelta [Zeiler,12]
 - AdaGrad や RMSProp の変形
- Adam [Kingma,14]
 - AdaGrad, RMSProp, AdaDelta の変形
 - Adam は全般的に最も良い

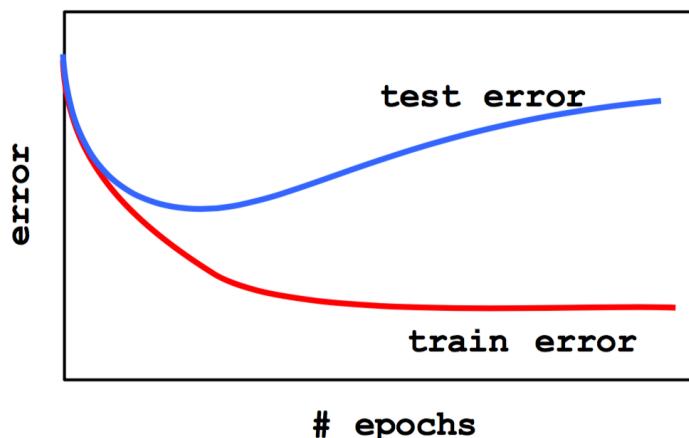


SDGに比べ、RMSpropや
Ada系はすぐに負の傾きを
たどる (作者: Alec Radford)

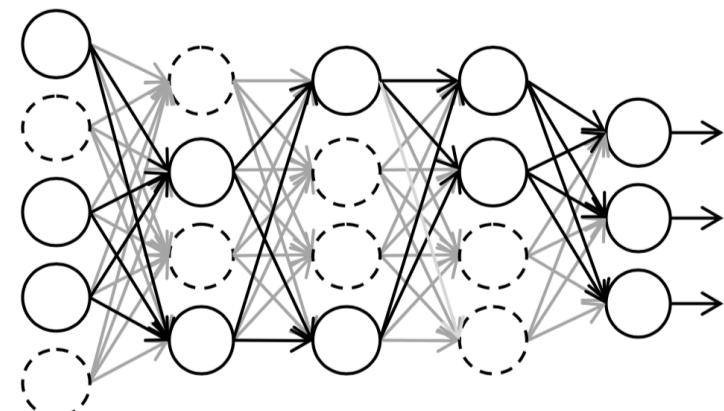
dropout [Hinton+,12]

- 学習時にネットワークの自由度を強制的に小さくし、過適合を避ける
 - 学習時: 一定の確率 p で中間層のユニットを「無効化」
 - 推論時: 全ての出力を一律に p 倍して補填

過学習の様子



ドロップアウト



ニューラルネットの歴史

- 黎明～終焉を繰り返し,近年は3度目のブーム

第1期	1940～	• McCullochとPittsが形式ニューロンモデルを発表 [McCulloch-Pitts,43]
	1950～	• Rosenblattがパーセプトロンを発表 [Rosenblatt,57]
	1960～	• MinskyとPapertが単純パーセプトロンの線形分離不可能問題への限界を指摘 [Minsky-Papert,69]
冬	1970～	冬の時代 (階層的構造の学習方法が未解決)
第2期	1980～	• Fukushimaらがネオコグニトロンを提案 [Fukushima,80]
		• Rumelhartらが誤差逆伝播法を提案 [Rumelhart+,86]
		• LeCunらが畳み込みニューラルネット Conv.net を提案 [LeCun,89]
冬	1990～	冬の時代 (学習時間や過学習に課題, 一方でSVMが流行)
第3期	2000～	• Hintonらが事前学習とオートエンコーダを導入した多層NNを提案 [Hinton+,06]
	2010～	• Seideらが音声認識のベンチマークで圧勝 [Seide+,11] • KrizhevskyらがReLUを提案し画像認識コンペで圧勝 [Krizhevsky,12]



冬の時代の終焉

- 最近の Deep Learning 成功の背景
 - 一定以上の規模のデータ → 改善
 - Webなどから十分な規模のデータを収集可能
 - 学習の難しさ → 改善
 - 様々なテクニック (事前学習, dropout 等)
 - 誤差逆伝搬法の計算量膨大 → 改善
 - 計算機能力の飛躍的向上
 - GPU, マルチコアCPU, PCクラスタの登場
 - 性能を引き出すのに必要なノウハウ → 未解決
 - 「黒魔術」のまま

事例1—音声認識での成功 (2011)

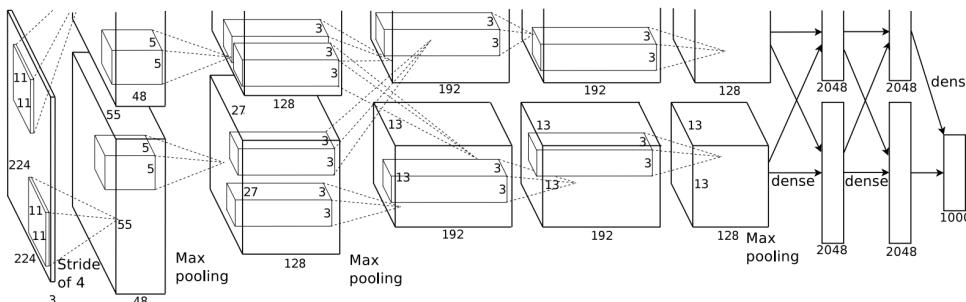
- Microsoft Research のグループ
 - 電話での会話音声の標準データセット
 - 入力(MFCC)-出力(HMM状態変数)の関係をDNNで学習
 - 従来 GMM-HMM → DNN-HMM (全結合7層, 事前学習あり)
 - 単語誤認識率で 10%前後の大幅な精度改善

acoustic model & training	recognition mode	RT03S		Hub5'00 SWB	voicemails		tele- conf
		FSH	SW		MS	LDC	
GMM 40-mix, ML, SWB 309h	single-pass SI	30.2	40.9	26.5	45.0	33.5	35.2
GMM 40-mix, BMMI, SWB 309h	single-pass SI	27.4	37.6	23.6	42.4	30.8	33.9
CD-DNN 7 layers x 2048, SWB 309h, this paper (rel. change GMM BMMI → CD-DNN)	single-pass SI	18.5 (-33%)	27.5 (-27%)	16.1 (-32%)	32.9 (-22%)	22.9 (-26%)	24.4 (-28%)

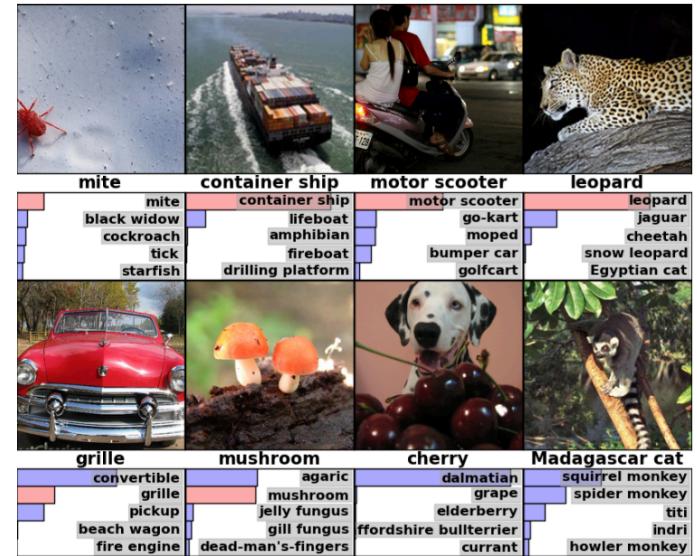
F. Seide, G. Li and D. Yu, "Conversational Speech Transcription Using Context-Dependent Deep Neural Networks." *Interspeech*. 2011.

事例2—画像認識での成功 (2012)

- 一般物体認識 (Hintonのグループ)
 - ImageNet Large-scale Visual Recognition Challenge 2012
 - 1000カテゴリ×約1000枚 = 100万枚 の訓練画像
 - 畳込み層5, 全結合層3, 2つのGPUで2週間 (AlexNet)
 - 誤識別率が10%以上減少 (過去数年間での向上は1~2%) ≈



Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton.
"Imagenet classification with deep convolutional neural networks."
Advances in neural information processing systems. 2012.
<http://image-net.org/challenges/LSVRC/2012/supervision.pdf>



ImageNet Large-scale Visual Recognition Challenge (IRSVRC)

- ImageNet のデータの一部を用いたフラッグシップコンペティション (2010年より開催)
 - ImageNet [Deng et al., 2009]
 - クラウドソーシングにより構築中の大規模画像データセット
 - 1400万枚, 2万2千カテゴリ (WordNet に従って構築)



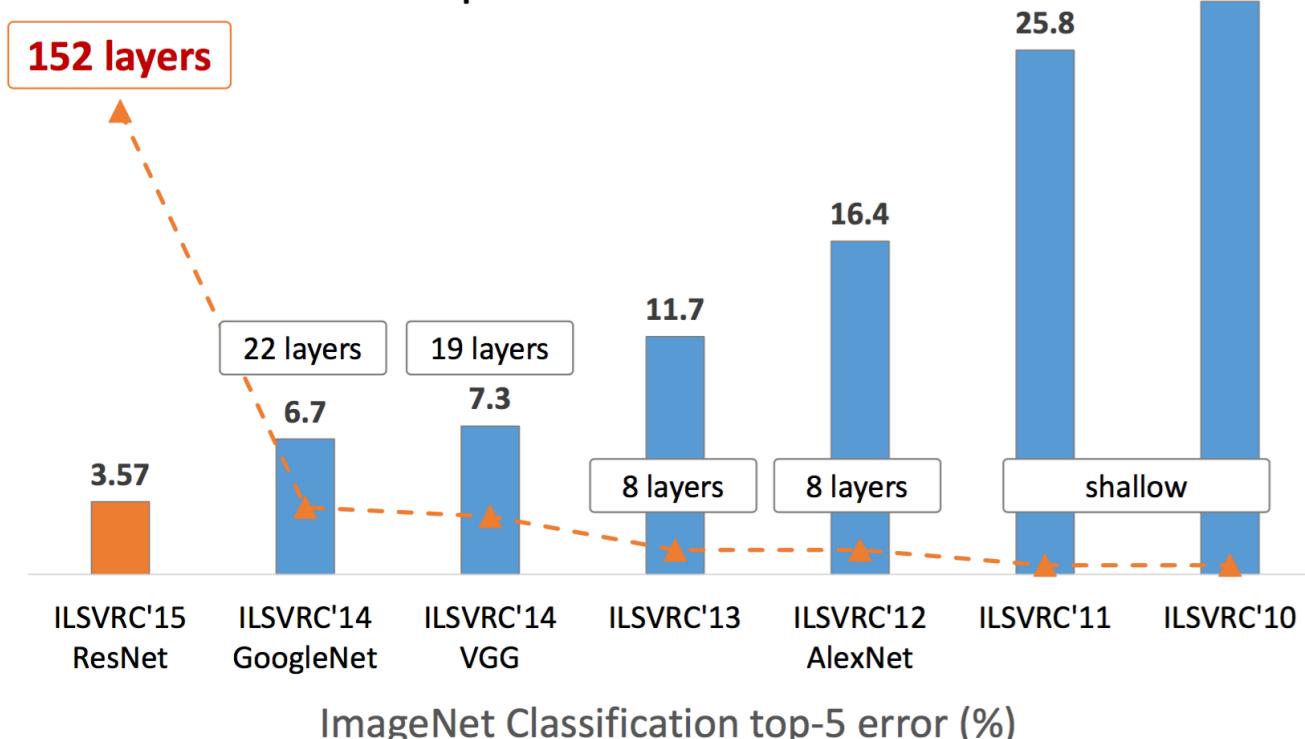
- コンペでのタスク
 - 1000クラスの物体カテゴリ分類
 - 学習データ120万枚, 検証用データ5万枚, テストデータ10万枚
 - 200クラスの物体検出
 - 学習データ45万枚, 検証用データ2万枚, テストデータ4万枚

中山英樹. "招待講演 深層畳み込みニューラルネットによる画像特徴抽出と転移学習 (音声)." 電子情報通信学会技術研究報告 = IEICE technical report: 信学技報115.146 (2015): 55-59.

IRSVRC 誤認識率の変遷

Microsoft
Research

Revolution of Depth



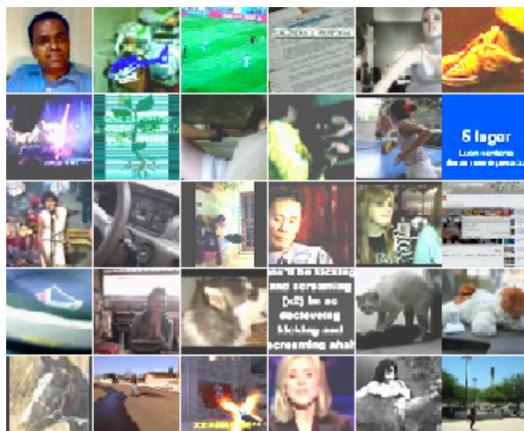
ICCV15
International Conference on Computer Vision

Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. "Deep Residual Learning for Image Recognition". arXiv 2015.

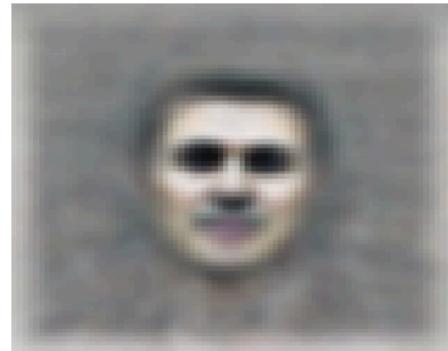
Kaiming, Xiangyu Zhang, Shaoqing Ren, Jifeng Dai, & Jian Sun. "ILSVRC 2015 workshop"

事例3—画像教師なし学習 (2012)

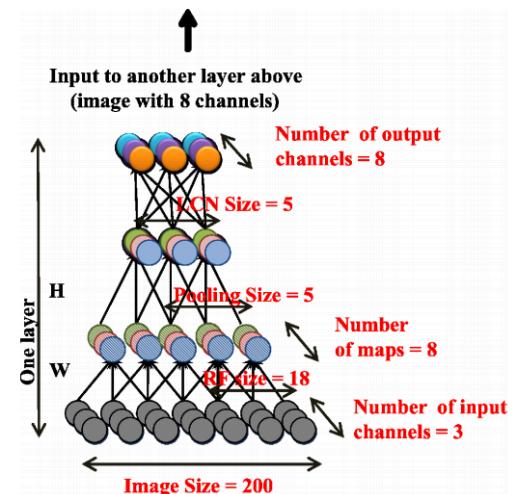
- 4層x3段=12層のAutoEncoder
 - パラメータ数10億個 (<人)
 - 16コアのPC 1000台 ×3日間
 - YouTubeの画像1000万枚 (顔画像は3%)



学習データ



顔に反応するニューロンに
対して最適化された画像



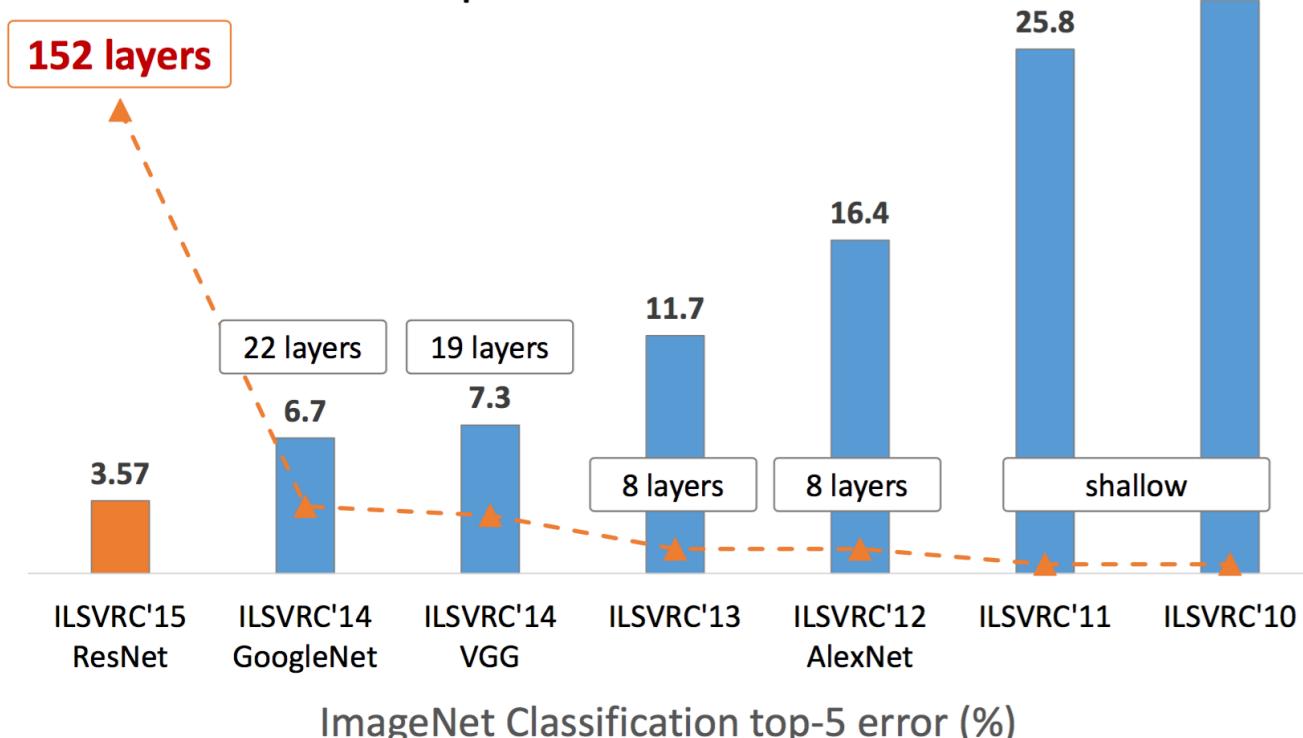
猫に反応するニューロンに
対して最適化された画像

Le, Quoc V. "Building high-level features using large scale unsupervised learning." In *ICML*, 2012.

ネットワークの変遷 (画像系)

Microsoft
Research

Revolution of Depth



ICCV15
International Conference on Computer Vision

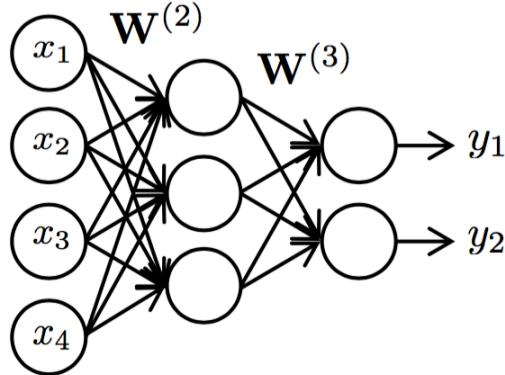
Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. "Deep Residual Learning for Image Recognition". arXiv 2015.

Kaiming, Xiangyu Zhang, Shaoqing Ren, Jifeng Dai, & Jian Sun. "ILSVRC 2015 workshop"

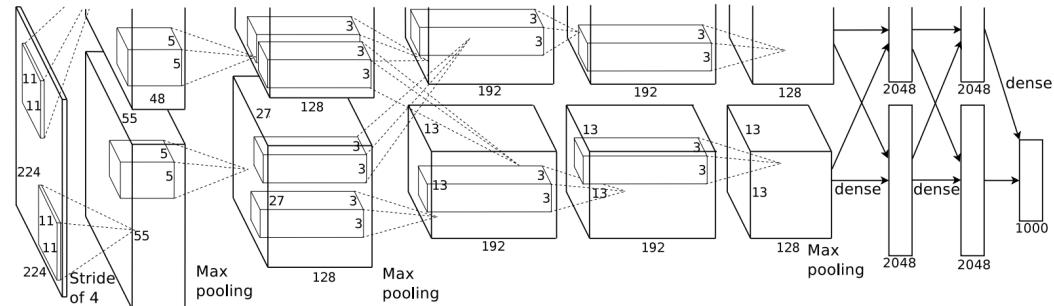
様々なニューラルネット

様々なニューラルネット

フィードフォワードネット

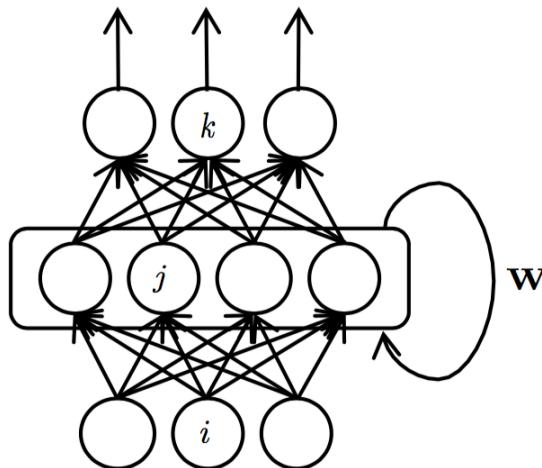


CNN (畳み込みNN)

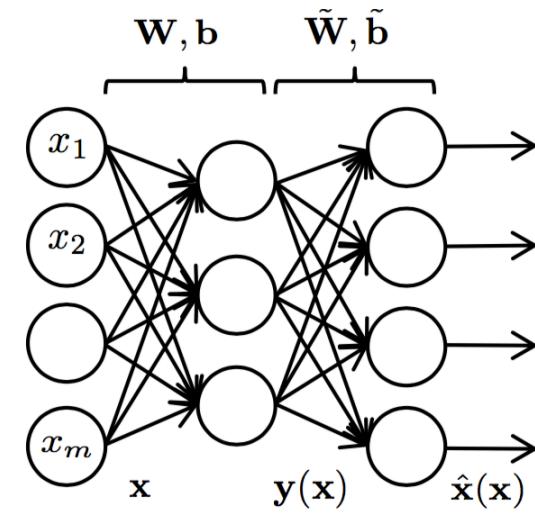


教師あり

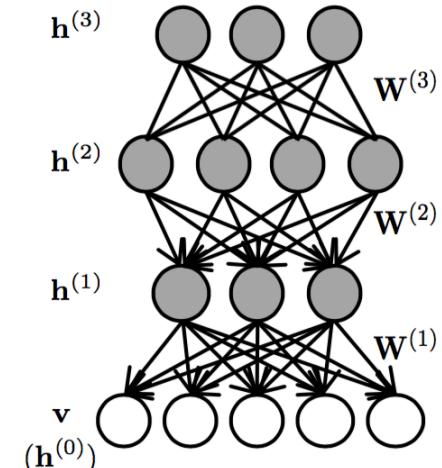
RNN (リカレントNN)



教師なし



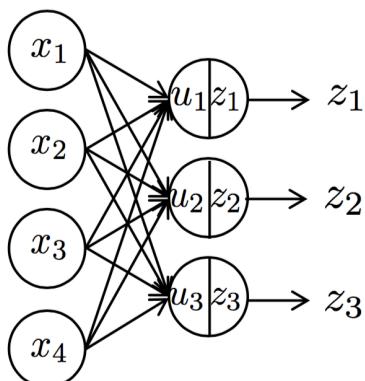
AutoEncoder (自己符号化器)



ボルツマンマシン

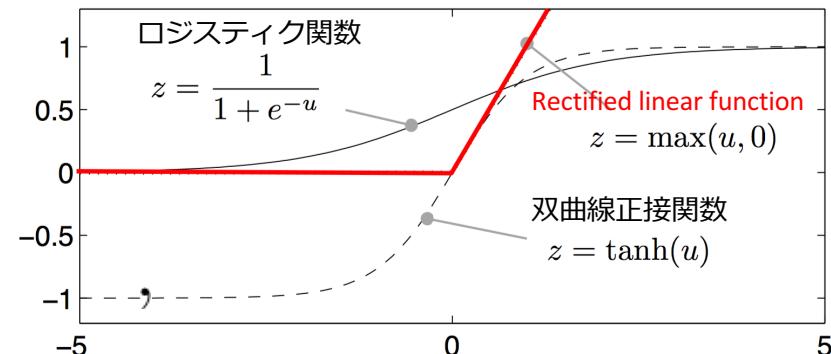
フィードフォワードネット

- 最も基本的かつよく使われているネットワーク
 - 入力側から出力側へ情報が一方向へのみ伝搬
 - 層状に並べたユニットが隣接層間のみで結合
 - 総入力 u に活性化関数を適用

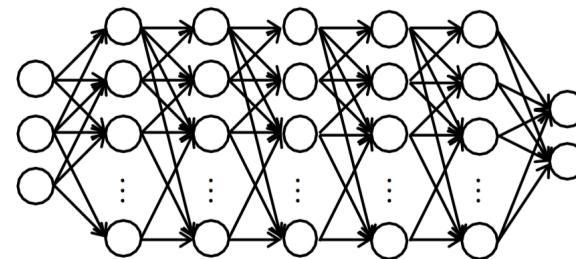


$$u_j = \sum_{i=1}^I w_{ji} x_i + b_j$$
$$z_j = f(u_j)$$

活性化関数 $f(u)$

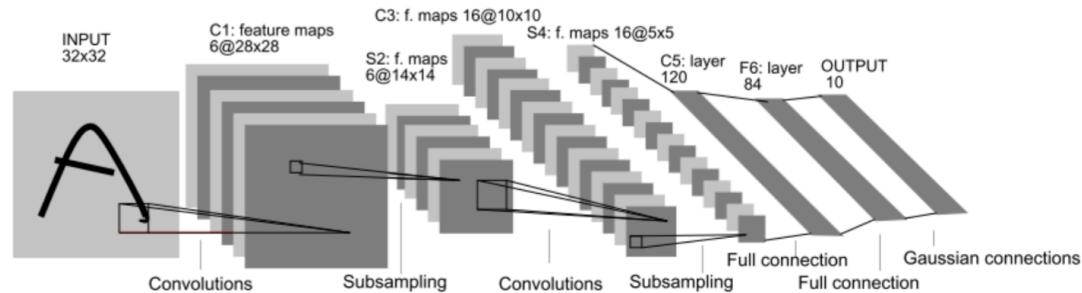
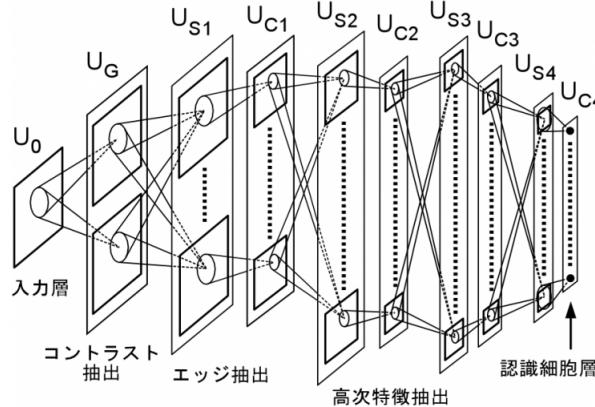


問題の種類	活性化関数	誤差関数
回帰	恒等写像	二乗誤差
二値分類	ロジスティック関数	尤度の対数を最小化
多クラス分類	ソフトマックス関数	交差エントロピー



Convolutional Neural Network (CNN)

- 画像を対象とする畠込みニューラルネット
 - ネオコグニトロンがルーツ [Fukushima,80]
 - 誤差逆伝播法による学習を取り入れ完成 [LeCun,98]
 - 畠込み(局所特徴抽出)+**プーリング**を繰り返し,
多様な変形に対して**シフト不変性**を獲得

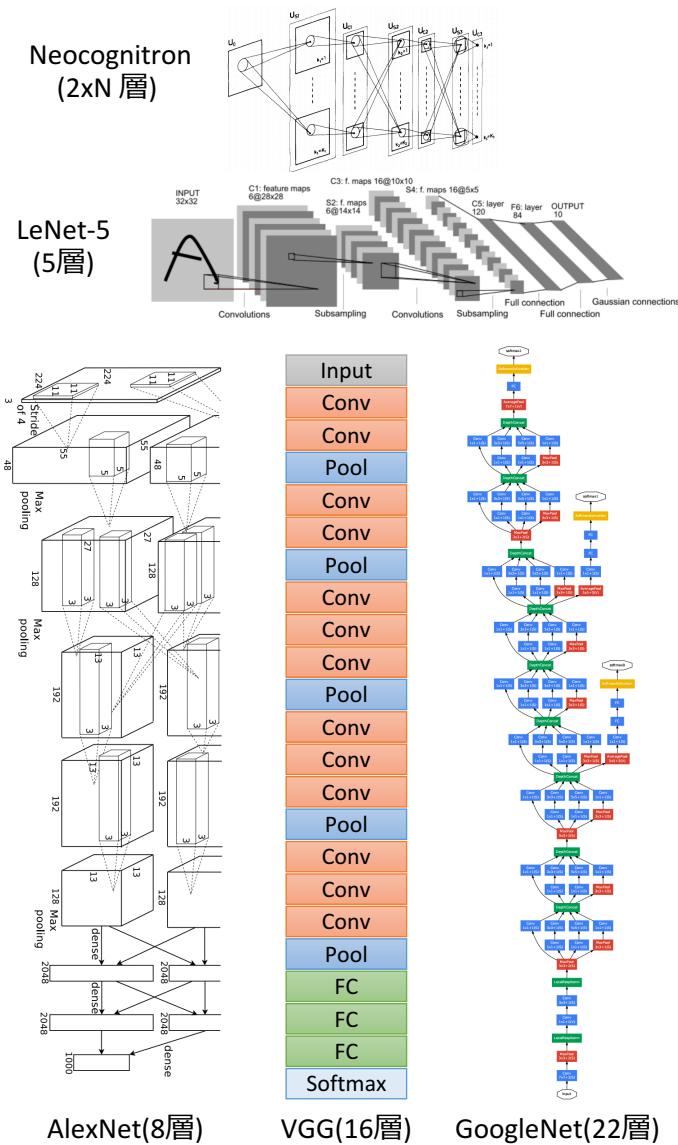


ネオコグニトロン: 神経科学の
階層仮説 [Hubel&Wiesel,59] を実装

LeNet-5 [LeCun,98]

CNNの変遷

畳み込み ニューラル ネットワー クの発見	Neocognitron (1980) [1]	<ul style="list-style-type: none"> 生物の覚神経路を模倣 特徴抽出と情報集約の処理を繰り返す
	LeNet-5 (1998) [2]	<ul style="list-style-type: none"> 畳み込みとプーリングの導入 誤差逆伝搬法によって学習した
プーリング, 活性化関数, 正則化	AlexNet (2012) [6]	<ul style="list-style-type: none"> Ave./Max Pooling, Normalization ,ReLU[4], Dropout[5] の導入 8層
	Network in Network (2013) [7]	<ul style="list-style-type: none"> 畳み込み層に非線型性を導入
	VGG Net (2014) [8]	<ul style="list-style-type: none"> 小さい畳み込みサイズ(3x3)を多段にした 19層
	GoogleNet / Inception (2014-2015) [9,10]	<ul style="list-style-type: none"> 複数の畳み込みフィルタを並列に用いる 22層
アーキテク チャの多様 化	SPP Net (2014) [11]	<ul style="list-style-type: none"> 様々なサイズの画像を許容 CNN 時のリサイズを回避
	All Convolutional Net, guided BP (2014) [12]	<ul style="list-style-type: none"> プーリングをストライド2の畳み込みに置き換え guided BPによる超高次層の特徴可視化
学習法の多 様化	Exemplar CNN (2014) [13]	<ul style="list-style-type: none"> Data Augmentation を利用して教師なし表現学習
	Triplet Network (2014) [14]	<ul style="list-style-type: none"> ユークリッド空間上でCNN上の特徴同士が同じクラスなら近くなるように,別クラスなら遠くなるようにした
超多層ア ーキテクチャ	Batch Normalization (2015) [15]	<ul style="list-style-type: none"> パラメータ付き正規化処理 複雑なアーキテクチャをスクラッチで学習させる必須技術
	Residual Network; ResNet (2015) [16]	<ul style="list-style-type: none"> 152層からなる超多層 途中の特徴マップを何層か先にバイパス



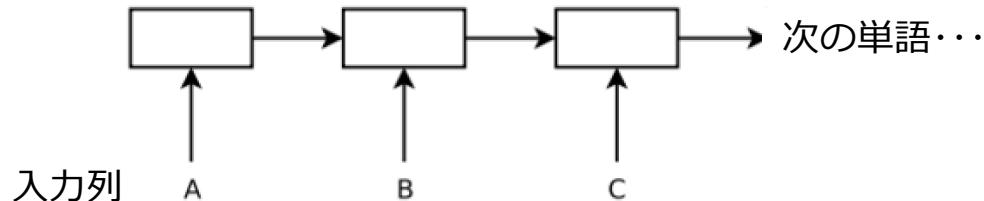
- [1] K. Fukushima. Neocognitron: A self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics* 36, 1980.
- [2] Y LeCun, L Bottou, Y Bengio, P Haffner. Gradient based learning applied to document recognition. *Proceedings of the IEEE* 86, 1998.
- [3] K. Jarrett, K. Kavukcuoglu, M. Ranzato, Y. LeCun. What is the best multi stage architecture for object recognition?. *CVPR*, 2009.
- [4] X. Glorot, A. Bordes, Y. Bengio. Deep Sparse Rectifier Neural Networks. *AISTATS* 11, 2011.
- [5] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, R. R. Salakhutdinov. Improving neural networks by preventing co adaptation of feature detectors. *arXiv: 1207.0580*, 2012.
- [6] A. Krizhevsky, I. Sutskever, G. E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. *NIPS*, 2012.
- [7] M. Lin, Q. Chen, S. Yan. Network In Network. *arXiv: 1312.4400*, 2013.
- [8] K. Simonyan, A. Zisserman. Very Deep Convolutional Networks for Large Scale Visual Recognition. *arXiv: 1409.1556*, 2014.
- [9] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich. Going deeper with convolutions. *arXiv: 1409.4842*, 2014.
- [10] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna. Rethinking the Inception Architecture for Computer Vision. *arXiv: 1512.00567*, 2015.
- [11] K. He, X. Zhang, S. Ren, J. Sun. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *arXiv: 1406.4729*, 2014.
- [12] J. T. Springenberg, A. Dosovitskiy, T. Brox, M. Riedmiller. Striving for Simplicity: The All Convolutional Net. *arXiv: 1412.6806*, 2014.
- [13] A. Dosovitskiy, P. Fischer, J. T. Springenberg, M. Riedmiller, T. Brox. Discriminative Unsupervised Feature Learning with Exemplar Convolutional Neural Networks. *arXiv: 1406.6909*, 2014.
- [14] E. Hoffer, N. Ailon. Deep metric learning using Triplet network. *arXiv: 1412.6622*, 2014.
- [15] S. Ioffe, C. Szegedy. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv: 1502.03167*, 2015.
- [16] K. He, X. Zhang, S. Ren, J. Sun. Deep Residual Learning for Image Recognition. *arXiv: 1512.03385*, 2015.

Recurrent Neural Network (RNN)

- 系列データを対象とするニューラルネット

例1) RNN 言語モデル [Mikolov+, 2010]

- 言語モデルとは文の尤もらしさをスコア化するモデル
- 過去の履歴を考慮して,現在の単語を予測



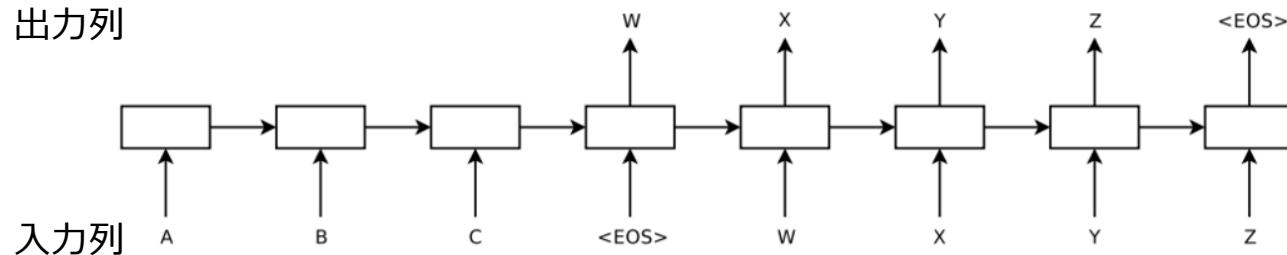
例: RNNにより文字単位で「吾輩は猫である」を学習し,文生成した

吾輩は猫である。鍋の中がちょっと立派な事はないから、そうしておやじや君は何と云う。甘らんですか。そりや、それですばからかったのは、つまり葉中だけ上げて冷淡に描き出した。今日は日曜だってもしかしたに違ない。

Recurrent Neural Network (RNN)

例2) Sequence to Sequence [Sutskever+, 2014]

- LSTM (Long-Short Term Memory) を2つ用い, 入力系列を固定長ベクトルに変換し, そのベクトルから出力系列を生成
- 機械翻訳も系列→系列を生成するモデルとして扱える



“ABC”という単語列から“WXYZ”という単語列への翻訳

例: Seq-to-Seq により文字単位で百人一首を文字単位で学習し, 下の句を生成した

これやこの行くも帰るも別れては

知るも知らぬぬ逢坂の関

これやこの行くも帰るも別れては

知るも知らぬ恋逢坂の思し

これやこの行くも帰るも別れては

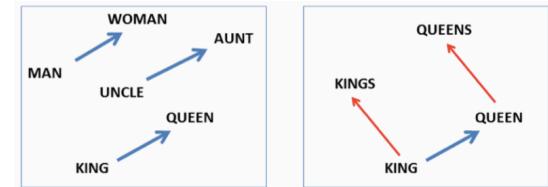
知るも知らぬも逢坂の関

AutoEncoder (自己符号化器)

- 入力と出力を同じデータで学習し,特徴量を圧縮

例1) word2vec [Mikolov+,2013] (単語ベクトル)

- 単語の持つ言語的な性質を含み,類似する単語が類似するベクトルを持つよう学習
- 著者がツールを公開→世界中で注目



上: 意味ベクトル演算の代表例

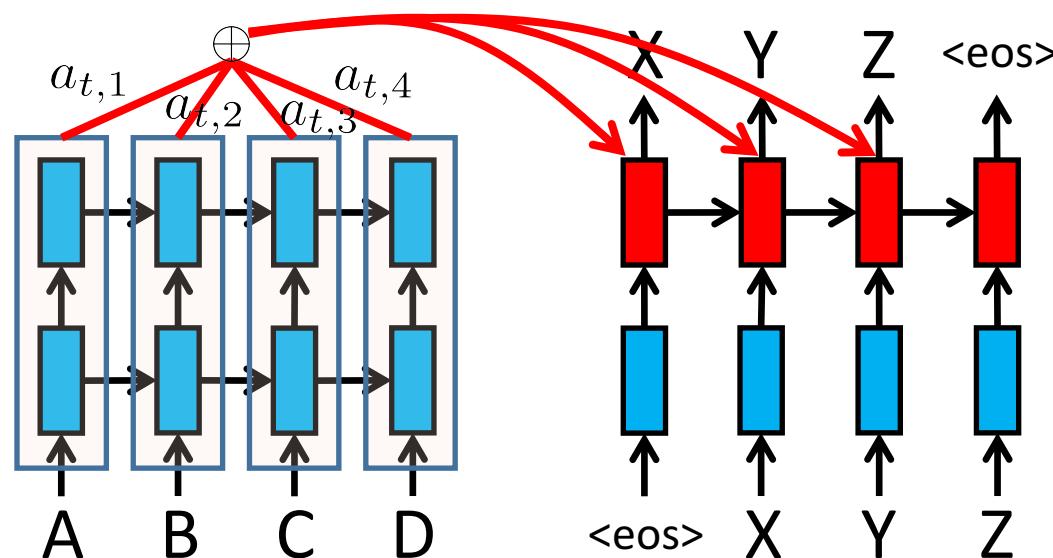
例: word2vec で各コーパスを学習し, 「パンク」検索した → コーパスやドメインによる差が出る

Wiki 日本語 2016/02 (6,235,945文)		Yahoo! 知恵袋:車 (1,439,341文)		教えてgoo:車 (391,912文)		みんなQA:QA (874,457文)	
word	distance	word	distance	word	distance	word	distance
パンク	1.000000	パンク	1.000000	パンク	1.000000	パンク	1.000000
パンク・ロック	0.741283	バースト	0.671790	バースト	0.614949	バースト	0.620422
ロック	0.738832	釘	0.656777	チューブレス	0.529546	応急	0.585563
エモ	0.723946	チューブレスタイヤ	0.610192	破裂	0.512507	テンパータイヤ	0.558305

※ word2vec は隠れ層を持たないのでDeepではないが, DLの事前学習に用いられるので DLの技術のひとつ

アテンション

- RNN の標準的なテクニックになりつつある
 - RNN翻訳 [Bahdanau+, 2015]
 - どこに「注意」して翻訳するかを学習
 - アテンションのモデル化によって,アライメントを明示的に与える必要がなくなった

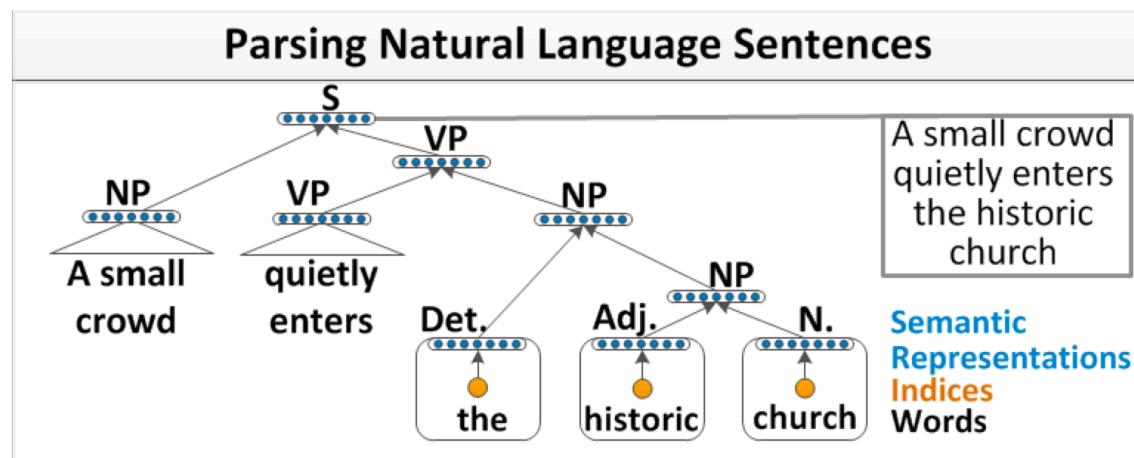


AutoEncoder (自己符号化器)

- 入力と出力を同じデータで学習し,特徴量を圧縮

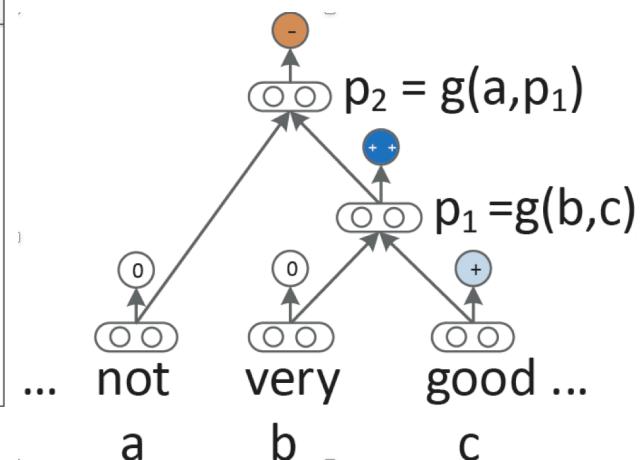
例2) Recursive Autoencoder [Socher+, 2011]

- 句や文の意味をそれを構成する単語の合成により得る
 - 単語ベクトルから文の意味を再帰的に計算
- 句の感情極性分類も実現 [Socher+, 2013]



構文構造を再帰的にNNで学習

句の極性判定



Deep Learning の実践

フレームワークが充実

- 論文で多く使われている Caffe や Torch
 - 国内は Chainer と TensorFlow がシェア拡大中

ライブラリ	開発	特徴	言語	ライセンス	GPU
Caffe	BVLC (UC Berkley)	<ul style="list-style-type: none">C++で高速 (Python, MATLABバイディング)画像処理のモデルが充実 (ModelZoo)	C++	BSD	<input type="radio"/> CUDA, cuDNN
Theano	Univ. Montreal	<ul style="list-style-type: none">Pythonベース, Numpyとの高い親和性独自のPython→C変換で高速DLライブラリの草分け	Python	BSD	<input type="radio"/> CUDA, cuDNN
Torch7	Facebook AI, DeepMind, Twitter	<ul style="list-style-type: none">独自言語 Lua (MATLABに似た実装)Lua→C変換で高速 (C/C++バイディング)拡張性が高い, ライブラリが充実	Lua	BSD	<input type="radio"/> CUDA, cuDNN
Keras	Univ. NY	<ul style="list-style-type: none">Theanoを隠蔽ライブラリが充実 → LSTM, RNN などにも対応	Python	MIT	<input type="radio"/> Theano と同じ
Chainer	PFI/PFN	<ul style="list-style-type: none">Pythonベース, Numpyとの高い親和性導入が簡単, Caffeモデルの読み込み可Python上でデバッグできる	Python	MIT	<input type="radio"/> CUDA, cuDNN
TensorFlow	Google	<ul style="list-style-type: none">Pythonベース, Numpyとの高い親和性分散環境で並列処理が可能 (Distributed Tensorflow)	C++/ Python/ Java	Apache 2.0	<input type="radio"/> CUDA, cuDNN

学習済みモデルが公開

- 転移学習
 - あるドメイン(データセット)で学習した識別器(特徴抽出器)を他ドメインでの識別器作成に役立てる
 - 特徴量抽出器として使う, 学習済みの CNN で再学習する
 - ILSVRC のカテゴリー認識のタスクを学習した CNN が転移学習に向くことが知られている
 - シーン認識 [Xiao,14], ポーズ推定 [Toshev,14], テクスチャマテリアル認識 [Cimpoi 15] など
- ModelZoo (http://caffe.berkeleyvision.org/model_zoo.html)
 - ILSVRC などの画像認識ベンチマークで好成績を収めた学習済みモデルが多数公開
 - AlexNet, VGG, GoogLeNet など, 最新の成果を簡単に試せる

転移学習の例

学習済みの VGG 19-layer をから推定した特徴量で分類器を作成し分類

1. train_pre-traind-mlp.py --batchsize 25 --plot
2. test_pre-traind-mlp.py --sample shiba.jpeg

shiba_inu



beagle



saint_bernard



1 100.00% shiba_inu

2 0.00% chihuahua

3 0.00% miniature_pinscher

4 0.00% basset_hound

5 0.00% pomeranian

1 100.00% beagle

2 0.00% american_bulldog

3 0.00% boxer

4 0.00% german_shorthaired

5 0.00% english_setter

1 100.00% saint_bernard

2 0.00% yorkshire_terrier

3 0.00% japanese_chin

4 0.00% american_pit_bull_terrier

5 0.00% basset_hound

参考書

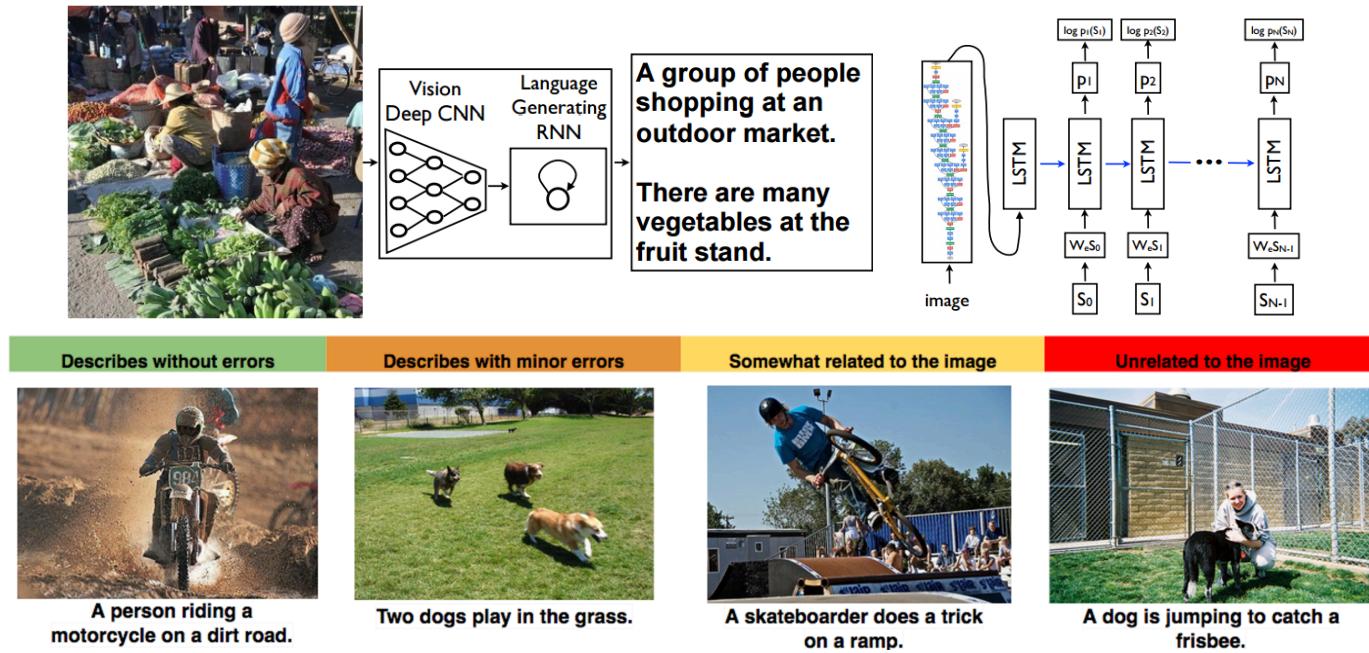
- 入門 山下隆義, 講談社, 2016/2/23
イラストで学ぶディープラーニング (KS情報科学専門書)
- 理論 岡谷貴之, 講談社, 2015/4/8
深層学習 (機械学習プロフェッショナルシリーズ)
- 開発 新納浩幸, オーム社, 2016/9/9
Chainerによる実践深層学習



デモンストレーション

画像キャプション

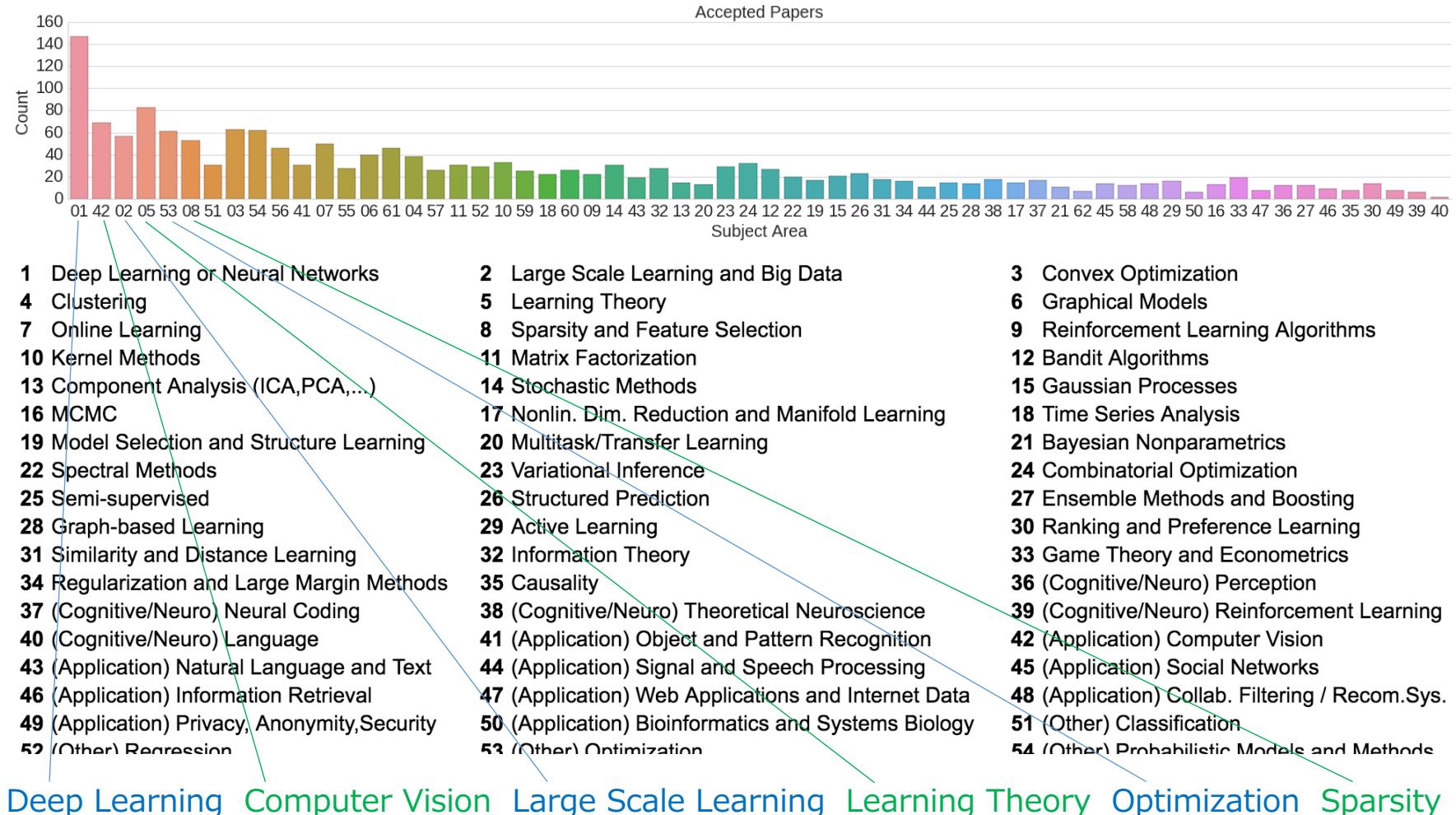
- Google NIC [Vinyals+, 2015]
 - 画像だけからキャプションを生成
 - 画像表現CNN + LSTM(文生成,翻訳)を直列に接続することで,マルチモーダルな入出力を自然に統合



最近の研究

NIPS2016 – 機械学習の国際会議

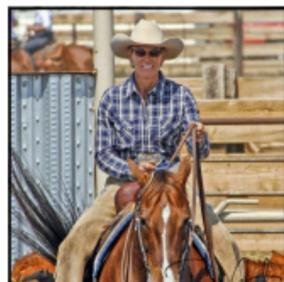
Deep Learning のブームが続き, GANなどの生成系と, 強化学習や RNN との組み合わせにシフト



物体検出

- R-CNN [Girshick+,14] (BVLC: Caffeのチーム)
 - 物体の領域候補を多数抽出 (別手法)
 - 無理やり領域を正規化し,CNNで特徴抽出
 - SVMで各領域を識別

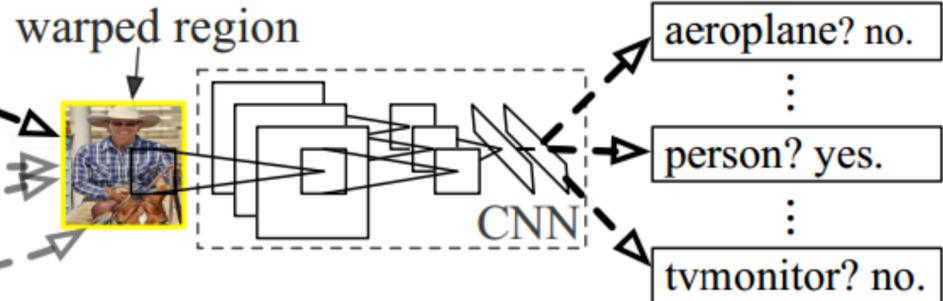
R-CNN: *Regions with CNN features*



1. Input image



2. Extract region proposals (~2k)

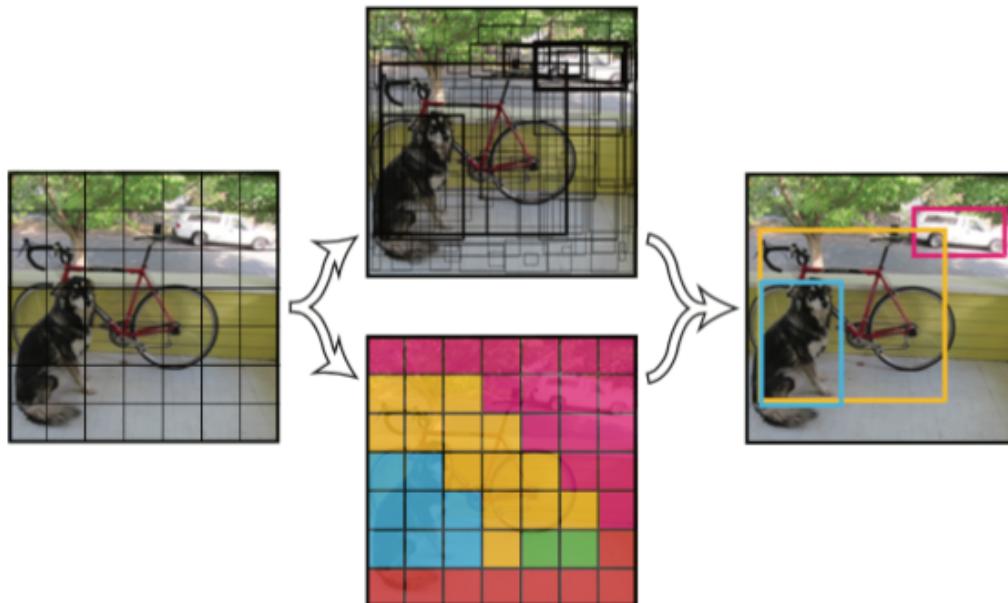


3. Compute CNN features

4. Classify regions

物体検出 (最近)

- ・ボックス候補の生成もニューラルネットを行う
 - Fast R-CNN [Girshick,15] → 1ヶ月 → Faster R-CNN [Ren,15]
 - YOLO [Redmon,15] は Faster R-CNN の倍速
 - YOLOv2 が公開されているらしい…

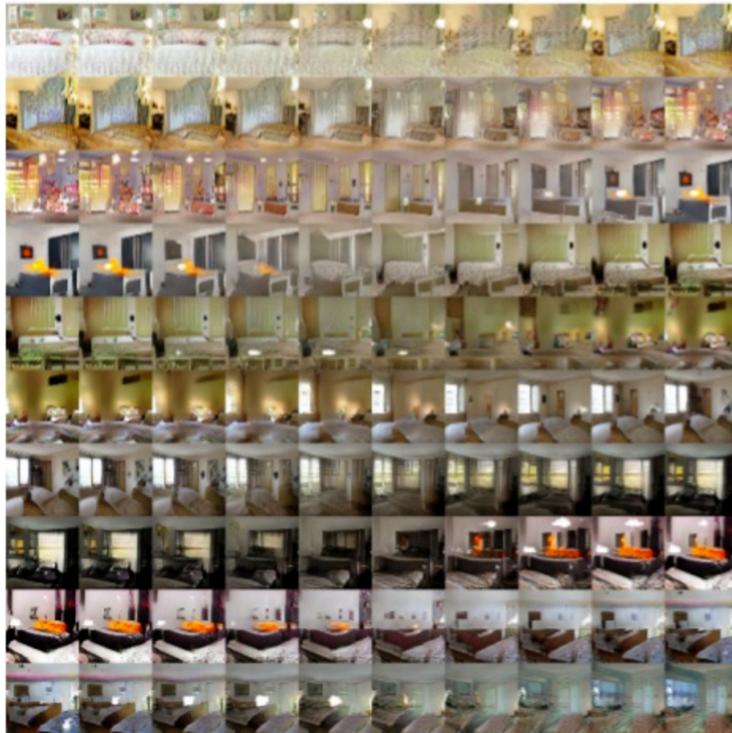


予め画像全体をグリッド分割しておき各領域ごとに物体のクラスと bounding box を求める

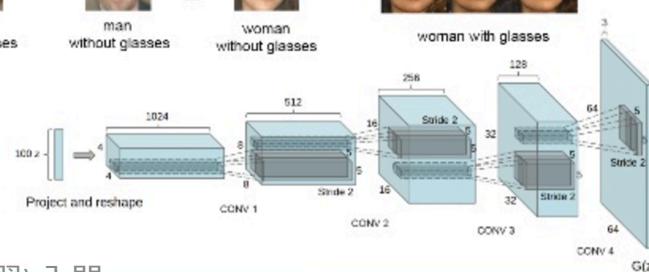
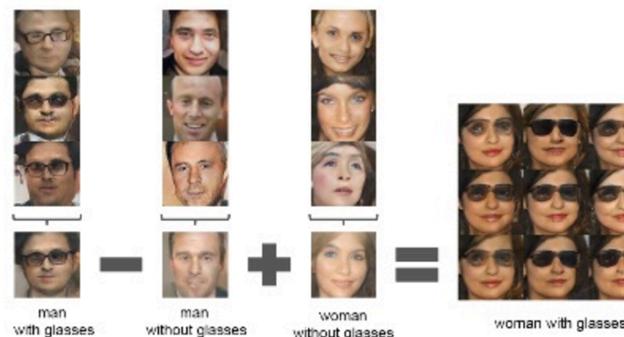
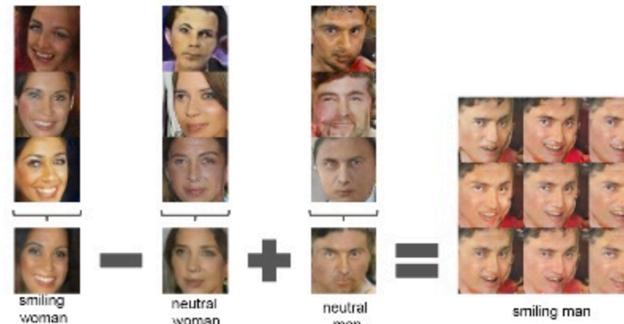
画像生成 (DCGAN)

- 自然画像のクリアな画像生成, 演算 [Radford+,16]

画像生成



画像演算



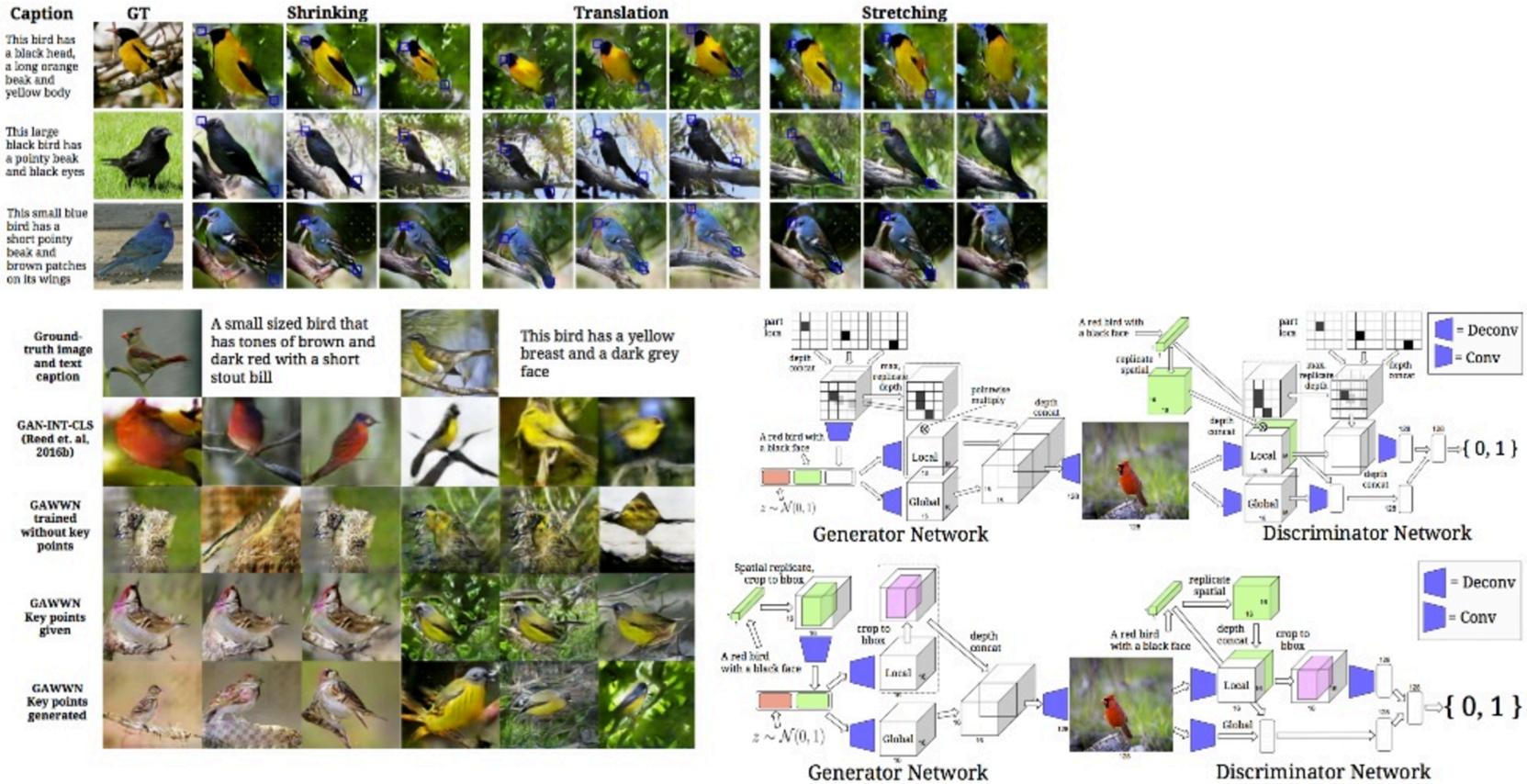
Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks.

Alec Radford, Luke Metz, Soumith Chintala.

arXiv:1511.06434. In ICLR 2016.

画像生成 (文章から画像)

- 文章で条件付けしたGAN [Reed+, 16]



Learning What and Where to Draw.

Scott Reed, Zeynep Akata, Santosh Mohan, Samuel Tenka, Bernt Schiele, Honglak Lee.
arXiv:1610.02454. In NIPS 2016.

まとめ

- “Deep learning” という言葉 (2006～)
 - [Hinton and Salakhutdinov, Science, 2006]
 - 従来より,さらに深く,幅広いニューラルネット
 - 2010年頃で7～8層程度, 現在は20層以上も
- 2012年の大ブレーク以来,アカデミア,産業界で利用
 - 2014～2015中に出された関連論文数は 1500 超える
- 音声認識,画像認識の分野で劇的な精度向上
 - 多くがすでに実用化
- 自然言語処理の分野でも,従来法の置き換えが加速
 - word2vec, RNN言語モデルなどで実用化
- 生データから目的変数まで end-to-end で構造学習
 - 従来の特徴量エンジニアリングを最小限にできる

おまけ

福島 邦彦

Kunihiko Fukushima

- ネオコグニトロン(視覚パターん認識に関する階層型神経回路モデルや選択的注意機構のモデル)などを提唱
- NHK放送技術研究所、その後大阪大学基礎工学部生物工学科

人工知能分野 で有名な人



ジェフリー・ヒントン

Geoffrey E. Hinton

- 1947年, UK生まれ
- 第一次AIの冬の時代にも研究を地道に積み上げ, Deep learning を生みだした. 誤差伝搬法の考案者の1人
- 現在は Google Brain プロジェクトの研究者とトロント大学教授を兼務

人工知能分野 で有名な人



ヤン・ルカン Yann LeCun

- 1960年,フランス生まれ
- 画像認識で知られるCNNの考案者
- 博士時代はヒントンの研究室に所属
- AT&Tベル研究所を経て,Facebook の人工知能研究所, ニューヨーク大学勤務



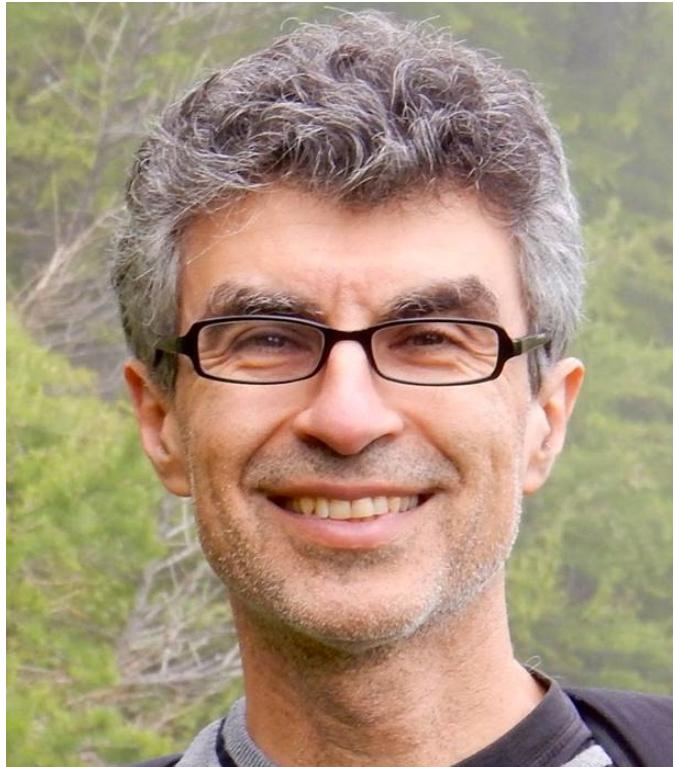
人工知能分野 で有名な人

ヨシュア・ベンジオ

Yoshua Bengio

- 1964年,フランス生まれ
- Stacked AutoEncoder の考案者
- モントリオール大学の LISA LAB (thanoやpylearn2を開発) を率い, IBM ワトソン研究所 と共同研究を行なっている
- AT&Tベル研究所では,ルカンと一緒に働いていた

人工知能分野 で有名な人



アンドリュー・ Ng

Andrew Ng

- 1976年, UK生まれ
- Google Brain を立ちあげ,
Googleの猫の論文の著者
- 今はシリコンバレーにある
Baiduの研究所で勤務
- Stanford大学, NCAPの研究員



人工知能分野 で有名な人

デミス・ハサビス Demis Hassabis

- 1976年、UK生まれ
- Deepmind (ディープマインド) の設立者
- ATARI のゲームを自動学習するDQNや囲碁の分野で世界1位の棋士に勝った Alpha GO が有名

人工知能分野
で有名な人

