

機械学習関連メモ

Ryoichi Matsumoto

2023 年 7 月 28 日

目次

1	はじめに	1
2	Deep Learning	1
2.1	Deep Learning の歴史	1
2.2	Batch Normalization	6
2.3	You Only Look Once	7
3	Section	7
3.1	Subsection	7
3.2	Subsection	7

表目次

図目次

1	AI のブームと冬の時代	2
2	ILSVRC の歴代優勝モデル (2010 年～2016 年)	2
3	人間の画像認識エラー率	3
4	2012 年以降の優勝モデル	3
5	LeNet アーキテクチャ	4
6	AlexNet アーキテクチャ	5
7	ZF Net アーキテクチャ	5

1 はじめに

機械学習に関して調べたことのメモ.

2 Deep Learning

2.1 Deep Learning の歴史

Deep Learning は AI を実現する手段のひとつであり、Deep Learning の歴史の前に AI 研究の歴史について概略を説明する。[1] [2] [3]

AI は図 1 に示すように、第一次ブームから第三次ブームがあり、これまでに 2 回、冬の時代が訪れている。第一次ブームは 1950 年代～1960 年代で、商用コンピュータの登場により AI の研究が進んだが、人間と同じ考え方を持たせるという理想への壁は高く、1970 年代にブームが冷め、1 回目の冬の時代が訪れた。第二次ブームは 1980 年代で、進化したコンピュータに知識を加えるアプローチが行われたが、コンピュータは知識の意味を理解するわけではないため、1990 年代にブームが冷め、2 回目の冬の時代が訪れた。第三次ブームは 2000 年代で、インターネットの発展により収集可能なデータの規模が劇的に増加し、AI の研究に活用できるようになり、機械学習・Deep Learning の研究が加速し、2020 年現在もなお、研究が進んでいる。

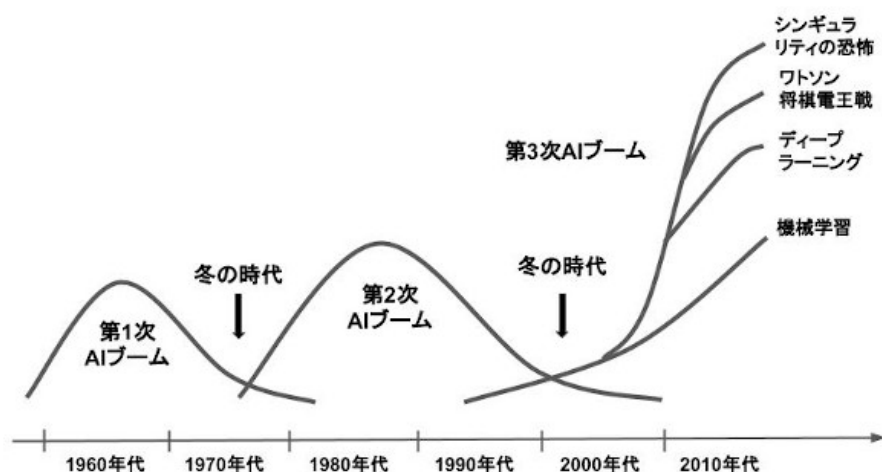


図 1 AI のブームと冬の時代

Deep Learning が最初に脚光を浴びたのは、画像認識コンテスト ILSVRC(Imagenet Large Scale Visual Recognition Challenge) [6] で、2012 年に 2011 年の優勝モデルのエラー率を約 10% 下げて優勝した [4]。以降、Deep Learning を活用したアルゴリズムが台頭し、2015 年には、人間のエラー率 5.1% [7] を下回るアルゴリズムが登場した^{*1} (図 2, 図 3)。

^{*1} 図 2 の引用元の AINOW が何を参照して人間のエラー率を 4% と定義したのかは不明

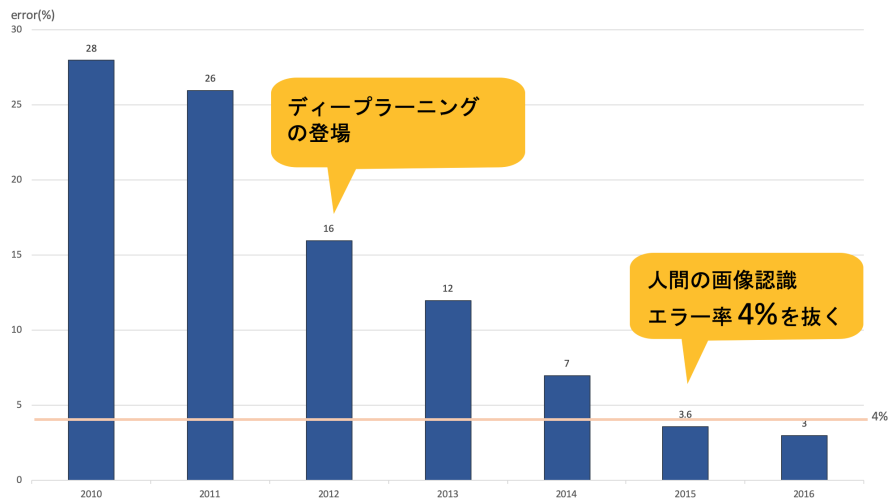


図2 ILSVRC の歴代優勝モデル (2010 年～2016 年)

Relative Confusion	A1	A2
Human succeeds, GoogLeNet succeeds	1352	219
Human succeeds, GoogLeNet fails	72	8
Human fails, GoogLeNet succeeds	46	24
Human fails, GoogLeNet fails	30	7
Total number of images	1500	258
Estimated GoogLeNet classification error	6.8%	5.8%
Estimated human classification error	5.1%	12.0%

Table 9 Human classification results on the ILSVRC2012-2014 classification test set, for two expert annotators A1 and A2. We report top-5 classification error.

図3 人間の画像認識エラー率

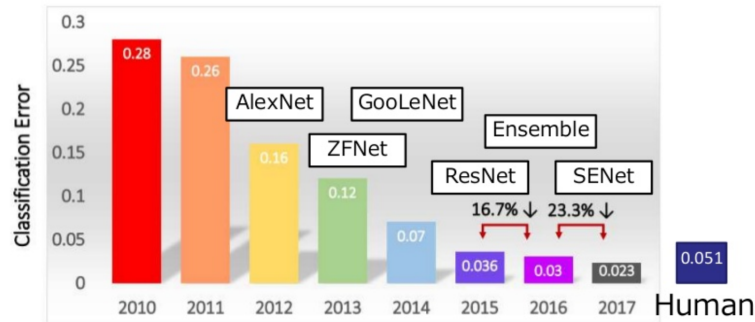
文字認識

2012 年以降の ILSVRC の優勝モデルは，図 4 に示す通り [8] *2.

*2 図中，2014 年の優勝モデルは GooLeNet となっているが，正しくは GoogLeNet

ILSVRC

- ・クラス分類タスクのエラー率 (top5 error) の推移



DeNA http://image-net.org/challenges/talks_2017/ILSVRC2017_overview.pdf

11

図 4 2012 年以降の優勝モデル

2.1.1 LeNet

LeNet は Deep Learning の根幹ともいえる CNN アーキテクチャで、図 5 に示す構造を持つ [9] [10]. 誤差逆伝播法による学習で、MNIST^{*3}などの文字認識で高精度を実現した.

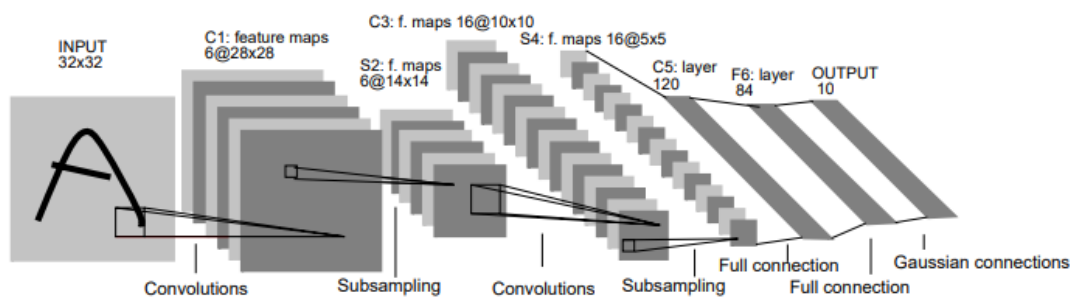


図 5 LeNet アーキテクチャ

2.1.2 AlexNet

2012 年の ILSVRC 優勝モデルで、図 6 に示す構造を持つ [11]. 下記特徴を持つ.

^{*3} THE MNIST DATABASE of handwritten digits, <http://yann.lecun.com/exdb/mnist/>

- Rectified Linear Units(ReLU) による非線形変換導入で勾配消失問題の解決と学習の高速化.
- Local Response Normalization (LRN) による正規化で学習の安定化と高速化. 当時, Local Contrast Normalization(LCN) と比較してより適切な正規化法と主張していたが, 後の VGG Net で効果が認められなかったことや, Batch Normalization の登場により近年のモデルでは利用されていない.
- Overlapping pooling によるプーリング層でのオーバーラップにより過学習を提言.
- 全結合層の Dropout による学習時のニューロン無効化によりモデルの汎化性能を向上.
- Momentum SGD による最適化で学習を安定化.
- 重み減衰による正則化で過学習を抑制.
- 学習率減衰で学習の収束度を向上.

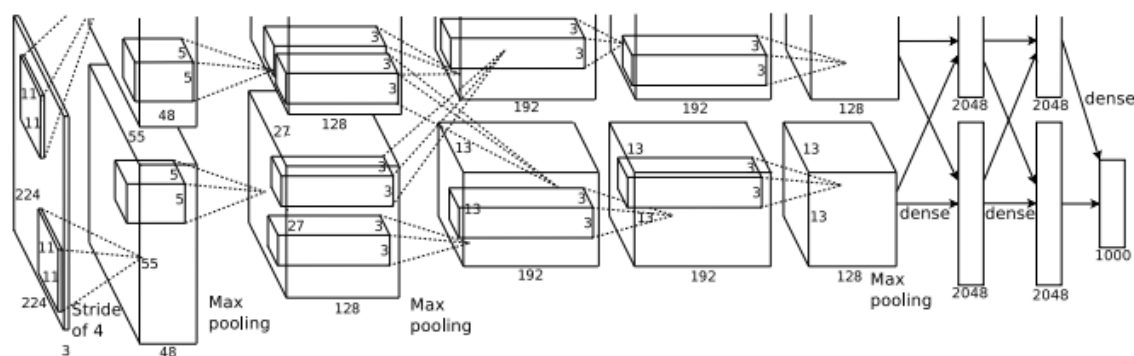


図 6 AlexNet アーキテクチャ

2.1.3 ZFNet

2013 年の ILSVRC 優勝モデルで図 7 に示す構造を持つ [12]. 下記特徴を持つ.

- Alex Net に対して, 第一層目のカーネルサイズを 11x11 から 7x7 に変更することで, 高周波及び低周波の情報量を抑制
- Alex Net に対して, 第一層目のストライドを 4 から 2 に変更することで, 第二層目に生じるエイリアシングを削減

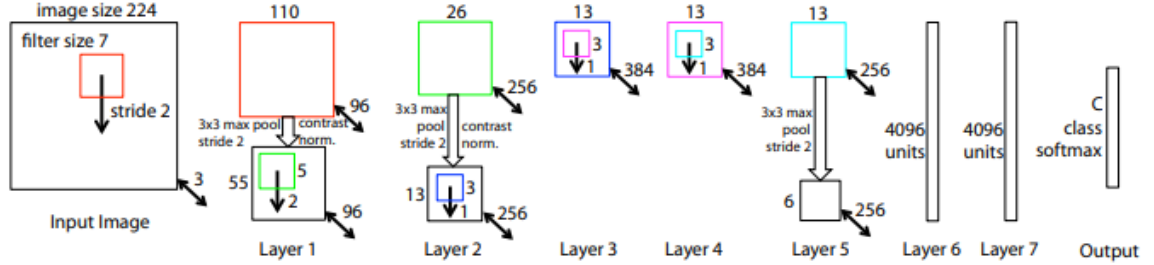


図7 ZF Net アーキテクチャ

2.1.4 GoogLeNet

2.1.5 ResNet

2.1.6 SENet

2.2 Batch Normalization

2.2.1 背景

訓練時とテスト時で入力（共変量）の分布が異なる状況は「共変量シフト」と呼ばれ、このような場合は良い学習結果が得られない。共変量シフト下においては、忘却型の学習アルゴリズムやパラメータ学習法、モデル選択法等が用いられる [13] が、Deep Learning においては、学習過程でパラメータ更新により各層の Activation の分布が変わる「内部共変量シフト (Internal Covariate Shift)」と呼ばれる問題がある [14] [15]。

2.2.2 概説

Deep Learning は画像、音声など様々な分野で急速な技術進化をもたらした。Deep Network の学習方法として確率的勾配降下法 (Stochastic gradient descent ; SGD) が効果的であることが知られている [16]。SGD の派生として、Momentum, AdaGrad, RMSprop, Adam, AdaBound, RAdam などがある。SGD は Network のパラメータ Θ を、loss を最小化するように最適化する。

$$\Theta = \arg \min_{\Theta} \frac{1}{N} \sum_{i=1}^N l(x_i, \Theta) \quad (1)$$

$x_{1...N}$ はデータ数 N サンプルの学習データセットで、SGD での学習ステップは m サンプルのミニバッチ $x_{1...m}$ に対して行う。ミニバッチは、パラメータについての損失関数の勾配近似に使用され、式 (2) で導出する。

$$\frac{1}{m} \frac{\partial l(x_i, \Theta)}{\partial \Theta} \quad (2)$$

SGD では、学習率をはじめとするハイパーパラメータや、モデルパラメータの初期値の調整が

重要である。各層の入力が一つ前の層のモデルパラメータの影響を受け、この影響がネットワークが深いほど大きくなるため、学習が複雑化する。学習によってモデルパラメータが更新され、各層の入力分布が変わり学習の系が変わることで、内部共変量シフトが引き起こされる。この内部共変量シフトを軽減することで学習を安定化する手法の一つが Batch Normalization である。

ネットワークの一部を切り出して、式 (3) の損失関数を考える。

$$l = F_2(F_1(u, \Theta_1), \Theta_2) \quad (3)$$

F_1, F_2 は任意の変換関数、 Θ_1, Θ_2 は損失 l の最小化により学習されるパラメータである。 Θ_2 の学習は、入力を $x = F_1(u, \Theta_1)$ とおき、

$$l = F_2(x, \Theta_2) \quad (4)$$

とすると、パラメータ Θ_2 の更新

$$\Theta_2 \leftarrow \Theta_2 - \frac{\alpha}{m} \sum_{i=1}^m \frac{\partial F_2(x_i, \Theta_2)}{\partial \Theta_2} \quad (5)$$

は、入力 x の独立したネットワーク F_2 と等価である。したがって、学習効果を高める入力分布の性質は、学習データとテストデータで同じ分布を持たせることと同様に、サブネットワークの学習に対して有効である。つまり、学習時間経過によらず x の分布を一定に保つことは有効であり、 Θ_2 の再調整を不要とすることができる。

そこで、各学習ステップで Activation の白色化を行うことを考える。学習後のバイアス b を加算し、学習データに対して計算された Activation の平均で減算した結果で正規化する入力 u を持つ層は $\hat{x} = x - E[x]$ で示される。ここで、 $x = u + b$ 、 $E[x] = \frac{1}{N} \sum_{i=1}^N x_i$ である。勾配降下ステップが $E[x]$ の b についての依存がないとき、 b の更新は $b \leftarrow b + \Delta b$ 、 $\Delta b \propto -\partial l / \partial [\hat{x}]$ で示され、 $u + (b + \Delta b) - E[u + (b + \Delta b)] = u + b - E[u + b]$ となる。

d 次元の入力 $x = (x^{(1)} \dots x^{(d)})$ を持つネットワークの層について

Batch Normalization では Activation の出力 x とペアのパラメータ γ, β を用いて、

$$y = \gamma x + \beta \quad (6)$$

2.3 You Only Look Once

You Only Look Once(YOLO) は物体検出手法の一つで、「検出」と「識別」を同時に行うことで、処理時間の遅延を解消することを特長とする。

3 Section

...

3.1 Subsection

...

3.2 Subsection

...

参考文献

- [1] AI(人工知能) の歴史 | 時系列で簡単解説, <https://ledge.ai/history-of-ai/>
- [2] 「Deep Learning」の進化で AI を活用した新しいビジネスが生まれる, https://www.nttcom.co.jp/comware_plus/img/201705_DeepLearning.pdf
- [3] 人工知能は人間を超えるか - ディープラーニングの先にあるもの, <https://www.ipa.go.jp/files/000050258.pdf>
- [4] 初心者でもわかるディープラーニング - 基礎知識から AI との違い、導入プロセスまで細かく解説, <https://ainow.ai/2019/08/06/174245/>
- [5] [サーベイ論文] 畳み込みニューラルネットワークの研究動向, http://mprg.jp/data/MPRG/F_group/F188_uchida2017.pdf
- [6] IMAGENET Large Scale Visual Recognition Challenge (ILSVRC) 2017 Overview, http://image-net.org/challenges/talks_2017/ILSVRC2017_overview.pdf
- [7] ImageNet Large Scale Visual Recognition Challenge, <https://arxiv.org/abs/1409.0575>
- [8] 畳み込みニューラルネットワークの高精度化と高速化, <https://www.slideshare.net/ren4yu/ss-145689425>
- [9] GradientBased Learning Applied to Document Recognition, http://vision.stanford.edu/cs598_spring07/papers/Lecun98.pdf
- [10] Understanding and Implementing LeNet-5 CNN Architecture (Deep Learning), <https://towardsdatascience.com/understanding-and-implementing-lenet-5-cnn-architecture-deep-learning-a2d531ebc342>
- [11] ImageNet Classification with Deep Convolutional Neural Networks, <https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>
- [12] Visualizing and Understanding Convolutional Networks, <https://arxiv.org/abs/1311.2901>
- [13] 共変量シフト下での教師付き学習, https://www.jstage.jst.go.jp/article/jnns/13/3/13_3_111/_article/-char/ja/
- [14] Batch Normalization : ニューラルネットワークの学習を加速させる汎用的で強力な手法, https://deepage.net/deep_learning/2016/10/26/batch_normalization.html
- [15] Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, <https://arxiv.org/abs/1502.03167>
- [16] Optimization: Stochastic Gradient Descent, <http://deeplearning.stanford.edu/tutorial/supervised/OptimizationStochasticGradientDescent/>
- [17] 【物体検出手法の歴史 : YOLO の紹介】, <https://qiita.com/mdo4nt6n/items/68dcda71e90321574a2b>

- [18] Object Detection YOLO v1 , v2, v3, <https://medium.com/@venkatakrishna.jonnalagadda/object-detection-yolo-v1-v2-v3-c3d5eca2312a>
- [19] You Only Look Once: Unified, Real-Time Object Detection, <https://arxiv.org/abs/1506.02640>
- [20] YOLO9000: Better, Faster, Stronger, <https://arxiv.org/abs/1612.08242>
- [21] YOLOv3: An Incremental Improvement, <https://arxiv.org/abs/1804.02767>