

機械学習関連メモ

Ryoichi Matsumoto

2023 年 8 月 14 日

目次

1	はじめに	2
2	Deep Learning	2
2.1	Deep Learning の歴史	2
2.2	エッジ AI	4
2.3	Deep Learning のモデル構造	9
2.4	Deep Learning の学習テクニック	11
3	Section	13
付録 A	エッジ AI 関連の市場規模・動向・関連企業	14

表目次

1	FLOPs vs Weights	8
---	------------------	---

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)	3
3	人間の画像認識エラー率	3
4	2012 年以降の優勝モデル	4
5	エッジ AI 関連市場の動向	5
6	モデルごとの FLOPs と Weights	9
7	LeNet アーキテクチャ	10
8	AlexNet アーキテクチャ	10
9	ZF Net アーキテクチャ	11

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 の研究に活用ができるようになり、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.

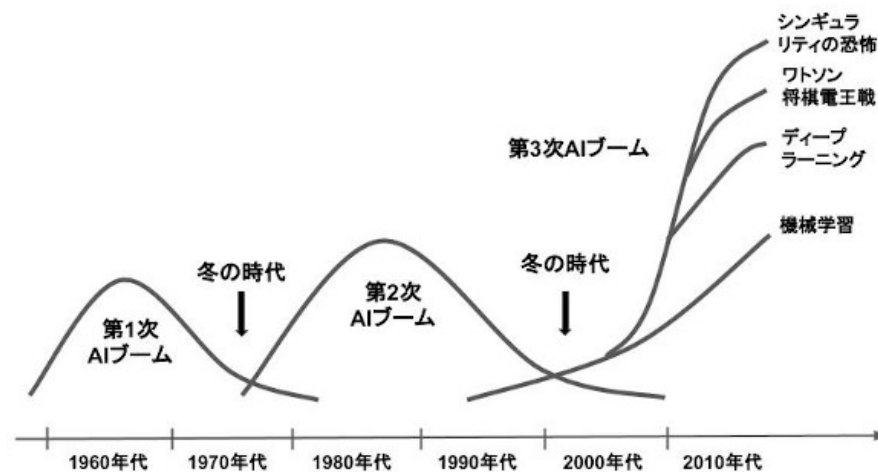


図 1 AI のブームと冬の時代

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).

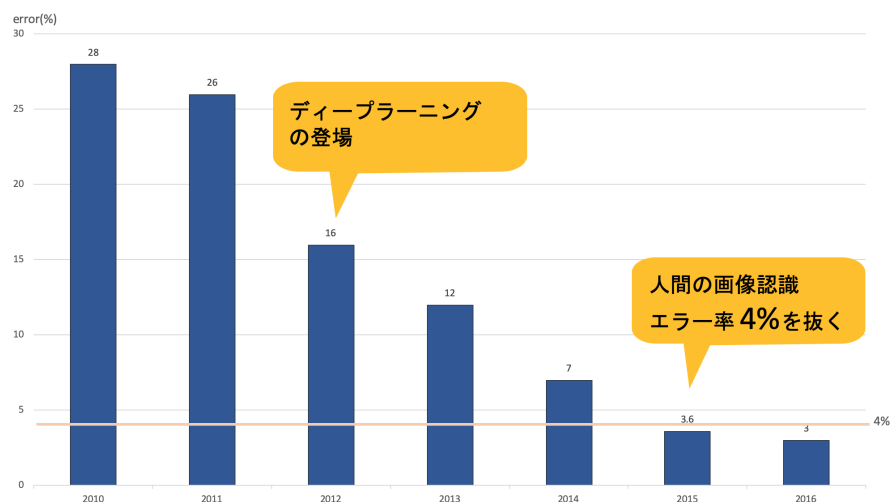


図 2 ILSVRC の歴代優勝モデル (2010 年～2016 年)

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

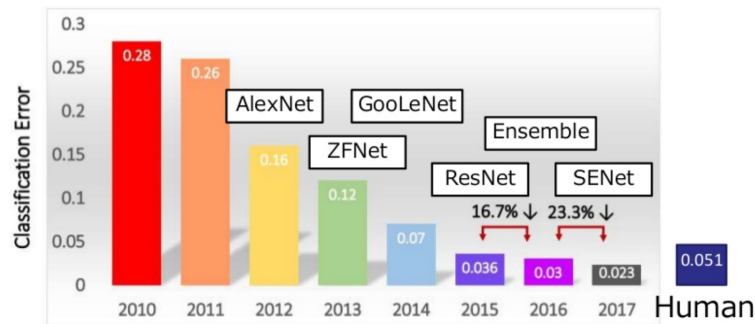
2012 年以降の ILSVRC の優勝モデルは, 図 4 に示す通り [8] *².

*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

*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

以降、様々なアーキテクチャが考案され、物体検出・音声認識・異常検知・画像/音声の生成など、多くの用途で活用が進んでいる。

2.2 エッジ AI

DeepLearning 技術の発展に伴い、車載や産業、スマートホームといった多くの分野への活用が期待されている。特に、エッジデバイス上で処理するエッジ AI は、2023 年現在、産業分野を中心に活用が進み、車載分野における自動運転は 2030 年以降に活用が広まるものと予想されている。

図 5 はエッジ AI 関連市場の予測を示すものである。2035 年頃まではスマートカメラやスマートホームなど産業・民生分野を中心に成長し、2030 年ごろから自動運転が実用化されることが予測されている。

なお本グラフは、Appendix 付録 A に示す数値をもとに、年平均成長率 (CAGR) で各年の数値を補完したものである。

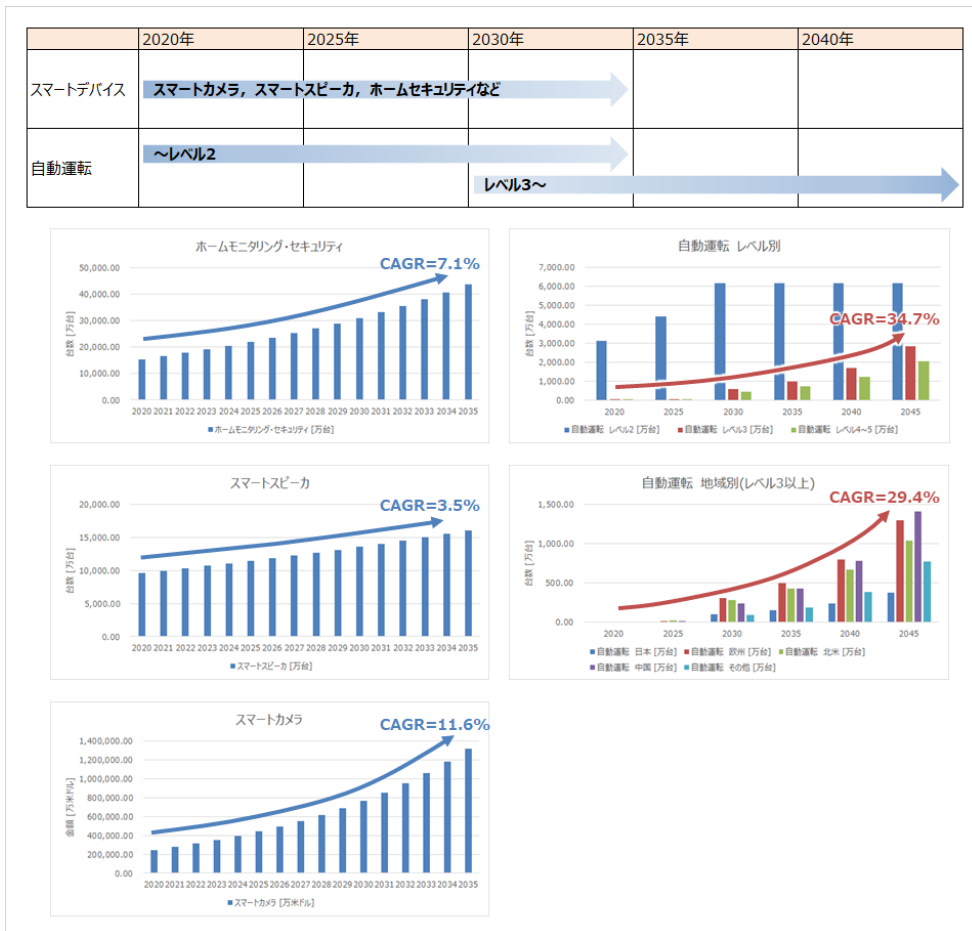


図5 エッジ AI 関連市場の動向

AI 処理をエッジ化することにより、下記のような効果が期待できる。

- リアルタイム性・低遅延
 - ー ネットワーク通信を介さないデータ処理で即時に応答
- 情報漏洩リスクの低減
 - ー クラウド利用がない為、機密情報の漏洩リスクを低減
- 通信負荷と通信コストの低減
 - ー データ処理がエッジで完結することにより、ネットワークの通信負荷とクラウド利用コストを低減
- 低消費電力
 - ー 消費電力が低いエッジデバイスでの処理で、電力コストと環境負荷を低減

- デバイスの軽薄短小化
 - － 小さな装置への実装や、設置容易性を向上

実際にカメラに AI を組み込んだ障害物検知や人流分析、製造装置に AI を組み込んだ異常検知や機器制御など、実証実験や運用が進んでいる。AI の導入が進み始めた中で、顔認証をはじめとし、人種的・性的などの側面から倫理問題が顕在化してきた。

実際に発生した AI の倫理問題は、例えば、

- Google フォトの人種差別：画像認識で黒人をゴリラと誤判定
- Amazon の人材採用 AI による性差別：男性の評価を高く評価
- リクナビによる内定辞退率の販売：AI を用いて算出した学生の内定辞退率のデータを企業に販売
- 東京 2020 オリンピック選手村での自動運転事故：トヨタ「e パレット」が自動運転中に選手と接触
- フェイクニュースの多発：DeepFake による嘘情報の拡散

が挙げられる。

AI が識別・判断した根拠を説明できず、実害が生じたにも関わらず責任の所在が曖昧になる場合も多い。これらの倫理問題を受け、政府や企業は倫理問題への対策を進めている。

- 日本政府が規定した AI 社会原則 (<https://www8.cao.go.jp/cstp/aigensoku.pdf>)
 - － 人間中心の原則
 - － 教育・リテラシーの原則
 - － プライバシー確保の原則
 - － セキュリティ確保の原則
 - － 公正競争確保の原則
 - － 公平性・説明責任及び透明性の原則
 - － イノベーションの原則
- AI 倫理の 7 原則 (Google, <https://ai.google/responsibility/principles/>)
 - － Be socially beneficial (社会への有益性)
 - － Avoid creating or reinforcing unfair bias (不公平なバイアスの防止)
 - － Be built and tested for safety (安全性確保のための開発)
 - － Be accountable to people (人々への説明責任)
 - － Incorporate privacy design principles (プライバシーの設計原則の組み込み)
 - － Uphold high standards of scientific excellence (科学手卓越性の高位平準化)
 - － Be made available for uses that accord with these principles (これらの原則に沿ったユーザへの技術提供)
- AI 倫理原則 (パナソニック, <https://monoist.itmedia.co.jp/mn/articles/2208/30/news073.html>)
 - － 「より良いくらしとより良い社会」を実現すること
 - － 安全のための設計、開発、検証を行うこと
 - － 人権と公平性を尊重すること
 - － 透明性と説明責任を重視すること

- お客様のプライバシーを保護すること
- 責任ある AI：AI 倫理とガバナンス (accenture, <https://www.accenture.com/jp-ja/services/applied-intelligence/ai-ethics-governance>)
 - 意図しないバイアスの最小化
 - AI の透明性を確保する
 - 従業員に機会を与える
 - データプライバシーとセキュリティを守る
 - 顧客と市場に利益をもたらす
- HPE における AI の倫理と原則 (HeulettPackard, <https://www.hpe.com/jp/ja/solutions/artificial-intelligence/ethics.html>)
 - AI のプライバシー対応のセキュリティ
 - AI の人間重視の原則
 - AI のインクルーシブ原則
 - AI の頑強な原則
 - 責任ある AI
- AI 倫理ガイドライン (AWL, <https://awl.co.jp/ai-ethics-guidelines/>)
 - 人間中心の AI 社会の実現
 - 法令遵守
 - AI の適正な利活用
 - 安全性とセキュリティの確保
 - プライバシーへの配慮
 - 公平性の尊重・非差別
 - 透明性等
 - AI の発展と人材育成
- OKI グループ AI 原則 (OKI, <https://www.oki.com/jp/technology/ai/principle.html>)
 - 人権の尊重
 - 説明と透明性
 - 対話と協調
 - 安全およびデータの取扱い
 - 人材育成

以上のことから、非常に多くの分野で活用されるようになった DeepLearning 技術だが、エッジへの組み込みに加え倫理問題への対策も必要である。

2.2.1 エッジへの AI 実装課題

モデルの性能向上に伴い計算に必要な資源も増加している。計算機の資源、特に GPU やスーパーコンピュータの性能も向上しているが、スマートフォンやスマートカメラなど計算資源が限られているエッジデバイスでは、実行可能なモデルも制限され、AI の小型化と性能維持が課題である。

表 1 FLOPs vs Weights

Model Name	FLOPs	Weights
Xception	8357403496	22801424
VGG16	15470264320	138357544
VGG19	19632062464	143667240
ResNet50	3857973248	25530472
ResNet101	7570194432	44496488
ResNet152	11282415616	60117096
InceptionV3	5713216096	23800136
InceptionResNetV2	13155794016	55782920
MobileNet	568740352	4210088
MobileNetV2	300774272	3470760
DenseNet121	2834161664	7895208
DenseNet169	3359843328	13991080
DenseNet201	4291365888	19784872
NASNetMobile	563638816	5253240
NASNetLarge	23783414658	88556482
EfficientNetB0	388121280	5246532
EfficientNetB1	690912160	7732136
EfficientNetB2	998832224	9042426
EfficientNetB3	1836129536	12145936
EfficientNetB4	4413319168	19216416
EfficientNetB5	10306979360	30217048
EfficientNetB6	19136716704	42816272
EfficientNetB7	37868782912	66037240
EfficientNetV2B0	719342144	7079096
EfficientNetV2B1	1198640192	8069980
EfficientNetV2B2	1700157088	10013798
EfficientNetV2B3	3015809440	14249190
EfficientNetV2S	8375340032	21304616
EfficientNetV2M	24609813184	53847324
EfficientNetV2L	56127521408	118002696
ConvNeXtTiny	25491808974	27011848
ConvNeXtSmall	50982829662	48625672
ConvNeXtBase	90635773534	85772648
ConvNeXtLarge	203929679454	191474728
ConvNeXtXLarge	362540942942	339054312

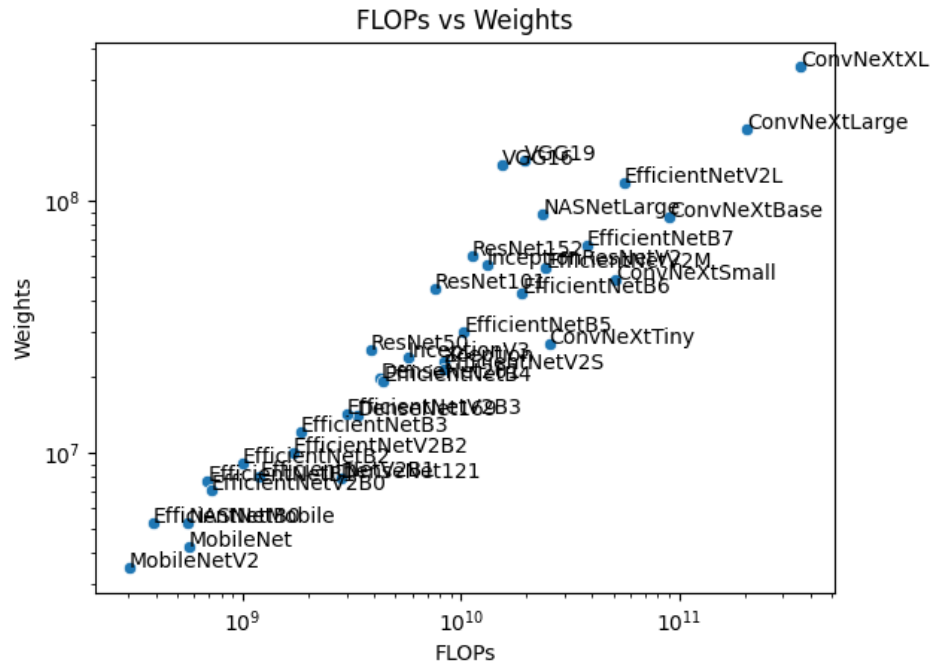


図6 モデルごとの FLOPs と Weights

2.2.2 AI 倫理問題対策の課題

AI の社会実装において、AI の判断に対して責任を持つことが基本となる。しかし、AI はブラックボックス的である為、判断根拠を人が解釈することが困難である。

AI を活用する範囲を人が解釈可能な範囲に限定したり、説明可能 AI(Explainable AI) 技術を活用したりするなどの対策が挙げられる。

2.3 Deep Learning のモデル構造

2.3.1 LeNet

LeNet は Deep Learning の根幹ともいえる CNN アーキテクチャで、図 7 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。

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

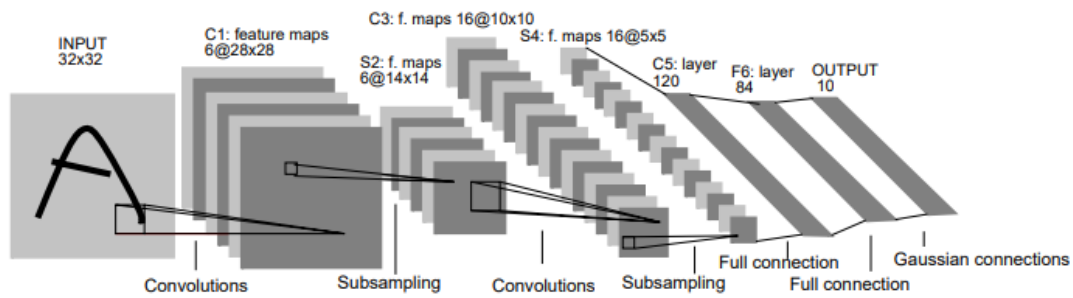


図 7 LeNet アーキテクチャ

2.3.2 AlexNet

2012 年の ILSVRC 優勝モデルで、図 8 に示す構造を持つ [11]。下記特徴を持つ。

- 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。

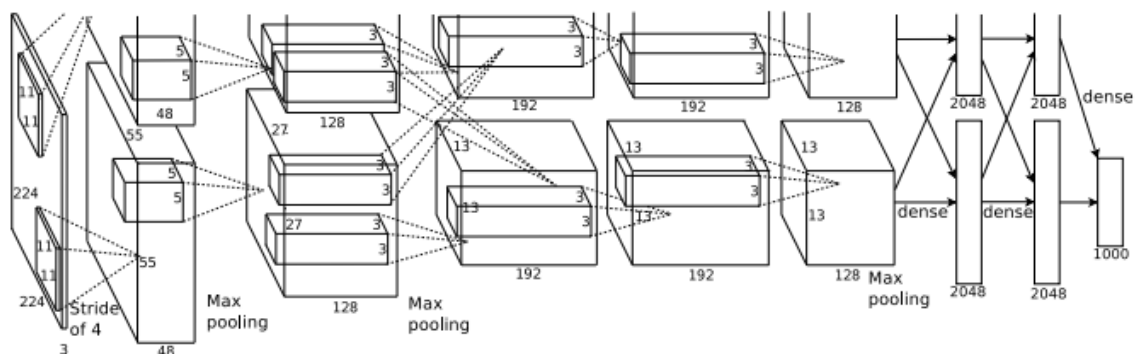


図 8 AlexNet アーキテクチャ

2.3.3 ZFNet

2013 年の ILSVRC 優勝モデルで図 9 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

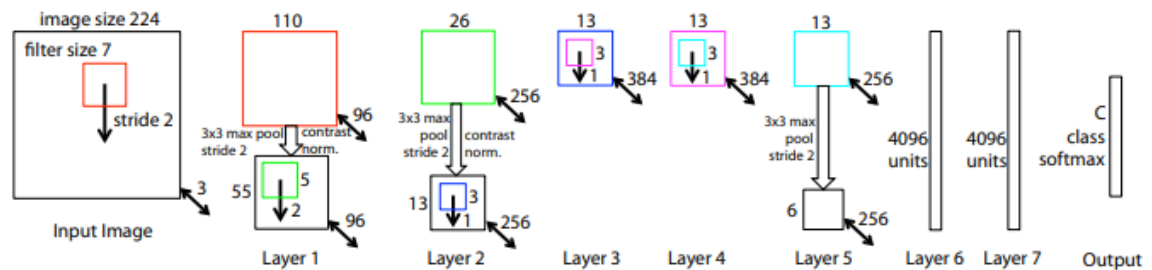


図 9 ZF Net アーキテクチャ

2.3.4 GoogLeNet

T.B.D

2.3.5 ResNet

T.B.D

2.3.6 SENet

T.B.D

2.3.7 YOLO(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.

2.4 Deep Learning の学習テクニック

2.4.1 Batch Normalization

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バ

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].

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,

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

3 Section

T.B.D

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>

付録 A エッジ AI 関連の市場規模・動向・関連企業

● 市場背景

- － 製造業におけるロボティクス需要の高まり (<https://prtimes.jp/main/html/rd/p/000002320.000072515.html>)
 - * 2022 年 83 億米ドル
 - * 2032 年 412 億米ドル
 - * CAGR 39.5% → Web サイトは 17.4% の誤植かもしれない ($((\frac{412}{83})^{\frac{1}{10}} - 1 \simeq 17.4)$)
- － スマートデバイスの普及 (<https://prtimes.jp/main/html/rd/p/000002320.000072515.html>)
 - * ビデオエンターテインメント (<https://japan.cnet.com/article/35205930/>)
 - ・ 2023 年 2 億 7,870 万台
 - ・ 2027 年 3 億 640 万台
 - ・ CAGR 2.4%
 - * ホームモニタリング・セキュリティ (<https://japan.cnet.com/article/35205930/>)
 - ・ 2023 年 1 億 9,180 万台
 - ・ 2027 年 2 億 5,220 万台
 - ・ CAGR 7.1%
 - * 照明 (<https://japan.cnet.com/article/35205930/>)
 - ・ 2023 年 1 億 190 万台
 - ・ 2027 年 1 億 6,970 万台
 - ・ CAGR 13.6%
 - * スマートスピーカ (<https://japan.cnet.com/article/35205930/>)
 - ・ 2023 年 1 億 690 万台
 - ・ 2027 年 1 億 2,250 万台
 - ・ CAGR 3.5%

- * 空調 (<https://japan.cnet.com/article/35205930/>)
 - ・ 2023 年 2,490 万台
 - ・ 2027 年 2,970 万台
 - ・ CAGR 4.5%
- * その他 (<https://japan.cnet.com/article/35205930/>)
 - ・ 2023 年 1 億 5,290 万台
 - ・ 2027 年 2 億 1,100 万台
 - ・ CAGR 7.7%
- * 合計 (<https://japan.cnet.com/article/35205930/>)
 - ・ 2023 年 8 億 5,710 万台
 - ・ 2027 年 10 億 9,150 万台
 - ・ CAGR 6.2%
- * スマート小売機器市場 (<https://newscast.jp/news/0853835>)
 - ・ 2021 年 255 億 8,000 万米ドル
 - ・ 2030 年 816 億 8,000 万米ドル
 - ・ CAGR 15.2%
- * スマートカメラ (<https://www.gii.co.jp/report/smrc1308705-smart-cameras-market-forecasts-global-analysis-by.html>)
 - ・ 2023 年 35 億 4,000 万米ドル
 - ・ 2030 年 76 億 3,000 万米ドル
 - ・ CAGR 11.57%
- 自動運転市場の拡大 (<https://prtimes.jp/main/html/rd/p/000002320.000072515.html>)
 - * 自動運転（レベル別台数） (<https://monoist.itmedia.co.jp/mn/articles/2208/04/news070.html>)
 - ・ レベル 2 2022 年 3,608 万台
 - ・ レベル 2 2030 年 6,176 万台
 - ・ レベル 2 2045 年 6,166 万台
 - ・ レベル 3 2022 年 3 万台
 - ・ レベル 3 2030 年 580 万台
 - ・ レベル 3 2045 年 2,847 万台
 - ・ レベル 4/5 2022 年 9 万台
 - ・ レベル 4/5 2030 年 433 万台
 - ・ レベル 4/5 2045 年 2,051 万台
 - * 自動運転レベル (参考) (<https://jidounten-lab.com/autonomous-level>)
 - ・ 自動運転レベル 1：運転支援
 - ・ 自動運転レベル 2：部分運転自動化
 - ・ 自動運転レベル 3：条件付き運転自動化
 - ・ 自動運転レベル 4：高度運転自動化
 - ・ 自動運転レベル 5：完全運転自動化
 - * 自動運転（地域別台数） (<https://monoist.itmedia.co.jp/mn/articles/>)

2208/04/news070.html)

- ・ 日本 2022 年 1 万台
- ・ 日本 2030 年 97 万台
- ・ 日本 2045 年 377 万台
- ・ 欧州 2022 年 2 万台
- ・ 欧州 2030 年 306 万台
- ・ 欧州 2045 年 1,294 万台
- ・ 北米 2022 年 5 万台
- ・ 北米 2030 年 278 万台
- ・ 北米 2045 年 1,042 万台
- ・ 中国 2022 年 4 万台
- ・ 中国 2030 年 237 万台
- ・ 中国 2045 年 1,411 万台
- － 低 遅 延 (<https://www.mordorintelligence.com/ja/industry-reports/edge-ai-hardware-market>)
- － 運用コスト削減 (<https://www.mordorintelligence.com/ja/industry-reports/edge-ai-hardware-market>)
- エッジ AI の特徴と活用シーン (https://www.toshiba-dme.co.jp/dme/dig/edge_solution/trend.htm)
 - － リアルタイム性
 - * ネットワークを介さないデータ処理で低遅延に実行可能
 - * 製造現場の不良品検出や自動運転などでの活用が期待される
 - － 情報漏洩リスクの低減
 - * クラウド利用がない為、機密情報の漏洩リスクが低減できる
 - * 防犯カメラ・金融サービス・入退室管理などでの活用が期待される
 - － 通信負荷とコスト削減
 - * データ処理がエッジで完結する、またはクラウドへアップロードするデータ量削減による通信量削減で、コスト低減
 - * 農業や水産業などの通信環境の良くないシーンでの活用が期待される
 - － 低消費電力
 - * 用途・要求に最適化したソリューションで消費電力を削減（対パソコンなど）
 - － デバイスの軽薄短小化
 - * 用途・要求に最適化したソリューションで設置面積を削減（対パソコンなど）
 - － 低 遅 延 (https://aismiley.co.jp/ai_news/edge-ai-software-market-expects-20-3-growth-by-2027/)
- 市場規模・予測
 - － エッジ AI 市場
 - － 日本 (<https://prtimes.jp/main/html/rd/p/000002320.000072515.html>)
 - * 2022 年 81.2 億米ドル
 - * 2032 年 421 億米ドル
 - － AI ビジネス国内市場 (<https://www.fuji-keizai.co.jp/press/detail.html?>

- cid=19039)
 - * 2018 年度 5,301 億円
 - * 2022 年度 1 兆 2,000 億円
 - * 2030 年度 2 兆 1,286 億円
- エッジ AI コンピューティング市場 (<https://www.fuji-keizai.co.jp/press/detail.html?cid=19039>)
 - * 2018 年 110 億円
 - * 2030 年 664 億円
- エッジ AI 市場 (<https://it.impress.co.jp/articles/-/24005>, <https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r05/pdf/n4800000.pdf>)
 - * 2021 年度 76.6 億円
 - * 2022 年度 117 億円
 - * 2026 年度 431 億円
 - * 通信コストの削減, 低遅延処理の実現, プライバシーリスクの低減メリット
- 国内 AI システム市場 (https://aismiley.co.jp/ai_news/idc-japan-2023/)
 - * 2023 年 4,930 億 7,100 万円
 - * 2027 年 1 兆 1,034 億 7,700 万円
 - * CAGR 23.2%
- エッジ コンピューティング市場 (<http://www.ex-press.jp/lfwj/lfwj-news/lfwj-biz-market/52519/>)
 - * 2030 年 1559 億米ドル
 - * 2023-2030 年 CAGR 37.9% 成長
- エッジ AI コンピューティング市場 (<http://ex-press.jp/lfwj/lfwj-news/lfwj-biz-market/45075/>)
 - * 2020 年 90 億 9600 万米ドル
 - * 2030 年 596 億 3300 万米ドル
 - * CAGR 21.2% 成長
- エッジ AI チップ (<https://newscast.jp/news/4431496>)
 - * 2030 年 132 億米ドル
- エッジ AI プロセッサ (<https://prtimes.jp/main/html/rd/p/000002320.000072515.html>)
 - * 2022 年 27.5 億米ドル
 - * 2032 年 110 億米ドル
 - * CAGR 15.56%
- エッジ AI プロセッサ タイプ別内訳 (<https://prtimes.jp/main/html/rd/p/000002320.000072515.html>)
 - * 中央処理装置：CPU
 - * グラフィック処理装置：GPU
 - * 特定用途向け集積回路：ASIC
 - ・ 2022 年 4 億米ドル
 - ・ 2032 年 16 億米ドル

- － エッジ AI ソフトウェア市場 (<https://www.value-press.com/pressrelease/267210>)
 - * 2020 年 5 億 9000 万米ドル
 - * 2026 年 18 億 3500 万米ドル
 - * CAGR 20.8%
- エッジ AI プロセッサ市場の成長を制限する課題 (<https://prtimes.jp/main/html/rd/p/000002320.000072515.html>)
 - － 半導体不足
 - － AI テクノロジーの成長に伴う実装コストの増加
- エッジ AI ソリューションの成長を制限する課題 (<https://www.value-press.com/pressrelease/267210>)
 - － プライバシー
 - － セキュリティ
 - * セキュリティ侵害・攻撃・DDoS の増加により、当該分野ではエッジ AI の採用を制限
- エッジ AI の注目トレンド
 - － <https://blogs.nvidia.co.jp/2023/01/17/edge-ai-trends-2023/>
- ベンチマーク
 - － エッジデバイス向け AI アクセラレータの比較検証 (<https://lab.mo-t.com/blog/benchmark-edge-ai-accelerator>)
- AI 倫理の 7 原則 (Google, <https://ai.google/responsibility/principles/>)
 - － Be socially beneficial (社会への有益性)
 - － Avoid creating or reinforcing unfair bias (不公平なバイアスの防止)
 - － Be built and tested for safety (安全性確保のための開発)
 - － Be accountable to people (人々への説明責任)
 - － Incorporate privacy design principles (プライバシーの設計原則の組み込み)
 - － Uphold high standards of scientific excellence (科学手卓越性の高位平準化)
 - － Be made available for uses that accord with these principles (これらの原則に沿ったユーザへの技術提供)
- オリンピックで再燃した AI の倫理問題を考える。最新の取り組み事例も紹介 (https://aismiley.co.jp/ai_news/what-is-the-ethical-issue-of-ai-that-has-become-more-important-due-to-the-evolution-of-technology/)
- AI 倫理原則 (パナソニック, <https://monoist.itmedia.co.jp/mn/articles/2208/30/news073.html>)
 - － 「より良い暮らしとより良い社会」を実現すること
 - － 安全のための設計、開発、検証を行うこと
 - － 人権と公平性を尊重すること
 - － 透明性と説明責任を重視すること
 - － お客様のプライバシーを保護すること
- 責任ある AI：AI 倫理とガバナンス (accenture, <https://www.accenture.com/jp-ja/services/applied-intelligence/ai-ethics-governance>)

- 意図しないバイアスの最小化
 - AI の透明性を確保する
 - 従業員に機会を与える
 - データプライバシーとセキュリティを守る
 - 顧客と市場に利益をもたらす
- HPE における AI の倫理と原則 (HewlettPackard, <https://www.hpe.com/jp/ja/solutions/artificial-intelligence/ethics.html>)
 - AI のプライバシー対応のセキュリティ
 - AI の人間重視の原則
 - AI のインクルーシブ原則
 - AI の頑強な原則
 - 責任ある AI
- AI 倫理ガイドライン (AWL, <https://awl.co.jp/ai-ethics-guidelines/>)
 - 人間中心の AI 社会の実現
 - 法令遵守
 - AI の適正な利活用
 - 安全性とセキュリティの確保
 - プライバシーへの配慮
 - 公平性の尊重・非差別
 - 透明性等
 - AI の発展と人材育成
- OKI グループ AI 原則 (OKI, <https://www.oki.com/jp/technology/ai/principle.html>)
 - 人権の尊重
 - 説明と透明性
 - 対話と協調
 - 安全およびデータの取扱い
 - 人材育成
- 関連団体
 - tinyML (<https://www.tinyml.org/>)
 - Edge AI and Vision Alliance (<https://www.edge-ai-vision.com/>)
- 関連アプリケーション (<https://www.value-press.com/pressrelease/267210>, <https://www.mapion.co.jp/news/release/dn0000270836-all/>, <https://news.mynavi.jp/techplus/article/20211214-2227346/>)
 - リモートワーク
 - リモート監視
 - プラントオートメーション
 - 遠隔医療
 - スマートホーム
 - 監視カメラ
 - 予知保全

- 遠隔計測
 - 混雑状況監視（人，車）
- エッジ AI 活用事例：小売店 (https://standard-dx.com/post_blog/edge-ai)
 - 課題
 - * 購買時のデータ不足，欠品などによる販売機会の損失
 - * ニーズの多様化
 - * レジ打ち，売り場確認などによる人件費増加
 - ソリューション
 - * リテール AI カメラ
 - ・ 購入順序の把握
 - ・ 棚前行動の分析
 - ・ 滞在時間の把握
 - ・ 欠品状況監視
 - ・ 補充タイミングの通知
 - ・ 棚割最適化
 - * スマートショッピングカート
 - ・ レジレス・キャッシュレス
 - ・ レコメンド・レシピ提案
 - 導入効果
 - * スマートショッピングカート
 - ・ 有人レジ (125 秒)，無人レジ (75 秒) に対して，スマートショッピングカートで 32 秒までレジの所要時間を短縮
 - ・ 来店頻度が 13.8% 増加
 - * リテール AI カメラ
 - ・ 棚回転効率が 28.7% 向上
 - ・ 欠品率が 12.2% 改善
 - 参考
 - * http://www.trial-net.co.jp/cp/mediakit_ssc_aicamera/
- エッジ AI 活用事例：美容室・アパレル店舗 (https://standard-dx.com/post_blog/edge-ai)
 - 課題
 - * 化粧品の色味を探す工程にかかる時間と手間
 - * 肌へのダメージ
 - * マーケティング
 - 解決策
 - * 明暗やアンダートーンを含む約 90,000 種類の肌色データを学習した DeepLearning モデルで利用者の肌に最適な色味の検出とレコメンドやバーチャルメイクを行う
 - * オフラインはスマートミラーで，オンラインはスマートフォンやパソコンで利用（スマートシェードファインダー）
 - 導入効果
 - * トライされた商品の売上が 2.5 倍上昇

- * スマートシェードファインダー利用による顧客ロイヤリティアップ
- 参考
 - * <https://www.perfectcorp.com/business/successstory/detail?id=6>
- エッジ AI 活用事例：生産ライン (https://standard-dx.com/post_blog/edge-ai)
 - 課題
 - * 巻き線機でフィルムの特性変化や付け替えによる固有振動数の変化に伴う振動抑制制御補正にかかる時間が長い (PID 制御では振動抑制するために約 10 秒)
 - 解決策
 - * 固有振動の予測と自動補正 (DeepBinaryTree アルゴリズムを採用)
 - 効果
 - * 機器特性変化に対する補正を 3 秒まで短縮 (10 秒→3 秒への 7 秒の短縮)
 - * 約 20m 発生していた不良品が 1/3 以下に現象
 - 参考
 - * <https://aising.jp/algorithms/>
- エッジ AI 活用事例：生産ライン (https://www.toshiba-dme.co.jp/dme/dig/edge_solution/trend.htm)
 - 自動化、省人化
 - * 課題
 - ・ 検査精度がバラつく
 - ・ 不良品が多い
 - ・ 危険
 - * 効果
 - ・ 検査精度の均一化
 - ・ 歩留まり向上
 - ・ 安全性向上
 - ・ 人件費削減
 - 熟練スタッフの技術継承
 - * 課題
 - ・ 属人化している工程がある
 - ・ 熟練スタッフの退職によりノウハウが損失する
 - * 効果
 - ・ 判断が自動化
 - ・ 省人化
 - ・ 品質向上
 - 立ち入り禁止エリアの人検知
 - * 課題
 - ・ 危険を即時に検知できない
 - ・ 安全対策コストが高い
 - ・ 稼働率が低下する
 - * 効果
 - ・ 緊急時の即応

- ・稼働率と安全性の両立
 - ・エリア内の異常検知を自動化
- － 設備の予知保全
 - * 課題
 - ・設備が突然故障する
 - ・定期保全作業費用が高い
 - * 対策
 - ・稼働率を安定化できる
 - ・設備保全人件費の低減
- 関連企業 (<https://newscast.jp/news/4431496>, <https://www.mordorintelligence.com/ja/industry-reports/edge-ai-hardware-market>)
 - － Advanced Micro Devices, Inc., Xilinx, Inc.
 - * Kria KV260 ビジョン AI スターター キット (<https://japan.xilinx.com/products/som/kria/kv260-vision-starter-kit.html>)
 - * Vitis AI (<https://japan.xilinx.com/products/design-tools/vitis/vitis-ai.html>)
 - ・ AI 推論開発プラットフォーム
 - ・ PyTorch, TensorFlow, ONNX フォーマットの学習済みモデルを AI Model Zoo で提供
 - ・ AI モデル構造を圧縮する Vitis AI Optimizer
 - ・ AI モデルの重みを量子化する Vitis AI Quantizer
 - ・ AI モデルを高効率な命令セットとデータフローへのマッピングやレイヤー融合・スケジューリングなどの最適化を行う Vitis AI Compiler
 - ・ AI 推論実装における効率性と使用率を分析可能な Vitis AI Profiler
 - ・ 高効率な推論処理を行うための Vitis AI Library
 - * Vivado ML
 - ・ 機械学習向け FPGA 開発環境
 - *
 - － Alphabet Inc. (Google LLC)
 - * Edge TPU (<https://cloud.google.com/edge-tpu?hl=ja>)
 - － Intel Corporation
 - * FPGA (<https://www.intel.co.jp/content/www/jp/ja/products/docs/programmable/ai-fpga-whitepaper.html>)
 - * Movidius VPU (<https://www.intel.co.jp/content/www/jp/ja/products/details/processors/movidius-vpu.html>)
 - * OpenVINO (<https://www.intel.co.jp/content/www/jp/ja/internet-of-things/openvino-toolkit.html>)
 - － Qualcomm Technologies, Inc.
 - * Edge AI Box (<https://www.qualcomm.com/products/technology/artificial-intelligence/edge-ai-box>)
 - － Apple Inc.

- * Core ML (<https://developer.apple.com/jp/machine-learning/core-ml/>)
- Mythic
 - * Mythic Analog Matrix Processors (<https://mythic.ai/>)
- ARM Ltd.
 - * Cortex CPU (<https://www.arm.com/ja/product-filter?families=cortex-a&families=cortex-m&use-cases=machine%20learning>)
 - * Ethos NPU (<https://www.arm.com/ja/product-filter?families=ethos%20npus>)
 - * Mali GPU (<https://www.arm.com/ja/product-filter?families=mali%20image%20signal%20processors,mali%20graphics%20processors&use-cases=machine%20learning&showall=true>)
- SAMSUNG ELECTRONICS CO., LTD.
 - * あまり注力してなさそう (<https://semiconductor.samsung.com/jp/solutions/applications/ai/>)
- NVIDIA Corporation
 - * エッジでの AI 向けソリューション (<https://www.nvidia.com/ja-jp/deep-learning-ai/solutions/ai-at-the-edge/>)
 - * Jetson Nano (<https://www.nvidia.com/ja-jp/autonomous-machines/embedded-systems/jetson-nano/product-development/>)
 - * Jetson AGX Xavier (<https://www.nvidia.com/ja-jp/autonomous-machines/embedded-systems/jetson-agx-xavier/>)
 - * Jetson AGX Orin Industrial (<https://developer.nvidia.com/ja-jp/blog/step-into-the-future-of-industrial-grade-edge-ai-with-nvidia-jetson-agx-orin-industrial/>)
- HiSilicon Technology Co., Ltd
 - * HUAWEI Ascend 310 (<https://www.hisilicon.com/en/products/Ascend/Ascend-310>)
 - * HUAWEI Ascend 910 (<https://www.hisilicon.com/en/products/Ascend/Ascend-910>)
- Huawei Technologies Co., Ltd.
 - * Atlas 500 AI エッジステーション (<https://e.huawei.com/jp/products/cloud-computing-dc/atlas/atlas-500>)
- MediaTek Inc.
 - * MediaTek AI プロセッシングユニット (APU) (<https://www.mediatek.jp/innovations/artificial-intelligence>)
 - * 6th Generation MediaTek APU (AI for Smartphones) (<https://www.mediatek.com/technology/ai-for-smartphones-6th-gen>)
 - * 5th Generation MediaTek APU (AI for Smartphones) (<https://www.mediatek.com/technology/ai-for-smartphones-5th-gen>)
- Imagination Technologies Limited

- * PowerVR, Neural Network Accelerators (<https://www.imaginationtech.com/products/ai/>)
- IBM Corporation
 - * Edge AI SDK (https://researcher.watson.ibm.com/researcher/view_group.php?id=10944)
- Amazon Web Services, Inc.
 - * Amazon SageMaker Edge (<https://aws.amazon.com/jp/sagemaker/edge/?sagemaker-data-wrangler-whats-new.sort-by=item.additionalFields.postDateTime&sagemaker-data-wrangler-whats-new.sort-order=desc>)
 - * AI /機械学習に活用できる AWS のエッジソリューションのご紹介 (https://pages.awscloud.com/rs/112-TZM-766/images/EV_awsloft-tko-iotloft14-1t4_Sep-2020.pdf)
- Synaptics Incorporated
 - * Edge Computing SoCs with AI (<https://www.synaptics.com/technology/edge-computing>)
 - * Synaptics Edge AI (<https://investor.synaptics.com/static-files/46bac204-81d8-45e1-a6b1-0e437c795bf9>)
- imagimob
 - * Development Platform for Machine Learning on Edge devices (<https://www.imagimob.com/>)
- TIBCO Software Inc.
 - * TIBCO Spotfire (<https://www.tibco.com/products/tibco-spotfire>)
 - * Make AI Real: Operationalize Data Science (<https://www.tibco.com/products/data-science>)
 - * Bringing Intelligence to Edge Computing Through Machine Learning (<https://www.tibco.com/resources/whitepaper/bringing-intelligence-edge-computing-through-machine-learning>)
- Octonion
 - * つぶれた? Web サイトにアクセスできない (<https://prtimes.jp/main/html/rd/p/000001113.000001337.html>)
- FogHorn System
 - * Johnson Controls に買収されたらしいが, その後どうなったかが分からない (<https://prtimes.jp/main/html/rd/p/000000127.000007712.html>)
- Gorilla Technology Group
 - * エッジ AI テクノロジー (<https://jp.gorilla-technology.com/Products/Technology/Edge-AI>)
- STMicroelectronics
 - * Edge AI - Artificial Intelligence at the Edge (https://www.st.com/content/st_com/ja/campaigns/artificial-intelligence-at-the-edge.html)
 - * X-CUBE-AI (<https://www.st.com/ja/embedded-software/x-cube-ai.html>)

- * X-LINUX-AI (<https://www.st.com/ja/embedded-software/x-linux-ai.html>)
- * FP-AI-SENSING1 (<https://www.st.com/ja/embedded-software/fp-ai-sensing1.html>)
- * FP-AI-VISION1 (<https://www.st.com/ja/embedded-software/fp-ai-vision1.html>)
- * FP-AI-NANOEDG1 (<https://www.st.com/ja/embedded-software/fp-ai-nanoedg1.html>)
- Renesas Electronics Corporation.
 - * 人工知能 (AI) (<https://www.renesas.com/jp/ja/application/key-technology/artificial-intelligence>)
- Texas Instruments Incorporated.
 - * エッジインテリジェンスの推進 (<https://www.ti.com/ja-jp/technologies/edge-ai.html>)
 - * AM68A (<https://www.ti.com/product/ja-jp/AM68A>)
 - * AM69A (<https://www.ti.com/product/ja-jp/AM69A>)
 - * AM62A3 (<https://www.ti.com/product/ja-jp/AM62A3>)
 - * AM62A7 (<https://www.ti.com/product/ja-jp/AM62A7>)
 - * EDGE-AI-STUDIO ([urlhttps://www.ti.com/tool/ja-jp/EDGE-AI-STUDIO](https://www.ti.com/tool/ja-jp/EDGE-AI-STUDIO))
- Idein Inc.
 - * 国内シェア No.1 エッジ AI プラットフォーム Actcast (<https://www.idein.jp/ja>)
- LeapMind Inc.
 - * あらゆるエッジデバイスをスマート化する超低消費電力 AI アクセラレーター EFFICIERA (<https://leapmind.io/>)
- Hitachi, Ltd.
 - * ビッグデータ × AI (https://www.hitachi.co.jp/products/it/bigdata/bigdata_ai/index.html)
 - * 業務改革や計画最適化などクラウドベースが中心とみられる
- TOSHIBA CORPORATION
 - * 東芝 AI 技術カタログ (<https://www.global.toshiba/jp/technology/corporate/ai.html>)
- Panasonic Holdings Corporation
 - * Panasonic × AI (<https://tech-ai.panasonic.com/jp/>)
- Fujitsu
 - * 応用領域: 画像, 音声, 自然言語, 数値解析 (<https://www2.fujitsu.com/jp/key-technologies/ai/>)
- Mitsubishi Electric Corporation
 - * AI (人工知能) とは (<https://www.mitsubishielectric.co.jp/it/it-words/keyword-ai/?split-page=1>)
 - * Maisart を構成する技術: DeepLearning, 強化学習, ビッグデータ分析, 知識

処理, メタヒューリスティック (<https://www.mitsubishielectric.co.jp/corporate/randd/maisart/index.html>)