ソフトウェアとハードウェアを融合し現実世界を計算可能にする

～AI技術の産業領域への取り組み～

高橋　正和
株式会社Preferred Networks
執行役員 最高セキュリティ責任者

Preferred Networks

Preferred Networks, Inc. (PFN)

- 設　立：2014年3月
- 所在地：東京都千代田区大手町（日本）、カリフォルニア州バークレー（米国）
- 取締役：西川 徹、岡野原 大輔、奥田 遼介
- 従業員数：約250名（2019年4月現在）
- ミッション：

  ソフトウェアとハードウェアを融合させ、
  現実世界を計算可能にする。

- 事業内容:
  交通システム、産業用ロボット、バイオヘルスケア、
  パーソナルロボット、スポーツ解析、クリエイティブ
  などの分野に深層学習を応用

Preferred Networks

Our Strategic Partners

and Collaborators

AGENDA

- 今日のDeep Learning応用領域
  - なぜDeep Learningが注目されるのか？
- Deep Learningが必要とする計算資源
  - Deep Learningに求められる膨大な計算資源と技術基盤
- AI応用領域に求められるセキュリティ
  - Trusted AI lifecycle への取り組み
- むすび

今日のDeep Learning応用領域

Preferred Networks


全自動お片付けロボット
Autonomous Tidying-up Robot
Preferred Networks

ロボットを題材に考察する

なぜDeep Learningが注目されるのか？

Preferred Networks

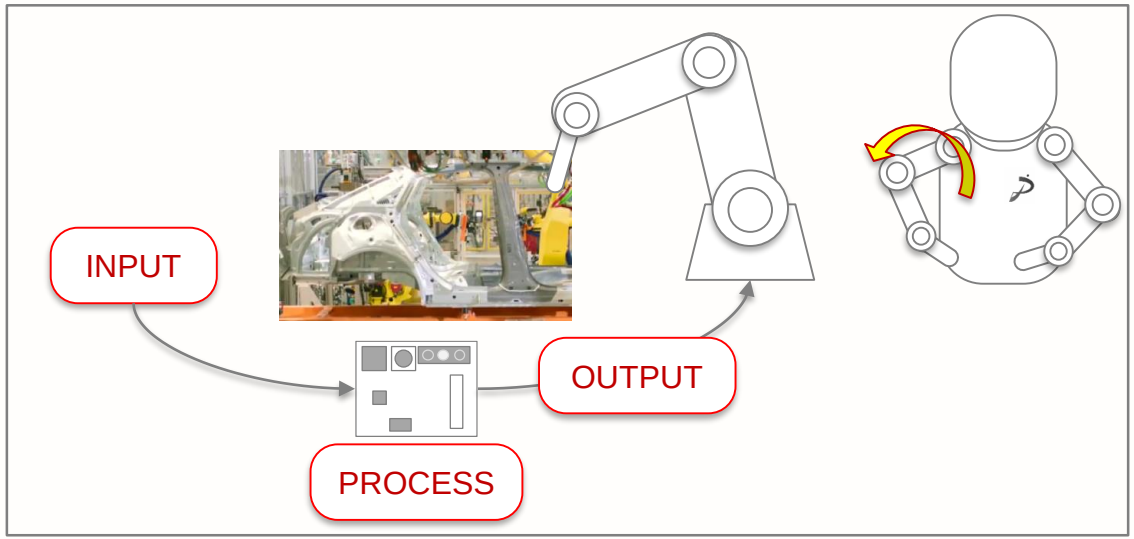
産業用ロボット

高度な制御が行われている。しかし、対象が常に決まった位置にある
ために実現できている**目の見えないロボット**

FANUC ipari robotok az AUDI Hungariánál
https://www.youtube.com/watch?v=NtQjQkDli-g

Preferred Networks

一般的な産業用ロボット



INPUT	PROCESS	OUTPUT
• 信号 （セット完了）	• プログラムに従った動作 （ティーチング）	• 動作の実行 （モーターの制御）

センサーにより、所定の位置にセットされたことが検出されると、あらかじめ決められたプログラム通りに動作を行う。

- 常に同じ動作
- 作業結果のフィードバックは生じない

つまり、**視覚情報なしに動作**している。
アルゴリズムで記述が可能

Preferred Networks

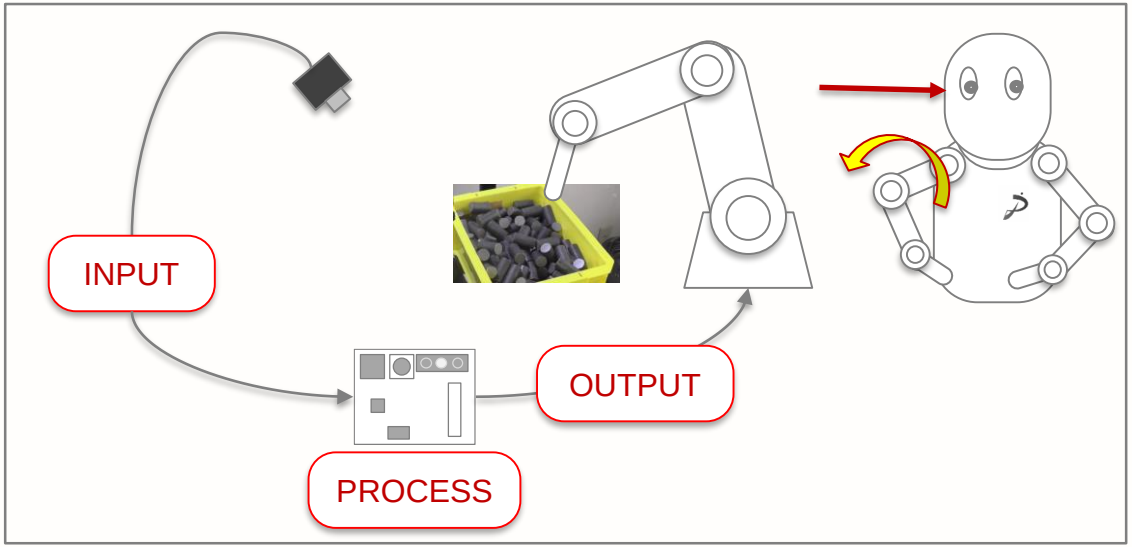
産業用ロボット（ばら積みロボット）

ばら積み(Random Pickup)では、事前に対象の位置を正確にプログラム（ティーチング）できないため、
画像などに基づいたフィードバックと、動作計画をリアルタイムで実施する必要がある。
つまり、**目が見える必要がある**

FANUC【バラ積みロボットの活用例】3Dビジョンセンサを用いた前段取りシステム
https://youtu.be/qHeDW7lCC5A

Preferred Networks

バラ積みロボット



INPUT
- 画像

PROCESS-1
- 対象の特定
- 位置の算出

PROCESS-3
- 成功・失敗の判断

PROCESS-2
- ピックアップの計画
- 動作のプログラム

OUTPUT
- 動作の実行
（モーターの制御）

画像から、自動的に対象群を認識し、そのなかからピックアップする対象を選択する。
対象をピックアップするためのシーケンスを生成し、これを実行する。
トレーニング時には、画像から、動作が成功したかどうかを判断する。

- 変化する状態に合わせた、３次元情報を算出
- 対象をピックアップするためのシーケンスを生成
- 訓練時には、成功・失敗を認識し、精度を高めていく

つまり、**視覚情報に基づいて作業を行うが、画像認識するアルゴリズムが極めて難しい。**
（形、光、方向、などで画像が変わってしまう）

Preferred Networks

画像と音声を計算可能にする取り組み

Preferred Networks

コンピューターの二つの問題

チェッカー	10の30乗
オセロ	10の60乗
九路盤囲碁	10の90乗
チェス	10の120乗
将棋	10の220乗
囲碁	10の360乗

- 現実的な時間で解けない問題がある
  - 計算能力を増大させ続ける一つの要因
    - ✓ ボードゲームは、可能な局面をすべて探索すれば必ず勝てるが、まともにすべてを計算するとチェッカーで 3億年かかる計算 (10の14乗/秒)
  - 必ずしもDeep Learningで取り組んではいないが、計算量を減らすために利用されている
- その問題を解く明確な手順が必要
  - 演繹的、アルゴリズムがかけないと解けない
    - ✓ アルゴリズムから外れたものは、解けないか、誤った解を導く（汎化性能）
    - ✓ 画像解析、音声認識、翻訳など
  - Deep Learningによる取り組みが行われている領域

ボードゲームの思考アルゴリズムの入門
http://alcuin.hatenablog.com/entry/2013/12/30/154841

Deep Learning とは？：摂氏から華氏への変換

普通の関数

仕様

入力: C
出力: F
ただし、FはCを華氏で表したもの

人が持つ
先験的知識

モデル

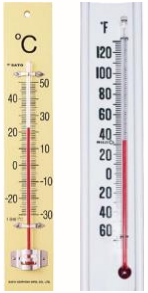
F = 1.8 * C + 32

アルゴリズム

```
double c2f(double c) {
  return 1.8*c + 32.0;
}
```

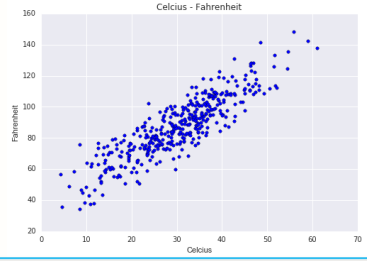
**モデルが既知
アルゴリズムが
構成可能である必要**

Deep Learning



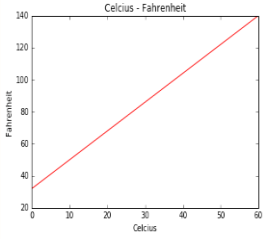
観測

訓練データセット



X → Y = f(X,θ) → Y

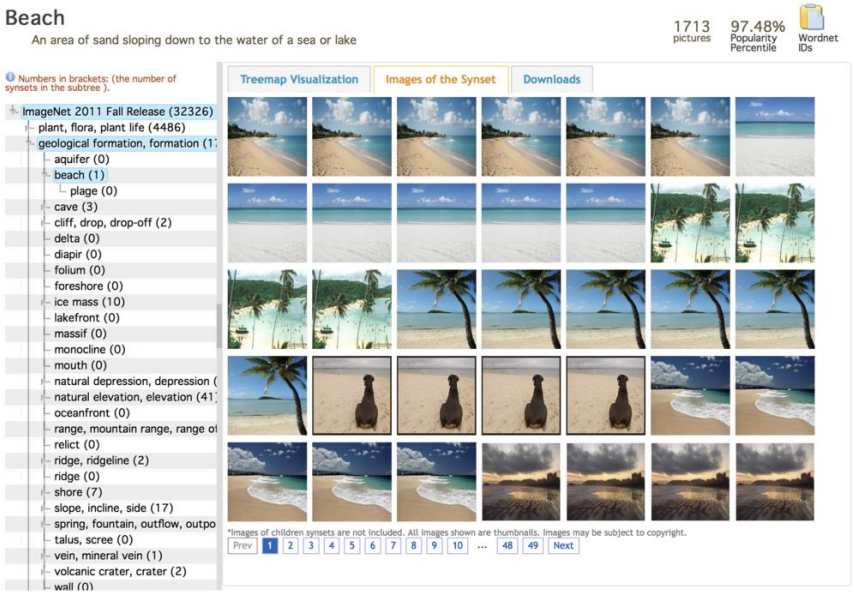
モデル： 線形回帰式

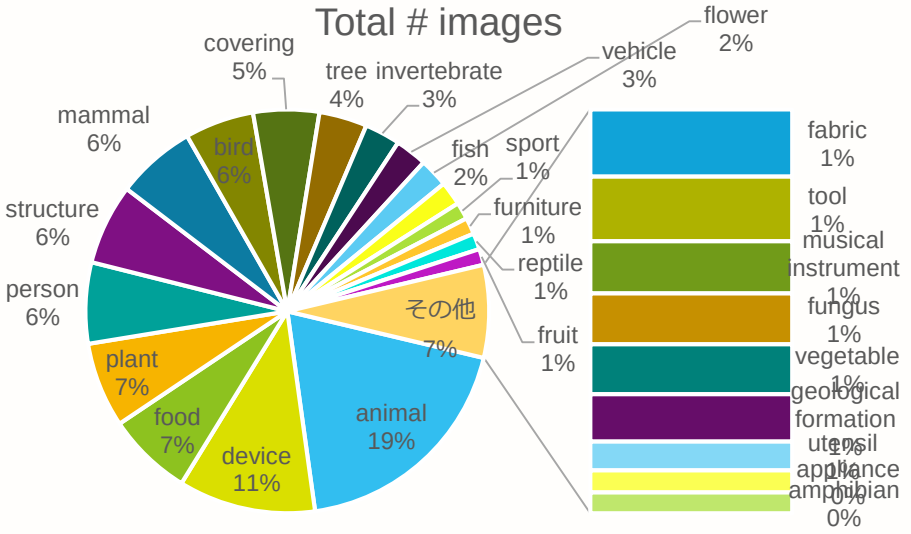


**モデル・アルゴリズムが
未知でよい**

Preferred Networks

Deep Learningへの注目：画像認識に関するコンテスト - IMAGENET





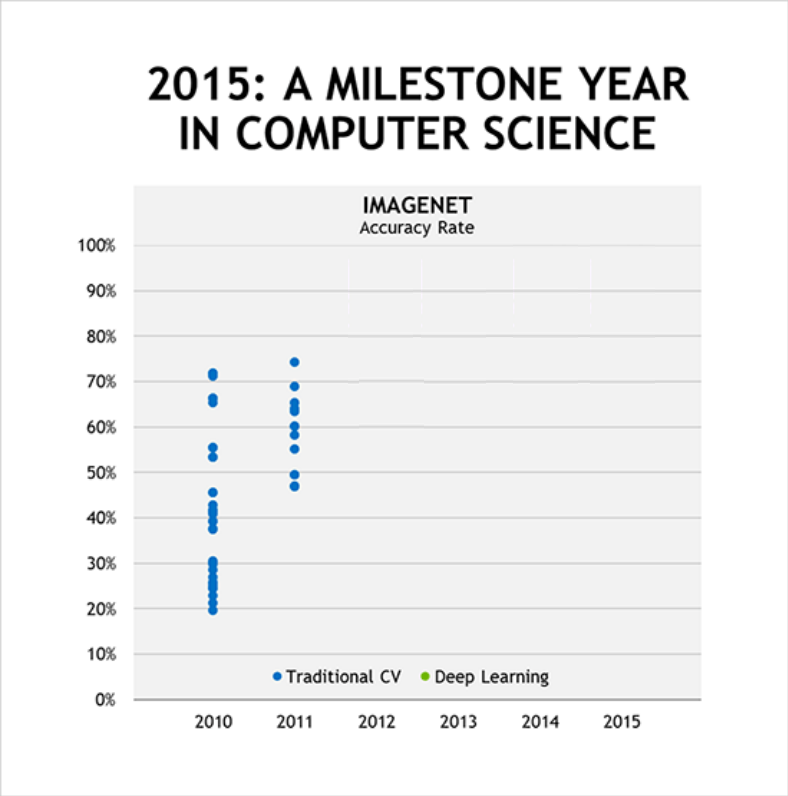
Total number of non-empty synsets	21,841
Total number of images	14,197,122
Number of images with bounding box annotations:	1,034,908
Number of synsets with SIFT features	1,000
Number of images with SIFT features	1.2 million

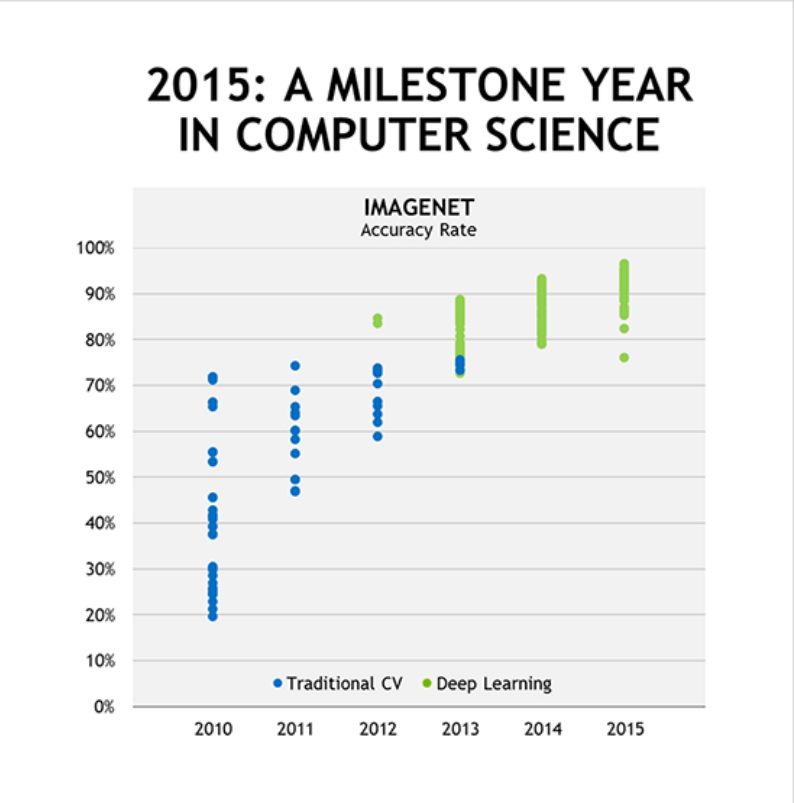
http://looseleaf0727.hatenablog.jp/entry/2017/12/17/233709

http://www.image-net.org/about-stats

Preferred Networks

Deep Learningによる画像認識のブレークスルー（IMAGENET)





https://blogs.nvidia.co.jp/2016/01/12/accelerating-ai-artificial-intelligence-gpus/

Preferred Networks

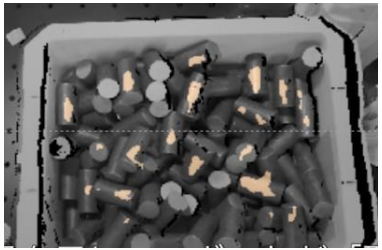
ロボットにおける
画像と音声・機能的プログラミングの利用

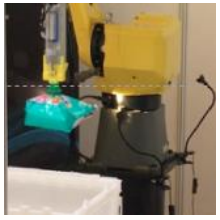
Preferred Networks

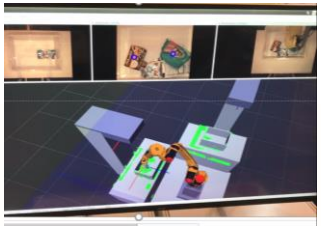
Deep Learningで帰納的に問題を解く

ティーチングなしでバラ積みされた部品を取り出す。
8時間の学習で９割の取得率（熟練者のチューニングに匹敵）

トレーニングをしていない、多様な形状のものを取り出す
液体の入った袋のような、移動に伴って重心が変わるモノや、
初めてのモノでも、取り出すことができる（高い汎化性能）







Preferred Networks

Deep Learningを使って物体を認識する

実世界における
インタラクティブな物体指示

羽鳥潤*, 菊池悠太*, 小林颯介*, 高橋城志*,
坪井祐太*, 海野裕也*, Wilson Ko, Jethro Tan
* Starred authors equally contributed

Preferred Networks

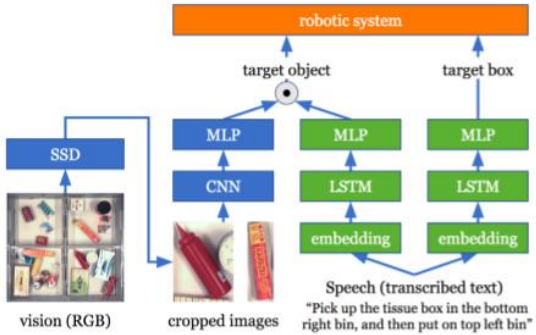
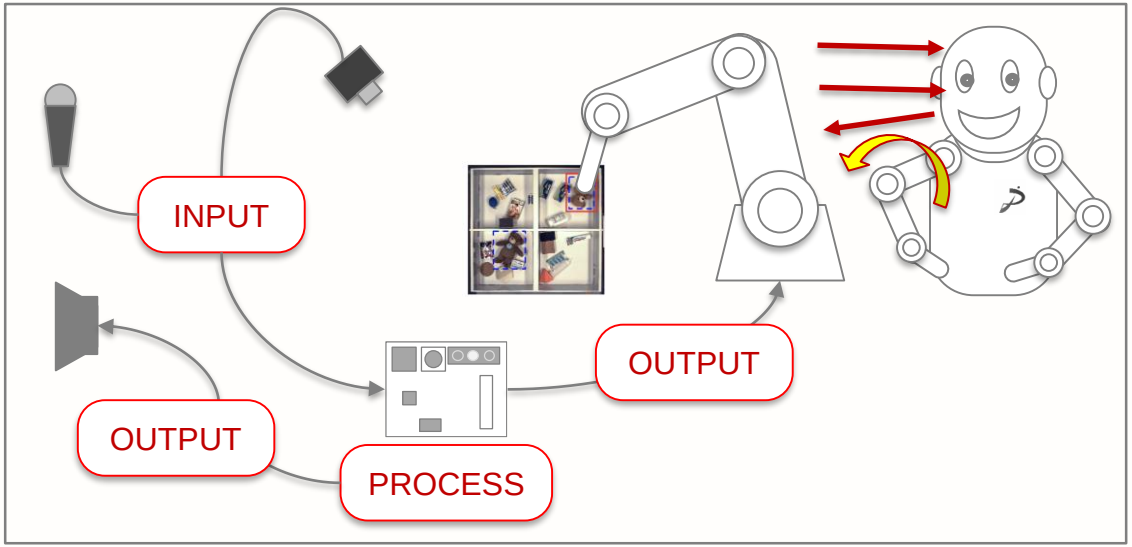


図3: 全体のネットワーク構成．入力画像は SSD によって検出した物体ごとに切り抜き，CNN によって特徴情報を抽出し，言語指示は LSTM によって同様に特徴情報を抽出する．両者を使って対象物と移動先を推定する．

https://anlp.jp/proceedings/annual_meeting/2018/pdf_dir/C5-1.pdf

そこにある“何か”ではなく、それが何かを認識する

インタラクティブな物体指示

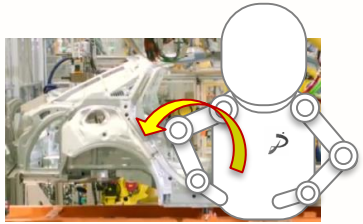


画像から、自動的に対象群を認識し、自然言語による曖昧な指示に従って移動する対象、移動先を認識し移動する。対象が特定できない場合は、問い合わせを行うことで、対象を特定する。
対象群が、形状だけではなく、名称や特徴として認識されている。

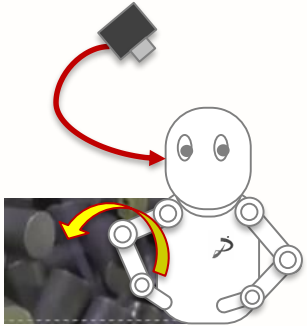
INPUT	PROCESS-1	PROCESS-2	OUTPUT
• 画像	• 物体検出 • 物体選択 • 移動先予測 • 指示の曖昧性判定と解消	• ピックアップの計画 • 動作のプログラム	• 動作の実行（モーターの制御） • 応答 • 指示の確認

Preferred Networks

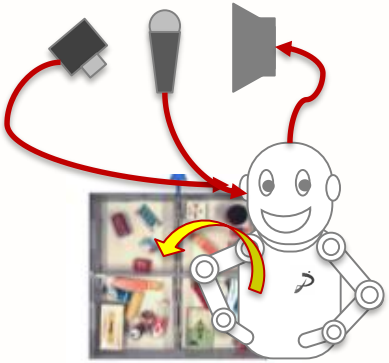
ロボットに見る
識別・認識・認知



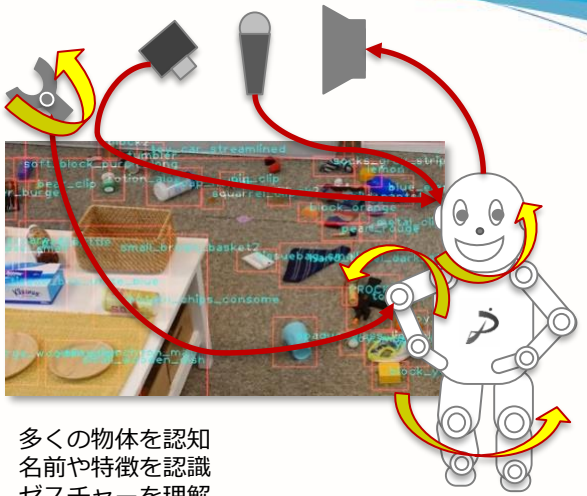
- 視覚を持たない
- 聴覚を持たない
- 発話をしない
- 計画をしない
- 移動しない（据え付け）



- 特定物体の形状を識別
- 名前や特徴は持たない
- 聴覚を持たない
- 発話をしない
- ピックアップを計画
- 移動しない（据え付け）



- 多くの物体を識別
- 名前や特徴を認識
- 音声の指示を理解
- 指示の不明瞭さを判断
- 指示の認識を応答
- 指示の不明瞭さの解決を促す
- 指示に基づいて、ピックアップと移動を計画
- 移動しない（据え付け）



- 多くの物体を認知
- 名前や特徴を認識
- ゼスチャーを理解
- 音声の指示を理解
- 指示の不明瞭さを判断
- ゼスチャーとの組み合わせを理解
- 指示の認識を応答
- 指示の不明瞭さの解決を促す
- お片付けを計画
- 物体をピックアップし指定の場所に移動
- 移動し、自身の位置を把握する
- 物体の位置を記憶し相対的な位置を理解する
- 握る・圧力を感じる

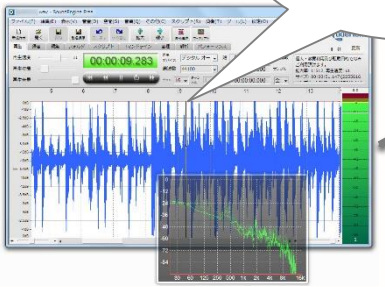
Preferred Networks

音声における計算レベル・認識レベル

Preferred Networks

Preferred Networks

音声の例



音データ

音声データ

再生

計算できない

背景音と分離
抽出データ

テキストデータ

ディクテーション
（聞き取り）

“ケンダマハドコ”

単語・文節の認識
“ケンダマ・ハ・ドコ”

意味の認識
“ケンダマ（名詞）・ハ（接続詞）・ドコ（副助詞）”にあるかを尋ねている

計算できる

文脈の認識
“ケンダマ”を認識（認識できなければ質問）
“ケンダマ”は部屋にひとつ（複数の場合は質問）
記録されている場所は（X,Y）
周辺で目印になる物体の選出（”Z1”,Z2,Z3）

単純な会話・指示
- 検索エンジンの入力
- カーナビの指示、音声案内

単純な会話・指示
- 自動翻訳、書き起こし
- チャットボット
- スマートスピーカー

単純な会話・指示
- お片付けロボット
- 高度な対話型のシステム

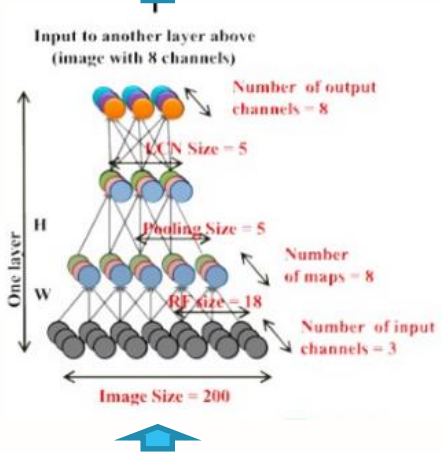
Preferred Networks

Deep Learningが必要とする膨大な計算資源

Deep Learning が必要とする計算リソース

汎用計算機構としての深層学習

出力（超多次元）



入力（超多次元）

桁違いに多いパラメタ

任意の多次元非線形関数を近似

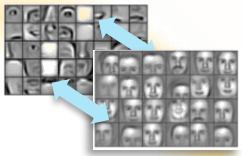
➔ **疑似的にチューリング完全！**

代償： 暴力的な計算リソース

学習を1日で終わらせるのに必要な計算リソース

P:Peta
E:Exa
F:Flops

画像/映像認識

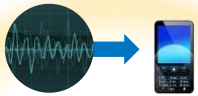


10P（画像）～ 10E（映像） Flops
学習データ：1億枚の画像 10000クラス分類
数千ノードで6ヶ月 [Google 2015]

バイオ・ヘルスケア

100P ～ 1E Flops
一人あたりゲノム解析で約10M個のSNPs
100万人で100PFlops、1億人で1EFlops

音声認識



10P～ Flops
1万人の5000時間分の音声データ
人工的に生成された10万時間の
音声データを基に学習 [Baidu 2015]

自動運転

1E～100E Flops
自動運転車 1 台あたり1日 1TB
10台～1000台, 100日分の走行
データの学習

ロボット/ドローン

1E～100E Flops
1台あたり年間1TB
100万台～1億台から得られた
データで学習する場合

機械学習、深層学習は学習データが大きいほど高精度になる
現在は人が生み出したデータが対象だが、今後は機械が生み出すデータが対象となる

各種推定値は1GBの学習データに対して1日で学習するためには
1TFlops必要だとして計算

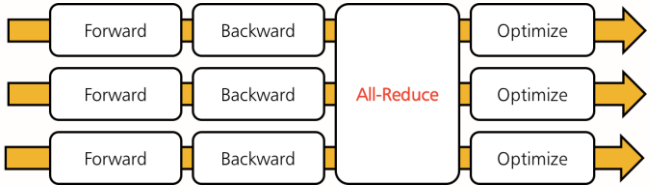
10PF　100PF　1EF　10EF　100EF

Preferred Networks

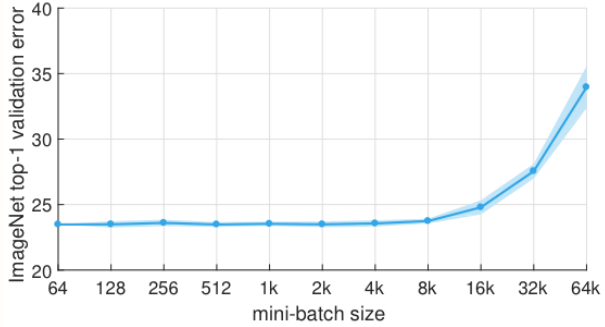
高速化・大規模化：分散学習の大規模を支える技術

- GPUとバッチサイズを増やせば高速化 ←→ モデル同期頻度が減るので精度悪化
- パラメータ調整、NVIDIAライブラリの有効活用、Infiniband等の高速通信…

GPUとデータセットを並列にしAll-Reduceで勾配情報の共有



各GPUの毎回の処理数＝バッチサイズが大きすぎると破綻



Accurate, Large Minibatch SGD: Training ImageNet in 1 Hour [Goyal+, 2017]

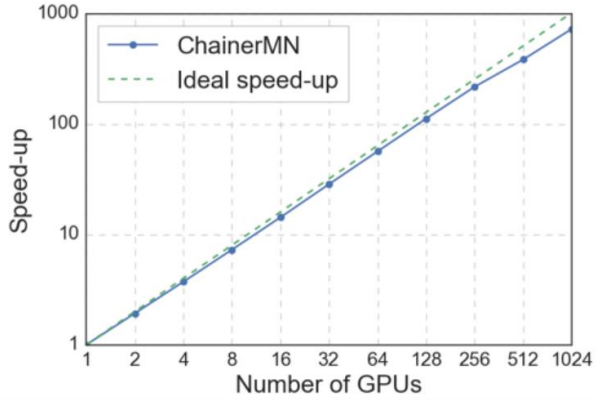
Chainer MN

Preferred Networks

分散学習

MN

ChainerMN: 分散深層学習用追加パッケージ
高いスケーラビリティ (1024GPUで1000倍の高速化)



MN-1: The GPU cluster behind 15-min ImageNet
https://preferredresearch.jp/2017/11/30/mn-1-a-gpu-cluster-behind-15-min-imagenet/

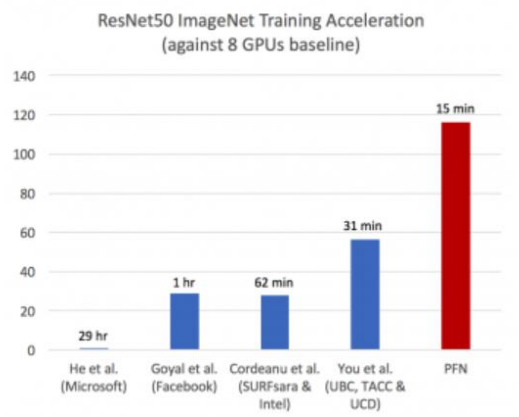
Preferred Networks

深層学習の学習速度において世界最速を実現

大規模な並列コンピュータを活用し、分散学習パッケージ ChainerMNでImageNetの学習を15分で完了

株式会社Preferred Networks（本社：東京都千代田区、代表取締役社長：西川徹、プリファードネットワークス、以下、PFN）は、大規模な並列コンピュータ「MN-1[※1]」を活用し、深層学習（ディープラーニング）の学習速度において世界最速を実現しました。

深層学習モデルの精度を向上させるため、学習データのサイズやモデルのパラメータ数が増加し、それにともなって計算時間も増大しています。1回の学習に数週間かかることも稀ではありません。複数のコンピュータを連携させて学習を高速化することは、新たなアイディアの試行錯誤や検証に要する時間を圧縮し、素早く研究成果をあげていくために非常に重要です。



https://www.preferred-networks.jp/ja/news/pr20171110

Preferred Networksのプライベート・スーパーコンピュータが Top 500リストのIndustry領域で国内１位に認定

2017年11月14日　タグ: MN-01, supercomputer　By preferred

株式会社Preferred Networks (本社: 東京都千代田区、代表取締役社長 最高経営責任者: 西川 徹、以下PFN)が占有利用するプライベート・スーパーコンピュータ「MN-1」が、LINPACK [※1] 性能測定の結果、約1.39ペタFLOPS [※2] を記録しました。これにより、2017年11月のスーパーコンピュータ性能ランキングを示すTOP500リスト (http://www.top500.org) において、産業領域 (Industry Segment) のスーパーコンピュータにおける世界12位、国内１位として登録されました。研究用等の全てのスーパーコンピュータを含むランキングにおいては、世界91位、国内13位となります。

PFNのプライベート・スーパーコンピュータMN-1の概要 [※3]

MN-1は、NTTコミュニケーションズ株式会社 (本社：東京都千代田区、代表取締役社長：庄司 哲也、以下NTT Com) と株式会社NTTPCコミュニケーションズ (本社: 東京都港区、代表取締役社長: 田中 基夫、以下NTTPC) の高速演算処理 (GPU) プラットフォームを採用し、計算ノードにはNVIDIA(R) Tesla(R) P100GPUを1,024基搭載しています。

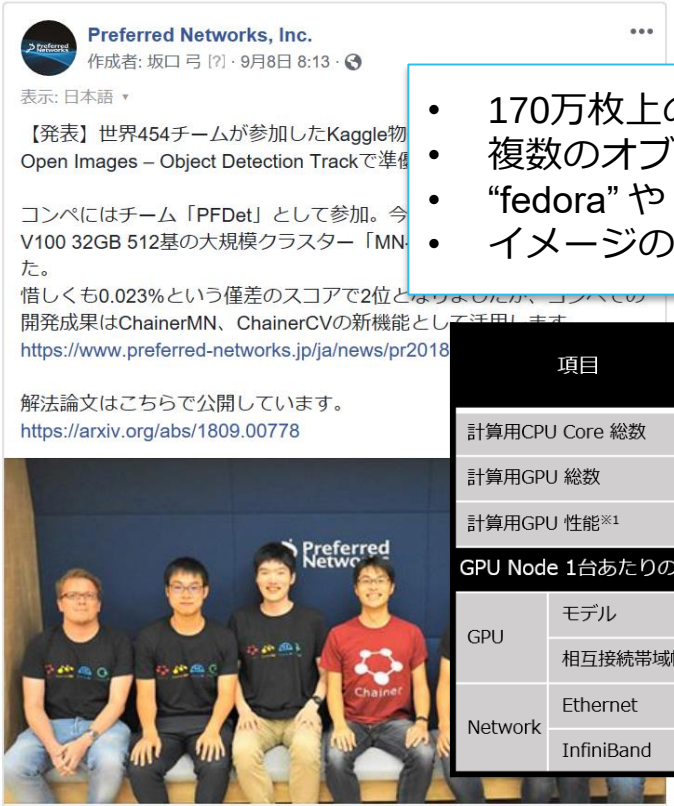
MN-1は、Mellanox社製 Infiniband インターコネクトを活用することで、PFNが開発する分散深層学習パッケージChainerMN （チェイナー・エムエヌ）による高速な分散深層学習 [※4] が可能です。

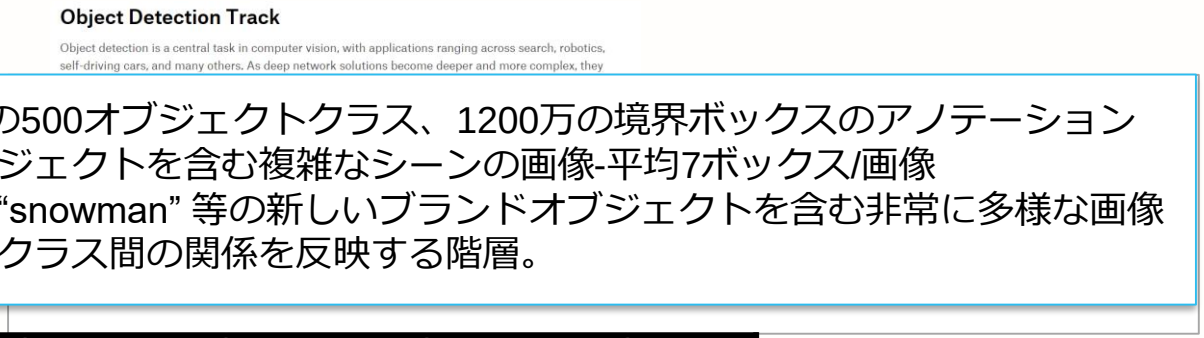
PFNはMN-1を使って、特に多くの計算機資源を必要とする交通システム、製造業、バイオ・ヘルスケア分野をはじめとしたさまざまな分野での深層学習の研究開発を、より一層加速させます。

https://www.preferred-networks.jp/ja/news/pr20171114

Preferred Networks

Google AI Open Images – Object Detection Track





- 170万枚上の500オブジェクトクラス、1200万の境界ボックスのアノテーション
- 複数のオブジェクトを含む複雑なシーンの画像-平均7ボックス/画像
- “fedora” や “snowman” 等の新しいブランドオブジェクトを含む非常に多様な画像
- イメージのクラス間の関係を反映する階層。

項目	MN-1	MN-1b	**MN-2**	**MN-1bとMN-2の性能比**
計算用CPU Core 総数	2,048	2,304	**5,760**	150 %UP
計算用GPU 総数	1,024	512	**1,024**	100 %UP
計算用GPU 性能※1	19.1 ペタフロップス	57.3 ペタフロップス	**128 ペタフロップス**	123 %UP
GPU Node 1台あたりの主な性能				
GPU	モデル	NVIDIA P100 PCI-e	NVIDIA V100 PCI-e	**NVIDIA V100 SXM2**	-
	相互接続帯域幅	32 GB/s	32 GB/s	**300 GB/s**	837 %UP
Network	Ethernet	1 GbE	10 GbE	**100 GbE x 4**	-
	InfiniBand	FDRx2 (112Gbps)	EDRx2 (200Gbps)	**なし**	

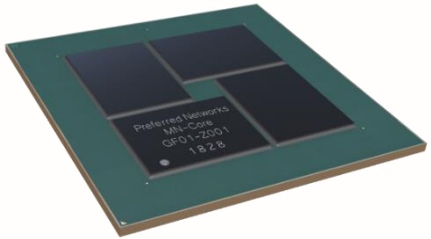
MN-2 (2019/03)

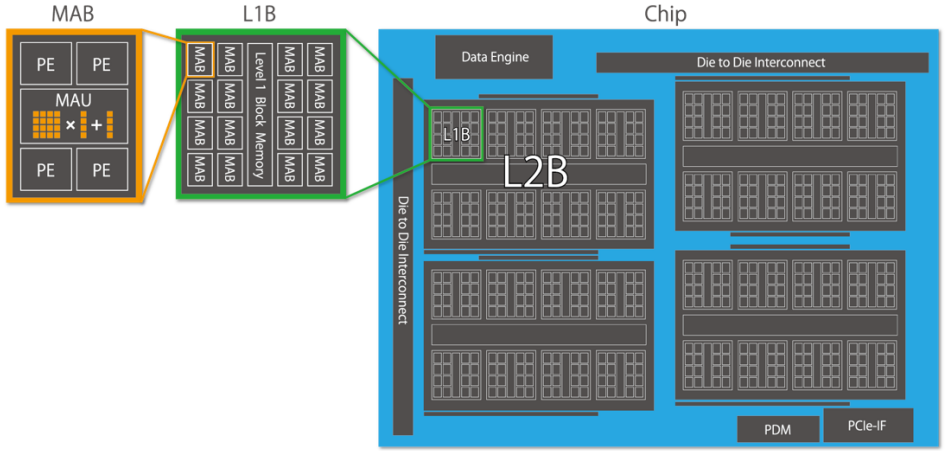
NVIDIA Tesla V100 x 512

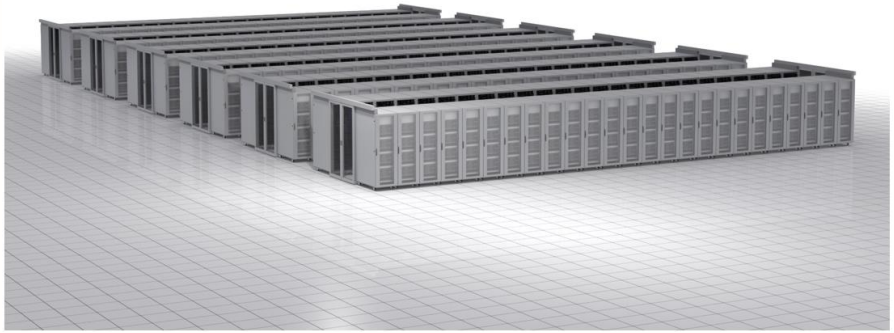
https://www.kaggle.com/c/google-ai-open-images-object-detection-track

Preferred Networks

MN-Core







- **MN-Coreチップのスペック**
- 製造プロセス ： TSMC 12nm
- 消費電力 （W、予測値） ： 500
- ピーク性能 （TFLOPS） ：32.8（倍精度） / 131（単精度） / 524 （半精度）
- 電力性能（TFLOPS / W、予測値） ： 0.066 （倍精度）/ 0.26（単精度） / 1.0（半精度）

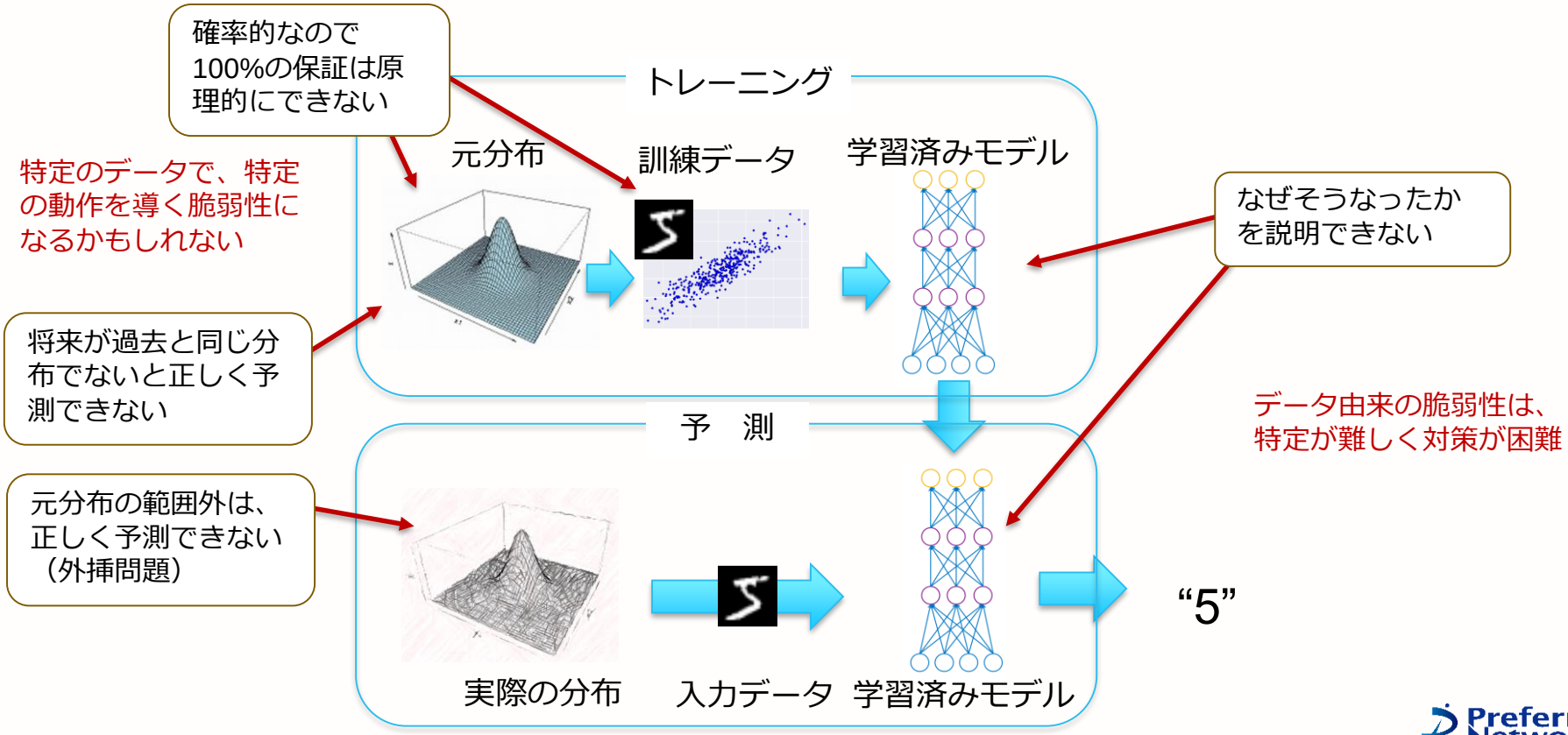
Preferred Networks

AI応用領域に求められる
セキュリティ

Preferred Networks

統計的機械学習の本質的限界

統計的機械学習は、トレーニングと予測は独立したフェーズとして構成される。
（トレーニングしながら予測を行うわけではない）



Preferred Networks

AI応用領域に求められるセキュリティ

IT環境のセキュリティ

AI応用製品のセキュリティ

Trusted AI Lifecycle
- (Traditional) SDLの展開
- Secure DL Engineering等の適用
- データ要件フレームワーク
- データ検証フレームワーク
- トレーニングフレームワーク
- 検査手法フレームワーク
- Trustレスポンスフレームワーク
- DL Software StackのSDL
- 教育とトレーニング

Threat Modeling for Trusted AI
- STRIDE for Trusted AI
- Factors Influencing Uncertaintyをベースにした、脅威モデルの構築
- Safety, Privacyの組み込み

CERT/CSIRT for AI
- 脆弱性全般のハンドリング
- CERT/CSIRTとの連携
  - 主にシステム領域
- DL Software Stack
  - 脆弱性等ハンドリング
  - ベストプラクティス
  - Incident/Accident事例
- データレピュテーション
  - トレーニングデータ
  - トレーニング済モデル

Data Security
流通するデータのセキュリティ

Security Product / Solution with AI

- AIを使った セキュリティ製品 ソリューション

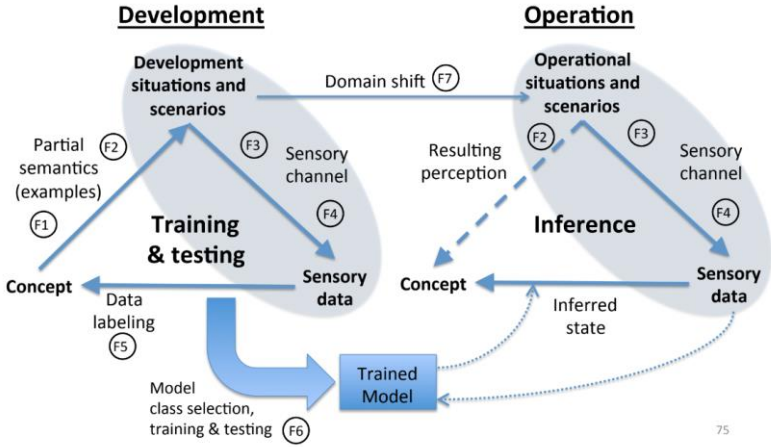
Preferred Networks

脅威分析に不確定性を取り入れてみてはどうか？

STRIDE

Spoofing なりすまし	Tampering 対破壊性	Repudiation 否認
Information disclosure 情報露呈	Denial of service サービス不能	Elevation of privilege 権限昇格

FIU: Factors Influencing Uncertainty



Safety Assurance of Machine Learning Based Func7ons in Automated Driving Systems (ADS)
Krzysztof Czarnecki, Waterloo Intelligent Systems Engineering Lab

Preferred Networks

不確かさに影響を与える要因（画像認識）

- **F1:コンセプトの不確かさ**
  - 歩行者とサイクリストの区別
  - 標準的なオントロジーの確立が必要
- **F2:開発シナリオのカバレッジ**
  - 様々な格好をした人々
  - どれだけのデータが必要か：正解/不正解、ニアヒット/ニアミス
- **F3:シーンの不確かさ**
  - 範囲、スケール、オキュレーションレベル、可視性、照明、集合・混雑レベル
  - テストセットの正確性と出力の確かさをこれらの指標で評価する必要がある
- **F4:センサーのプロパティ**
  - 成熟した工学的ディシプリン:
    モード、範囲、解像度、感度、配置、等
  - ML アルゴリズムとセンサ特性の相互作用
- **F5:ラベルの不確かさ**
  - 専門家のレビューとラベラーの不一致
- **F6: モデルの不確かさ**
  - 説明手法、堅牢性評価、ベイズによる検証
- **F7: 運用ドメインの不確かさ**
  - 新しい歩行者のポーズ、車の形状の新しいタイプ、カメラミスキャリブレーション
  - 稼働時の新規状況の評価、インパクトレベル、センサーパラメータの評価

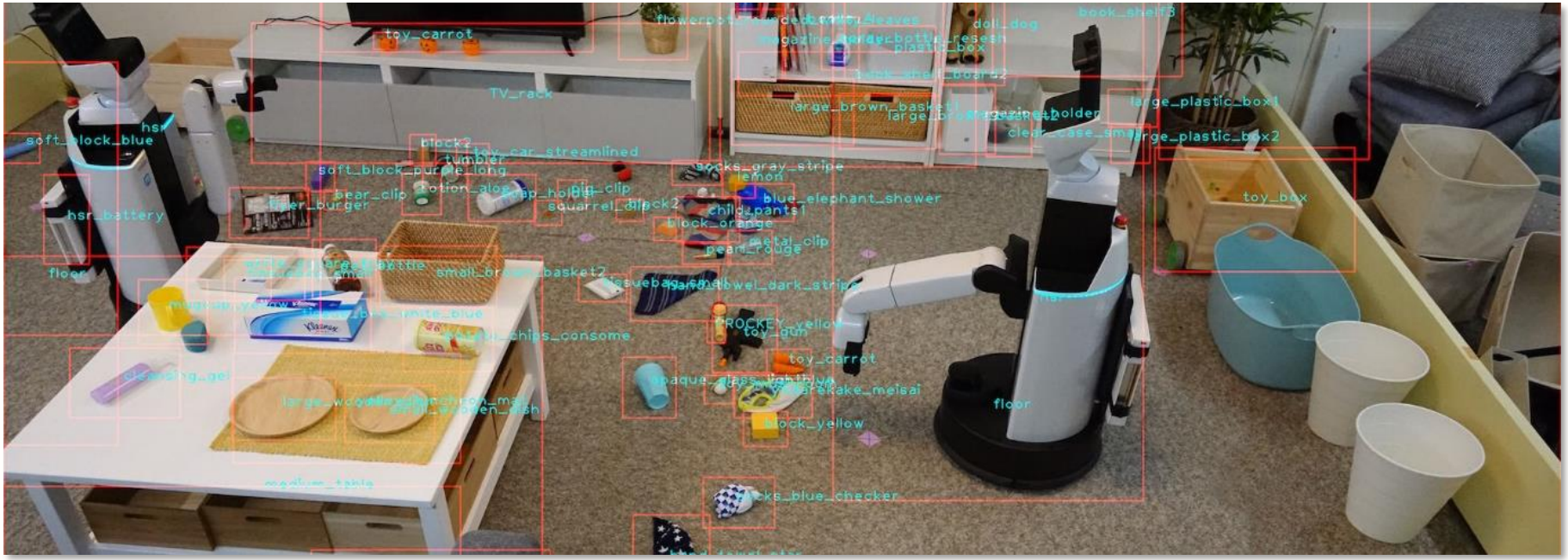
Preferred Networks

むすび

AI応用技術とセキュリティ

- AI応用技術技術は確実に利用が広がっている
  - これまでのロジックでは現実的な解がなかった領域に、現実的な解をもたらしている
  - 「現実世界を計算可能にする」ための取り組みが進んでいる
- AI応用技術は膨大な計算リソースを必要としている
  - 総合的な計算資源の強化が求められている
- AI応用技術のセキュリティを体系化することが求められている
  - ITレベルのセキュリティ
  - 製品レベルのセキュリティ
  - データレベルのセキュリティ

ソフトウェアとハードウェアを融合させ、現実世界を計算可能にする



Preferred Networks

Chainer

Preferred Networks

Chainer

Retail

Using big data from diverse devices, we recognize customers' interests and attributes through analysis of their in-store behavior.

Advertising

Our technology provides an advanced O2O solution by integrating online and offline data for prediction of consumer behaviors and attributes.

Manufacturing

We automate manufacturing operations to maximize production efficiency and recommend counteractions in emergency situations.

Network Security

Using our distributed and collaborative computing methods, we provide advanced attack detection and threat avoidance.

Health Care

Our framework enables early detection of disease, through analysis of genomic, vital signs, environmental, and infectious disease trends data.

Life Science

Our integrative, multi-purpose system detects and predicts new phenomena, advancing developments in biotechnology, drug discovery, and health care.

Public Safety

Powered by deep learning, our leading-edge video analytics system provides real-time threat detection for any environment.

Transportation

We automate vehicle and transportation operations through computer vision and collaborative, real-time analysis of multiple types of data.