社会に浸透するAIロボットの現状と未来

～ロボットはなぜお茶を持ってきてくれないのか ～

中京大学　工学部
橋本 学

Advanced Sensing & Machine Intelligence Group,
Chukyo University

略歴

1987　大阪大学大学院工学研究科 博士前期課程 修了
1987　三菱電機(株)　生産技術研究所，先端技術総合研究所，開発本部（R&D戦略・研究企画）

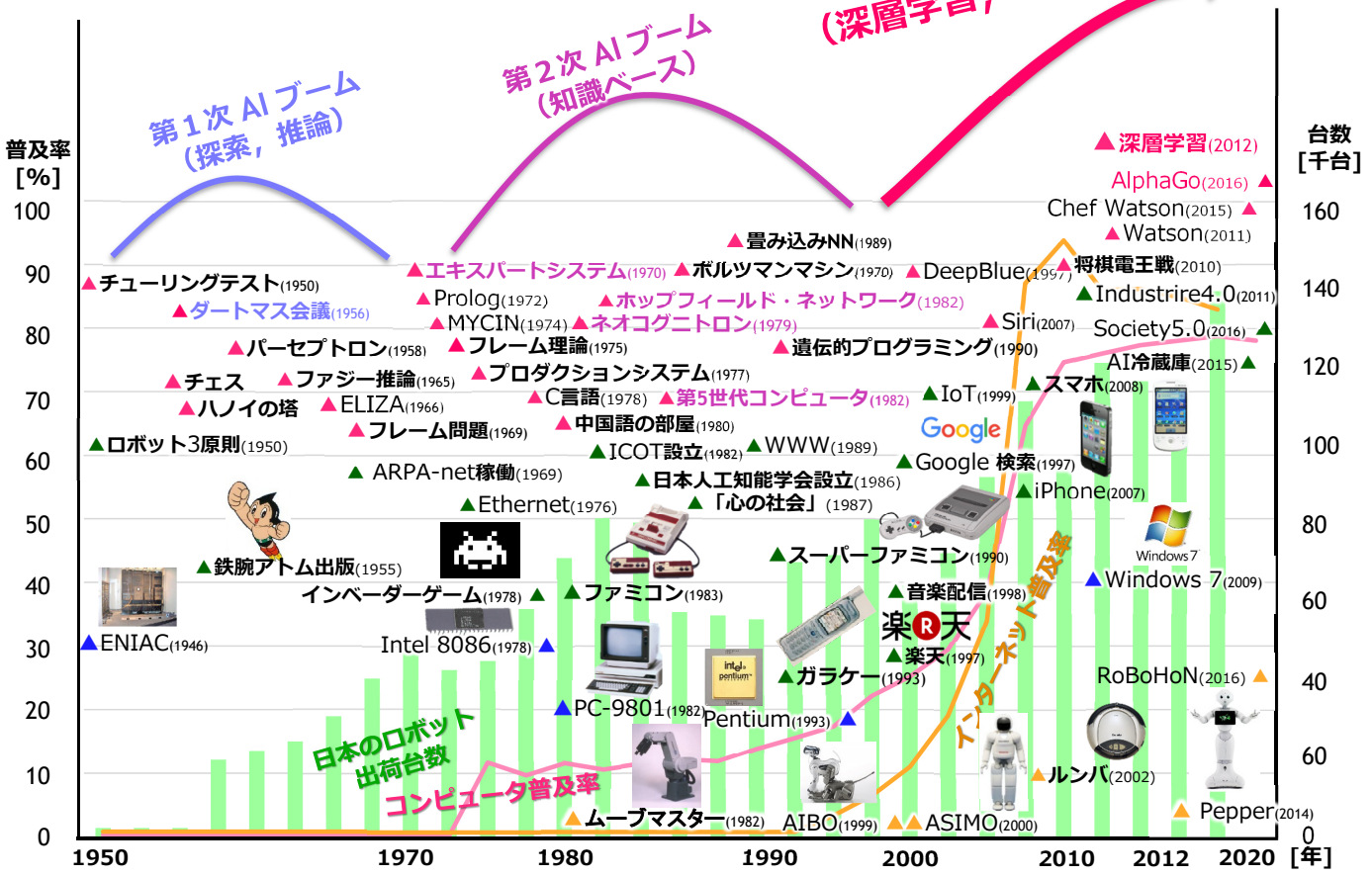
2008　中京大学 情報理工学部 教授
2013　中京大学 工学部（改組）教授

2017～　中京大学　工学部長

兼務　産業技術総合研究所 人工知能研究センター 客員研究員
　　中京大学 人工知能高等研究所 研究員
専門　ロボットビジョン，3次元物体認識，ヒューマンセンシング

AIへの大きな期待

【橋本研究室作成】　参考データ：主要耐久消費財の普及（総務省），通信利用動向調査平成26年報告書（総務省），年間統計推移表（日本ロボット工業会）



GAFAの動向　v.s.　日本型知能化

GAFAにおけるロボティクス開発

Google

Princeton, MITらとTossingBot

分解ロボットDaisy

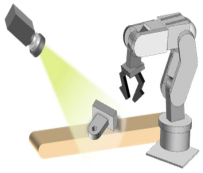
amazon

ロボット工学研究所を設立

Amazon Robotics

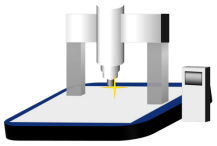
日本型知能化

自動ピッキング



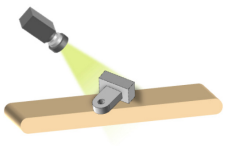
ロボットの知能化

自動加工



加工機の知能化

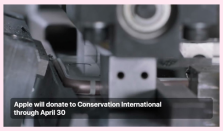
自動検査



検査装置の知能化

（資料）GAFA の動向

買収された企業
自社による研究ビジネス

	2010	2015			2019
Google	人型ロボット Schaft.inc ロボットアーム Redwood Robotics ロボット Meka Robotics ロボット Holomni	ロボット型カメラ Bot & Dolly ロボット Boston Dynamics ロボット SCHAFT	スマートホーム Nest Labs スマートホーム Dropcam スマートホーム Revolv	ロボット Boston Dynamics売却（ソフトバンクが買収） 自動車 Lumedyne Technologies 	大学（MIT,Princeton,Colimbia）と共同のロボット研究 Robots Learning to Toss
Amazon.com	生産用機械 Kiva Systems ↓ 棚を運ぶ AGV Kiva 		住宅用セキュリティ機器 Blink スマートホーム ドアベル Ring	Amazon Robotics Challenge 開催 2015～2017 amazonrobotics CHALLENGE	Kiva を改良した AGV Xanthus と Pegasus
Facebook		ハードウェア Oculus VR 航空機（ドローン） Ascenta	2機のチャットボット ボブとアリス	機械学習用の仮想ホーム 	6足ロボットを使用した研究
Apple Inc.	スマートフォン iPhone タブレット iPad		リサイクルロボット Liam（新事業） ↓ 分解ロボット Daisy（自社開発） 		モーションキャプチャ IKinema 自動運転 Drive.ai

でも，ロボットはお茶を持ってきてくれない！

本日の話題

1．イントロダクション
- 機械学習とディープラーニング

2．AI搭載ロボットの現状
- 国際ロボット競技大会に見る世界レベル

3．AIとロボット研究の最先端
- End-to-End の広がり
- お茶会ロボットが目指すもの

4．AI・ロボットの未来像
- AIの強みと弱み
- AI・ロボットのビジネス展開

本日の話題

1．イントロダクション
- 機械学習とディープラーニング

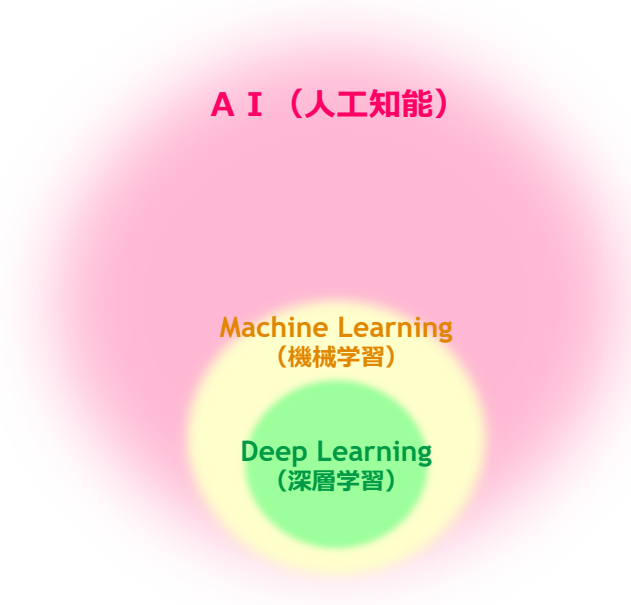
2．AI搭載ロボットの現状
- 国際ロボット競技大会に見る世界レベル

3．AIとロボット研究の最先端
- End-to-End の広がり
- お茶会ロボットが目指すもの

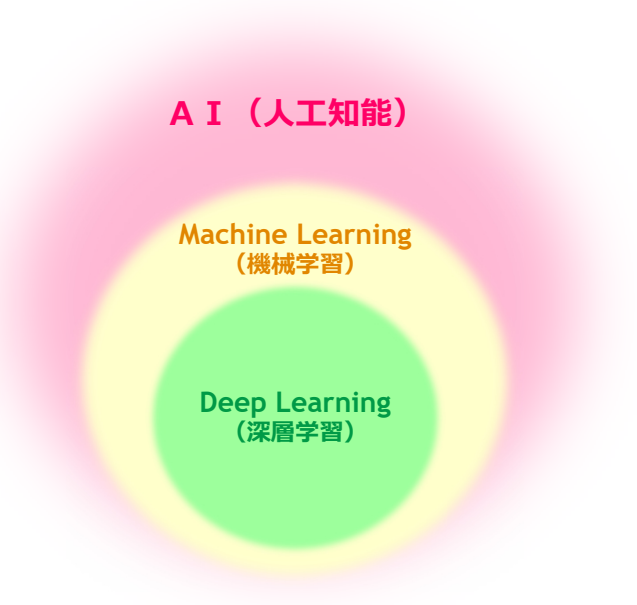
4．AI・ロボットの未来像
- AIの強みと弱み
- AI・ロボットのビジネス展開

AI・機械学習・Deep Learning

ベテラン？



若手？



画像認識問題

例題：写真に写っている５種類の動物の名前を推定する．

1.Dog　2.Cat　3.Lion　4.Panda　5.Giraffe

入力データ（画像 x ）

$y = f(\mathbf{x})$ → $y = 1$ (Dog)

$y = f(\mathbf{x})$ → $y = 1$ (Dog)

$y = f(\mathbf{x})$ → $y = 2$ (Cat)

$y = f(\mathbf{x})$ → $y = 4$ (Panda)

考え方：理想的な関数 f が決まれば，画像 x に対する正しい y が出力される．
では，いったい，どうやって f を決めればよいのか？

関数 f の求め方

入力データ（画像 x ）

$$y = f(\mathbf{x})$$

$y = 1$ (Dog)

もし，事前に f のカタチがわかっていたら？

‥‥いくつかの (x, y) の組を与えればよい．

（例）

$y = f(\boldsymbol{x}) = ax + b$ ならば，少なくとも２組の (x, y)

$y = f(\boldsymbol{x}) = ax^2 + bx + c$ ならば，少なくとも３組の (x, y)

しかし，もし $f(x)$が n 次の多項式じゃなかったら？

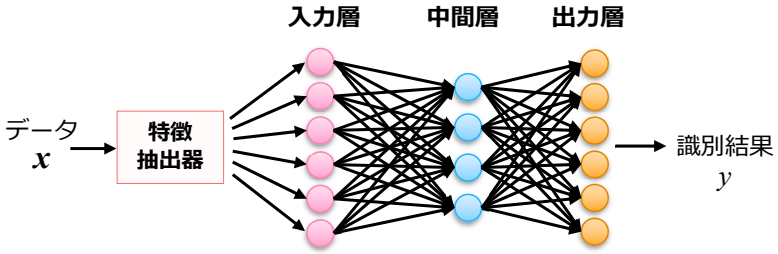
そもそも，$\boldsymbol{x}$ と y の関係は，関数で表せるのか？

‥‥何らかの解法が必要．

機械学習によって f を決める

Neural Network（従来）

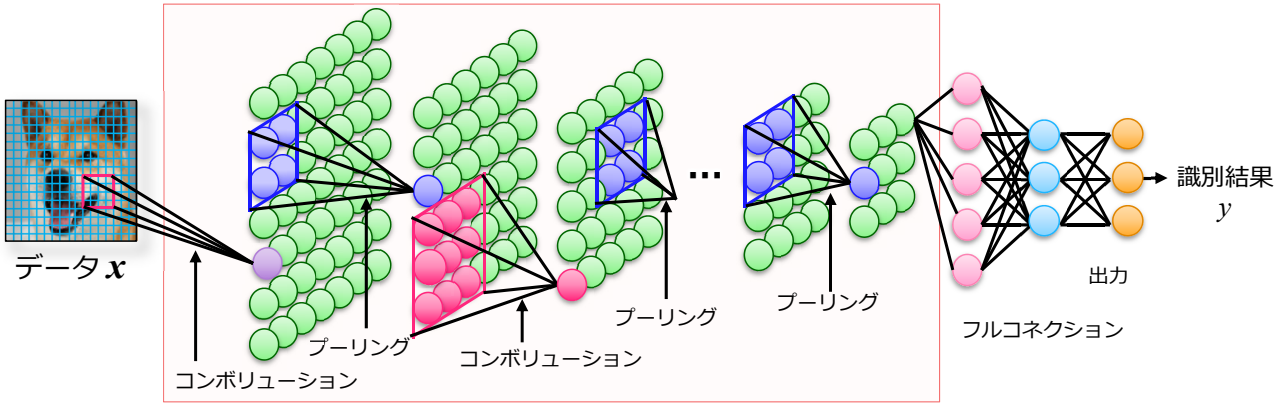
脳の神経回路網（イメージ）



例：多層パーセプトロン（原型は1958）

学習法：誤差逆伝播法（1986）

Deep Learning（現在）



30年のコンピュータの進歩

1987年

Sun Microsystems Sun-4

CPU: 25MHz ---100倍---→ CPU: 2.65GHz（x6コア）

メモリ: 128MB ---35倍---→ メモリ: 4GB

2019年

Apple iPhone 11

大量の画像群を用いた機械学習

学習のために使用する模範解答つき画像群

y=1 (Dog) y=2 (Cat) y=3 (Lion) y=4 (Panda) y=5 (Giraffe) ・・・

入力データ（画像 x）

$$y = f(\mathbf{x})$$

y = 1 (Dog)

x（ベクトル）と y のセットをたくさん与えることによって，f を決める．

未知画像 x

$$y = f(\mathbf{x})$$

y = 1 (Dog)

Deep Learning の開発環境

■ Deep Learning 向きシステム（NVIDIA社の例）

DGX-1

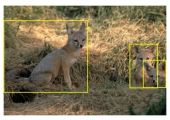
DGX station

DGX-2

■ Deep Learning に向いた大量データベースの整備

ImageNet http://image-net.org/index
[スタンフォード・プリンストン大学]

- 2009年から公開
- 総画像数：14,197,122 枚
- カテゴリ数：21,841

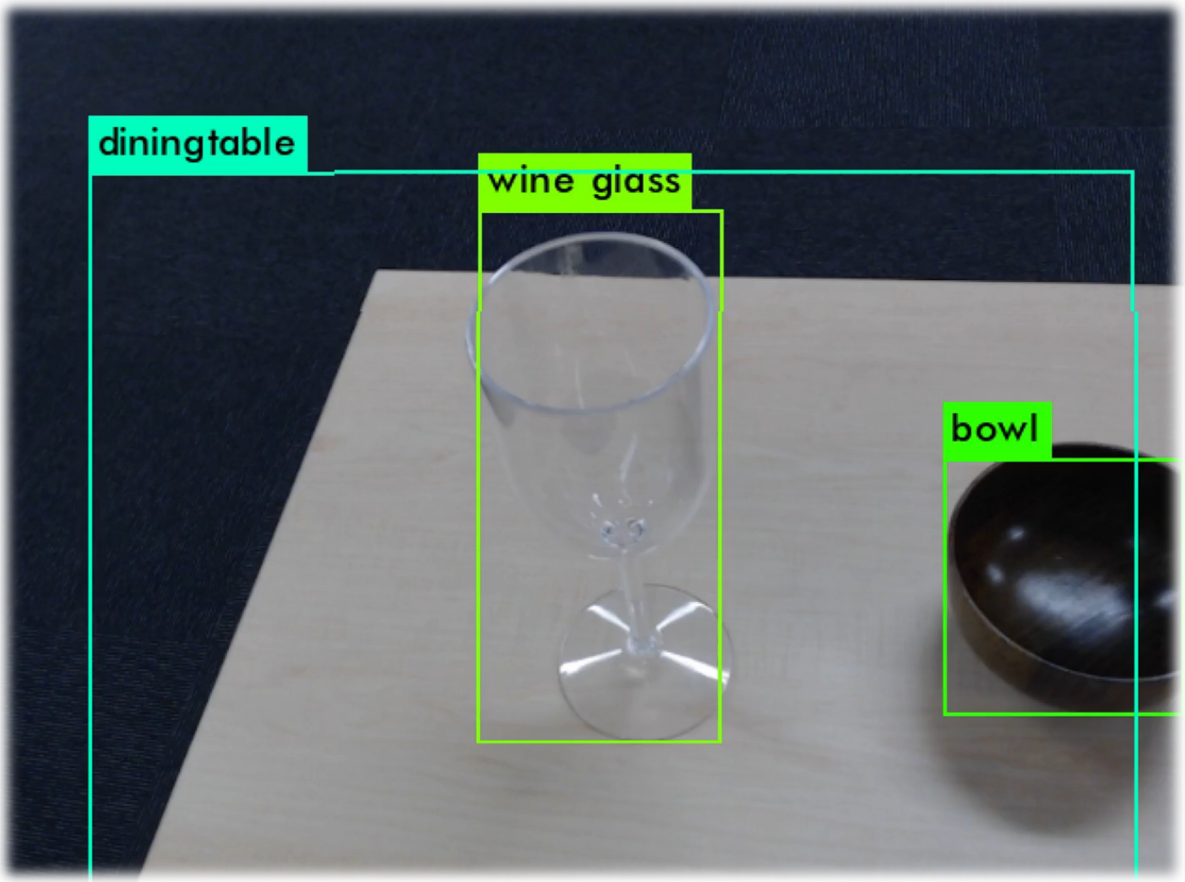


ShapeNet http://shapenet.cs.stanford.edu/
[プリンストン・スタンフォード大学，TTIC]

- 2015年から公開
- モデル数：3,000,000 個
- カテゴリ数：3,135

Deep Learning による物体認識

（0分30秒）



Deep Learning は，とりわけ画像認識の分野では，
予想以上の実力が確認されている

本日の話題

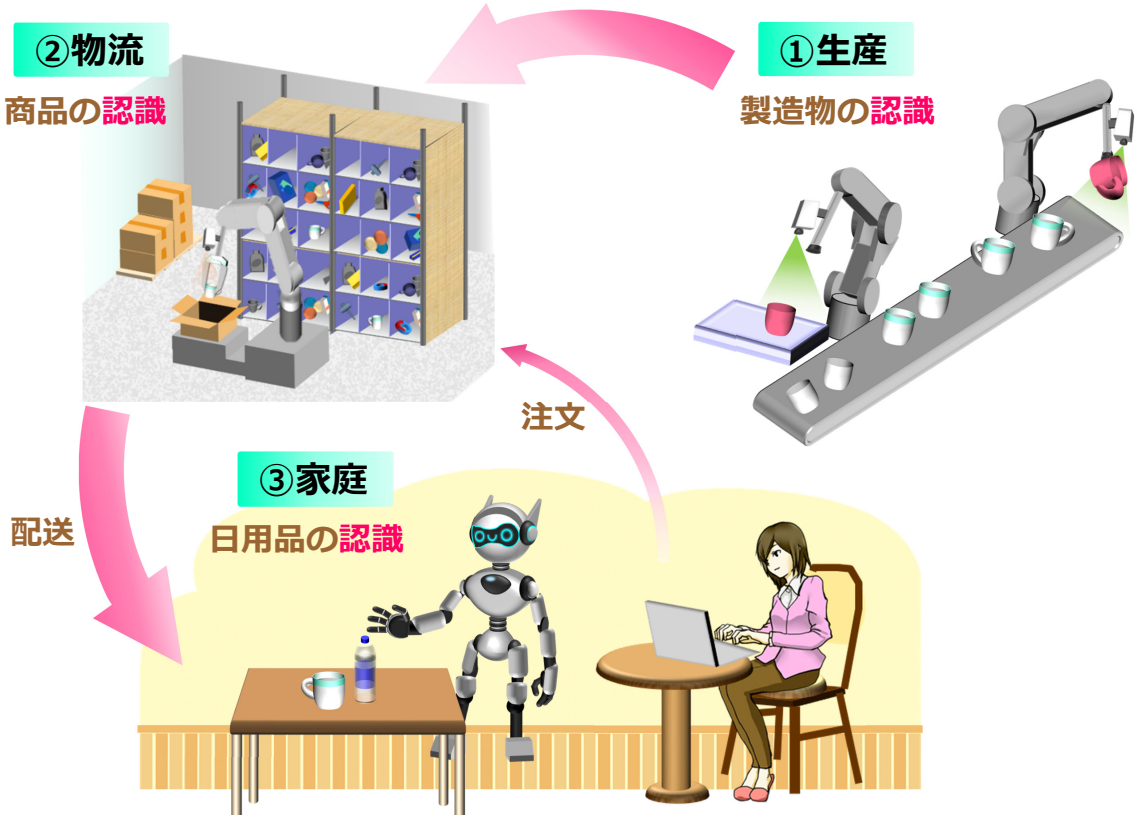
1．イントロダクション
- 機械学習とディープラーニング

2．AI搭載ロボットの現状
- 国際ロボット競技大会に見る世界レベル

3．AIとロボット研究の最先端
- End-to-End の広がり
- お茶会ロボットが目指すもの

4．AI・ロボットの未来像
- AIの強みと弱み
- AI・ロボットのビジネス展開

AIロボットの利用分野

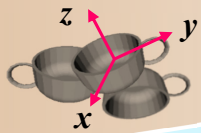


主要３分野におけるAIロボットの課題整理

Pick and Place	Pick and Operation
動作パターン種類 少（事前に生成）	動作パターン種類 多（その場で生成）
物体モデル利用 可	物体モデル利用 難
品種 複数だがほぼ既知	品種 多量かつ未知

特定物体認識

モデルを利用して，
対象物の位置と姿勢を認識する



一般物体認識

モデルを使わず，
クラス名（モノの名前）を認識する

“コップ” というクラス

事例①　Amazon Robotics Challenge

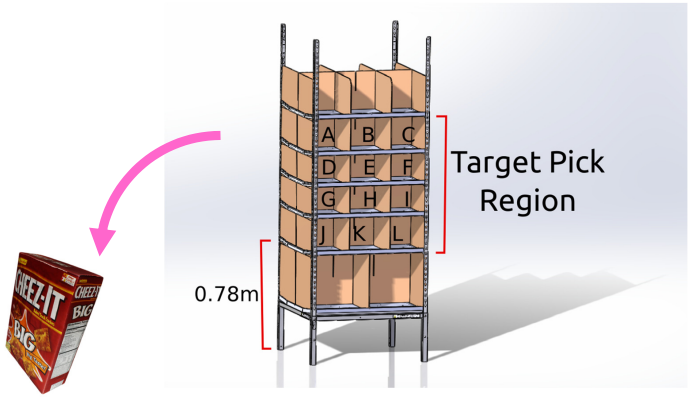
物流ロボット

Amazon チャレンジへの挑戦履歴　（三菱電機，中部大とコラボ）

2015年　第１回大会（シアトル）
2016年　第２回大会（ライプツィヒ）
2017年　第３回大会（名古屋）

2017年度 Stowタスク部門世界3位

制限時間内にどれだけ正確にアイテムをピッキングできたかを競う



アイテムの例（39種類）・・・・2016年度

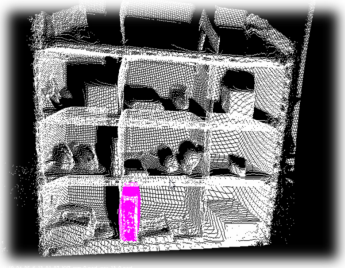
アイテムの例（39種類）・・・・2016年度

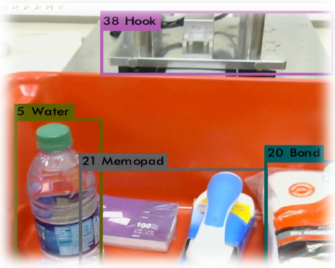
- 形状や重量，表面の硬さが異なる多様なアイテム
- 認識や把持が困難なアイテムには加点あり

Amazon Robotics Challenge 2017：16 Finalists

ロボットビジョンの基本課題

どこに，　なにがある？ → どう掴む？





Amazonチャレンジで開発した物体認識アルゴリズム

センサデータ

Deep Learning
による物体認識

低信頼

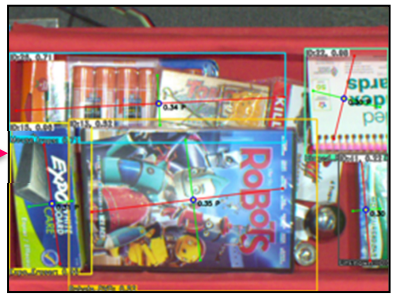
高信頼

異種データを利用
した再識別

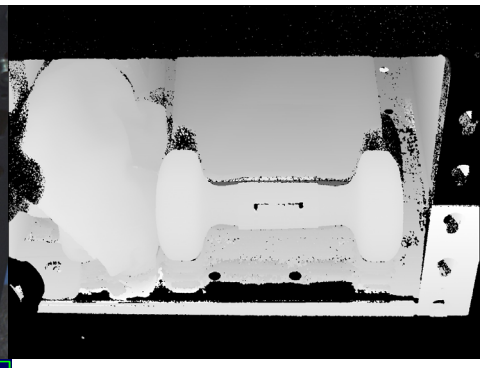
重さ・カラー
ヒストグラム

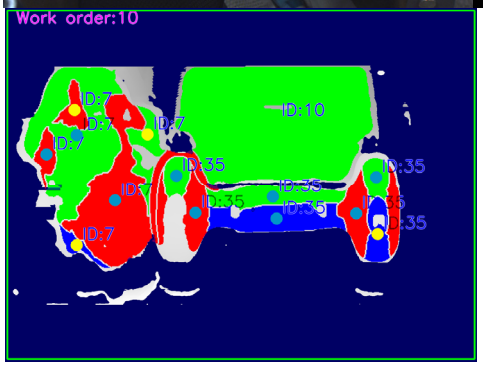
Deep Learning
による
把持パラメータ推定

認識結果



Deep Learning による認識結果の例



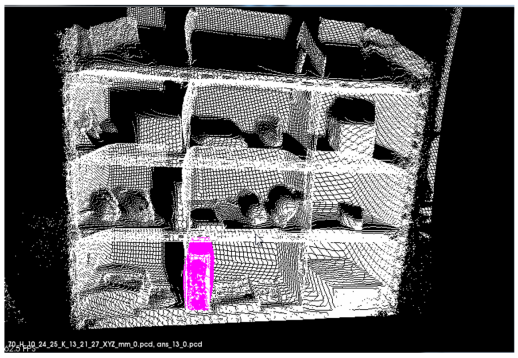


- オーダーアイテム
  ID 10 (カーテン)
- Binの中のアイテム
  ID 7 (Tシャツ)
  ID 10 (カーテン)
  ID 35 (ダンベル)
- **認識成功率**
  14 / 14

物体認識結果

箱状物体

入力シーン外観

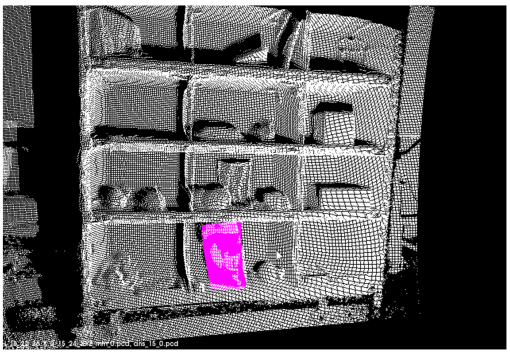


認識結果

対象物

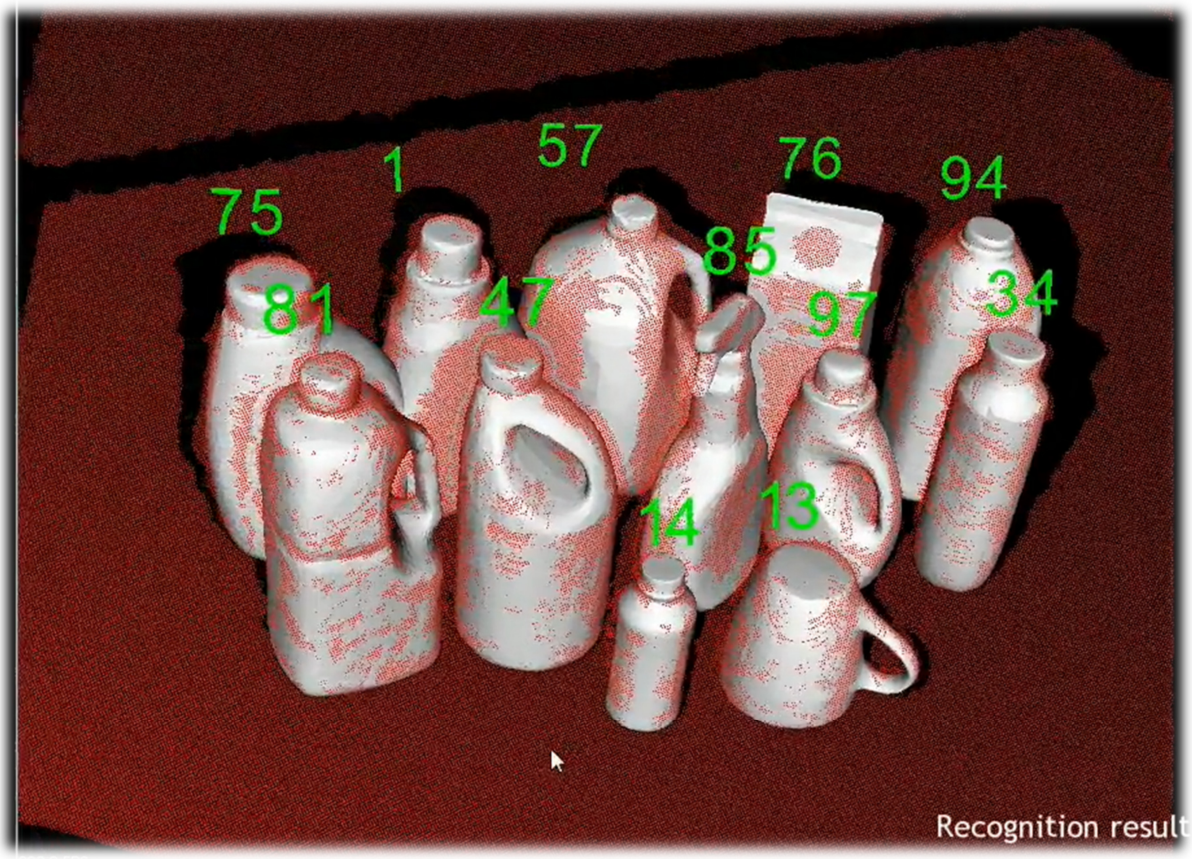
形が変形しやすい袋状物体

入力シーン外観

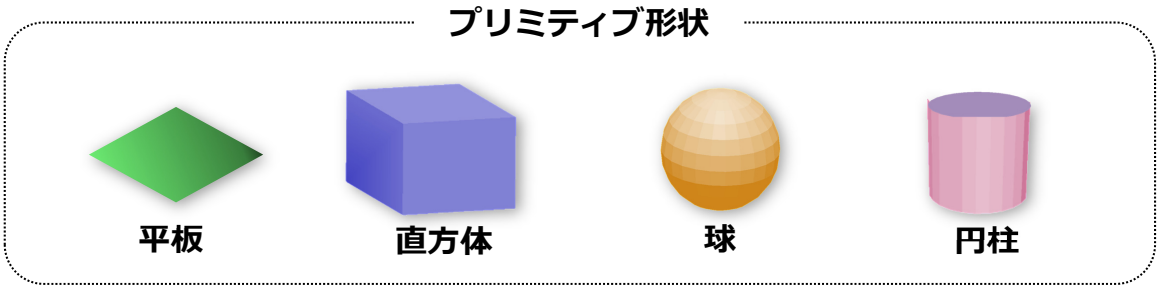


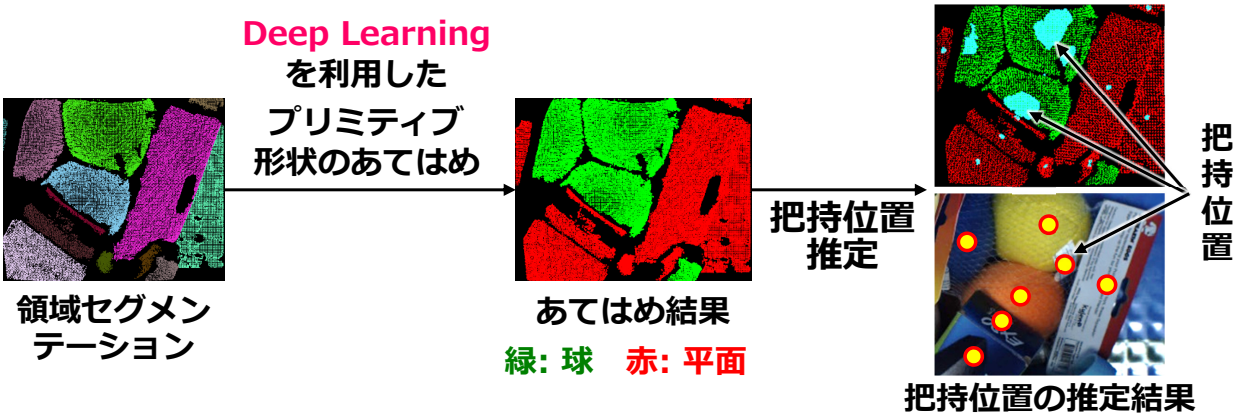
認識結果

複数物体の同時位置姿勢認識

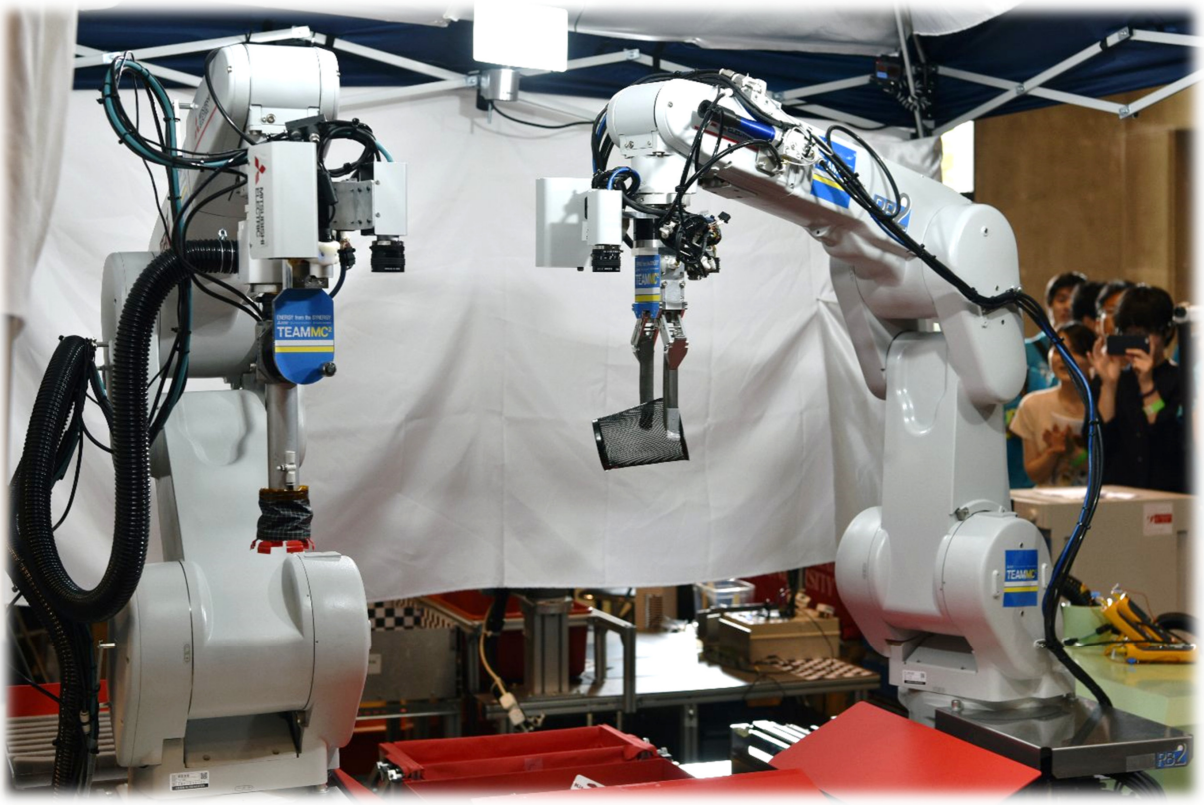


プリミティブ形状近似に基づく把持位置推定





Amazon チャレンジに出場した協調ロボット



競技開始

Amazon チャレンジ　名古屋大会

（2分21秒）

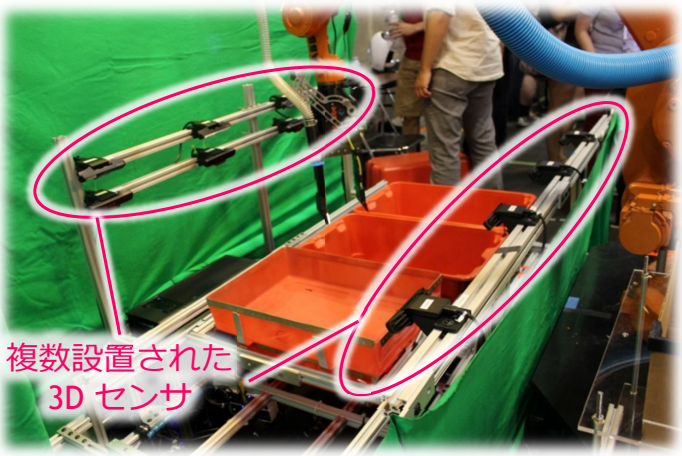
出典：Amazon Story, https://www.amazon.co.jp/b?node=3573739051

ARC2017　Stowタスク部門の 順位表

順位	チーム名	得点	所属
1	MIT-Princeton	160	マサチューセッツ工科大学，プリンストン大学
2	Nanyang	125	南洋理工大学
3	MC^2	120	中京大学，三菱電機，中部大学
4	NAIST-Panasonic	110	奈良先端科学技術大学大学院，パナソニック
5	IITK-TCS	105	インド工科大学カーンプル校，タタ・コンサルタンシーサービシズ
6	ACRV	55	クイーンズランド工科大学
7	IFL PiRo	25	カールスルーエ工科大学，ヘルムートシュミット大学
8	NimbRo Picking	20	ボン大学
9	PLAID	20	カーネギーメロン大学
10	UJI RoblnLab	20	ジャウメ1世大学
11	Team T2	10	鳥取大学，東芝
12	Team Duke	10	デューク大学
13	Team K	5	東京大学
14	Applied Robotics	5	スマートロボティクス，シドニー大学
15	TKU M-Bot	5	淡江大学
16	GMU-Negev	0	ジョージメイソン大学，ネゲヴ・ベン=グリオン大学

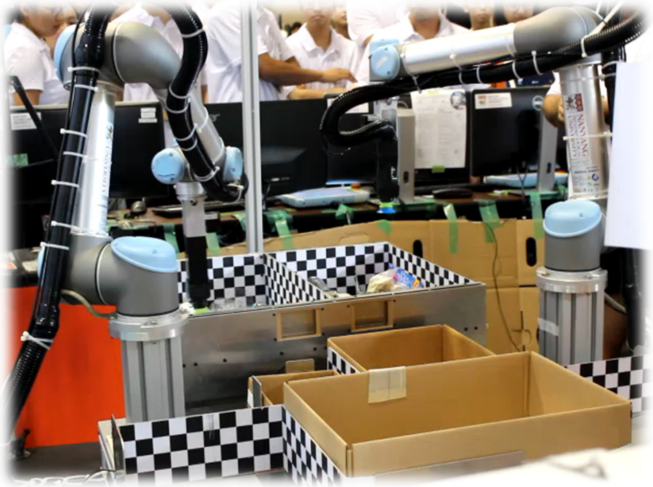
Stow優勝：　MIT-Princeton（アメリカ）

- 特徴
  - ロボットシステム全体
    - 単腕ロボットアーム
  - ビジョン部
    - 16 個の 3D センサを利用
    - Deep Learning を用いた物体認識
  - ハンド部
    - 吸着＋挟持



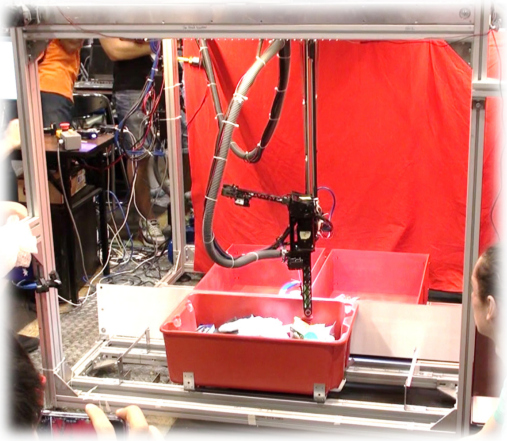
Pick優勝：　Nanyang（シンガポール）

- 特徴
  - ロボットシステム全体
    - 双腕ロボットアーム
    - 拡張可能なストレージ
  - ビジョン部
    - キーポイントマッチングをメインとしたアイテム認識
  - ハンド部
    - 吸着



Final優勝：　ARCV（オーストラリア）

- 特徴
  - ロボットシステム全体
    - XYZ ステージによる制御
  - ビジョン部
    - Deep Learning を用いた物体認識
  - ハンド部
    - 吸着＋挟持



事例②　World Robot Summit

コンビニロボット

World Robot Summit（WRS）

Future **C**onvenience **S**tore **C**hallenge（コンビニロボット）

コンビニを模した競技会場
（セブンイレブン様が協力）

サンドイッチの「陳列・廃棄」タスク

生産・物流分野では,
AI搭載ロボットが着実に実用域に近づいている

本日の話題

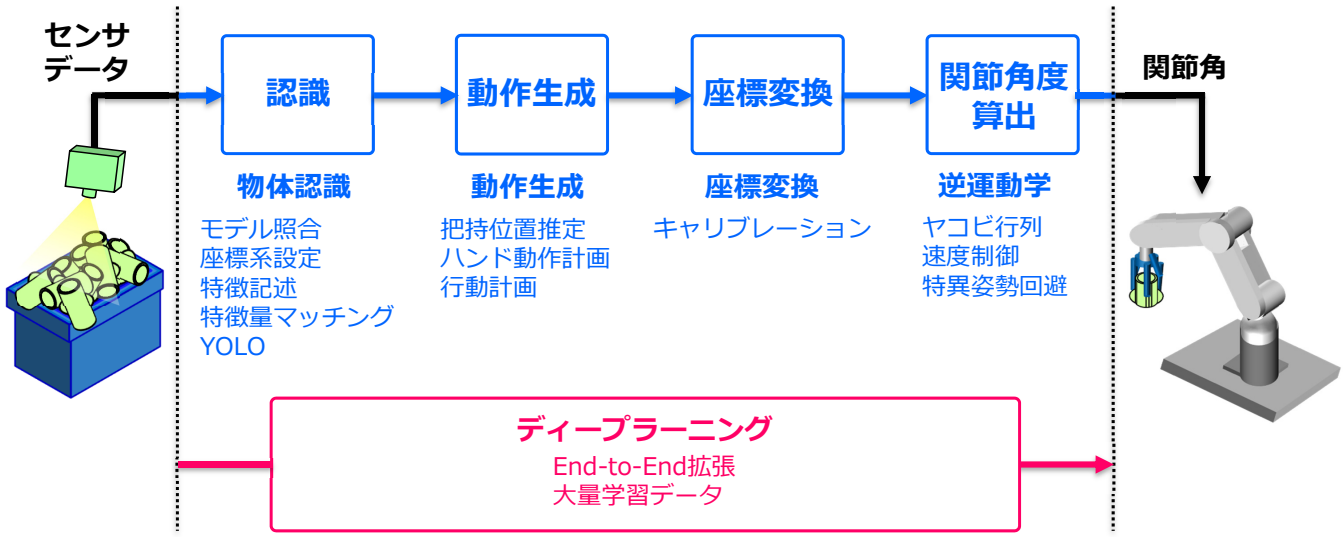
1．イントロダクション
- 機械学習とディープラーニング

2．AI搭載ロボットの現状
- 国際ロボット競技大会に見る世界レベル

3．AIとロボット研究の最先端
- End-to-End の広がり
- お茶会ロボットが目指すもの

4．AI・ロボットの未来像
- AIの強みと弱み
- AI・ロボットのビジネス展開

従来のロボット制御　v.s.　End-to-End 機械学習

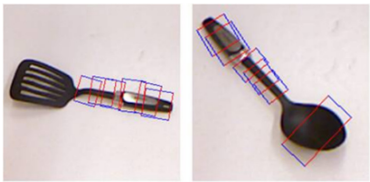


- End-to-End が広がっている
- ただし，ネットワークが複雑化し，必要な学習数も大きい

RGBD画像 ⇒ ハンド位置＋角度

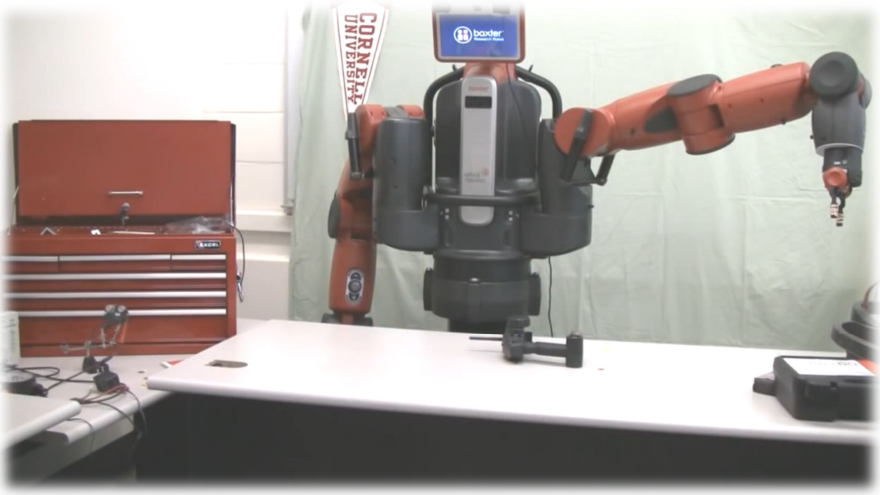
RGB 画像と Depth 画像からハンドの把持位置と角度（４次元）を推定

- Deep Learning による把持位置検出に関するおそらく初めての論文
- **RGB画像とDepth画像を効率よく学習**するためのマルチモーダル正則化
- ２段階カスケード構造を持つ Deep Learning
- 学習枚数：1035 枚



教示タグ付き画像の例

把持位置の推定結果



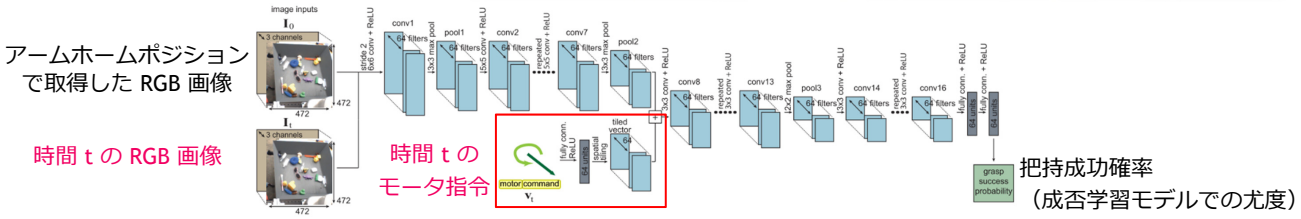
出典： I Lenz et al., “Deep Learning for Detecting Robotic Grasps”, IJRR, 2014.　http://y2u.be/f9Cuzql1SkE

画像 ⇒ モータ指令

[Google 2016]

3000 時間（80万回）×14 台のロボットの把持試行により，入力画像とモータ指令をもとに把持成功率を推定するモデルを構築

- 対象物への接近中に様々なモータ指令を与えて，把持成否サンプルを記録
- サンプルを利用してネットワークで学習（把持成否は 2 値）
- 把持成功率が高いモータ指令を採用

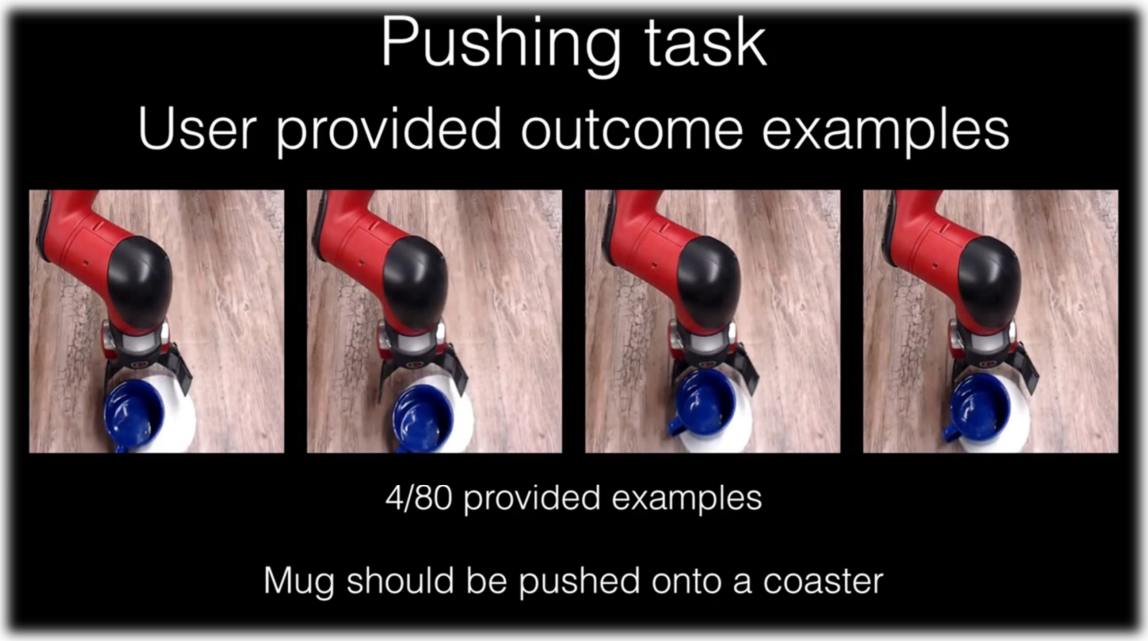


モータ指令と RGB 画像を同時入力する**逆強化学習**ネットワーク

S. Levine et al., "Learning Hand-Eye Coordination for Robotic Grasping with Large-Scale Data Collection", ISER, pp.173-184, 2016.
S. Levine et al., “Learning Hand-Eye Coordination for Robotic Grasping with Deep Learning and Large-Scale Data Collection” IJRR, 2018.

画像 ⇒ モータ指令　（強化学習）

- マグカップをコースターの上に移動させると「成功」
- ゴールを教えておき，成否によって報酬を与える（強化学習）
- End-to-End の学習により，正しい行動パターンを獲得.



University of California, Berkeley, 出典：https://www.youtube.com/watch?v=9pWJzb4G-CA

全自動お茶会ロボット

ロボットはなぜお茶を持ってきてくれないのか

お茶をたてる動作（ロボットの場合）

お茶を入れる → **お湯を入れる** → **かきまぜる**

- 茶さじを探す
- **持ち方を決定**
- **茶さじを持つ**
- ナツメを探す
- すくい方を決定
- **お茶をすくう**
- 茶碗を探す
- 入れ方を決定
- 茶碗に入れる

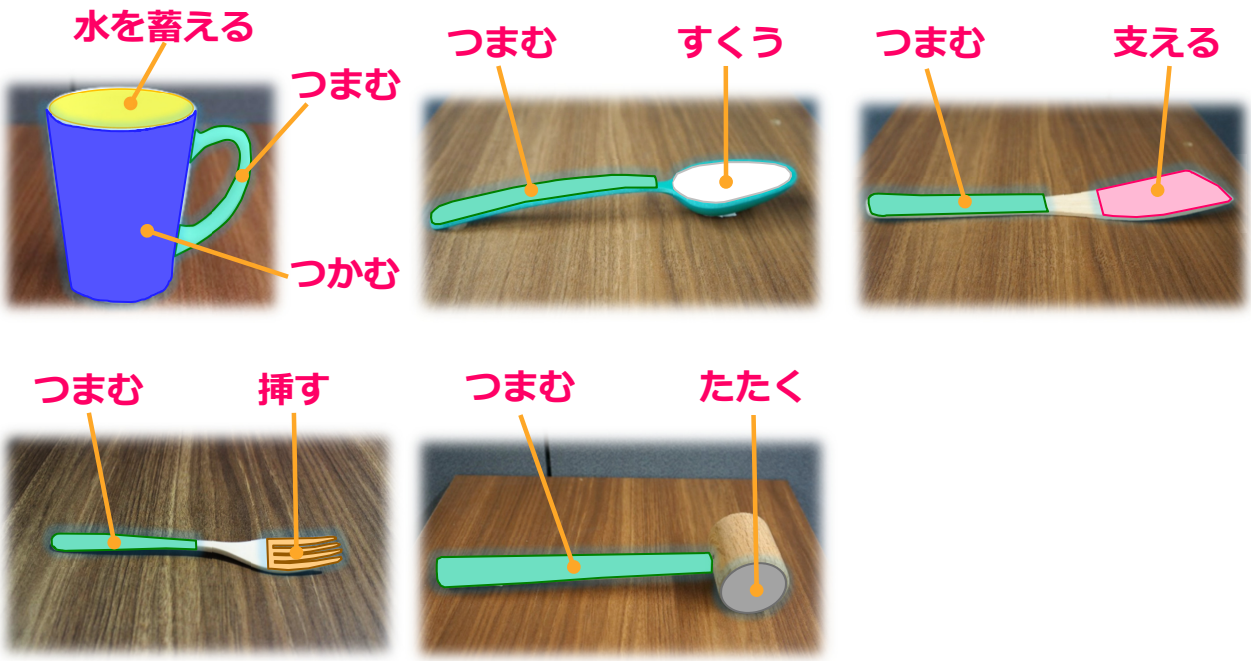
- 柄杓を探す
- **持ち方を決定**
- **柄杓を持つ**
- すくい方を決定
- **お湯をすくう**
- 茶碗を探す
- 入れ方を決定
- 茶碗に入れる

- 茶筅を探す
- **持ち方を決定**
- **茶筅を持つ**
- 茶碗を探す
- まぜ方を決定
- **まぜる**

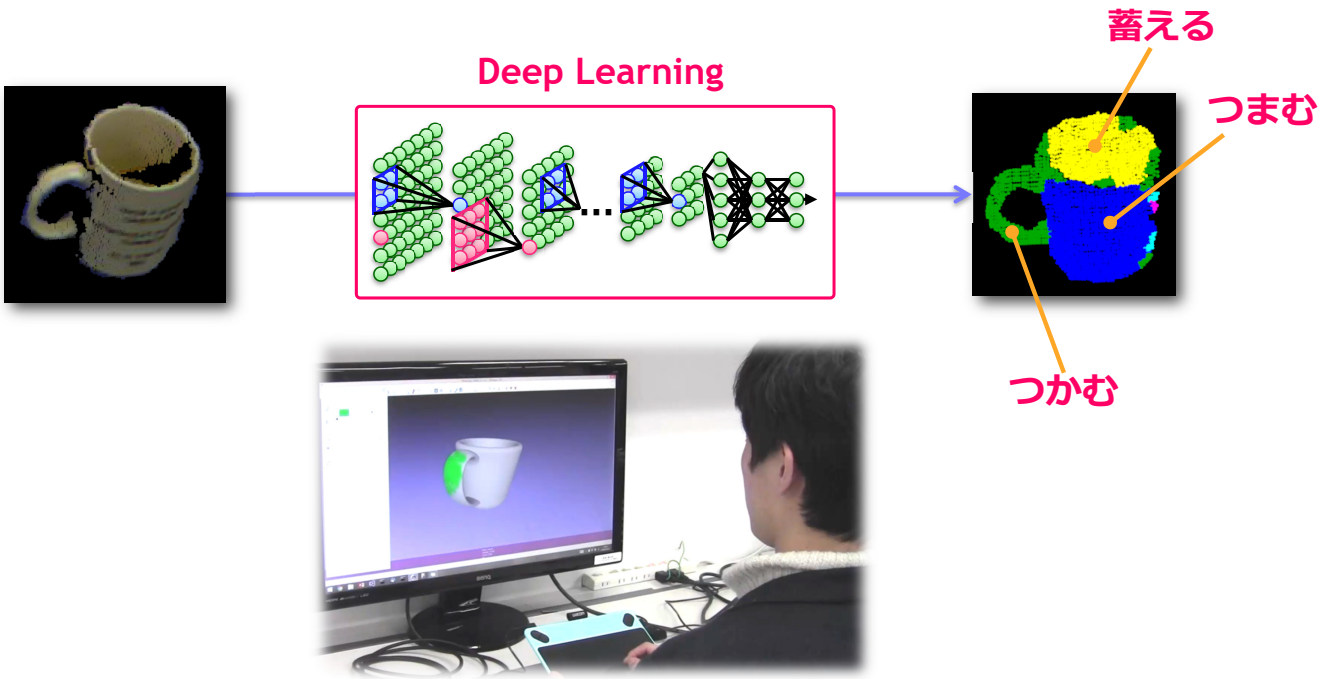
人間なら，道具の**使用目的**を想定して**持つ**ことができる．

⇒ **道具の使い方がわかるロボット**

日用品は “形” に “機能” を持っている



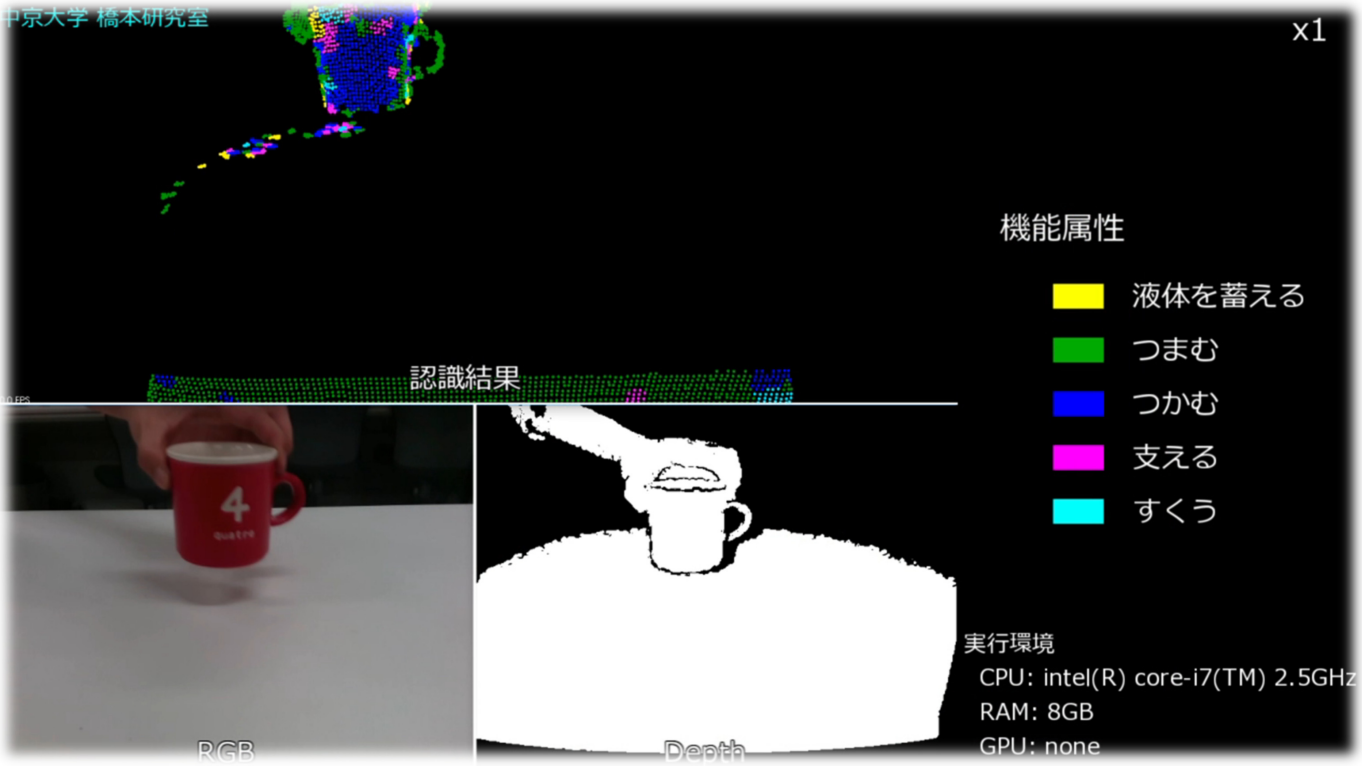
モノの “機能” を認識するロボットビジョン



模範解答は，人間が教える

機能を認識しているようす

（44秒）



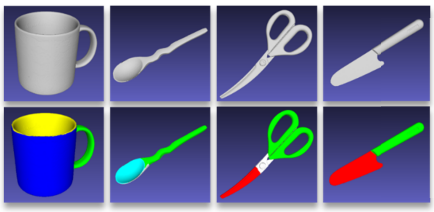
機能属性ラベル付き全周囲３Ｄデータセット（2019年5月24日公開）

データセット公開のねらい：　機能認識に関する研究活性化

http://isl.sist.chukyo-u.ac.jp/archives/nedopro/

a. 日用品 全周囲 3D モデル [モデル16品種321個，機能9クラス]

日用品の例



機能属性ラベル付き 3D モデルの例

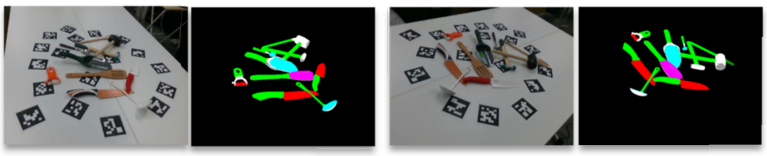
データ形式

3Dモデル	ply 形式 obj 形式 stl 形式
機能属性ラベル	ply 形式

使用例

全周囲データを用いたセグメンテーション【PointNet など】

b. 日用品シーン（2.5D） [モデル10品種11014枚，機能9クラス]



日用品シーンと機能属性ラベルの例

データ形式

RGB画像 2.5Dデータ	png 形式 pcd 形式
機能属性ラベル	png 形式

使用例

画像・点群ベースの機能認識（セグメンテーション）【SegNet，U-Net など】

a. のデータを用いたモデルマッチング

c. 機能属性セグメンテーション用シーン（2D）[モデル6品種2799枚，機能6クラス]

茶道具シーンと機能属性ラベルの例

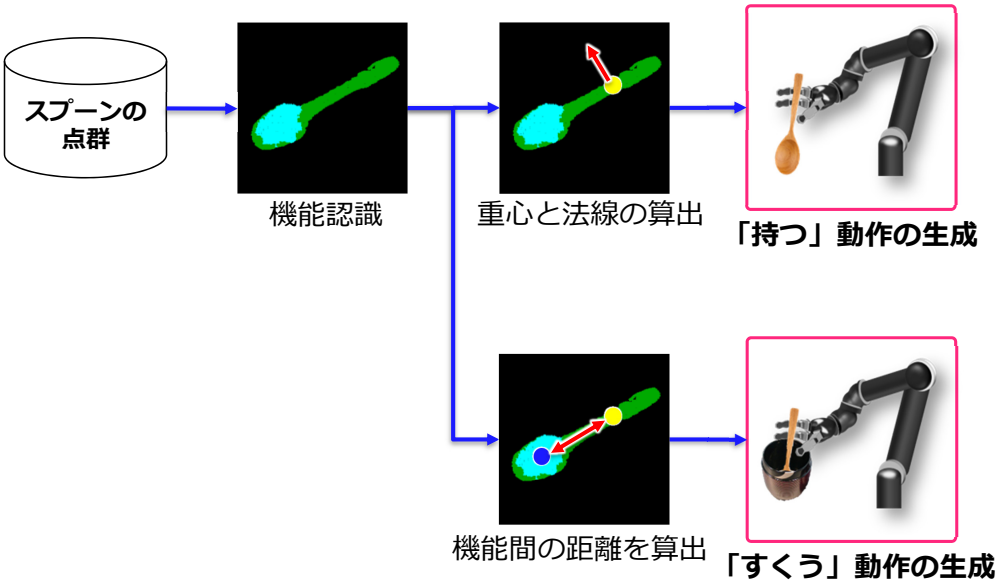
データ形式

RGB画像	jpg 形式
機能属性ラベル	png 形式

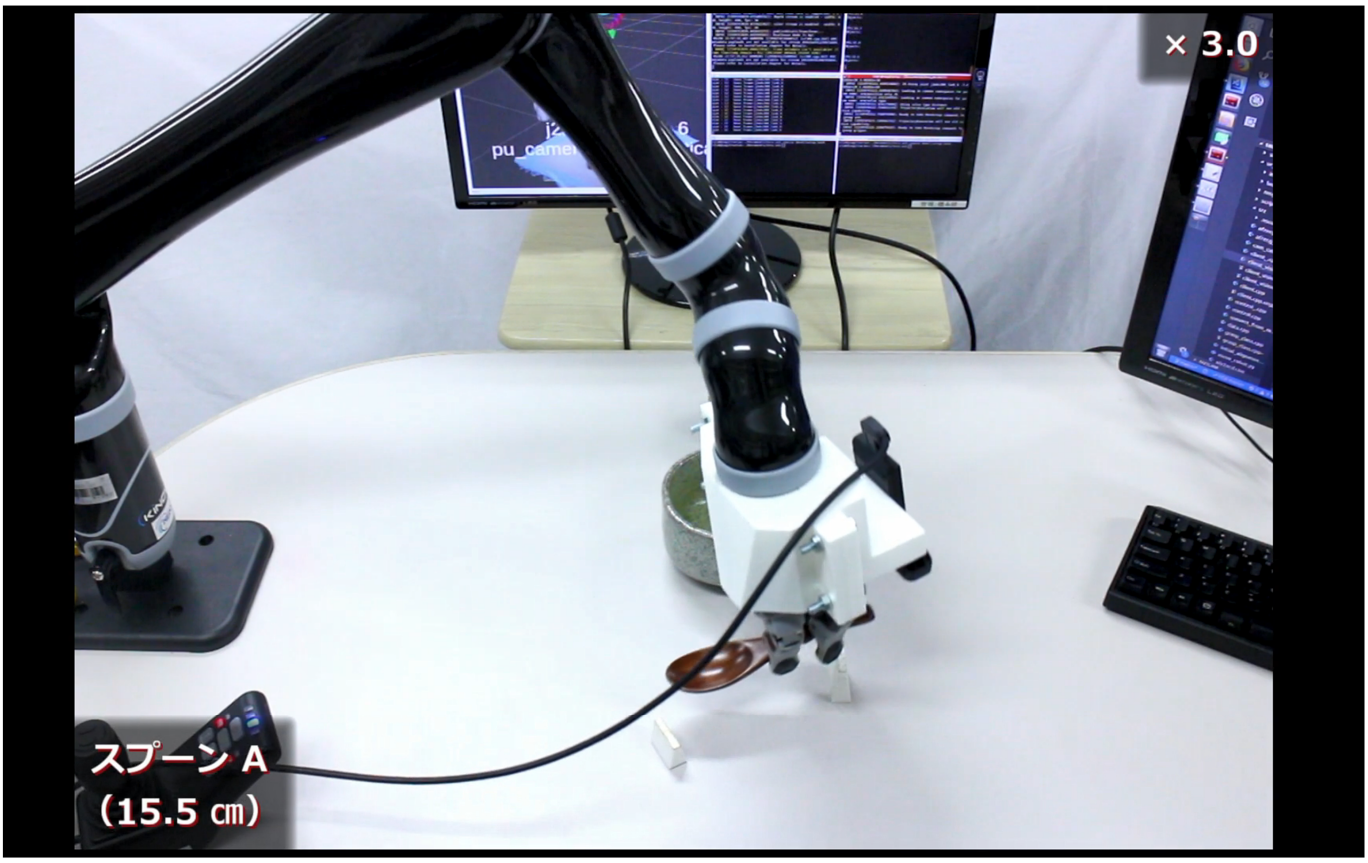
使用例

画像ベースの機能認識（セグメンテーション）【SegNet，U-Net など】

機能情報を利用した動作パラメータ自動生成

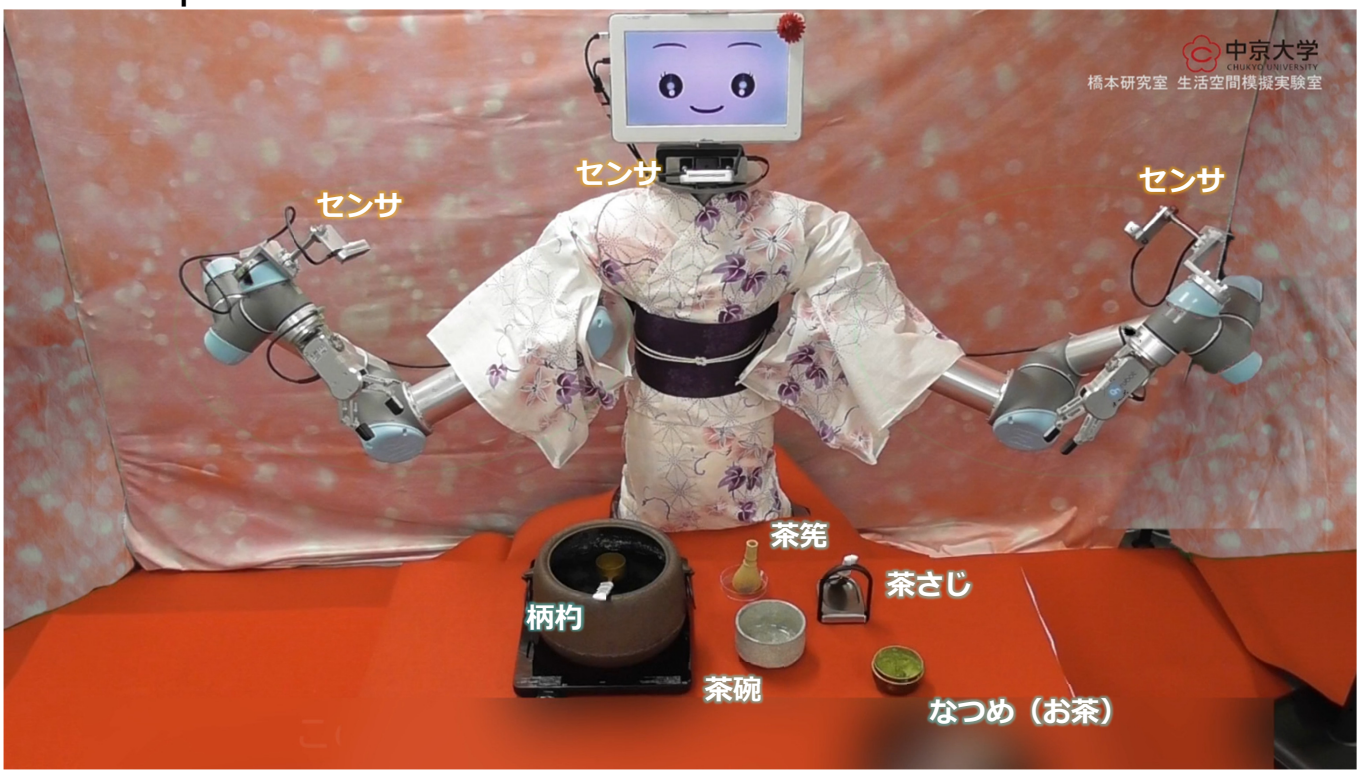


機能情報を利用した動作パラメータ自動生成



お茶会ロボットシステムの試作

CEATEC Japan 2018



CEATEC Japan 2018 （2018年10月16日～19日）

人間ってナンだ？超AI入門　シーズン２ 第9回「暮らす」

NHK Eテレ　2019年3月7日放送　http://www4.nhk.or.jp/aibeginner/x/2019-03-07/31/13497/1723021/

試作したお茶会ロボット

YouTube で公開中　（2分46秒）

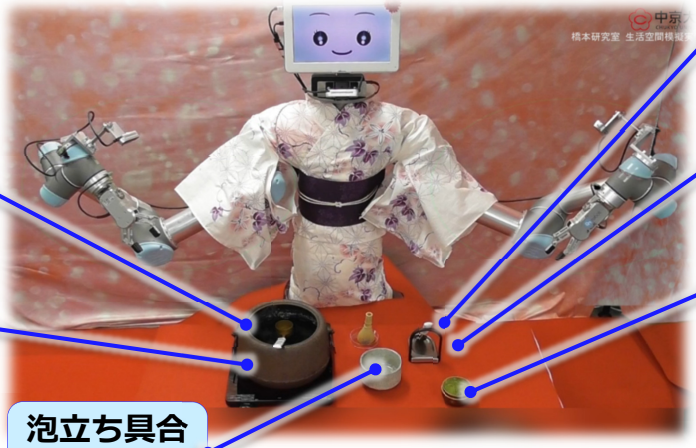
ロボットはなぜお茶を持ってきてくれないのか

「お茶を持ってきて」‥‥この一言に，まだ多くの難題が残っている！

そもそも，この言葉を，細かいタスクに分解できない
- 必要ならお湯を沸かす
- ナツメに蓋があったら開ける
- 足らない道具があったら取ってくる

⋮
⋮

お茶会ロボット（2019）にできないこと



- お湯を沸かす方法を知らない
- 余ったお湯を茶釜に戻すことができない
- 泡立ち具合は運次第
- 茶さじが裏向きだったら掴めない
- 小さな茶さじは掴めない
- お茶の量の調節ができない

AIとロボットの融合は，かつてないレベルに．
だが，人間のようなロボットの実現には，課題が多い

本日の話題

1．イントロダクション
- 機械学習とディープラーニング

2．AI搭載ロボットの現状
- 国際ロボット競技大会に見る世界レベル

3．AIとロボット研究の最先端
- End-to-End の広がり
- お茶会ロボットが目指すもの

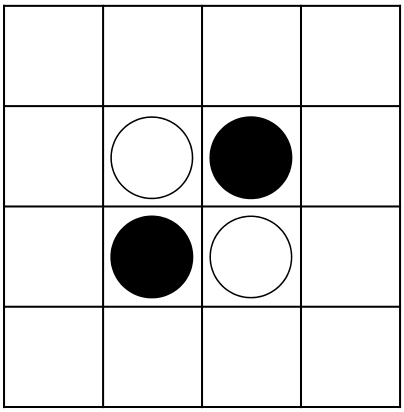
4．AI・ロボットの未来像
- AIの強みと弱み
- AI・ロボットのビジネス展開

AIの強み・・・・パターン処理が得意

- **入出力の間に，一定のパターンがある対象には強い**
- データの分析，相関の発見，人間の戦略策定支援に向いている．

- AIはなぜゲームに強いのか？
  → パターン化・数値化できるから（決定論的）．
- 膨大だが無限ではない探索空間から，近似最適解を求めればよい．

オセロ（実際は 8x8）



1. 最初は●からスタート．可能性は12通り．
2. 次は○で，11通り．
3. ・・・・
4. すべての場合の数は，12! ≒ 4.8億 通り．
5. 実際のオセロでは，60! ≒ 8×10^{80} 通り．
6. 可能性は有限（事実上は無限だが）

∴コンピュータがすべての手を記憶できたら，じつは，勝敗は最初から決まっている．

AIの強み・・・・パターン処理が得意

- **入出力の間に，一定のパターンがある対象には強い**
- データの分析，相関の発見，人間の戦略策定支援に向いている．

チェス

1997 AI vs カスパロフ

IBM Deep Blue

将棋

2012　AI vs 米長永世棋聖
2017　AI vs 佐藤名人

PONANZA

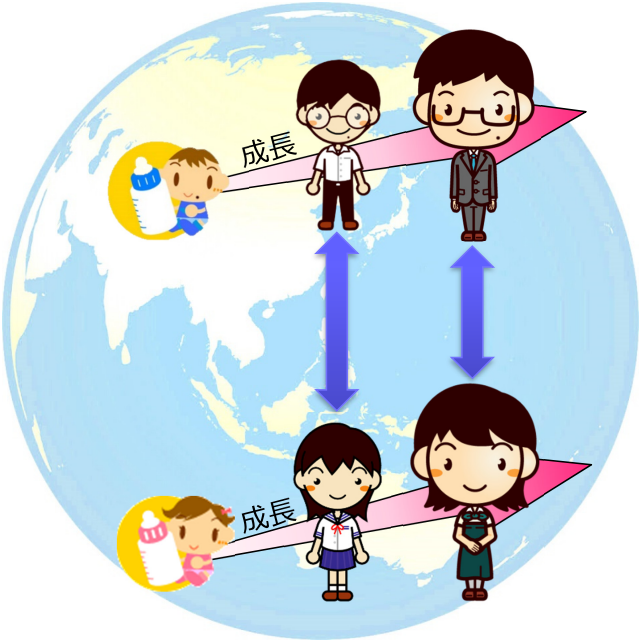
囲碁

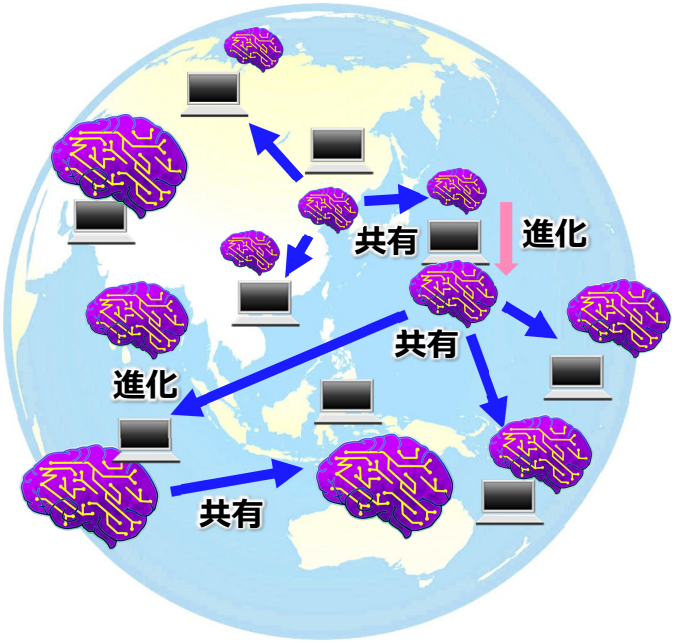
2016　AI vs イ・セドル九段
2017　AI vs カ・ケツ九段

AlphaGo Google DeepMind

AIの強み・・・・高速かつ高効率な進化サイクル

人間の成長　V.S.　AIの進化





①発達の初期値がリセットされる	→	①優秀な結果をもとにして進化する
②知識の伝達にコストがかかる	→	②よい結果は瞬時にシェアされる

「好奇心」をもつロボット

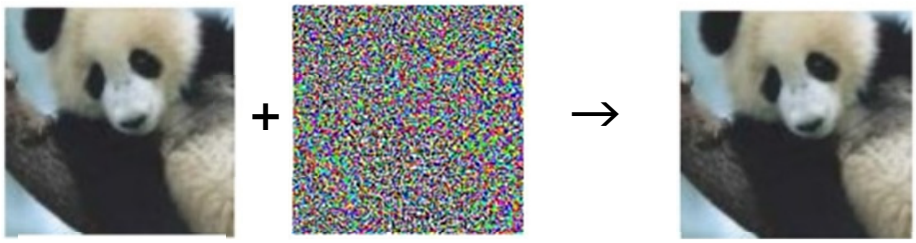
- 現在のAIは，まだ「自ら学ぼう」とはしない．
- しかし，ロボットに「好奇心」を持たせるための研究は，もう始まっている．実現した場合には，爆発的な進化が予想される．

University of California, Berkeley
出典：https://www.youtube.com/watch?v=Q4bMcUk6pcw

AIの弱み・・・・脆弱な面がある

認識結果を混乱させる画像（Adversarial example）を生成する技術

- **画像全体**の画素値を操作し，Deep Learning を欺く手法 [Goodfellow 2015]



パンダ 57.7%　特定パターン付与　**テナガザル 99.3%**

人とDLの判断基準の違いを利用して**人が気づけない画像加工**でDLを欺く

- 画像の一部の領域に特別なパターンを付与する手法 [T.B.Brown 2017]

バナナ 97.0%
ナメクジ 0.5%

特別なパターン
（小領域）

トースター 99.0%
バナナ 1.0%

任意パターンを映り込ませる（実物or画像加工）ことによってDLの誤識別を誘発

出典：G.Szegedy et al.,” Intriguing properties of neural networks” arXiv 2017. https://arxiv.org/pdf/1312.6199.pdf, Ian J. Goodfellow et al., “EXPLAINING AND HARNESSING ADVERSARIAL EXAMPLES”, ICLR 2015. https://arxiv.org/pdf/1412.6572.pdf, Tom B. Brown et al., “Adversarial Patch”, arXiv 2017. https://arxiv.org/pdf/1712.09665.pdf

AIの弱み・・・・抽象化が苦手

- Deep Learning は，現状では「シンボル」を扱っているに過ぎない
- 人間は，単語（シンボル）の**意味や概念**を理解して正しく解釈できるが，AIにはできない．

シンボルグラウンディング問題

曲がりネギ ＝ **「曲がっている」** ＋ **ネギ**

栃木県，宮城県などの特産

三省堂 辞典
[1] まっすぐでない
[2] 向きを変える

（意味，概念）

AIの弱み・・・・高次の解釈が苦手

TAE CAT

これも一種のシンボルグラウンディング問題

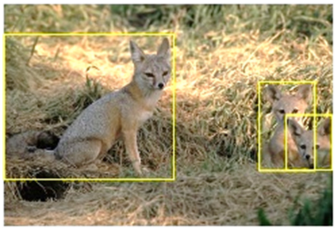
ただし，セマンティクス（意味論），文脈を利用した解釈をおこなうことによって，解決しつつある．

AIの弱み・・・・膨大な学習データが必要

ImageNet

[スタンフォード・プリンストン大学]

- 総画像数：14,197,122 枚
- カテゴリ数：21,841



データ拡充（Data Augmentation）

実在するデータをもとに，学習用データを人工生成する（≒水増し）

- 学習サンプルに幾何変換やノイズを施してデータを生成
- 近年は，２つ以上の学習サンプルを混合した新サンプル生成が主流に
- 生成より，それに付与すべき教師信号のほうが重要（品質は大丈夫？）

AIの弱み・・・・説明責任の不足

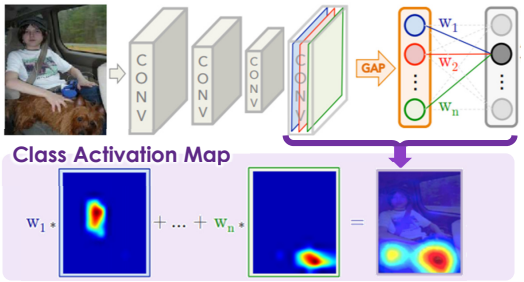
- **注目領域の可視化が可能になってきた**
- **しかし，それで十分か？　なぜ，そこに注目したのかは不明なまま．**
- **一方で，説明責任は本当に必要か？（事業性とのバランス）**

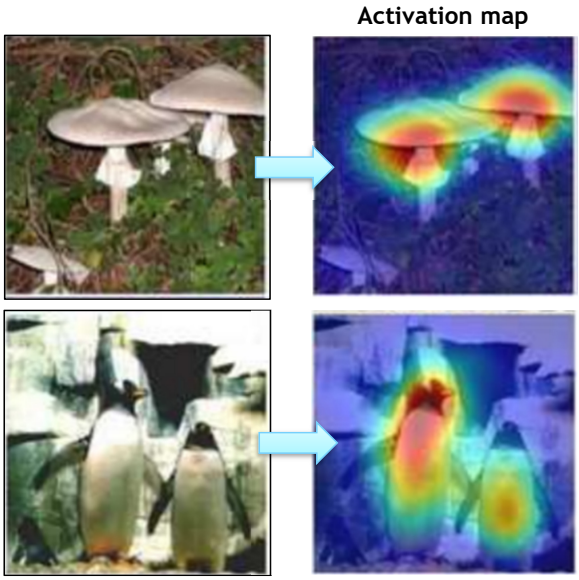
Class Activation Map　[B.Zhou 2016]

多クラス識別の推論時に Deep Learning が注目していた領域を可視化

可視化方法

畳み込み層の出力値と重みからマップを算出





AI・ロボットのビジネス展開

下線は想定する実用化・ビジネス例

難易度

ソフト

パターン認識主体

- 翻訳，記事作成，音声認識
- Deep Learning の本領発揮
- 入出力は定義しやすい
- 学習データ準備が鍵
- コモディティ化もしやすい

人間心理との融合

- UX，商品リコメンド，官能検査
- 入力は制約大きく，出力定義も困難
- マルチモーダル化が必須
- 言語，心理学との融合
- 「魅力」のセンシング

ソフト＋センサ群

データ分析（IoT，BigData）

- データ解析そのものがビジネス（工作機械メーカがコンペティタに）
- 入力データ収集の環境は整ってきている（まずは収集から始めるのが吉！）
- 出力をうまく設定できるかどうかが鍵
- 分析法はDLにこだわる必要なし（アカウンタビリティ重視の顧客が多そう）
- データシェア効率とセキュリティのバランス

ソフト＋身体（ロボ）

定型動作主体

- 生産ライン
- Pick&Operationは高難度
- 経済性と性能要求がシビア
- 汎用化にはセンサもネック
- 自律進化に期待大

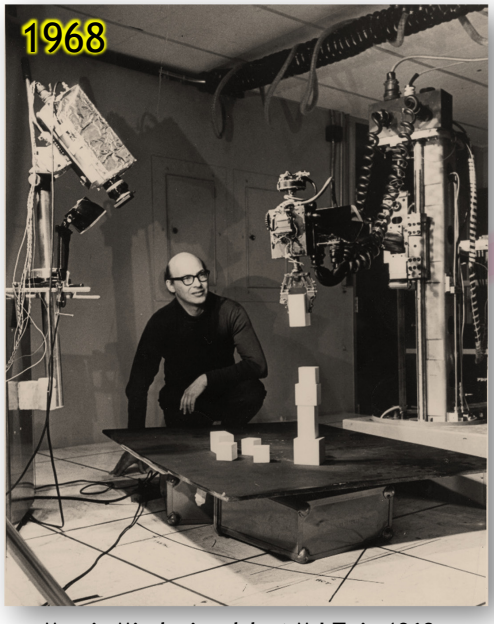
定型＋臨機応変

- 物流・配送
- 認識・ハンドともに多品種対応必須
- 経済性と性能要求

高自律性

- 生活支援
- タスク生成，エラーリカバリ，人間との対話必須
- 保守性・経済性もネック

まとめ　ロボットビジョンのさらなる進化に向けて



Marvin Minsky in a lab at M.I.T. in 1968.
出典：The New York Times, Jan. 25, 2016.

20XX
お茶をもって
きてくれる
ロボット