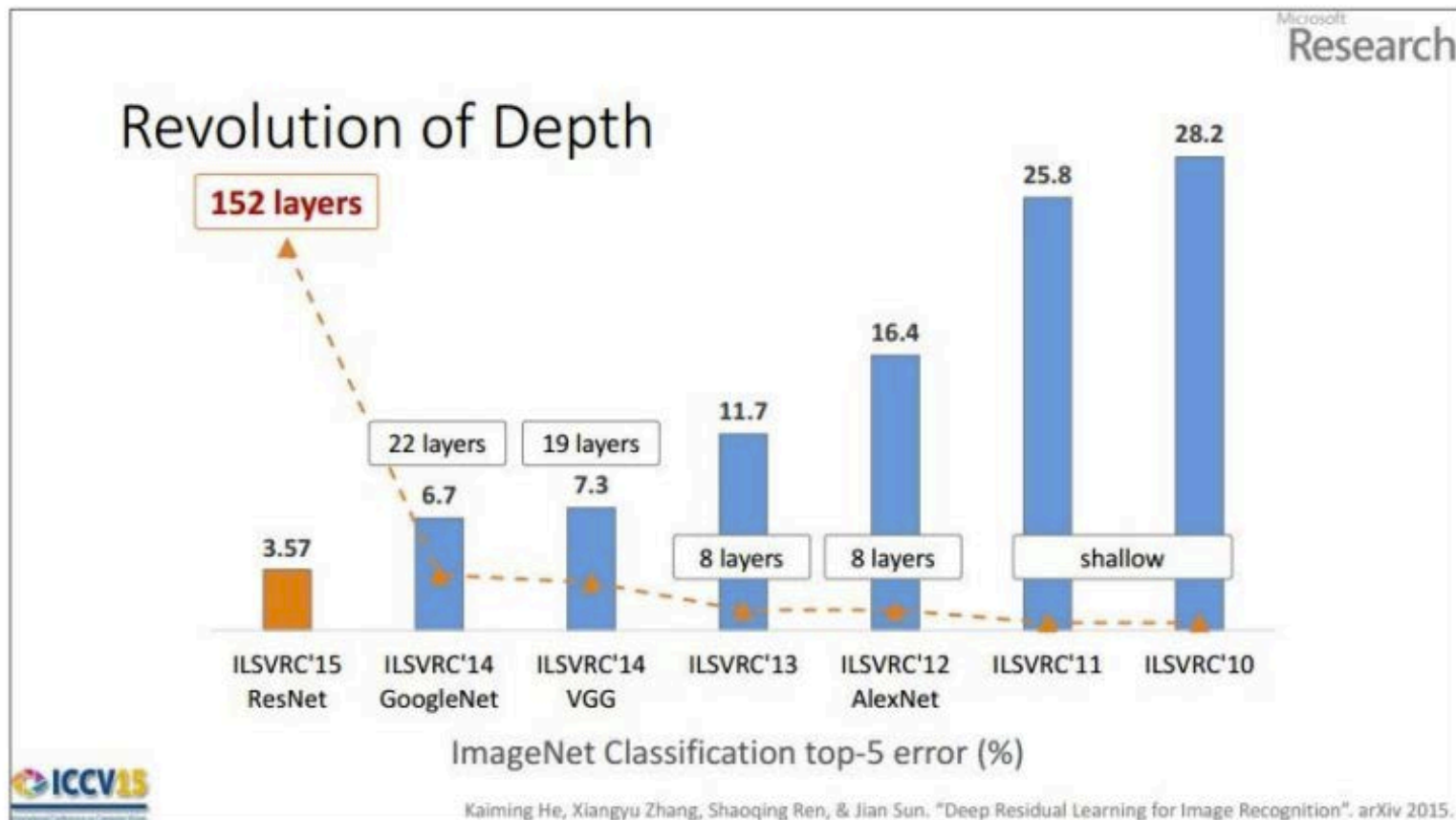


DL 産業論

松尾 豊

ディープラーニングによる画像認識の性能の急激な伸び



(slide from Kaiming He's recent presentation)

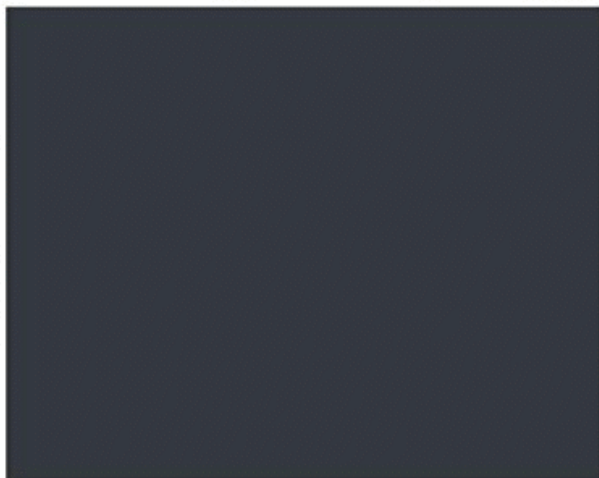
性能が向上し、層がますます深くなっている。
2015年には人間を超えた。Human Error ~5.1%。

未来の画像を予測する

Actual

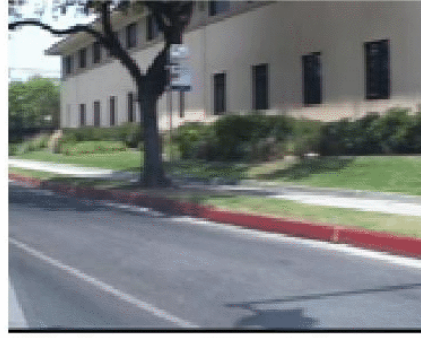


Predicted

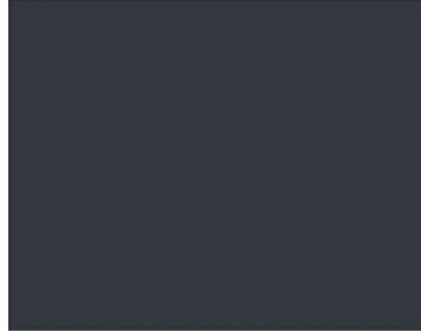
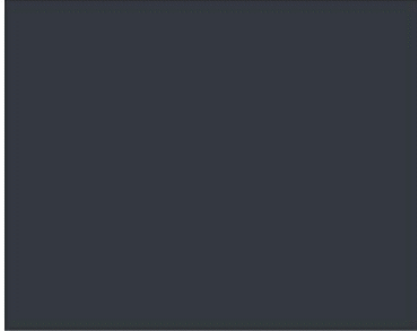
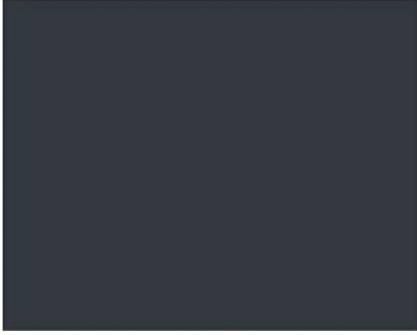


- KITTIデータセット。41,000のフレームから学習。
- 次のフレームを予測する。10フレーム=1秒。4レイヤーのモデル。

Actual



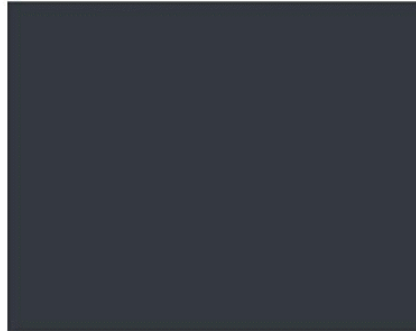
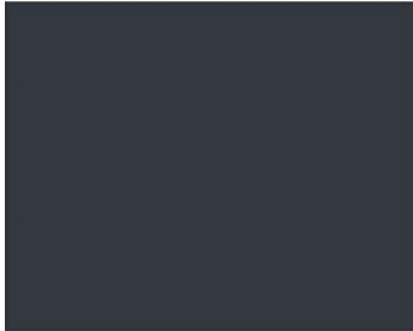
Predicted



Actual



Predicted





- より先の未来を予測。5フレーム(0.5秒後)になると、ややぼやけてくる。

静止画だけを与えて、1秒後を予測する

- ラベルなしの動画から、動画認識(行動分類)と動画生成(未来予測)の両方に使えるシーンダイナミクスを学習する。
- CNNによる時空間の畳み込み+GANを使って、背景と前景を切り分ける。
- 1秒までの短いビデオをフルフレームで生成できる。

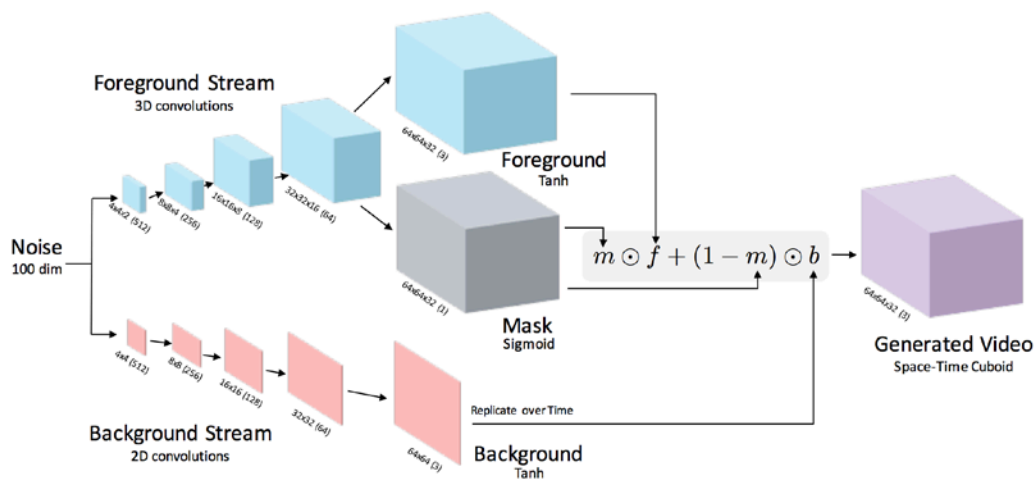


Figure 1: **Video Generator Network:** We illustrate our network architecture for the generator. The input is 100 dimensional (Gaussian noise). There are two independent streams: a moving foreground pathway of fractionally-strided spatio-temporal convolutions, and a static background pathway of fractionally-strided spatial convolutions, both of which up-sample to create the generated video using a mask from the motion network and the number of channels in parenthesis.

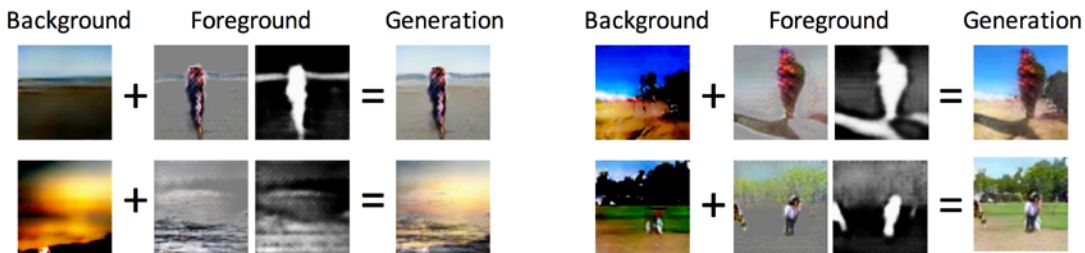


Figure 3: **Streams:** We visualize the background, foreground, and masks for beaches (left) and golf (right). The network generally learns to disentangle the foreground from the background.

Beach



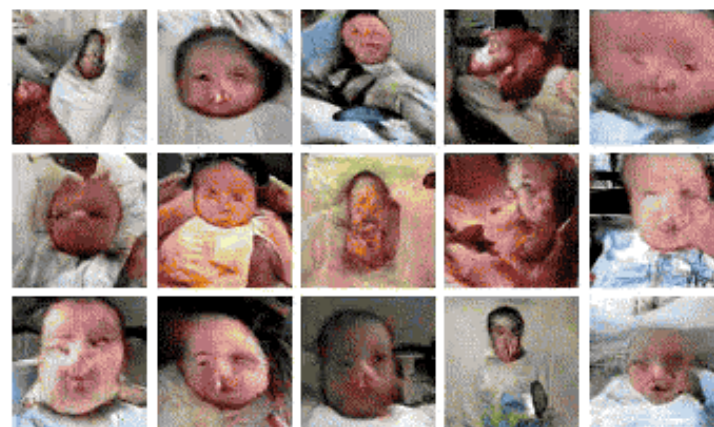
Golf



Train Station



Baby



- 最初のフレームだけ与え、次の1秒を生成することができる。
- 200万の動画(Flickr)から5000時間分のデータで学習。

動作とその帰結の予測



- Unsupervised Learning for Physical Interaction through Video Prediction (2016)
- 5万本のロボットのインタラクション(押す動作)の動画でも学習。
- CNNに、ロボットのアクションを途中で入れて、見える画像を予測する。

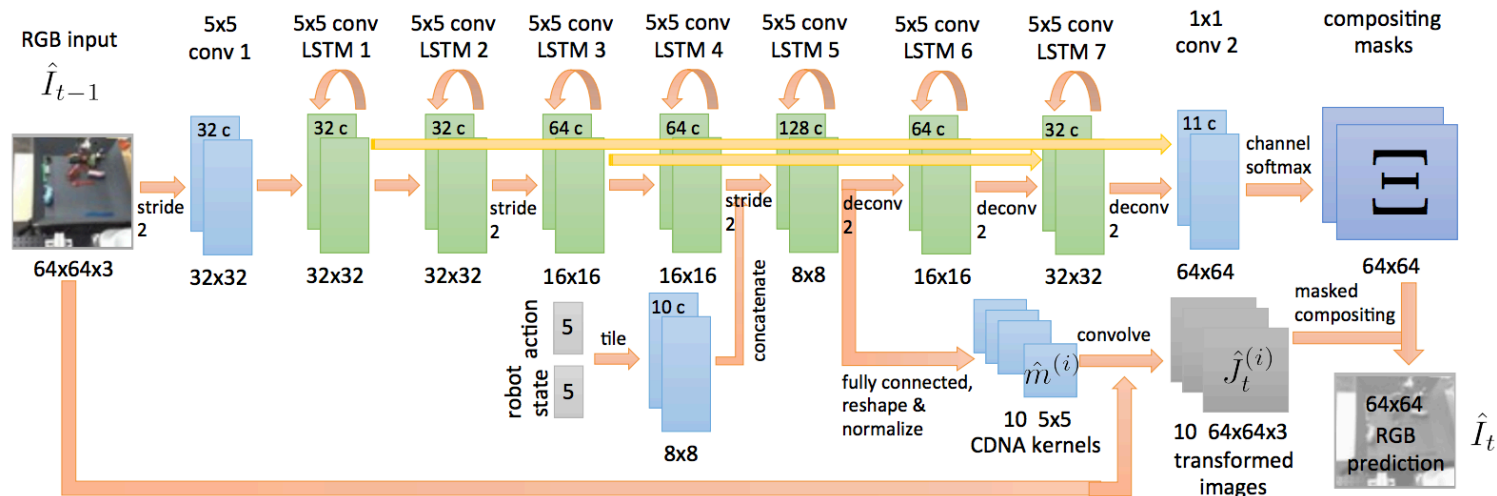


Figure 1: Architecture of the CDNA model, one of the three proposed pixel advection models. We use convolutional LSTMs to process the image, outputting 10 normalized transformation kernels from the smallest middle layer of the network and an 11-channel compositing mask from the last layer (including 1 channel for static background). The kernels are applied to transform the previous image into 10 different transformed images, which are then composited according to the masks. The

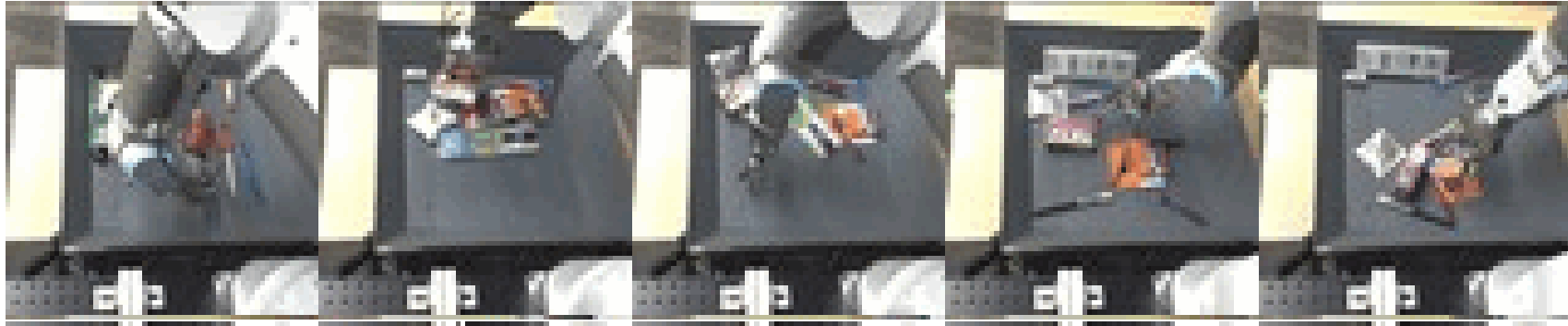
True



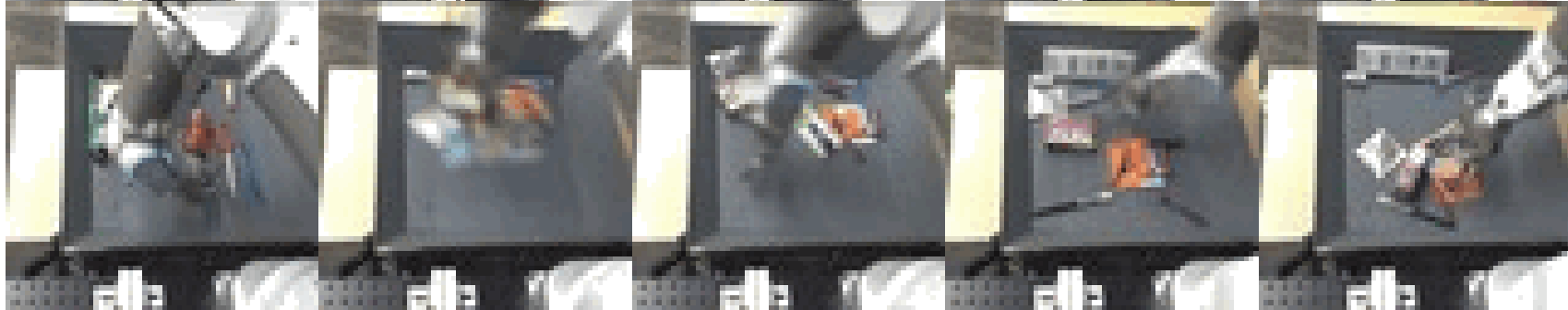
Predicted



True



Predicted

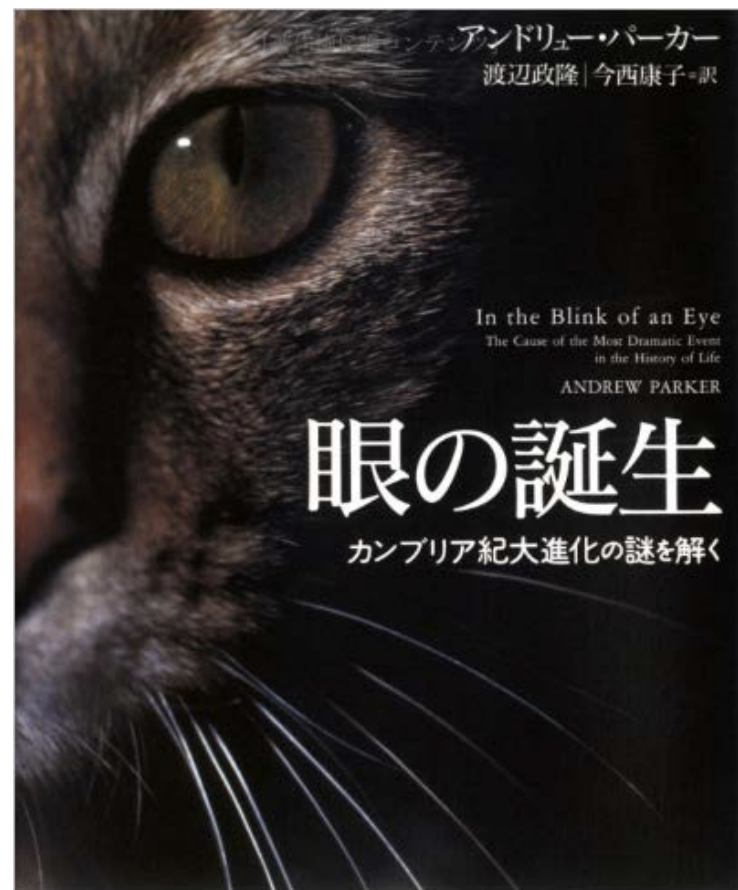


眼の誕生

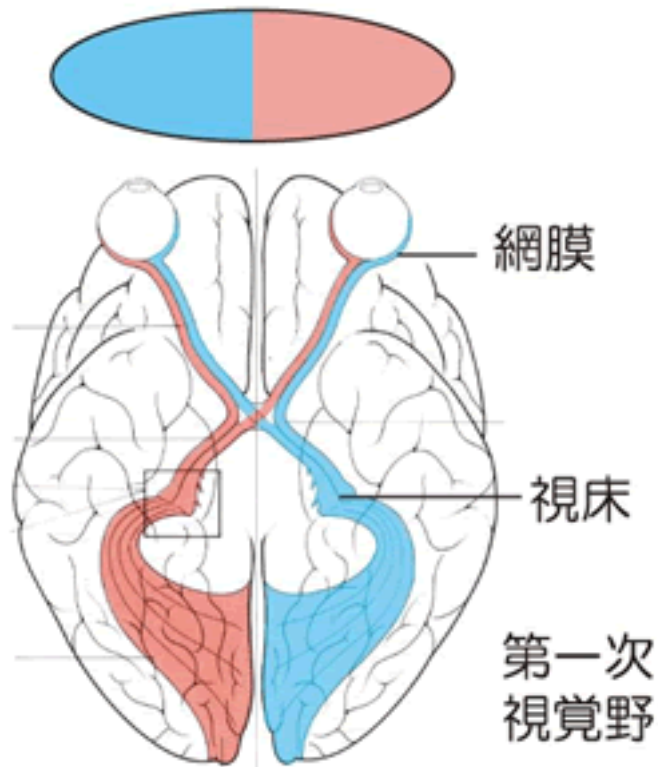
- カンブリア爆発
 - 5億4200万年前から5億3000万年前の間に突如として今日見られる動物の「門」が出そろった現象
 - 古生物学者アンドリュー・パーカーは、「眼の誕生」がその原因だったの説を提唱
- ディープラーニングにより、見えるようになる
 - さらに、次に何が起こるかを予想して動けるようになる。
- 「眼をもった機械」が誕生する。
 - 機械・ロボットの世界でのカンブリア爆発が起こる。
 - これを日本企業が取れるか？



三葉虫: 史上初めて眼をもった生物



眼のしくみ



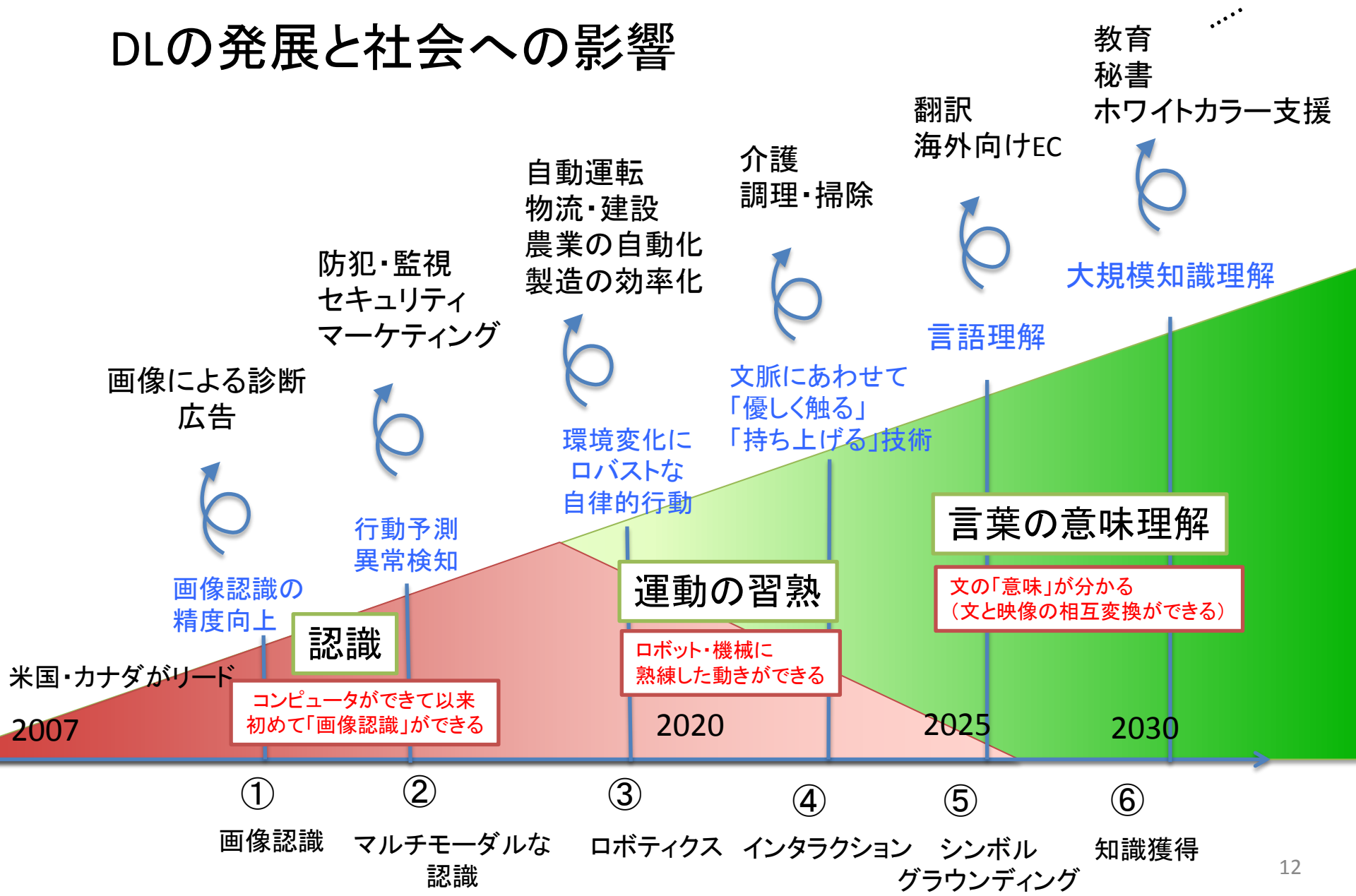
←イメージセンサ

←ディープラーニング
(CNN: 畳み込みニューラルネットワーク)

次に何が起こるかの予測も。

?

DLの発展と社会への影響



認識

運動の習熟

言葉の意味理解

コンピュータができて以来初めて「画像認識」ができる

ロボット・機械に熟練した動きができる

文の「意味」が分かる (文と映像の相互変換ができる)

①

②

③

④

⑤

⑥

画像認識

マルチモーダルな認識

ロボティクス

インタラクション

シンボルグラウンディング

知識獲得

Deep Learningをベースとする技術の進展

既存産業の発展

農業

収穫判定

トラクター、コンバインの
適用範囲拡大、効率向上
選別調製等の自動化

自動での収穫
自動での耕うん

建設

測量

掘削、基礎工事、
外装内装作業等の
効率向上

多くの作業の
自動化・効率化

食品
加工

振り分け
確認

カット、皮むき等
の自動化
食洗機に入れる

多くの加工工程の
自動化

組み立て
加工

目視確認の
自動化

動作効率の向上

段取りの自動化
セル生産の自動化

⋮

A: 画像認識

B: 運動の習熟

C: 計画立案を伴う運動

眼をもった機械・ロボットの典型例 単独の製品から入る

- 農業: トマト収穫ロボット
 - トマトは市場規模も大きく、収穫の工数も大きい。
 - 現状の技術で、トマトの認識ができる。上手にもぎ取ることも可能。
 - 先進的な農場から試しに入れる。
- 建設: 自動溶接機械
 - 建設の工程(例えば溶接)を自動化する
 - 現状の技術で、接合面の状態等の認識ができる。上手に溶接することも可能。機械を当てれば熟練した人でなくとも熟練の人のような溶接ができる。
 - 一部の建設現場で試しに入れる。
- 食品加工: 食洗機にお皿を入れるロボット
 - 食品加工に関わる仕事、まずは食洗機にお皿を入れることを自動化する
 - 現状の技術で、お皿の位置、把持位置の認識ができる。まずは、食器が下げられたところから、食洗機に入れるところを自動化する。(混雑時に重要)
 - ファミリーレストラン等の一部の店舗で試しに入れる。

製品を一刻も早く市場に投入する

眼をもった機械・ロボットの典型例 サービス化へ

- 農業: トマト収穫ロボット(異常発見機能つき)
 - トマトの収穫作業に使いながらデータをためる。(ネットワークに接続)
 - 使われる農場数を増やすとともに、コスト削減。
 - 「病気の判定」ができるようになる。それに対して、新たにチャージすることができる。
- 建設: 移動型の自動溶接機械
 - 溶接作業に使いながらデータをためる。(ネットワークに接続)
 - 使われる現場の数を増やすとともに、コスト削減。
 - そのうち、移動しながらの溶接作業や、すでに終わった溶接のチェックができるようになる。それに対して新たにチャージできる。
- 食品加工: お皿管理ロボット
 - お皿を入れながらデータをためる。(ネットワークに接続)
 - 使われる店舗の数を増やすとともに、コスト削減。
 - そのうち、残飯の処理、食器を棚に戻す、食器に盛り付ける等が可能になってくる。それに対して新たにチャージできる。

製品をネットワークにつなぎ、ソフト面の機能向上とともにサービス課金へ

眼をもった機械・ロボットの典型例 プラットフォーム化へ

- 農業: トマト栽培管理プラットフォーム
 - トマトの水やり、施肥、害虫駆除等を自動でできるようになる。
 - トマトの収穫全体を管理するプラットフォームを構築できる。このプラットフォームを利用すれば、品質のいいトマトがたくさん取れるというビジネス。
- 建設: 建設現場(躯体工事)プラットフォーム
 - 溶接作業を中心にして、鉄筋を組む作業、コンクリートを入れる作業などを次々に自動化していく。
 - 建設現場の作業全体のプラットフォームを構築できる。プラットフォームを利用すると、自動化された建設現場の施工管理ができるというビジネス
- 食品加工: 調理プラットフォーム
 - 食器の管理から食材管理、調理全般を行うまで自動化していく。
 - 調理全体のプラットフォームを構築できる。プラットフォームを利用すると、店舗の応じた調理が自動で提供されるというビジネス

その「場」の機能全体をプラットフォームとして提供

眼をもった機械・ロボットの典型例 海外展開へ

- 農業: グローバルなトマト栽培管理プラットフォーム
 - プラットフォームをそのまま海外に展開し、海外の事業者から課金する。
 - 日本の品質のトマトを、欧米の食卓に届けることができる。
 - 日本はきめ細やかな栽培方法を工夫し続けることで競争力を維持。
- 建設: グローバルな建設現場(躯体工事)プラットフォーム
 - プラットフォームをそのまま海外に展開し、海外の事業者から課金する。
 - 日本のレベルの高い施工技術を、機械化したプラットフォームで提供できる。
 - 日本は建物の耐久性や省エネ、デザイン等、付加価値の高い工夫を続けることで競争力を維持。
- 食品加工: グローバルな調理プラットフォーム
 - プラットフォームをそのまま海外に展開し、海外の事業者から課金する。
 - 日本のレベルの高い食を、世界の飲食店で提供できる。
 - 日本は高い食のレベルを維持し続ける。おいしいものを開発し続ける。

プラットフォームを世界に展開し、高い収益を維持する

モノ売りからの脱却へ

- 「眼のある機械」は、データの継続的収集が不可欠
 - 製品からデータが戻るようにしないと、継続的な品質向上につながらない
 - つまり、製品がネットワークに接続されることがほぼ確定している
- すると、眼のある機械の「稼働」に対して課金できるようになる
 - 「学習ずみモデル」の品質が上がれば、価格を上げることができる
 - 内部コストを下げれば、利益を上げることができる
 - モノ売りからサービス売りへの転換が容易にできる
- さらに、眼のある機械を起点とする「場」全体のプラットフォーム化へ
 - 製品が置かれるオフィス、家、商業施設、工場、農場、建設現場など、製品が取得するデータ・提供するサービスを起点として、その周りのお金・情報の流れに広げ、事業チャンスをとっていくことができる。
- それを世界展開し、日本品質でサービスを提供する
 - 「学習ずみモデル」は日本で作り続け、競争力を維持し続ける

眼のある機械の市場投入→サービス化→周辺を含んだプラットフォーム化
→海外へ大きく展開という流れが王道

日本なりのプラットフォーム戦略

- DLの技術はコモディティ化する。
 - 競争力をもつのは、データとハードウェア。
 - 早くDLの技術を取り入れてしまえばよい。
 - DLの技術とハードウェアのすりあわせになった瞬間、日本企業が再度、力を取り戻せる。
- 欧米のスタートアップ(とDL研究者)は、意外なほどハードウェアに対する抵抗感がある
 - そもそも、産業用ロボットの導入台数は日本が(ほぼ)トップ
 - また、ロボットに対する社会的抵抗感もある。
 - 米国は雇用を守らないといけない。日本は人手が足りない。
- ものが関連しないプラットフォームは無理
 - 英語圏でやったほうが絶対に強い。
 - 広告費規模でも10倍、ECの規模でも3倍以上
- 日本には、検索エンジンもECもSNSもあった。
 - GoogleやAmazon, Facebookの位置の企業を出せなくはなかった。
 - しかし、結果はそうになってない。原因は明確で、英語圏でなかったから。

ものづくりを起点に眼をもった機械を作り、
プラットフォーム化するのは日本ならではの戦略

機械・ロボットのカンブリア爆発

- 介護施設や病院等での見守り・介護ロボット
- 医療（X線、CT、皮膚、心電図、手術ロボット）
- 警備、防犯技術
- 顔による認証・ログイン・広告技術、表情読み取り技術（サービス業全般に重要）
- 国家の安全保障、入国管理、警察業務、輸出入管理業務における活用
- 防災系（河川、火山、土砂崩れを見張る）
- 重機系（掘削、揚重）、建設現場系（セメント固め、溶接、運搬、取り付け）
- 農業系（収穫、選果、防除、摘花・摘果）
- 自動操縦系（ドローン、小型運搬車、農機、建機）
- 自動運転系、物流
- 産業用ロボット系（特に組み立て加工等）
- 調理系（牛丼、炊飯、ファミリーレストラン、外食全般）
- ペットロボット系
- 片付けロボット（家庭、オフィス、商業施設）
- 新薬発見や新素材の開発（遺伝子の認識・分析、実験ロボット）
- 廃炉系（深海や鉱山、宇宙も含めた極限環境）

農業・建設・食品加工だけでなく、医療や介護、製造、廃炉なども。

DLに関わる海外ベンチャー

- Deep Mind (英) : DLの技術力をもった企業。DQNによるゲーム、アルファ碁、医療など。2011創業。Googleが2014に£ 400Mで買収。
- Enlitic : 医療画像(X線)におけるDL活用。2014創業、15M調達。
- Nervana Systems : 医療、農業、金融、自動車、エネルギー等における画像処理。24M調達後、インテルが2016買収。
- Emotient : 顔の表情を認識する会社。2012創業、6M調達後、Appleが2016買収。
- Affectiva : 映画やTV番組のどこで表情が変わったのかを読み取る。2009創業、34M調達。
- Perceptio : DLによる写真分類アプリ開発。創業、調達額不明。Appleが2015買収。
- VocalIQ (英) : DLによる音声認識。1M調達後、Appleが2015買収。
- Atomwise : ドラッグディスカバリーへのDL活用。新薬の候補物質を見つける。YC卒業生。2012創業。6M調達。
- Descartes Labs : DLによる衛星画像の分析。農業への適用。2014創業、8M調達。
- Canary : DLによるホームセキュリティ。2012創業。41M調達。
- Netatmo : 家電。DLによる監視カメラも。2011創業。38M調達。
- Pilot AI Labs : DLの画像認識を使ったドローンの自動操縦。まだ小さいが、動画が面白い。
- MetaMind : 画像認識一般。2014創業、8M調達。
- SkyMind : JavaベースのDL提供。2014創業。3M調達。
- AlchemyAPI : DLによる言語処理と画像認識。クラウドで提供。2005創業、2M調達。IBMが2015に買収。
- ZenRobotics (フィンランド) : ごみの選別ロボット。2007創業、17M調達。

急がないといけない

DLに関わる海外大手企業

- 先導するネット・IT系企業

- Google: J. Hinton氏の引き抜きからDeepMindの買収まで抜かりがない。研究開発力も最強。
- Facebook: Y. LeCun氏を引き抜き、FAIR (Facebook AI Research) はDL技術で強い。
- Microsoft: 学術研究にも力を入れ、有力研究者が多数。画像認識のコンペでも優勝。
- Apple: DLのベンチャー企業を何社か買収。
- Amazon: ピッキングチャレンジ開催。デルフト大(蘭)優勝。
- IBM: ワトソン。DL技術もプラットフォームに加える。
- Baidu(中): 世界2位の検索エンジン。A. Ng氏率いる強い研究チーム。DLによる音声認識、広告、画像のキャプションつけ等。

- 製造業

- GE: DLによる医療画像の診断。
- Dyson: 掃除機に眼をつけたものを開発。インペリアル・カレッジにRoboticsラボ設立。
- Kuka(独): 産業用ロボットへのDL適用
- Mobileye(蘭): 車用の画像認識を提供。1999イスラエルで創業。2014年上場。時価総額10B。
- LG(韓): インチョン空港で、DLを使ったロボットでの案内の実験
- Samsung(韓): DLを使った胸部エコー検診の医療機器

急がないといけない

必要になる「学習工場」

- 学習工場のイメージ
 - 機械学習を使える高度な人材
 - 高性能な計算機
 - データを準備する環境
- 学習工場で出荷されるもの
 - 「学習ずみのモデル」が作られる。視覚野の部分。
 - これが最終製品に載せて出荷される。
- 実例
 - トヨタやFanucに対してのPFNがそれに近い
 - それを、各産業分野、各企業ごとに持つ必要がある。
 - ベンチャーでも社内組織でもよいが、社内人事との整合性を考えるとベンチャーのほうが有望
- 「工場」なので、数十億～数千億円(数兆円)の投資規模
 - R&Dではない。最終製品を作るものである。
 - 10倍生産したいのであれば10倍の規模に。
 - この規模でようやく諸外国と勝算ある状態で戦える
- これを新たな設備投資の形とすることができないか。
 - 国内企業には、内部留保が大変多い。
 - 何に投資していいのかわからない状態。

それぞれで必要なもの

- 学習データ製造
 - アノテーション付きのデータをたくさん作り出す事業者(2次サプライヤー)
 - そこには大量の雇用が発生
 - 3次サプライヤー
 - そのためのツールを開発する事業者
 - そのためにクラウドソーシングを活用する事業者
 - 模擬データを出すためのシミュレータを開発する事業者
 - 極限環境でのデータをとる事業者
- 計算機・開発環境
 - HPC:スパコン? 海外が強い
 - ライブラリ開発: 海外が強い
 - アルゴリズム開発: 海外が強い
 - ここは本来は、国の研究所(産総研・理研)が頑張るべきところだが、戦いは厳しいので、ひとまず輸入モデルをしっかりと。あまり国産にこだわっても良くない。
- 学習職人(高度なスキル・知識をもった人材)
 - いま、これが圧倒的に足りない。
 - 大学で人材育成を急ぐ。
 - 民間の事業者でも、研修プログラムの提供、資格制度作りを急ぐべき。

大量の雇用、
たくさんの事業者を生む
産業としての広がり大きい

しっかり輸入
(いずれ独自の競争力)

ハイスペック人材

国としてやるとよいこと

- 眼のある機械からプラットフォームを作り、海外展開に至るまでのそれぞれの産業分野でのロードマップ作り
 - 特に、最初の入口はコストとの勝負。リアリティを持って。
 - 最初の段階から海外でやったほうがよいものも多いはず。
- いくつかの分野での先行事例作り
 - すでに水面下ではいろいろと動いているものをしっかり調査
 - 先進的な事例を作りに行く
 - 多くの企業に分かりやすく実例を示す
- 人材の育成とVC機能をセットにした学習工場づくり
 - 各産業分野で多数の学習工場が必要
 - トヨタやFanucに対するPFNというのがひとつの例。こうしたものが10倍～100倍のスケールで出てくる必要がある
 - それを促進するための制度や仕組み作り