

AIと自動運転の展望と課題



Toyota Technological Institute at Chicago 学長

東京工業大学名誉教授

古井 貞熙

furui@ttic.edu

講演概要

- TTIC (Toyota Technological Institute at Chicago)
- 深層学習(DNN)による人工知能(AI)の急速な進歩
 - 音声認識・画像認識の例
- AIの最近の進歩の特徴
- スマートロボット
- 自動運転のインパクトと課題
- AIの今後の展望
- まとめ

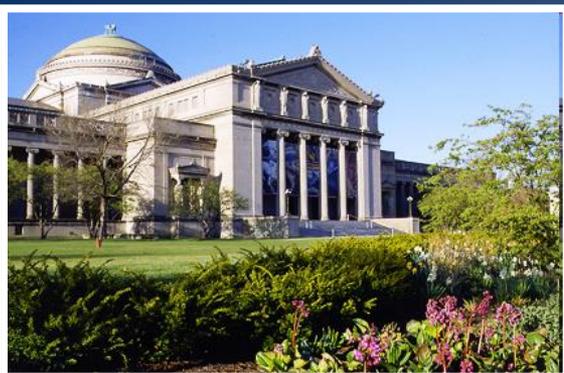
TOYOTA TECHNOLOGICAL INSTITUTE AT CHICAGO



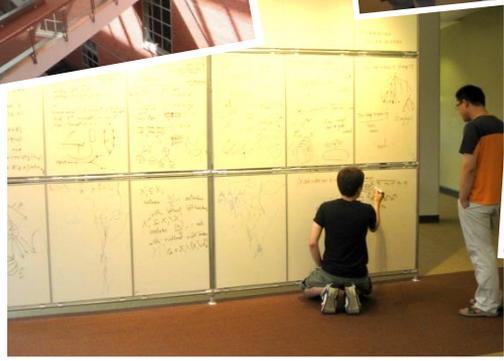
TTIC (Toyota Technological Institute at Chicago)

- **ミッション**: コンピュータ科学の基礎分野及び情報技術の分野、特に人工知能(機械学習)における研究と教育活動を通じ、人類に貢献する
- **トヨタ自動車**の寄付(当初\$105M、現在\$200M、2017年には\$265Mの予定)により、豊田工大とシカゴ大学が連携して、2001年10月に設立、2003年秋より開学
- **場所**: イリノイ州シカゴ市ハイドパーク内
(シカゴ大学キャンパス内)
- **現員**: 教員 (Full time) 23 事務職員 7
博士課程学生 28 インターン (20)

Hyde Park



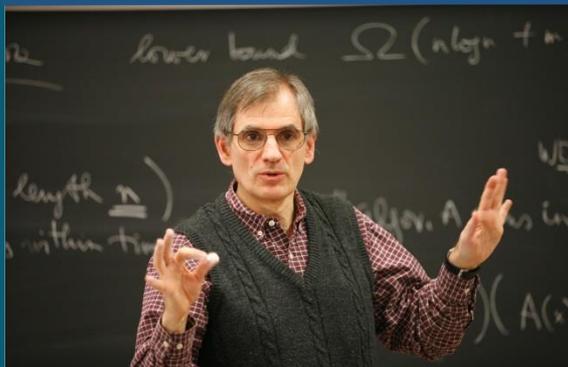
TTIC



世界トップレベルの教員が、次世代の コンピュータ科学と情報技術のリー ダーを教育している



10 tenured and tenure-track faculty



TTICの学生は授業料が無料で、生活費が支給される

TTICは、世界トップレベルのシカゴ大学のComputer Science学科と、授業の相互提供を行っており、生物学科、統計学科、数学科、ビジネススクールなどとも連携している

TTICの教員と学生は、シカゴ大学の種々のシステムを使うことができる:

- 図書館
 - 運動施設
 - カフェテリア
 - キャンパス内バス
 - 学生寮
 - 健康管理センター
 - キャリアセンター
 - その他
-

人工知能(AI: Artificial Intelligence)の研究

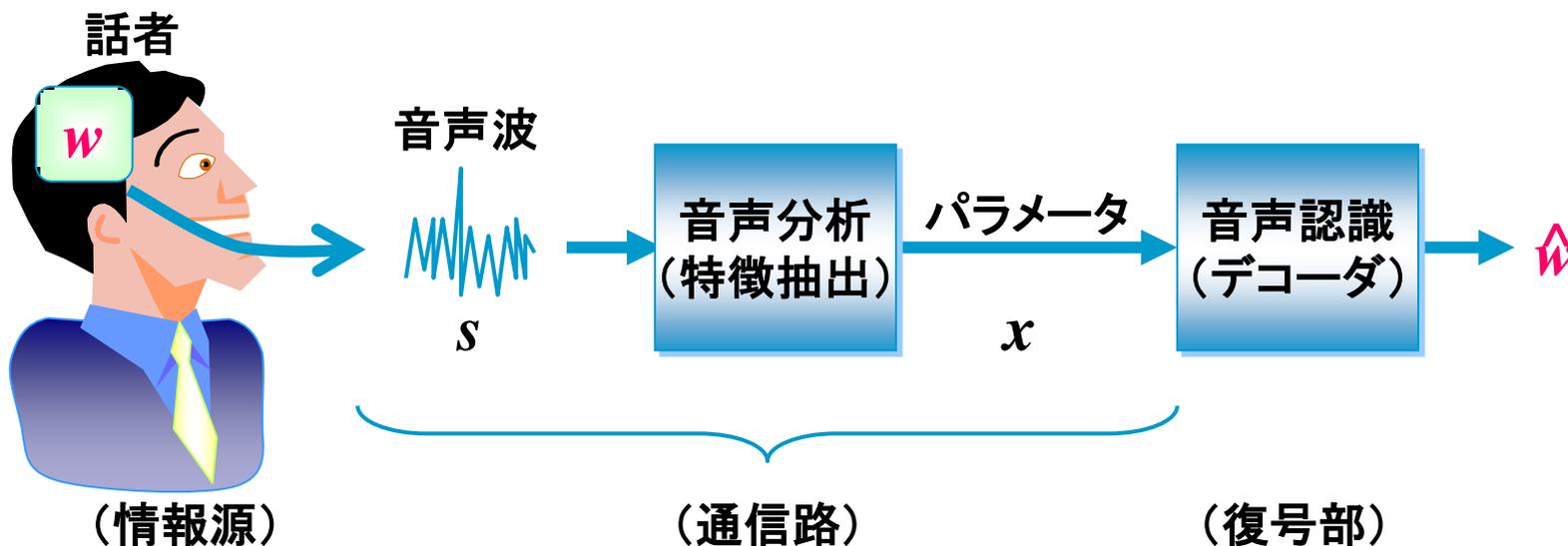
- 機械学習の理論

- 統計的学習理論
- 深層学習(DNN: Deep Neural Network)

- 機械学習理論の応用

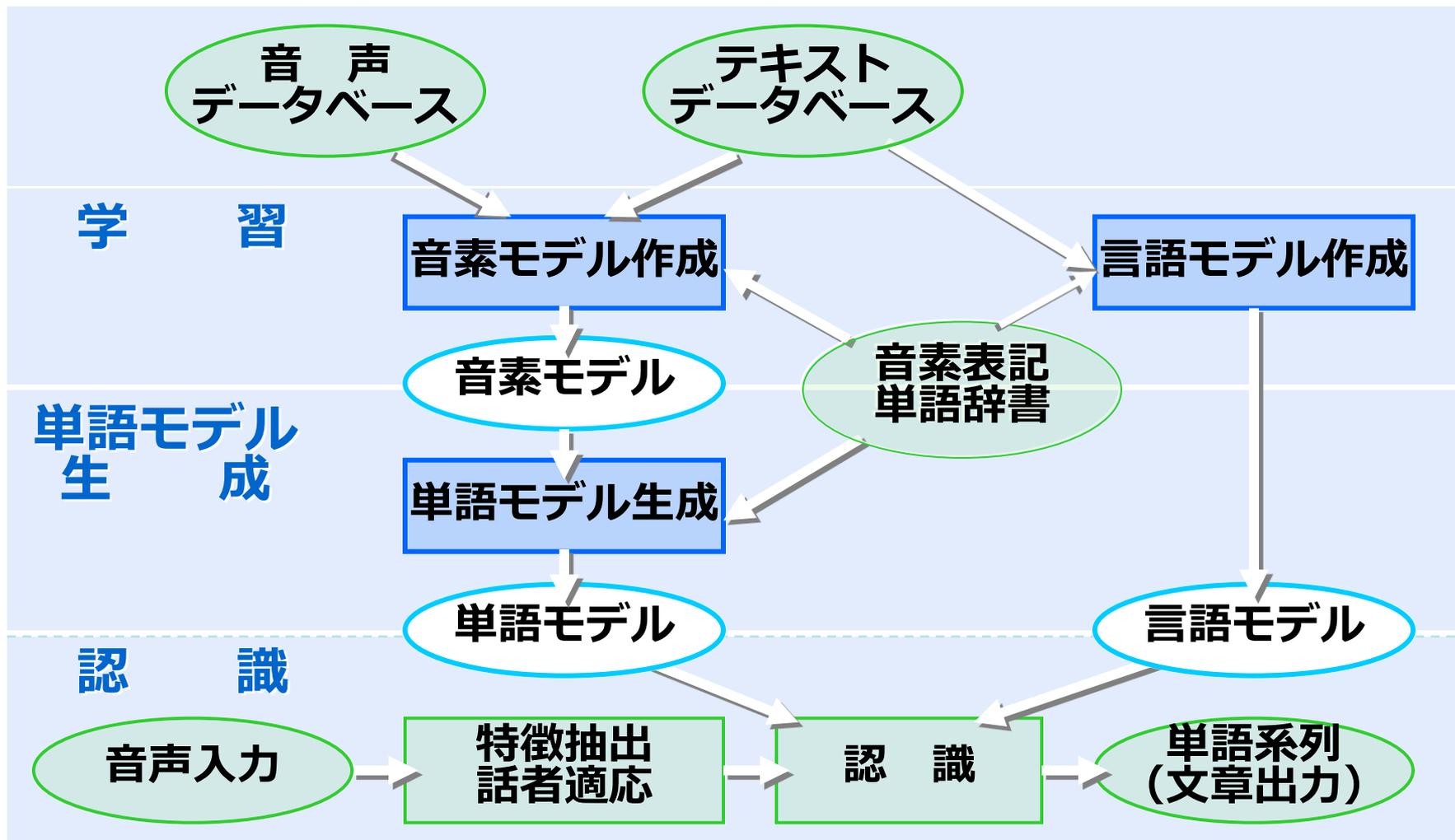
- 自然言語処理(情報検索・質問応答・自動翻訳など)
- 音声処理(音声認識・話者認識・音声合成など)
- 映像・画像処理
- ロボット
- 車の自動運転、安全運転
- 計算生物学(遺伝子解析・創薬)
- 病気の早期発見や予防
- 投資
- その他、種々のビジネスへの適用など

確率モデルに基づく音声認識の原理 (w : 単語列)

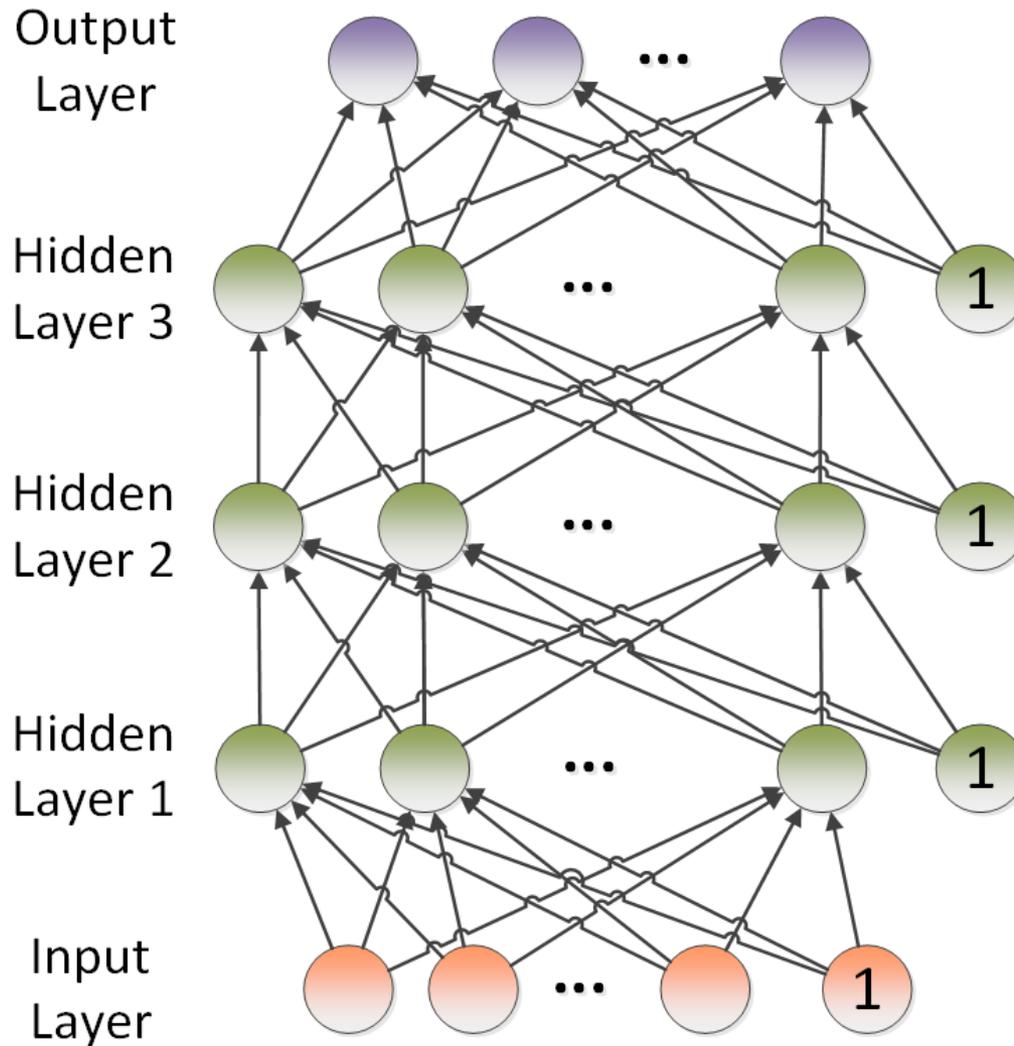


$$\hat{w} = \operatorname{argmax}_w P(w | x) = \operatorname{argmax}_w P(x | w)P(w)$$

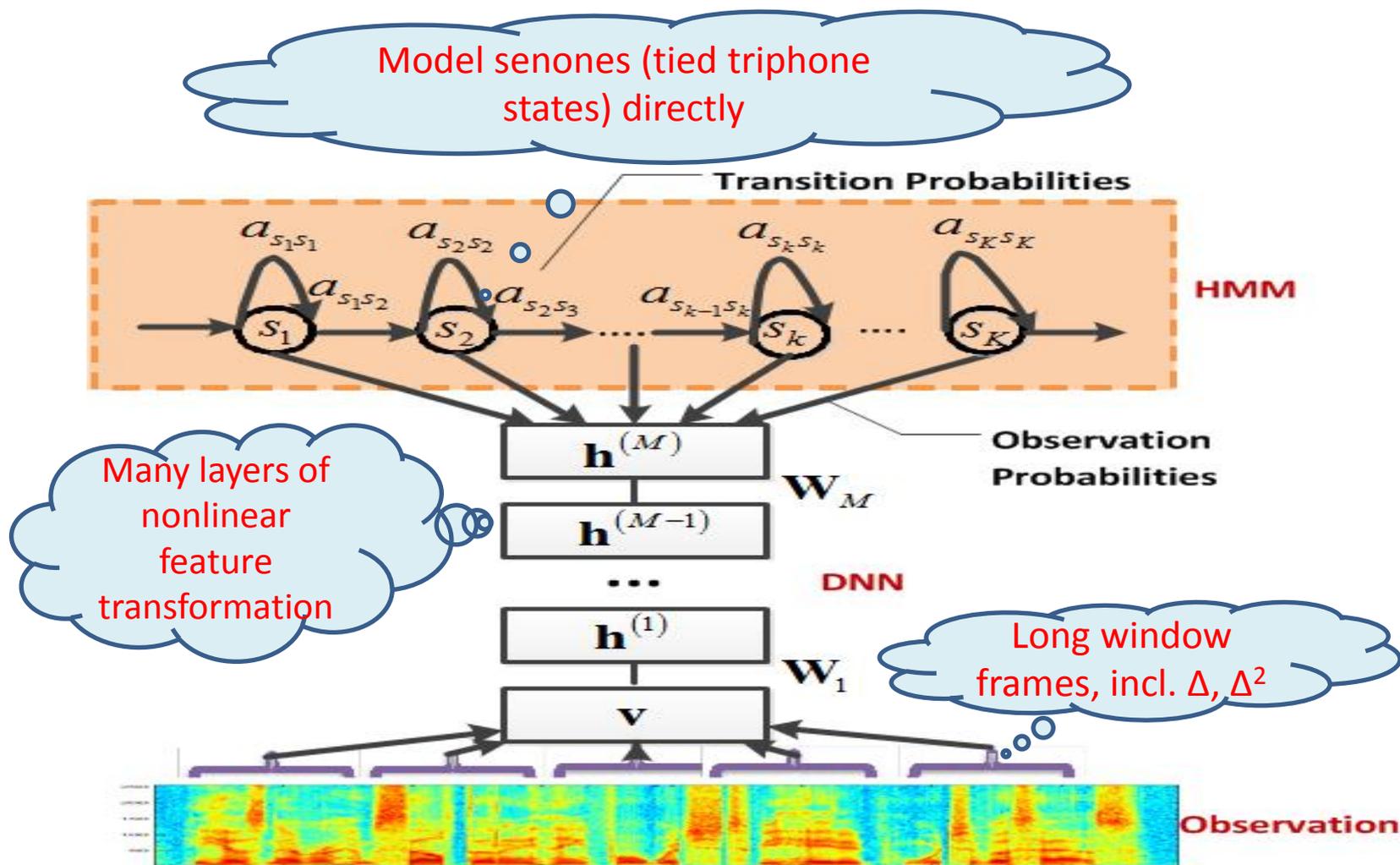
音声認識の方法



ディープニューラルネットワーク (DNN)



DNNによる音声認識



DNNによる音声認識の性能向上

(Dahl, Yu, Deng, Acero 2012, Seide, Li, Yu 2011 + new result)

- Voice Search SER (24 hours training)

| AM | Setup | Test |
|---------|-----------------|--------------|
| GMM-HMM | MPE | 36.2% |
| DNN-HMM | 5 layers x 2048 | 30.1% (-17%) |

- Switch Board WER (309 hours training)

| AM | Setup | Hub5'00-SWB | RT03S-FSH |
|---------|----------------------|--------------|--------------|
| GMM-HMM | BMMI (9K 40-mixture) | 23.6% | 27.4% |
| DNN-HMM | 7 x 2048 | 15.8% (-33%) | 18.5% (-33%) |

- Switch Board WER (2000 hours training)

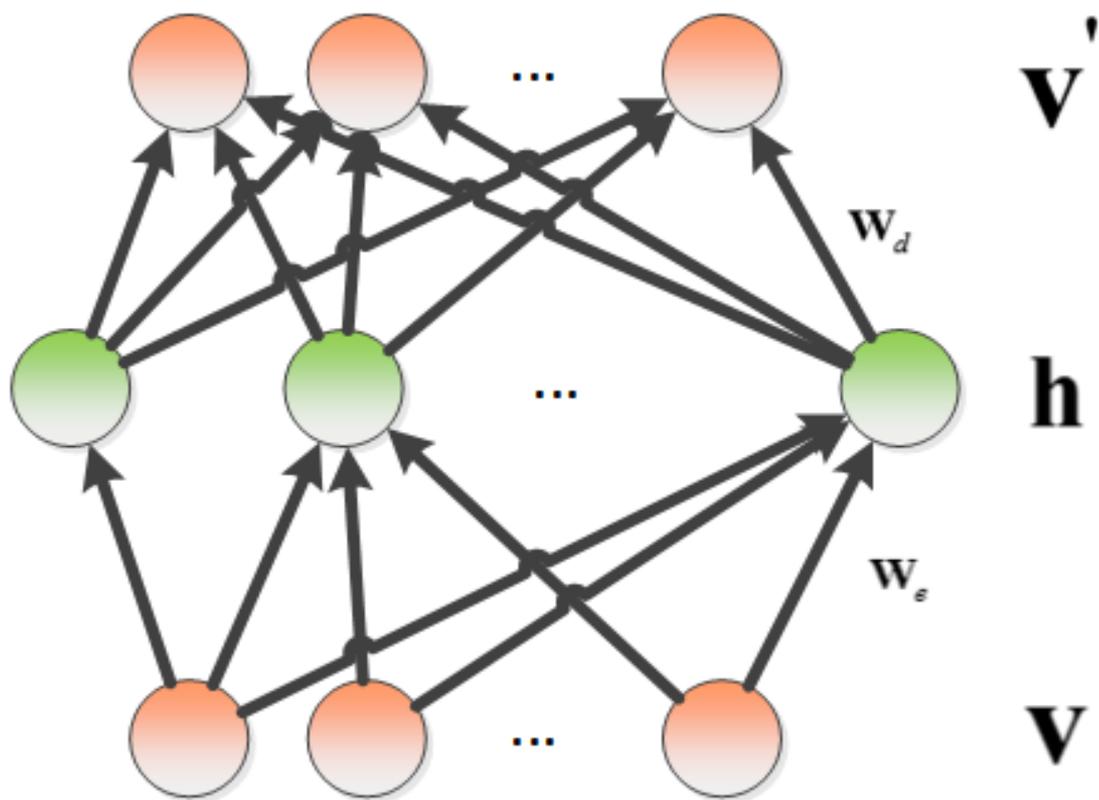
| AM | Setup | Hub5'00-SWB | RT03S-FSH |
|-------------|-----------------------|----------------------------|----------------------------|
| GMM-HMM (A) | BMMI (18K 72-mixture) | 21.7% | 23.0% |
| GMM-HMM (B) | BMMI + fMPE | 19.6% | 20.5% |
| DNN-HMM | 7 x 3076 | 14.4% (A: -34% B: -27%) | 15.6% (A: -32% B: -24%) |

(Dong Yu, 2012)

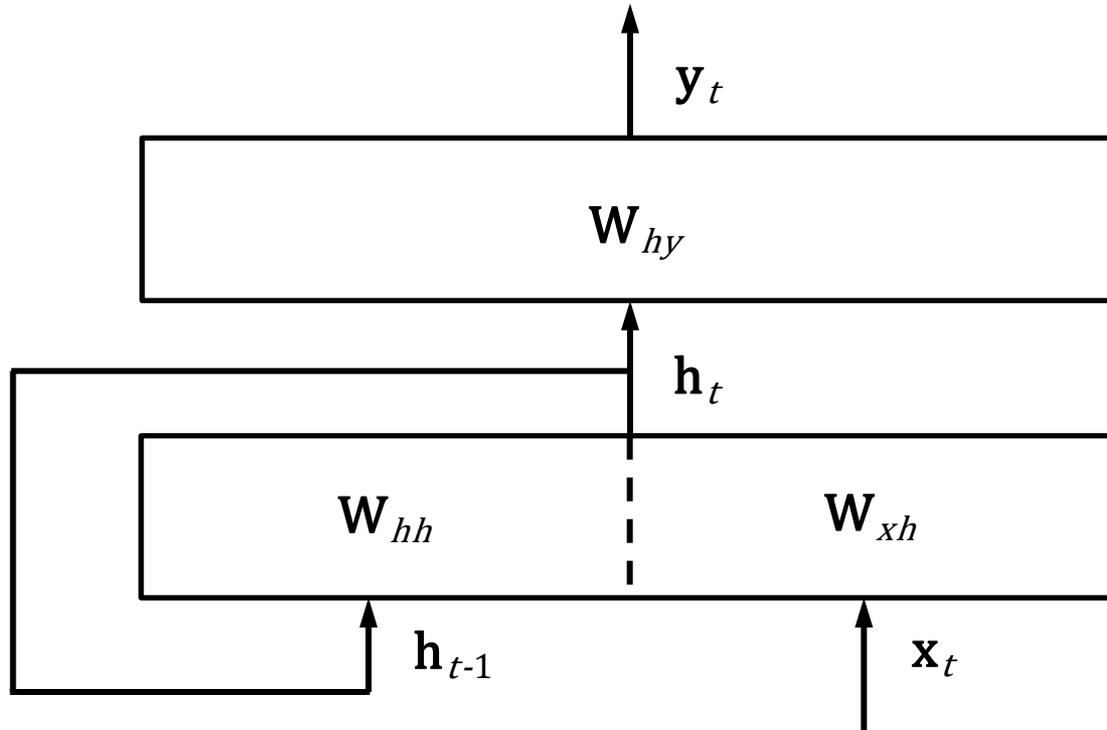
なぜDNNが有効なのか

- 単純な非線形ネットワークを多段に積み重ねることによって、**複雑な非線形処理**が実現できる。
- ビッグデータを用いることにより、入力から目標出力までのプロセスを、**全体として最適化**することができる。
- **表現学習**(representation learning)、すなわち与えられたデータのどこに注目し特徴量として表すかという特徴抽出方法の学習が、自動的にできる。
- 従来からの誤差逆伝搬(backpropagation)学習法に加えて、DropOutなどの新技術と、GPUを用いた並列計算により、**膨大なデータを用いた頑健(robust)な学習**ができるようになった。

自己符号化器 (Autoencoder)



再帰型ニューラルネットワーク (RNN: Recurrent NN)

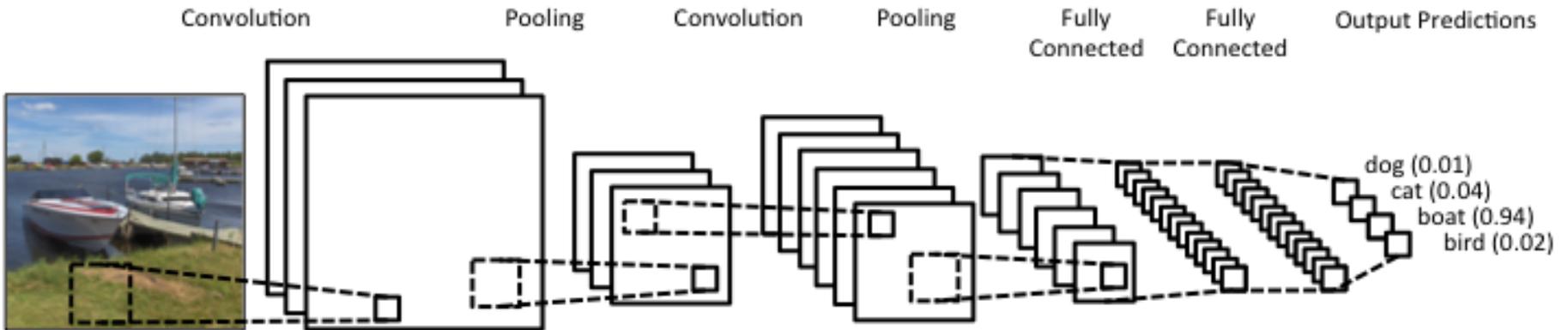


$$\mathbf{h}_t = \sigma(\mathbf{W}_{xh}\mathbf{x}_t + \mathbf{W}_{hh}\mathbf{h}_{t-1})$$

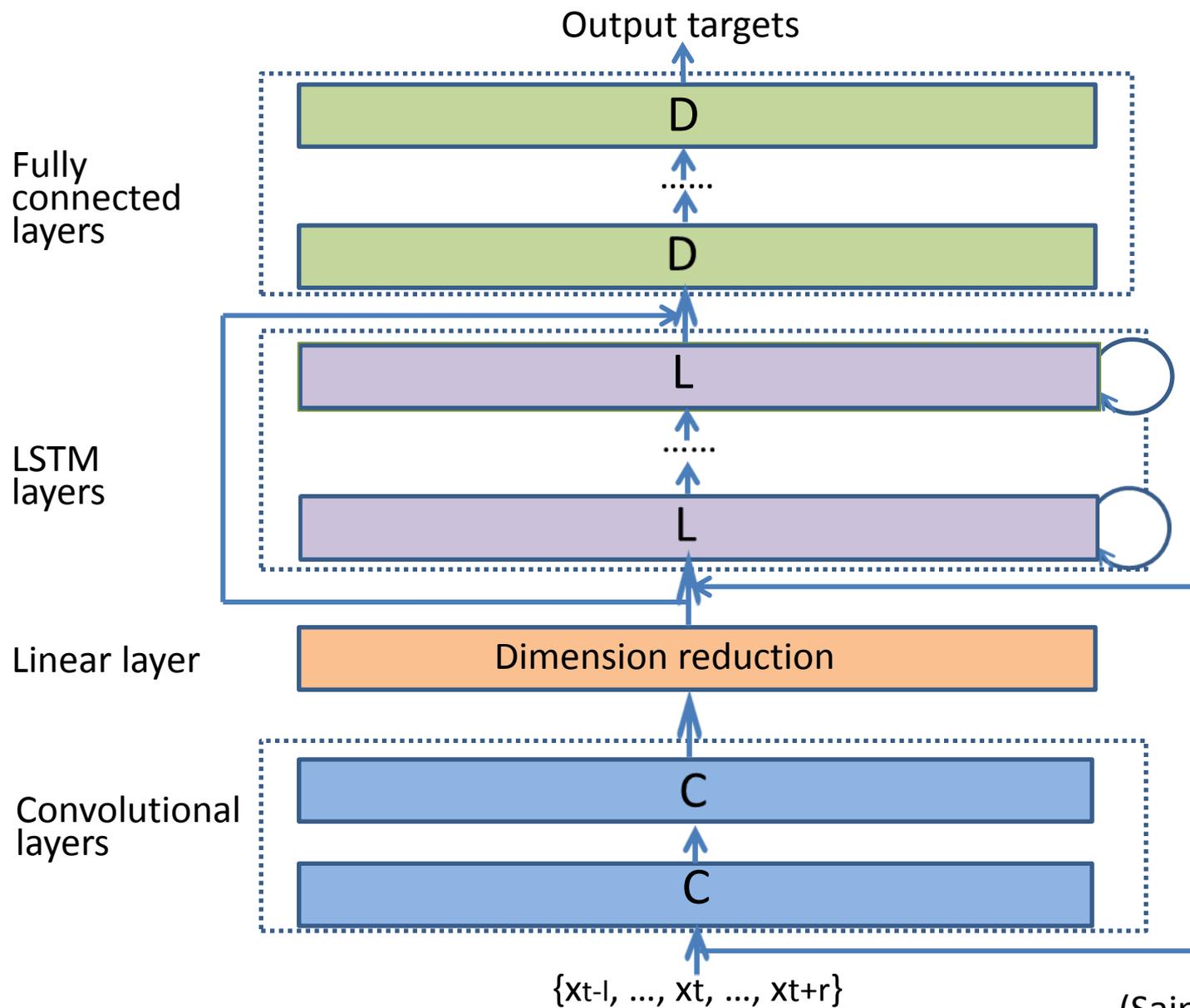
$$\mathbf{y}_t = \text{softmax}(\mathbf{W}_{hy}\mathbf{h}_t)$$

Bi-directional RNN, LSTM (Long Short-term Memory) RNN, Hierarchical RNN,
Deep RNN

畳み込みニューラルネットワーク (CNN: Convolutional NN)



CLDNN (Convolutional, LSTM-DNN)



(Sainath et al., 2015)

音声認識技術の現状と今後の展望

- **確率モデルと機械学習**による音声認識技術の進歩
 - 大規模データベース(「ビッグデータ」)と学習理論(特に「深層学習」)の融合
- 多様な**アプリケーション**が展開している
 - 音声対話(音声入力)システム
 - 音声文字変換
- 音声認識にはまだ**解決すべき課題**が多い
 - 雑音や声の個人差への適応
 - 話し言葉の音素モデルと言語モデル
 - 未知語の取り扱い
 - コンテキストや韻律情報(アクセント、イントネーション)の利用
 - 音声認識から音声理解、要約、情報抽出へ
 - マルチモーダル環境での音声の利用
 - 大規模データからの効果的学習法(能動、強化、教師あり・なし学習)

Progress of object recognition (1k ImageNet)



2012 - 2015

Super-deep: 152 layers



Announced
Dec.2015

DNNの一層の深層化

AlexNet, 8
layers
(ILSVRC 2012)



VGG, 19 layers
(ILSVRC 2014)



ResNet, 152
layers
(ILSVRC 2015)

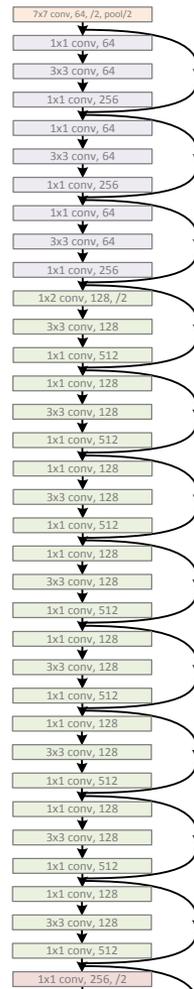


ILSVRC (Large Scale Visual Recognition Challenge)

(slide credit: Jian Sun, MSR)

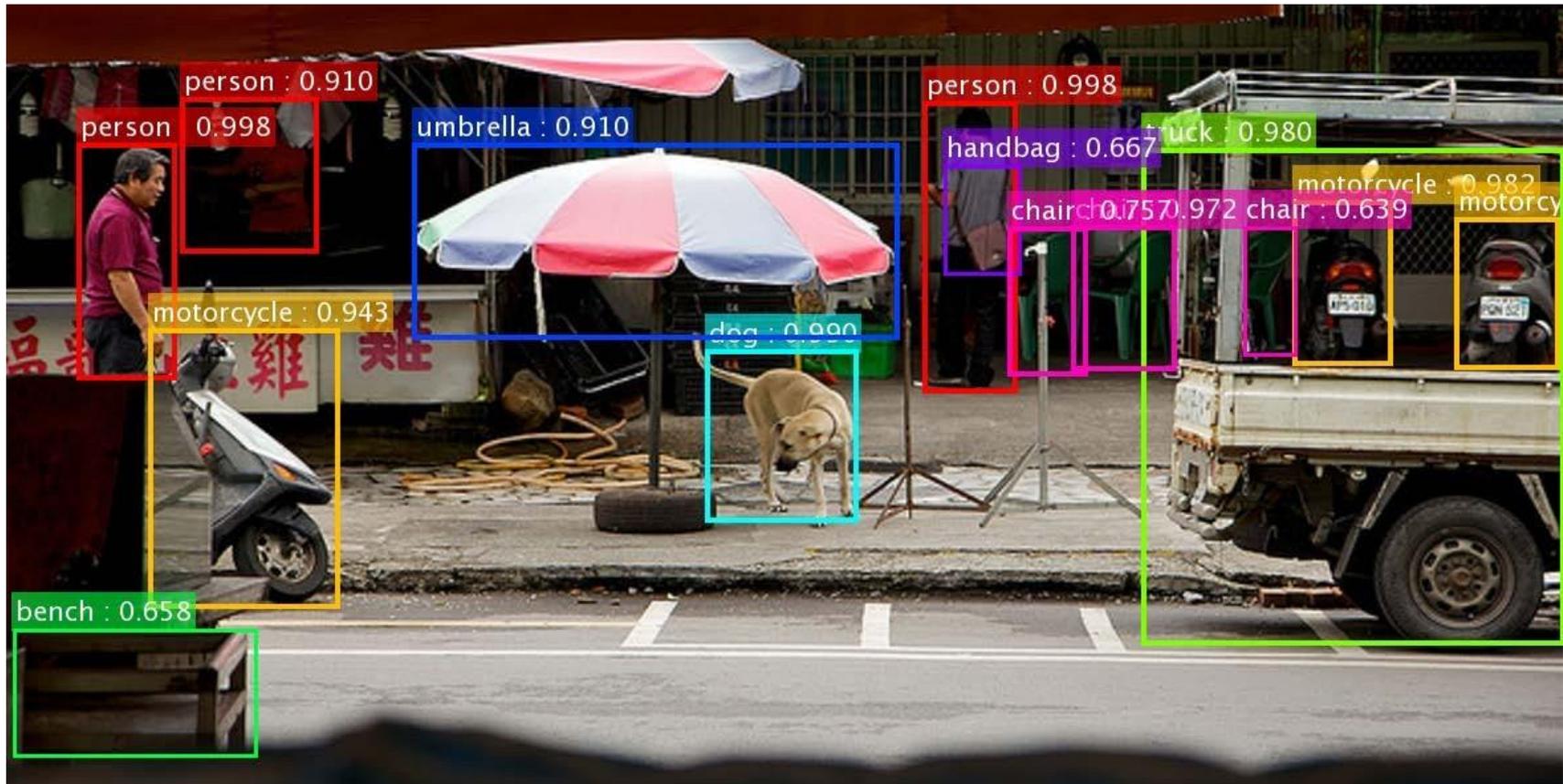
DNNの一層の深層化

ResNet, 152
layers



(slide credit: Jian Sun, MSR)

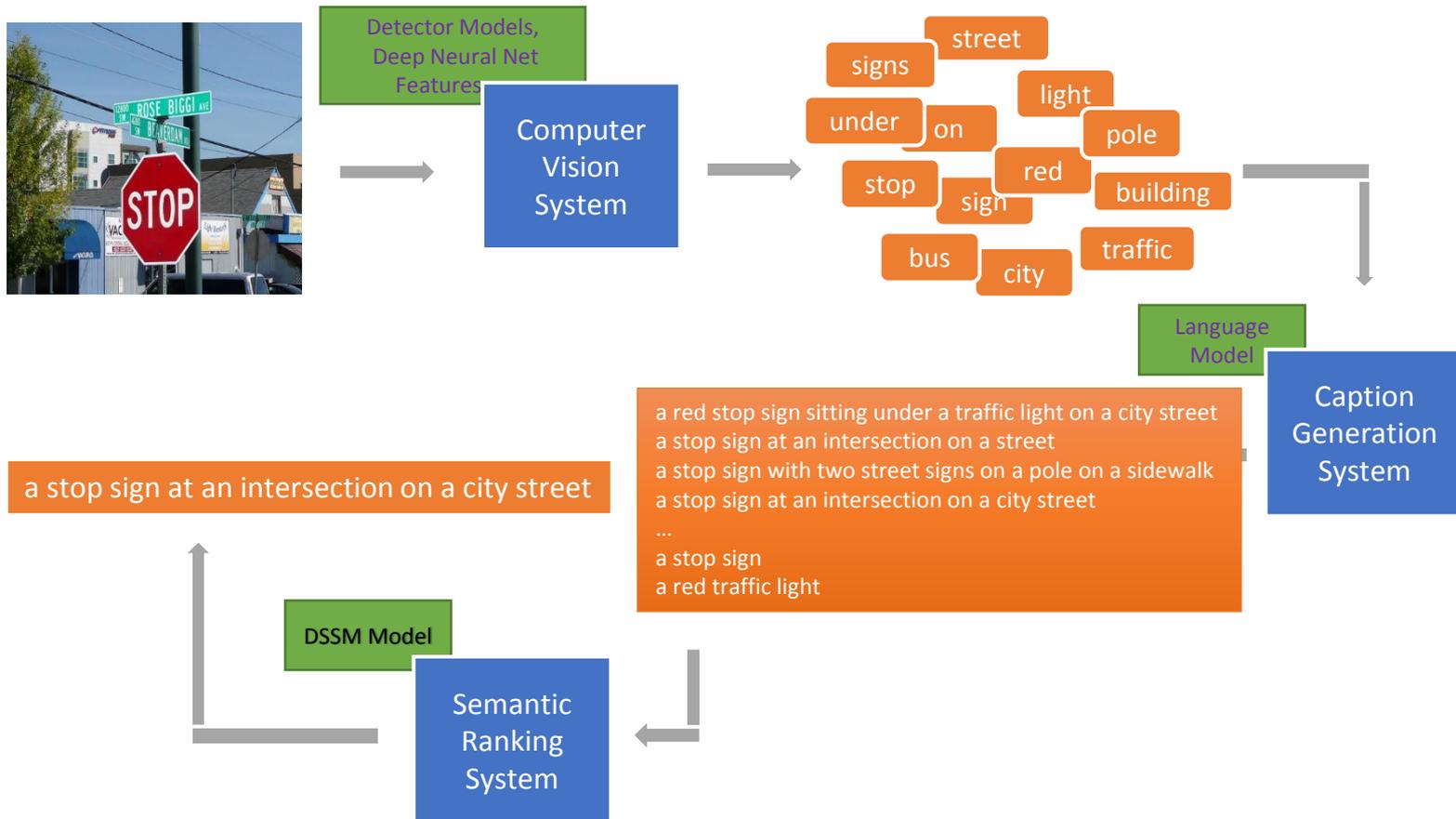
画像認識



*the original image is from the COCO dataset

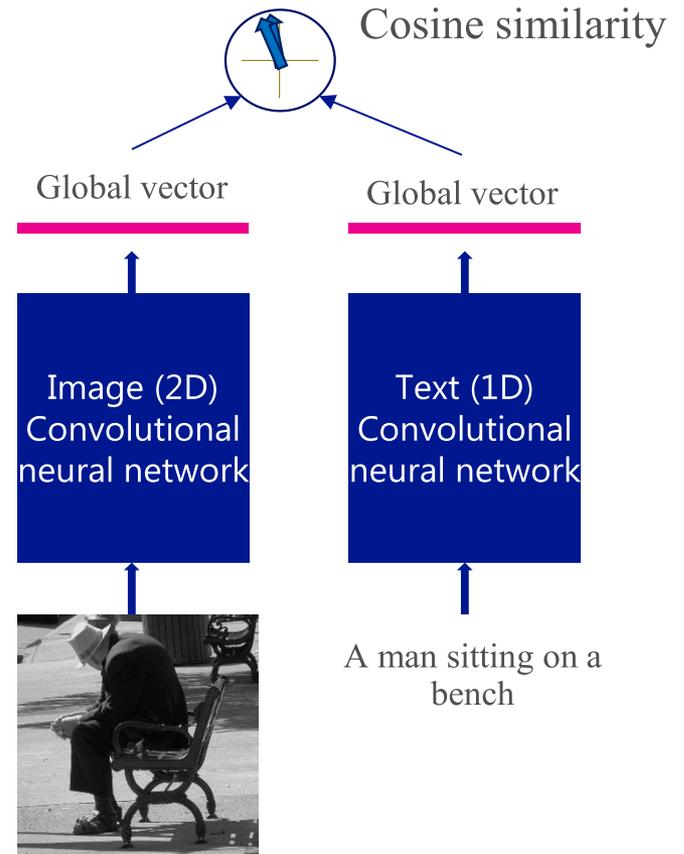
Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. "Deep Residual Learning for Image Recognition". arXiv 2015. Shaoqing Ren, Kaiming He, Ross Girshick, & Jian Sun. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks". NIPS 2015.

画像への自動キャプション付与



マルチモーダル意味適合

- Project a sentence and image into the same semantic vector space
- where similarity can be measured
- Model + basic features → re-ranked caption list



$Q = \text{image}, D = \text{caption}, R = \text{relevance}$

Relevance: $R(Q, D) = \text{cosine}(y_Q, y_D) = \frac{y_Q^T y_D}{\|y_Q\| \|y_D\|}$

Caption probability: $P(D|Q) = \frac{\exp(\gamma R(Q, D))}{\sum_{D' \in \mathcal{D}} \exp(\gamma R(Q, D'))}$

Candidate captions ↑ ↓ Smoothing factor

Objective: $L(\Lambda) = -\log \prod_{(Q, D^+)} P(D^+|Q)$

Correct caption ↑



A

a woman in a kitchen preparing food

B

woman working on counter near kitchen sink preparing a meal

(Li Deng, 2016)



Machine:

a woman in a kitchen preparing food

Human:

woman working on counter near kitchen sink preparing a meal

(Li Deng, 2016)

人工知能(機械学習)の最近の進歩

- 従来のコンピュータは、人間がプログラムやデータの内容を与えなければならなかったが、最近では、**機械学習によりデータの内容を自動的に理解し、適切な行動をするようになってきた**(IBMの質問応答システム「Watson」、コンピュータ将棋・囲碁など)。
- **Deep Learning(深層学習)、強化学習などの機械学習アルゴリズムと、コンピュータとネットワークの進歩に支えられたデータ処理量の急激な増大(Big Data)が、それを支えている。**
- Googleが世界中の本をスキャンしているのは、人々に読ませるためではなく、コンピュータに読ませるため。
- Googleや主要大学は、人間の生体情報や遺伝子情報を収集・解析し、病気の早期発見・治療に役立てる研究を始めている。
- **データや知識に基づく判断は、人間よりコンピュータの方が性能が高くなる**(弁護士、医者、ホワイトカラーなどの多くの仕事)。

スマートロボット

- 車を含め身の回りのすべてが、(進化した)スマートフォンで相互につながり(IoT)、制御されるようになる(ロボット化)。
- AIと組み合わせた次世代ロボットの開発が、シリコンバレーの新たなロボット産業の潮流になっている(例: Toyota's TRI)。
- 自動運転車、掃除ロボットなど
- 製造プロセスを導く情報の費用を含め、製品のコストのかなりの部分を情報(ソフトウェア)が占めるようになっている。
- 自動運転車は、川下にあるすべての産業(タクシー業界、運送業界など)へ波及していく。
- 人型ロボットの「不気味な谷(Uncanny Valley)」: 人間的な特徴を中途半端に持つようになると、ロボットへの好意が反感に変わり、恐怖を感じるようになる。これを越えることが鍵。

Toyota Research Institute (TRI)

- トヨタ自動車今年1月に米国に設立した、AI技術の研究開発を行う研究会社 (California州Palo Alto、Massachusetts州Cambridge、Michigan州Ann Arbor)
- Stanford大学およびMITとのAIの連携研究を強化する
- 最高経営責任者: Gill A. Pratt
- TRIの主なメンバー: Eric Krotkov, 岡島博司, Larry Jackel, James Kuffner, Russ Tedrake, John Leonard, Brian Storey
- Advisory Board: John Roos, Rodney Brooks, Fei-Fei Li, Daniela Rus他
- 予算: 当面5年間で約10億ドル(約1,200億円)、200人規模
- **4つの目標:**
 - a. 事故を起こさない車という究極の目標に向け、車の安全性を向上させる
 - b. 誰でも移動(運転)できる車を実現する
 - c. モビリティ技術を活用した屋内用ロボットを開発する
 - d. AIや機械学習の知見を利用し、科学的・原理的研究を加速させる

自動運転



自動運転の定義

- **レベル0**:ドライバーが常にすべての主制御系統(加速・操舵・制動)の操作を行う。運転支援システム。
- **レベル1**:加速・操舵・制動の**いずれか**をシステムが行う。自動ブレーキなどの安全運転支援システム。
- **レベル2**:加速・操舵・制動の内、**複数**をシステムが行う。アダプティブクルーズコントロール(ステアリングアシスト付き)など。ドライバーが常に運転状況を監視操作する。
- **レベル3**:加速・操舵・制動を**全て**システムが自動的に行い、システムが要請した時(緊急時やシステムの限界時)はドライバーが対応する。
- **レベル4**:加速・操舵・制動を**全て**システムが**完全**自動的に行い、ドライバーは全く関与しない。有人と無人がある。

自動運転の技術

- センサー（物体形状、距離、速度ベクトルなど）
 - 障害物・歩行者など
 - 交通標識・信号
 - 車線
 - * ミリ波レーダー、レーザレーダー、カメラなど
- 測位技術（GPS、3D地図、ジャイロ）
- 外部連携（クラウド、車車間、路車間）
- 認識・判断（AI）
- エンジン、ブレーキ、ハンドル制御
- システムの安全性・信頼性（フェイルセーフ）
- ユーザインタフェース（表示技術など）

自動運転の利点

- 運転からの解放。誰でも車で移動できる。
- 乗り心地が向上する。
- 車が移動するパーソナル空間になる。
- 交通事故が減る。
- 車間距離短縮により、道路容量が増え、交通流量制御が容易になる。
- 最高速度規制が緩和できる。
- カーシェアリングにより、自動車総数が削減できる。
- 駐車場不足が緩和できる。
- ぴったり寄せられるので、駐車スペースが削減できる。
- 自動車保険や交通警察の必要性が減る。
- 物理的な道路標識が削減できる。
- 車両盗難が減る。

自動運転車は世の中の仕組みを変える

- 車は所有するものではなく、利用するものになる
- 車の生産数が大幅に減るとともに、高級車の需要は減り、関連産業が縮小する(生産・販売・サービスの融合)
- エネルギー消費が減る(現在の米国ではエネルギーの40%が交通消費)
- ぶつからなければ、車の形や、作りも変わってくる
- ICT (AI)、ハードウェア、エネルギーが、競争の3要素になる
- 車の生産コストの半分以上が、インフラを含むソフトウェアシステムになる
- ユーザインタフェースも、自動車の使い勝手の重要なファクターになる
- 運転免許証がいらなくなる
- 自動運転車は、通信事業者や、電力会社から買うことになるかもしれない
- タクシー運転手は、いらなくなる
- 家に、車庫がいらなくなる
- 社会システムを構成する規制や官庁、産業構造、都市の設計などに変化をもたらす
- 日本の自動車会社は、グーグルに買われてしまうかもしれない
- 日本がこの変化を主導できるかどうか、日本の未来の鍵

自動運転の課題

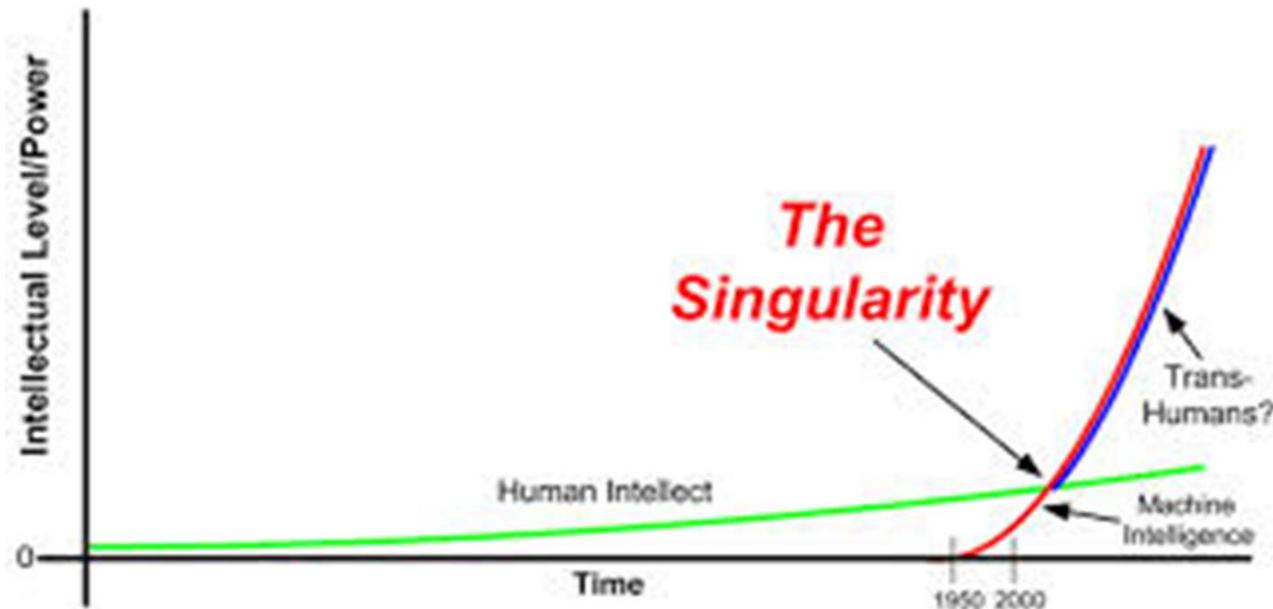
- **技術的課題**
 - 天候の影響、地図の精度、車外の人（ジェスチャー、合図）とのインタラクション、無線通信周波数帯域の確保、多様なマニュアル運転車との共存
- **トラブルへの懸念**
 - ソフトウェアの信頼性、ハッキング、マニュアル運転が必要になった時のドライバーの経験不足、事故が不可避の時に複数の事故のどれを選択するか
- **制度上の問題**
 - 損害賠償責任、自動運転車の法的枠組み
- **社会的影響**
 - テロリストによる悪用など

AIの今後の展開

- Google (AI) + 人間 → 一種の**人間の能力の拡張**
- **人の能力を凌駕するAIシステム**が、これからどんどん出てくる。
- 3.11の原発過酷事故も、**AI**を活用すれば防げたかもしれない。
- **人の脳に近いAGI (Artificial General Intelligence)**が開発される。
- **AI**が市場を出し抜き、発明、特許取得で人間の研究者を圧倒し、大衆操作で人間のリーダーを上回ることになるかもしれない。
- 人が解決できない難しい国際問題なども、**AI**に任せた方がよくなるかもしれない。
- アメリカでは、我々が知らない**秘密AIプロジェクト**が、沢山走っている。Siriも、元々DARPAの資金を受けているSRIのプロジェクトを分離したもので、「Stealth Company」だった。
- 難しいこと(推論)は簡単で、**簡単なこと(知覚や行動)**は難しい。

Singularity (技術的特異点)

- Ray Kurzweil: “The singularity is near: When humans transcend biology” 「ポストヒューマン誕生—コンピュータが人類の知性を超えるとき」
 - 遺伝学、ナノテクノロジー、人工知能(AI)で社会が大きく変わる。AIが自分で進化するようになり、指数関数的に発展して、2045年にはAIが人類全体の知能を追い抜く。「人工超知能(Artificial Super-Intelligence)」



Singularity (技術的特異点)

- Ray Kurzweil: “The singularity is near: When humans transcend biology” 「ポストヒューマン誕生—コンピュータが人類の知性を超えるとき」
 - 遺伝学、ナノテクノロジー、人工知能(AI)で社会が大きく変わる。AIが自分で進化するようになり、指数関数的に発展して、2045年にはAIが人類全体の知能を追い抜く。「人工超知能(Artificial Super-Intelligence)」
- AIと人の知能が結びつくことにより、人類の生活が変容する。
- 脳が埋め込みデバイスによって、どこにいてもワイアレスでクラウド・コンピュータと接続できるようになり、新しい知識や技術をダウンロードできるようになる。考えるだけで、答えが得られるようになる？ (本来の)脳が退化するのではないか？
- AIにできない判断力、創造性を持てるか、AIをどれだけ使いこなせるかがこれからのリーダーの鍵。

AIは心(意識)を持てるか

- 我々の素晴らしい心は、物理を基礎にしている。物理的に可能なことは、**すべてテクノロジーで実現**できる？
- 脳のreverse engineeringやsimulationには、まだ時間がかかる。**意識**は、脳細胞が実行する論理的なアルゴリズムではなく、膨大な量の混沌とした心理作用が新皮質で統合された結果。
- 人の**主観的な世界**について客観的に確認できない→AIの内部世界(**意識**の所在)も知ることができないかもしれない。
- **人間が理解できない人工超知能**(ASI: Artificial Super Intelligence)が出現したら、どうする？ 世界が征服される恐れはないのか？
- 映画「Transcendence」: AIの世界的研究者の死後に、その脳を移植したコンピュータが暴走を始めて止められなくなり、ウィルスを注入して止めたところ、つながっていた世界中のシステムが崩壊して、世界中が荒廃してしまう話。

まとめ

- **音声・画像認識技術**は、**人工知能(AI)**研究の牽引車として、**深層学習(DNN)**の効果により急速に進歩しているが、人の能力に近づくには、まだ研究すべき課題がたくさんある。
- **スマートロボット**(自動運転を含む)も、AIの重要分野の一つ。
- **自動運転**のために、多様な技術が研究・開発されている。
 - 自動運転は、世の中の仕組みを変える。
 - 自動運転には、まだ沢山の課題がある。
- 今後、**AI**が加速度的に進歩し、生活を変える。データや知識に基づく判断はコンピュータに任せ、人間は**AIを使いこなす**とともに、**AIにできない判断力**や**創造性**を持てることが必要。
- **AGI**や**人工超知能(ASI)**の出現に、どう対応・利用するか？
- これからはハードウェアでなく、**(情報)システム**の戦い。
- 我が国が存在感を失わないためには、**研究戦略の国際化**が必須。