



第20回情報論的学習理論ワークショップ (IBIS2017)

The 20th Information-Based Induction Sciences Workshop

2017.11.8~11 東京大学 本郷キャンパス

IBIS 2017 紹介(一部)

Artificial Intelligence



Tokyo AI Tech. Alliance

https://github.com/xurong1981/ai_room.git

第20回情報論的学習理論ワークショップ (IBIS2017)



情報論的学習理論ワークショップ (IBIS,Information-based Induction Sciences)は機械学習に関する日本最大の学際的フォーラムであり、初回の1998年から数え今年で第20回という節目の年を迎えます。今から20年前、情報理論、統計学、統計物理学等の諸問題を「学習」という視点から総合的に捉える新しい知識情報処理の体系を目指してIBISは発足しました。現在では、画像処理、言語処理、音声・音響処理、医療情報などの計算機科学・情報処理にとどまらず、医学、生物学、脳科学、神経科学、化学、材料科学、実験物理学、天文学などのサイエンスや様々なビジネスの現場へとその重要性は広がっています。

現在我々が直面している技術革新は、日々新しい技術を生み出し、様々な分野に瞬時に広がっていく非常に流動性の高い状況をもたらしました。今年は、広がりゆく機械学習の重要性と新しいイノベーションを確かなものにするために、その中でもなお揺るがぬものを見つめることをテーマに議論して行きたいと思います。

安田講堂

東京大学 本郷キャンパス



実社会への機械学習の応用

IBIS2017 企画セッション

IBIS2017 企画セッション
実社会への機械学習の応用

東京大学 本郷キャンパス 安田講堂
11月9日 16:00～18:00

竹中一仁 (デンソー)

実社会への機械学習の応用

IBIS2017 企画セッション

実社会への機械学習の応用

概要

実社会への応用の中でも特に物理的実体を伴った「自動車」「ロボット」にフォーカスし
機械学習応用のアプローチ・先端事例や その課題について講演いただく

自動車分野への応用

「自動車の自律化におけるロボティクスと学習」

名古屋大学 竹内栄二郎

産業界でのロボットへの応用

「実社会・実環境におけるロボットの機械学習」

Preferred Networks 高橋城志

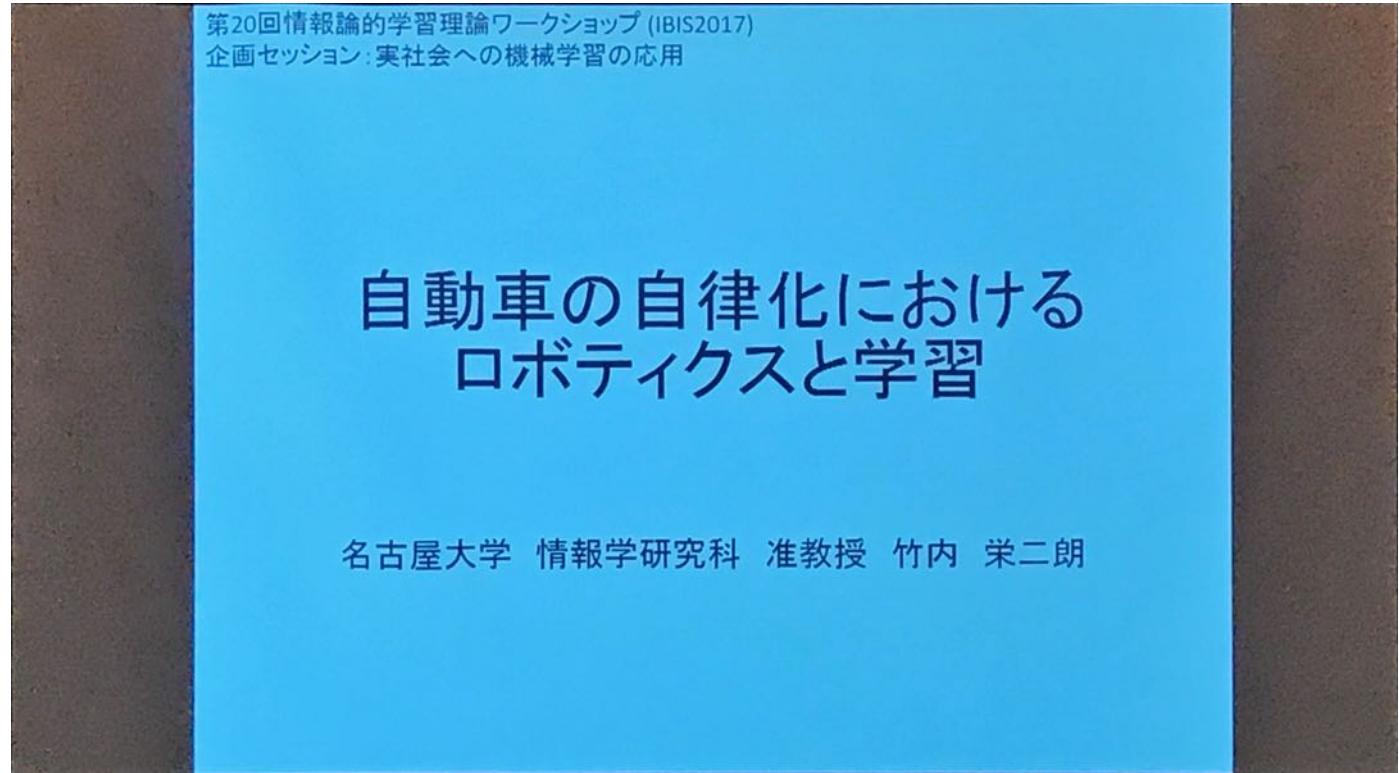
インターラクションと概念獲得への応用

「マルチモーダルカテゴリゼーション：階層ベイズモデルに基づくロボットによる概念・言語獲得」

電気通信大学 中村友昭

自動車の自律化におけるロボティクスと学習

名古屋大学 情報学研究科 准教授 竹内 栄二郎

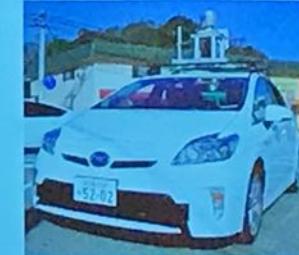


自動車の自律化におけるロボティクスと学習

名古屋大学 情報学研究科 准教授 竹内 栄二郎

自己紹介

- 2002–2008 筑波大学 知能ロボット研究室
 - 移動ロボット機能のモジュール化
 - 位置推定・SLAMに関する研究
- 2008–2014 東北大学 田所研究室
 - 屋外移動ロボット
 - レスキュー・ロボット
- 2014–2016 名古屋大学 未来社会創造機構
 - 車の運転支援と自動運転の研究
- 2016–現在 名古屋大学 情報学研究科
武田研究室
 - 自動運転
 - 学習による自動運転



自動車の自律化におけるロボティクスと学習

名古屋大学 情報学研究科 准教授 竹内 栄二郎

近年の自動運転研究開発の活発化

- DARPA GrandChallenge(2004,2005)／UrbanChallenge(2007)
 - 自動運転車の競技
 - Velodyne LiDARの開発
 - 欧米での自動運転研究の活発化
- Google
 - UrbanChallenge経験者が集結
 - 公道での自動運転190万キロ以上
- Daimler
 - 2013年8月 100kmの市街地自動運転
 - 量産型のセンサを利用(ステレオカメラ、LiDAR、RADAR等)
 - 高精度地図の積極的利用(HERE)

多くの場合で公道での自動運転は可能

自動車の自律化におけるロボティクスと学習

名古屋大学 情報学研究科 准教授 竹内 栄二郎

名古屋大学の自動運転研究環境

- ・自動運転研究設備
 - ・3台の自動運転対応車両
 - ・豊富な整備・実験設備
- ・走行実験環境
 - ・豊田市交通安全学習センターの協力による模擬市街地での実験等
- ・自動運転実験ガイドライン
 - ・自動運転実験をより安全に実施できる体制づくり

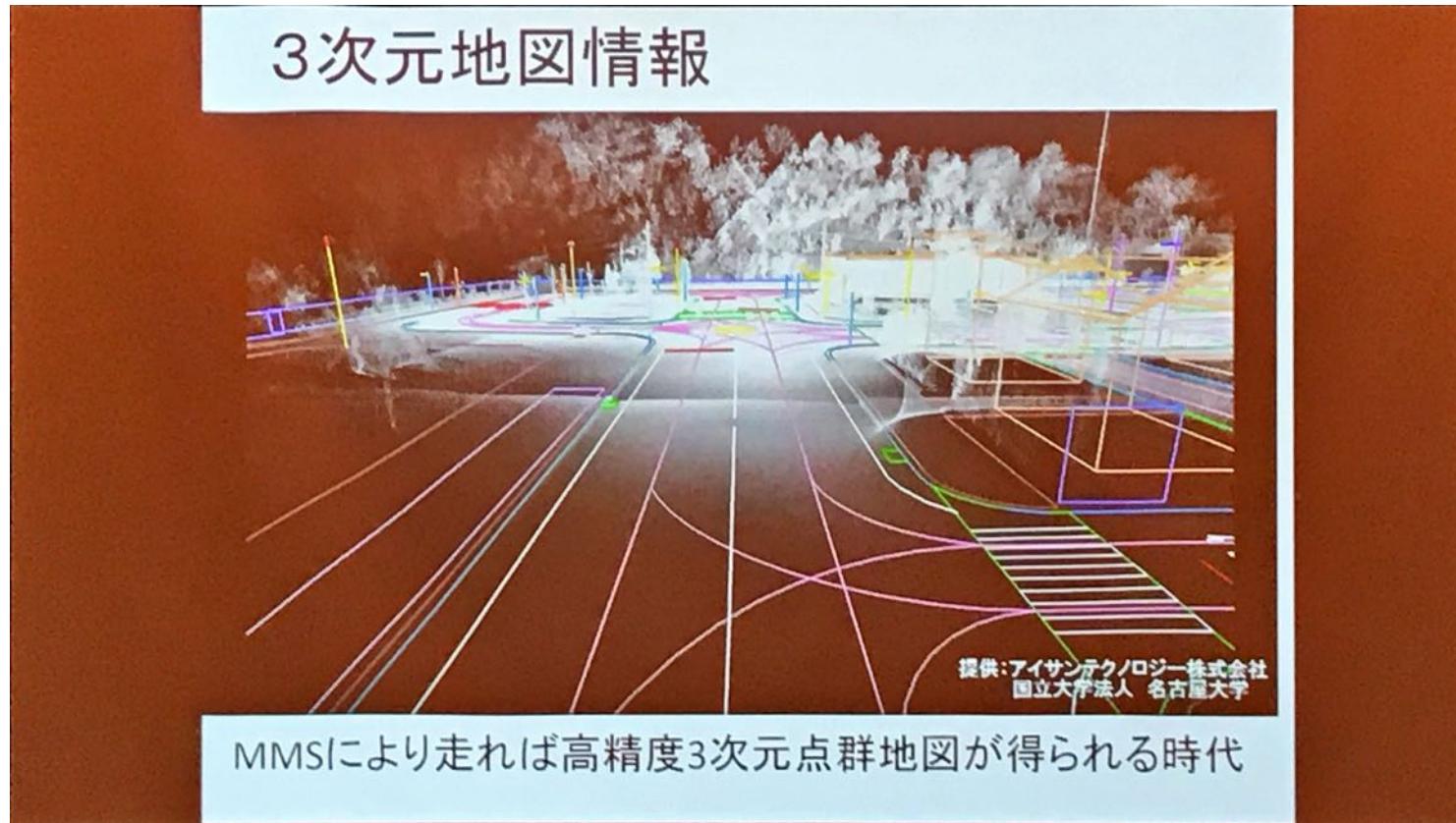


センサを搭載した自動運転対応車両 シャシーダイナモ 模擬市街地での自動運転実験

本格的に自動運転研究ができる環境の整備

自動車の自律化におけるロボティクスと学習

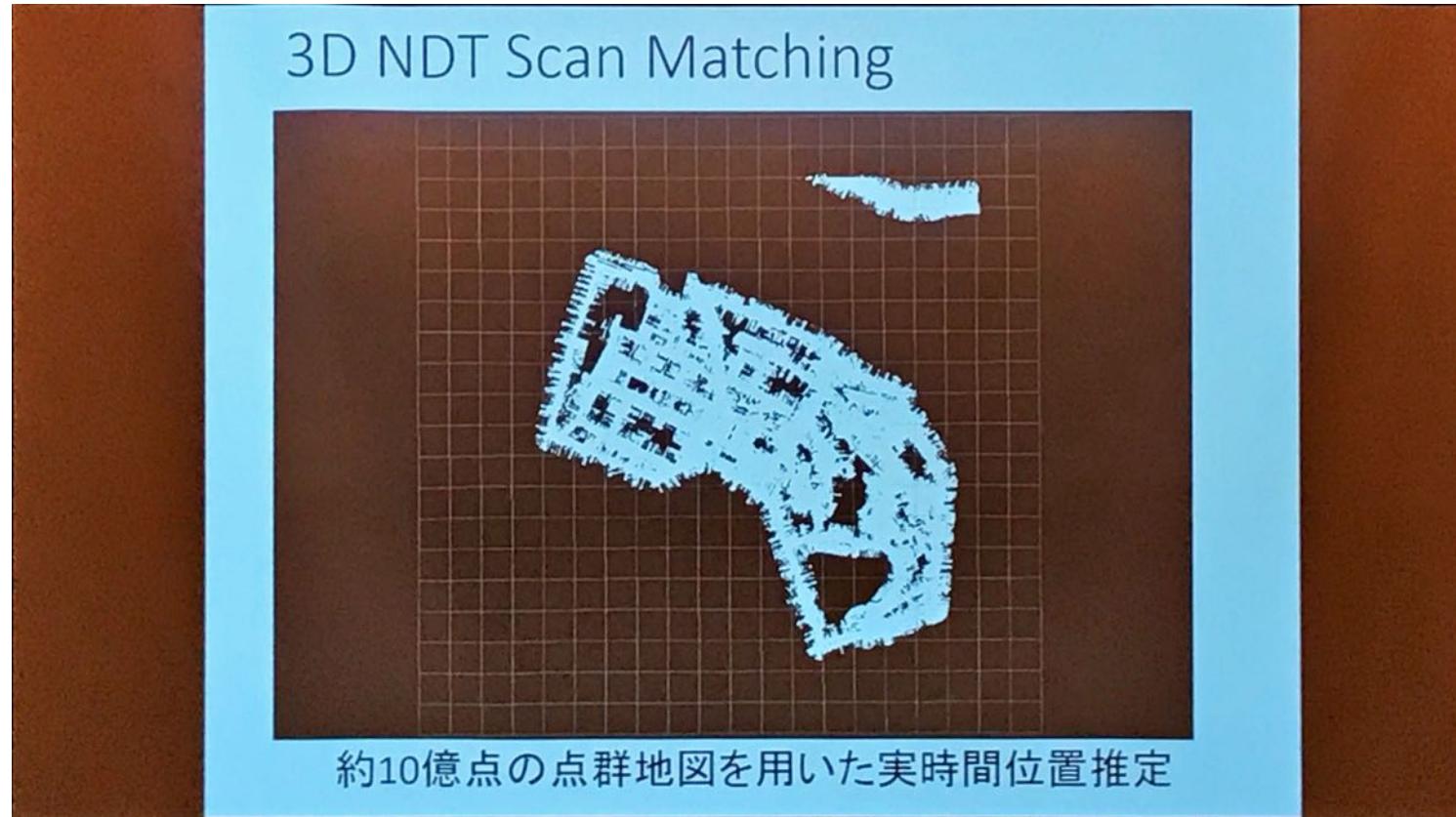
名古屋大学 情報学研究科 准教授 竹内 栄二郎



自動車の自律化におけるロボティクスと学習

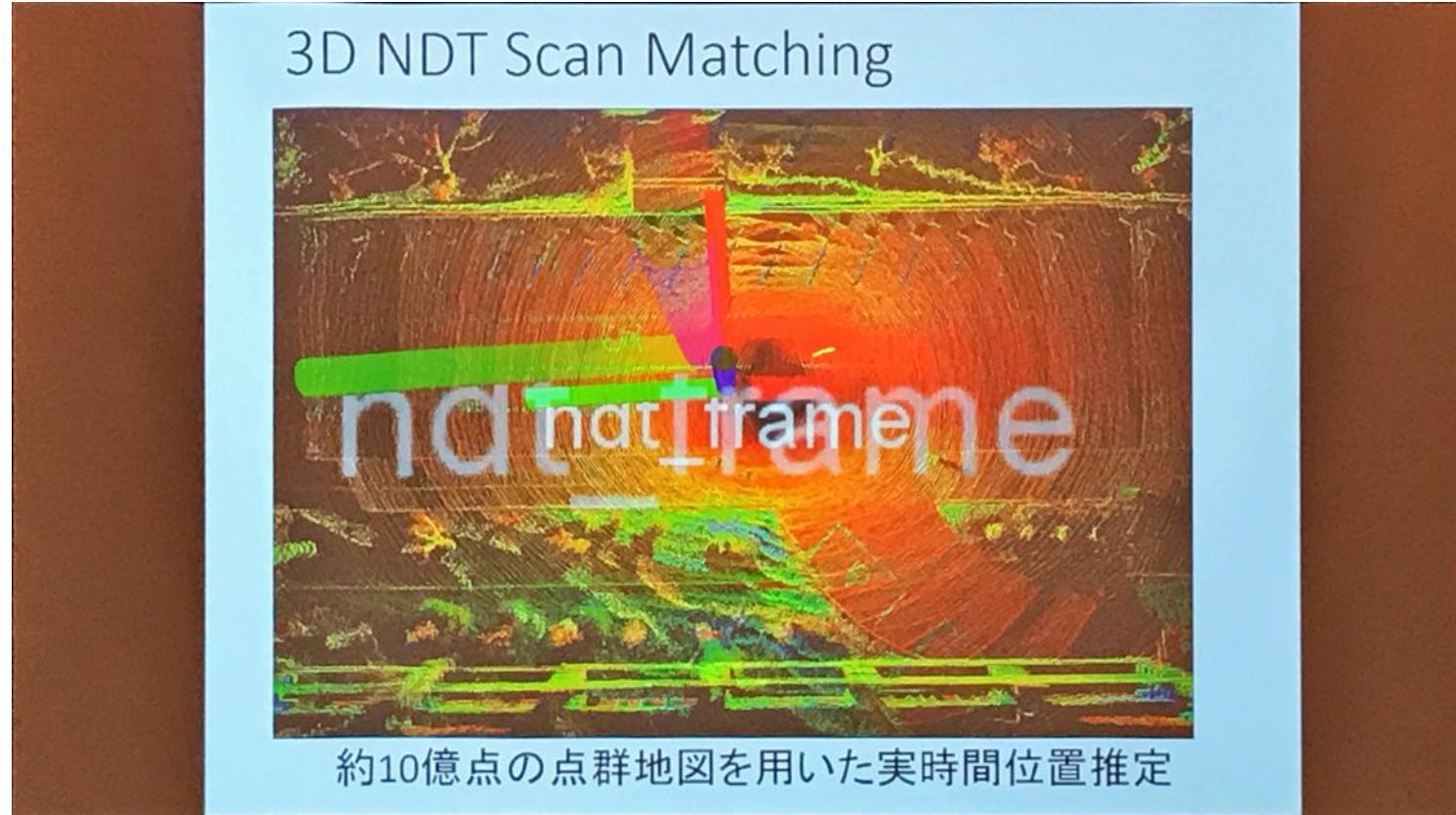
11

名古屋大学 情報学研究科 准教授 竹内 栄二郎



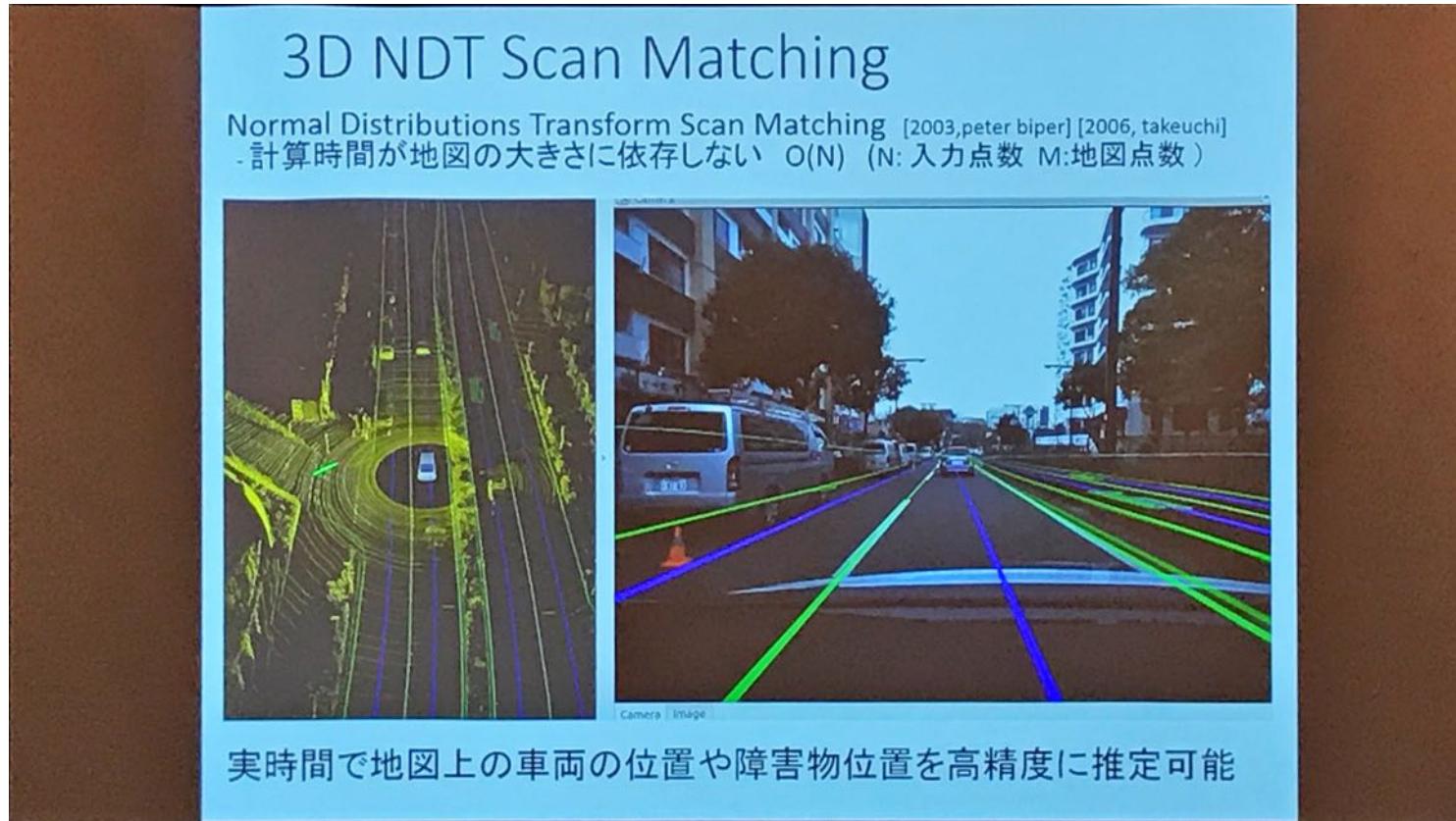
自動車の自律化におけるロボティクスと学習

名古屋大学 情報学研究科 准教授 竹内 栄二郎



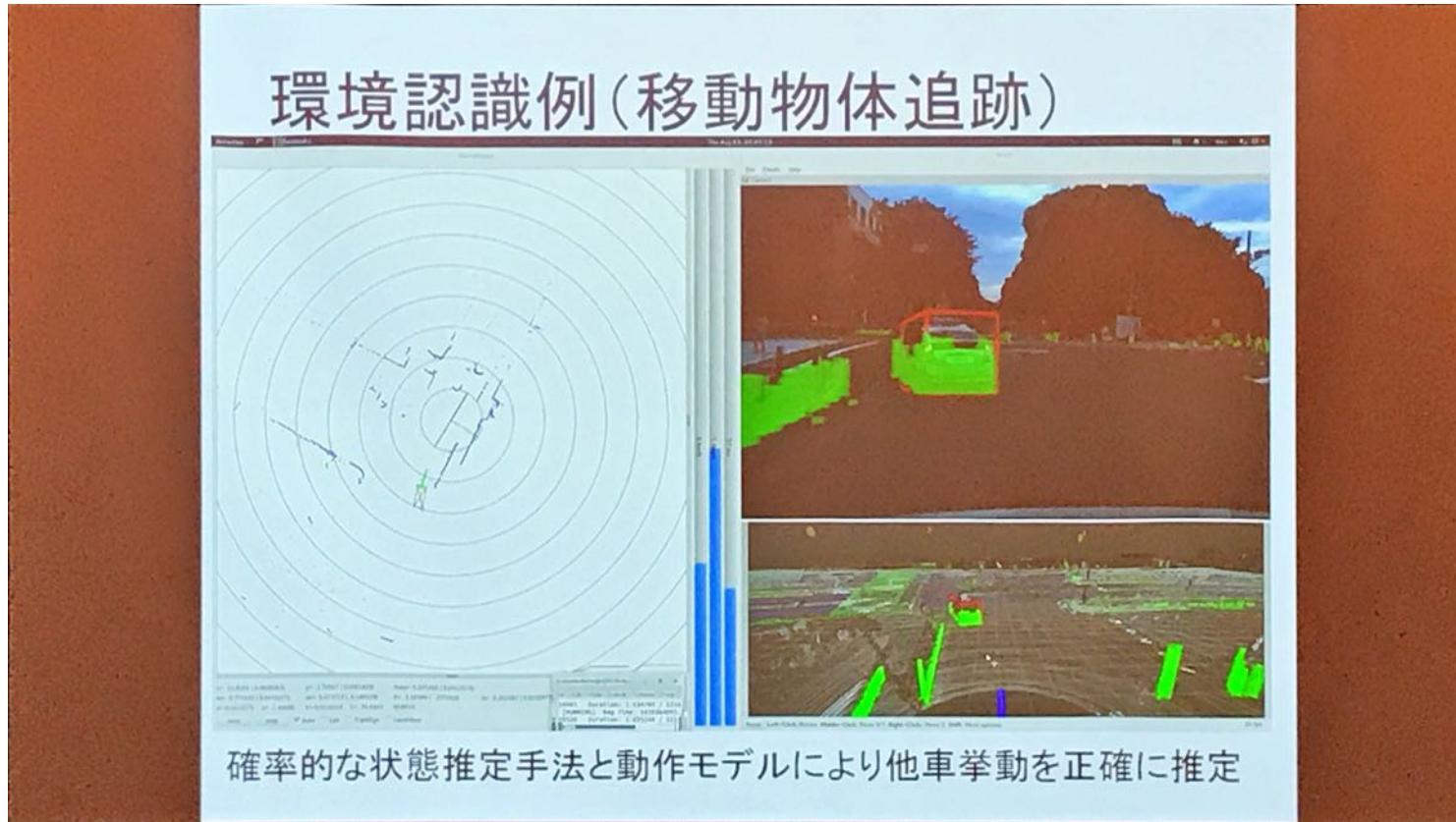
自動車の自律化におけるロボティクスと学習

名古屋大学 情報学研究科 准教授 竹内 栄二郎



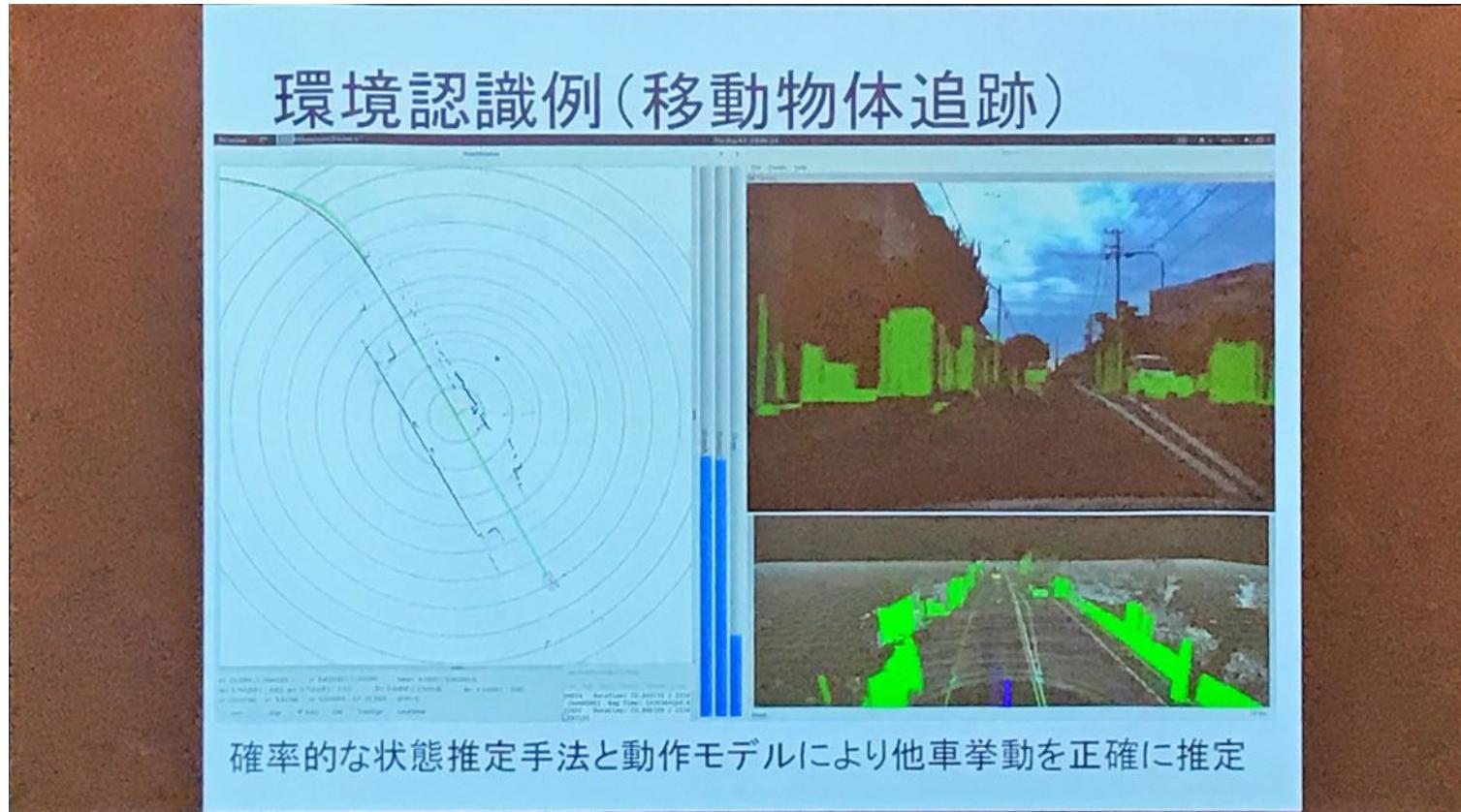
自動車の自律化におけるロボティクスと学習

名古屋大学 情報学研究科 准教授 竹内 栄二郎



自動車の自律化におけるロボティクスと学習

名古屋大学 情報学研究科 准教授 竹内 栄二郎



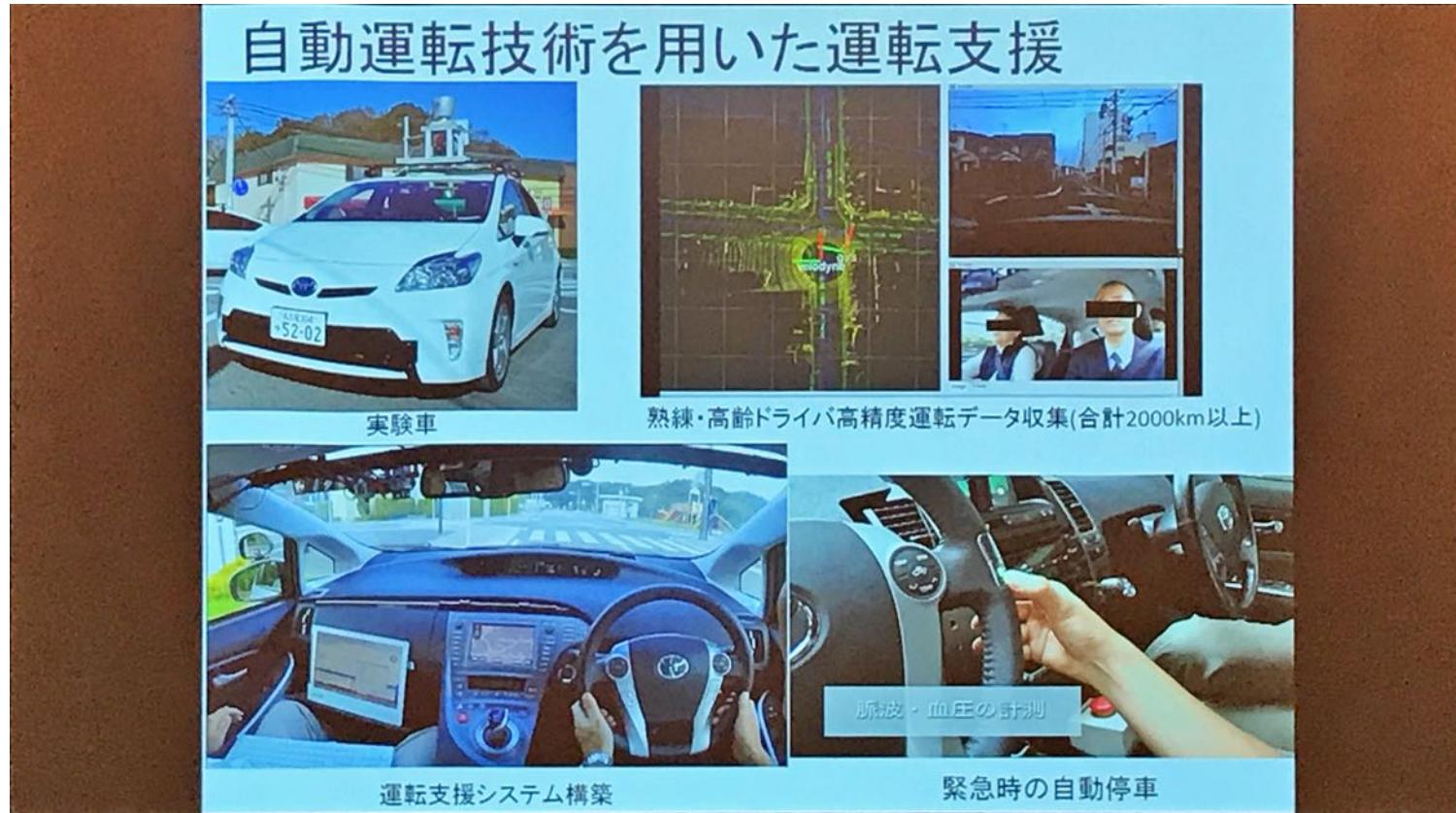
自動車の自律化におけるロボティクスと学習

名古屋大学 情報学研究科 准教授 竹内 栄二郎



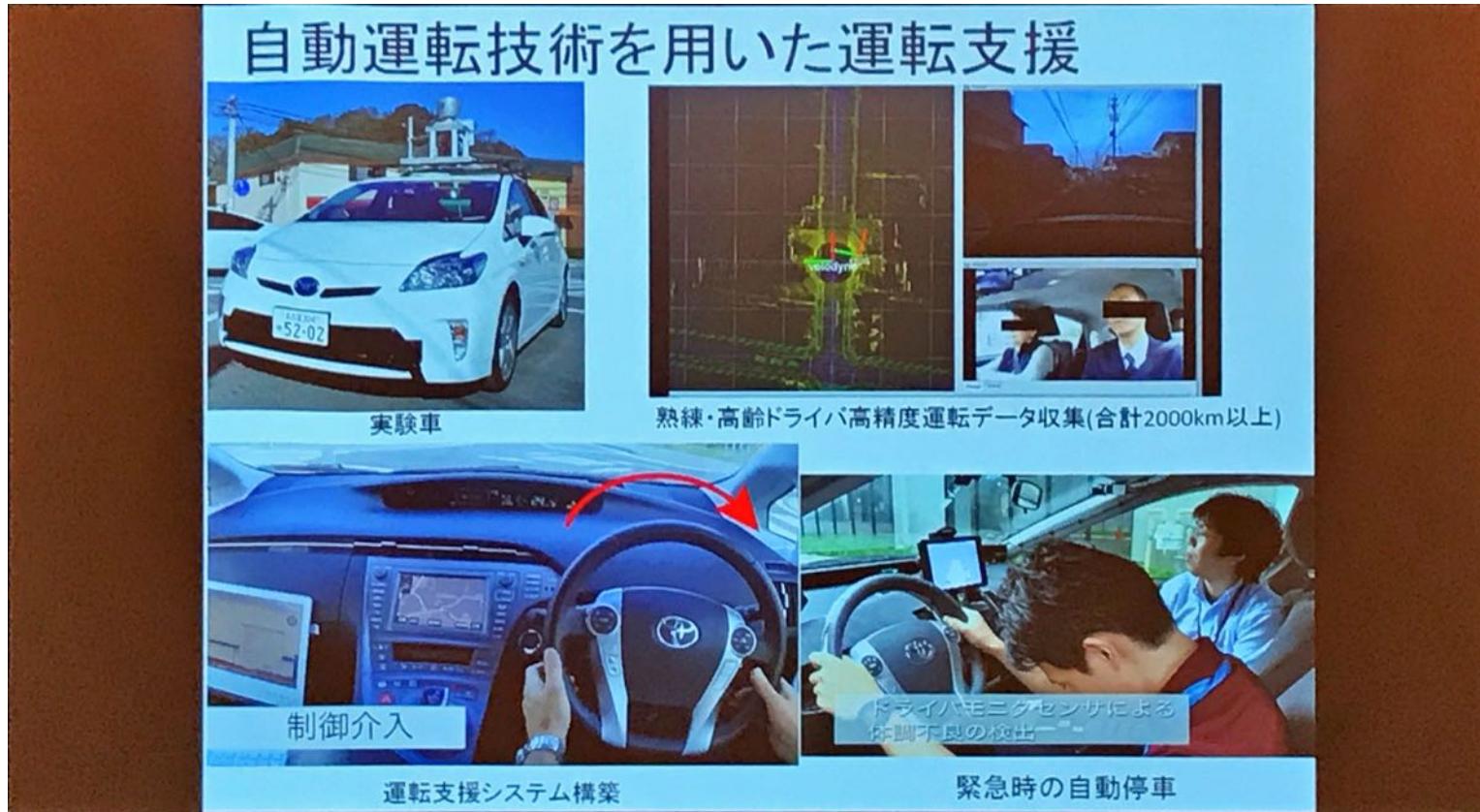
自動車の自律化におけるロボティクスと学習

名古屋大学 情報学研究科 准教授 竹内 栄二郎



自動車の自律化におけるロボティクスと学習

名古屋大学 情報学研究科 准教授 竹内 栄二郎



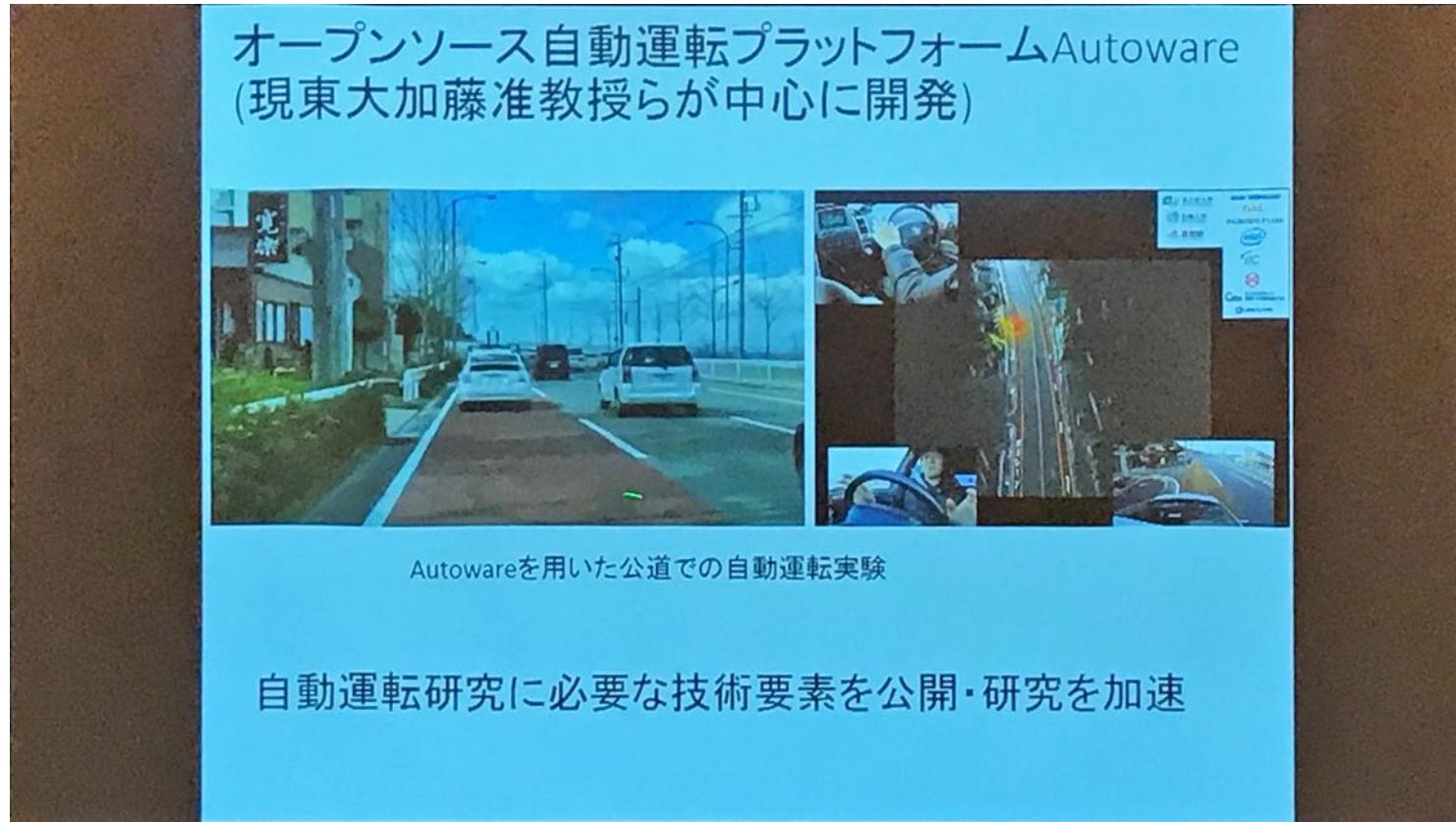
自動車の自律化におけるロボティクスと学習

名古屋大学 情報学研究科 准教授 竹内 栄二郎



自動車の自律化におけるロボティクスと学習

名古屋大学 情報学研究科 准教授 竹内 栄二郎



自動車の自律化におけるロボティクスと学習

名古屋大学 情報学研究科 准教授 竹内 栄二郎

運転行動信号処理(武田研)

- ・運転により得られる信号からドライバの個性や運転行動を分析
 - ・音響・音声信号処理を運転行動信号処理に応用
 - ・連続信号の機械学習に多くの知見
 - ・1000人以上の運転データを収集

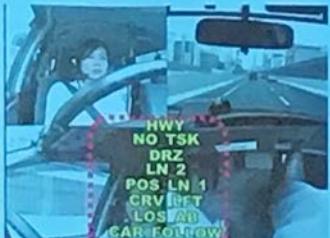
The diagram illustrates the evolution of sensor technology in two Toyota vehicles. On the left, the 'TOYOTA REGIUS 1999 – 2005 (retired)' is shown with basic sensors: Velocity (speedometer icon), GPS (GPS receiver icon), Pedal pressure (gas, brake) (pedal icon), and Steering angle (steering wheel icon). On the right, the 'TOYOTA Hybrid ESTIMA 2006 –' is shown with a much larger number of sensors: Velocity (speedometer icon), GPS (GPS receiver icon), Pedal pressure (gas, brake) (pedal icon), Steering angle (steering wheel icon), Speech and noise (microphone icon), Video (head, face, feet) (camera icon), Video (360-degree view) (360-camera icon), 3D acceleration (accelerometer icon), Laser scanners (front, back) (laser scanner icon), Skin conductance (skin conductance meter icon), Heart rate (heart rate monitor icon), and Perspiration (palm and sole) (hand icon).

自動車の自律化におけるロボティクスと学習

名古屋大学 情報学研究科 准教授 竹内 栄二郎

音声信号処理と運転行動信号処理

- 運転信号: 可変長の時系列信号
 - 音声・対話等で利用されている可変長・時系列信号処理法が適用可能



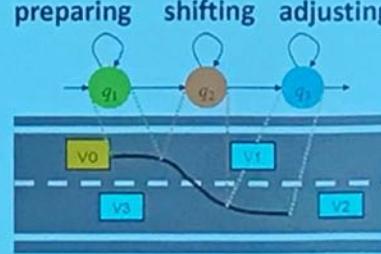
Tags

HWY, NO_TSK, DRZ,
LN_*, POS_LN_*, LOS_AB:
LN_ENDS_IN, CAR_FOLLOW,
FACE_POS, HEAD_OTHER

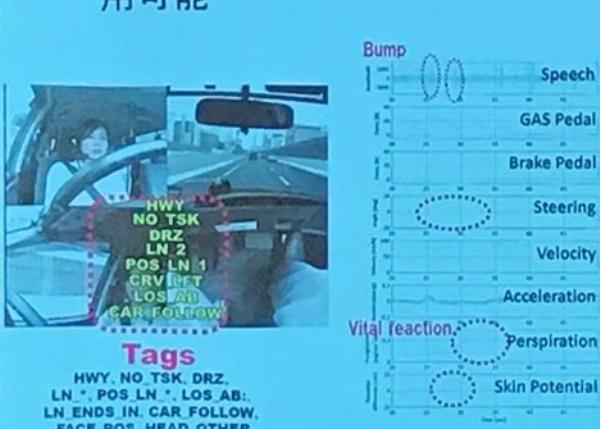
Bump

Speech
GAS Pedal
Brake Pedal
Steering
Velocity
Acceleration
Vital reaction
Perspiration
Skin Potential

preparing shifting adjusting



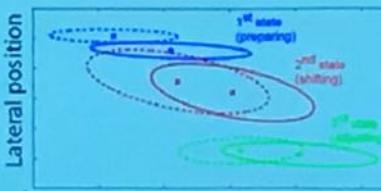
運動データからの運転行動分類



Lateral position

Forward speed

隠れマルコフモデルによる車線変更推定



自動車の自律化におけるロボティクスと学習

名古屋大学 情報学研究科 准教授 竹内 栄二郎

運転データからの運転技能学習

多量の運転データから自動運転技能は獲得できるのか？

- NvidiaのBojarskiらが、運転データを用いたEnd-to-End learningによる自動運転を実現
 - CNN(Convolutional Neural Network)を用いて前方カメラの映像からステアリング角を生成
 - 様々な運転状況下の走行データ72時間分で学習し、公道でのレーン内走行を実現

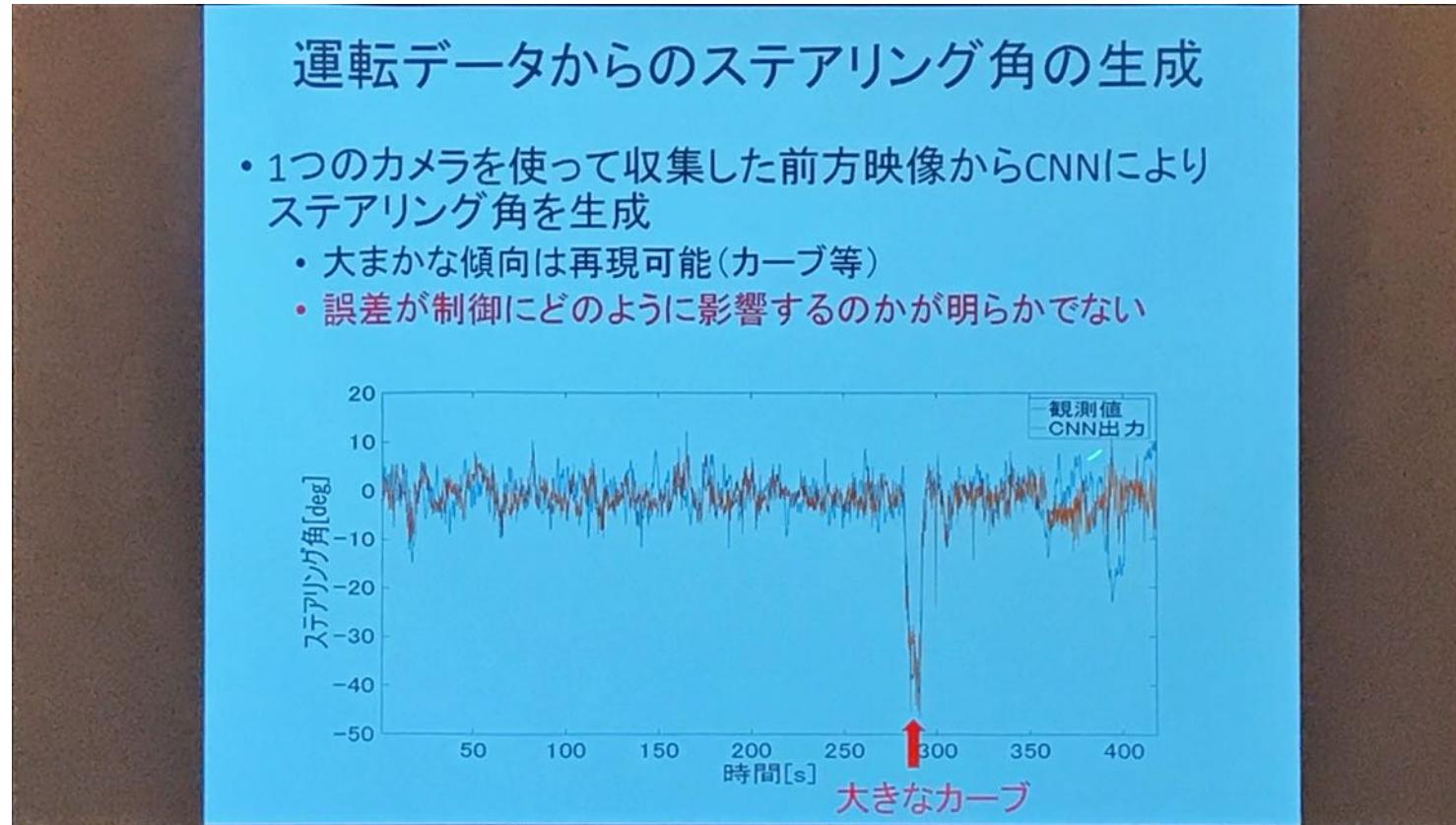
前方映像 → CNN → 車両制御

[Mariusz Bojarski, et al."End to End Learning for Self-Driving Cars"]

16

自動車の自律化におけるロボティクスと学習

名古屋大学 情報学研究科 准教授 竹内 栄二郎

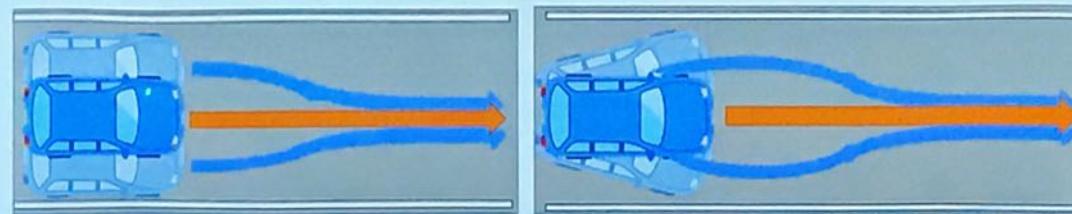


自動車の自律化におけるロボティクスと学習

名古屋大学 情報学研究科 准教授 竹内 栄二郎

学習データの生成

- ・中央と左右に前向きに取り付けた3つのカメラを使用
 - ・視線の方向を変化させる視点変換を行い、進行方向や車体がずれた場合の学習データを生成
 - ・レーン中心に戻るような教師データを付与し、学習データを増加



中心のカメラだけでなくずれた位置の画像も利用

[Mariusz Bojarski, et al."End to End Learning for Self-Driving Cars"]

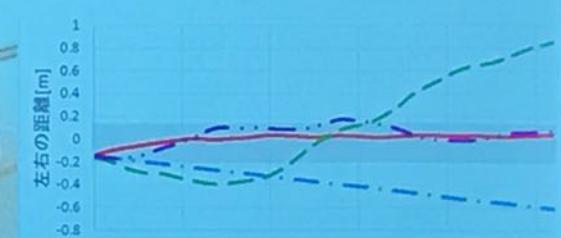
18

自動車の自律化におけるロボティクスと学習

名古屋大学 情報学研究科 准教授 竹内 栄二郎

追実験

- BojarskiらのCNNと同様の構造用いて前方映像から運転操作を生成
 - 入力: 前方画像 出力: 操舵角
 - ネットワーク: 置み込み5層+全結合3層
- 移動体を利用し4種類の学習データを用意
 - カメラ1つ+視点変換あり・なし
 - カメラ3つ+視点変換あり・なし
- 学習データを学習したCNNを用いて、直線路における走行軌跡とレーン中心からの誤差を比較



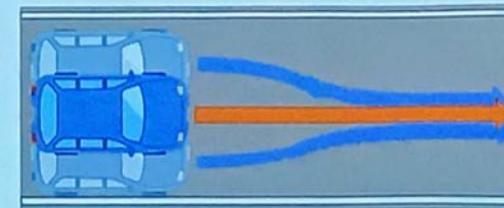
自動車の自律化におけるロボティクスと学習

27

名古屋大学 情報学研究科 准教授 竹内 栄二郎

学習による制御の獲得

- 制御：ずれた状態からどんな出力をすれば戻るのかを最適化問題として解く
- 学習時は前方カメラ画像を視点変換し、角度がずれた状態を模擬し、本来の経路に戻る操舵角を学習データに付与
 - 正しい状態 → 維持する動作
 - ずれた状態 → 戻る動作
- 制御をするためには、通常の運転データだけでなく、ずれた状態と、そこから正しい状態に戻る動作情報のセットが必要

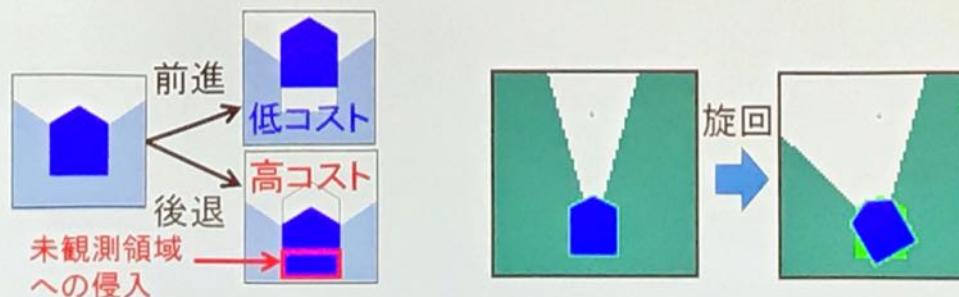


自動車の自律化におけるロボティクスと学習

名古屋大学 情報学研究科 准教授 竹内 栄二郎

視野を考慮した経路計画

- ・行動: 移動
- ・観測状態: 占有度グリッドマップ(未知・自由・占有)
- ・行動に必要な観測状態
 - ・動作でどれだけ未観測領域に侵入するか
- ・行動で変化する観測状態
 - ・どこが可視になるかを予測



自動車の自律化におけるロボティクスと学習

名古屋大学 情報学研究科 准教授 竹内 栄二郎

視野制限下での経路計画

- ・経路計画の状態変数に観測状態を付加 (x, y, θ, map)
- ・安全を確認してから移動する動作が自然に発生

The slide features three diagrams illustrating path planning under field-of-view constraints:

- 後ろの目的地への移動**: A vehicle moves from a red start point to a blue destination through a narrow corridor. A yellow arrow indicates the vehicle's path.
- 狭い視野で前方に移動**: A vehicle moves forward in a narrow field of view. A yellow arrow indicates the path. A blue shaded area labeled "未観測領域" (Unobserved area) is shown to the left.
- 移動体を考慮し道路を横断**: A vehicle crosses a road at an intersection. A yellow arrow indicates the path. Labels include "リスク" (Risk), "目的地" (Destination), "始点" (Start point), and "交差点通過時に左右を確認" (Check left and right when crossing the intersection).

観測状態を考慮することで計画のみで動作を生成

自動車の自律化におけるロボティクスと学習

30

名古屋大学 情報学研究科 准教授 竹内 栄二郎

結果から得られた知見

- ・動作計画に観測状態を追加するという、シンプルな拡張で左右確認動作等の特徴的な動作を生成
 - ・各行動に必要な情報を定義(移動方向が自由空間)
 - ・各行動により得られる情報の予測
 - ・目的地へ到達する最短コストの経路探索
- ・経験的な解法(準最適解)でも良い
 - ・予測が完全ではないため最適解は最適ではない
 - ・計画に時間をかけるくらいなら動いて情報を得たほうがよい

自動車の自律化におけるロボティクスと学習

名古屋大学 情報学研究科 准教授 竹内 栄二郎

経路計画問題の深層学習による解法

- 動作計画問題は、評価値が最小となる動作群の組み合わせを求める問題。
 - 最適性の原理から、各状態における最適値を学習すればよい。
 - この状態で、この状態になりたい場合は、この動作をする

A*の結果を学習した例

自動車の自律化におけるロボティクスと学習

名古屋大学 情報学研究科 准教授 竹内 栄二郎

自律移動機能の学習への置き換え

- 物体認識
 - すでに多くの事例
- 位置推定
 - どの状態にあるかを認識する問題(特定物体認識と類似)
 - Place recognition
- 制御
 - ずれた状態と戻り方の対応付け(Nvidia End-to-End)
 - モデル予測制御の深層学習による高速化
- 経路計画問題
 - 最適性の原理
 - 現在状態と目的状態に対応する解群の圧縮

自動車の自律化におけるロボティクスと学習

名古屋大学 情報学研究科 准教授 竹内 栄二郎

車の自律化におけるロボティクスと学習

- ・物理・確率モデル等をもとにある種の最適化問題として解いてきたロボティクスは、ある程度の自律移動を実現
- ・自律移動に利用される機能の多くは、学習における最適化問題でも説明可能
　画像認識や音声認識のように学習によるロボティクスが性能を超えることも近い？

何が置き換えられ、学習のためにどのようなデータが必要であるかを、両側面から考えていくことが重要

実社会・実環境におけるロボットの機械学習

Preferred Networks, Inc. 高橋 城志



IBIS 2017, 11月9日(木):ワークショップ
実社会への機械学習の応用

実社会・実環境におけるロボットの機械学習

高橋 城志 (Kuniyuki Takahashi)
Preferred Networks, Inc.

Preferred Networks

実社会・実環境におけるロボットの機械学習

Preferred Networks, Inc. 高橋 城志

自己紹介:高橋城志 (Kuniyuki Takahashi)

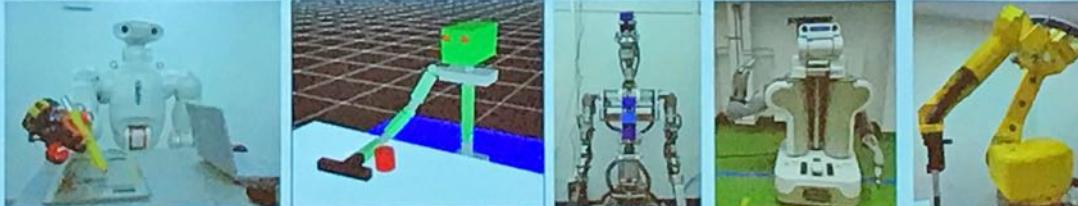
<https://sites.google.com/site/kuniyukitakahashi/home>

<経歴>

- 2007.4 - 2017.3
早稲田大学 総合機械工学科 (B.S., M.S., Ph.D.), 菅野研究室 & 尾形研研究室
- 2015.2 - 2016.3, 2017.2-2017.3
ミュンヘン工科大学, ドイツ; (external Ph.D.), Prof. Gordon
- 2017.4
Preferred Networks, Inc., リサーチャー

<研究内容>

- ロボットの運動生成のための機械学習 (RNN, deep learning)



Preferred Networks

2

実社会・実環境におけるロボットの機械学習

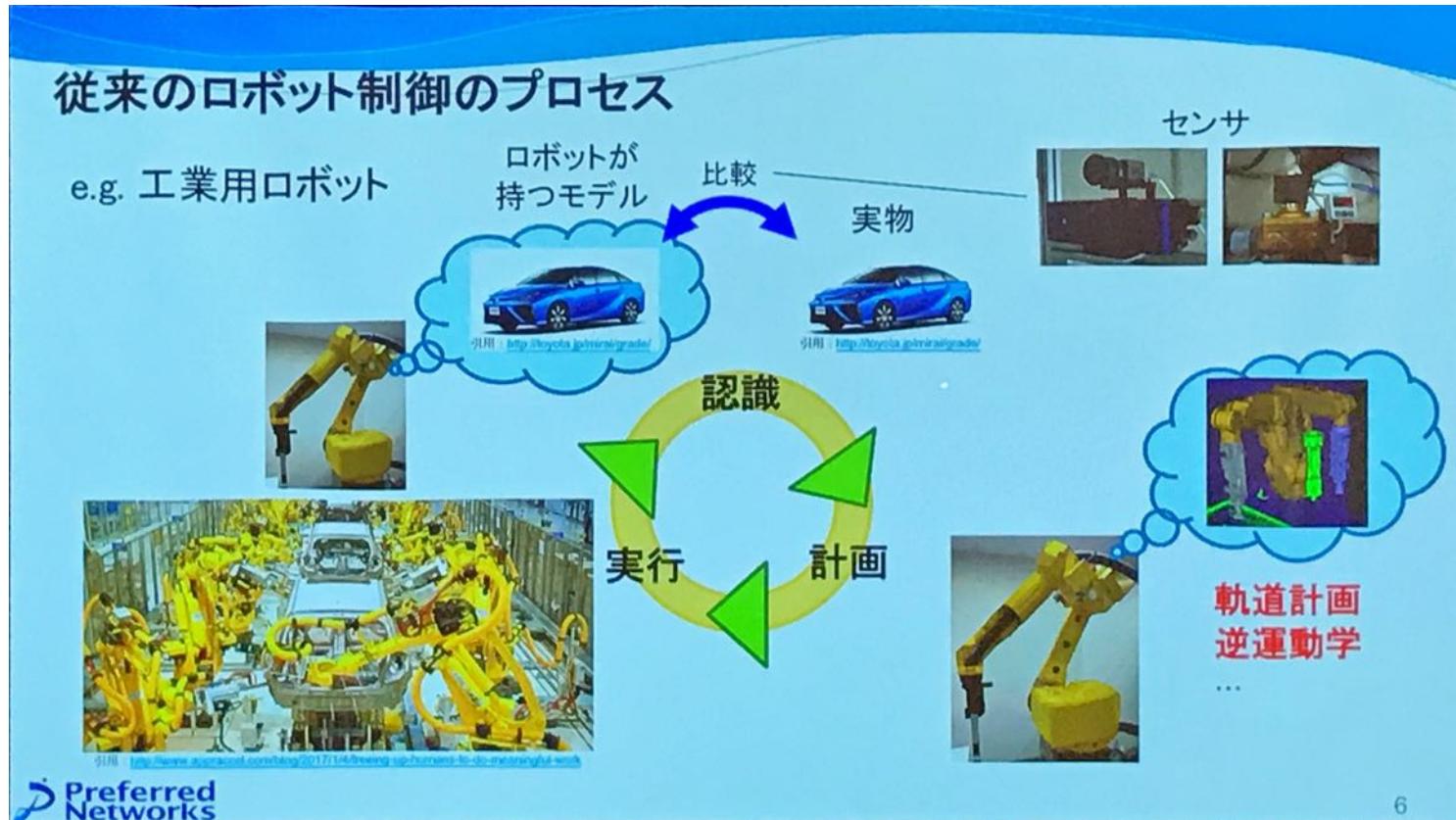
Preferred Networks, Inc. 高橋 城志

本日の内容

- 実社会・実環境ロボットにおける機械学習
 - これまでのロボット制御
 - 機械学習による認識を用いたロボット制御
 - 機械学習を用いたロボットの運動生成

実社会・実環境におけるロボットの機械学習

Preferred Networks, Inc. 高橋 城志



実社会・実環境におけるロボットの機械学習

Preferred Networks, Inc. 高橋 城志

物体操作 & 自由度(Degree of Freedom (DOF))

物体の位置を一意に決める情報:

- 物体の座標: x, y, z
- 物体の姿勢: α, β, γ

→ 物体を操作するための最小の自由度は6
 $(q_1, q_2, q_3, q_4, q_5, q_6)$

*必要とされる自由度はタスクに依存
e.g. 書字: 3自由度

Preferred Networks

7



実社会・実環境におけるロボットの機械学習

Preferred Networks, Inc. 高橋 城志

逆運動学(Inverse Kinematics)

関節角度
 (θ_1, θ_2) ?

手先座標
 (x, y)

θ_1, θ_2 : 余弦定理で計算可能

$$\theta_1 = \tan^{-1}(y/x) + \beta$$

$$\theta_2 = \pi - \alpha$$

$$\alpha = \cos^{-1}\left(\frac{l_1^2 + l_2^2 - x^2 - y^2}{2l_1l_2}\right)$$

$$\beta = \cos^{-1}\left(\frac{l_1^2 + x^2 + y^2 - l_2^2}{2l_1\sqrt{x^2 + y^2}}\right)$$

複数解

実社会・実環境におけるロボットの機械学習

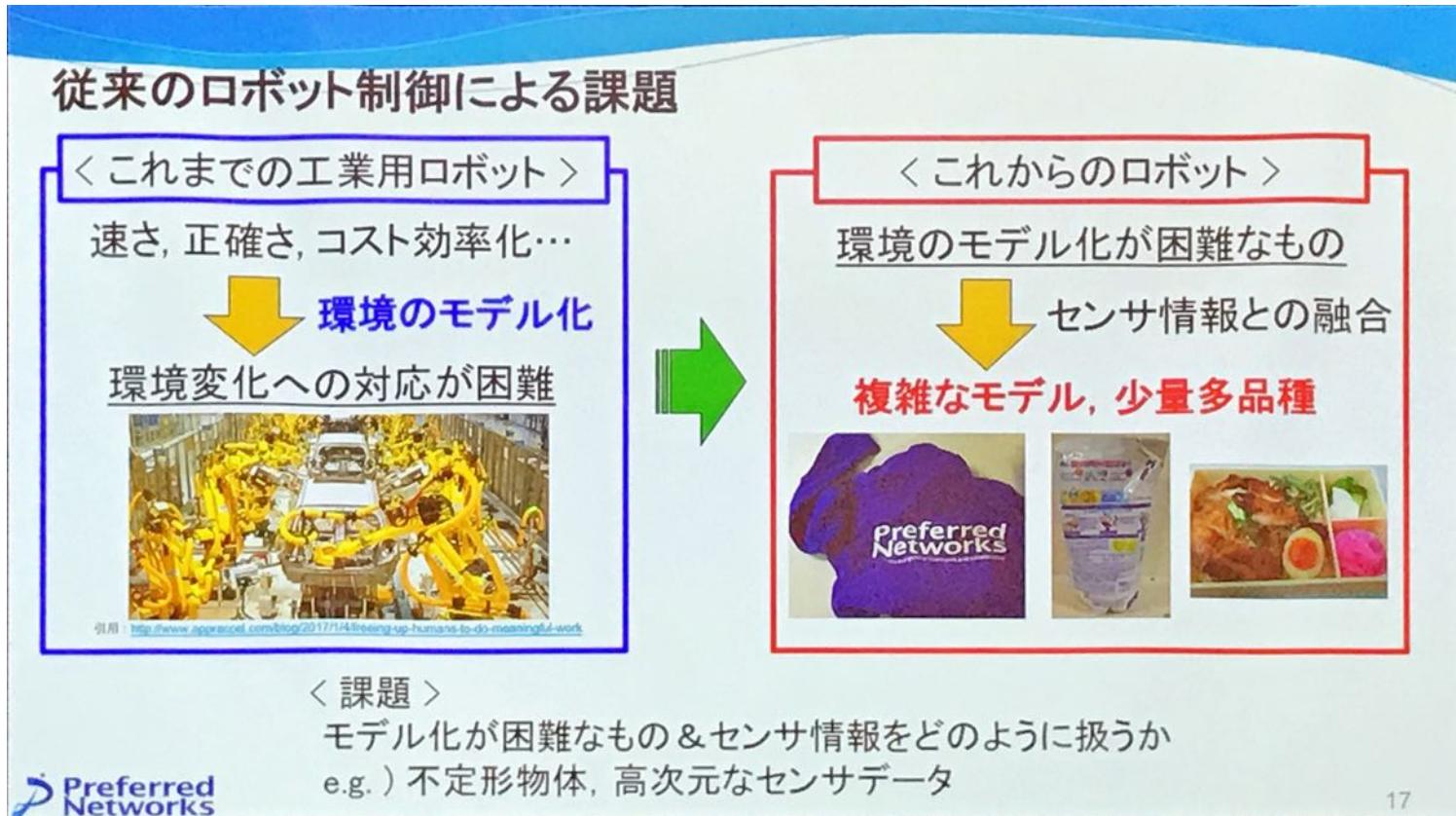
Preferred Networks, Inc. 高橋 城志



16

実社会・実環境におけるロボットの機械学習

Preferred Networks, Inc. 高橋 城志



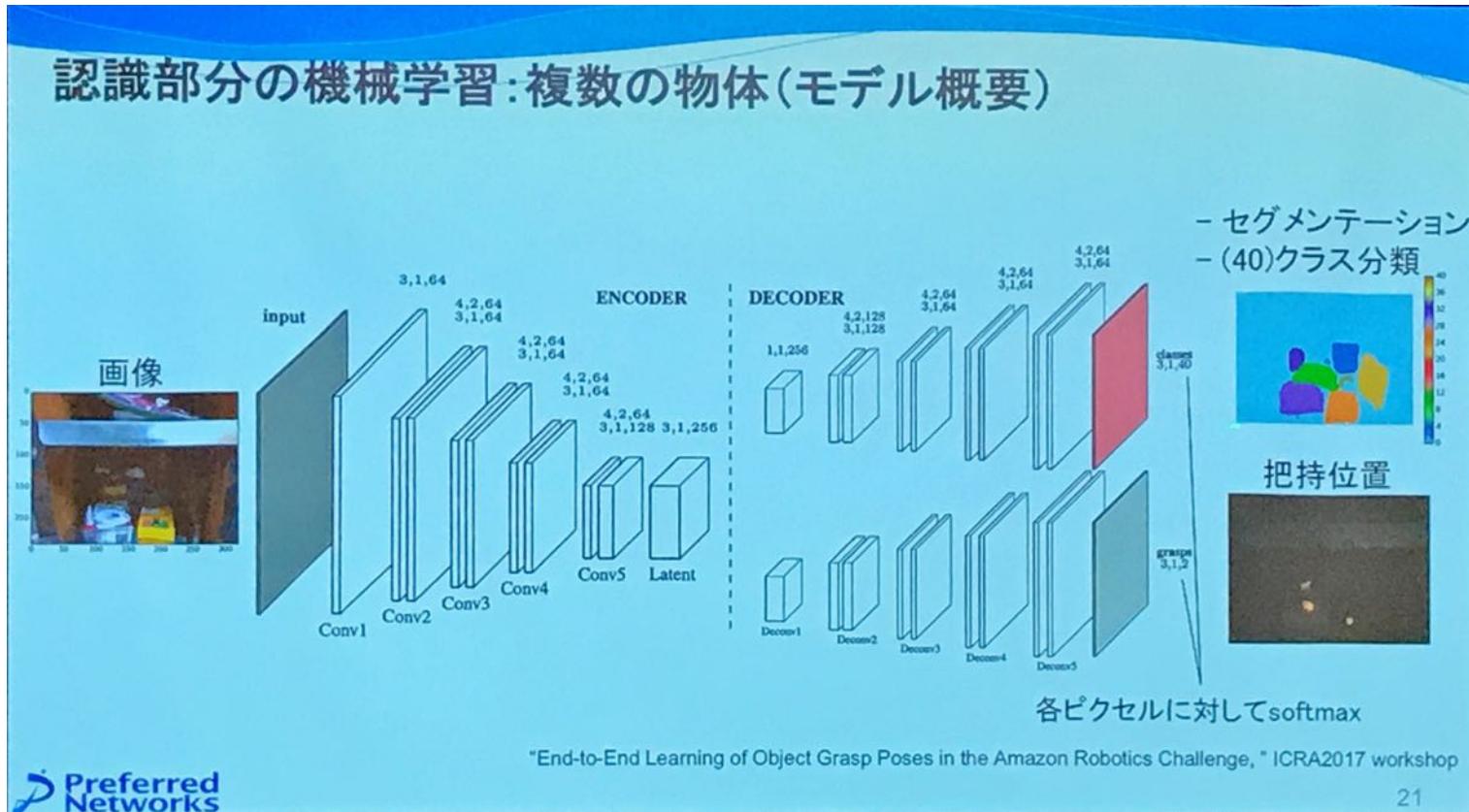
実社会・実環境におけるロボットの機械学習

Preferred Networks, Inc. 高橋 城志



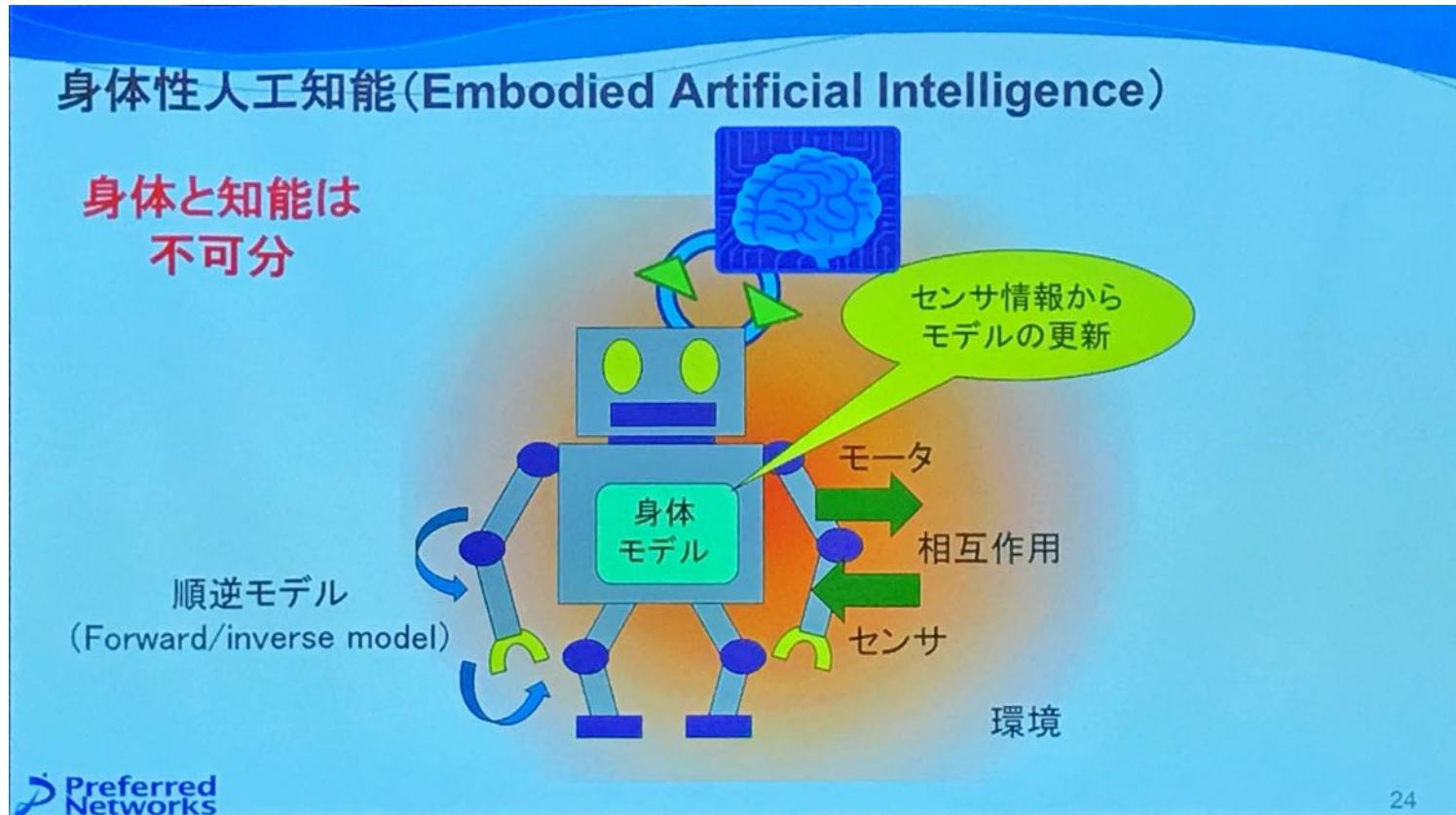
実社会・実環境におけるロボットの機械学習

Preferred Networks, Inc. 高橋 城志



実社会・実環境におけるロボットの機械学習

Preferred Networks, Inc. 高橋 城志



実社会・実環境におけるロボットの機械学習

Preferred Networks, Inc. 高橋 城志

身体性人工知能(Embodied Artificial Intelligence)

Subsumption architecture^[1] (行動規範型A. I.)

e.g. Roomba

The diagram shows a Roomba robot on a wooden floor. A yellow arrow points from the robot to a circular model of a robot. The circular model has two blue squares labeled 'モータ' (Motors) at the bottom, connected by a line. Two yellow rectangles labeled 'センサ' (Sensors) are positioned above the circle. Red lines connect the sensors to the center of the circle, representing sensor input. The word '単純化' (Simplification) is written next to the arrow.

Credit: iRobot Corporation

[1] R. Brooks, "A robust layered control system for a mobile robot," 1986.

25

実社会・実環境におけるロボットの機械学習

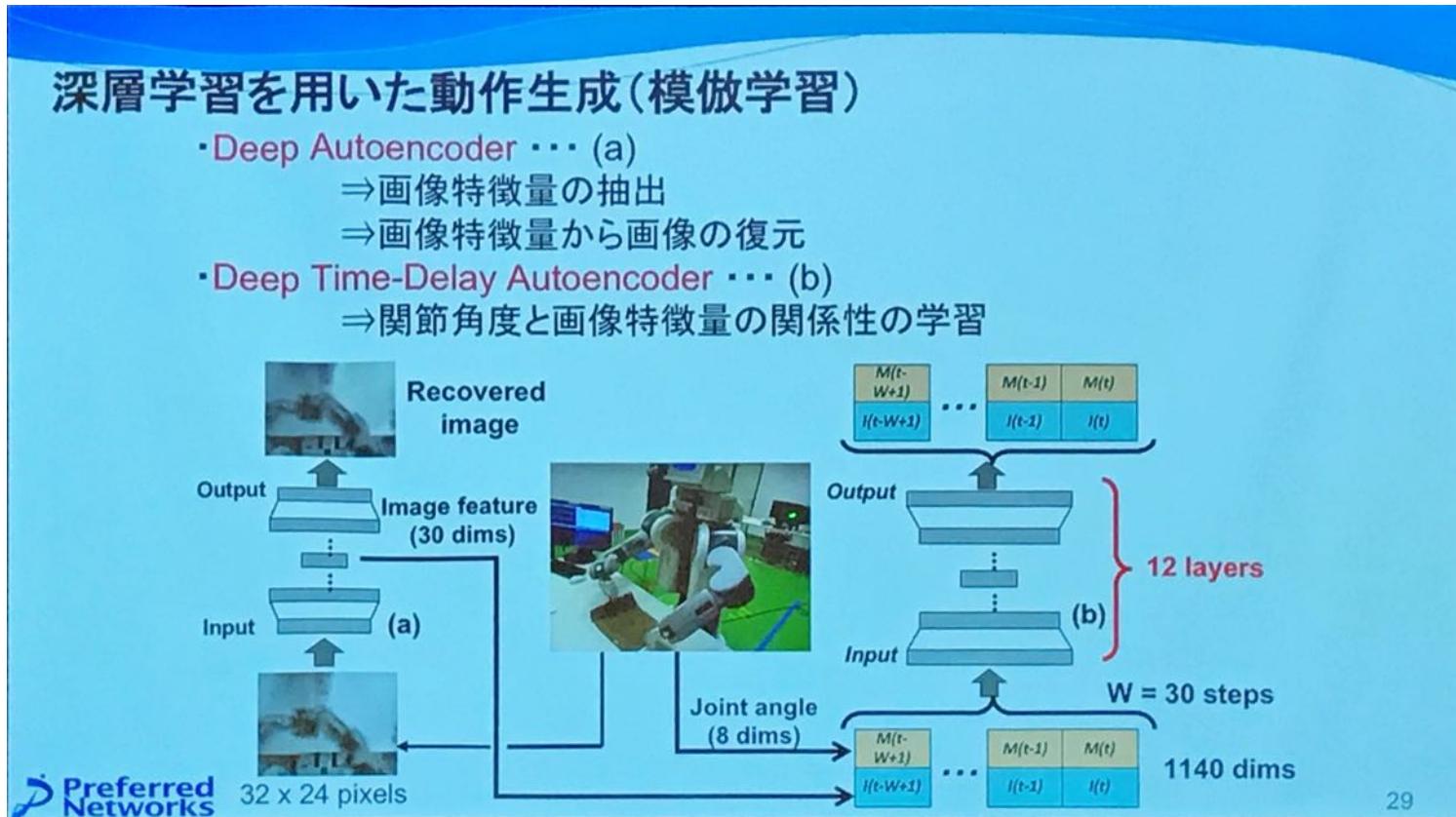
Preferred Networks, Inc. 高橋 城志

機械学習をロボットに適応するときの課題

- 方策を獲得するまでの**試行回数の多さ**
->実際のロボットで行うと壊れる
- シミュレーションと実機との**差異**
->シミュレーション後の再学習の必要性
- 未知な環境への**低適応性**
->再学習では即時対応が困難

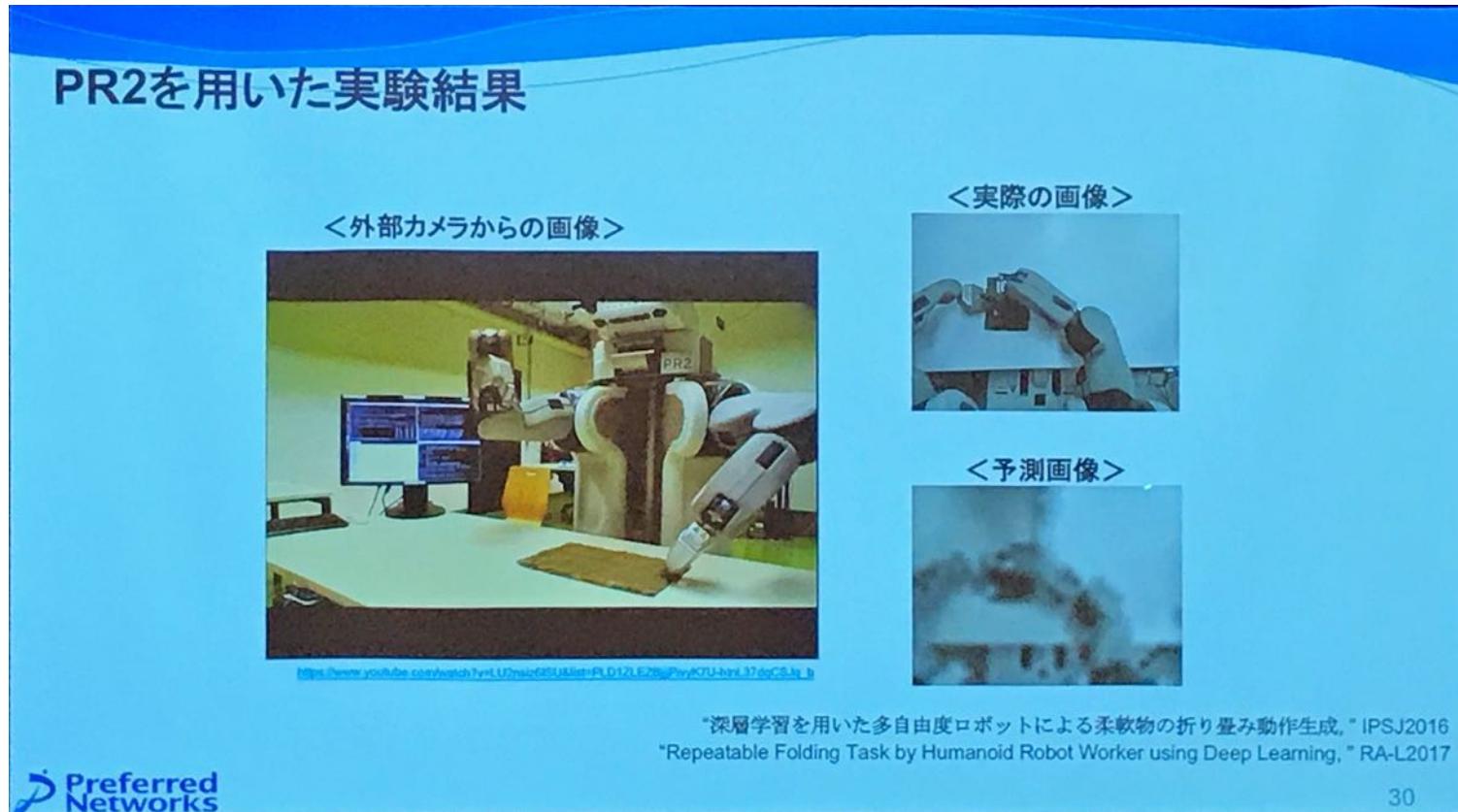
実社会・実環境におけるロボットの機械学習

Preferred Networks, Inc. 高橋 城志



実社会・実環境におけるロボットの機械学習

Preferred Networks, Inc. 高橋 城志

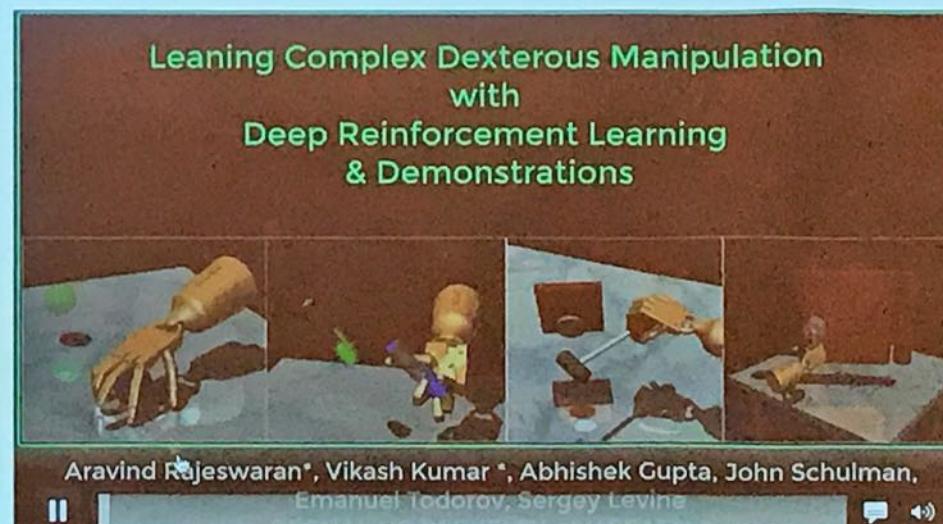


実社会・実環境におけるロボットの機械学習

Preferred Networks, Inc. 高橋 城志

試行回数の削減(教師あり+強化学習)

人の実演データから次の動作を真似るポリシーを学習し、そこから強化学習すると学習時間と試行回数を大幅に削減



Learning Complex Dexterous Manipulation
with
Deep Reinforcement Learning
& Demonstrations

Aravind Rajeswaran*, Vikash Kumar *, Abhishek Gupta, John Schulman,
Emanuel Todorov, Sergey Levine

https://www.youtube.com/watch?v=jIBii8l_OM

"Learning Complex Dexterous Manipulation with Deep Reinforcement Learning and Demonstrations," 2017
<https://arxiv.org/abs/1709.10087>

Preferred Networks

32

実社会・実環境におけるロボットの機械学習

Preferred Networks, Inc. 高橋 城志

シミュレーションと実験との差異の低減

シミュレーション時に画像の色情報を複数変更することで、実世界でも対応

課題：
環境との相互作用を行うようなものはシミュレーションとの差異が顕著に出る
e.g. 把持、二足歩行

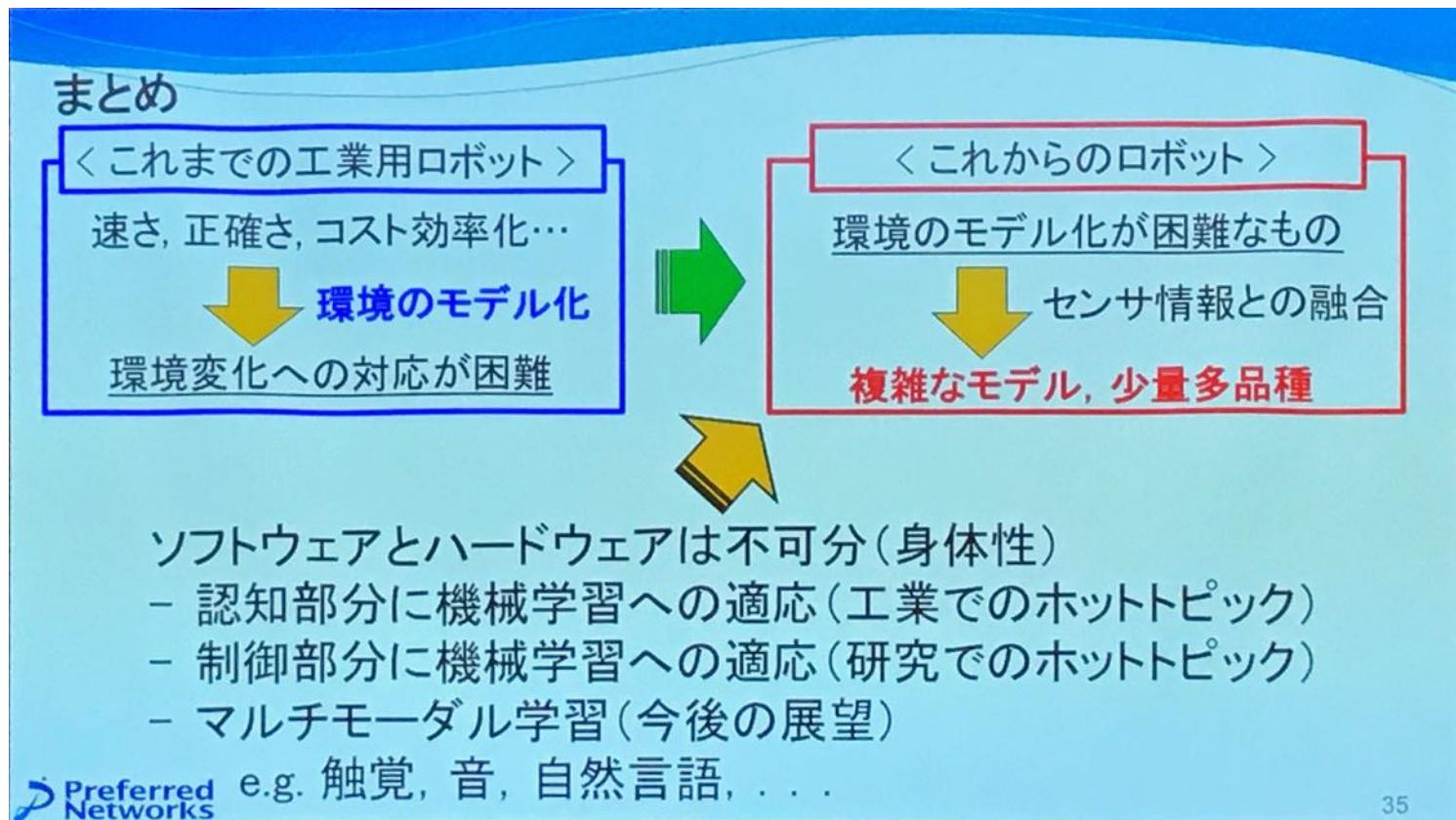
Training	Test
...	

"Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World," IROS2017

33

実社会・実環境におけるロボットの機械学習

Preferred Networks, Inc. 高橋 城志



実社会・実環境におけるロボットの機械学習

Preferred Networks, Inc. 高橋 城志

今後の展望:マルチモーダル学習

e.g. 自然言語 + ロボット

曖昧な言語指示に対して、
聞き返しを行うアプローチ

Human:
hey move that brownish fluffy
stuff to the lower right bin.

Robot:
which one?
(two objects highlighted in [])

Human:
the one next to the green and
blue box.

Robot:
I got it.
(one object highlighted in [])

<https://www.youtube.com/watch?v=Uvv1XIIjghk&t=19s>
<https://arxiv.org/abs/1710.06280>
Under review in ICRA2017

PFN

Preferred Networks

Thank you!! 36

実社会・実環境におけるロボットの機械学習

Preferred Networks, Inc. 高橋 城志

今後の展望:マルチモーダル学習

e.g. 自然言語 + ロボット

曖昧な言語指示に対して、
聞き返しを行うアプローチ

Human:
hey move that brownish fluffy
stuff to the lower right bin.

Robot:
which one?
(two objects highlighted in [])

Human:
the one next to the green and
blue box.

Robot:
I got it.
(one object highlighted in [])

<https://www.youtube.com/watch?v=Uvv1XIIjghk&t=19s>
<https://arxiv.org/abs/1710.06280>
Under review in ICRA2017

PFN

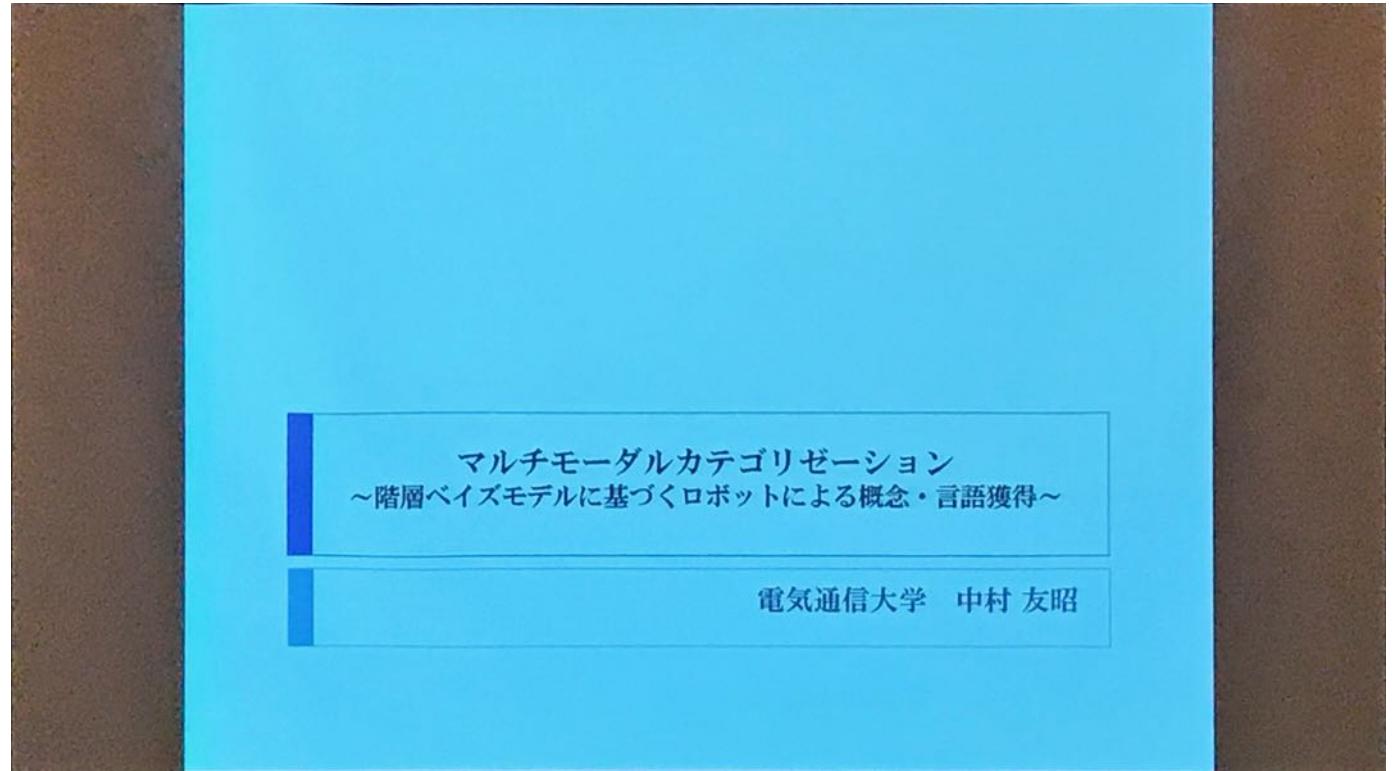
Preferred Networks

Thank you!! 36

階層ベイズモデルに基づくロボットによる概念・言語獲得

54

電機通信大学 中村 友昭



階層ベイズモデルに基づくロボットによる概念・言語獲得

55

電機通信大学 中村 友昭

中村 友昭

- ▶ 所属:電気通信大学 助教
- ▶ 研究内容:
 - ▶ 知能ロボティクス, 知能システム
 - ▶ 自律移動ロボット(ロボカップ@ホーム)
 - ▶ 言語理解ロボット
 - ▶ 概念・言語学習ロボット

A photograph of a robotics competition booth at the DiGO RoboCup 2010 World Cup Camp in Singapore. A banner in the background reads "DiGO RoboCup 2010 World Cup Camp Singapore".

A screenshot of a video showing a robot interacting with a vending machine. The robot is asking for a drink, and the text on the screen says "ロボット: ドリンクをホワイトシェルフから貰ってきます。いいですか?" (Robot: I will get a drink from the white shelf. Is it okay?).

A screenshot of a video titled "The Mobile Manipulation Experiment" showing a robot arm performing tasks on a cluttered table. The video has two zoom levels: ×10 and ×400.

A screenshot of a video showing a 3D point cloud visualization of an object recognition task. The interface includes labels for "対象物体" (Target Object), "分離精度" (Separation Accuracy), and "誤検出" (False Positives).

階層ベイズモデルに基づくロボットによる概念・言語獲得

56

電機通信大学 中村 友昭

概念・言語学習

- ▶ 人のように言語を獲得するロボットの実現
 - ▶ 人や環境とのインタラクションにより自律的に獲得
 - ▶ 概念形成
 - ▶ 語彙の獲得(単語辞書)
 - ▶ 語意の獲得(記号接地)
 - ▶ 文法の獲得

音声認識(語彙)

これわぼーるだよ

概念形成

接地

言語的知識
(語彙・文法)

- ▶ 言語獲得アルゴリズムを確率モデルを用いて実現
 - ▶ 人間のような知能の実現
 - ▶ 人間の言語獲得過程の解明

▶ 3

階層ベイズモデルに基づくロボットによる概念・言語獲得

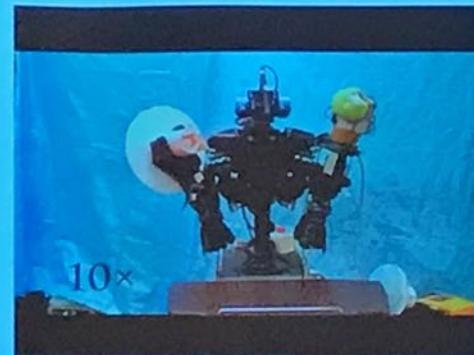
57

電機通信大学 中村 友昭

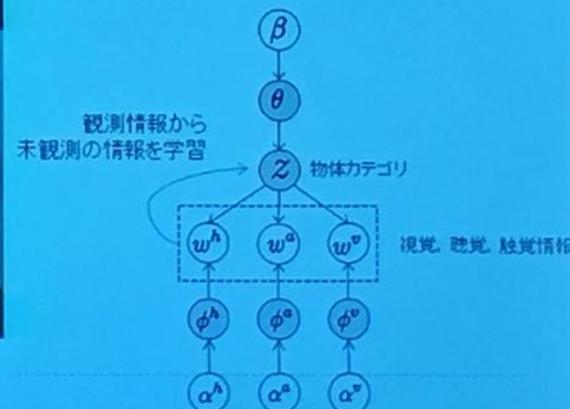
概念形成

▶ 概念の工学的定義

- ▶ 概念=知覚情報のクラスタリングによって形成されたカテゴリ
- ▶ 知覚情報: ロボットの視覚、聴覚、触覚情報
- ▶ これらの情報を確率モデルにより教師なしで分類



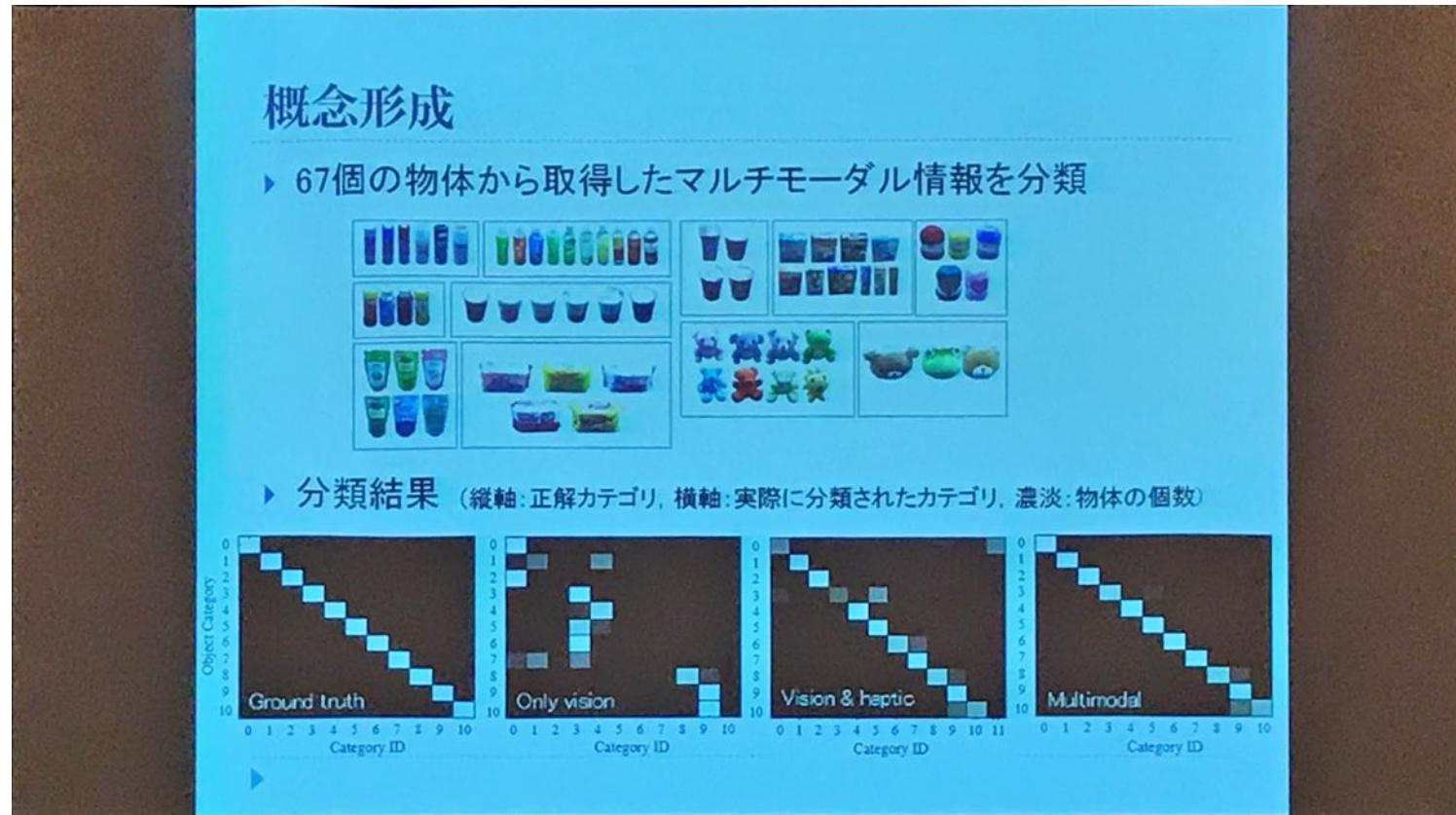
物体に見て・触ることで物体の概念を形成



階層ベイズモデルに基づくロボットによる概念・言語獲得

58

電機通信大学 中村 友昭



階層ベイズモデルに基づくロボットによる概念・言語獲得

59

電機通信大学 中村 友昭

語意の獲得

- マルチモーダル情報として単語を追加
- 単語も含めたマルチモーダルな概念を形成

それは
ペットボトルだよ

$P(w^v, w^a, w^t | w^w)$
マルチモーダル情報 単語

単語を"感覚的"に理解可能
⇒ 単語の意味の理解

階層ベイズモデルに基づくロボットによる概念・言語獲得

60

電機通信大学 中村 友昭

語彙の獲得

- ▶ 前の研究では語彙を持っていることが前提
 - ▶ 語彙=音声認識・単語分割で使用する単語辞書(言語モデル)
- ▶ 語彙を獲得する際の問題
 - ▶ 語彙を持たないため
音声が正しく認識できない
 - ▶ ぼーる? ぼーう? ごーる?
 - ▶ 単語の切れ目が分からず
ぼーる? わぼー? ぼーるだ? るだよ?
- ▶ これらの問題を言語的知識と概念を相互学習することで解決
 - ▶ 言語のパターンに基づく単語分割
 - ▶ 同じ概念に含まれる物体には同じ単語が教示される可能性が高い

これわぼーるだよ

これわぼーるだよ

ぼーるがあるよ

これわ | ぼーる | だよ

言語的知識

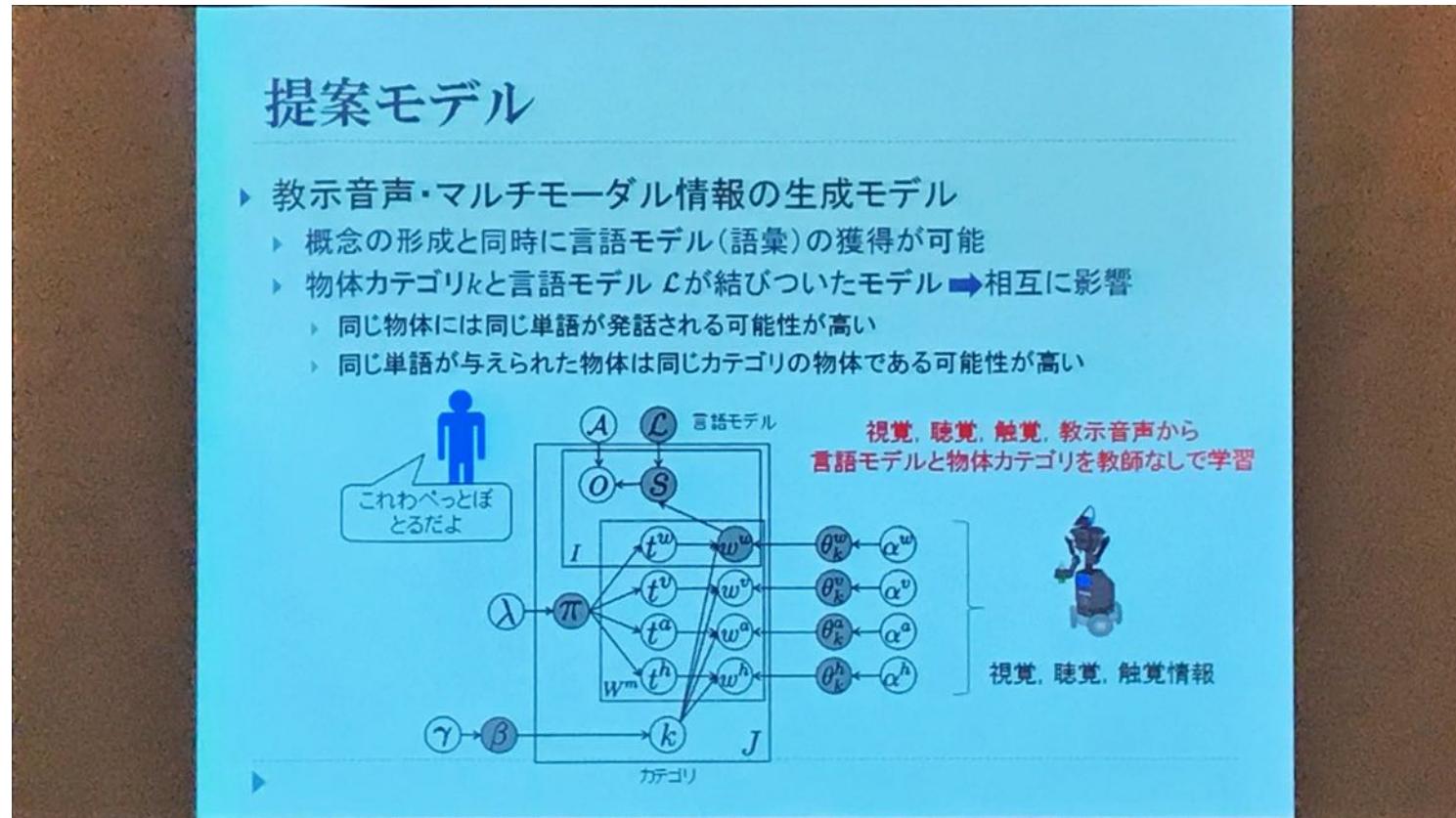
相互に学習

概念形成

階層ベイズモデルに基づくロボットによる概念・言語獲得

61

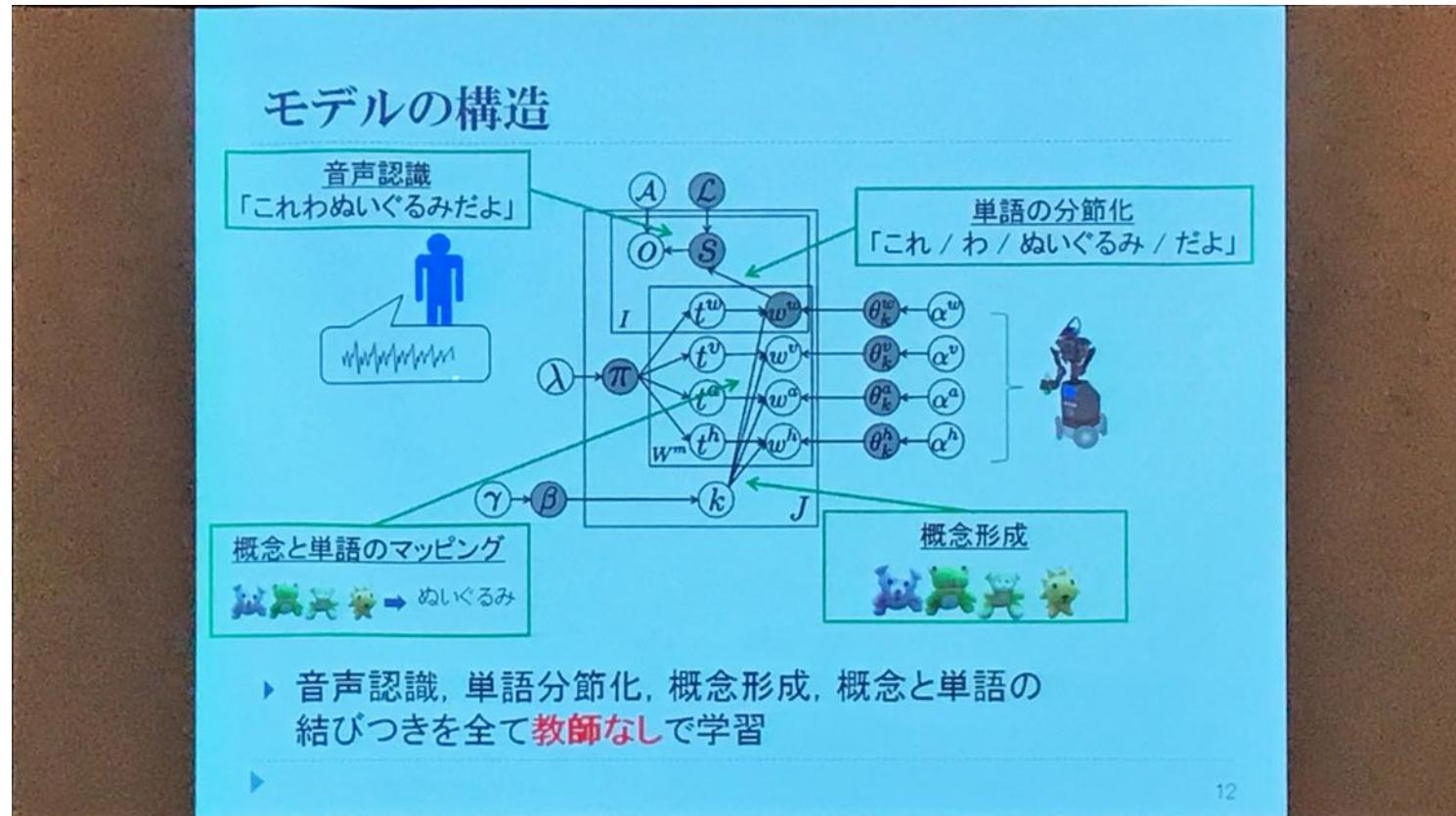
電機通信大学 中村 友昭



階層ベイズモデルに基づくロボットによる概念・言語獲得

62

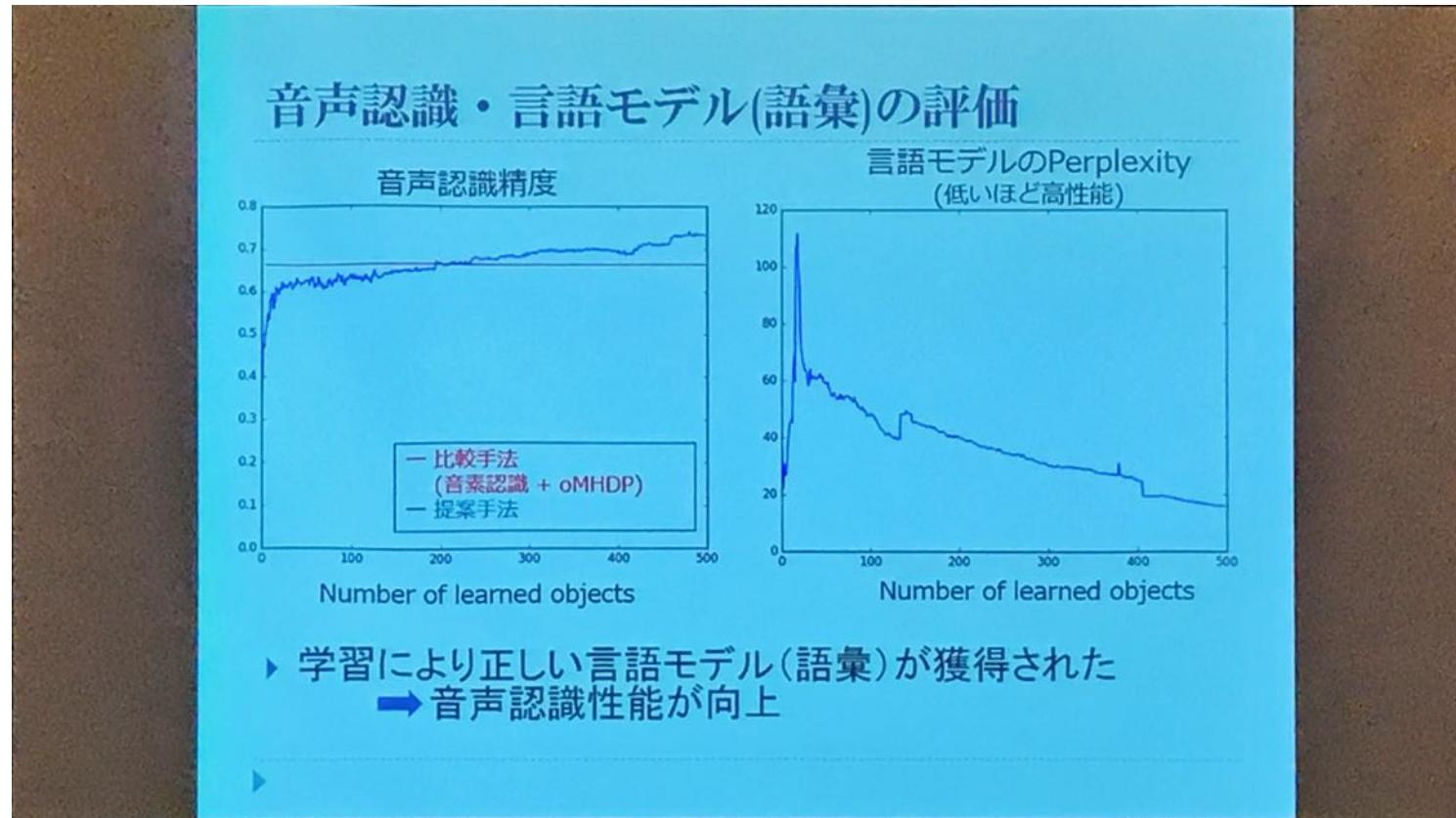
電機通信大学 中村 友昭



階層ベイズモデルに基づくロボットによる概念・言語獲得

63

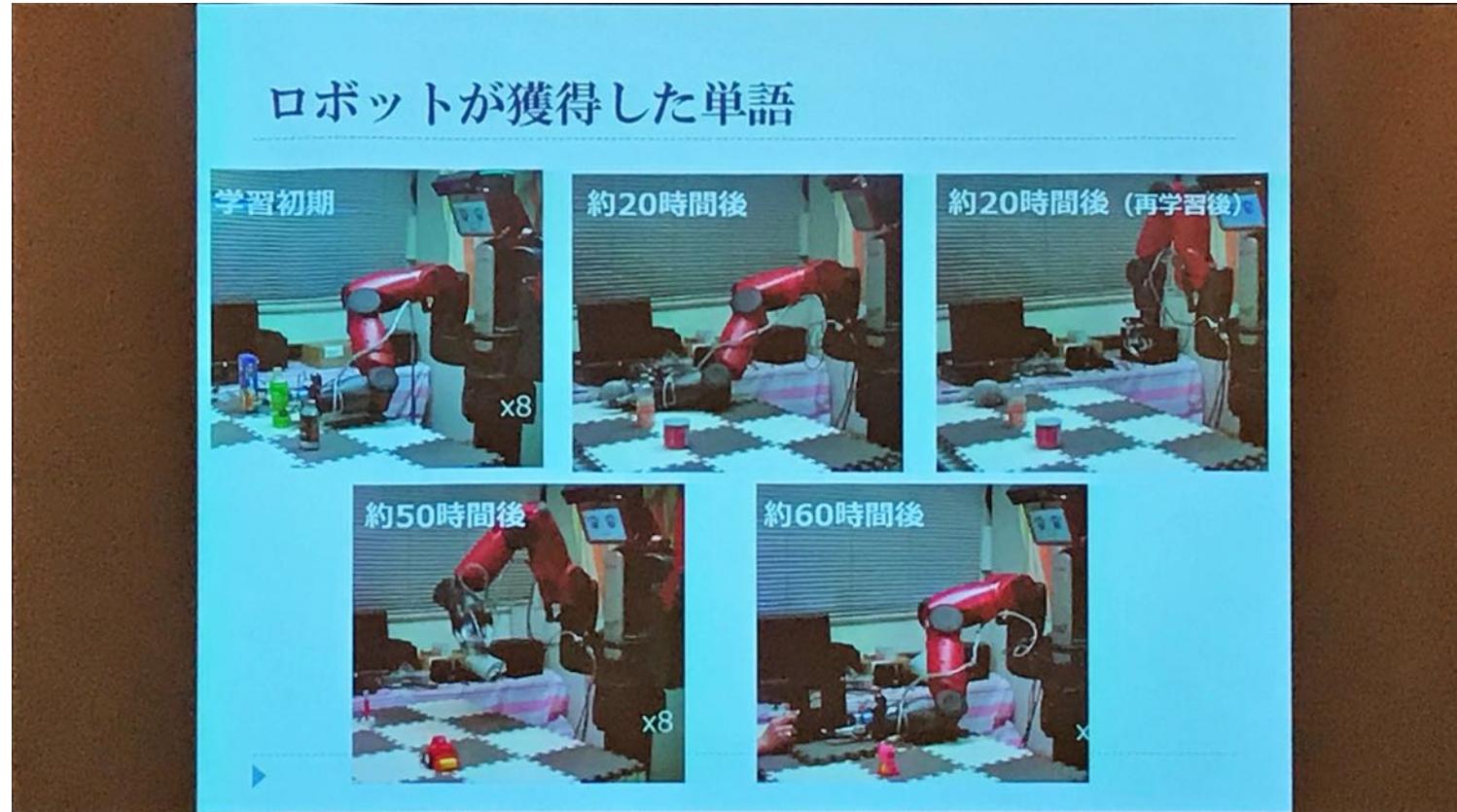
電機通信大学 中村 友昭



階層ベイズモデルに基づくロボットによる概念・言語獲得

64

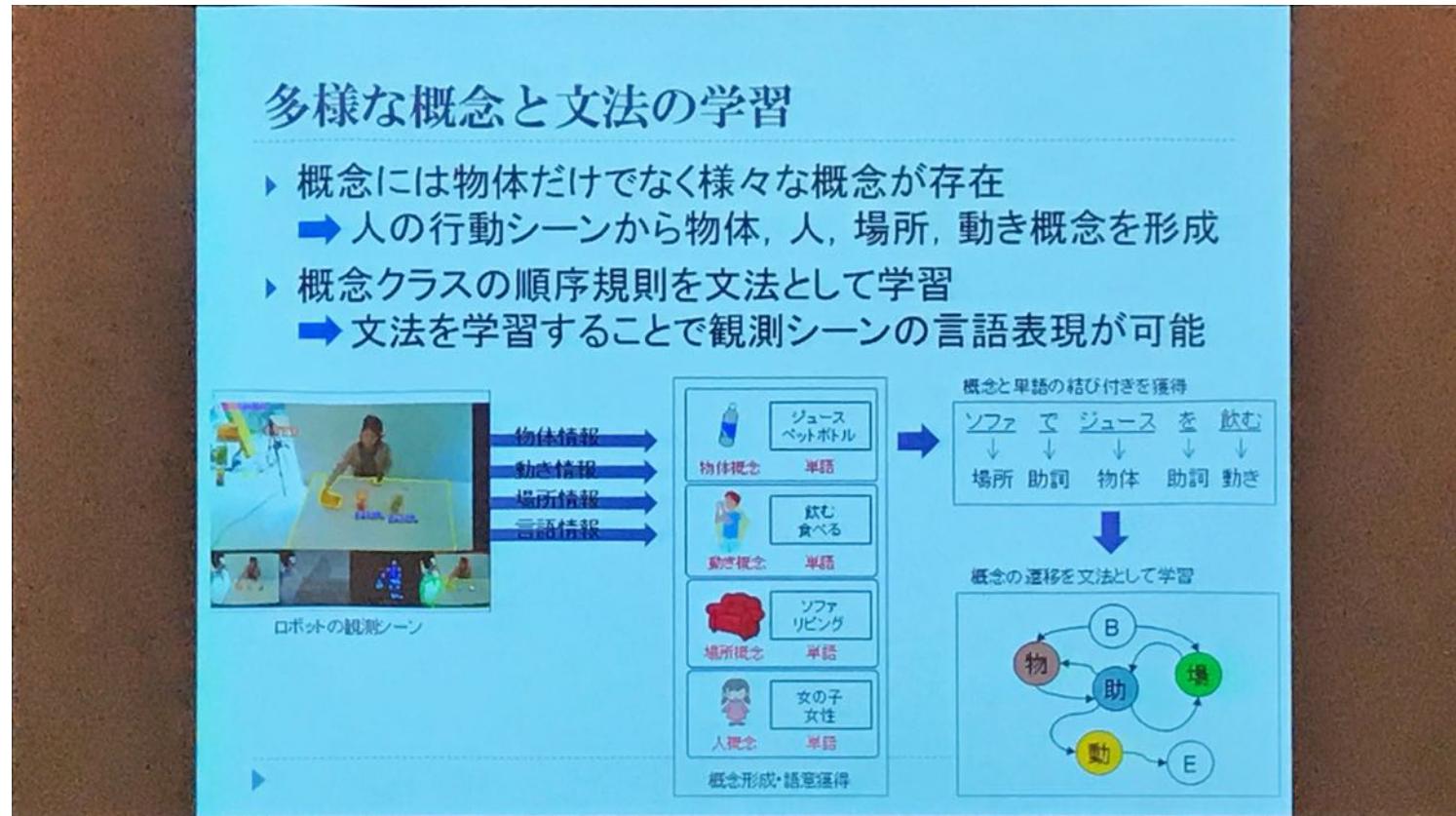
電機通信大学 中村 友昭



階層ベイズモデルに基づくロボットによる概念・言語獲得

65

電機通信大学 中村 友昭



階層ベイズモデルに基づくロボットによる概念・言語獲得

66

電機通信大学 中村 友昭

複数概念モデルと文法モデル

- ▶ 複数概念モデル(mMLDA)
 - ▶ 場所, 動き, 物体, 人概念の生成モデル
 - ▶ ロボットが観測した情報から概念を教師なしで学習可能
 - ▶ 単語と概念の結び付きも学習
- ▶ 文法モデル(HSMM)
 - ▶ 概念の結び付きから概念クラスの遷移順を文法として学習
 - ▶ 概念モデルの結果を初期値として教師なし学習
- ▶ 文法と概念を相互に学習
 - ▶ 言語的制約と知覚情報との共起性に基づく単語の接地

The diagram illustrates the hierarchical Bayesian model (mMLDA) and the HSMM model. At the top, a '総合概念' (General Concept) node α influences four latent concept nodes z^L, z^M, z^O, z^P , which in turn influence word nodes $w^{iL}, w^{iM}, w^{iO}, w^{iP}$ through parameters $\theta^L, \theta^M, \theta^O, \theta^P$. Below this, the model is divided into four categories: 地点概念 (Place Concept), 動き概念 (Movement Concept), 物体概念 (Object Concept), and 人概念 (Person Concept). Each category contains a set of nodes: $\phi^{iL}, \beta^{iL}, \alpha^{iL}$ for place, $\phi^{iM}, \beta^{iM}, \alpha^{iM}$ for movement, $\phi^{iO}, \beta^{iO}, \alpha^{iO}$ for object, and $\phi^{iP}, \beta^{iP}, \alpha^{iP}$ for person. At the bottom, a sequence of words is shown: B (始) -> ソファ (sofa) -> で (in) -> お茶 (tea) -> を (to) -> 飲む (drink). A double-headed arrow between the two main sections indicates the interaction between linguistic constraints and perceptual information.

階層ベイズモデルに基づくロボットによる概念・言語獲得

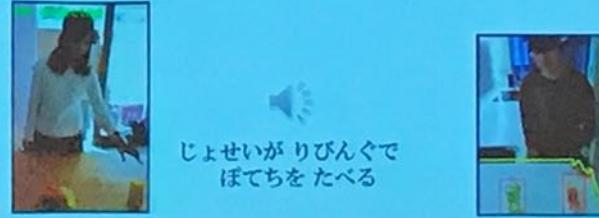
67

電機通信大学 中村 友昭

文生成

- ▶ 獲得した概念・文法からシーンを説明する分を生成
- ▶ 生成文の評価結果
 - ▶ 文法と概念の相互学習により精度が向上
- ▶ 生成例

評価基準		比較手法 (相互学習なし)	提案手法 (相互学習あり)
文法	意味		
正	正	16.06%	69.23%
正	誤	78.73%	23.76%
誤	正	0.91%	1.31%
誤	誤	4.30%	5.66%



じょせいがりびんぐで
ぼてちをたべる

せいねんがきっちんで
そそぐ

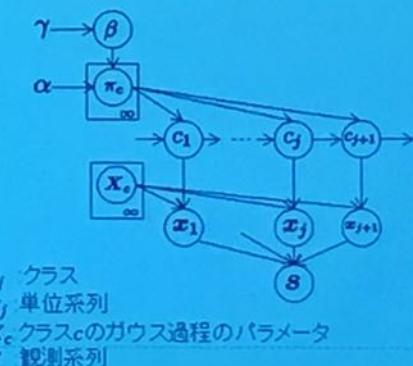
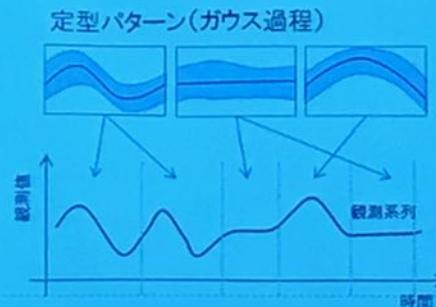
階層ベイズモデルに基づくロボットによる概念・言語獲得

68

電機通信大学 中村 友昭

提案モデル

- 時系列データはガウス過程(GP)を出力分布とする隠れセミマルコフモデル(HSMM)によって生成されると仮定
 - ガウス過程: 分節化された定型パターンを表現
 - HSMM: 定型パターンの長さも隠れ変数としたHMM
- HSMMとGPのパラメータ推定することで、連続的な情報の分節・分類が可能



階層ベイズモデルに基づくロボットによる概念・言語獲得

69

電機通信大学 中村 友昭

ガウス過程

- 時刻 t の出力 x を予測する回帰関数の確率モデル
- 学習データ x が与えられると各時刻 t' での予測分布がガウス分布で得られる
$$p(x'|t', x, t) \propto N(x|k^T C^{-1} i, c - k^T C^{-1} k)$$
- 軌道の一致度、観測値の曖昧性を確率で評価可能

学習データ点
予測分布

学習データ点
予測分布

学習データ点
予測分布

学習データ点
予測分布

- ガウス過程で時系列情報の定型パターンを表現

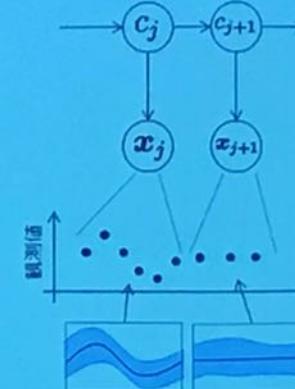
階層ベイズモデルに基づくロボットによる概念・言語獲得

70

電機通信大学 中村 友昭

GP-HSMMのパラメータ推定

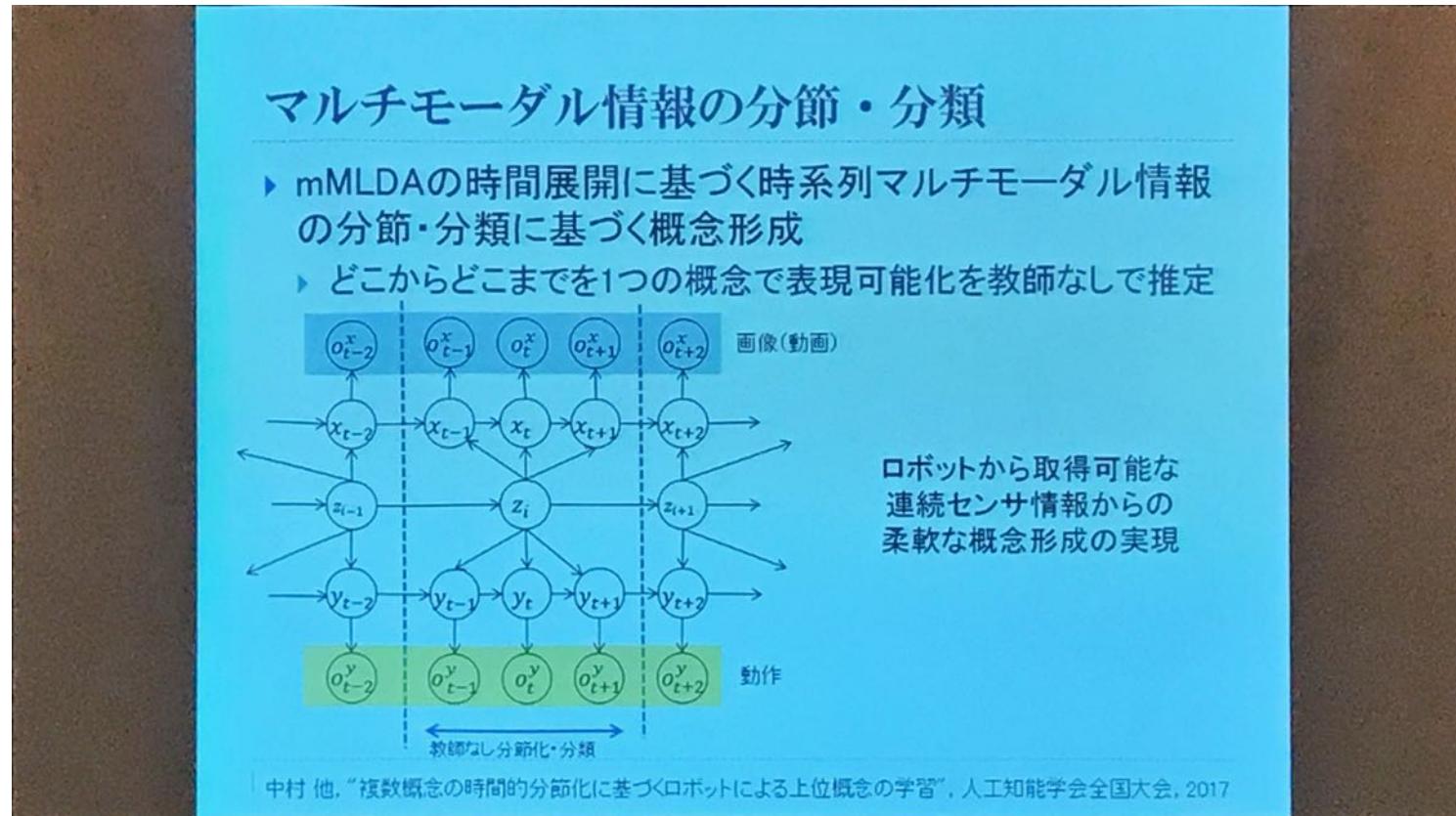
- ▶ HSMMでは1つの状態に分類される
系列長は状態によって異なる
→ 系列長も推定する必要がある
- ▶ パラメータ推定
 - ▶ 時刻tのデータ点を終点とした長さkの
系列のクラスがcである確率から
サンプリングすることで決定
 - ▶ あらゆるkとcの組み合わせの確率を計算する必要がある
= 動的計画法(Forward filtering–Backward sampling)を利用



階層ベイズモデルに基づくロボットによる概念・言語獲得

71

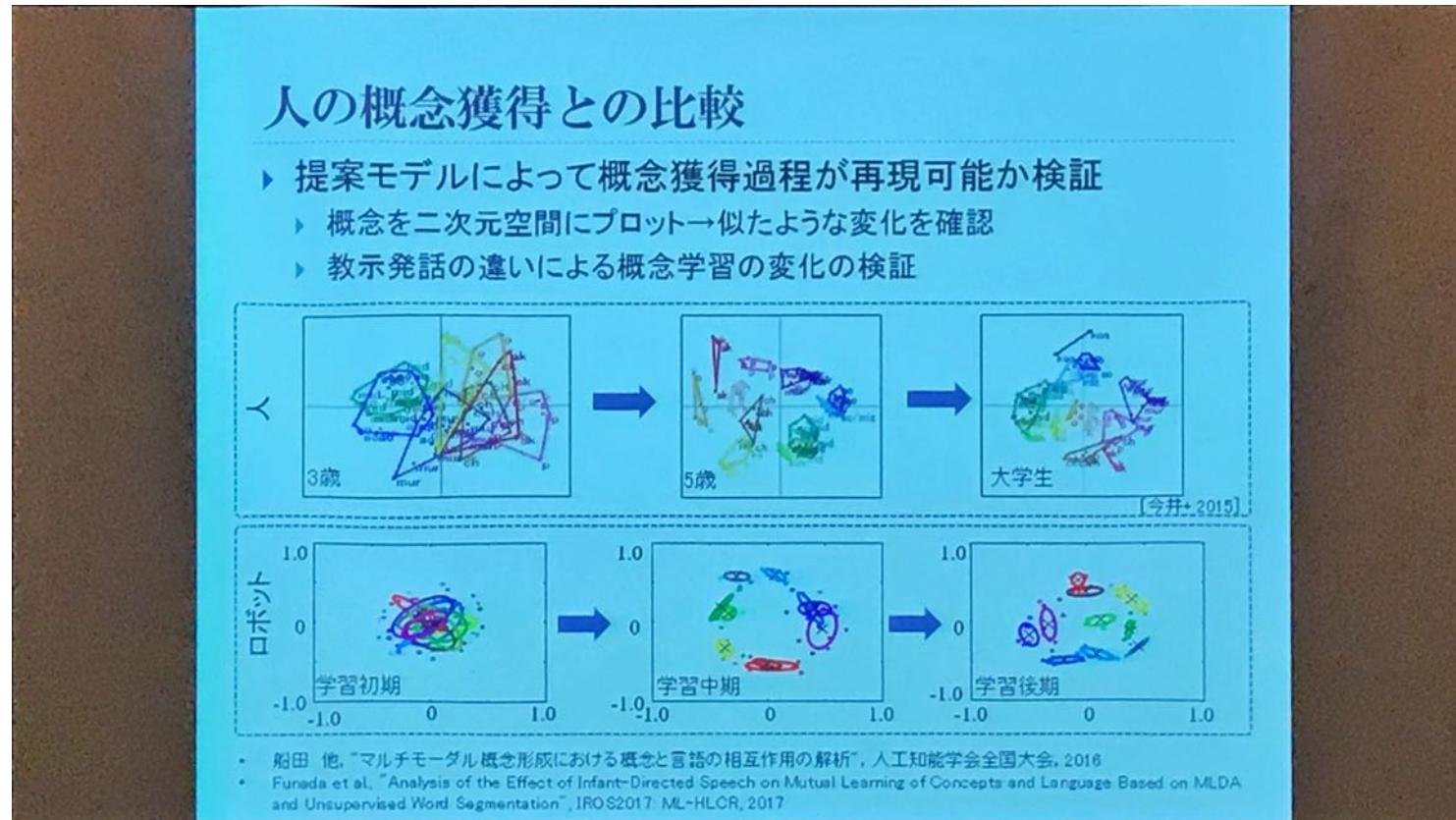
電機通信大学 中村 友昭



階層ベイズモデルに基づくロボットによる概念・言語獲得

72

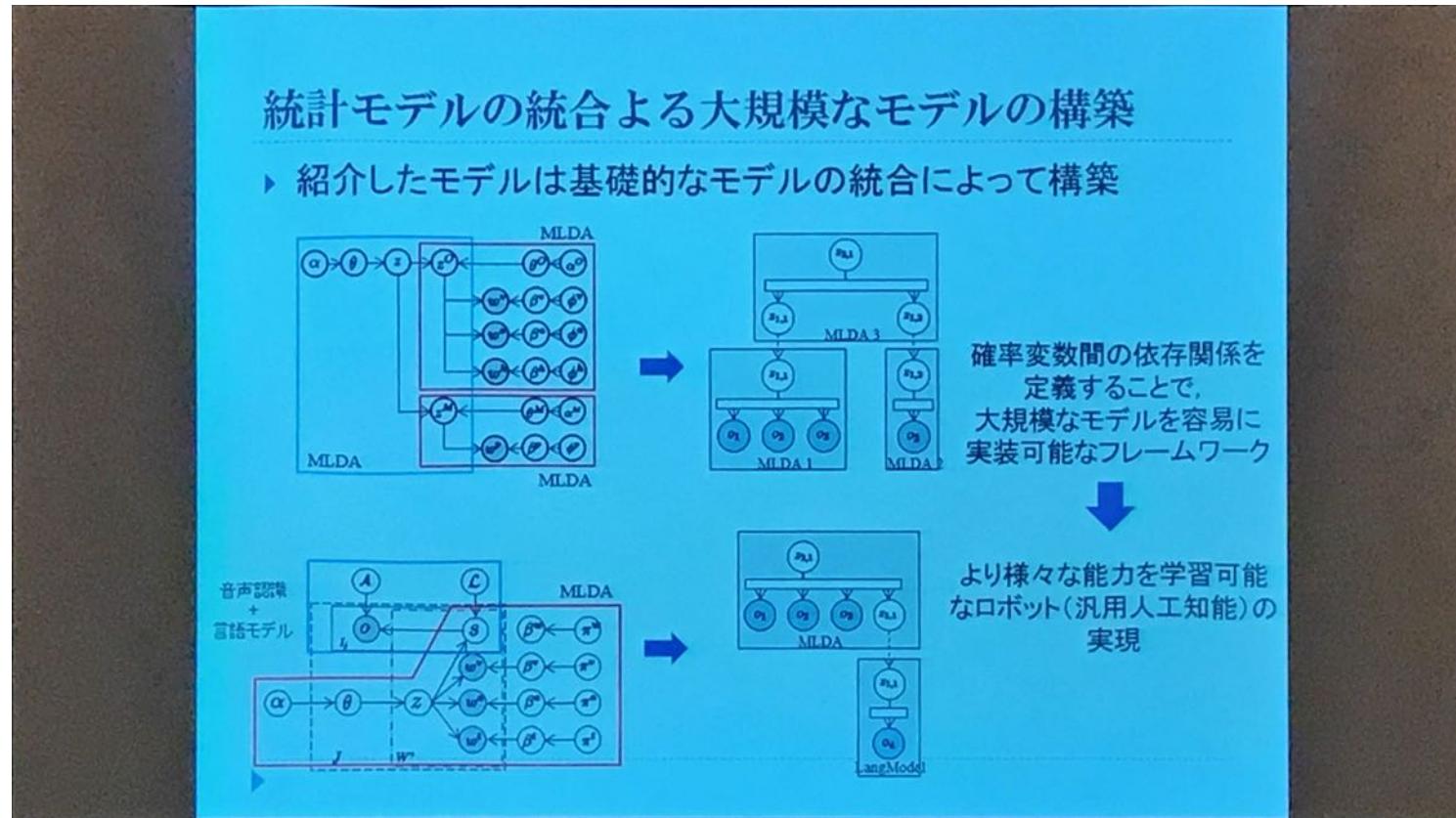
電機通信大学 中村 友昭



階層ベイズモデルに基づくロボットによる概念・言語獲得

73

電機通信大学 中村 友昭



階層ベイズモデルに基づくロボットによる概念・言語獲得

74

電機通信大学 中村 友昭

まとめ

- ▶ 本発表ではロボットによる概念・語彙・語意・文法獲得に関する研究を紹介
 - ▶ ロボットが取得したマルチモーダルな情報から教師なしでボトムアップに概念を形成
 - ▶ 人の教示発話から語彙・文法を学習
 - ▶ 単語と概念を結びつけることで単語の意味の学習
 - ▶ 単語と結びついた概念の遷移規則を文法として学習
 - ▶ 獲得した言語の情報がトップダウンに作用することでより人の感覚に近い概念の形成が可能

画像処理への機械学習の応用

IBIS2017 企画セッション

9:30-11:30 企画セッション

画像処理への機械学習の応用

画像処理への機械学習の応用

IBIS2017 企画セッション

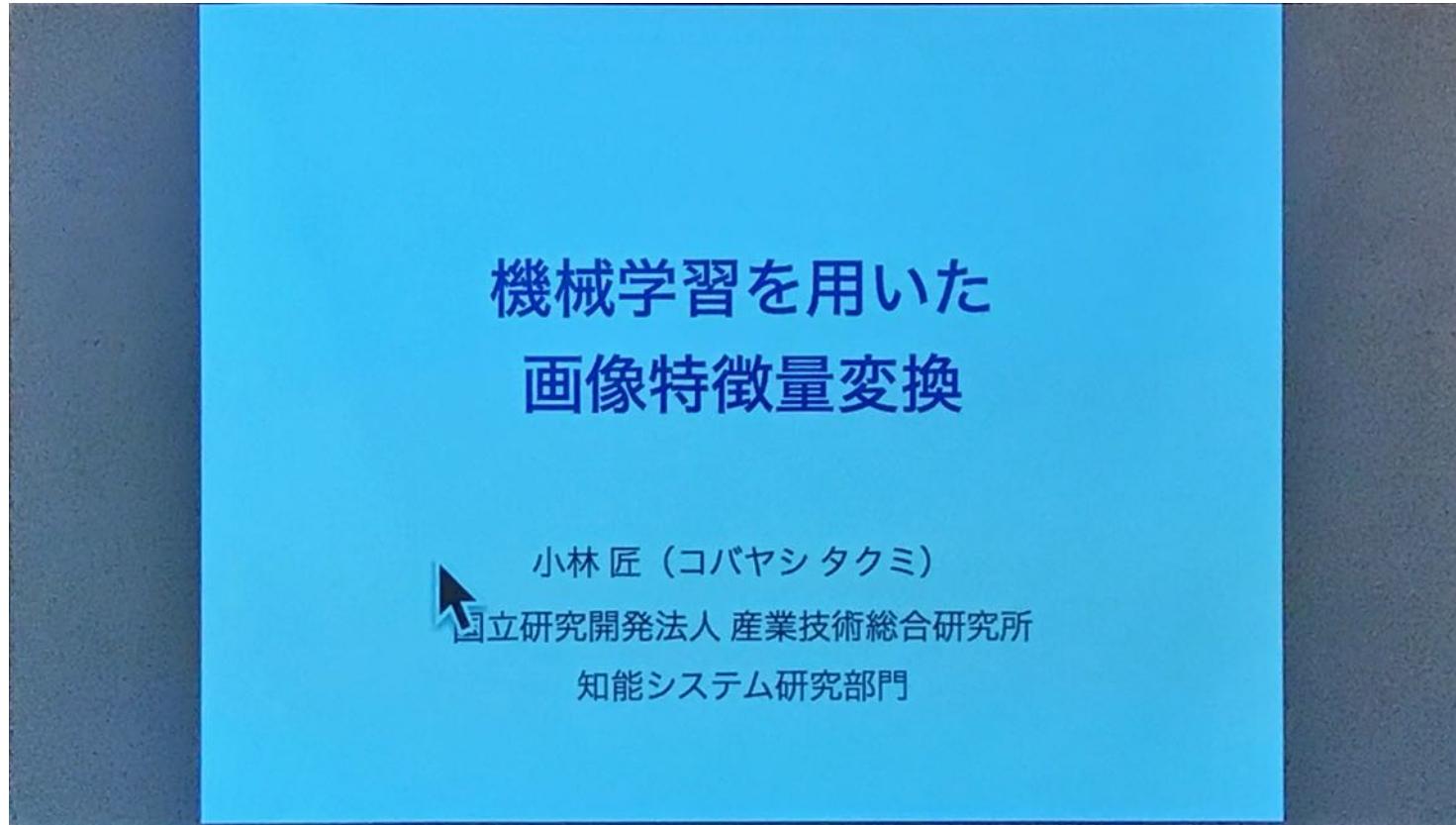
画像処理への機械学習の応用

産業技術総合研究所 小林 匠 氏 大阪大学 菅野 裕介 氏 早稲田大学 飯塚里志 氏

機械学習を用いた
画像特徴量変換 機械学習による視線推定
とその実世界応用 ディープラーニング
による画像変換

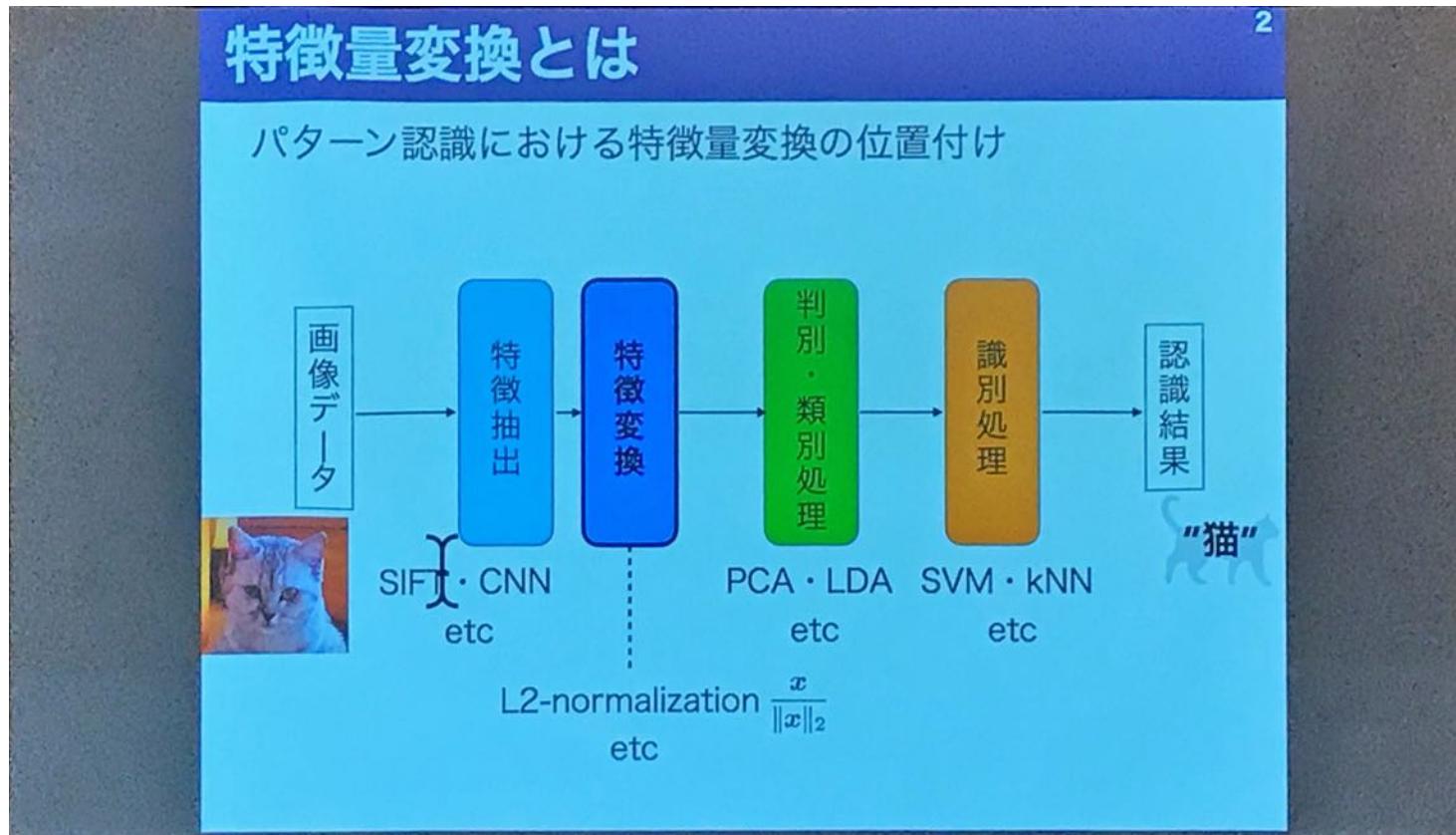
機械学習を用いた画像特徴量変換

産総研 知能システム研究部門 小林 匠



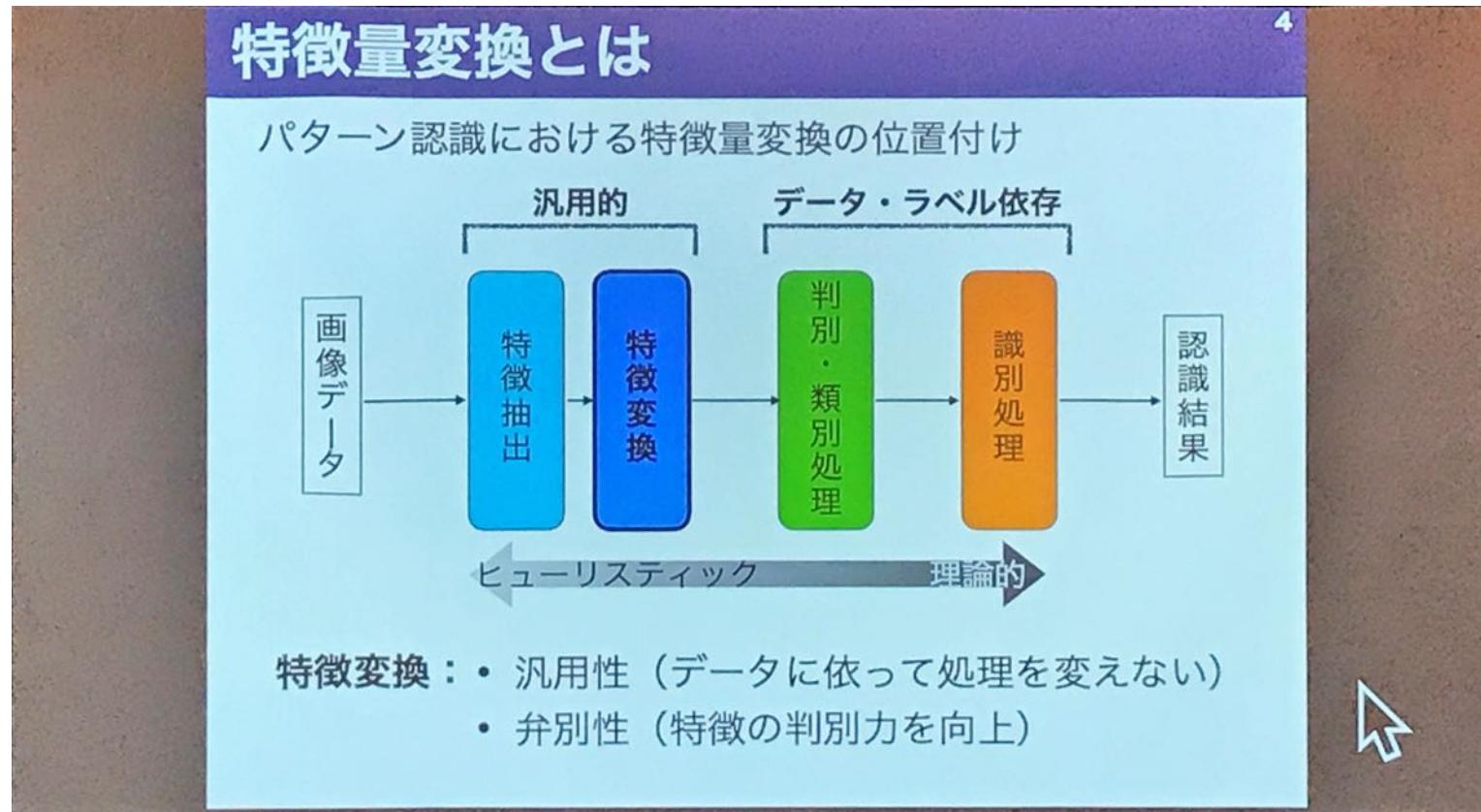
機械学習を用いた画像特徴量変換

産総研 知能システム研究部門 小林 匠



機械学習を用いた画像特徴量変換

産総研 知能システム研究部門 小林 匠



機械学習を用いた画像特徴量変換

産総研 知能システム研究部門 小林 匠

画像特徴量

5

人為的特徴量 = Hand-crafted特徴

例： image patch → gradient x → 勾配方向ヒストグラム (SIFT)

勾配方向ヒストグラム (SIFT)

學習的特徴量 = 置み込みニューラルネット(CNN)特徴

例：

転移された (pre-trained on ImageNet) CNNモデルを用いる



機械学習を用いた画像特徴量変換

産総研 知能システム研究部門 小林 匠

画像特徴量の変換 6

人為的特徴量 = Hand-crafted特徴

T. Kobayashi,
“Structured Feature Similarity with Explicit Feature Map”,
CVPR2016

學習的特徴量 = 置み込みニューラルネット(CNN)特徴

T. Kobayashi,
“Learning Additive Kernel For Feature Transformation
and Its Application to CNN Features”,
BMVC2016

科研費「パターン認識のための特徴量変換に関する研究」15K00261

機械学習を用いた画像特徴量変換

産総研 知能システム研究部門 小林 匠

距離尺度

主にヒストグラム特徴に対する手法が提案されてきた

- χ^2 (カイ²乗)-distance
- Earth Mover's distance (EMD)
- Faster EMD (高速版EMD)
- SiftDist (SIFTでのEMD)
- Diffusion distance (ヒストグラム上の拡散過程)

ここでは特徴の物理的構造に着目した特徴変換を考える

ポイント

1. 特徴量のテンソル構造
2. SSIMに基づく距離尺度

機械学習を用いた画像特徴量変換

産総研 知能システム研究部門 小林 匠

特徴量のテンソル構造 (1) 9

画像特徴量は多くの場合にテンソル構造を内包する
画像 ($x \times y$: 2次元) から特徴量 (1次元) を抽出
→ 3階テンソル ($x \times y \times$ 特徴)
→ (従来は) ベクトル

画像の局所領域から 特徴抽出 3階テンソル 高次元ベクトル

機械学習を用いた画像特徴量変換

産総研 知能システム研究部門 小林 匠

特徴量のテンソル構造（2）10

テンソル構造を保持した表現を考える
テンソルの各軸（ $x, y, \text{特徴}$ ）に沿った特徴束を一つの単位として、そこでの距離尺度を考える

- 特徴束を大きくしすぎると、元のベクトル表現
- 特徴束を小さくしすぎると、特徴要素1個ずつ

The diagram illustrates the process of extracting features from a local image region. It starts with a small 3D volume labeled "画像の局所領域から 特徴抽出" (Extracting features from a local image region). This is transformed into a 3D tensor labeled "3階テンソル" (3rd-order tensor) with axes labeled "y軸" (y-axis), "x軸" (x-axis), and "I" (feature dimension). Finally, it is shown as three separate vertical stacks of feature vectors, labeled "3軸それぞれに沿った 特徴束" (Feature bundles along each of the three axes).

機械学習を用いた画像特徴量変換

産総研 知能システム研究部門 小林 匠

SSIMに基づく距離尺度 (1) 12

Structural Similarity Index Measure (SSIM)
画像の品質評価のための定量的指標として用いられる

Image	MSE	SSIM
(a)	0	1
(b)	306	0.928
(c)	309	0.987
(d)	309	0.578
(e)	315	0.730
(f)	309	0.580
(g)	308	0.641
(h)	694	0.505
(i)	871	0.404
(j)	873	0.399
(k)	590	0.549
(l)	577	0.551

Z. Wang and A. C. Bovik. Mean squared error: love it or leave it? - a new look at signal fidelity measures. *IEEE Signal Processing Magazine*, 26(1):98–117, January 2009.

機械学習を用いた画像特徴量変換

産総研 知能システム研究部門 小林 匠

SSIMに基づく距離尺度 (2)

特徴束においてSSIMを距離尺度として採用

特徴束 x y に対して

$$\mathcal{S}_{org}(x, y) = \mathcal{M}(x, y) \times \mathcal{V}(x, y) \times \mathcal{C}(x, y) \rightarrow \mathcal{C}(x, y) = \frac{(x - u(x))^\top (y - u(y))}{\|x - u(x)\|_2 \|y - u(y)\|_2},$$

相関係数

明るさ \downarrow コントラスト \downarrow 構造パターン

$$k(u(x), u(y)), \quad k(q(x), q(y)), \quad \text{where 類似度関数 } k(a, b) = \frac{2ab}{a^2 + b^2}$$

平均値の類似度 \downarrow 標準偏差の類似度

特徴変動への頑健性を高めるため乗法的表現から加法的表現へ

- 上記の元表現 \mathcal{S}_{org} は、 \mathcal{M} , \mathcal{V} , \mathcal{C} のわずかな変化にも敏感

乗法的 : $\mathcal{S}_{org} = \mathcal{M} \times \mathcal{V} \times \mathcal{C}$

加法的 : $\mathcal{S}_{add} = w_{\mathcal{M}}\mathcal{M} + w_{\mathcal{V}}\mathcal{V} + w_{\mathcal{C}}\mathcal{C}$

	$w_{\mathcal{M}}$	$w_{\mathcal{V}}$	$w_{\mathcal{C}}$
\mathcal{M}	(2, 2, 1)		
\mathcal{V}		(0, +1) [0, +1] [-1, +1]	
\mathcal{C}			(-1, +1) [0, +1] [-1, +1]
non-neg. feat. [-1, +1] [0, +1] [-1, +1]			
real feat. [-1, +1] [0, +1] [-1, +1]			
value range			
	(1, 2, 1)		

機械学習を用いた画像特徴量変換

産総研 知能システム研究部門 小林 匠

効率的表現 \Leftrightarrow 特徴"変換"の実現 (1)¹⁷

類似度 S を陽写像 (explicit map) の内積へ展開

- 線形表現になり、高速演算かつ線形識別器を構築可能

$$\begin{aligned} S(x, y) &= w_{\mathcal{M}} \mathcal{M}(x, y) + w_{\mathcal{V}} \mathcal{V}(x, y) + w_{\mathcal{C}} \mathcal{C}(x, y) \\ &= \underline{\phi(x)^{\top} \phi(y)} \end{aligned}$$

陽写像 (explicit map) $\in \mathbb{R}^d$

$$\phi = \sqrt{w_{\mathcal{M}}} \phi_{\mathcal{M}} \oplus \sqrt{w_{\mathcal{V}}} \phi_{\mathcal{V}} \oplus \sqrt{w_{\mathcal{C}}} \phi_{\mathcal{C}}$$

直和

✓ $\mathcal{C}(x, y) = \frac{(x - \mathbf{u}(x))^\top (y - \mathbf{u}(y))}{\|x - \mathbf{u}(x)\|_2 \|y - \mathbf{u}(y)\|_2},$

?

$$\begin{aligned} \mathcal{M}(x, y) &= \underline{k}(\mathbf{u}(x), \mathbf{u}(y)), \\ \mathcal{V}(x, y) &= \underline{k}(\mathbf{q}(x), \mathbf{q}(y)), \end{aligned}$$

$k(a, b) = \frac{2ab}{a^2 + b^2}$

機械学習を用いた画像特徴量変換

産総研 知能システム研究部門 小林 匠

効率的表現 ⇄ 特徴"変換"の実現 (2)¹⁸

類似度関数 k を陽写像 (explicit map) の内積へ展開

$$k(a, b) = \frac{2ab}{a^2 + b^2} = \frac{2\text{sgn}(ab)}{\left|\frac{a}{b}\right| + \left|\frac{b}{a}\right|} = \frac{2\text{sgn}(ab)}{e^{-\omega} + e^{\omega}} = \text{sgn}(ab)\text{sech}(\omega)$$

where $\omega = \log \left|\frac{b}{a}\right|$

関数 sech のフーリエ展開に基いて

$$\begin{aligned} k(a, b) &= \text{sgn}(ab)\text{sech}(\omega) = \text{sgn}(ab) \int_{-\infty}^{\infty} e^{-i\omega\lambda} \kappa(\lambda) d\lambda \\ &= \int_{-\infty}^{\infty} [\text{sgn}(a)e^{-i\lambda \log |a|} \sqrt{\kappa(\lambda)}]^* [\text{sgn}(b)e^{-i\lambda \log |b|} \sqrt{\kappa(\lambda)}] d\lambda, \end{aligned}$$

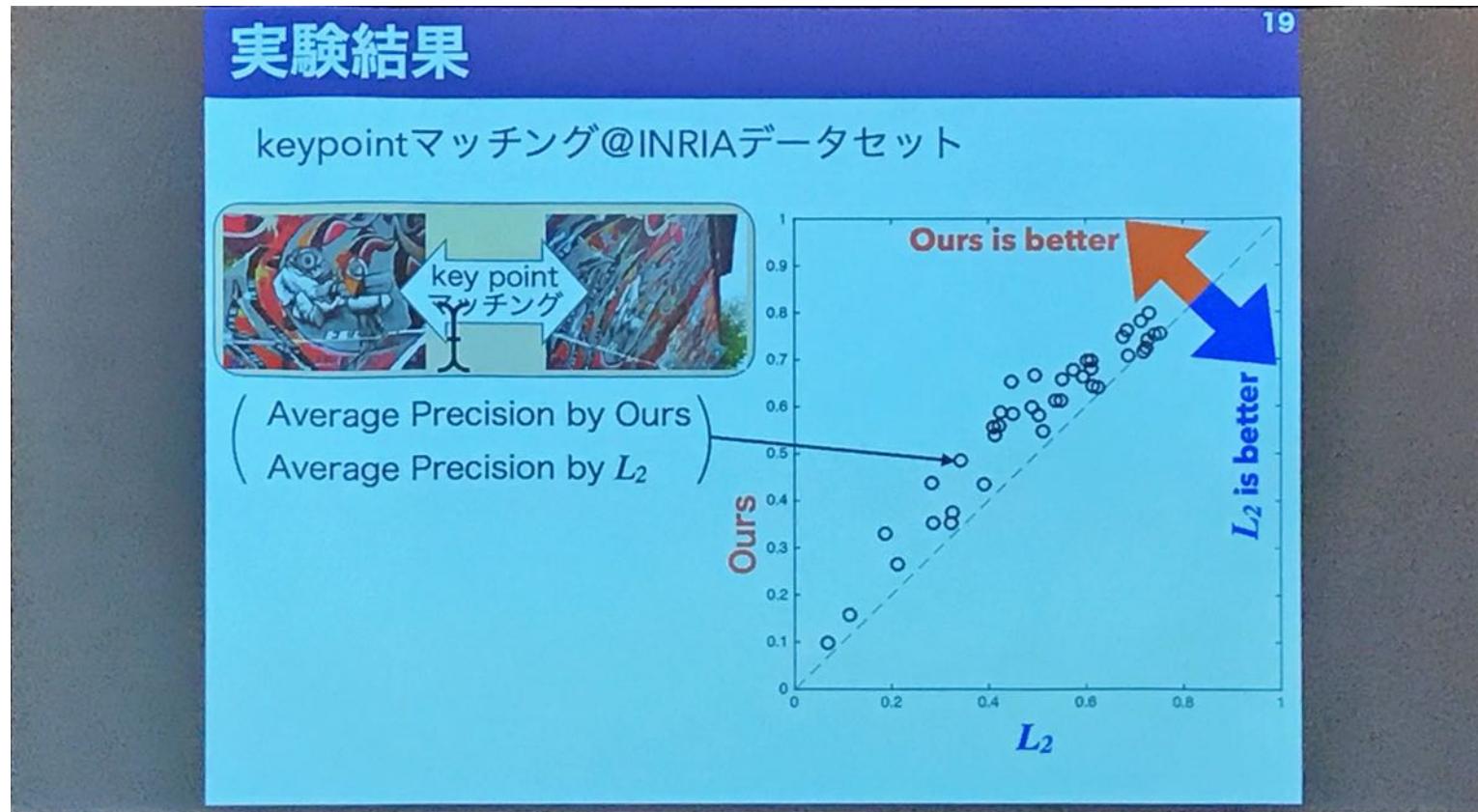
$\Rightarrow \phi_k = \begin{bmatrix} [a = 0] \\ \tilde{g}_k(\lambda; a) = \text{sgn}(a)e^{-i\lambda \log |a|} \sqrt{\frac{1}{2} \text{sech}\left(\frac{\pi\lambda}{2}\right)} \end{bmatrix} \in \mathbb{R}^{d_k}$

$\tilde{g}_k(\lambda; a) = \text{sgn}(a)e^{-i\lambda \log |a|} \sqrt{\frac{1}{2} \text{sech}\left(\frac{\pi\lambda}{2}\right)}$

λを複数個サンプリング

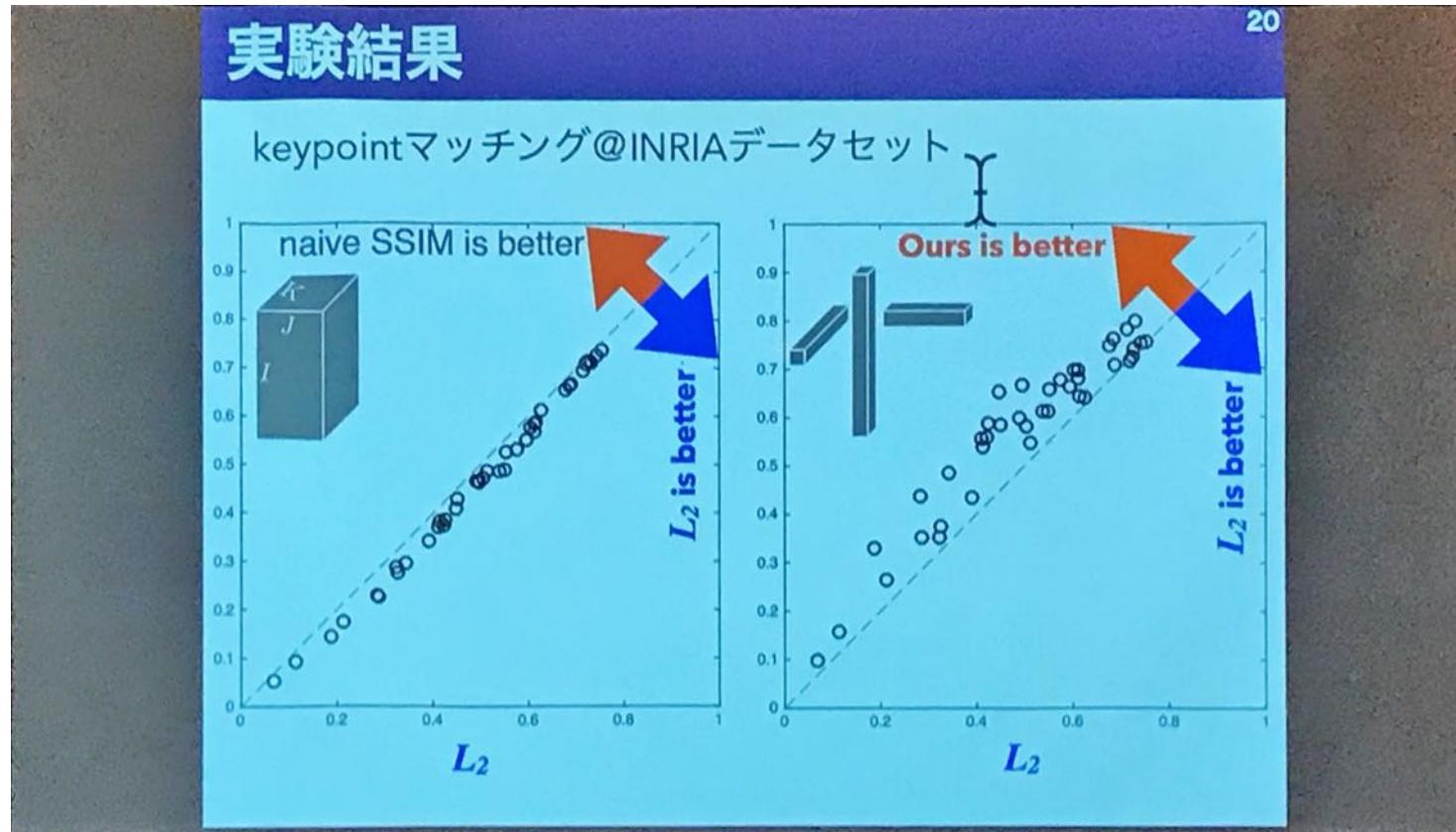
機械学習を用いた画像特徴量変換

産総研 知能システム研究部門 小林 匠



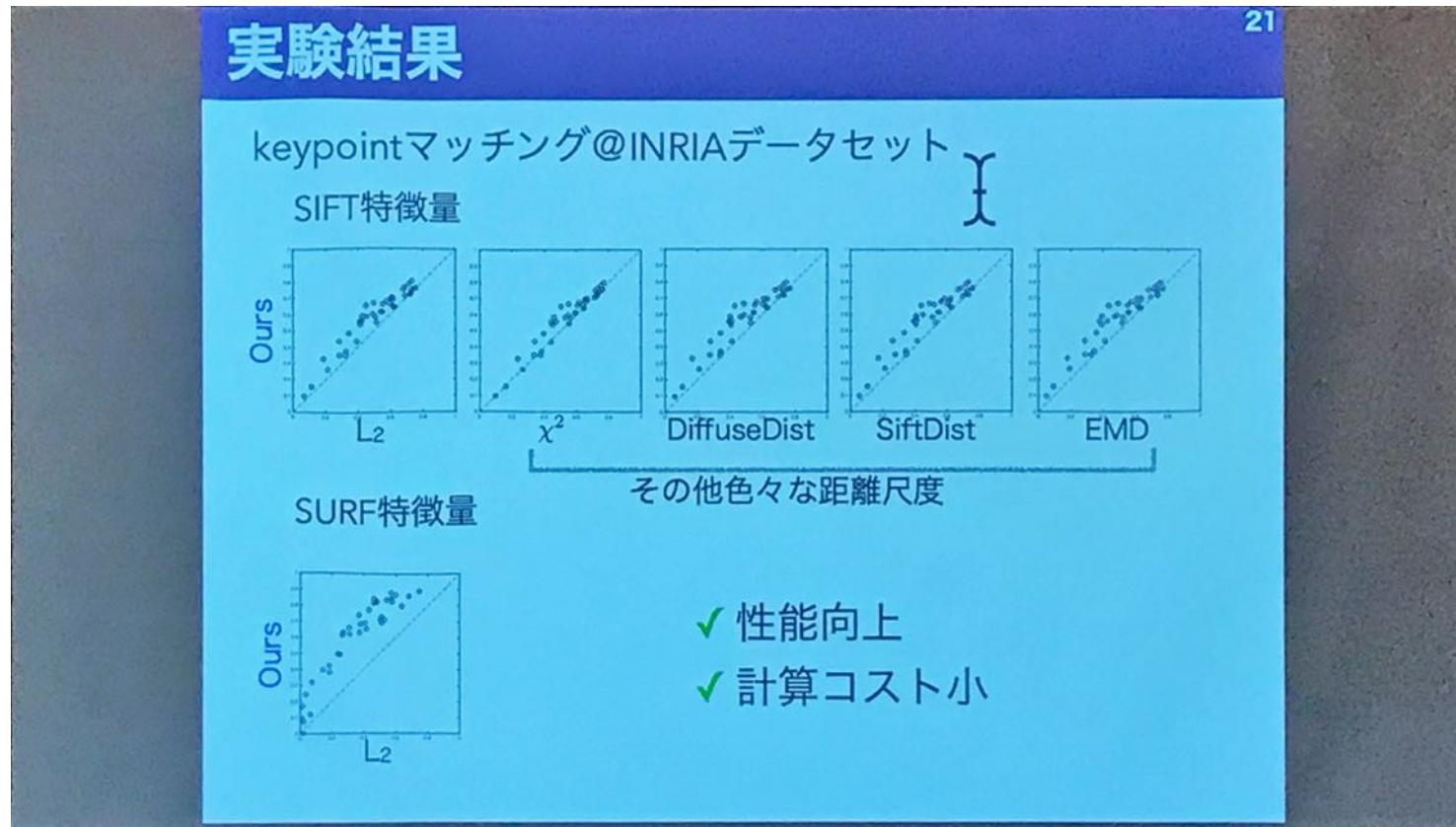
機械学習を用いた画像特徴量変換

産総研 知能システム研究部門 小林 匠



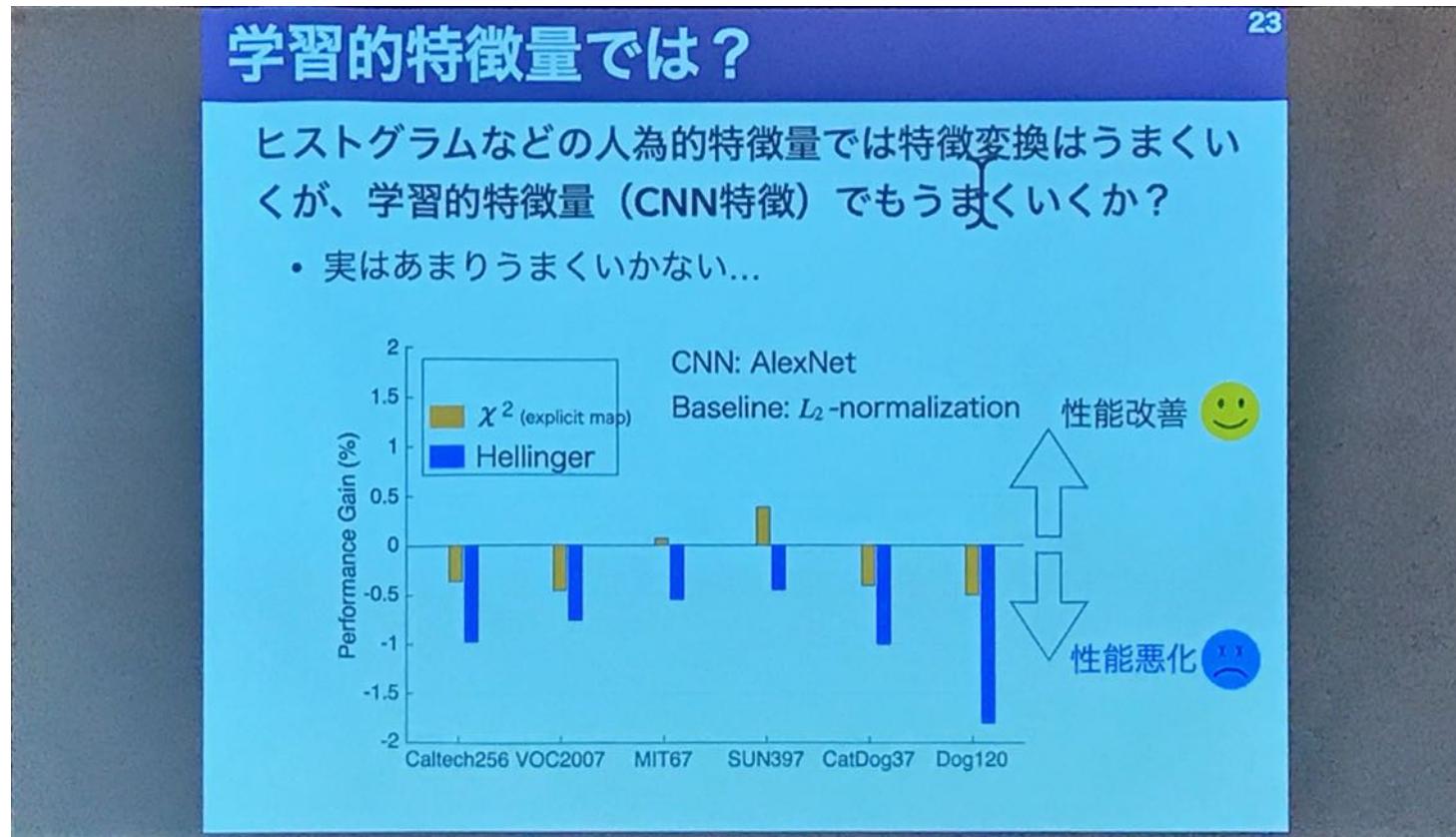
機械学習を用いた画像特徴量変換

産総研 知能システム研究部門 小林 匠



機械学習を用いた画像特徴量変換

産総研 知能システム研究部門 小林 匠



機械学習を用いた画像特徴量変換

産総研 知能システム研究部門 小林 匠

なぜか？

距離尺度＝非線形（kernel）関数は、対象の特徴量の特性に沿って決めていた。

CNN特徴のような“よくわからない”特徴に対して強引に適用するのは困難であるらしい。

そこで、学習的特徴量に対する距離尺度も学習する
ポイント

1. 距離尺度のデータからの学習
2. 弁別性・汎化性の高い変換方式

24

→

機械学習を用いた画像特徴量変換

産総研 知能システム研究部門 小林 匠

加法的カーネル表現

25

加法的カーネルから出発

D 次元特徴ベクトル $x, y \in \mathbb{R}^D$ に対して

加法的カーネル: $\bar{k}(x, y) = \sum_{i=1}^D k(x_i, y_i)$, 要素毎のカーネル関数の和

カーネル関数の陽写像展開:

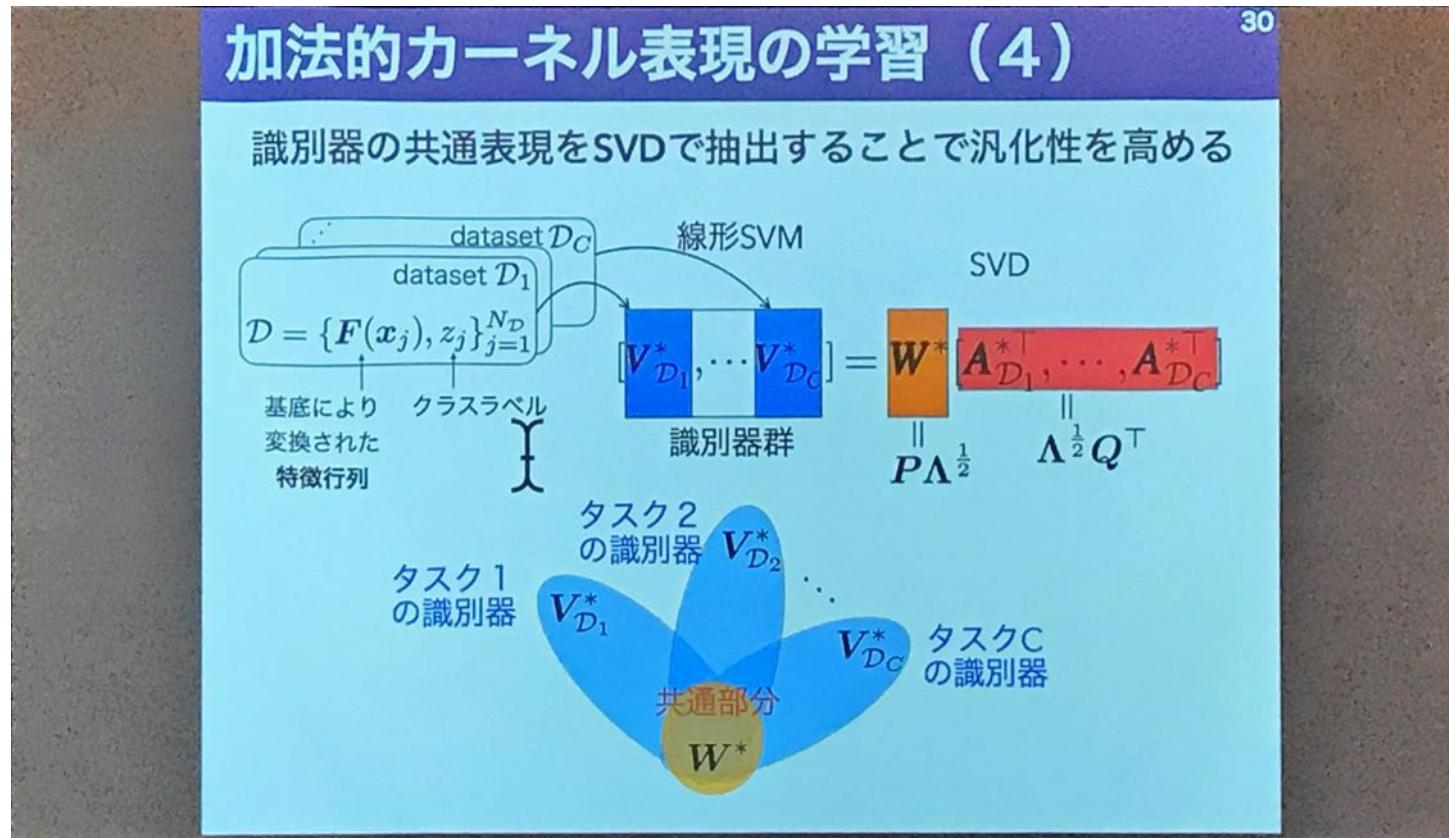
$$\begin{aligned} k(x, y) &= \int \phi(\gamma; x)\phi(\gamma; y)d\lambda \approx \sum_{l=1}^K \phi(\gamma_l; x)\phi(\gamma_l; y)\Delta_l \\ &= \phi(x)^\top \phi(y) \end{aligned}$$

加法的カーネルの学習 → 関数 $\phi(x)$ の学習



機械学習を用いた画像特徴量変換

産総研 知能システム研究部門 小林 匠



機械学習を用いた画像特徴量変換

産総研 知能システム研究部門 小林 匠

基底関数

フーリエ基底を考えるのが自然。

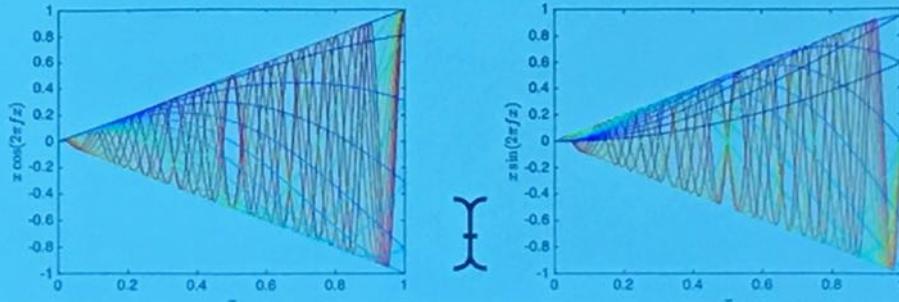
ここではさらに、特徴の根本的な性質を考慮

↓

(magnitudeの) 大きい特徴値はより特徴的な情報を含む

$$f_{2m-1}(x) = \underline{x} \cos(2\pi\eta_m x), \quad f_{2m}(x) = \underline{x} \sin(2\pi\eta_m x),$$

where $\eta \in \{0, 0.1, \dots, 0.9, 1, \dots, 10\}$,



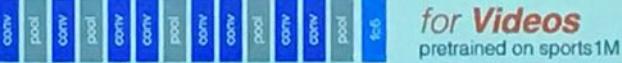
31

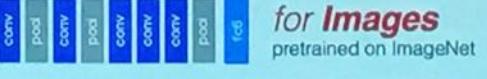
機械学習を用いた画像特徴量変換

産総研 知能システム研究部門 小林 匠

実験結果 32

3種類のCNNモデル

C3D [Tran, 2015] 

Alex [Krizhevsky, 2012] 

VGG [Simonyan, 2014] 

データセット

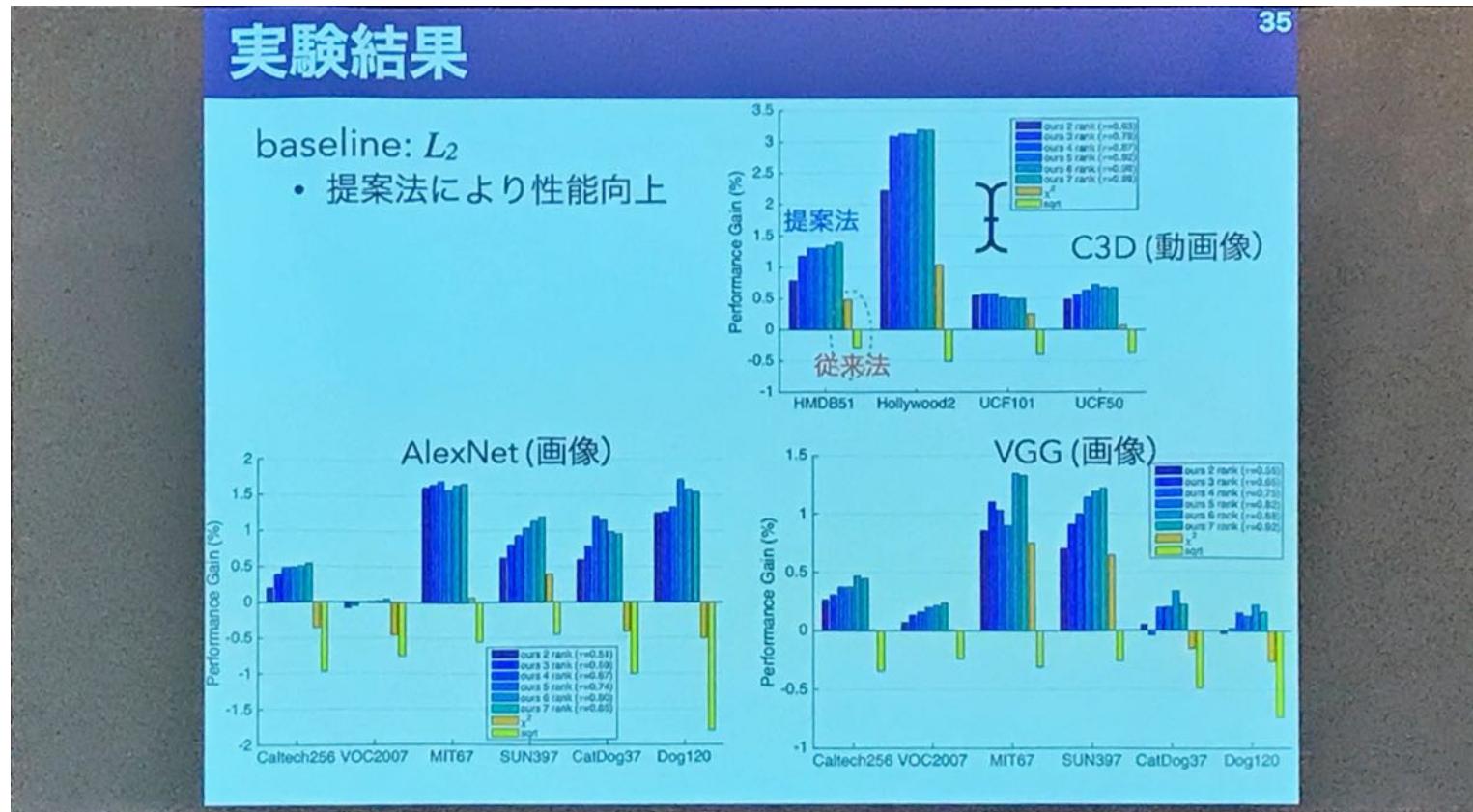
Video: HMDB51, UCF101/50, Hollywood2 (動作認識)

Image: Caltech256, VOC2007, (物体認識)
MIT67, SUN397, (シーン認識)

CatDog37, Dog120 (詳細な種別認識)

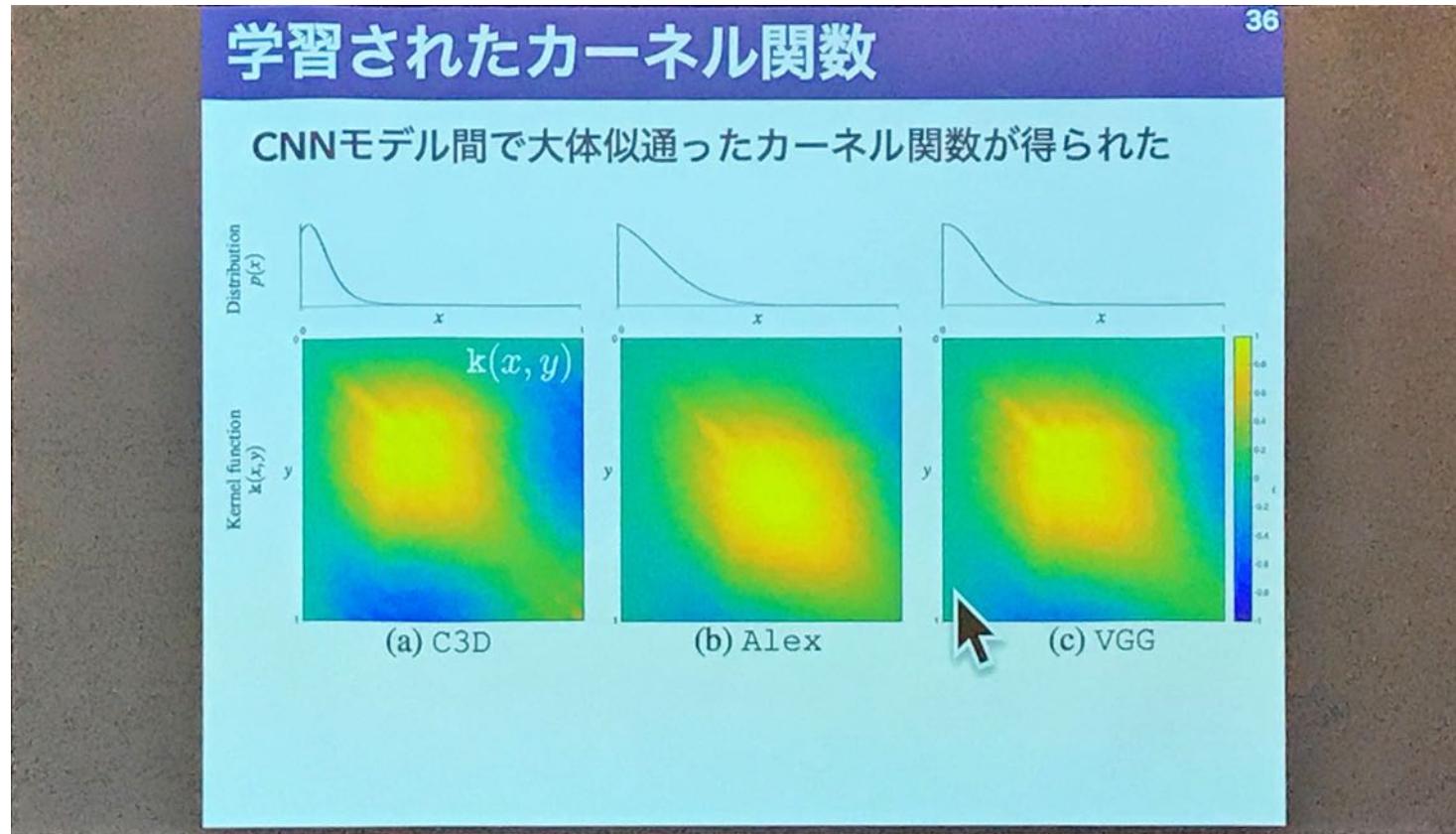
機械学習を用いた画像特徴量変換

産総研 知能システム研究部門 小林 匠



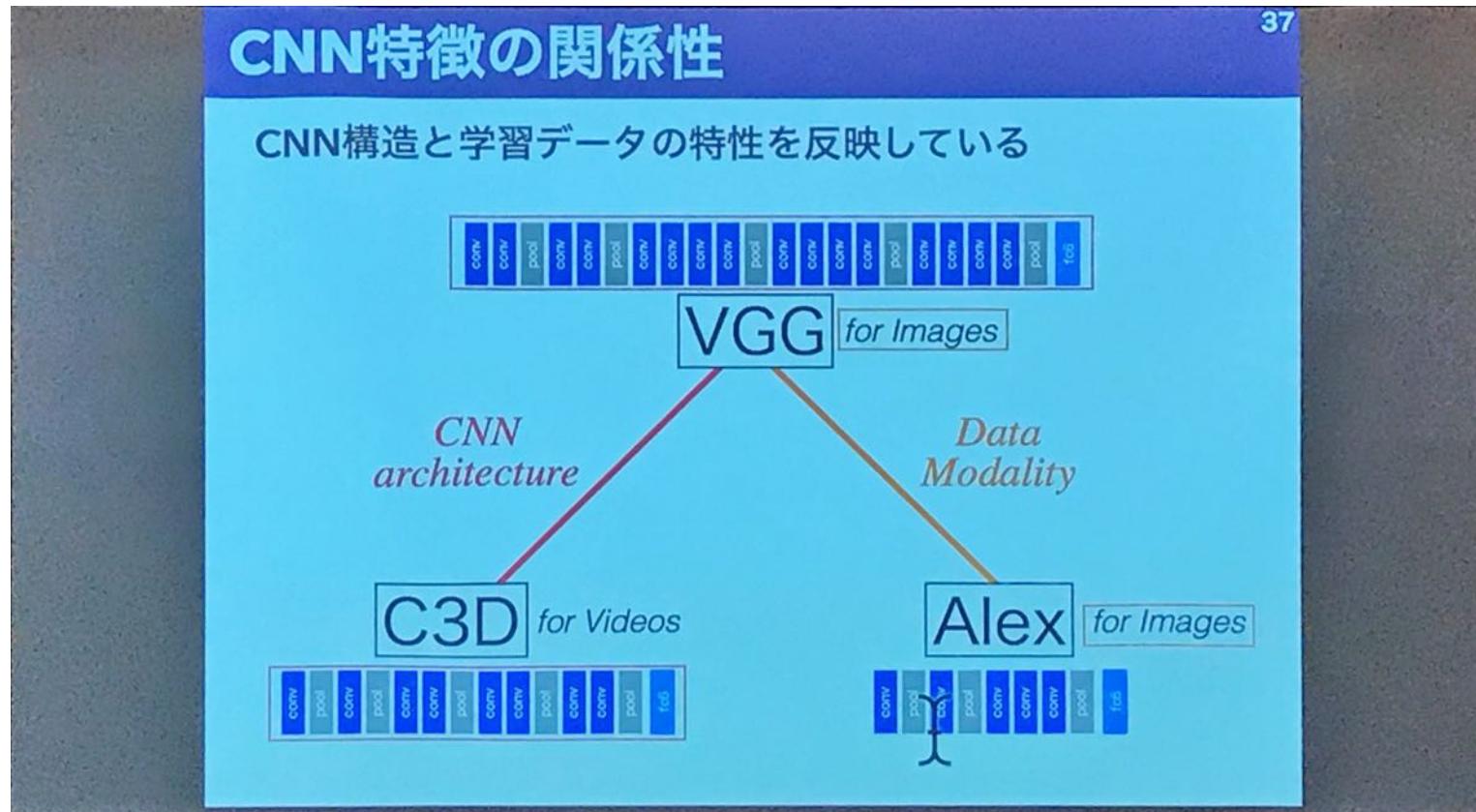
機械学習を用いた画像特徴量変換

産総研 知能システム研究部門 小林 匠



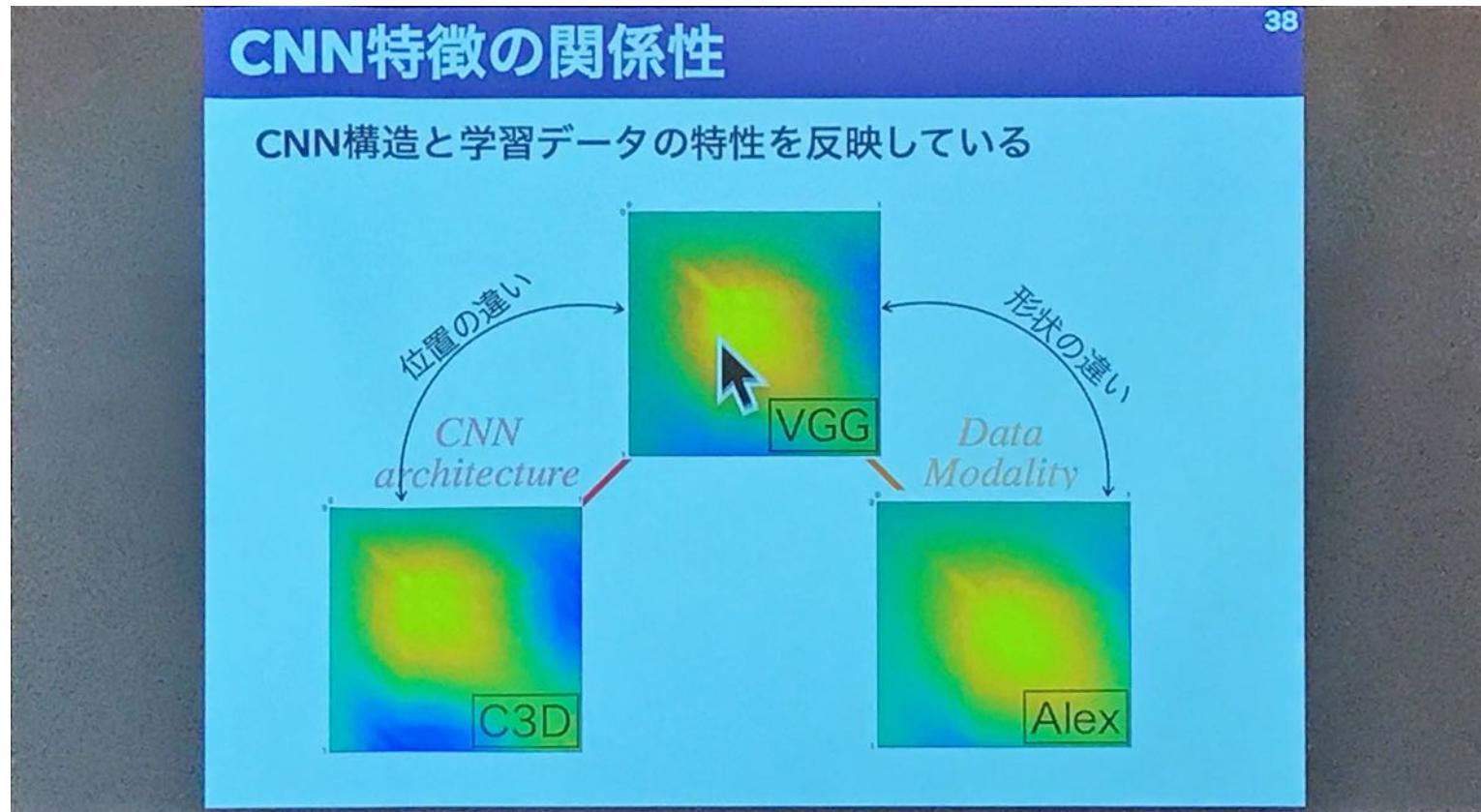
機械学習を用いた画像特徴量変換

産総研 知能システム研究部門 小林 匠



機械学習を用いた画像特徴量変換

産総研 知能システム研究部門 小林 匠



機械学習を用いた画像特徴量変換

産総研 知能システム研究部門 小林 匠

まとめ

39

人為的特徴量と学習的特徴量に対しての特徴変換

人為的特徴 = hand-crafted特徴 : SIFT, HOGなど

学習的特徴 = CNN特徴 : AlexNet, VGGなど

人為的特徴量に対しては内在しているテンソル構造に着目し、学習的特徴量に対しては距離尺度の学習を通して、それぞれの特徴変換手法を示した。

学習されたカーネル関数（の類似性）は、CNN特徴の特性解析に使えるかも...

機械学習を用いた画像特徴量変換

産総研 知能システム研究部門 小林 匠

まとめ

39

人為的特徴量と学習的特徴量に対しての特徴変換

人為的特徴 = hand-crafted特徴 : SIFT, HOGなど

学習的特徴 = CNN特徴 : AlexNet, VGGなど

人為的特徴量に対しては内在しているテンソル構造に着目し、学習的特徴量に対しては距離尺度の学習を通して、それぞれの特徴変換手法を示した。

学習されたカーネル関数（の類似性）は、CNN特徴の特性解析に使えるかも...

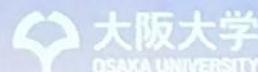
機械学習による視線推定とその実世界応用

大阪大学大学院 情報科学研究科 菅野 裕介

機械学習による視線推定と その実世界応用

菅野 裕介

大阪大学大学院 情報科学研究科



機械学習による視線推定とその実世界応用

大阪大学大学院 情報科学研究科 菅野 裕介

自己紹介

菅野裕介

2016年 - 現在 : 大阪大学大学院 情報科学研究科 准教授
2014年 - 2016年 : Max Planck Institute for Informatics ポスドク



◀ PERCEPTUAL
▶ USER INTERFACES
perceptual mpi-inf.mpg.de



Xucong Zhang Andreas Bulling

2010年 - 2014年 : 東京大学生産技術研究所 特任助教
2010年 : 東京大学大学院 情報理工学系研究科 博士課程修了

OSAKA UNIVERSITY

2

機械学習による視線推定とその実世界応用

大阪大学大学院 情報科学研究科 菅野 裕介

視線推定・アイトラッキング

- 人物がどこを見ているかを計測・推定する
 - Attentive User Interfaces
 - 広告・マーケティング調査
 - 人間の内部状態を推定するための特徴として
- 従来の手法は専用のハードウェアを必要とする



OSAKA UNIVERSITY

3

機械学習による視線推定とその実世界応用

大阪大学大学院 情報科学研究科 菅野 裕介

カメラベースの視線推定

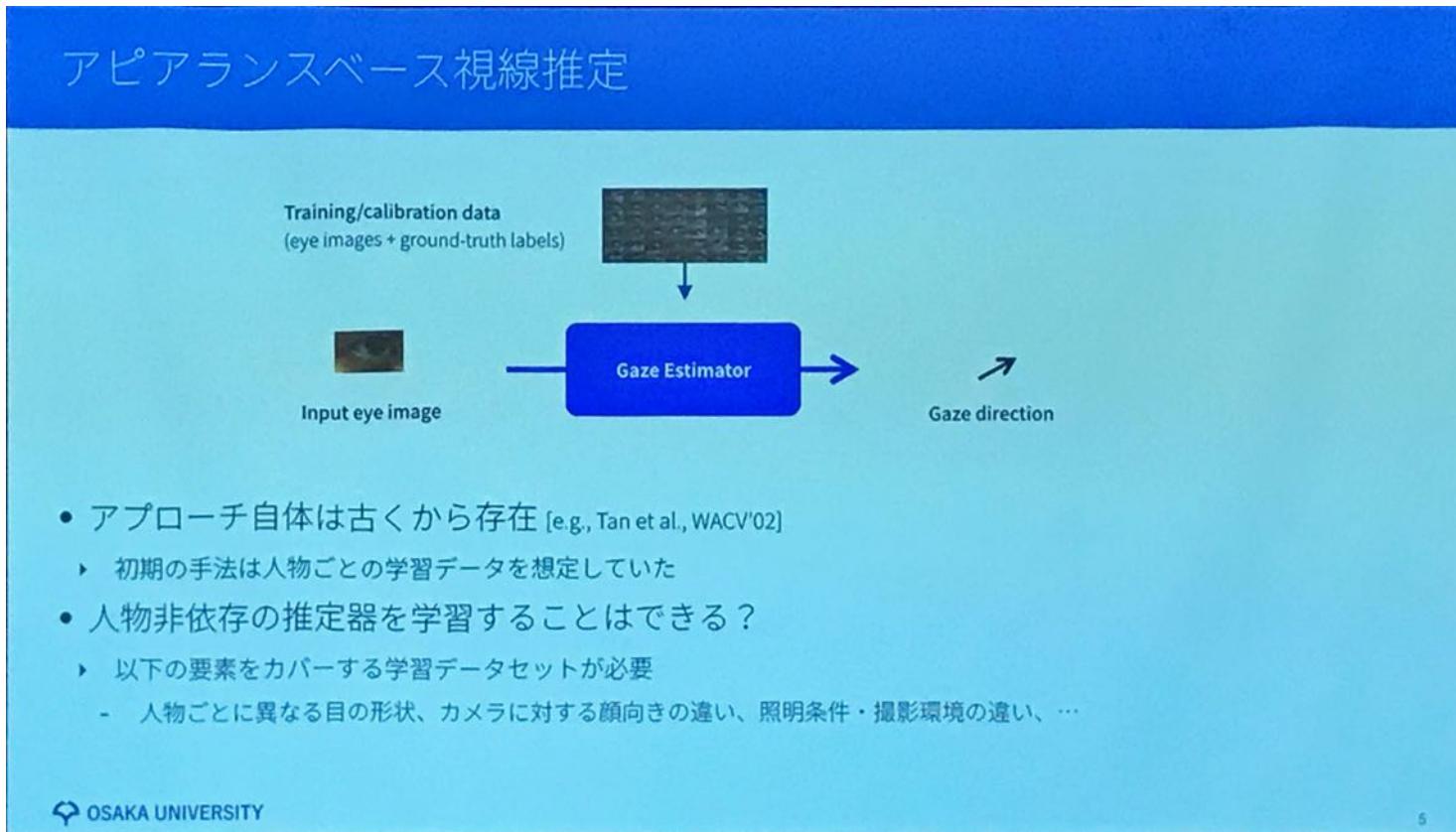


- 通常のカメラから、画像中に写っている人物の(3次元)視線方向を推定する
 - ウェブカメラ、ウェアラブルカメラ、…
- 従来手法では実現できないアプリケーションが多数
 - HRI、一人称視点映像解析、公衆空間での注意推定

OSAKA UNIVERSITY

機械学習による視線推定とその実世界応用

大阪大学大学院 情報科学研究科 菅野 裕介

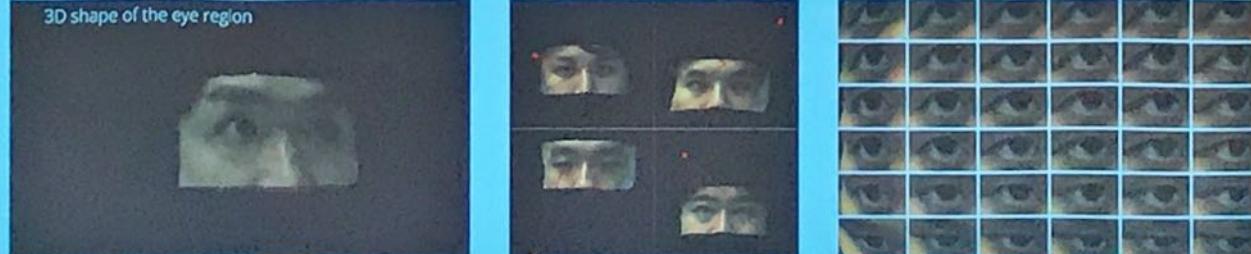


機械学習による視線推定とその実世界応用

大阪大学大学院 情報科学研究科 菅野 裕介

Multi-view gaze dataset [Sugano et al., CVPR'14]

- 8台のカメラを同期して撮影した視線真値付きの目画像データセット
 - 50人 × 160 (16 × 10) 視線方向
- 3D reconstruction can be used to synthesize training data
 - $50 \times 160 \times 144$ 姿勢 = 1,152,000枚の目画像
- 人物非依存の視線推定関数を学習することが可能



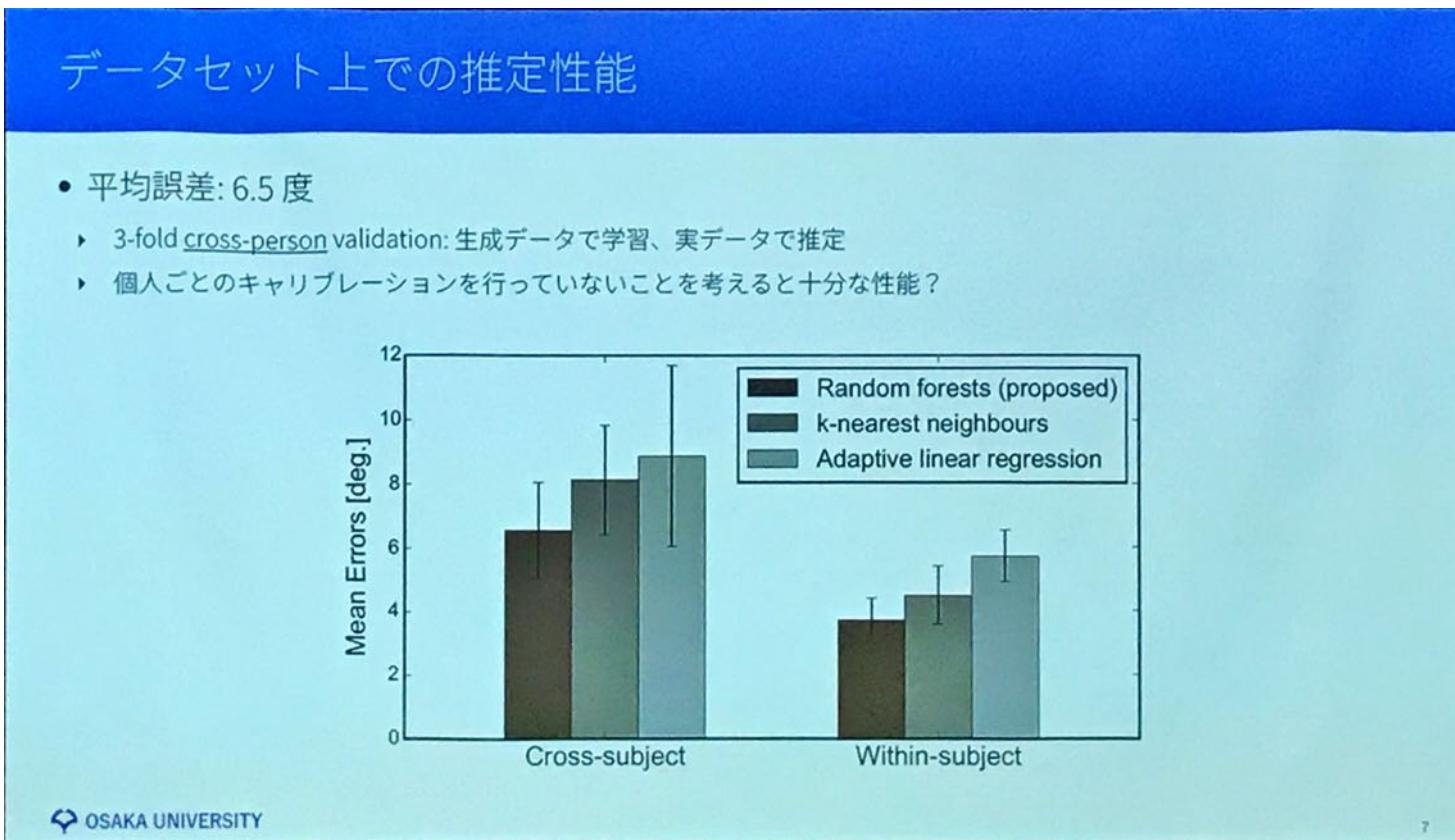
3D shape of the eye region

OSAKA UNIVERSITY

6

機械学習による視線推定とその実世界応用

大阪大学大学院 情報科学研究科 菅野 裕介



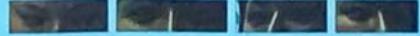
機械学習による視線推定とその実世界応用

大阪大学大学院 情報科学研究科 菅野 裕介

アピアランスベース視線推定 “in the wild”

- 実応用上はこれよりも遙かに多様な入力画像を想定する必要がある

Cross-domain training



- 外部データセットで学習
- 未知の環境に対する頑健性

Within-domain training



- 対象環境で獲得したデータで学習
- 性能上限の評価

Daily life environment



機械学習による視線推定とその実世界応用

大阪大学大学院 情報科学研究科 菅野 裕介

MPIIGaze dataset [Zhang et al., CVPR'15, TPAMI (accepted)]

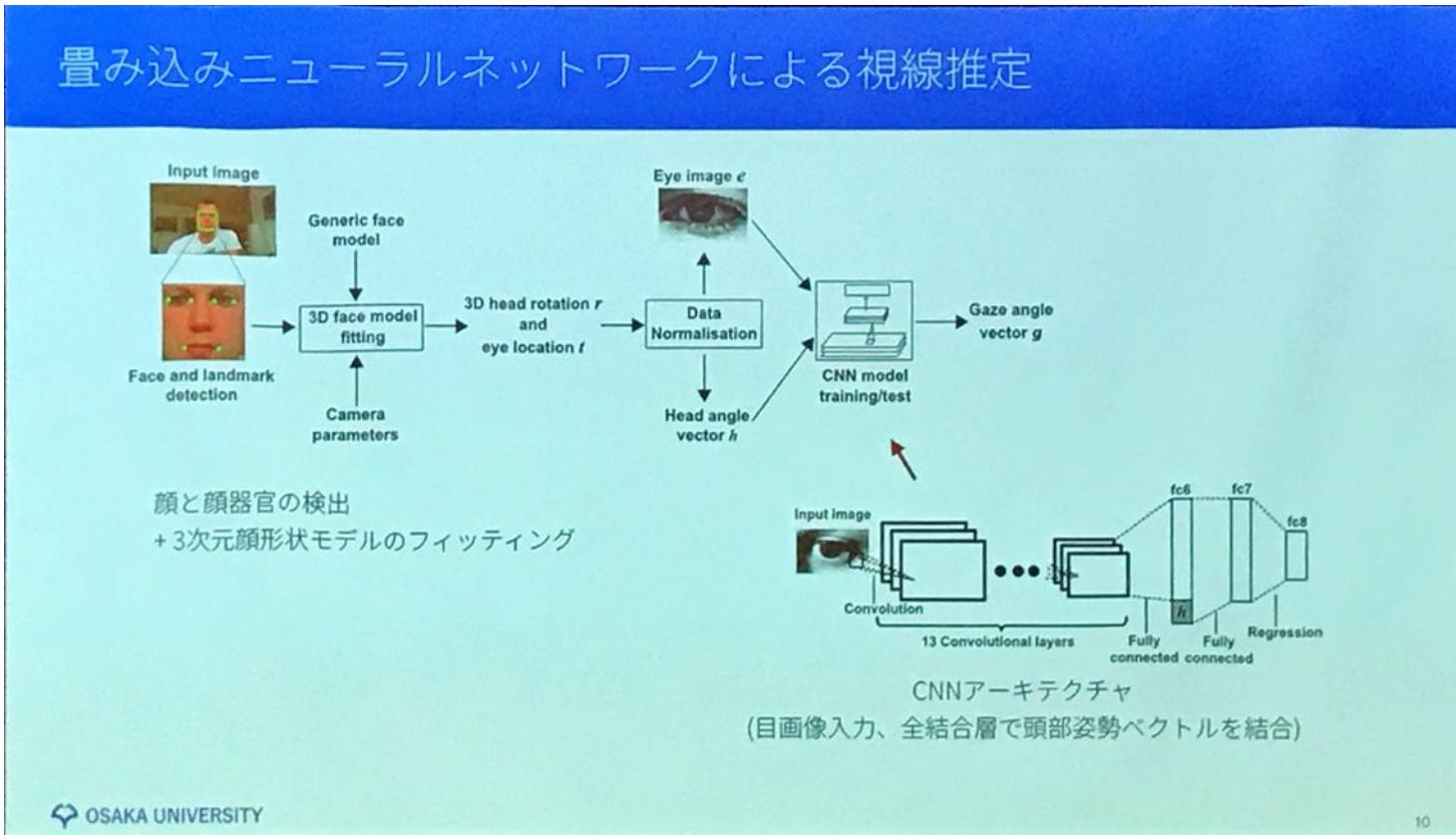


- より実環境に近い多様な照明条件下で撮影したデータセット
 - 15人、～3ヶ月、計213,659画像
 - ラップトップPCにインストールしたソフトウェアで視線真値付きの画像を定期的に撮影
 - 日常生活内で定期的に撮影が行われることで撮影環境の多様性を担保

OSAKA UNIVERSITY

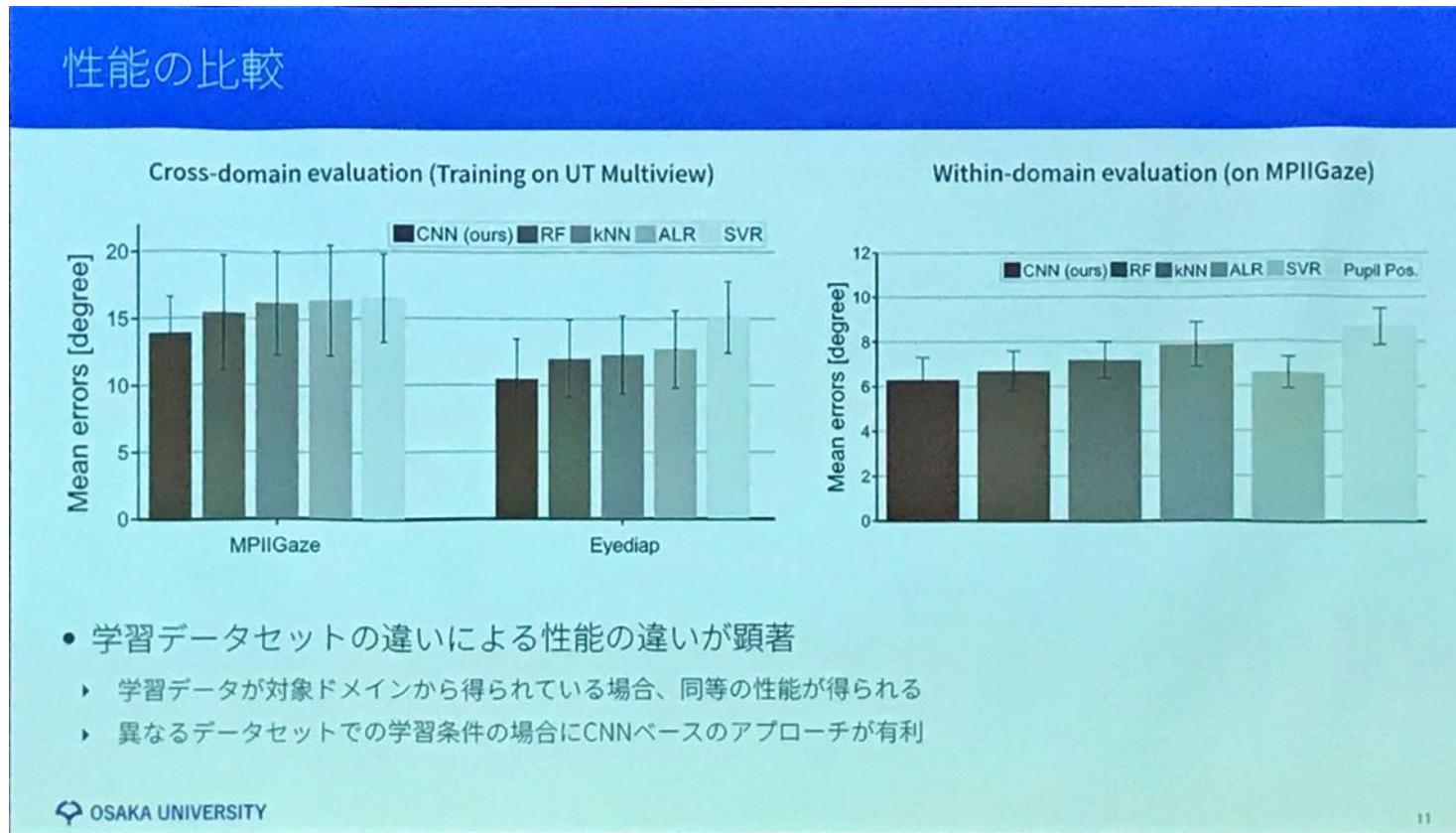
機械学習による視線推定とその実世界応用

大阪大学大学院 情報科学研究科 菅野 裕介



機械学習による視線推定とその実世界応用

大阪大学大学院 情報科学研究科 菅野 裕介



機械学習による視線推定とその実世界応用

大阪大学大学院 情報科学研究科 菅野 裕介

学習データの獲得

- 照明条件の変化が重要な要素の一つ
- 対象となる環境から学習データを獲得できれば解決するが、現実的には難しい
- 新しい環境に対応できる推定器を事前学習するには？
- カメラ/頭部姿勢だけではなく、学習データの照明環境もコントロールする
- 3DCGモデルを使った学習データ生成

OSAKA UNIVERSITY 12

機械学習による視線推定とその実世界応用

大阪大学大学院 情報科学研究科 菅野 裕介

SynthesEyes Dataset [Wood et al., ICCV'15]

The visualization shows a grid of images from the SynthesEyes dataset. The top row displays full faces of various individuals, each with five smaller views of their eyes labeled f1 through f5 below them. The bottom row shows close-up views of four different eyes, each with a corresponding scene image to its right: a landscape, a cityscape at night, a street scene, and another cityscape.

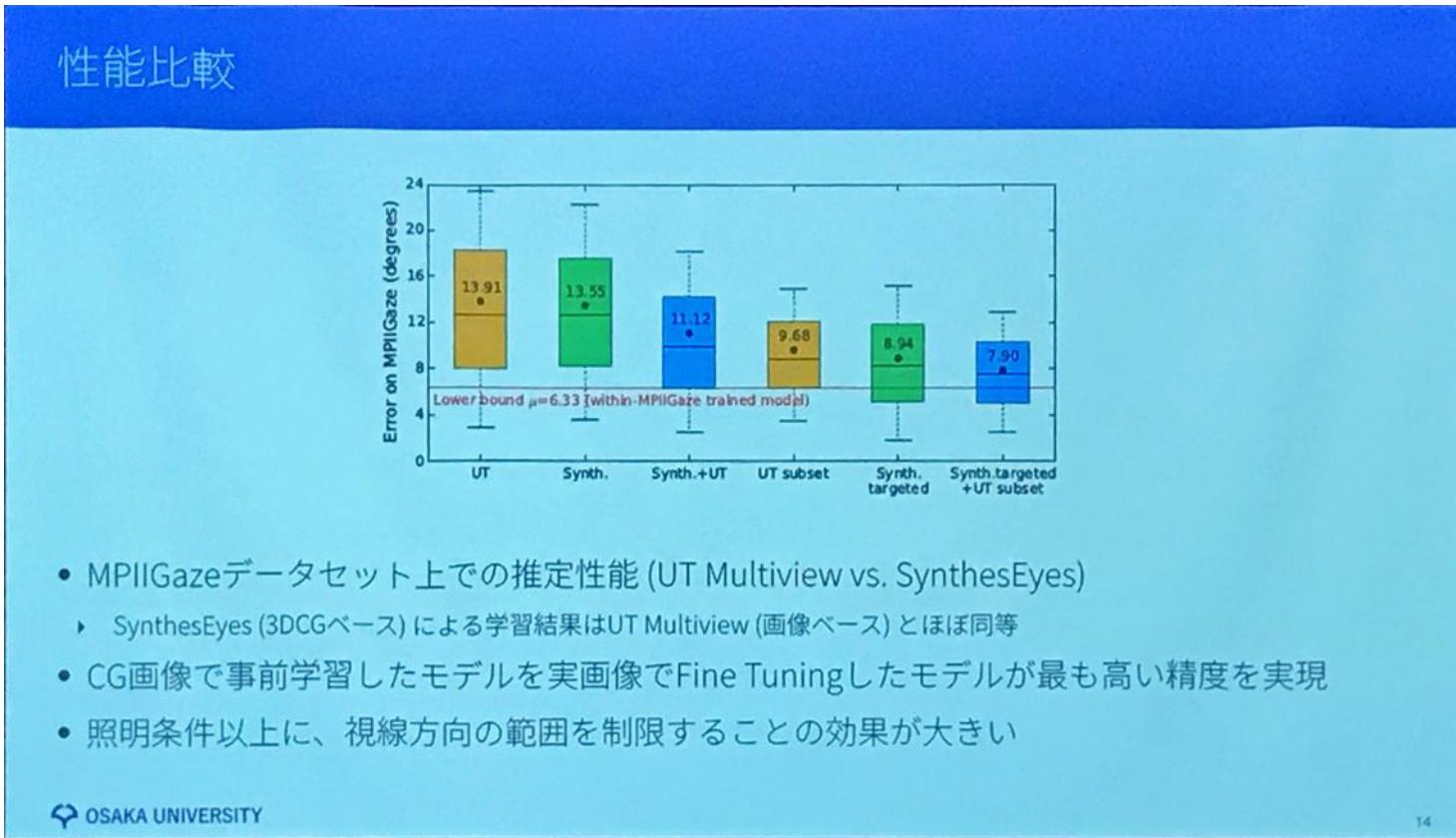
- 顔の3次元形状・テクスチャモデルと眼球モデルの組み合わせ
 - イメージベースライティングによる照明条件コントロール
- Multiviewデータセットと比較して、
 - 照明も変化させながら、正確な視線方向と共に目画像を生成することができる
 - 顔・目領域形状の多様性は制限される(形状変化のモデルは無い)

OSAKA UNIVERSITY

13

機械学習による視線推定とその実世界応用

大阪大学大学院 情報科学研究科 菅野 裕介



機械学習による視線推定とその実世界応用

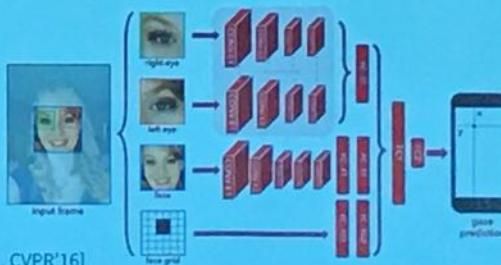
大阪大学大学院 情報科学研究科 菅野 裕介

顔画像入力からの視線推定 [Zhang et al., CVPRW'17]

- これまでのモデルは片目毎に視線方向を推定
 - 物理的には正しいが、多くの場合視線方向が両目で一致しない
- 顔画像全体をそのまま入力にすれば良いのでは？
 - パート毎の入力を結合するアーキテクチャによる性能向上事例
 - 最適な組み合わせを手動で設計している

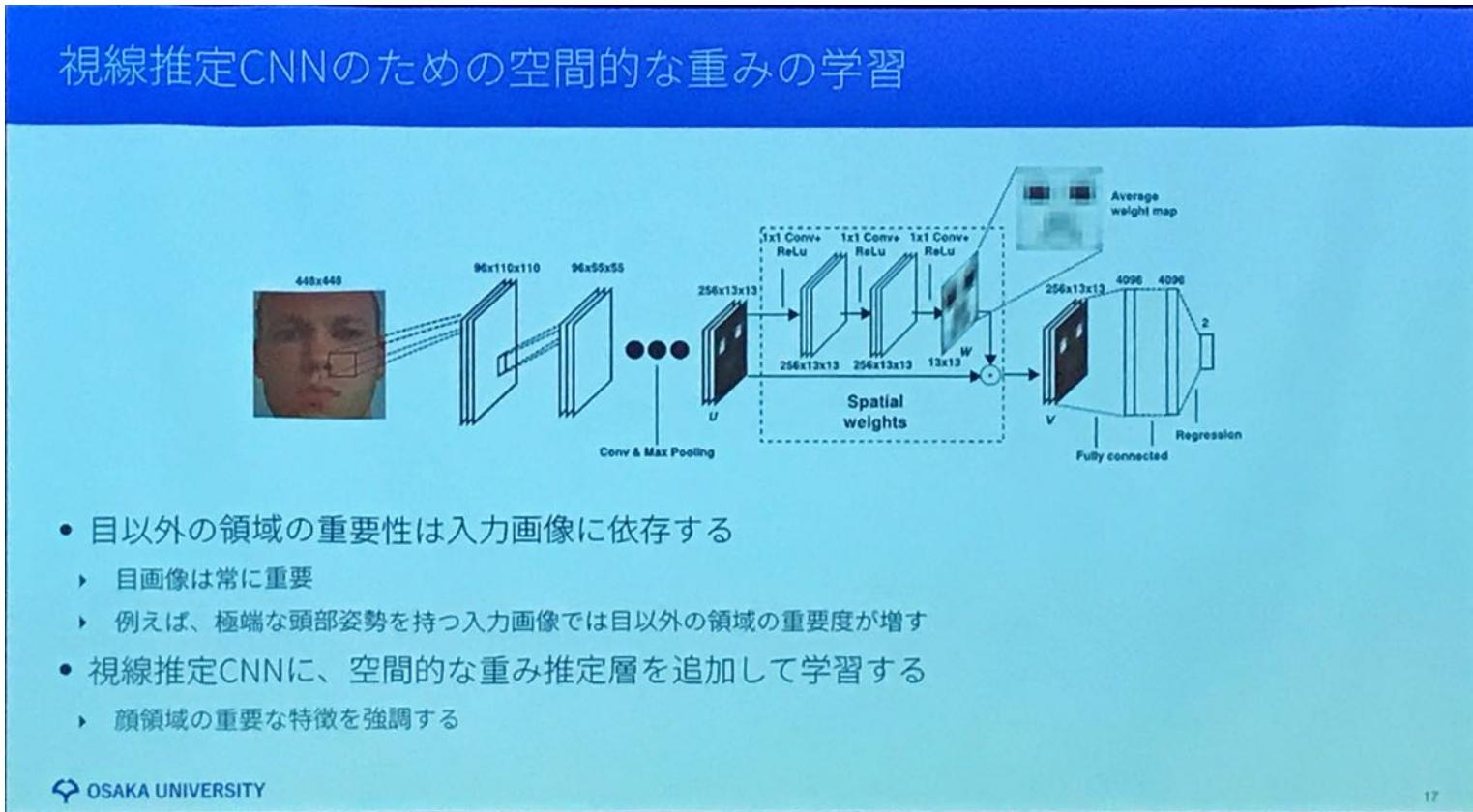


[Khosla et al., CVPR'16]



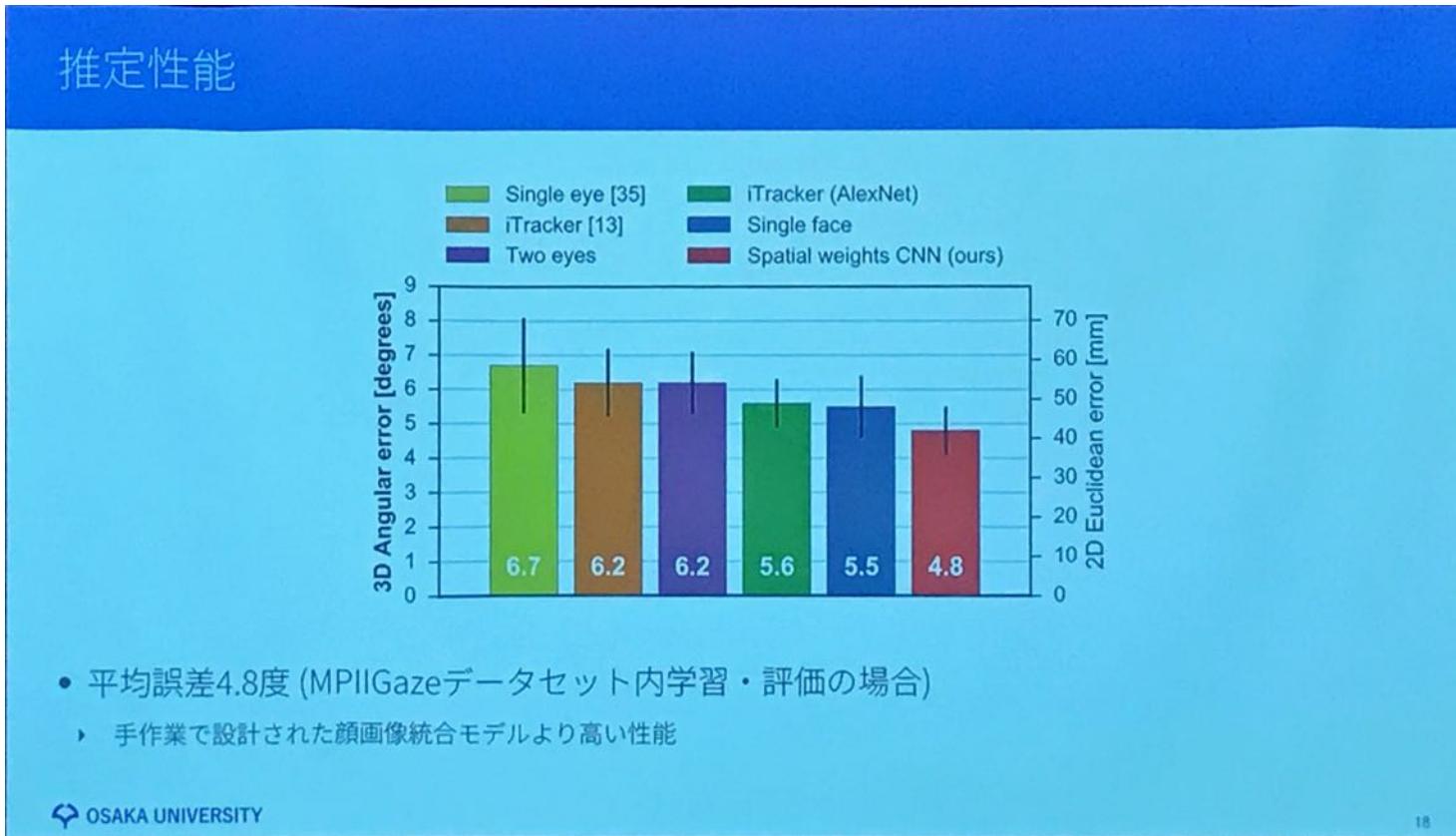
機械学習による視線推定とその実世界応用

大阪大学大学院 情報科学研究科 菅野 裕介



機械学習による視線推定とその実世界応用

大阪大学大学院 情報科学研究科 菅野 裕介



機械学習による視線推定とその実世界応用

大阪大学大学院 情報科学研究科 菅野 裕介

アピアランスベース視線推定の実世界応用

- 画像ベースの認識結果を実際のアプリケーションに使うことはできる？



OSAKA UNIVERSITY

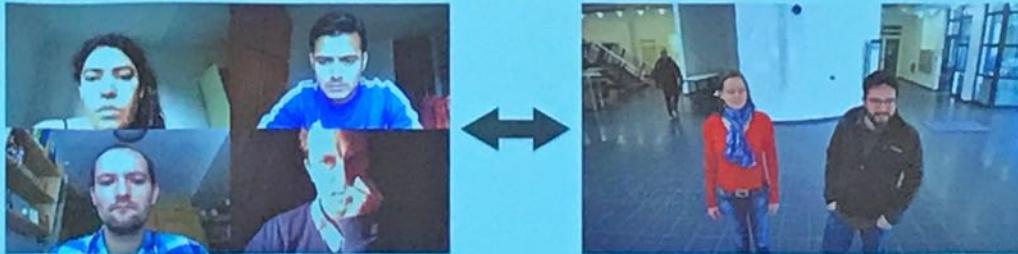
19

機械学習による視線推定とその実世界応用

大阪大学大学院 情報科学研究科 菅野 裕介

技術的課題

- 学習環境とテスト・応用環境の違い
 - ▶ ドメイン間の違いに起因する問題は完全には解決できていない



- 理想的な条件でも精度は限られる
 - ▶ 従来の視線推定応用が想定している推定誤差は～1度

OSAKA UNIVERSITY

20

機械学習による視線推定とその実世界応用

大阪大学大学院 情報科学研究科 菅野 裕介

HCIの観点から適応的なシステムを設計するアプローチ

- 対象の環境・ユーザから学習データを獲得できるよう、システムの設計自体を工夫する
 - システムが意図している「視線方向真値」にユーザの注意を誘導する
 - ユーザの行動や人間的特性からその人がどこを見ているかを予測する



視覚的顕著生マップ
(ボトムアップな注意推定)
[Sugano et al., CVPR'10, TPAMI'13, UIST'15]

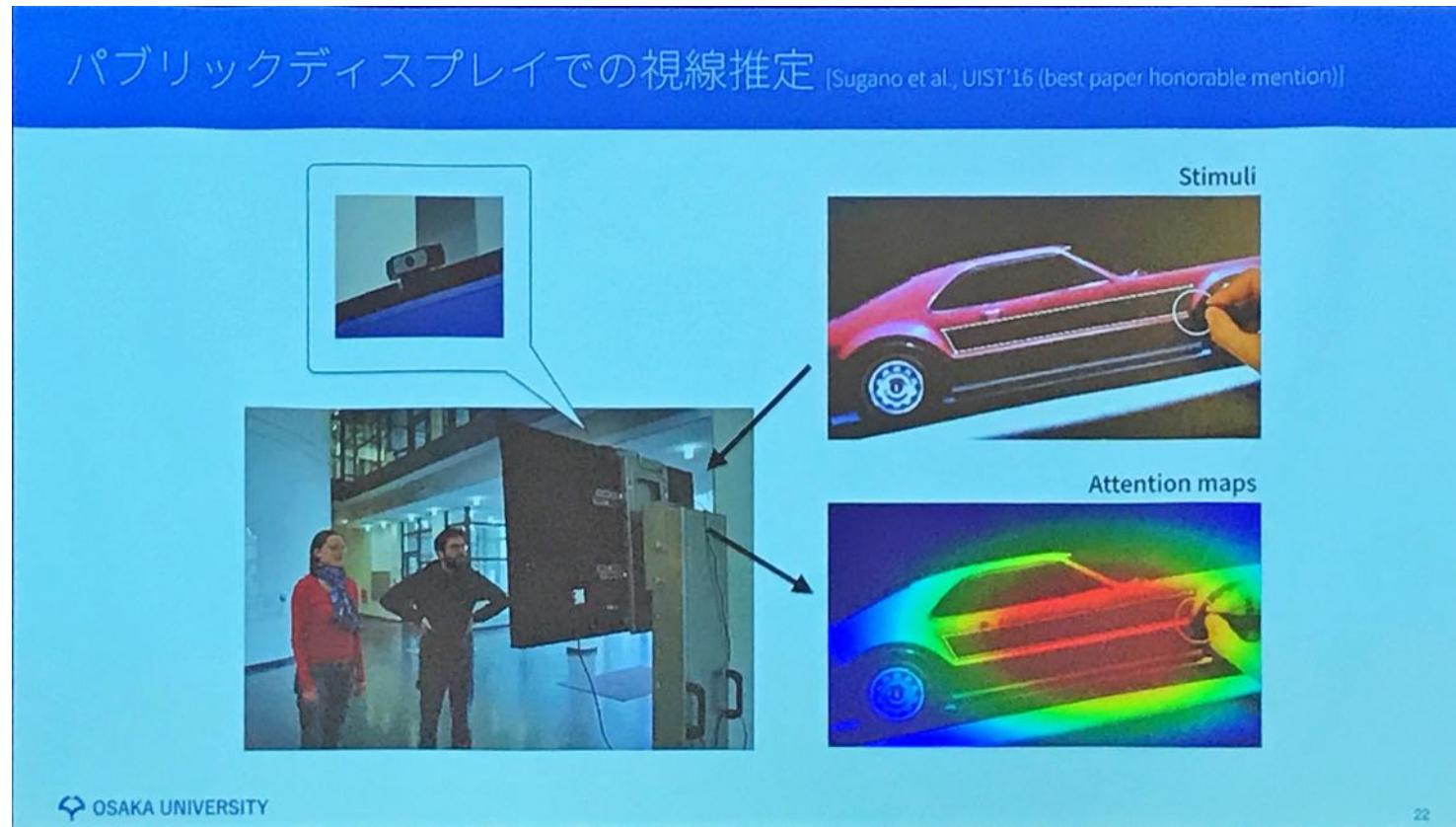


インタラクション情報
(マウスクリックなど)
[Sugano et al., ECCV'10, THMS'15]

OSAKA UNIVERSITY

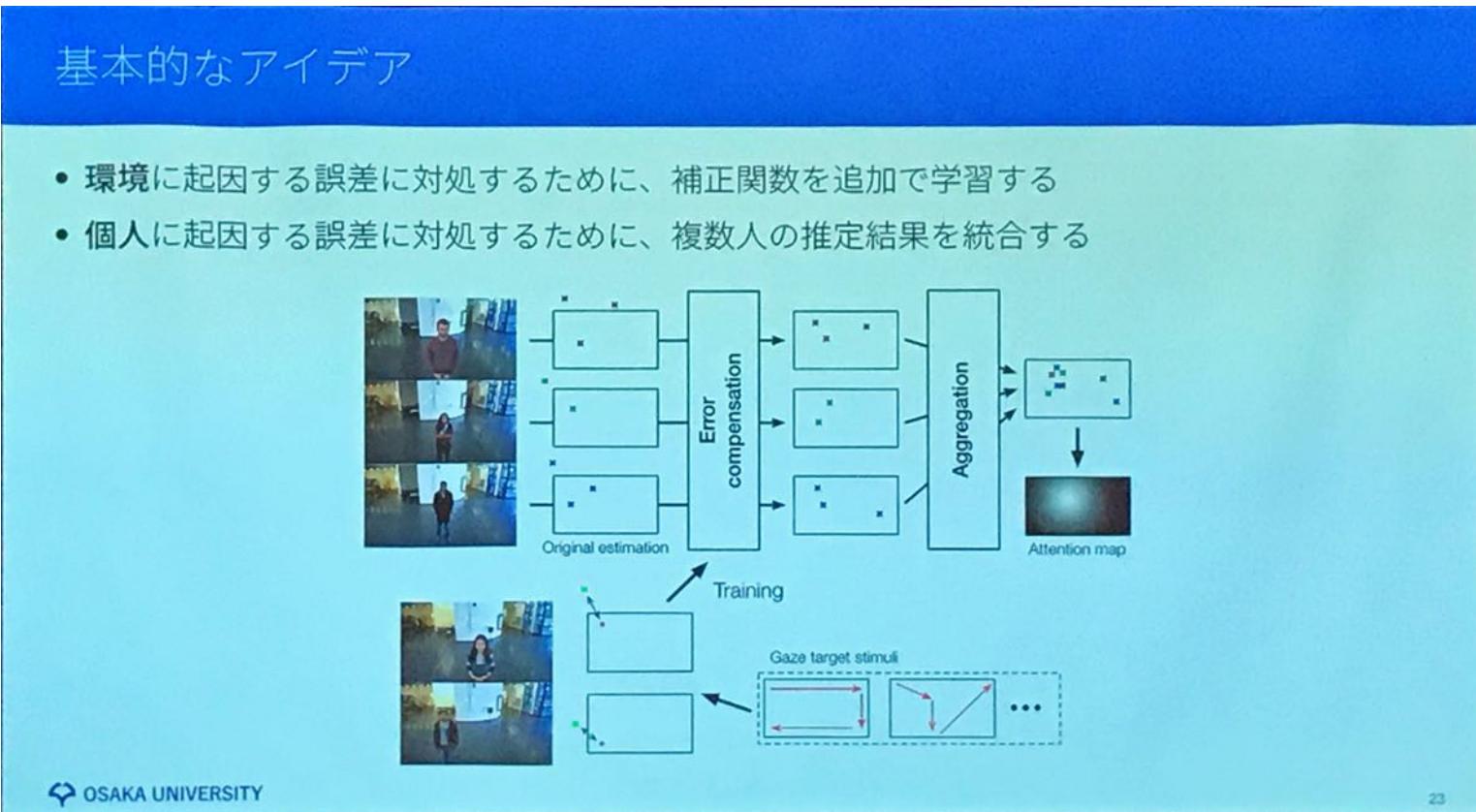
機械学習による視線推定とその実世界応用

大阪大学大学院 情報科学研究科 菅野 裕介



機械学習による視線推定とその実世界応用

大阪大学大学院 情報科学研究科 菅野 裕介

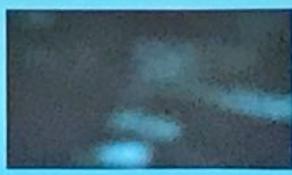


機械学習による視線推定とその実世界応用

大阪大学大学院 情報科学研究科 菅野 裕介

誤差補正関数の学習データ

- 運用前に「キャリブレーション」を行う
 - 何人かテストユーザを募って、通常の視線推定学習と同様にターゲットを見てもらえば良い
- 運用中にユーザの注視位置が予測できるコンテンツを挿入する

Stationary	Edge	Gaze template/Saliency
		
Pursuit	Text	
		

OSAKA UNIVERSITY

機械学習による視線推定とその実世界応用

大阪大学大学院 情報科学研究科 菅野 裕介

運用・評価実験

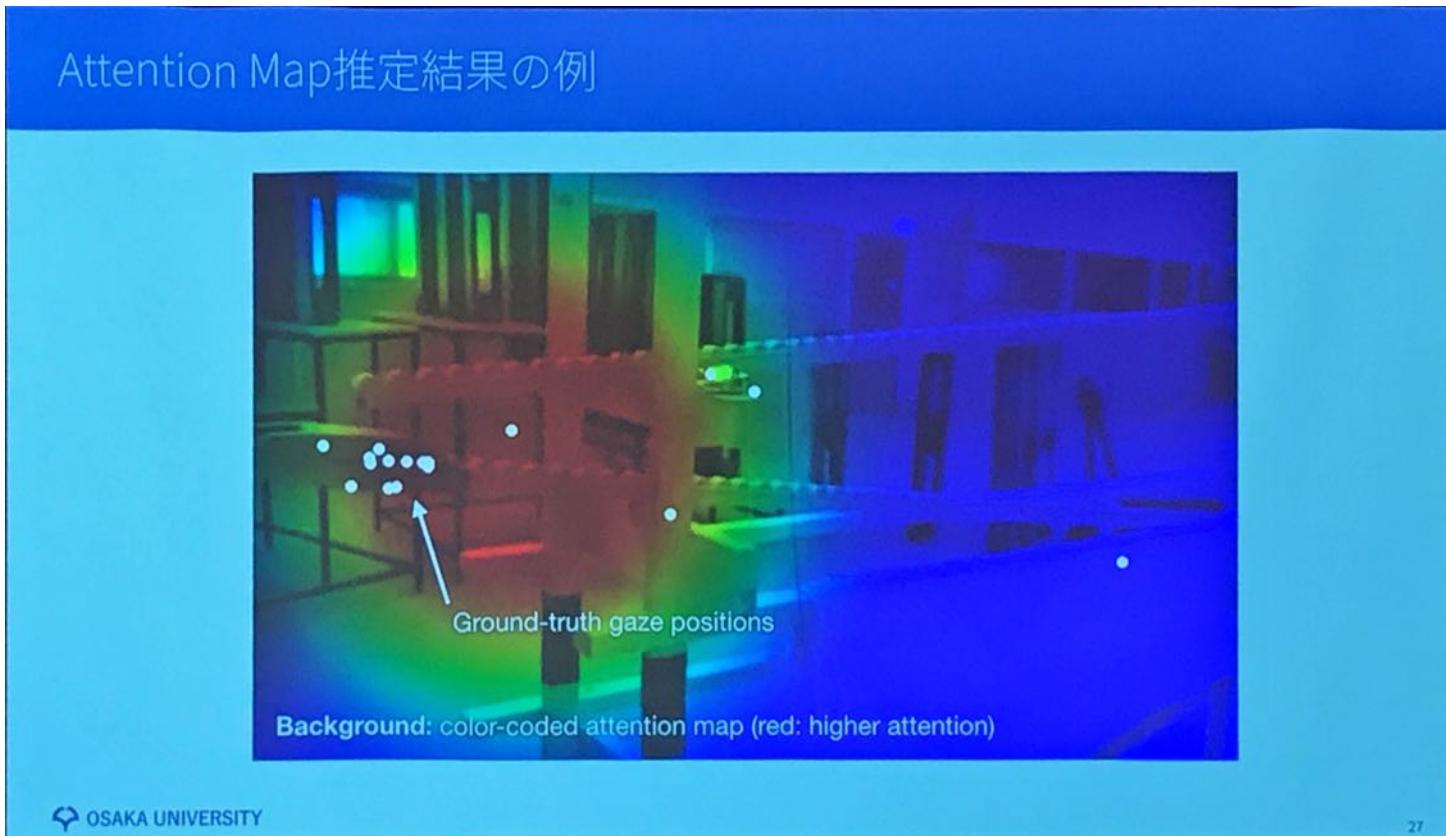


- パブリックスペースに設置したシステムで不特定多数の注視を計測
 - 複数のビデオをループで流し続けている大型ディスプレイにウェブカメラを設置
 - 12時間×13日、1ビデオあたり平均44人の顔を検出・統合
 - 商用アイトラッカーによる推定結果と比較

OSAKA UNIVERSITY

機械学習による視線推定とその実世界応用

大阪大学大学院 情報科学研究科 菅野 裕介



機械学習による視線推定とその実世界応用

大阪大学大学院 情報科学研究科 菅野 裕介

アイコンタクト検出 [Zhang et al., UIST'17 (best paper honorable mention)]

- ユーザが対象物体を見ているかどうかを識別する
 - HRIなど、アプリケーションによってはこれで十分なケースもある
- アイコンタクト検出は視線推定より容易なタスク?
 - 連続値の推定→二値分類
 - アイコンタクト検出器を事前学習するアプローチ [e.g., Smith et al., UIST'15]
- 識別境界は常に対象物体に依存する

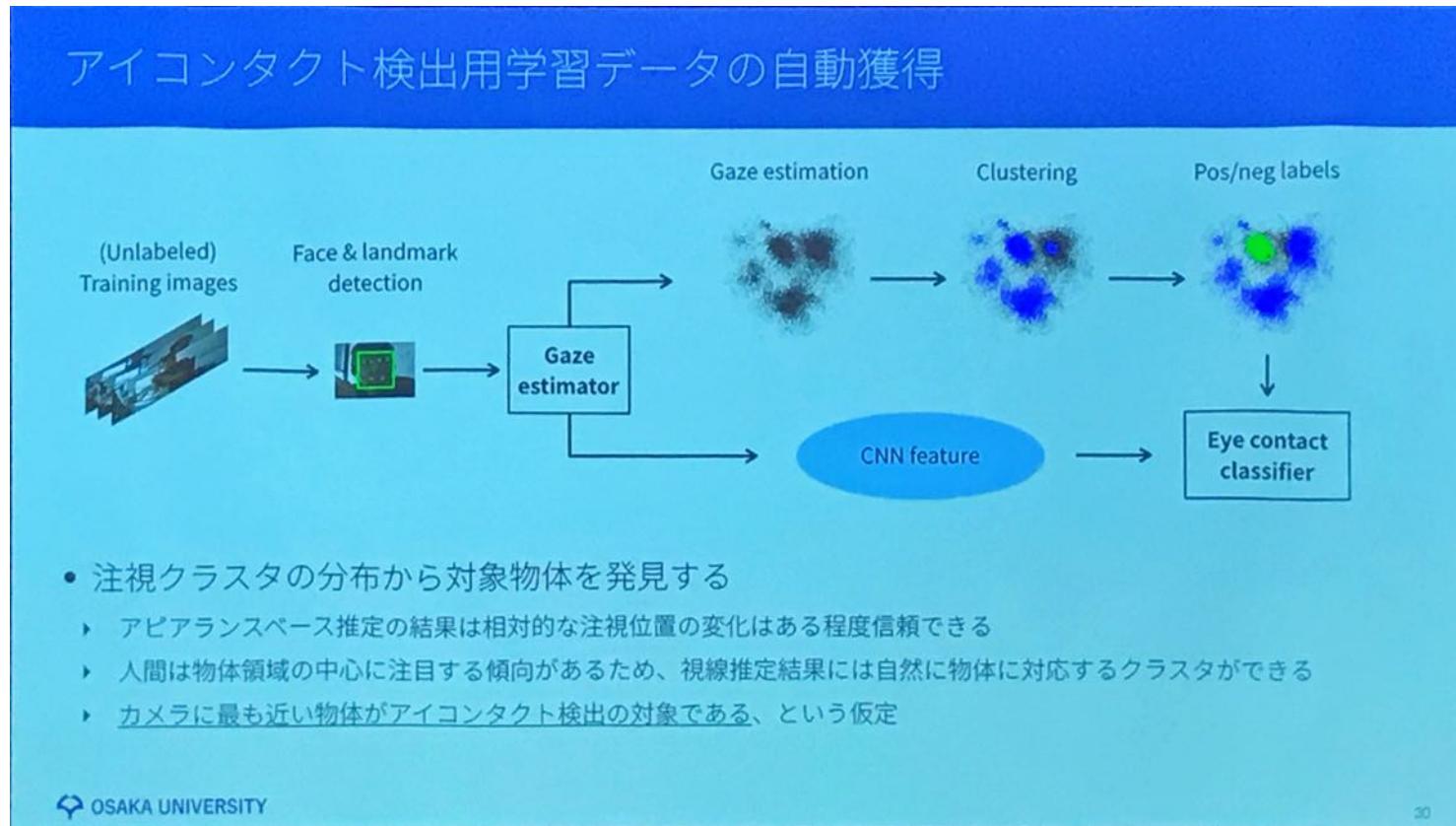


- 対象物体とカメラの位置関係が既知であれば視線推定より容易
- 事前学習することはほぼ不可能

OSAKA UNIVERSITY

機械学習による視線推定とその実世界応用

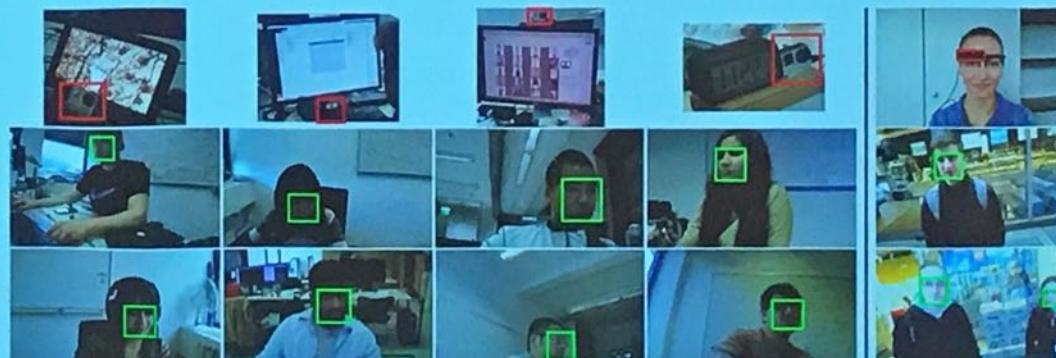
大阪大学大学院 情報科学研究科 菅野 裕介



機械学習による視線推定とその実世界応用

大阪大学大学院 情報科学研究科 菅野 裕介

運用・評価実験



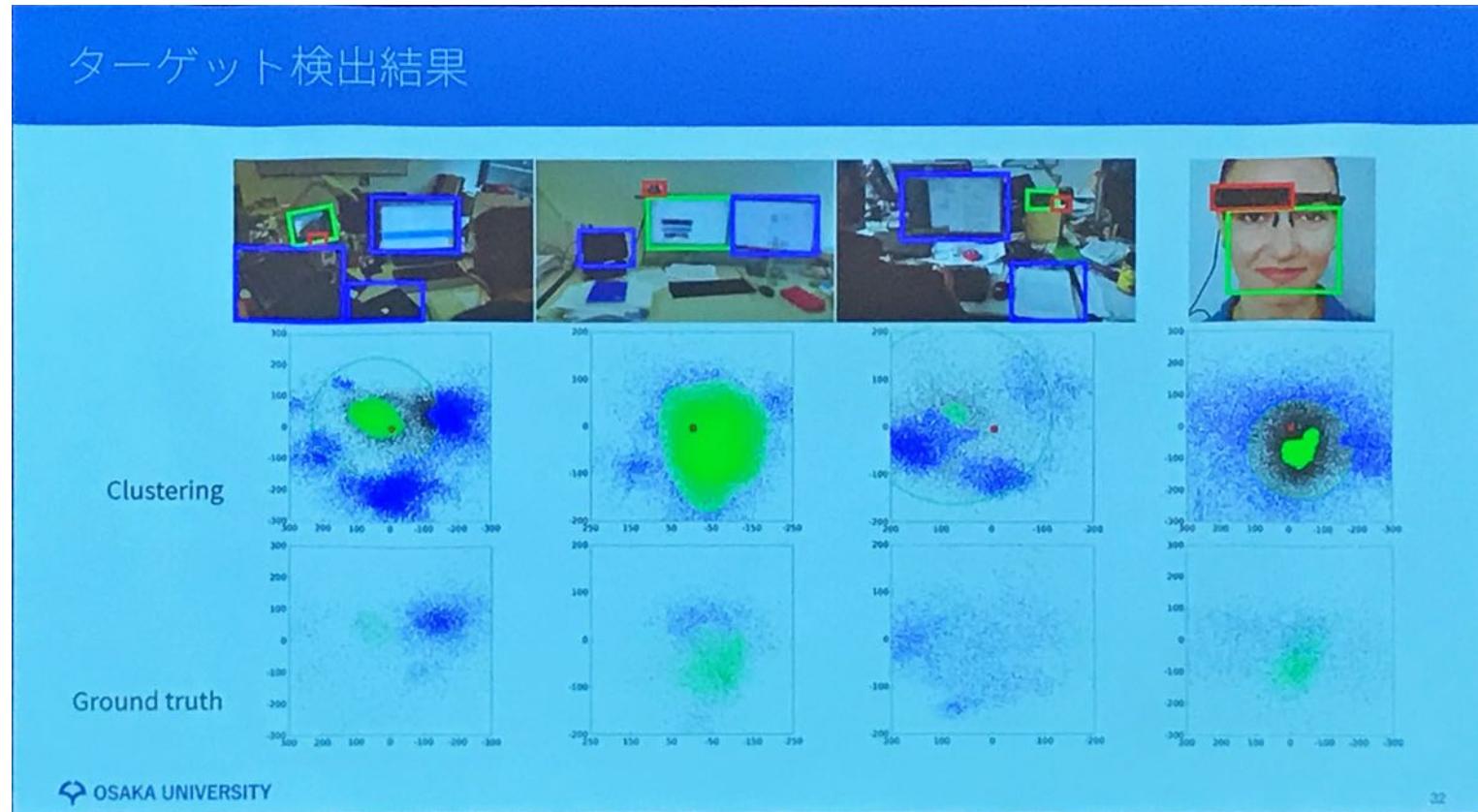
- 二つの運用シナリオでの評価
 - Object-mounted: 4通りの物体・カメラ位置の組み合わせ、14人の被験者、各3~7時間
 - Head-mounted: 3人の撮影者がおよそ28人の人物にインタビュー、合計~5時間
- 真値のアノテーションはテストデータのみ

OSAKA UNIVERSITY

31

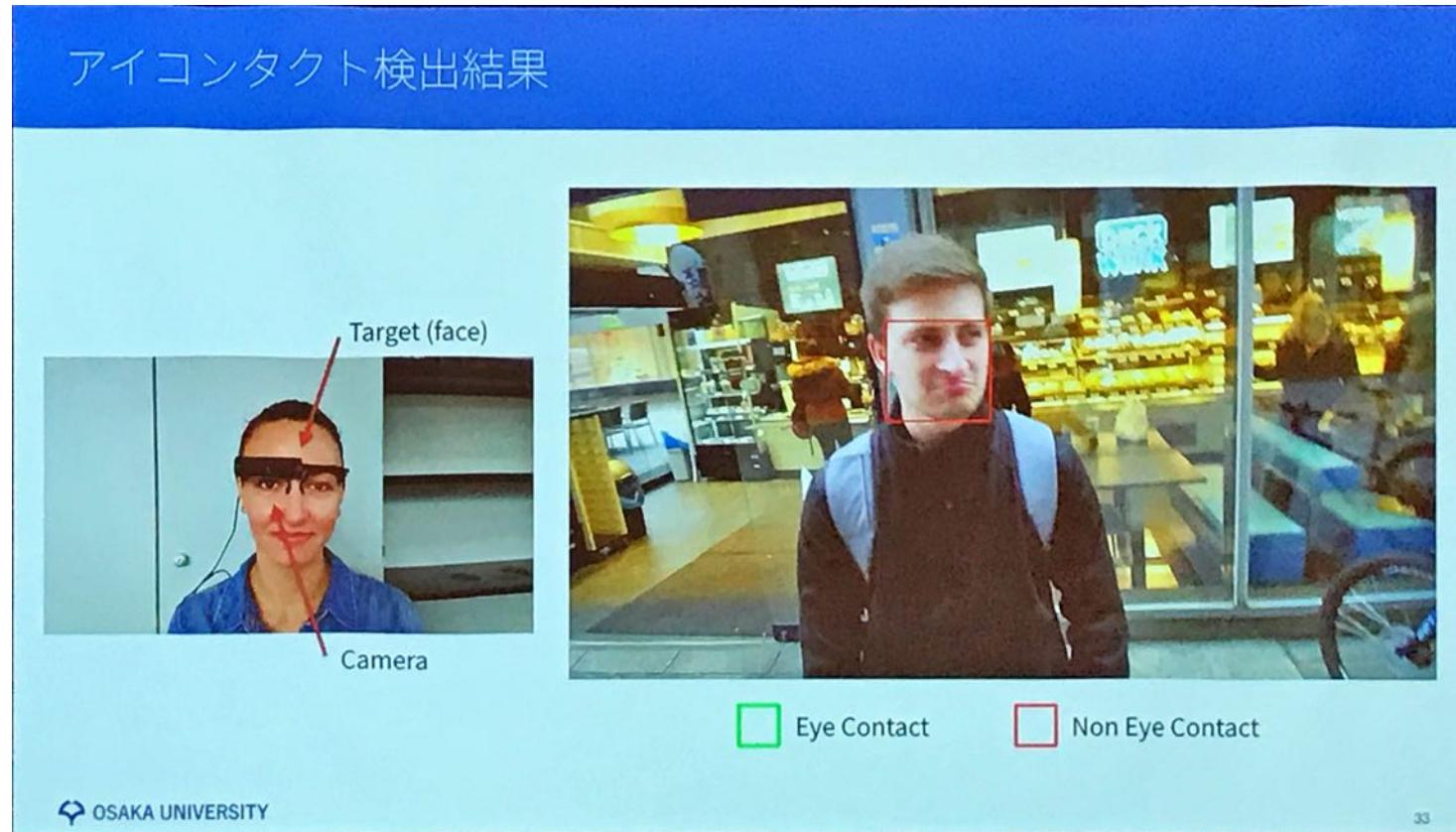
機械学習による視線推定とその実世界応用

大阪大学大学院 情報科学研究科 菅野 裕介



機械学習による視線推定とその実世界応用

大阪大学大学院 情報科学研究科 菅野 裕介



機械学習による視線推定とその実世界応用

大阪大学大学院 情報科学研究科 菅野 裕介



機械学習による視線推定とその実世界応用

大阪大学大学院 情報科学研究科 菅野 裕介

まとめ

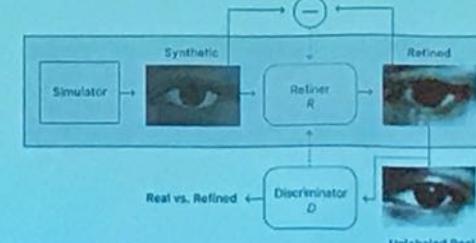
- 機械学習アプローチによる視線推定の試み
 - 大規模なデータセットから、人物非依存の視線推定器を学習する
 - 学習・評価データセット構築が重要な研究タスクの一つ
 - 3DCGによる生成画像の利用
 - 顔画像全体を入力特徴とした視線推定のためのCNNアーキテクチャ
- 実環境応用に向けたシステム設計
 - HCIの観点から学習環境とテスト環境の違いに対処する
 - 対象ユーザ・設置環境から自動的に学習データを獲得するための枠組み

機械学習による視線推定とその実世界応用

大阪大学大学院 情報科学研究科 菅野 裕介

今後の展望

- シミュレーション・生成データによる学習
 - 学習データ不足、ドメイン適応の困難さはCV全体でも重要な課題の一つ
 - GANによる画像変換を行う例
- 顔画像からの視線推定
 - 顔全体をCG生成して学習することは可能？
 - 実画像データセットも不足している
- 学習データ獲得のためのシステムデザイン
 - ユーザ・環境の多様性にHCI的な観点から取り組むことは他の認識タスクでも重要
 - 学習アルゴリズムからユーザインターフェースまで一貫した視点で取り組むことで新たな地平が開けるのでは？



[Shrivastava et al., CVPR'17]

ディープラーニングによる画像変換

早稲田大学 飯塚 里志



ディープラーニングによる画像変換

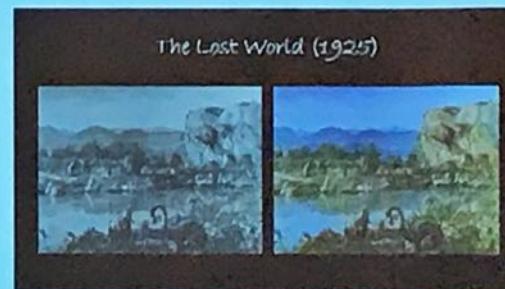
早稲田大学 飯塚 里志

自己紹介

- 飯塚里志 (IIZUKA Satoshi)
 - 早稲田大学 研究院講師
- 研究分野
 - コンピュータグラフィクス、画像処理、コンピュータビジョンなど



経年変化の再現



白黒写真の自動色付け

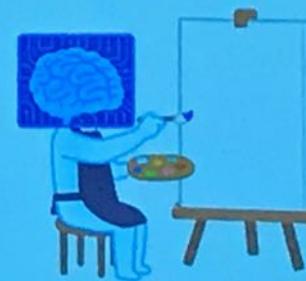
2

ディープラーニングによる画像変換

早稲田大学 飯塚 里志

今日お話しする内容

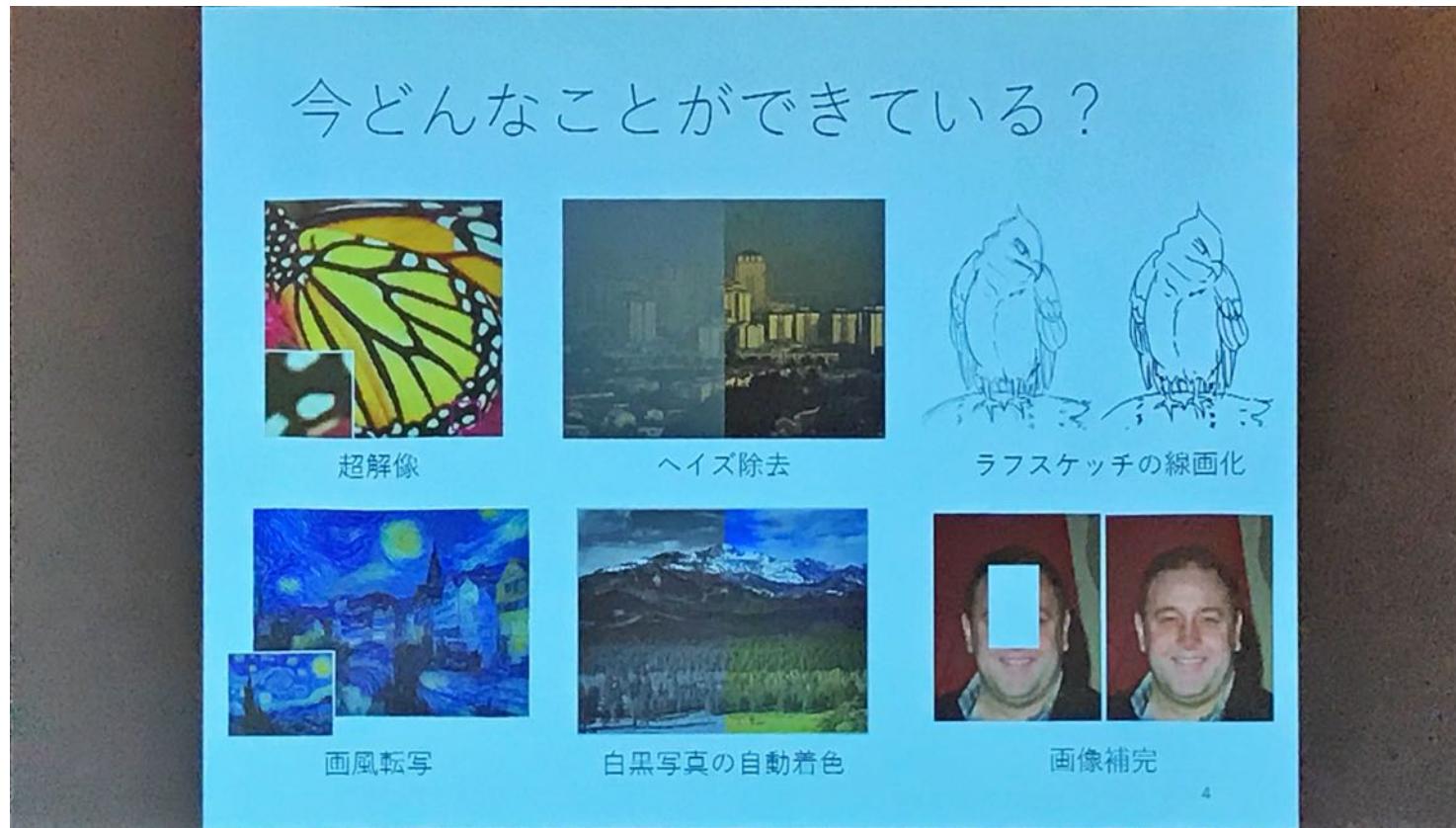
- ・画像を対象としたコンピュータグラフィクス (CG)へのディープラーニングの応用
 - ・どんな問題に使われている？
 - ・課題と解決策
- ・具体例を紹介
 - ・白黒写真的自動色付け
 - ・画像補完
 - ・ラフスケッチの線画化



3

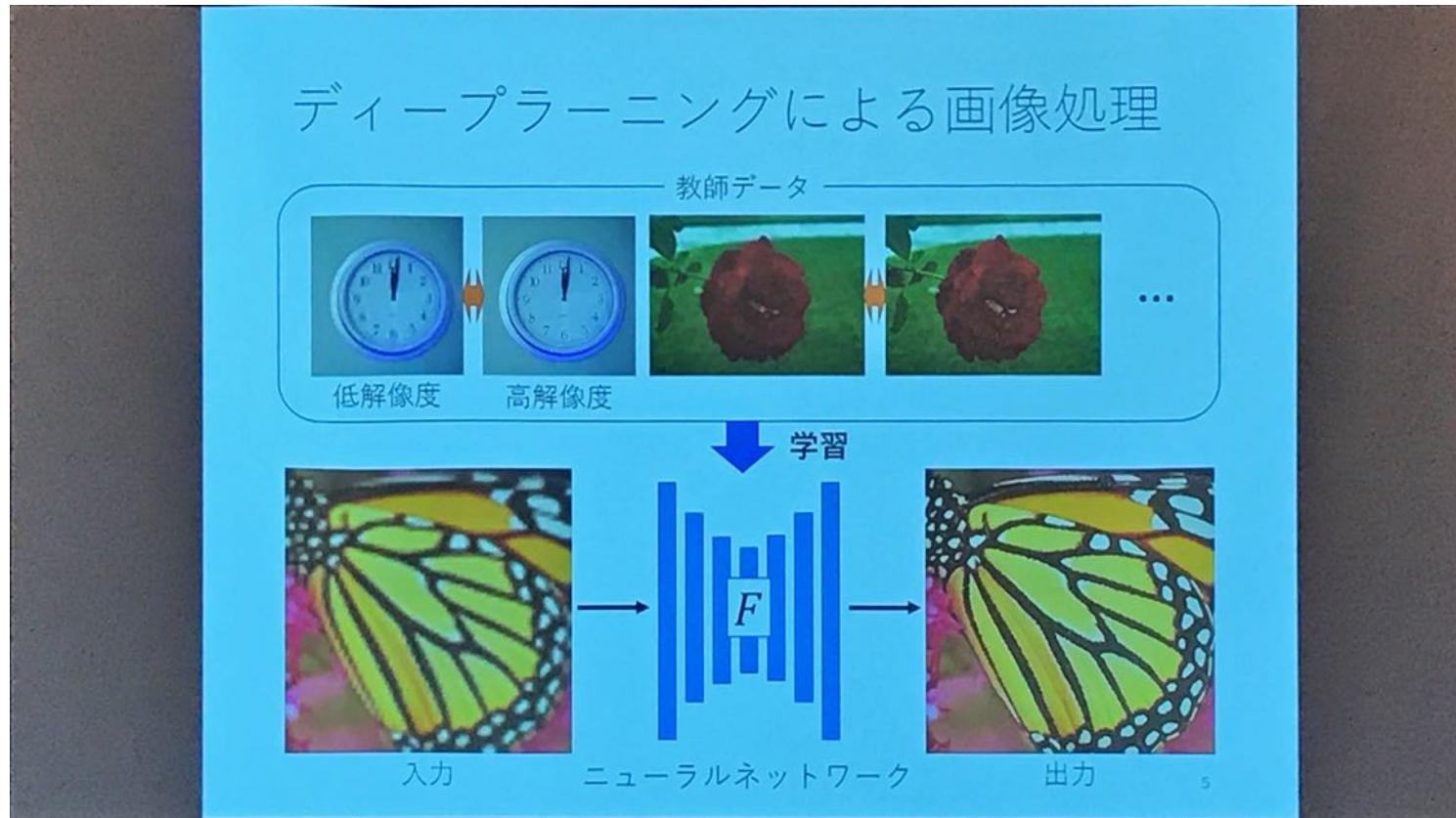
ディープラーニングによる画像変換

早稲田大学 飯塚 里志



ディープラーニングによる画像変換

早稲田大学 飯塚 里志

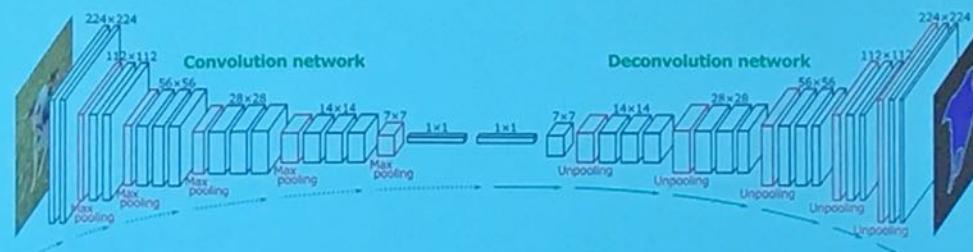


ディープラーニングによる画像変換

早稲田大学 飯塚 里志

画像処理で重要な技術

- Fully Convolutional Neural Network
 - すべての層が畳み込み層
 - 目的の画像を直接出力
- Batch Normalization
 - 各層をミニバッチ毎に正規化
 - 深いネットワークの学習に必須



Semantic Segmentation [Noh+ CVPR '15]

6

ディープラーニングによる画像変換

早稲田大学 飯塚 里志

なぜ画像処理にCNN？

- 複雑な画像処理には高レベル特徴が重要
 - 手動設計するのが困難
 - CNNは複雑な特徴を学習可能

なぜ画像処理にCNN？

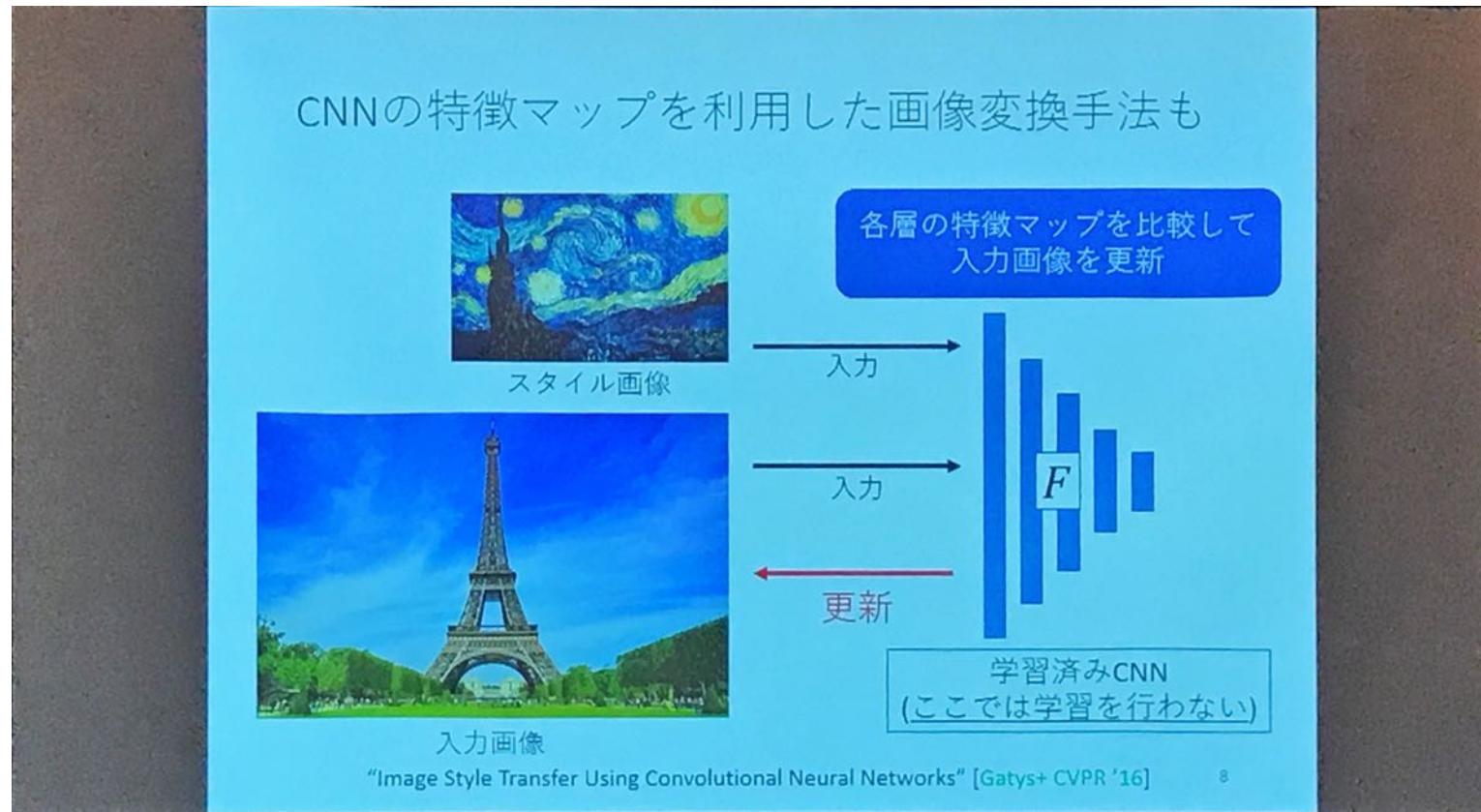
- 複雑な画像処理には高レベル特徴が重要
 - 手動設計するのが困難
 - CNNは複雑な特徴を学習可能

1層目	2層目	3層目	4層目
コーナー、エッジ、色 (低レベル特徴)	模様 (中レベル特徴)	犬の顔、鳥の足 (高レベル特徴)	

"Visualizing and Understanding Convolutional Networks" [Zeiler+ ECCV '14]

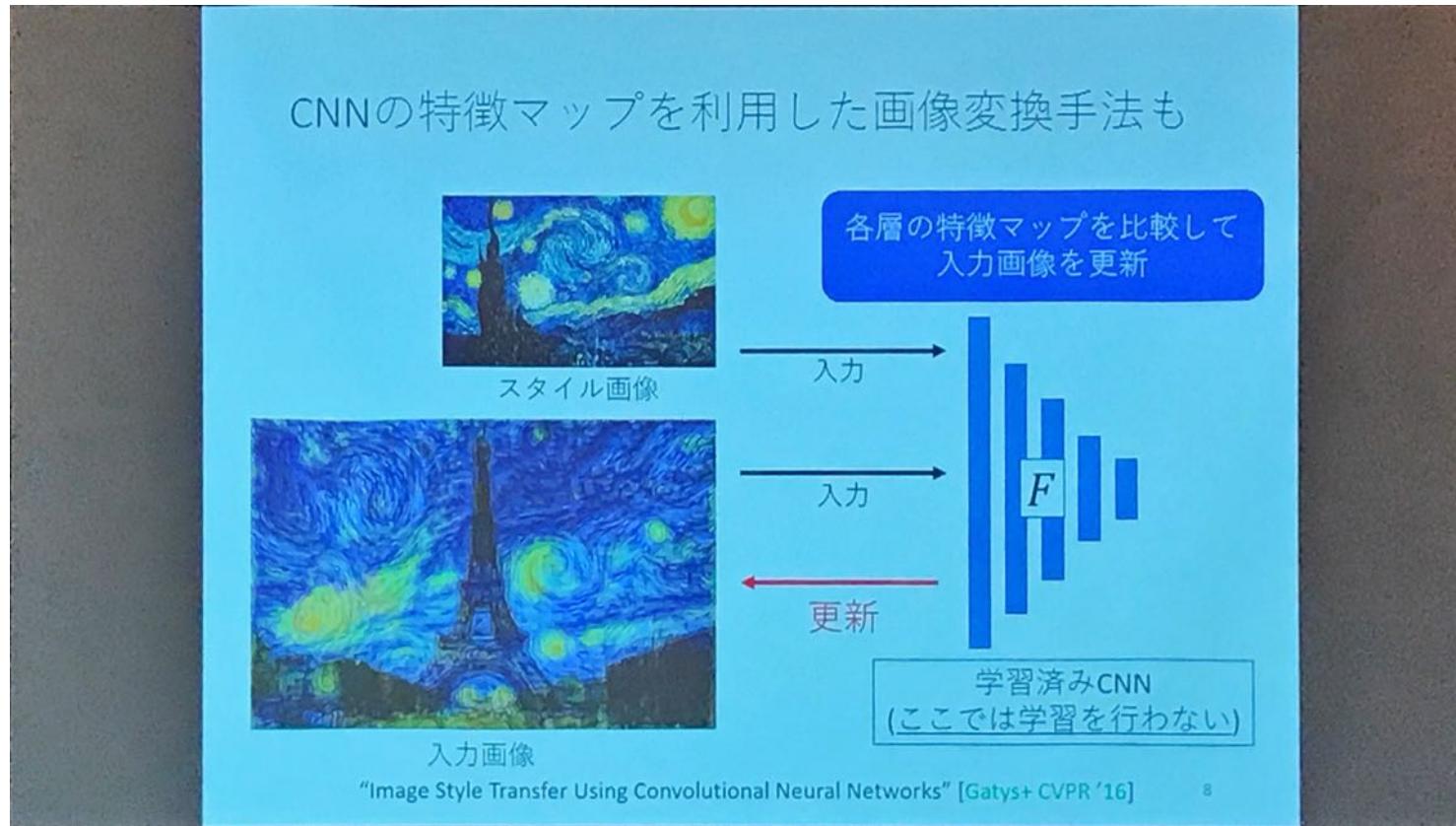
ディープラーニングによる画像変換

早稲田大学 飯塚 里志



ディープラーニングによる画像変換

早稲田大学 飯塚 里志



ディープラーニングによる画像変換

早稲田大学 飯塚 里志

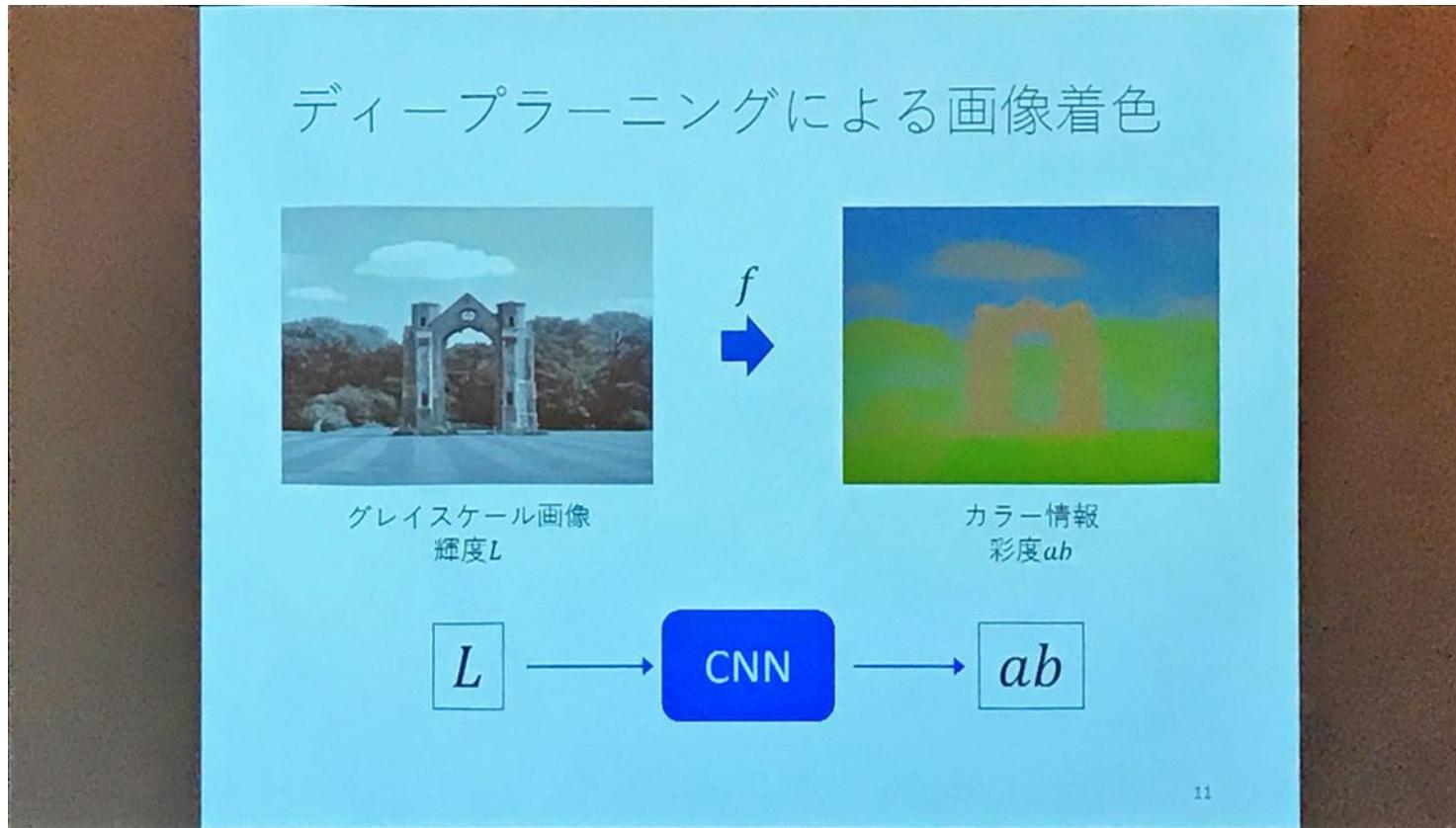
目的の画像変換を実現するための課題

- モデル構造
 - どんな変換？
 - どんな特徴が必要？
- 学習方法
 - 一般的なロス関数で学習できない場合は？
- データセット
 - データが少ない場合は？
 - E.g., イラスト

9

ディープラーニングによる画像変換

早稲田大学 飯塚 里志



ディープラーニングによる画像変換

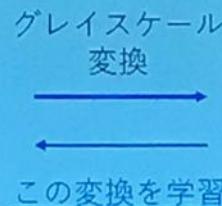
早稲田大学 飯塚 里志

教師データの作成は容易

- カラー画像をグレースケールに変換するだけ
 - 自動で大規模なデータセットを作成可能



カラー画像
(いくらでも用意可能)



グレースケール画像

12

ディープラーニングによる画像変換

早稲田大学 飯塚 里志

自動着色を実現するためには

- 各領域が何であるか
 - セグメンテーション+認識
- それが何色であるか
 - 葉っぱなら緑、空なら青など
- 大域的な情報も重要
 - シーンの認識

これらを学習できる
モデルを考える

The diagram shows a building facade and a pool of water. Four regions are labeled with thought bubbles: '木' (tree) pointing to the foliage, '空' (sky) pointing to the clear blue sky above, '壁' (wall) pointing to the building's exterior, and '水' (water) pointing to the surface of the pool. Below this, a room interior is shown with a blue square highlighted, and a question mark '空?' (Is it sky?) is placed next to it, with an arrow pointing to the ceiling.

Fully CNNによる着色

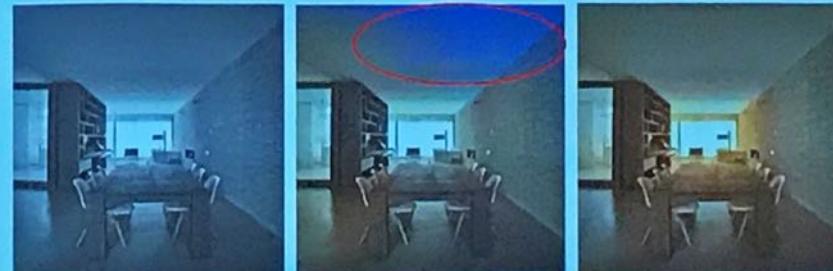
13

ディープラーニングによる画像変換

早稲田大学 飯塚 里志

大域特徴と局所特徴による自動着色
[Iizuka and Simo-Serra+ SIGGRAPH '16]

- 大域特徴と局所特徴を同時に学習し、色付けを行う
畳み込みネットワークモデルを提案



入力画像

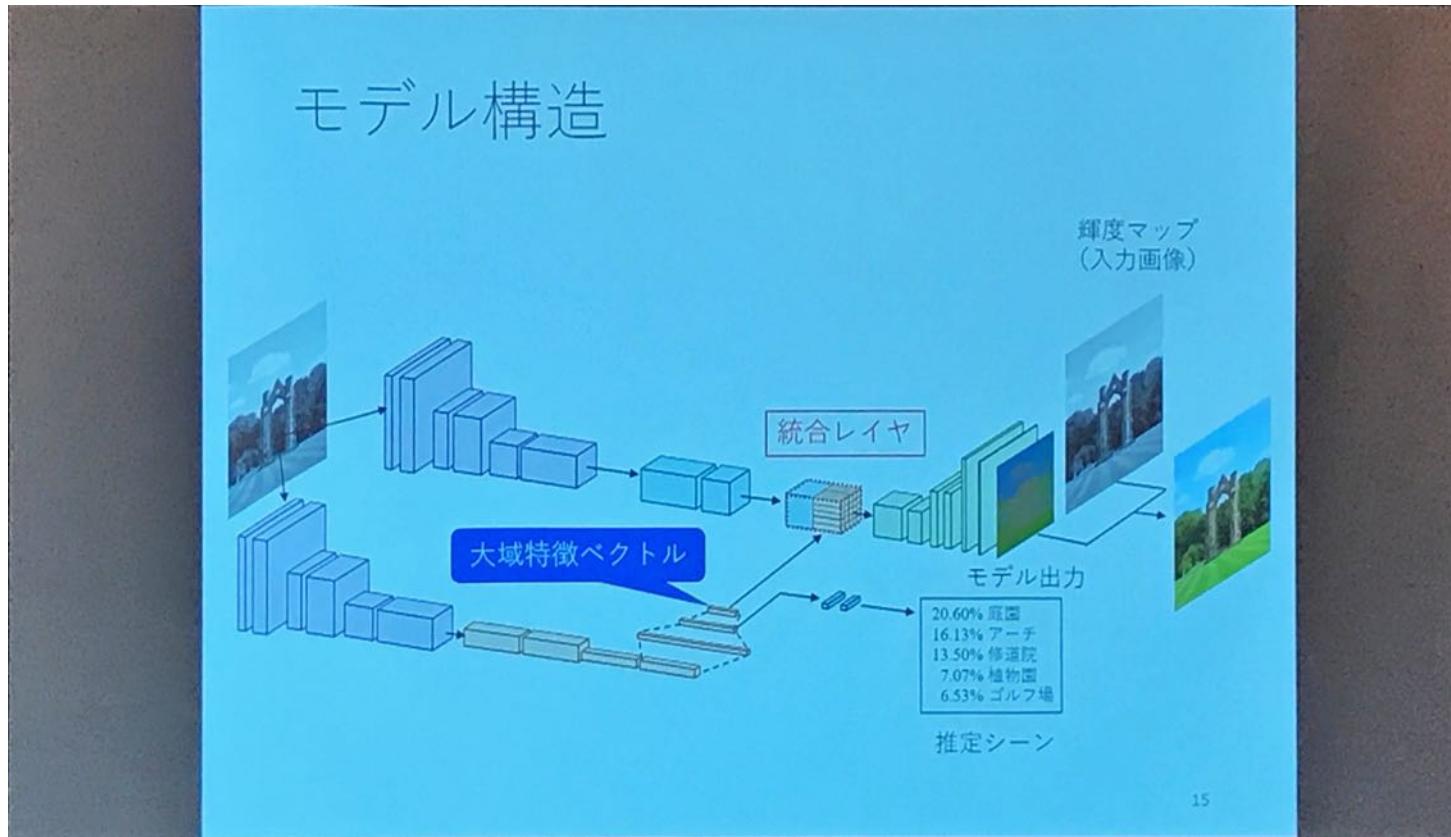
Fully CNN
(大域情報なし)

提案手法

14

ディープラーニングによる画像変換

早稲田大学 飯塚 里志



ディープラーニングによる画像変換

早稲田大学 飯塚 里志

統合レイヤ

- 特徴マップに大域特徴ベクトルを統合
- どんなサイズの特徴マップにも大域特徴を統合可能

$$y^{\text{fusion}}_{u,v} = \sigma(\mathbf{b} + W[y^{\text{mid}}_{u,v} | y^{\text{global}}_{u,v}])$$

16

ディープラーニングによる画像変換

早稲田大学 飯塚 里志



ディープラーニングによる画像変換

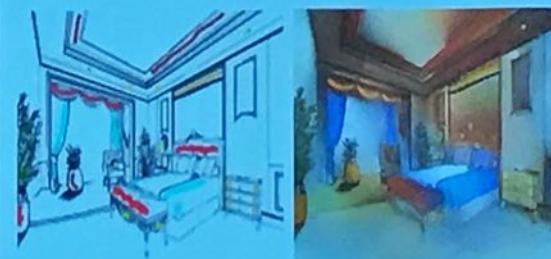
早稲田大学 飯塚 里志

ユーザ入力付きの着色

- ユーザが指定した色を出力に反映するように学習
 - 対話的な編集が可能に
 - 今後の注目テーマのひとつ



[Zhang+ SIGGRAPH '17]



[Sangkloy+ CVPR '17]

18

ディープラーニングによる画像変換

早稲田大学 飯塚 里志

ユーザ入力の学習

- 入力を「グレイスケール画像 + カラー点」に
 - 4チャンネル ($L + a + b + mask$)
 - もしくは単純にRGB
- 学習時はランダムに色をサンプリング



ディープラーニングによる画像変換

早稲田大学 飯塚 里志

目的の画像変換を実現するための課題

- モデル構造
 - どんな変換？
 - どんな特徴が必要？
- 学習方法
 - 一般的なロス関数で学習できない場合は？
- データセット
 - データが少ない場合は？
 - E.g., イラスト

→ **画像補完**

20

ディープラーニングによる画像変換

早稲田大学 飯塚 里志

従来のアプローチ

- パッチベース [Criminisi+ '04; Wexler+ '07; Simakov+ '08; Barnes+ '09; Darabi+ '12; Huang+ '14]
 - 小さな画像パッチを合成
 - 大域的な構造を考慮できない
 - 新しい物体をつくれない



パッチベースの画像補完 [Barnes+ '09]

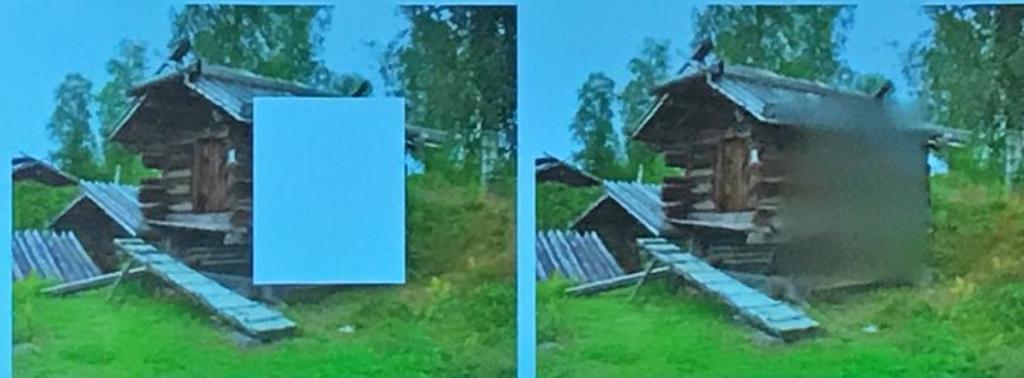
22

ディープラーニングによる画像変換

早稲田大学 飯塚 里志

一般的な学習方法だと...

- ぼやけた画像しか作れない
 - 平均二乗誤差（Mean Squared Error, MSE）を最小化
 - 複雑な画像生成は困難



24

ディープラーニングによる画像変換

早稲田大学 飯塚 里志

“Globally and Locally Consistent Image Completion” [Iizuka+ SIGGRAPH ’17]

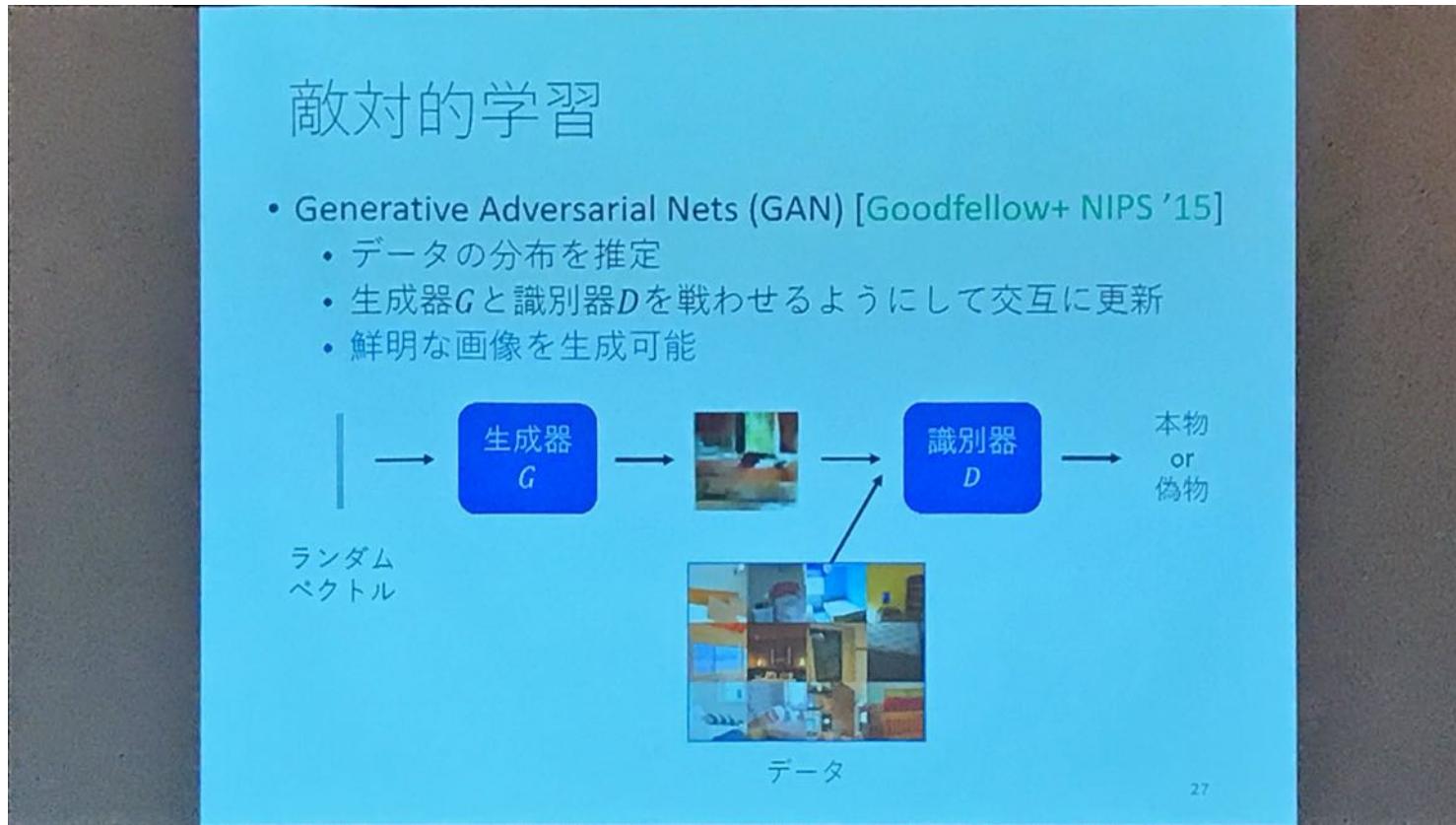
- 2つの補助ネットワークを用いた敵対的学習による補完ネットワークを提案
 - 大域的・局所的に自然な画像補完
 - 新しい物体を生成することも可能



25

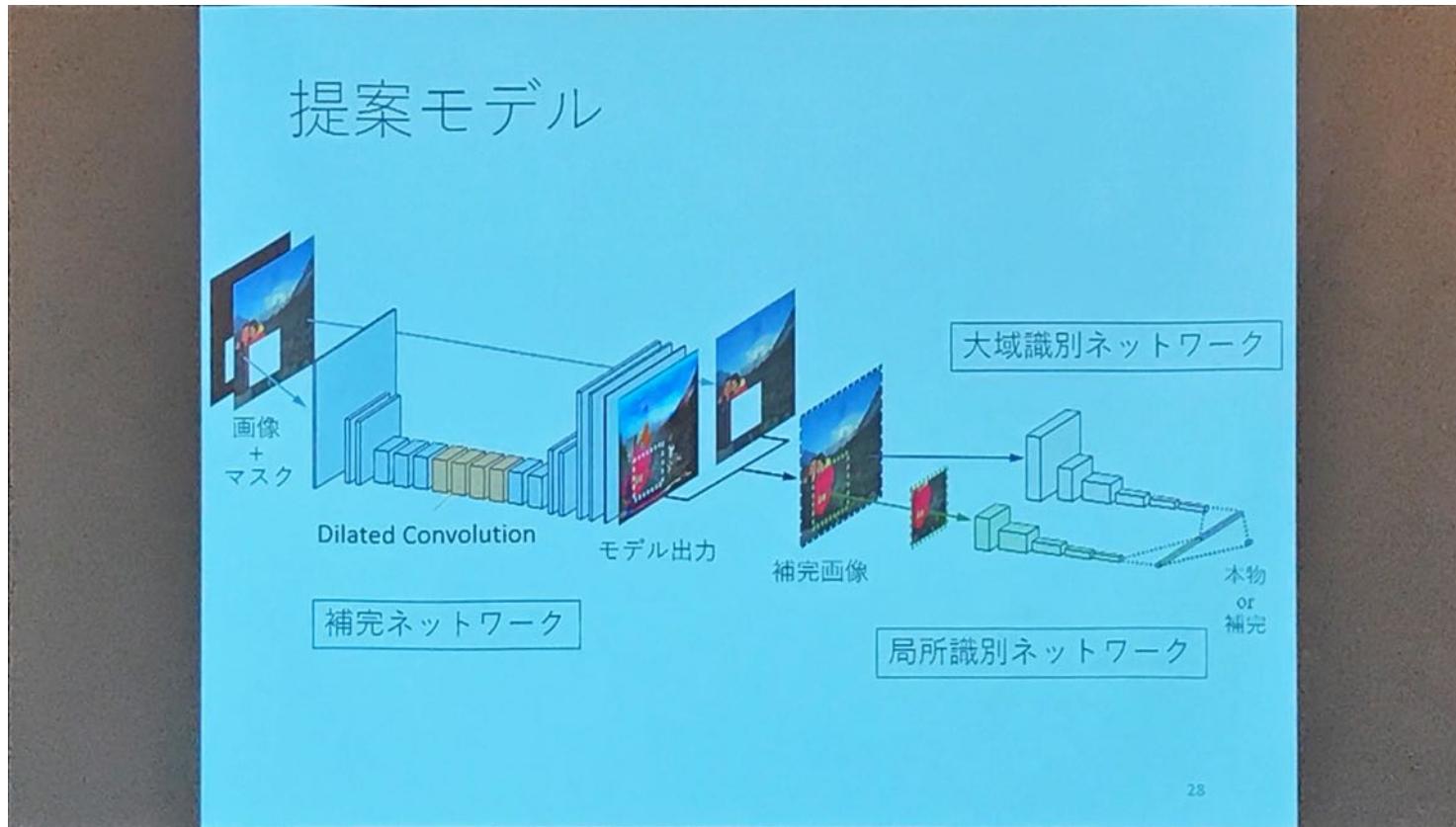
ディープラーニングによる画像変換

早稲田大学 飯塚 里志



ディープラーニングによる画像変換

早稲田大学 飯塚 里志



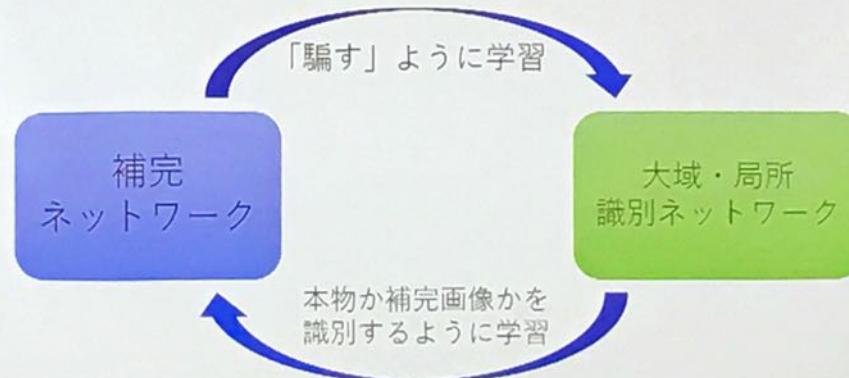
28

ディープラーニングによる画像変換

早稲田大学 飯塚 里志

学習フレームワーク

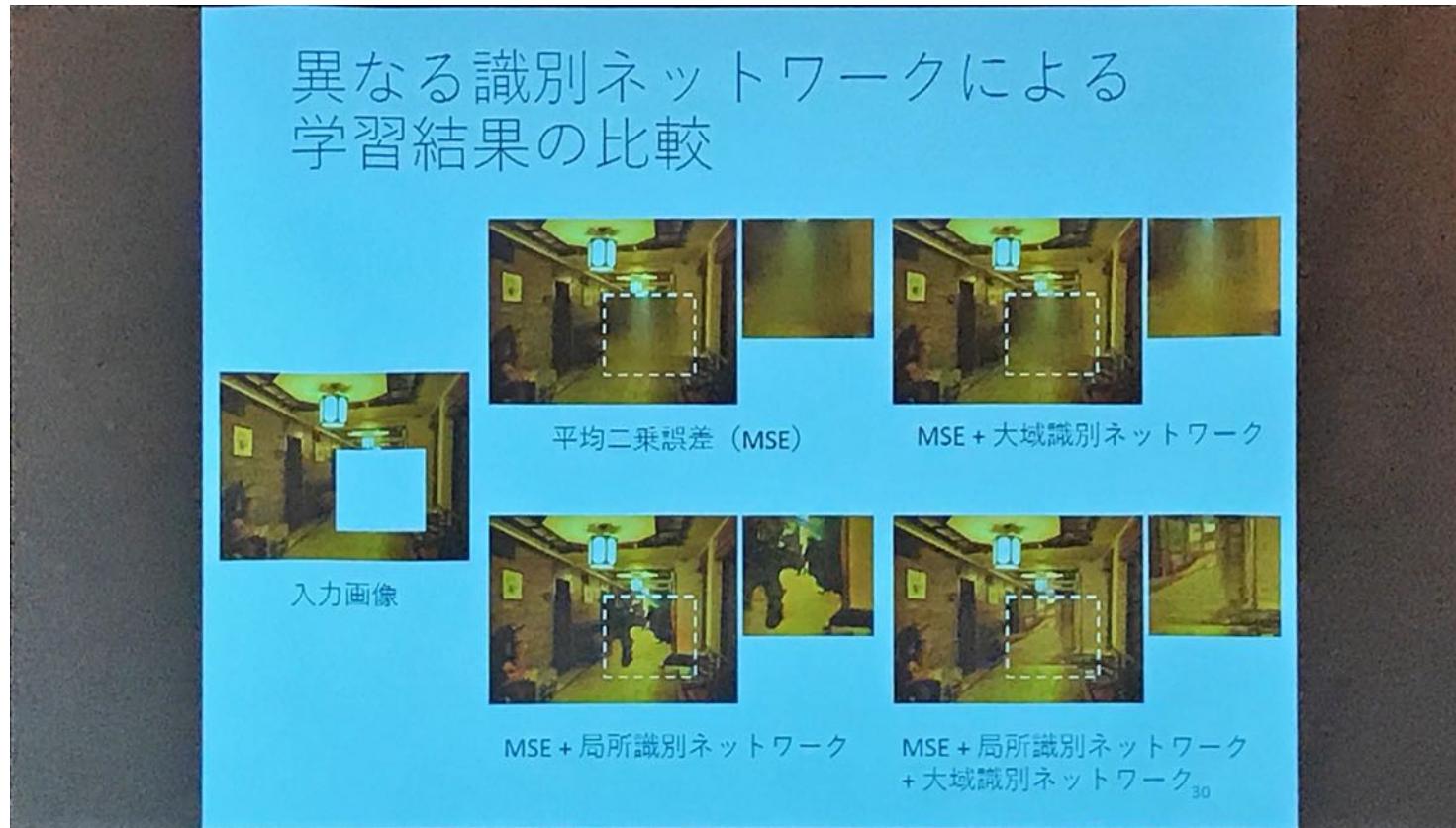
- 補完ネットワークと識別ネットワークを交互に更新
 - 敵対的学習 + MSEロス



29

ディープラーニングによる画像変換

早稲田大学 飯塚 里志



ディープラーニングによる画像変換

早稲田大学 飯塚 里志

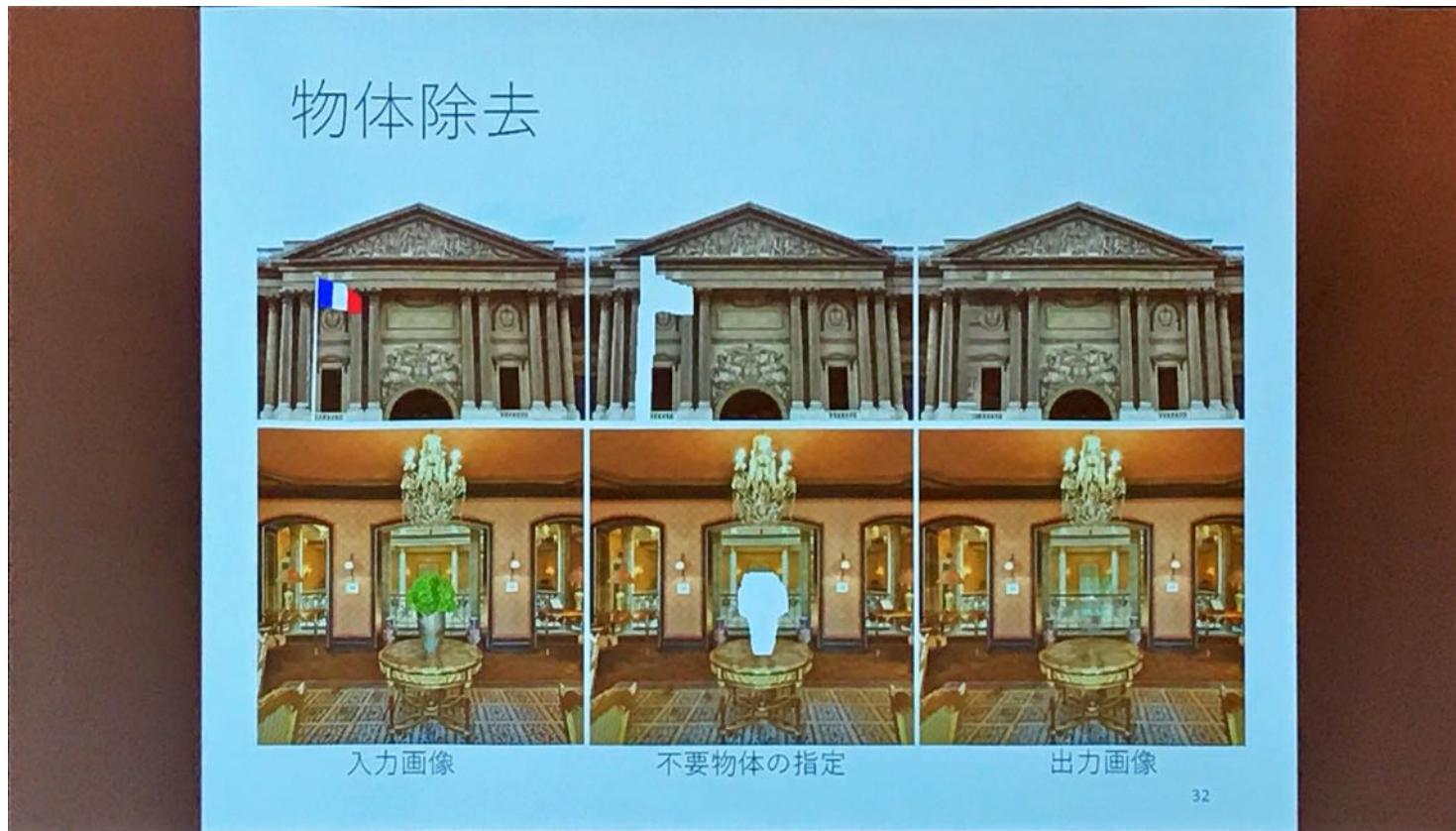
結果：自然画像の補完



31

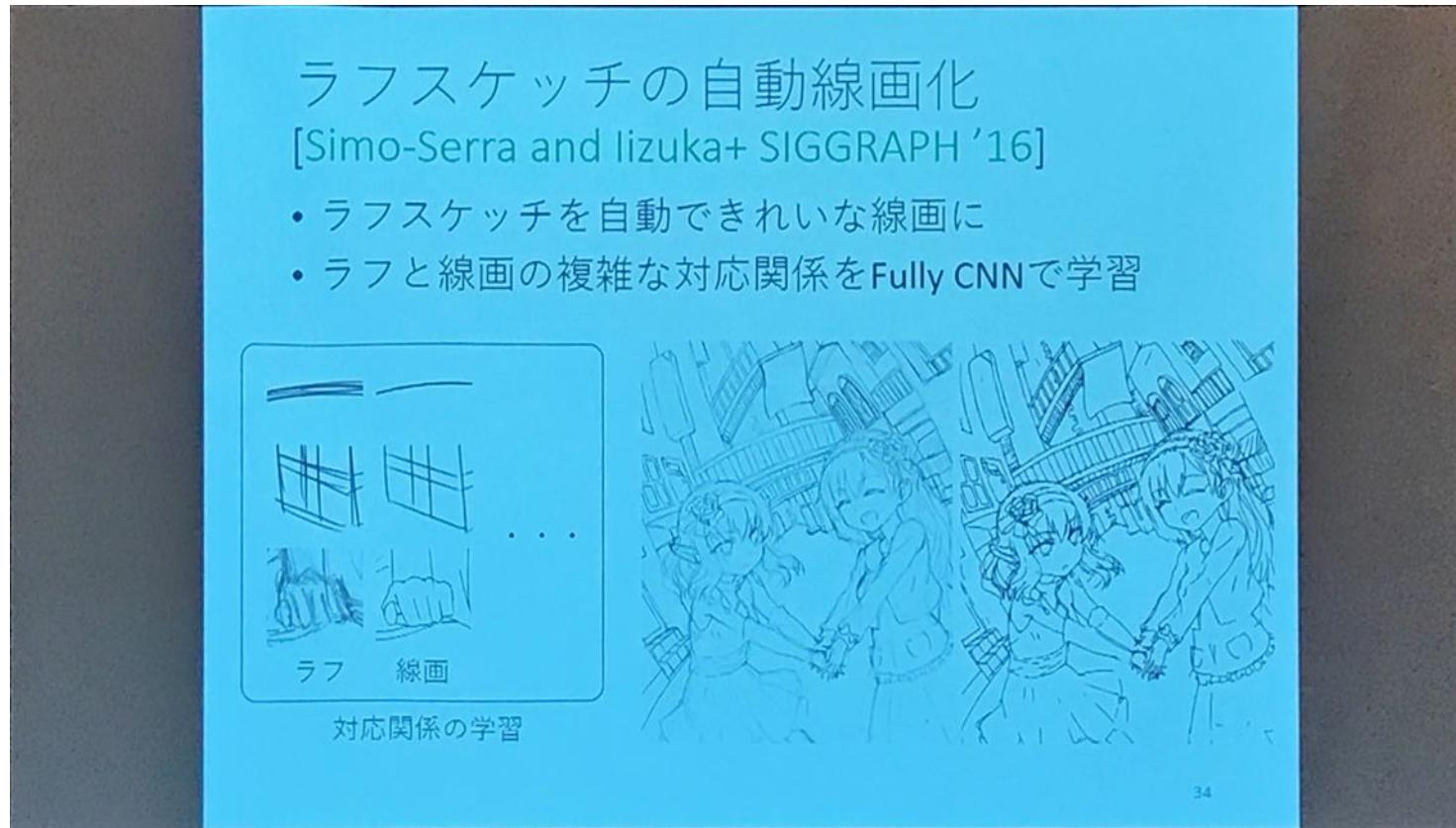
ディープラーニングによる画像変換

早稲田大学 飯塚 里志



ディープラーニングによる画像変換

早稲田大学 飯塚 里志



ディープラーニングによる画像変換

早稲田大学 飯塚 里志

教師データの難しさ

- ・ラフと線画が合っていない
- ・大量に用意するのが難しい



ラフスケッチと線画（赤線）

35

ディープラーニングによる画像変換

早稲田大学 飯塚 里志

データセットの作成

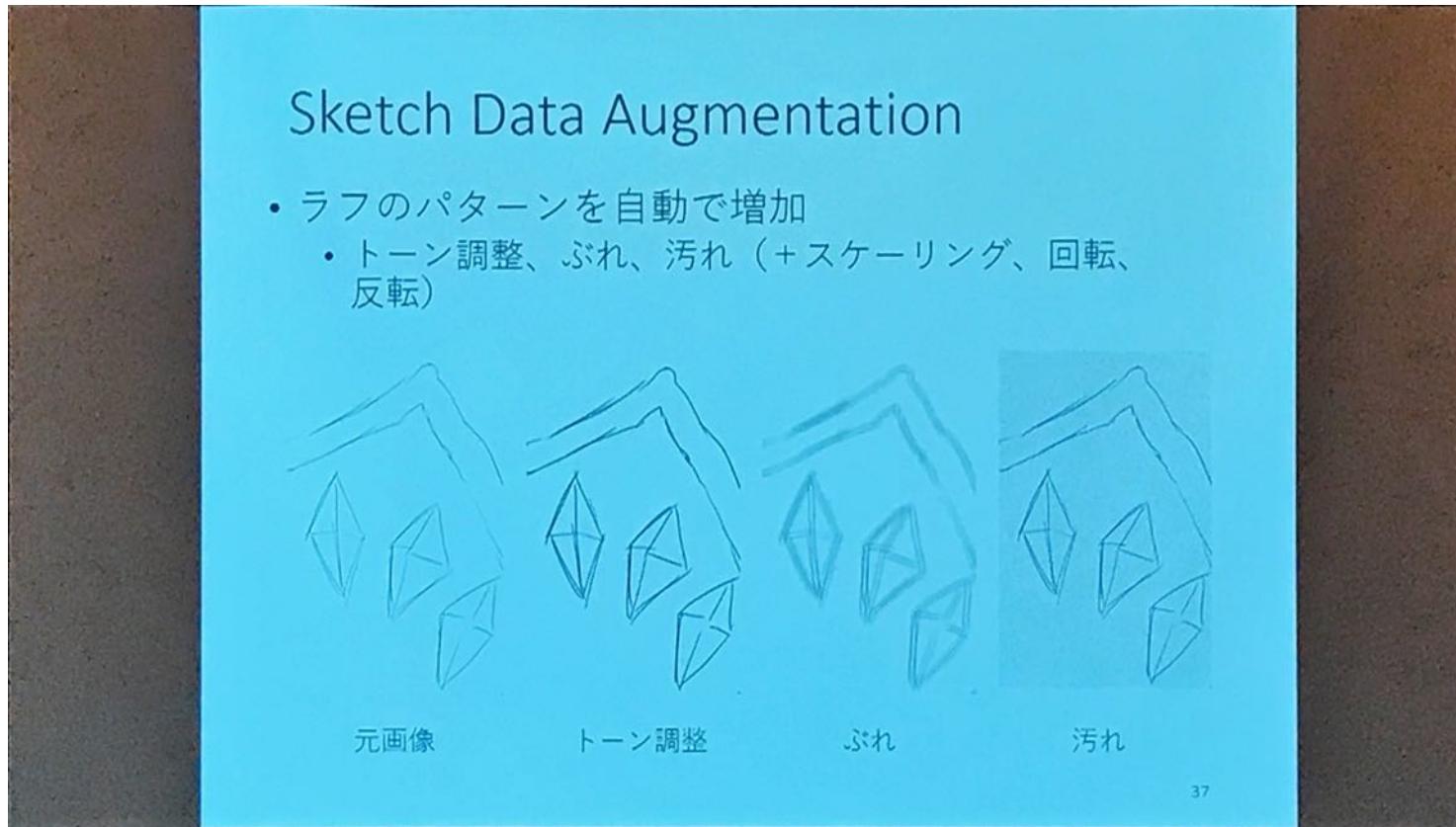
- 逆方向データ作成
 - 線画からラフスケッチを作成
 - 5人のイラストレーター、68枚（全然足りない！）



36

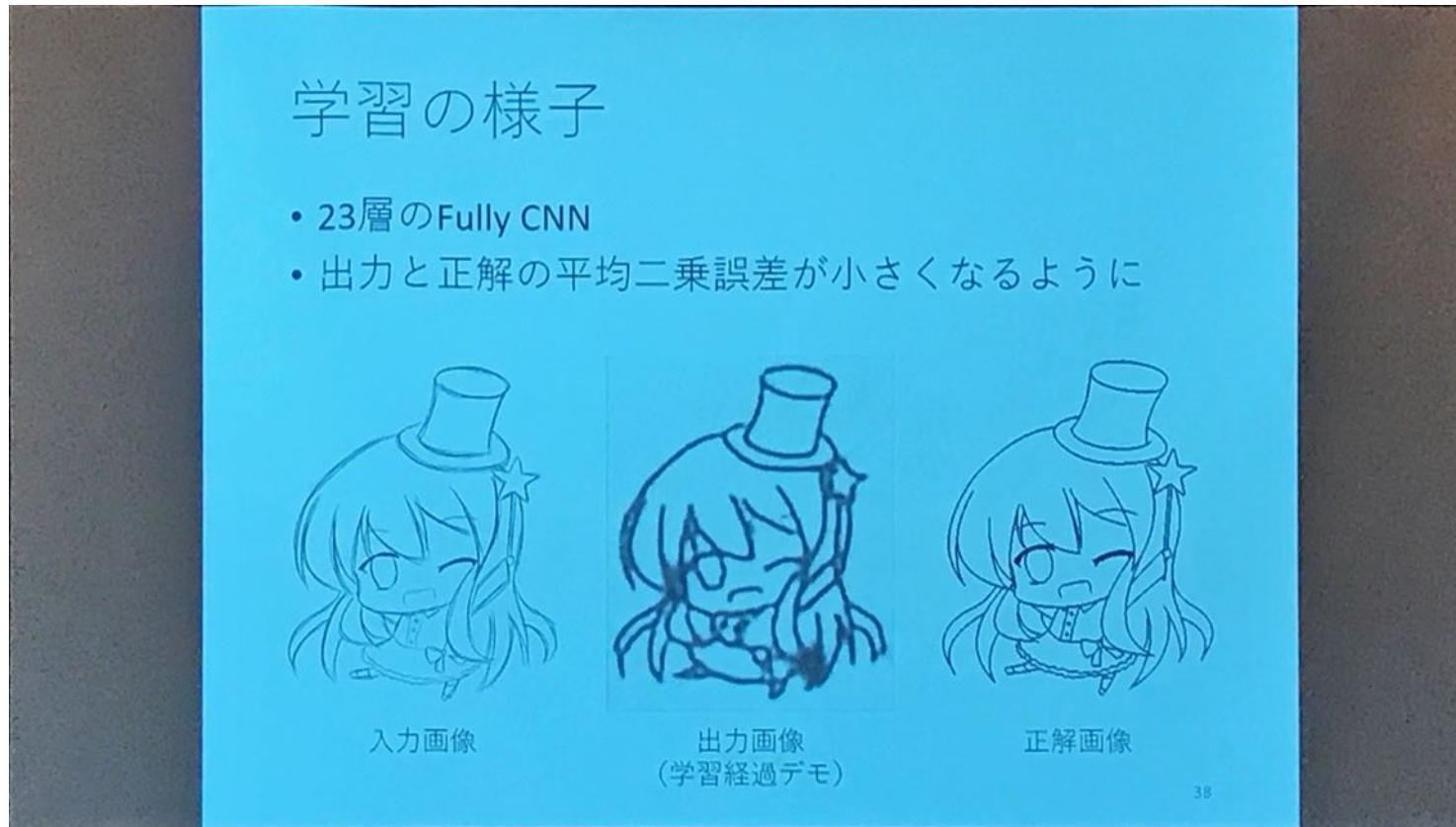
ディープラーニングによる画像変換

早稲田大学 飯塚 里志



ディープラーニングによる画像変換

早稲田大学 飯塚 里志



ディープラーニングによる画像変換

早稲田大学 飯塚 里志



ディープラーニングによる画像変換

早稲田大学 飯塚 里志

しかし...

- ・実世界には大量のイラストデータが存在
- ・なんとか対応していないデータを学習できないか



さまざまなラフスケッチと線画

40

ディープラーニングによる画像変換

早稲田大学 飯塚 里志

Adversarial Augmentation
[Simo-Serra and Iizuka+ TOG '17]

- 敵対的学习を使って線画化の精度を向上
 - 鮮明な線画を出力
 - 教師ありデータと教師なしデータを同時に学習

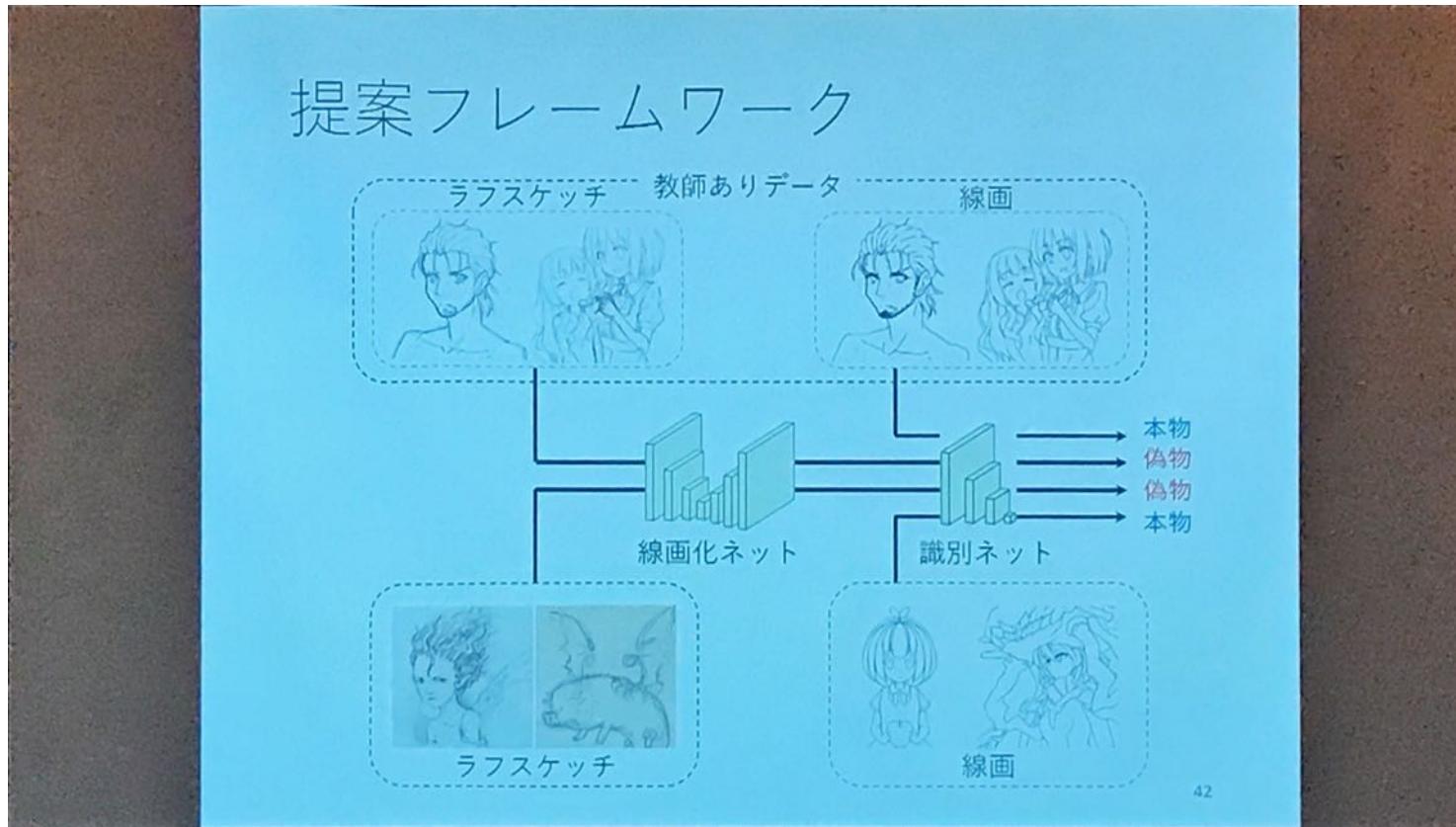


入力画像 [Simo-Serra and Iizuka+ '16] 提案手法

41

ディープラーニングによる画像変換

早稲田大学 飯塚 里志



ディープラーニングによる画像変換

早稲田大学 飯塚 里志



ディープラーニングによる画像変換

早稲田大学 飯塚 里志

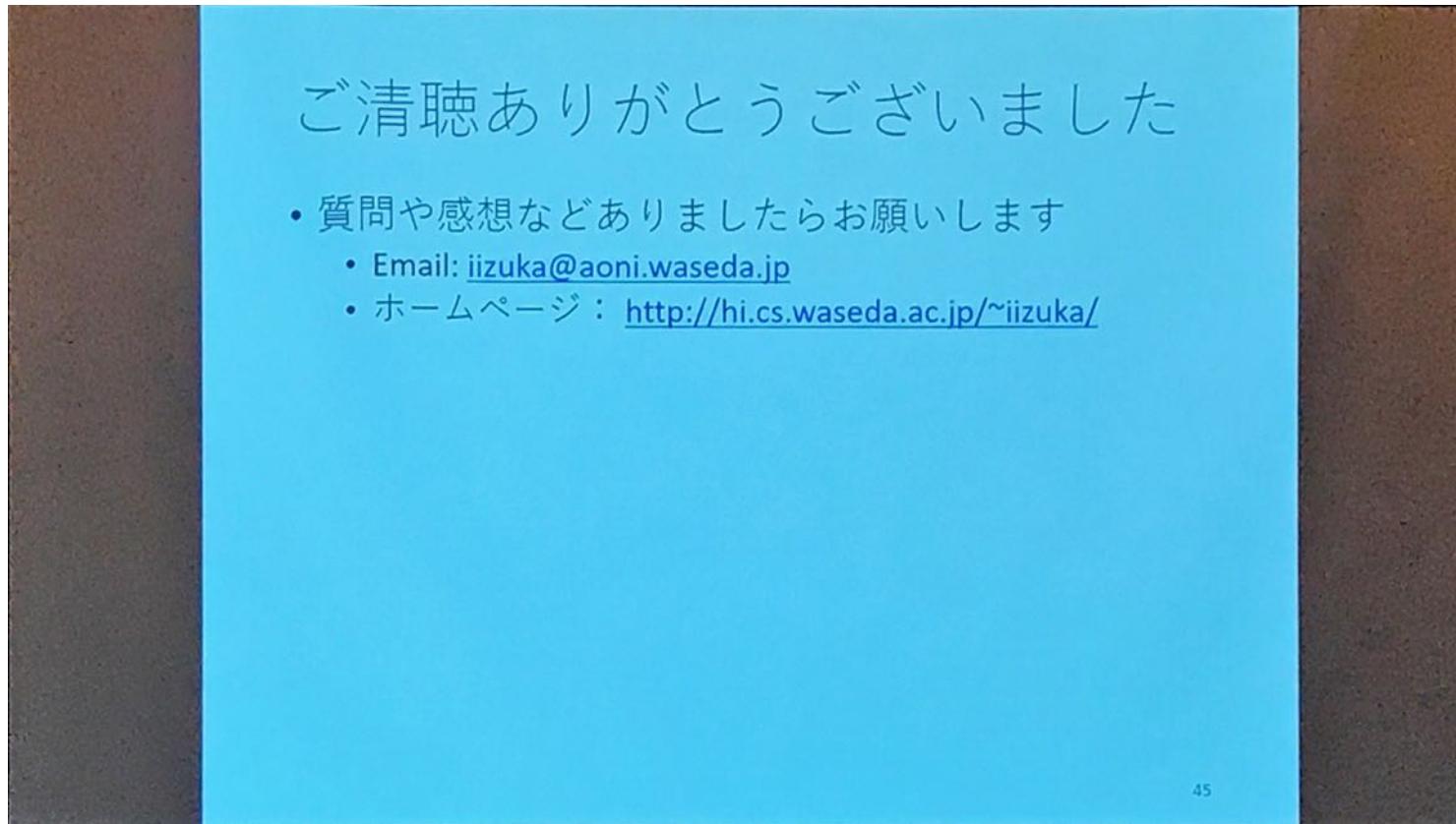
まとめ

- ディープラーニングの画像変換への応用を紹介
 - モデル構造、学習方法、データセットの重要性
 - 具体的な研究事例を解説
- 今後はより難しい問題がテーマに?
 - 扱いやすい研究テーマはどんどんやられている
- 実用化にはまだ課題も多い

44

ディープラーニングによる画像変換

早稲田大学 飯塚 里志





ARTIFICIAL INTELLIGENCE

Thank You So Much !

Tokyo AI Tech. Alliance

Github

https://github.com/xurong1981/ai_room