B3スタートアップ

コンピュータビジョンの現在と未来にやるべきこと

片岡 裕雄

http://hirokatsukataoka.net/

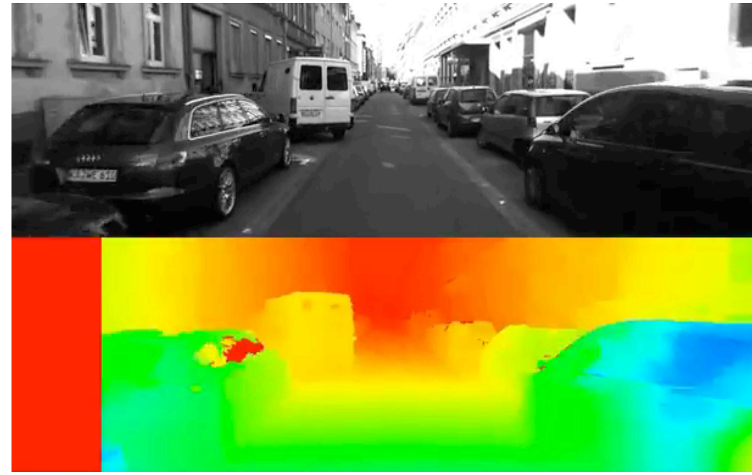
Computer Vision（コンピュータビジョン）

人間の視覚機能を工学的に（再）実装する学問分野，大雑把に「ロボットの目をつくる」ための学術分野である

どんな技術が今までに研究されたきた？

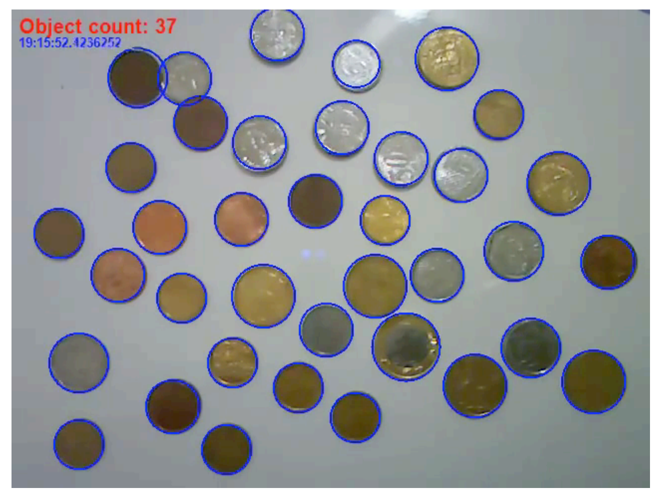
背景差分

ステレオビジョン



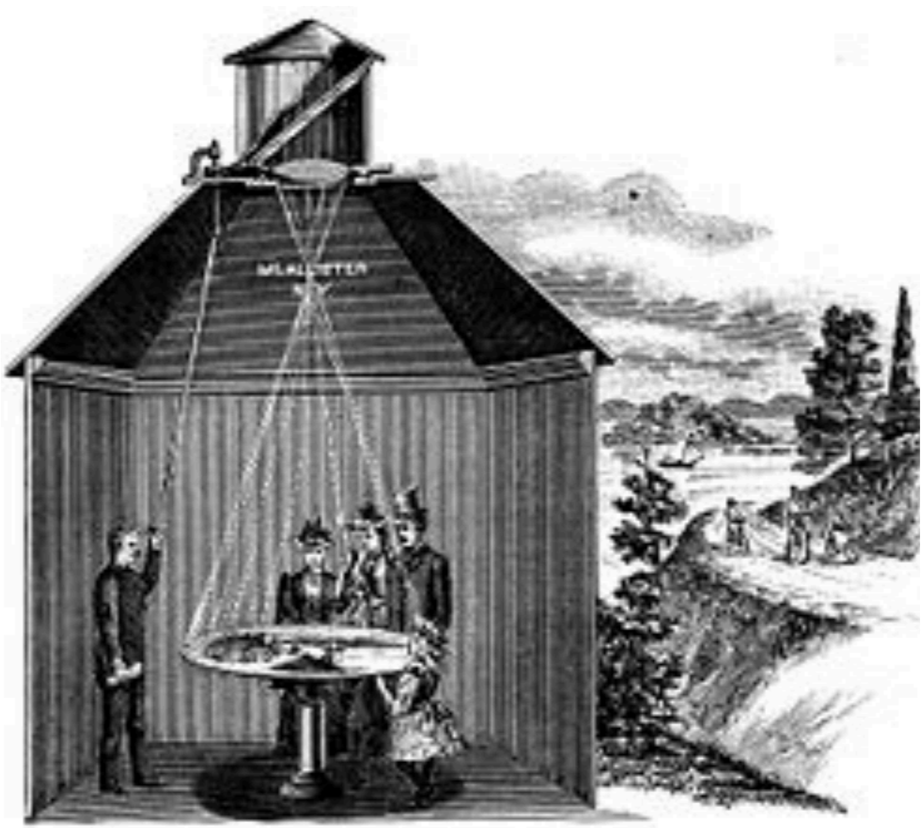
動き検出（オプティカルフロー）

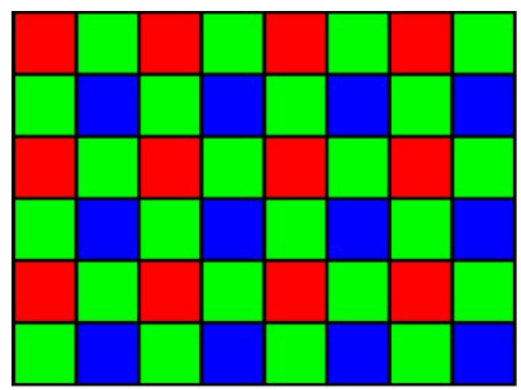
形状認識（ハフ変換）



カメラ/センサ技術

カメラオブスクラ（1600s），デジタルカメラ（1975）



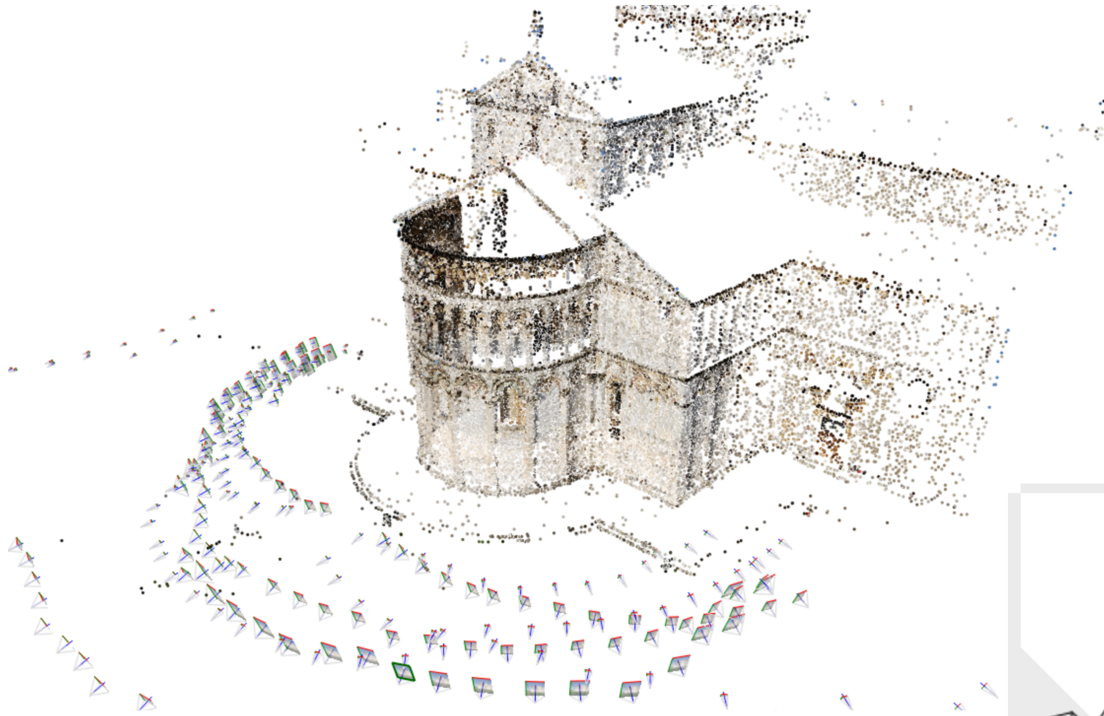


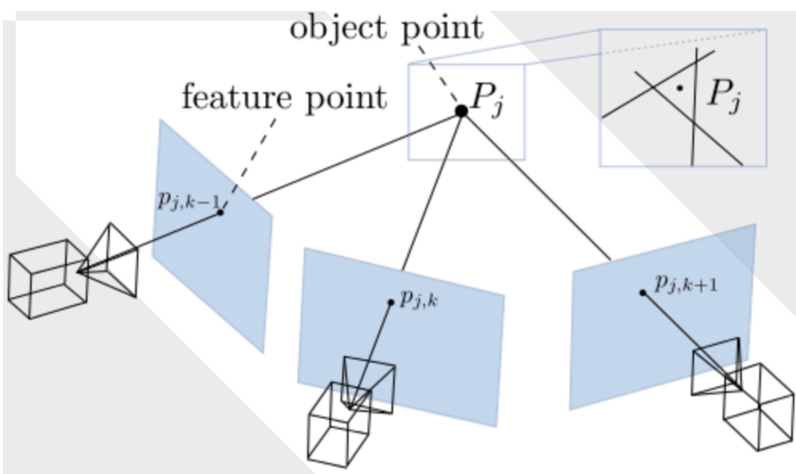
H. P. Gage. “Optic projection, principles, installation, and use of the magic lantern, projection microscope, reflecting lantern, moving picture machine," in 1914.

カメラ幾何

如何に現実世界を捉えるか？

– Keywords: Multiple View Geometry, Epipolar, Stereo Vision

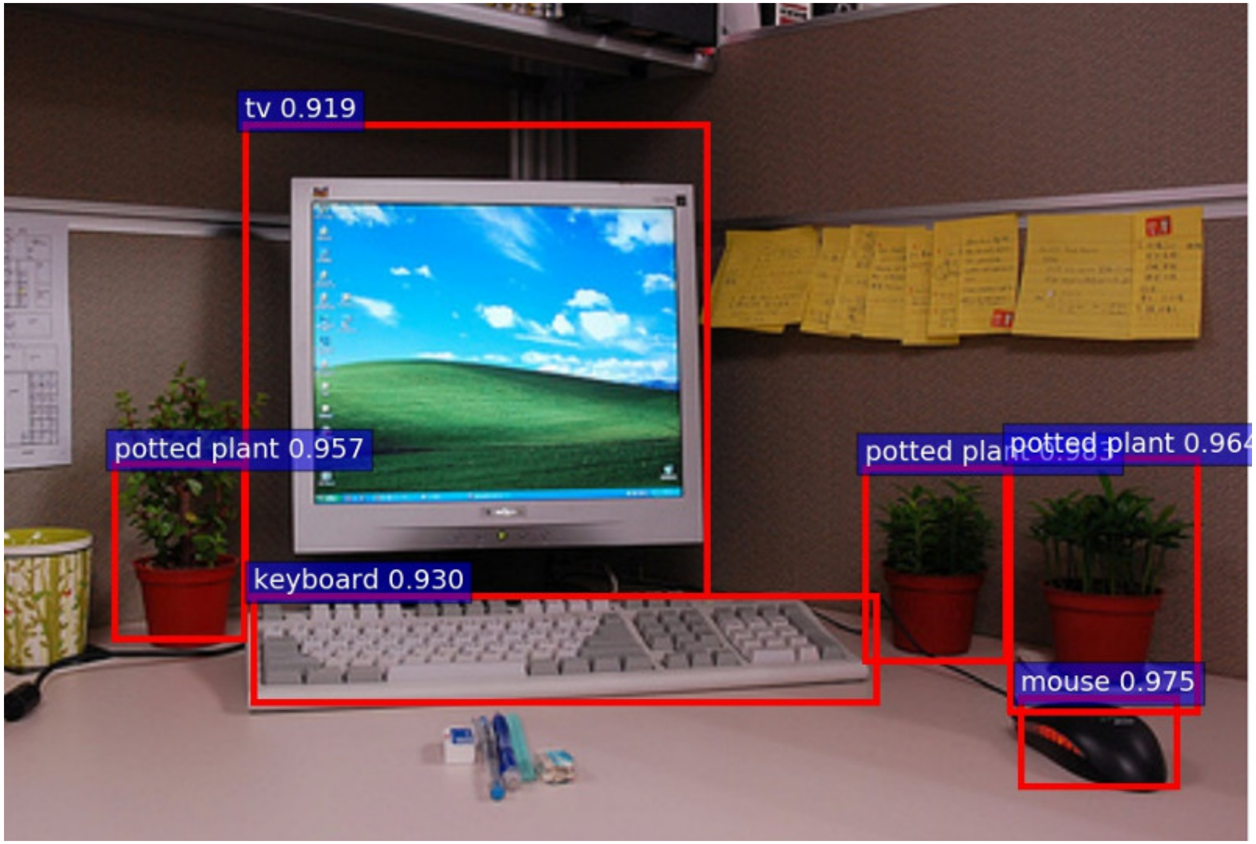




認識

画像内に映り込んだものをカテゴリ分け

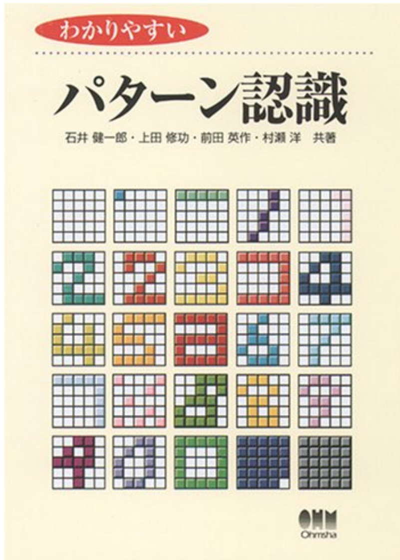
– Keyword: Learning, Supervised Learning, Clustering

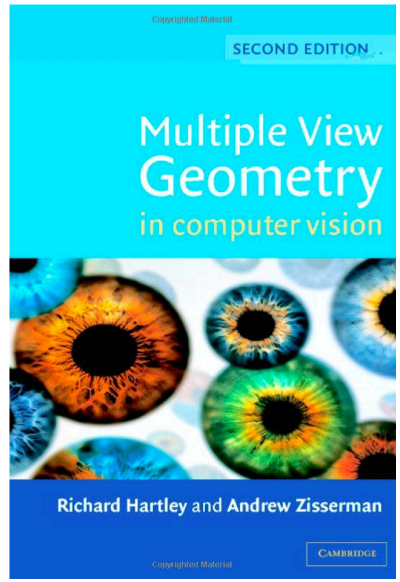


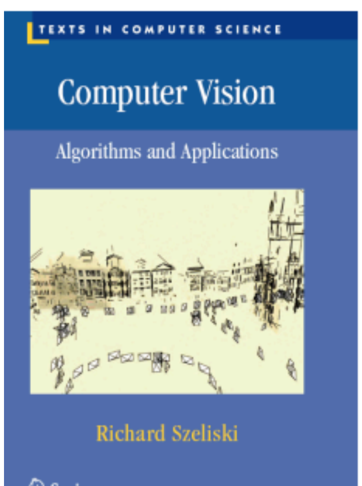
CVの基礎は他文献に譲ります

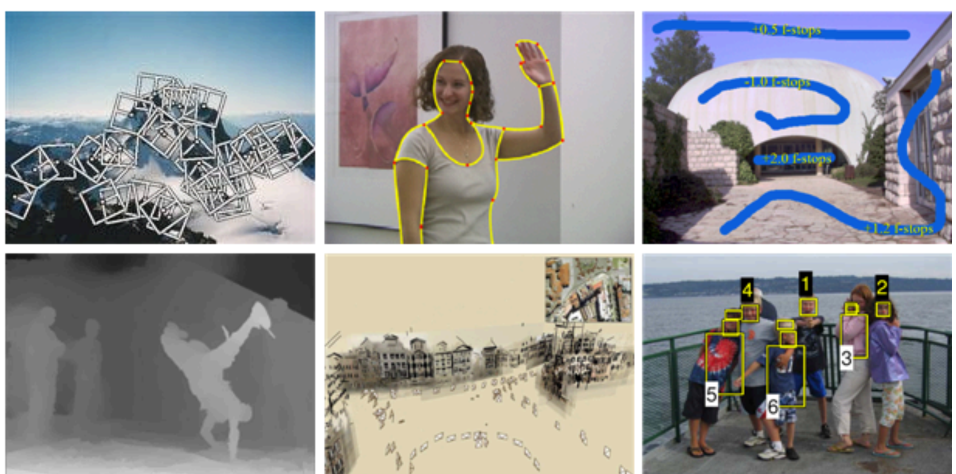
コンピュータビジョンの良書は多い

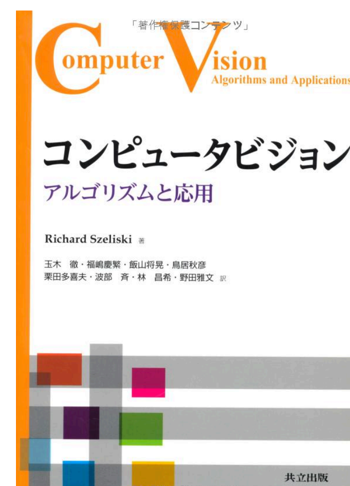
– 詳細な内容は下記など参照











コンピュータビジョンの分野では人の目に憧れ，模倣し実装してきたが，一方で非常に簡単なルールに従う技術が多かった

が，十分発展した機械学習との
結びつきにより覆されようとしている

Deep Neural Networks (DNN)
ー不連続な変化ー

深層学習の隆盛

DNNによりアルゴリズム破壊が生じた

– 本質はデータから最適な特徴を自動で抽出
  - 従来型の特徴抽出の知識が通用しなくなった

従来型の識別

入力

画像/動画

特徴抽出

HOG/SIFT

エンコード

BoW/FVs
部分空間

識別

SVM/RF

出力

ラベル/タグ

深層学習の識別

入力

データから特徴を自動学習
人間はデータ収集/構造設計をすれば良い

識別

出力

例によってニューラルネットやDNNの詳細理論は他文献に譲ります

一般物体認識の解決

実世界の物体を普遍的に捉える問題

– コンペでは1,000クラスを同時に識別

– 従来法（前ページ）を圧倒してDNNが勝利

http://karpathy.github.io/assets/cnntsne.jpeg

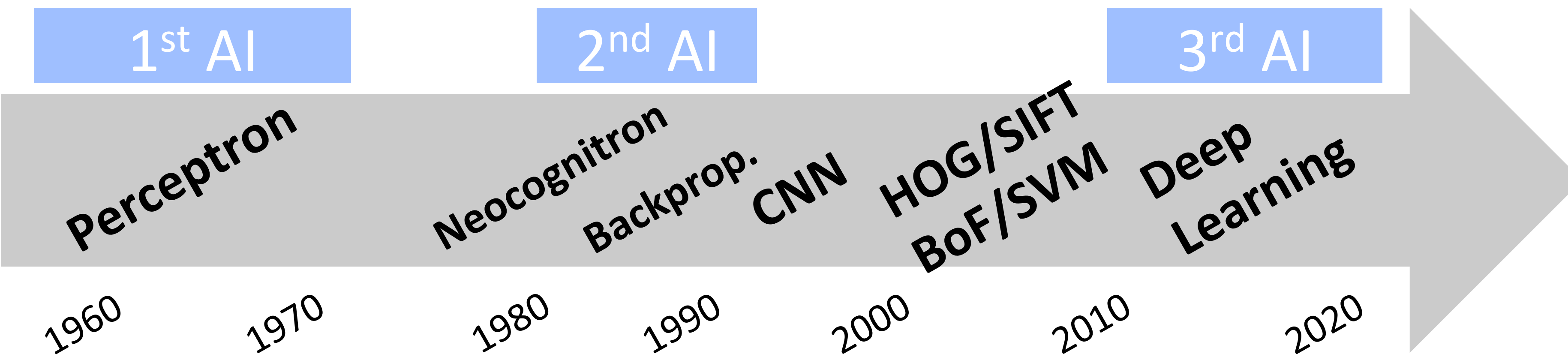
2012年10月，CVのアルゴリズムを置き換える不連続な変化が始まった

かなり大雑把なDNNの概要

DNNの動向（1/8）

DNN時代以前の動向

– Perceptron, MLP, Neocognitron, BackProp, CNN
– DNNが流行る前の画像認識では局所特徴が使用

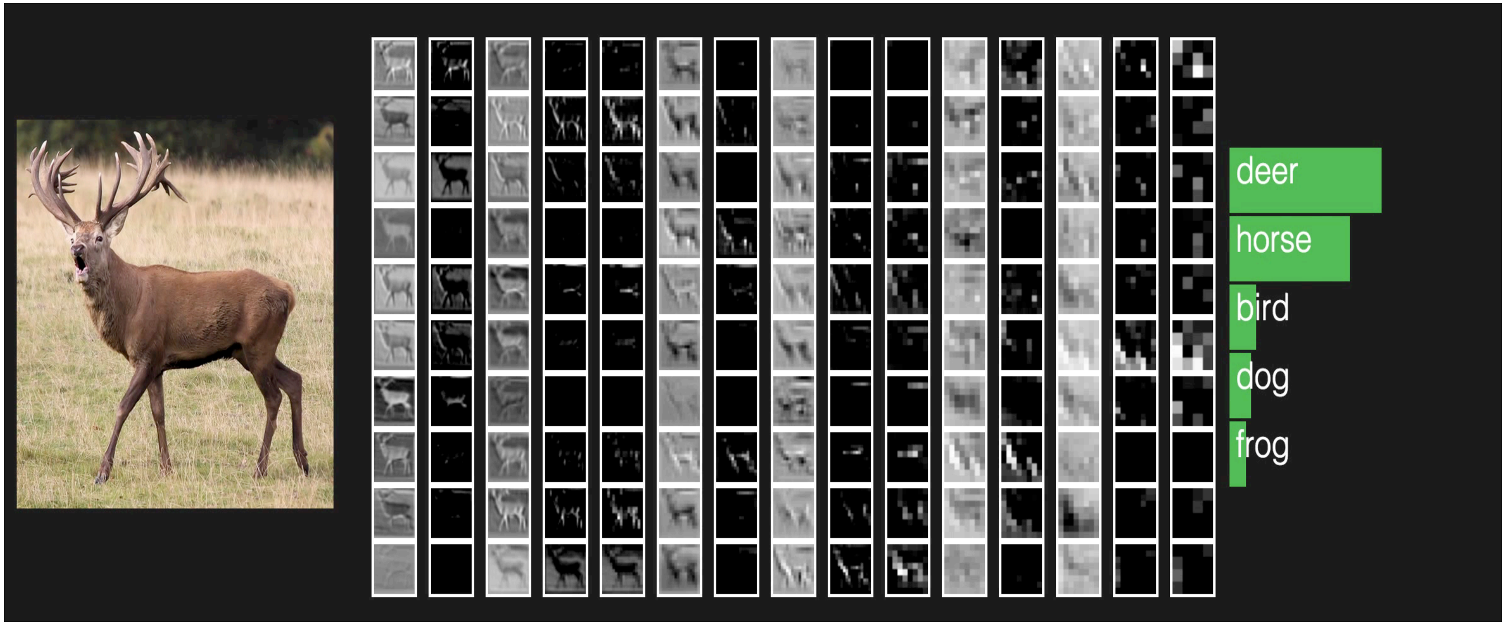


F. Rosenblatt et al. “Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms” in 1961.
Rumelhart et al. “Learning representations by back-propagating errors” in Nature 1986.
K. Fukushima, “Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position”, in 1980
Y. LeCun et al. “Gradient-based learning applied to document recognition” in IEEE 1998.

DNNの動向（2/8）

ILSVRCを発端とする画像識別タスクへの応用

- AlexNet @画像認識コンペILSVRC2012
  - 第一著者Alexさんのネットワーク（仕掛け人のHintonNetになってたかも？）
- 背景にはBelief Propagation, ReLU, SGD, Dropoutなど構造をDEEPにする技術が揃ってきた

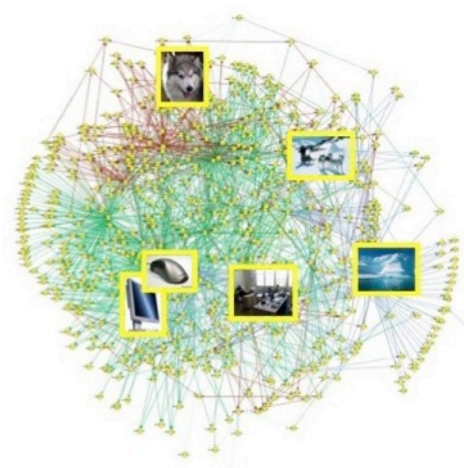


https://www.youtube.com/watch?v=zj_JlVqWK1M

DNNの動向（3/8）

DNNが勝てた背景

– ImageNet!（データが最も重要）
– NVIDIA!（圧倒的な計算力）



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

×

http://cvpr2017.thecvf.com/

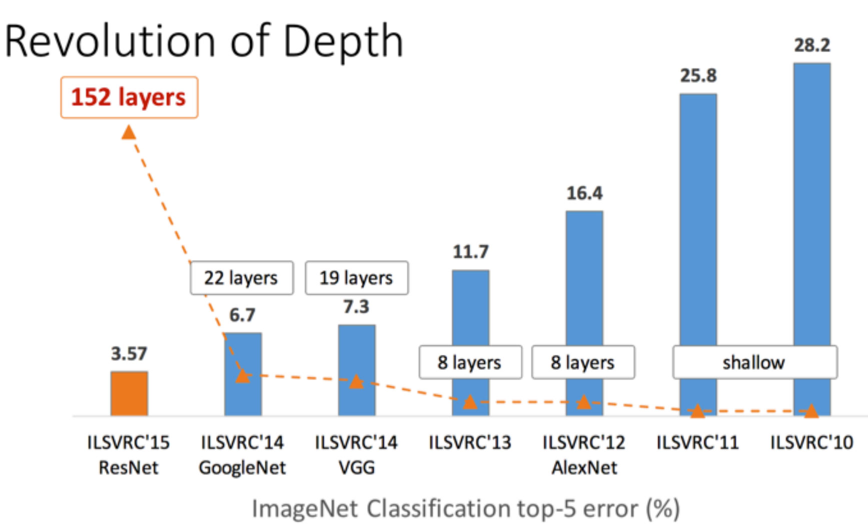
DNNの動向（4/8）

構造の深化

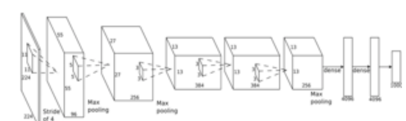
– 2014年頃から「構造をより深くする」ための知見が整う

– 現在（主に画像識別で）主流なのはResidual Network

- 2015年には人間を超える精度を実現

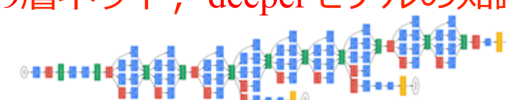


Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. "Deep Residual Learning for Image Recognition". CVPR 2016.

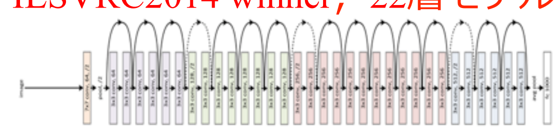


AlexNet [Krizhevsky+, ILSVRC2012]
ILSVRC2012 winner，DLの火付け役

VGGNet [Simonyan+, ILSVRC2014]
16/19層ネット，deeperモデルの知識



GoogLeNet [Szegedy+, ILSVRC2014/CVPR2015]
ILSVRC2014 winner，22層モデル



ResNet [He+, ILSVRC2015/CVPR2016]
ILSVRC2015 winner， 152層！(実験では10^3+層も)

DNNの動向（5/8）

他タスクへの応用（画像認識・動画認識）

– R-CNN: 物体検出

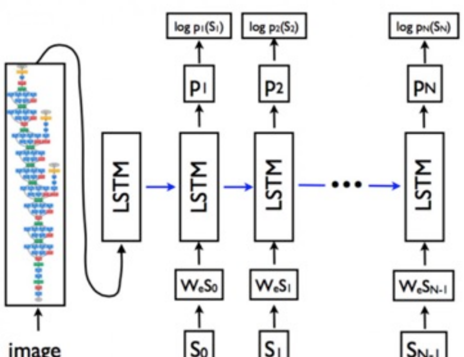
– FCN: セマンティックセグメンテーション

– CNN+LSTM: 画像説明文

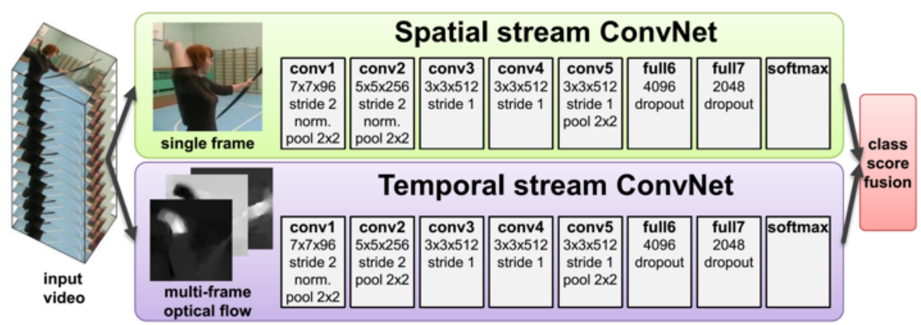
– Two-Stream CNN: 動画認識

ResNet + FRCNN
[He+, CVPR16]
https://www.youtube.com/watch?v=WZmSMkK9VuA

PSPNet [CVPR17]



Show and Tell [Vinyals+, CVPR15]

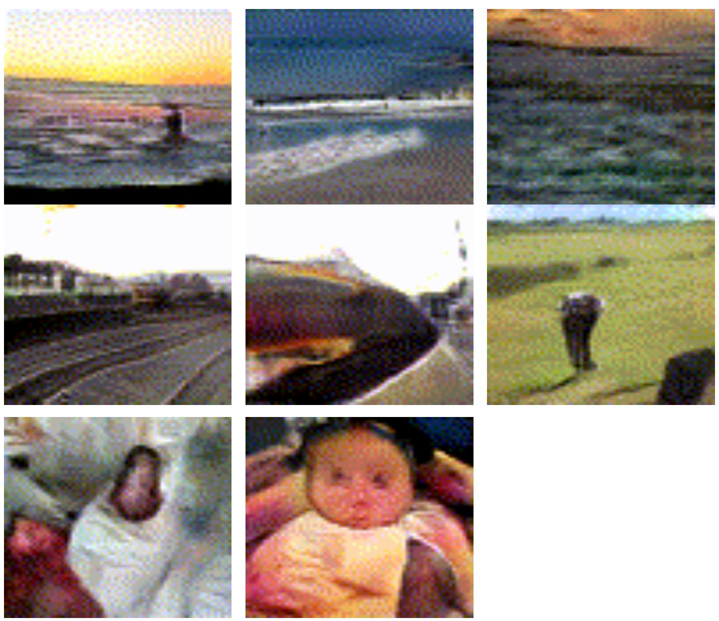


Two-Stream CNN [Simonyan+, NIPS14]

DNNの動向（6/8）

画像生成・強化学習への移行

- GAN: 敵対的学習，Generator（G）と Discriminater（D）が競い合いリアルな画像を生成
- DQN: 強化学習モデル



Video GAN [Vondrick+, NIPS16]
http://carlvondrick.com/tinyvideo/

https://www.youtube.com/watch?v=V1eYniJ0Rnk

DNNの動向（7/8）

DNNのフレームワークが次々にリリース

- Caffe/Caffe2, Theano, Chainer, TensorFlow, Keras, Torch/PyTorch, MatConvNet, Deeplearning4j, CNTK, MxNet, Lasagne（順不同，その他多数）
- 特に，Caffeが出てきてからCVにおけるDNNの研究は爆発的に広がった

https://chainer.org/images/logo.png

https://www.tensorflow.org/_static/images/tensorflow/logo.png

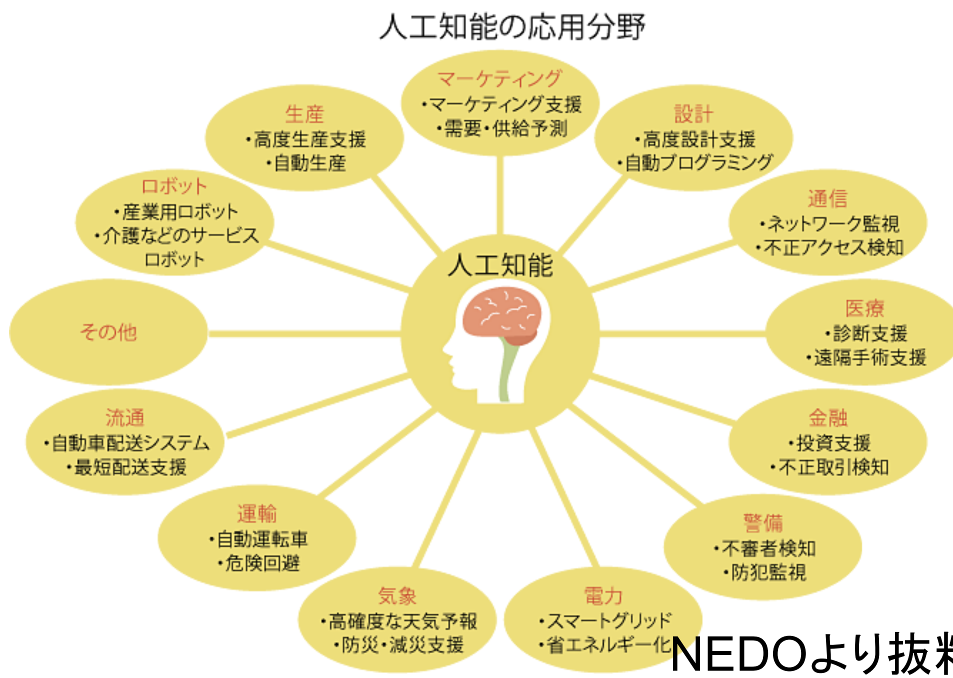
http://pytorch.org/docs/master/_static/pytorch-logo-dark.svg

\# 日本ではChainer？ EuroではMatConvNet？ 世界的にはTensorFlow？ コード量の少なさからKerasもよく聞く
\# 当グループではPyTorchでほぼ統一して共有（Facebookでも研究はPyTorch，プロダクトはCaffe2）

DNNの動向（8/8）

現在も進化の一途を辿り，社会実装が進む

– 自動運転/ADAS
– ロボティクス
– ファッション
– 画像/動画検索
– 物流（ピッキング等）
– 等



NEDOより抜粋
http://www.nedo.go.jp/tokushu/imgs/fn56/img_otuki01_l.png

研究者としては「こんなこともできる」を世に出したい

DNN全盛期，激動の時代

- 面白いことが出てきている！(&これからもできそう)
  - ぜひ当事者になりましょう！
- ただ，情報過多ではある
  - ぼーっとしていると，ついていくことすらできない！

こんな時代に備えるべき能力とは？

DNN隆盛時代を生き抜くための研究/開発力

大別すると,

– サーベイ力：現状を正しく把握する能力
– 発想力：新しいコンセプトを生み出す
– 実装力：アイディアから付加価値を生む

この２年半の取り組みをベースに

cvpaper.challengeに基づいて紹介

論文読破・まとめ・発想・議論・実装・論文執筆・（社会実装）に至るまで取り組むCVの今を映す挑戦である

- 人員：産総研，筑波大，東京電機大，慶應大による20名弱
- 知識共有と洗練された論文化（・社会実装）
- 2017はトップ会議※に１０+本以上の投稿を目標に活動中

※ HP, Twitter, SlideShareもご覧ください
HP: https://sites.google.com/site/cvpaperchallenge/
Twitter: @CVpaperChalleng
SlideShare: @cvpaperchallenge

※ Google Scholar Top-20にリストアップされている国際会議や論文誌

cvpaper.challengeの２年半を簡単に

- 2015年，CVPRにある600論文（英語で5,000ページ超）を読破してまとめ公開
- 2.5年でまとめた論文数2,000超，学会論文10本超，招待講演8件
- 論文速報arXiv:cv 週間お気に入り数世界一位
- スライド年間アクセス数 約90,000（Google/Facebook/Amazon/Apple等世界からアクセスあり）
- メジャーな国内外学会・シンポジウムで学生が毎年受賞（ViEW/MIRU/ECCVWS）
- 2017年はGスカラーTop20論文に17本投稿中（7本採択，7本審査中）

サーベイカ

サーベイとは？

ひとことでいうと，その分野の動向を把握すること

- 現在どんな技術が流行っている？
- どういう歴史を辿ってきた？
- 自分のやっていることに最も近いものは？

なぜ，サーベイが必要？

トレンドの把握

- 知識がないと既存研究の劣化版を作りかねない
- トレンドを知らないと(天才でない限り)最先端の研究を生み出すことは難しい

自身の研究の立ち位置を確認

- 何が違う？なぜやる？どこが良いのか？という哲学

究極的には次のトレンドを作るため

- 分野の方向性を自ら定める
- より良く，正しい方向へ導く

なぜ，そこまでサーベイするのか？

網羅的に調べることで分野の状況を精度よく映す

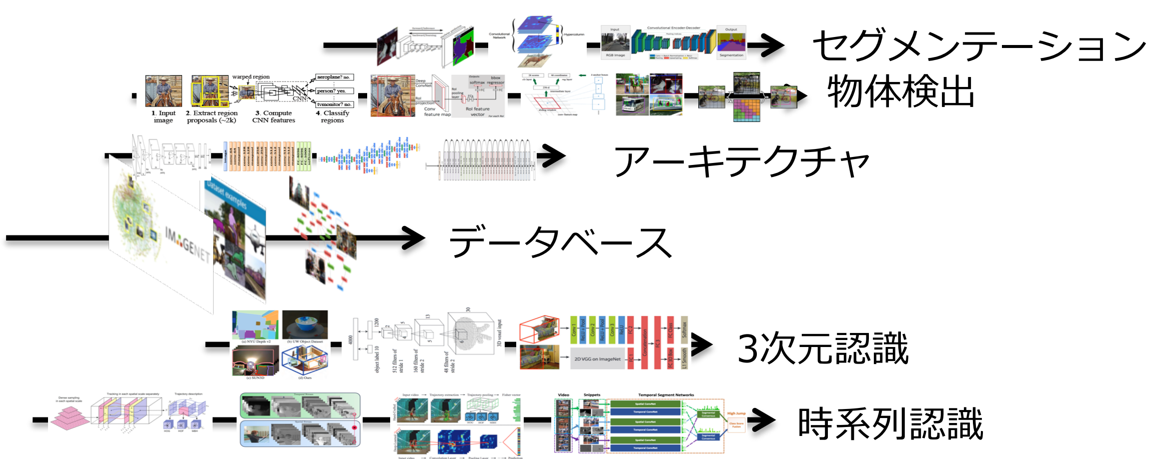
– サーベイの質や量が研究テーマ考案・選定に直結

– 年間1,000本くらいは論文を読むポテンシャルが備わった

技術の流れがわかる

– 点より線/面で技術を捉える（歴史を追う，他分野から学ぶ）

– 次に何をすれば良いかがなんとなくわかる



どう読む？

目的に合わせた読み方を実践

- 論文は体系付けられているから読み方を身につける方が効率が良い
  - どんな研究しているか？：概要（+序章）だけ
  - どう実装する？：手法と実験
  - 満遍なく分野を抑えたい：サーベイ論文
- 情報の発生源から読んだ方が無駄がない
  - トップ会議やジャーナルから探していく
  - 朝刊よりもネットニュース！（情報を判断できる目を養う）
- cvpaper.challengeはトップ会議の情報をいち早く日本語で読めるように工夫

ここは長くなるので後で聞くか，熟練の先輩にご質問ください

情報を共有しよう！

1人で頑張るより，ツールを使い協力して進めるのが現代流

– クラウド（Google/Dropbox），チャット（Slack/Skype/Line）

– モチベーション，集合知，作業分割と統合，ディスカッション等グループで活動するメリットは多い

[redacted] :45 PM
遅ればせながらReadingListのURLをCVFのものに置換しました。
今年のCVFのページにはarXivのリンクもついていますね．年次の投稿数遷移が取れるかと思いましたが，ICCV2015にはリンクなかったです．残念．
👍 1

[redacted] ☆
ワークショップの動向:
若いワークショップが多い?印象です
> ざっくり数えたところ今回が4回目以下のWSが16件/44件
面白そうなワークショップ
> COMPUTER VISION PROBLEMS IN PLANT PHENOTYPING (CVPPP)
> https://www.plant-phenotyping.org/CVPPP2017
> 日本は農業に力を入れてる印象であり，国内学会(View，SSIIなど)だと農業分野の外観検査関連の研究を見るので，このWSは今後ねらい目ではないでしょうか
(edited)

[redacted] :04 PM
葉っぱセグメンテーション＆カウントのチャレンジですね。
かなり成熟した感じのある画像認識技術を，アプリケーション寄りにすそ野を広げていきたいという意図が働いているのでしょうか？
非CV屋さんを取り込みたい？

[redacted]
GANチュートリアル（スライド含む） link: https://sites.google.com/view/iccv-2017-gans/schedule

hirokatsu.kataoka 2:13 PM
1. Mask-RCNN

は前述の通りResNet, Faster RCNNで有名な Kaimingさんの著書です。物体 検出とセマンティックセグメンテーションを同時に解いた方が良い、という知見に基づいています。これ以降、Kaimingさんの成果の速度がゆるくなっているので、実はこの論文が（検出やセグメンテーションの）完成版で、彼はすでに違うことにフォーカスして 研究しているのでは？と見ています。

2. Kd-network

ICCV 2017 速報の資料作成

	Erhan, Vincent Vanhoucke, Andrew Rabinovich		Anton van den Hengel, Chris Russell, Anthony Dick, John Bastian, Daniel Pooley, Lachlan Fleming, Lourdes Agapito
1505	Understanding Image Representations by Measuring Their Equivariance and Equivalence Karel Lenc, Andrea Vedaldi	1505	SUN RGB-D: A RGB-D Scene Understanding Benchmark Suite
			Shuran Song, Samuel P. Lichtenberg, Jianxiong Xiao
1505	Deep Neural Networks Are Easily Fooled: High Confidence Predictions for Unrecognizable Images Anh Nguyen, Jason Yosinski, Jeff Clune	1505	Small-Variance Nonparametric Clustering on the Hypersphere Julian Straub, Trevor Campbell, Jonathan P. How, John W. Fisher III

Monday June 8, 10:10am-12:30pm

	Poster Session
1505	Going Deeper With Convolutions Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich
1506	Propagated Image Filtering Jen-Hao Rick Chang, Yu-Chiang Frank Wang
1506	Web Scale Photo Hash Clustering on A Single Machine Yunchao Gong, Marcin Pawlowski, Fei Yang, Louis Brandy, Lubomir Bourdev, Rob Fergus
1506	Expanding Object Detector's Horizon: Incremental Learning Framework for Object Detection in Videos Alina Kuznetsova, Sung Ju Hwang, Bodo Rosenhahn, Leonid Sigal
1506	Supervised Discrete Hashing Fumin Shen, Chunhua Shen, Wei Liu, Heng Tao Shen
1505	What do 15,000 Object Categories Tell Us About Classifying and Localizing Actions? Mihir Jain, Jan C. van Gemert, Cees G. M. Snoek
1508	Landmarks-Based Kernelized Subspace Alignment for Unsupervised Domain Adaptation Rahaf Aljundi, Rémi Emonet, Damien Muselet, Marc Sebban

CVPR2015 完全読破

資料をまとめてみよう

テンプレートに従って１論文１ページ

- （自分だけでなく他の人が）素早く研究の肝をつかむ
- 【概要】【新規性・差分】（【手法】【結果】）【Links】

【1】 Jianxiong Xiao, James Hays, Krista A. Ehinger, Aude Oliva, Antonio Torralba, “SUN Database: Large-scale Scene Recognition from Abbey to Zoo”, in CVPR2010.

Keywords: Dataset, Scene Categorization, Benchmark, Recognition

概要

コンピュータビジョンにおいてシーン認識のデータベースである Scene UNderstanding (SUN) databaseを提案. シーン認識の裾野を広げた.

新規性・差分

それまでの物体認識のデータセットでは数百クラスの識別クラスが用意されていたが, シーン認識では15種類程度しか含まれていなかった. SUN databaseでは, それまでのデータセットをさらに拡大させ, 397クラスのシーンを含む, 大規模なデータセットである.

データセットの概要

シーン認識に関する397クラス, 130,519枚の画像が含まれる.
画像例は次ページ. 比較した特徴量は, HOG, denseSIFT, self-similarity (ssim), LBP, GIST, textonなど.

結果

次ページの図の通り. 全ての特徴量を統合するのが最も精度が高いことが判明した(38.0%). 次いでHOG2x2 (27.2%), geometry texton hist (23.5%), ssim (22.5%), dense SIFT (21.5%)であった.

Links

論文ページ:
http://cs.brown.edu/~hays/papers/sun.pdf

プロジェクトページ:
http://vision.princeton.edu/projects/2010/SUN/

HOG https://hal.archives-ouvertes.fr/inria-00548512/document
GIST http://cvcl.mit.edu/scene_understanding.html
SSIM
http://www.researchgate.net/profile/Eli_Shechtman/publication/221362526_Matching_Local_Self-Similarities_across_Images_and_Videos/links/02e7e520897af25746000000.pdf
DenseSIFT http://www.vision.caltech.edu/Image_Datasets/Caltech101/cvpr06b_lana.pdf
LBP http://www.outex.oulu.fi/publications/pami_02_opm.pdf
Sparse SIFT http://www.robots.ox.ac.uk/~vgg/publications/papers/sivic04b.pdf
Texton http://www.ics.uci.edu/~fowlkes/papers/mftm-iccv01.pdf

まとめ資料はどのように使えば良い？

基本的には自由

- とにかく読む
- 気になった論文があれば原本を読んでみる
- ブレインストーミングに使う
- 考案したアイディアへの関連付けに使う
- 引用論文を探す
- など

論文は「こんな凄い/面白いができるようになった」をまとめているので，ツライものではない

発想力

発想とは？

ここでは研究テーマに結びつくまでの過程

- 個人でアイディアを考案
- 他人と議論（ブレインストーミング; ブレスト）
- 追加サーベイ・アイディアの洗練
- 以上を繰り返す

キーワードは「集合知」

個人作業とブレスト

– ひとりよりも複数人でアイディアを考えた方が
  - 劇的に効率良い
  - より強いアイディアが生まれる
  - 広い分野をカバーしたアイディアができる

– ブレストを最大限に活かすための個人作業
  - 個人はユニークなので人が多いほど発想のポテンシャルは向上
  - 調査を事前に行い各人が知識を増やしておく

サーベイとブレストの繰り返し

まずは個人作業

- サーベイ（動画，スライド，まとめから探すのがポイント）
- 良い論文/技術はまとめてみる（自分のものにする）
- アイディアが出たら常にメモ（忘却防止）

ブレスト

- アイディアを増強（事前の個人作業が重要）
- ポジティブに，躊躇わず，頭フル回転
- 集合知を発揮（集合天才は一人の天才を超える）

ブレストのポイント

ネタを準備しておく

– 直前に輪講/研究紹介/サーベイ発表等を行う
– まとめ資料を眼前にばらまく
– 元手となるアイディアを蓄積

グループを常に組み替える

– 3～4人のグループを単位時間ごとにシャッフル

アイディアの洗練

追加サーベイと繰り返し

– アイディアに対して肉付けするのが追加サーベイ
– アイディアと追加サーベイを見て洗練化
– 繰り返しにより研究テーマレベルに昇華

例：Fashion Culture Database (FCDB)

– スタバの自撮り画像を収集すれば意識高い解析？
– City Perception（ECCV14）：世界各地の都市解析

・・・

– SNS画像収集して世界の動的ファッション解析！

立場は関係ない

- バリエーションが豊かな方が良い発想は出る
  - 熟練者と初心者の脳を同時に働かせる
  - 学生も研究テーマ考案に参加
  - 趣味趣向国籍所属性別年齢等無関係

発想もチームプレイ
「個人作業」と「チーム連携」

実装力

実装とは？

ここでは考案したアイディア/研究テーマを実行し価値を生み出す（e.g. 論文執筆，システム開発）こと

- 研究テーマを細分化，仕事を割り当て
- ミーティング
- 完成形に落とし込む

手を動かす

何でもいいから，まずは自分で手を動かす

- 論文の追実装
- プログラミングで遊ぶ
- 自分流のアイディアを形にする

分割と統合

- 割り振られた個人作業（分割）と，
- ミーティング/チャット（統合）により進行
  - 学生同士による相補チェック
  - 常にタスクを与えることで適度なプレッシャー

重要な戦略（１）

問題設定と同時にデータベース（DB）を作る

- 自ら問題とDBを作ることで世界で唯一の問題を扱うことになる
- スピード勝負になりづらい，新規性をグループ全員が享受

時代により戦略を変える

重要な戦略（２）

複数人で複数本の関連テーマを走らせる

- 手法，問題など複数の成果にできるように分割
- ここにおいても分割と統合の考え方が役立つ

実装

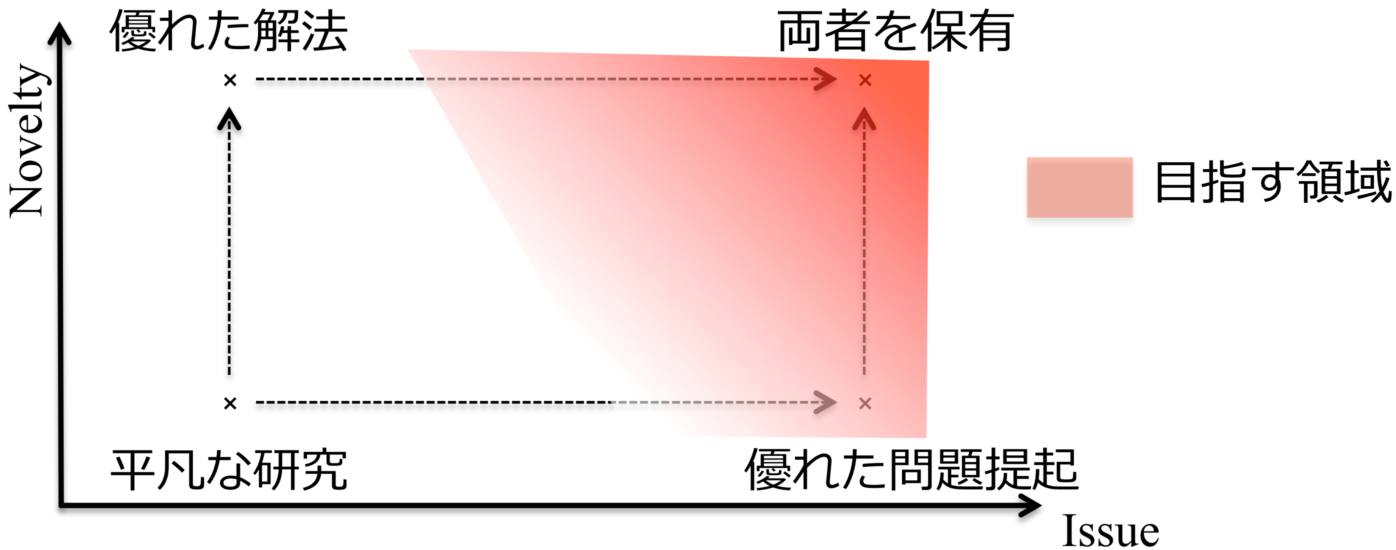
個人作業（分割）と
ミーティング/相談（統合）
の繰り返し，実装においても
チーム力を重視

cvpaper.challengeの研究

目指すは視点を変える新しい解法を与えること

- 【問題設定: Issue】と【解の質: Novelty】
- 既存手法の改善ではない

２年間，サーベイ法・発想法・実装法に従い研究



目指すべき姿勢

次世代のCV研究を創ってみせる

– 問題がさき，解法はあとの{Concept-, Vision-}Driven
– どうやってモデルを改善するか？ではなく今の技術+αで何ができるか？を考える

当グループでも問題設定とともに「データを自ら作る」を徹底
画像・動画の収集，アノテーションやクロスチェックもできる限り迅速かつ大規模，網羅性を考慮

cvpaper.challengeの目指すもの

技術で面白いことをやる・見せる

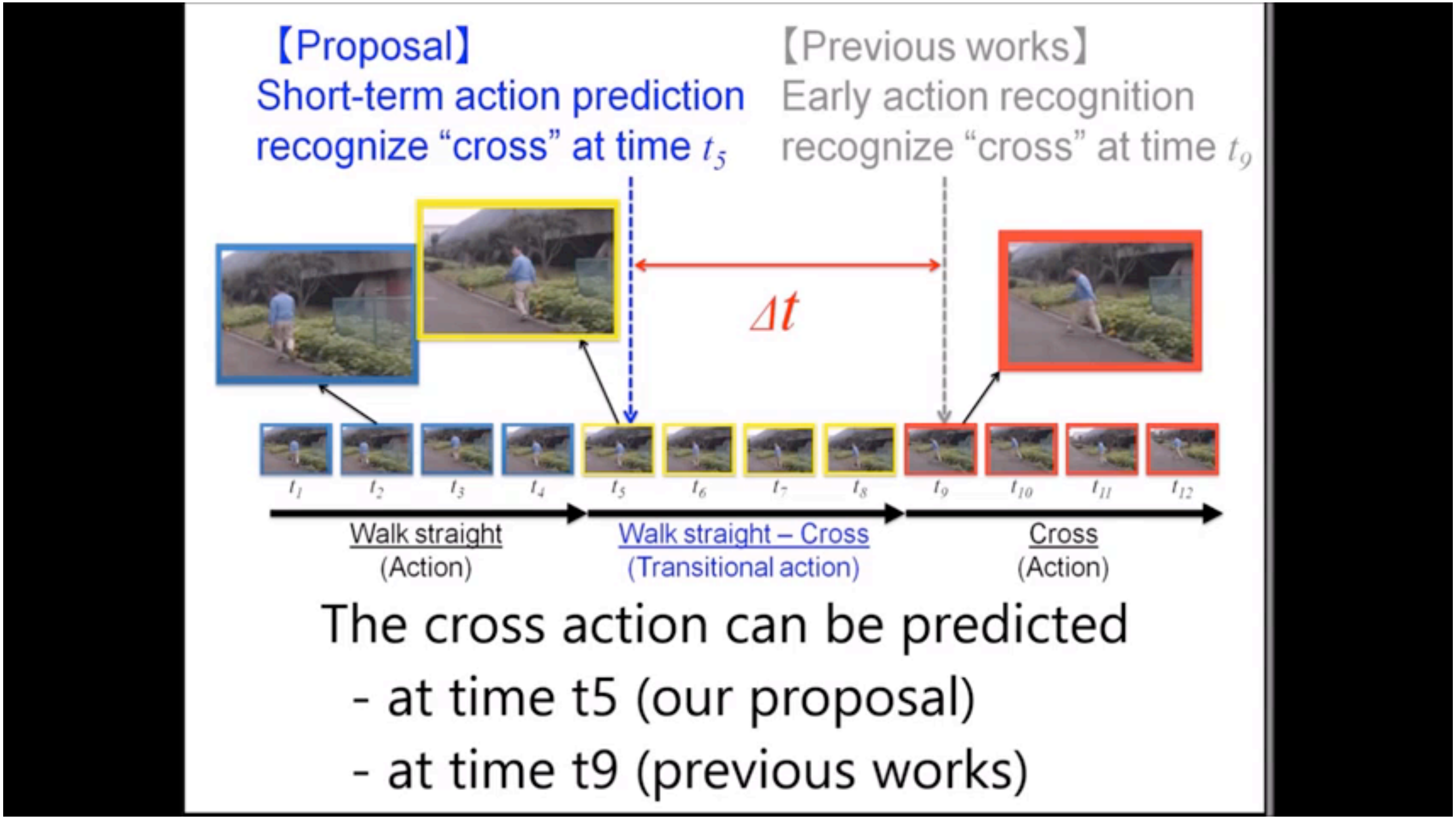
– こんな面白いことができる！と未来を見せる論文
– 今の技術を「使い，組み合わせ，洗練させ」ギリギリ実現可能な将来ビジョンを創り出す
– 新しい技術はどんどん使うが，その代わりそれでどんな未来ができるのかを想像して研究する

以下，既に公開済みの研究について紹介

Transitional Action Recognition

BMVC2016

遷移状態を把握して行動予測



H. Kataoka et al. "Recognition of Transitional Action for Short-Term Action Prediction using Discriminative Temporal CNN Feature" in BMVC, 2016.

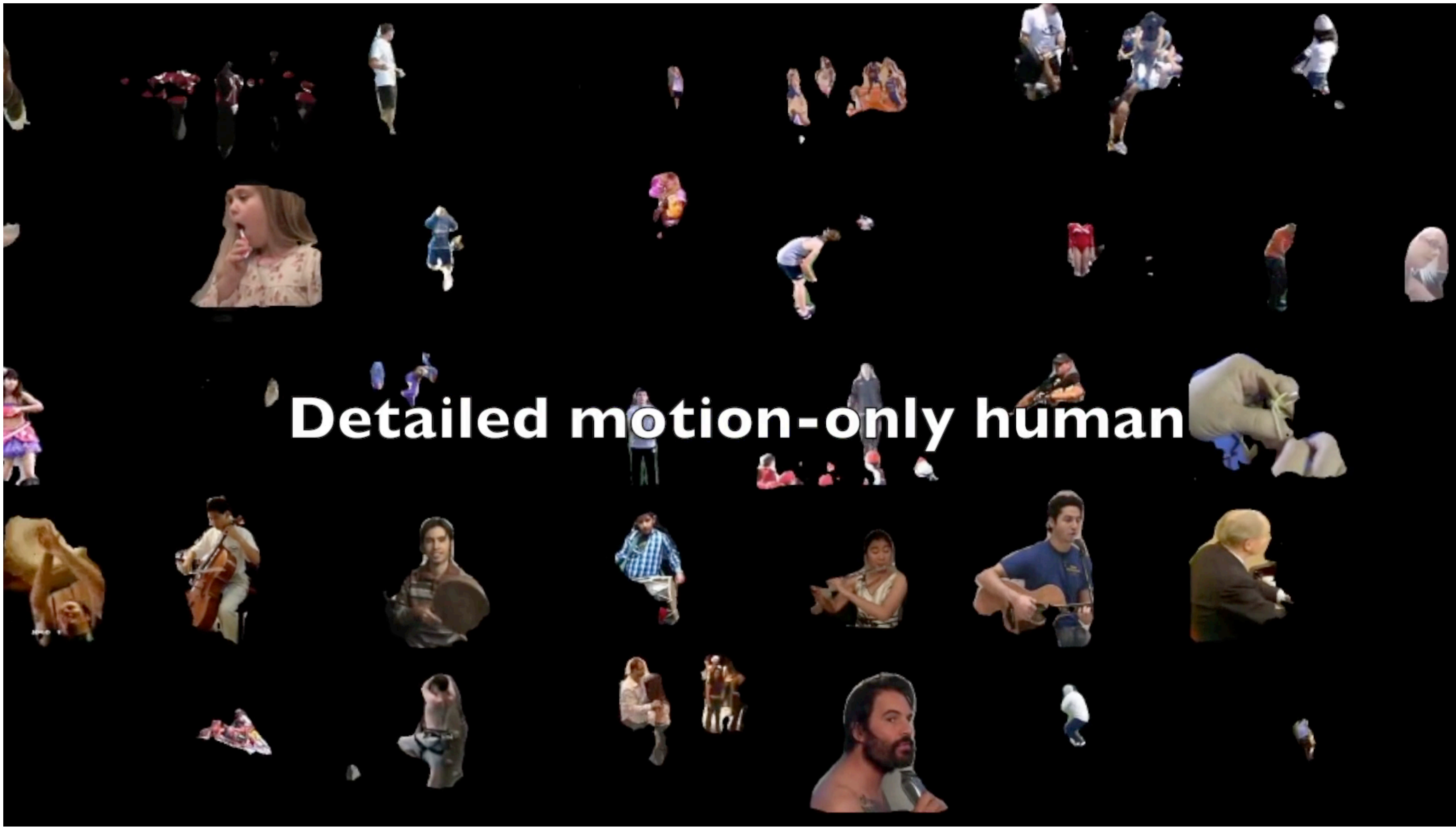
https://www.youtube.com/watch?v=m1AmdIYGaBY&vl=en

(Acceprance rate: 39%)

Human Action Recognition without Human

ECCVW2016

背景領域が人物行動認識に与える影響を評価



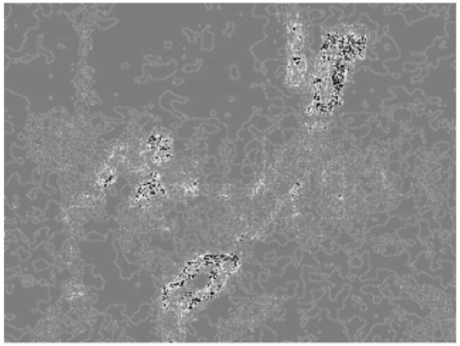
H. Kataoka et al. “Human Action Recognition without Human” ECCV 2016 Workshop Oral (Best Paper) & MIRU 2017 Oral.

Accelerated Images

加速度を2D画像に投影して特徴抽出

– 位置 (RGB), 速度 (Flow)に対して加速度を追加
– 特徴抽出はDNNに任せる





Physics quantity	I	I' (1st-order diff.)	I'' (2nd-order diff.)
Input	RGB	Flow image	Acceleration image
Stream	Spatial-stream	Temporal-stream	Acceleration-stream

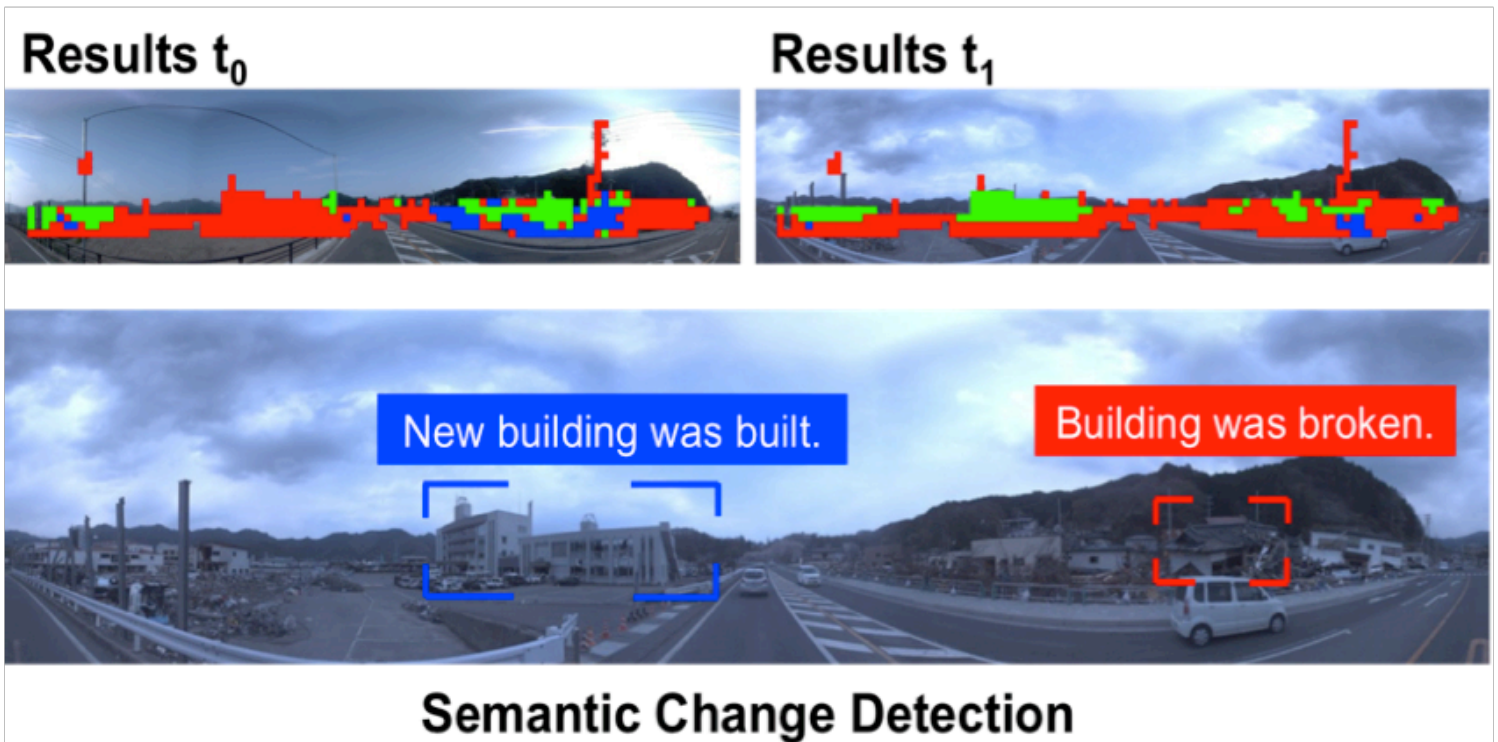
H. Kataoka et al. "Motion Representation with Acceleration Images" in ECCV 2016 Workshop, 2016.

Semantic Change Detection

意味的な変化を捉える

– 直感的には「変化検出」+「セマンティックセグメンテーション」

– 「何がどのように」変化したかを理解



H. Kataoka et al. “Semantic Change Detection with Hypermaps” in arXiv 1604:7513, 2016.

Movie Poster Database (MPDB)

映画ポスターのみからアカデミー賞を予測

Let's try!

- Can you correct the Academy Award 2017?
- Which movie poster do you like, and why?

Nominated movies in 89th Academy Awards

Y. Matsuzaki et al. "Could you guess an interesting movie from the posters?" in MVA, 2017.

Pedestrian Near-Miss Detection

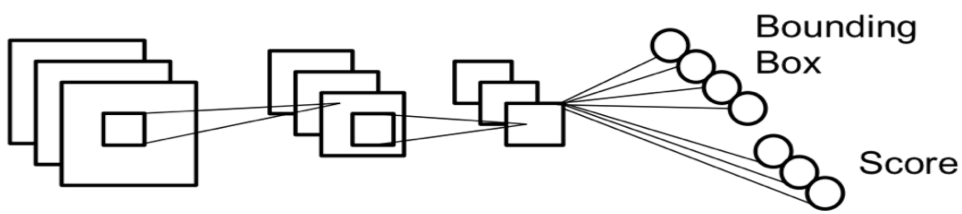
MVA2017

歩行者のニアミス状態を検出

– 通常状態とニアミス状態の歩行者の動作や姿勢は異なる

– ニアミス位置やその危険度も含めて検出可能

Input (224 × 224)



CNNs

Fully Connected

Output Example

Risk high

Risk low

T. Suzuki et al. “Pedestrian Near-Miss Analysis on Vehicle-Mounted Driving Recorders” in MVA, 2017.

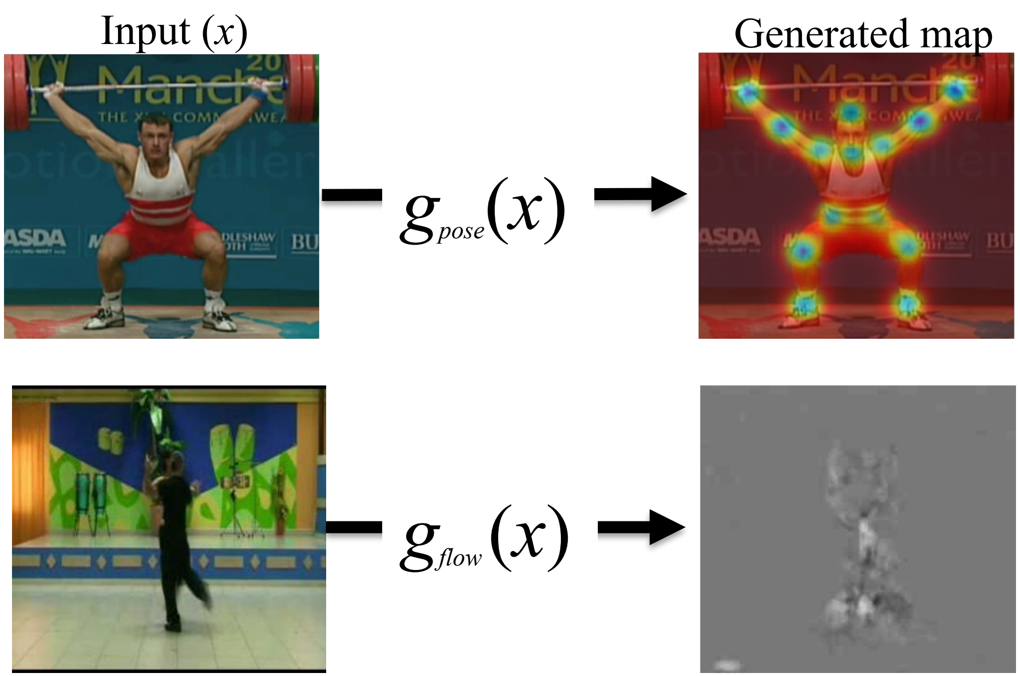
Generated Motion Maps

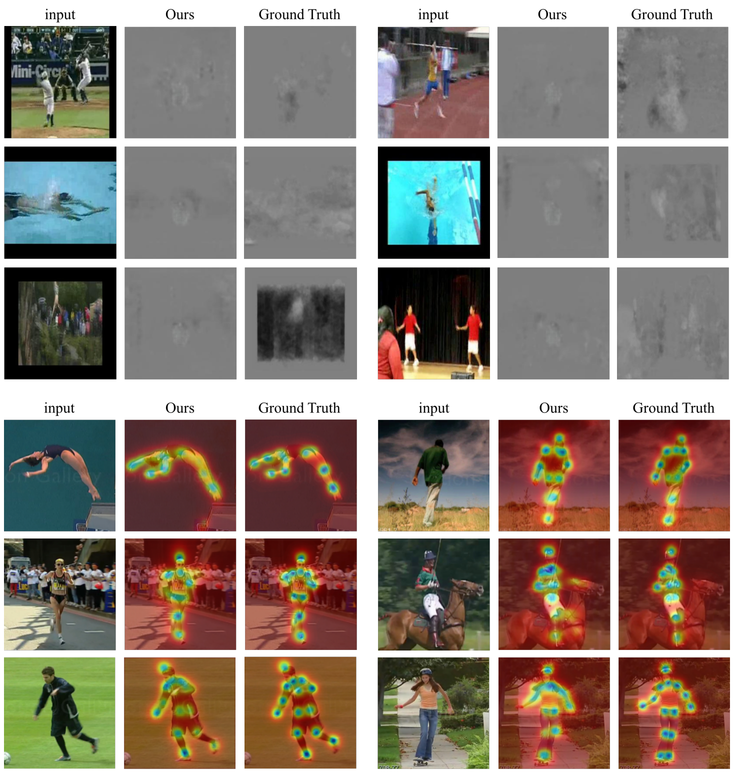
CVPRW2017

入力から直接 動作解析に必要なマップに変換

– フレームワークにはpix2pixを使用

– Poseのヒートマップ，Stacked Flow画像を生成



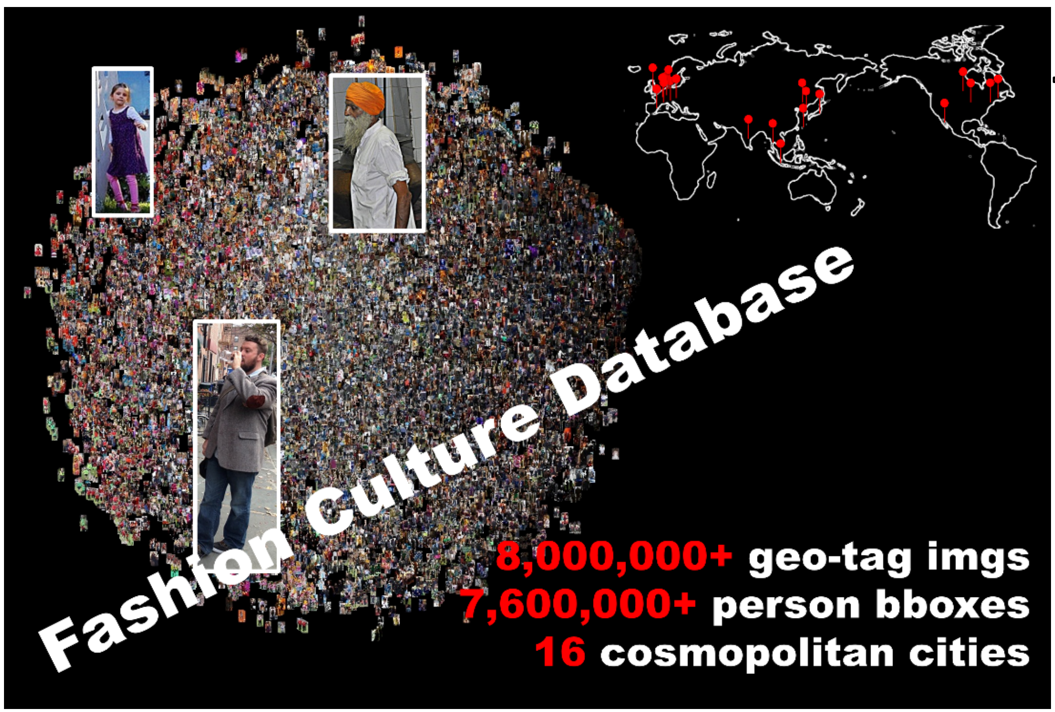


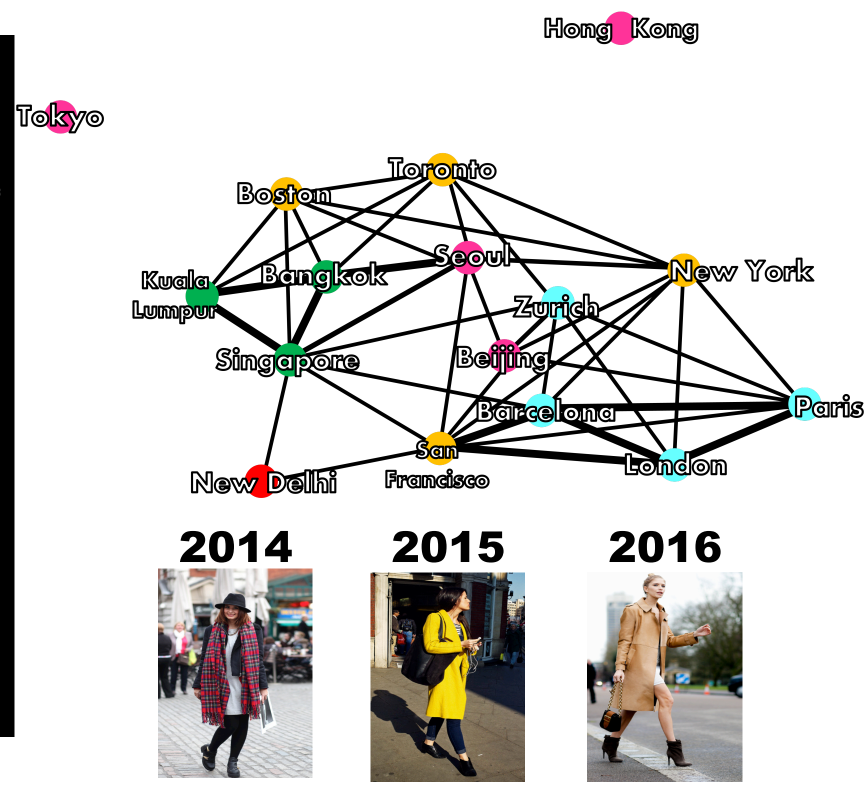
Y. Matsuzaki et al. “Generated Motion Maps” in CVPRW, 2017.

Dynamic Fashion Cultures

大規模DB（76M）からファッショントレンドを発見

– 動的に変化するファッション（服装）を捉える
– 世界主要都市の動向をSNS画像から推定





K. Abe et al. "Dynamic Fashion Cultures", arXiv, 2017 & MIRU 2017 Oral. (Best Student Paper)

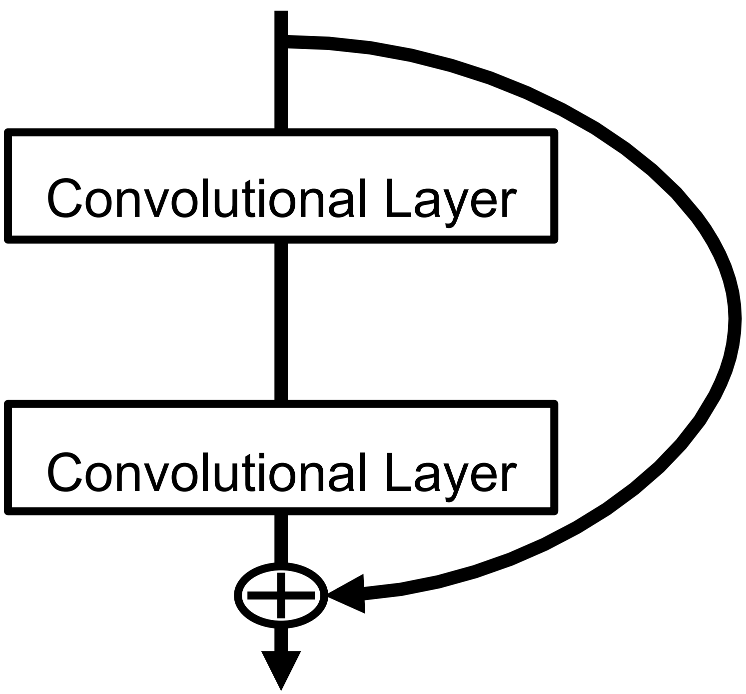
3D ResNets

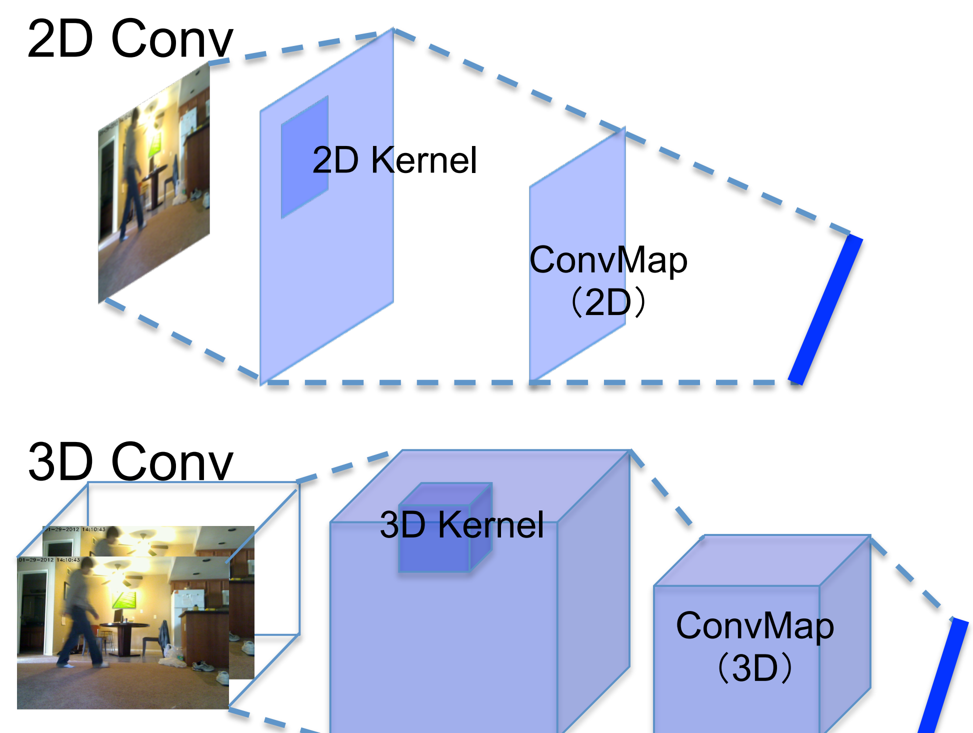
ICCVW2017

ResNetsのカーネルに3D Convを使用

- 動画認識用（xyt）のResNetsを構築
- [公開済み] Kinetics Dataset Pre-trained Model

https://github.com/kenshohara/3D-ResNets





K. Hara et al. "Learning Spatio-Temporal Features with 3D Residual Networks for Action Recognition", ICCVW, 2017.
（コード&モデル公開済）