@teamLab

片岡 裕雄

http://hirokatsukataoka.net/

cvpaper.challenge

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

- 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

チャレンジを開始したきっかけ

- 論文を意識して読んでこなかった
  - 視野狭かった
- 知識量の重要性に気づくチャンスが増えた
  - AISTに来て周囲にプロ研究者が増えた
- 現状維持ではいけないと強く感じていた
  - アメリカ/ドイツへの留学
  - ポスドクという立ち位置，重なる就職の失敗
- 研究の楽しさを当時研究連携の学生に伝えたかった
- 誰もやらないことをやりたい！
  - キチガイでもなんでもいいから

2,843

169,390

17

cvpaper.challengeの２年半を簡単に

@2015
CVPR2015の論文
計
602
本を完全読破

@2016
論文読破
計
1,000
本を達成

– 2.5年で論文まとめ2,000超，論文10本超，招待講演8件
– arXiv:cv 週間お気に入り数世界一位
– スライドアクセス数 約90,000/年（Google/Facebook/Amazon/Apple等世界からアクセスあり）
– 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（7本採択，7本審査中）

この2年半で何が変わった？

個人戦から
戦略的思考
への転換

この2年半で何が変わった？

2015年以前：個人プレー

- 論文調査：自分の狭い分野のみ
- 研究：従来法の単純な改善
    - e.g. CoHOG => Extended CoHOG [Kataoka+, ACCV14]（共起性 + 強度）

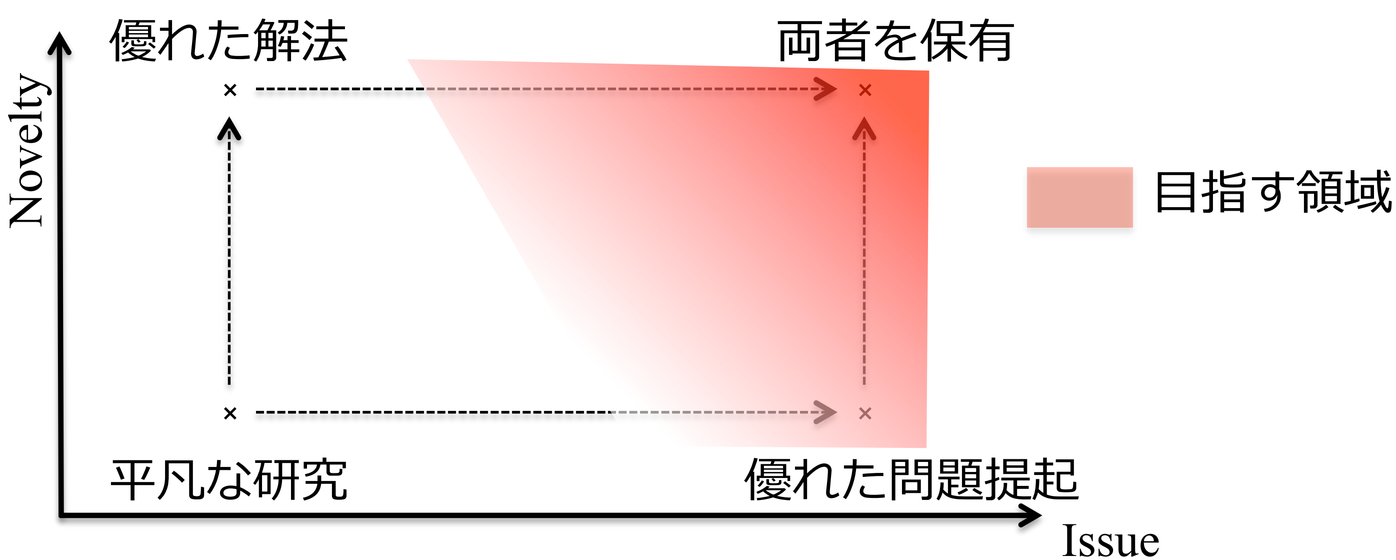
当プロジェクト開始後：組織的戦略

- 論文調査：網羅的に分野を把握
- 研究：物事の本質に迫るような問いを意識
- サーベイ/テーマ/実験に至るまで「質・効率を高める努力」を徹底

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



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

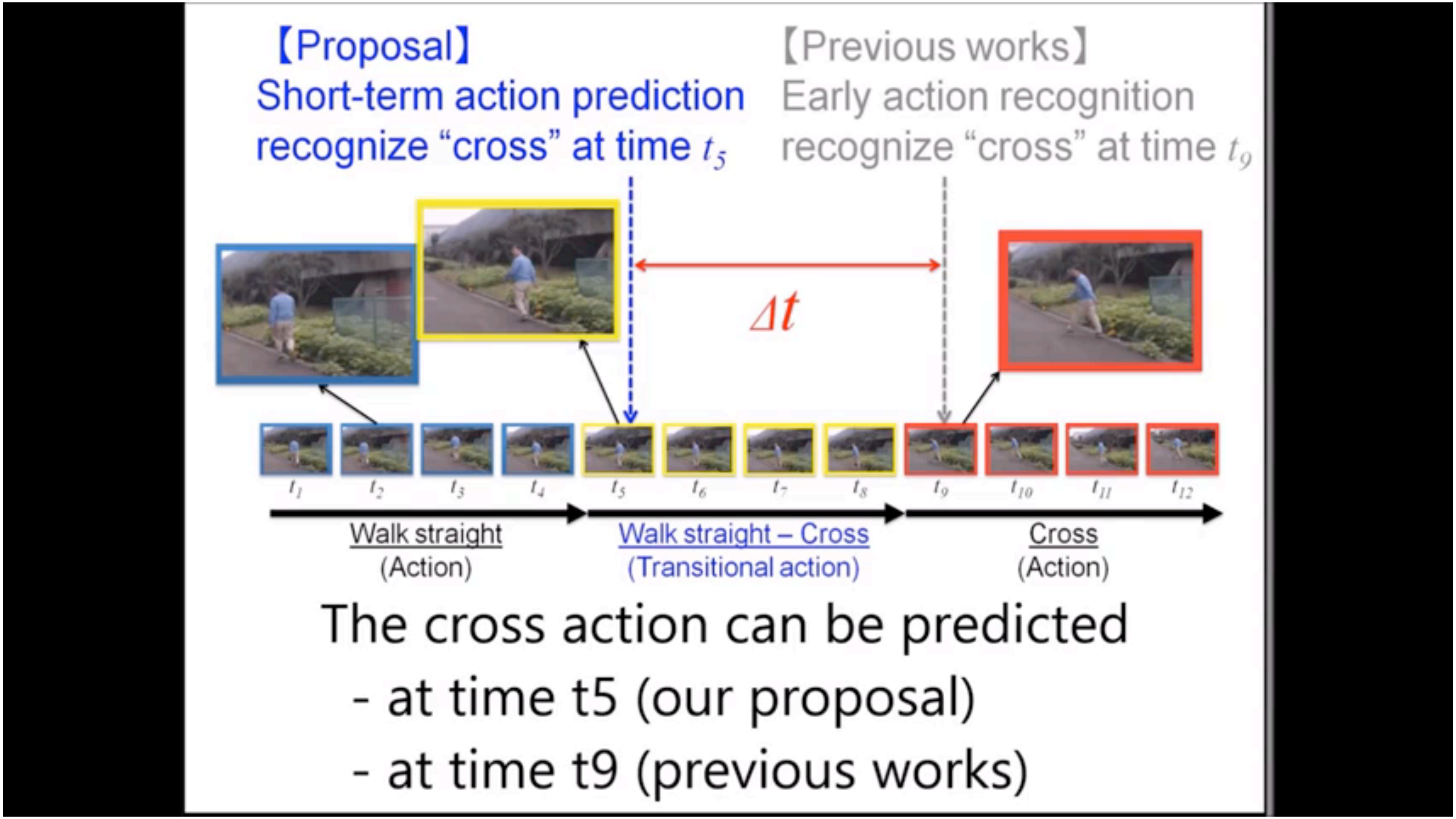
以下，既に公開済みの研究について紹介

メンバーからの研究紹介

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

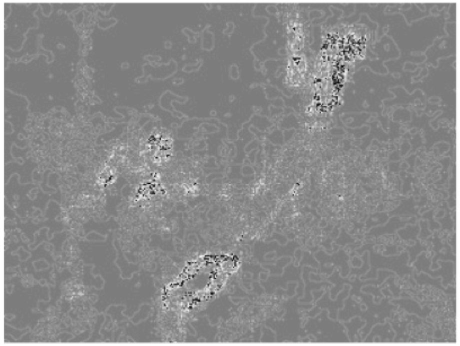
H. Kataoka et al. “Human Action Recognition without Human” ECCV 2016 Workshop Oral (Best Paper) & MIRU 2017 Oral.

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.

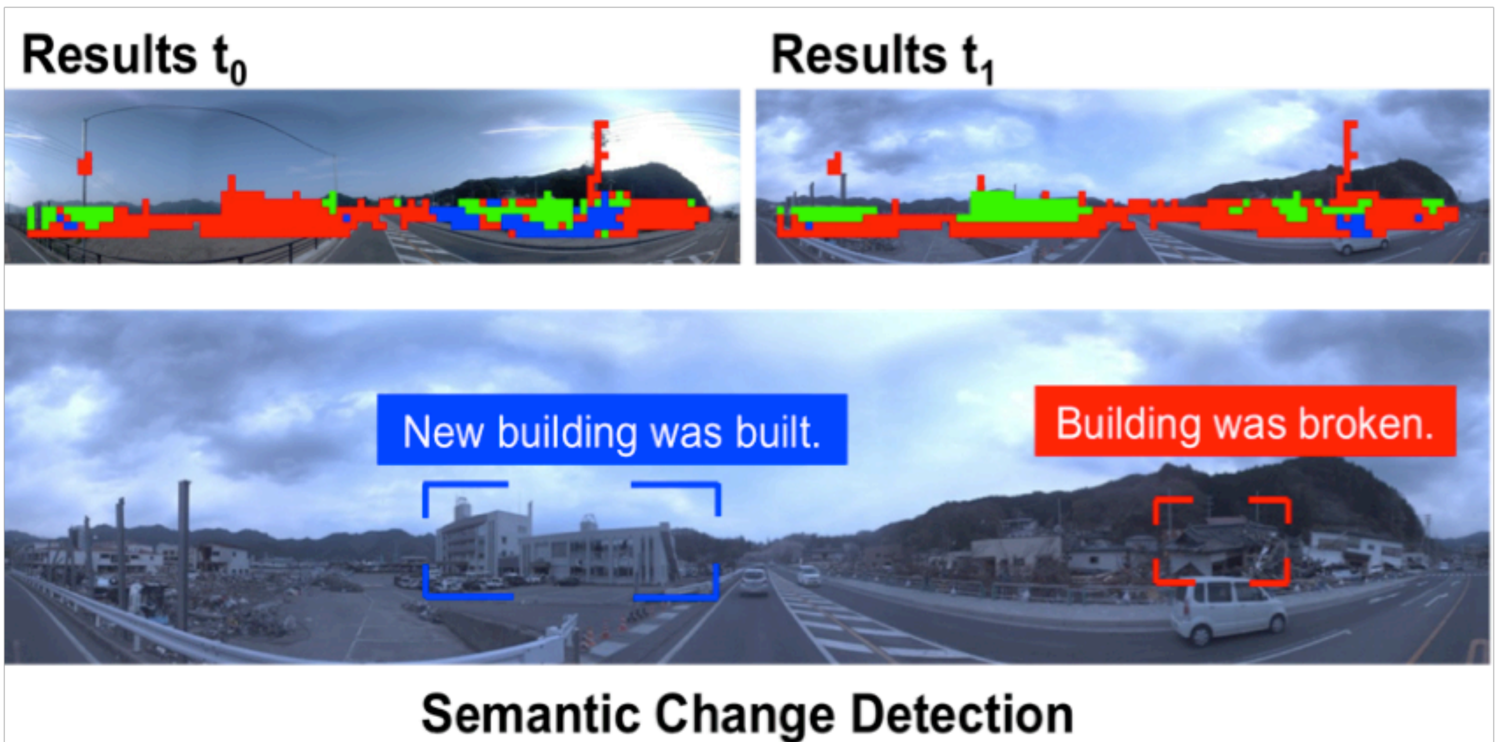
Semantic Change Detection

arXiv2016

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 89^{th} Academy Awards

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

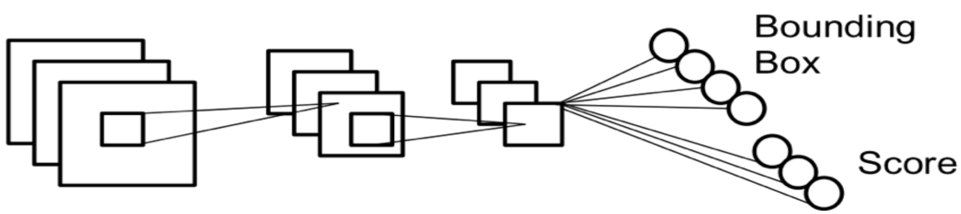
Pedestrian Near-Miss Detection

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

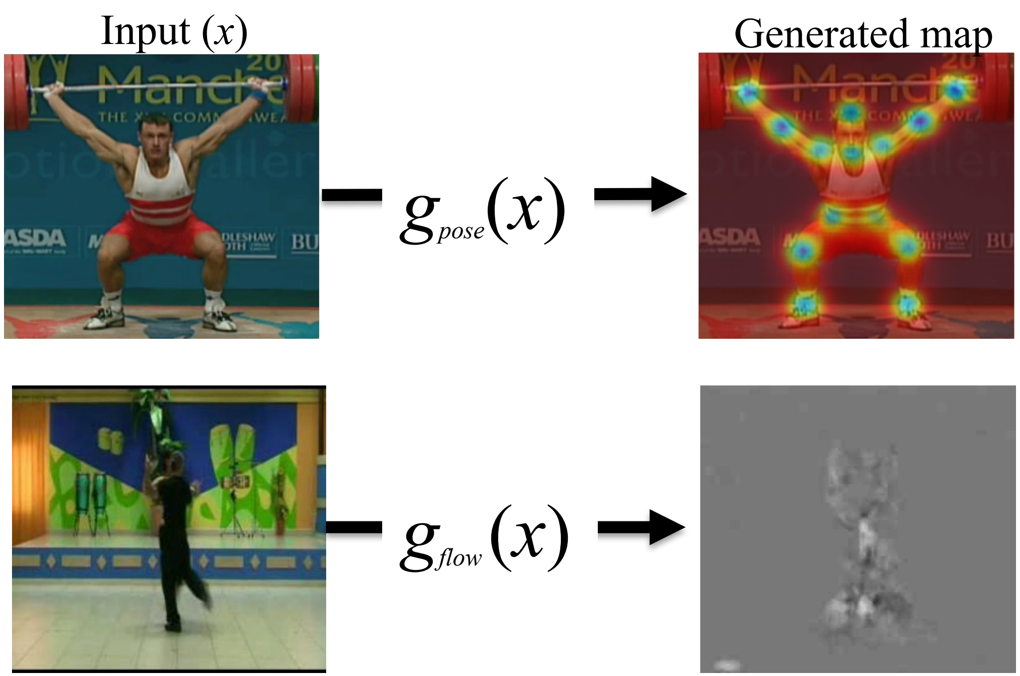
T. Suzuki et al. “Pedestrian Near-Miss Analysis on Vehicle-Mounted Driving Recorders” in MVA, 2017.

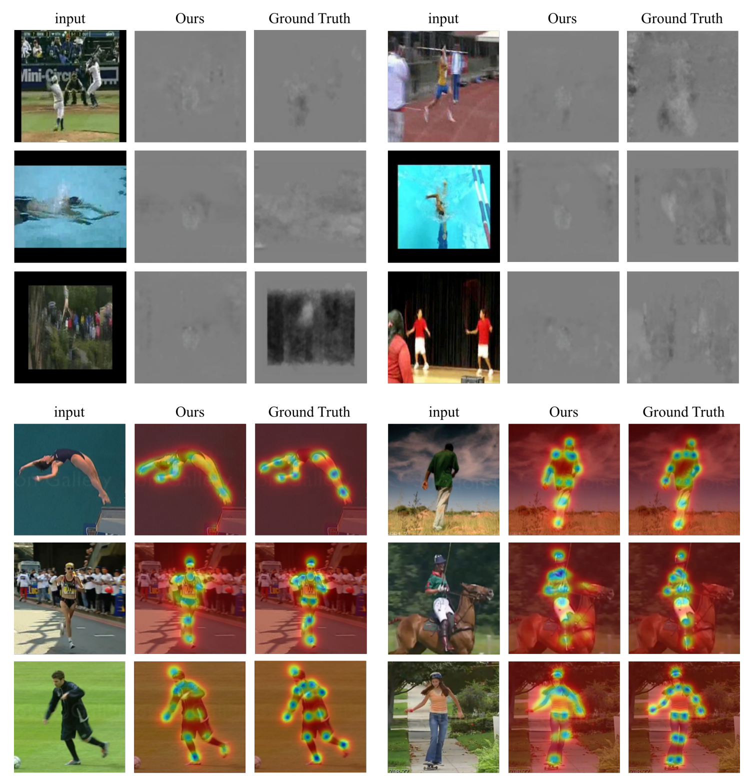
Generated Motion Maps

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



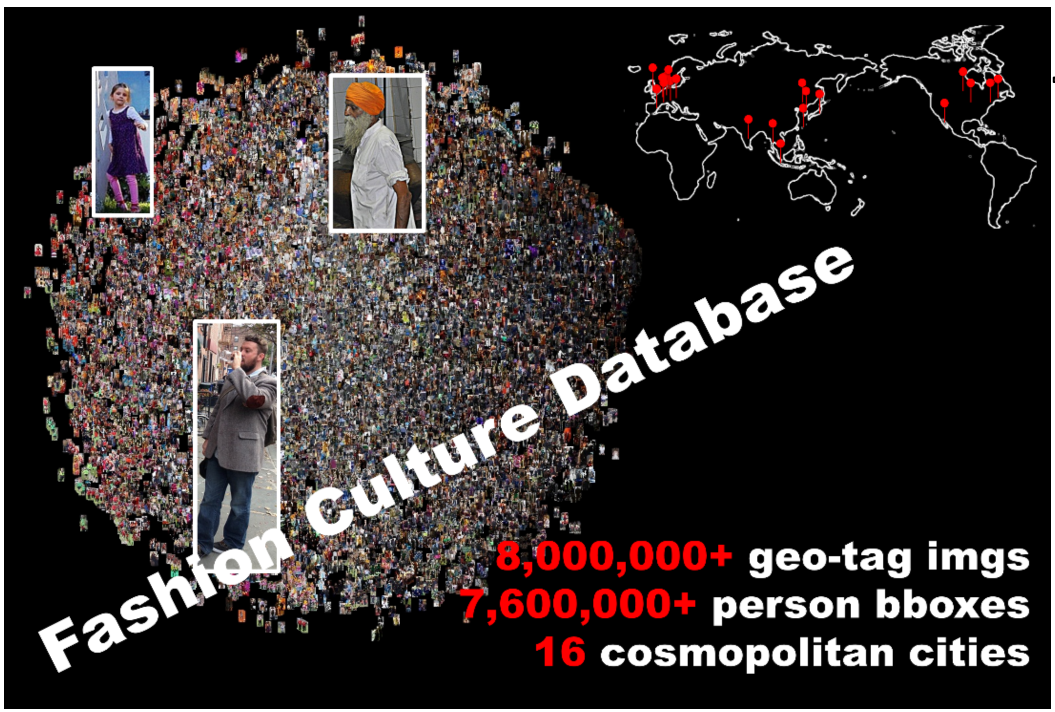


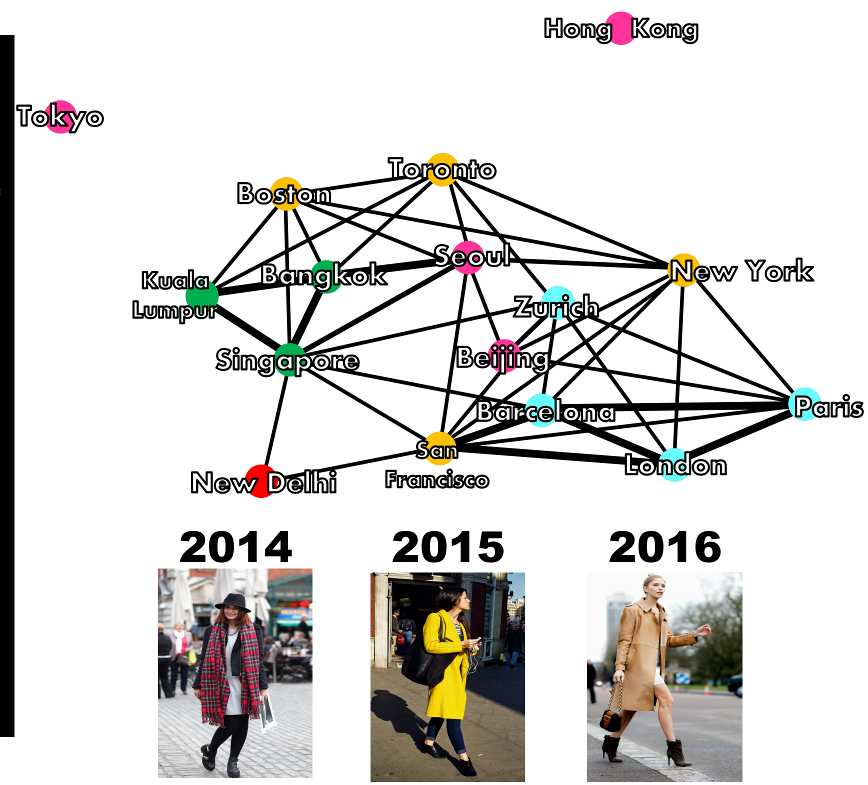
Y. Matsuzaki et al. “Generated Motion Maps” in CVPRW, 2017.

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





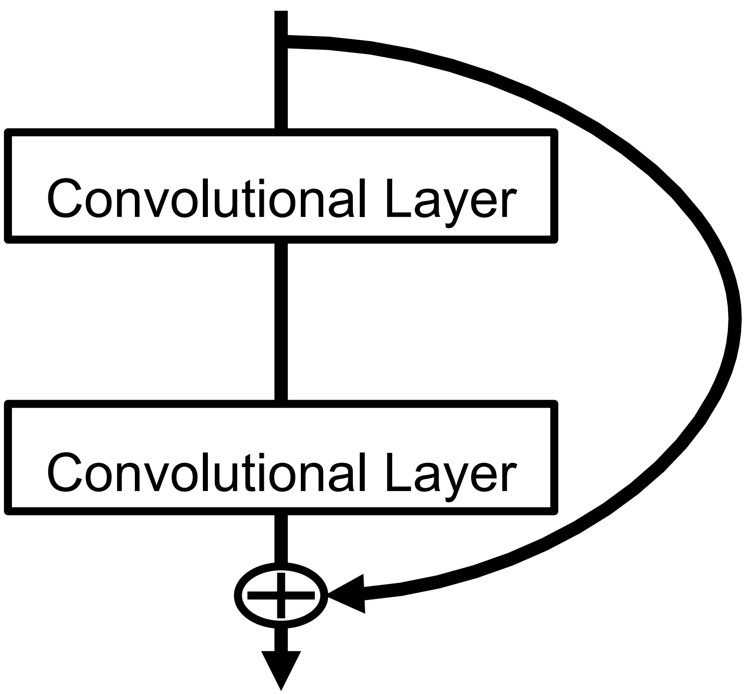
K. Abe et al. “Dynamic Fashion Cultures”, arXiv, 2017 & MIRU 2017 Oral. (Best Student Paper)

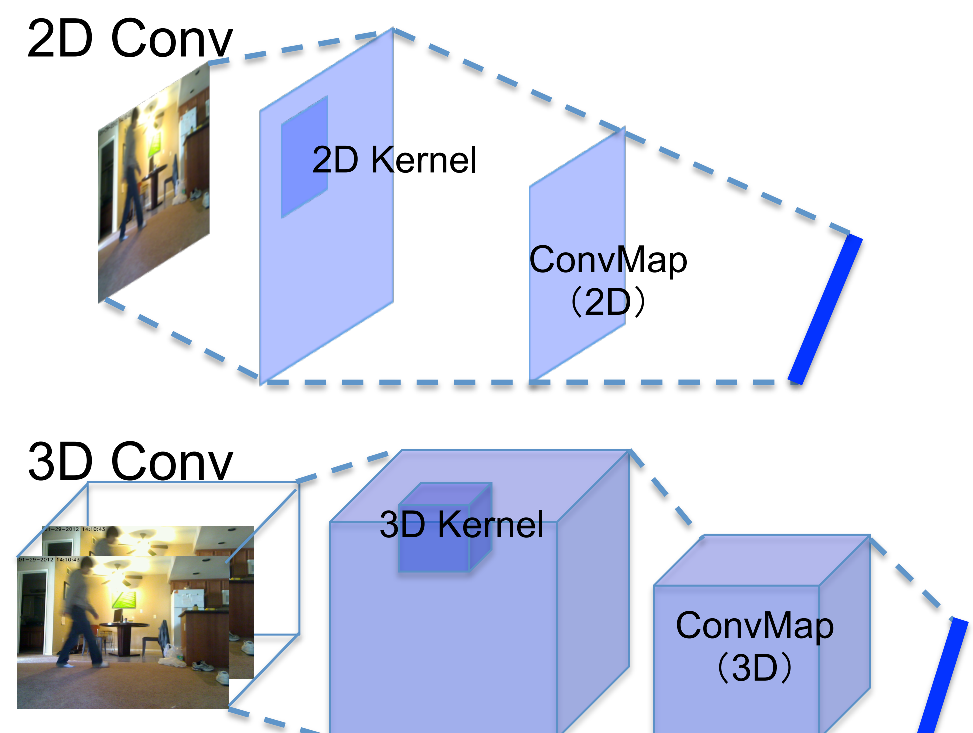
3D ResNets

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）

リクエスト

国際会議の採点

（問題設定）－（ミス）

を点数化して上位N*%に入るか

*Nは採択率

投稿時に気を使うこと

- 問題設定
  - 著者も持ち点は自分で決められる
  - 新分野開拓？解法考案？従来法の改善？
- ミス
  - 構成から問題設定の欠陥，実験の欠落など
  - ストーリーに無理がない？文法のミスは？
  - 大きく言っても良いが，無理があると大減点！（問題設定のミスは挽回できないかも？）
- 採択率
  - 世界トップ層が投稿するので（それ以外の会議と比較して）数字以上に困難

良い研究をすると同時に，ミスをしないこと

メタな技術を身につけるのが一番の近道？

国際会議の通し方は確かに存在

- 会議の性質/分野の性格を熟知
- アカデミック・ライティングは重要
- 論文構成・ストーリー・結果のインパクト・議論
- 査読者にとって面白ければ通る！？

普段からその分野の論文を読む，会議に参加する

通しやすい領域

- これと言ってないが，重要なのに誰もやらない内容が通る
  - 分野の流れを読む
  - 視点を変える
  - 自分オリジナルの領域を開拓

投稿する体制

通しやすい＝常時投稿する体制

– 組織にアイディア考案と試行する体力があるか
– 研究のレシピ（後述）

投稿で気をつけていること

ニッチでも良いから世界でトップの研究をして，それが証明できたら査読を突破できる！

現在，最もホットだと思う分野

ICCV 2017 速報より抜粋

フォーカスすべき研究分野（1/8）

3D Convolution

– より高次なカーネルでの畳み込みが重要
– 時系列処理（xyt），実空間処理（xyz），ライトフィールド（xyuv）など2D+α
– 当プロジェクトでも3DカーネルのPyTorchのコードや学習済み時系列モデルを提供
  - 3Dカーネルを持つResNetsをKinetics Human DBで学習
  - 3D-Resnets-PyTorch: https://github.com/kenshohara/3D-ResNets-PyTorch
  - 3D-ResNets-Torch: https://github.com/kenshohara/3D-ResNets

フォーカスすべき研究分野（2/8）

より高次なラベルを返却するモデル

- 人間の知識や直感等をコンピュータに実装
  - 危険，異常，不快，など
- 画像キャプション/質問回答はより進化する？
  - 画像の事象を総括的に捉える，組み合わせて初めて明らかになる知識の獲得
- 画像に見えないものを推定
  - 行動予測，未来の状態推定

フォーカスすべき研究分野（3/8）

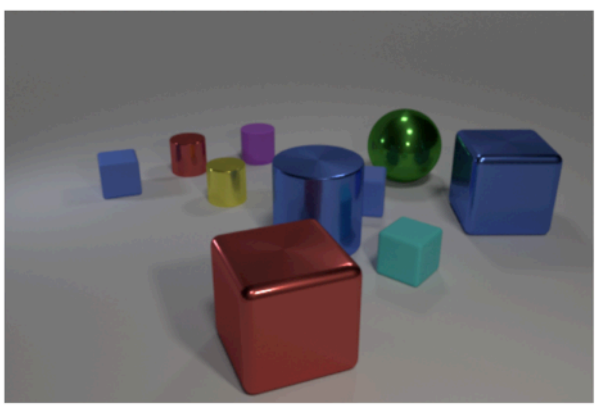
高度な教師なし学習

- 人間を真似て教師あり学習を倒せるような学習を考案
- 教師なしでパラメータ学習して，画像識別タスクでImageNet学習済モデルに勝てたら高度と言っていい？
- ImageNetに置き換わるような特徴表現学習を実施！

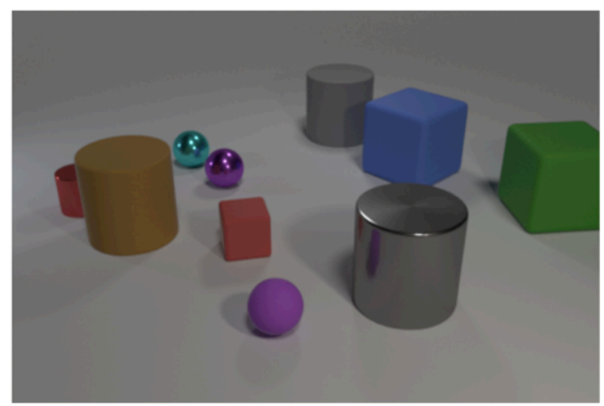
フォーカスすべき研究分野（4/8）

説明できるコンピュータ

- Beyond 100%?
  - https://www.slideshare.net/SeiichiUchida/cnn-65229102
- 100%認識できるよりも，80%の性能でも残りの20%を「説明できること」（なぜなぜ質問をコンピュータにさせる？）
- 今回もGrad-CAMやVisual Reasoning（下図）など



Q: What shape is the object reflected in the blue cylinder?
A: cube



Q: What number of cylinders share the same color?
A: 2

フォーカスすべき研究分野（5/8）

アプリケーション（社会実装）

- 技術と同時に，実装に使うところまで試してみる
- Deploy or Die!

フォーカスすべき研究分野（6/8）

成熟したアルゴリズムをベースにして研究する

- 研究分野ではないが重要な考え方
- 組み合わせるでも良いし，高精度なモデルを正解値としてもよい
- Keywords: Object Detection, Semantic Segmentation, Instance Segmentation, Video Recognition etc...

フォーカスすべき研究分野（7/8）

昔に学ぼう！

- 高度になったアルゴリズムを用いて以前の問題を解決するものを考案
- 低次元な局所特徴
  - Learning Compact Geometric Features: 32次元（約10分の1のサイズ）で従来よりも高精度な3次元点群マッチングを実現

フォーカスすべき研究分野（8/8）

データを集めよう！

- 自分の問題設定を作って研究する
- データも同時に集めて（できれば）公開する

レシピを紹介！

大別すると,

– サーベイ：現状を正しく把握する能力
– 発想：新しいコンセプトを生み出す
– 実装：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

サーベイ

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？

II. Related work

A. Traffic data and approaches to its representation

Several practical databases for pedestrian detection, such as the French Institute for Research in Computer Science and Automation (INRIA) Dataset [3], Caltech [4], and the KITTI Vision Benchmark Suite for self-driving cars [2]) have been proposed in the past decade. The information contained in the KITTI database, which has been used to set meaningful vision problems for self-driving cars [2] as well as problems related to stereo vision, optical flow, visual odometry, semantic segmentation, two- and three-dimensional (2D/3D) object detection, and 2D/3D tracking, has proven especially useful.

In 2015, these problems were updated for stereo and optical flow [5]. Thanks to the development of sophisticated approaches, such as fully convolutional networks (FCN) [6] and region-based convolutional neural networks (R-CNN) [7], there has been improved performance of solving these problems using the KITTI benchmark dataset. In addition, a manner of geometry allows us to improve the rate of object detection [8] and optical flow [9] not only in stereo [10]. As for semantic segmentation, we can now obtain knowledge about dense connections and use this information with graphical models [11], [12].

Unfortunately, none of these datasets contain scenes of near-miss incidents in which pedestrians, cyclists, or other vehicles must be avoided. Thus, there is an urgent need for a collection of incident scenes that can be used to train self-driving cars on how to safely navigate such dangerous situations.

論文にもRelated workを書くこと多し

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

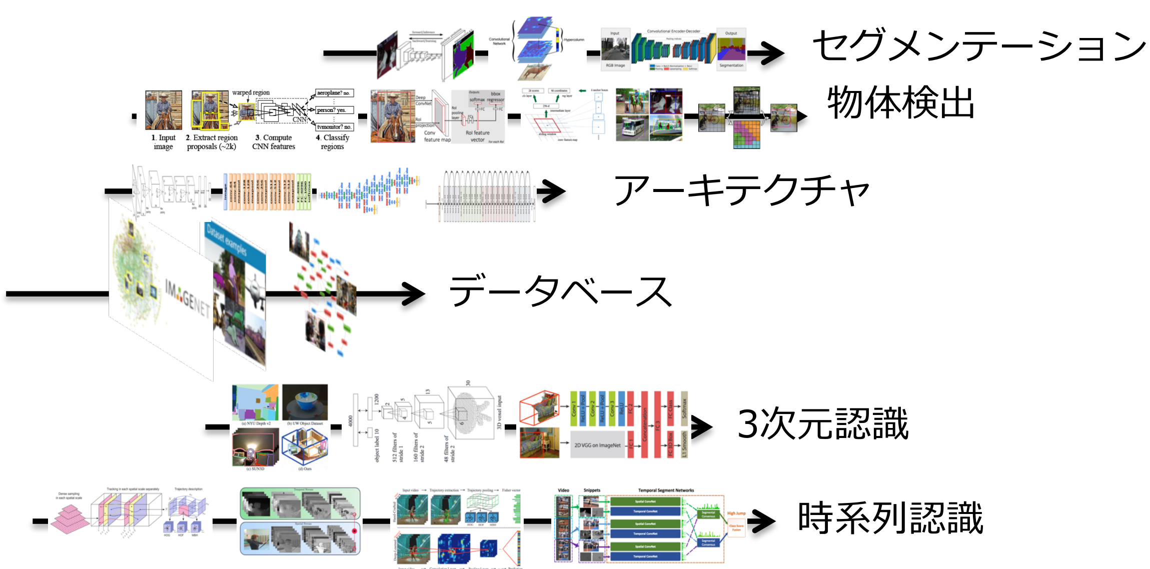
– サーベイの質や量が研究テーマ考案・選定に直結

– 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（参考にするなら歴史を追う，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！

一人で読むよりも，みんなで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

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

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

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