

コンピュータビジョンの今を映す —CVPR2017速報より—

片岡裕雄

原健翔, 阿部香織, Yue Qiu

鈴木亮太, 鈴木智之, 大喜周平, 張雨辰, 範龍男

<https://sites.google.com/site/cvpaperchallenge/>

コンピュータビジョンの今を映す挑戦

論文読破・まとめ・アイデア考案・議論・実装・論文執筆・
(社会実装) に至るまで広く取り組む

- 人員：産総研，筑波大，東京電機大，慶應大による約20名
- 読む：多読・体系化・共有
- 書く：洗練された問題設定を考案・論文化（・社会実装）
- 2017年はトップ会議※への投稿を目標に活動中



cvpaper.challenge
@CVpaperChalleng

産総研片岡(@HirokatuKataoka)と電大(is.fr.dendai.ac.jp, HQ宮下), 筑波大(HQ白壁)による合同プロジェクトです。現在, CVPR2015の論文602件をまとめております。
mail:cvpaper.challenge[at]gmail[dot]com

📍 Tokyo, Japan
🌐 slideshare.net/cvpaperchallen...

cvpaper.challenge

Edit profile

16 SlideShares
23 Followers
0 Clipboards

👤 Researchers, Graduate Students

🌐 <https://twitter.com/CVpaperChalleng>

📄 産総研片岡裕雄(@HirokatuKataoka)と電機大中村研(<http://www.is.fr.dendai.ac.jp/>)による合同プロジェクト「cvpaper.challenge」です。現在, CVPR2015の論文602件を読破するチャレンジ中。キーワード: コンピュータビジョン, パターン認識, トップカンファレンス, 人工知能, CVPR2015

※ HP, Twitter, SlideShareもご覧ください

HP: <https://sites.google.com/site/cvpaperchallenge/>

Twitter: [@CVpaperChalleng](https://twitter.com/CVpaperChalleng)

SlideShare: [@cvpaperchallenge](https://slideshare.net/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

@2015

CVPR2015の論文

計

602

【主な業績】

・招待講演 x 1 (PRMU研究会)

本を完全読破

@2016

論文読破

計

1,000

【主な業績】



本を達成

- ECCV2016 WS x 2
- BMVC2016
- ViEW2016 oral/spotlight oral
- 招待講演 x 2 (東大, IAIP)

@2017

TOP会議 (GスカラーTOP20論文) への投稿

計

10

【主な業績】

MIRU2017 学生優秀賞
Dynamic Fashion Cultures



本を達成

- ICCV2017 WS x 1
- MIRU2017 (oral x2, poster x1)
- MVA2017 x2
- 招待講演 x4 (中京大, 慶大, 早稲田大, ABEJA)

ではなぜ、そこまでサーベイするのか？

世界水準の研究のため

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
- ティンドを知らないと(天才でない限り)最先端の研究を生ま出すことは難しい

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

現状を変える力になる？

日本人研究者はサーベイが足りない？

- よく言われるが、そうは思わない（読んでる人は読んでいる）
- しかし海外の（大学院生含めた）研究者は年間数百本読んでる

このプロジェクトは好例になれるか？

- 日本のプレゼンスを少しでも高める
- 一足飛びに追いつくには思考が飛んだことをやらないといけない
- 論文を膨大に読んで動向を網羅的に把握する

CHI勉強会2016

The ACM CHI Conference on Human Factors in Computing Systemsの論文（Proceedingsに入っているPapers&Notes）、全545本を6時間で一気に読破します！

- CHIはユーザインタフェース界のトップ会議
- 2006年から12年間続く取り組み
- 日本のHCI分野の好調を支えている（CHI2017は日本がアジアで1位）

cvpaper.challengeはCV分野をさらに盛り上げるための起爆剤になれるか？

CVの現状

CVPRのみならずCVの現状についてご説明します

CVの現状：CVPR2017について

コンピュータビジョン, パターン認識分野のトップ会議

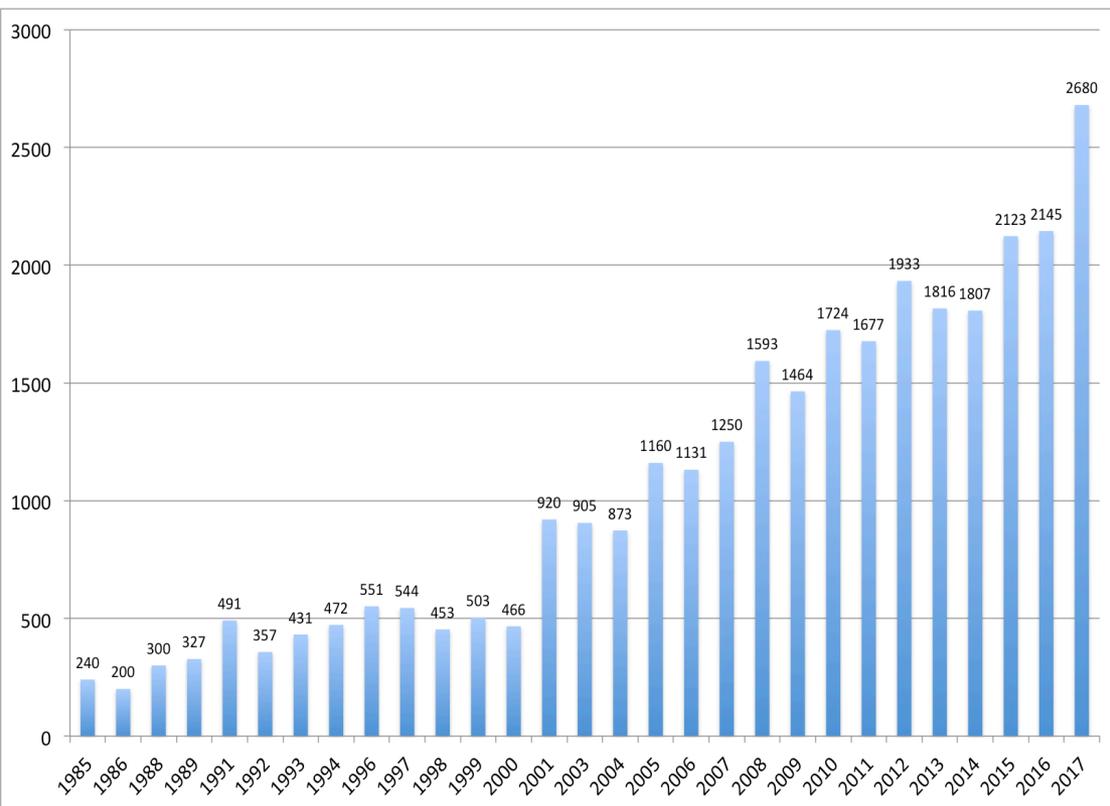
- Google Scholar (CV&PR)の統計にて常に一位
- 採択率は25~30%
- 2017年はハワイ・ホノルルにて開催



CVの現状：CVPR2017について

AIブームの流れで投稿数・参加者数等増加傾向

- 投稿数は過去最高（2,620論文）を記録
- 参加者数 約5,000人, スポンサー数127



http://img.mp.itc.cn/upload/20170727/a94c7cfda6c34d2fb3f52825c3c6b928_th.jpg

arXiv.org

30 - 50/day @CS:CV
7,500+ papers/year?



783 papers!

<http://cvpr2017.thecvf.com/images/CVPRLogo3.jpg>

621 papers!



<http://iplab.dmi.unict.it/acvr2017/img/conjunction.png>

主要2会議で1,400+本/年

学会

速報

密な議論、人脈

とにかく速い

(信憑性はコードで測定)

学会は査読を突破した

という承認



認められてから査読を突破するルート?

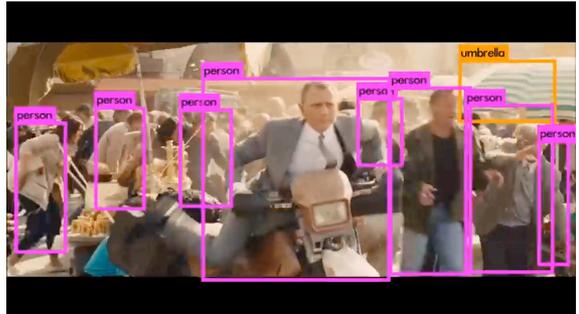
ライブに行く感覚

速報はライブの広告・宣伝

キーワードはarXiv (論文) GitHub (コード) 公開DB (データ)

論文採択前に「質の高い」全ての情報が出揃ってしまう!

YOLO_v2 (物体検出)



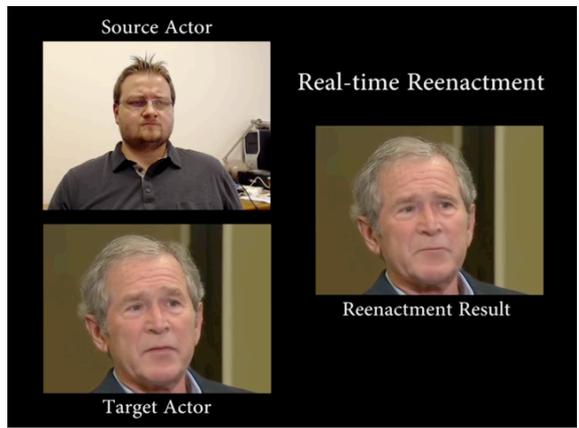
PSPNet (意味ラベル)



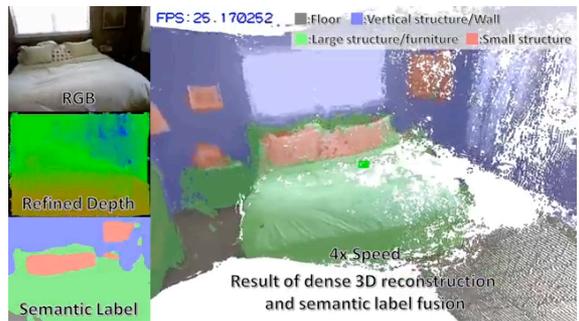
OpenPose (姿勢推定)



Face2Face (顔器官追跡)



CNN-SLAM (自己位置推定&マッピング)



Lip Reading (読唇術)



加速する進化, 世界中の誰もが当事者になるチャンスが与えられた?

このような現状も踏まえて動向や今後すべき行動について解説！

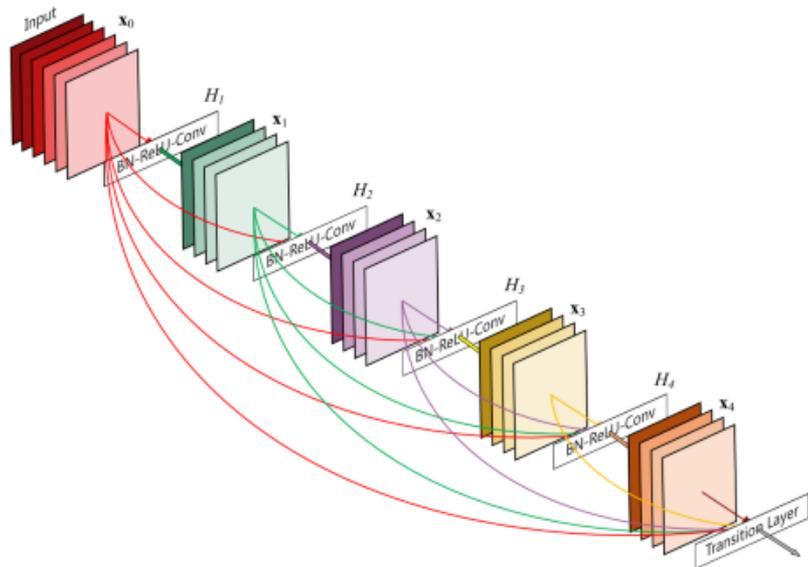
CVPR 2017の動向・気付き

- # 今までの蓄積, 今回CVPR2017論文を読んで気付いたことや海外の研究者の動向をまとめました
- # あまり繋がりがなく, 列挙しただけですがご了承ください

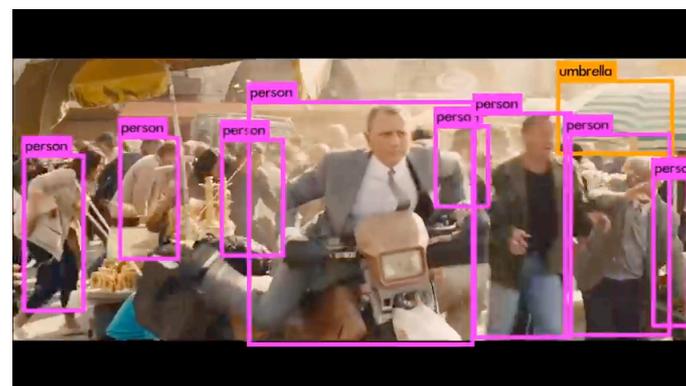
CVPR2017の動向・気付き

– (DNNは) タスク特化に加えて高精度化

- 精度を改善するための取り組み
- DNN初期からある画像識別/物体検出を例にすると...



DenseNet [Huang+, CVPR17]
CVPR Best Paper



	YOLO								YOLOv2
batch norm?		✓	✓	✓	✓	✓	✓	✓	✓
hi-res classifier?			✓	✓	✓	✓	✓	✓	✓
convolutional?				✓	✓	✓	✓	✓	✓
anchor boxes?				✓	✓				
new network?					✓	✓	✓	✓	✓
dimension priors?						✓	✓	✓	✓
location prediction?						✓	✓	✓	✓
passthrough?							✓	✓	✓
multi-scale?								✓	✓
hi-res detector?									✓
VOC2007 mAP	63.4	65.8	69.5	69.2	69.6	74.4	75.4	76.8	78.6

YOLO_v2 [Redmon+, CVPR17]
CVPR Honorable Mention Award

CVPR2017の動向・気付き

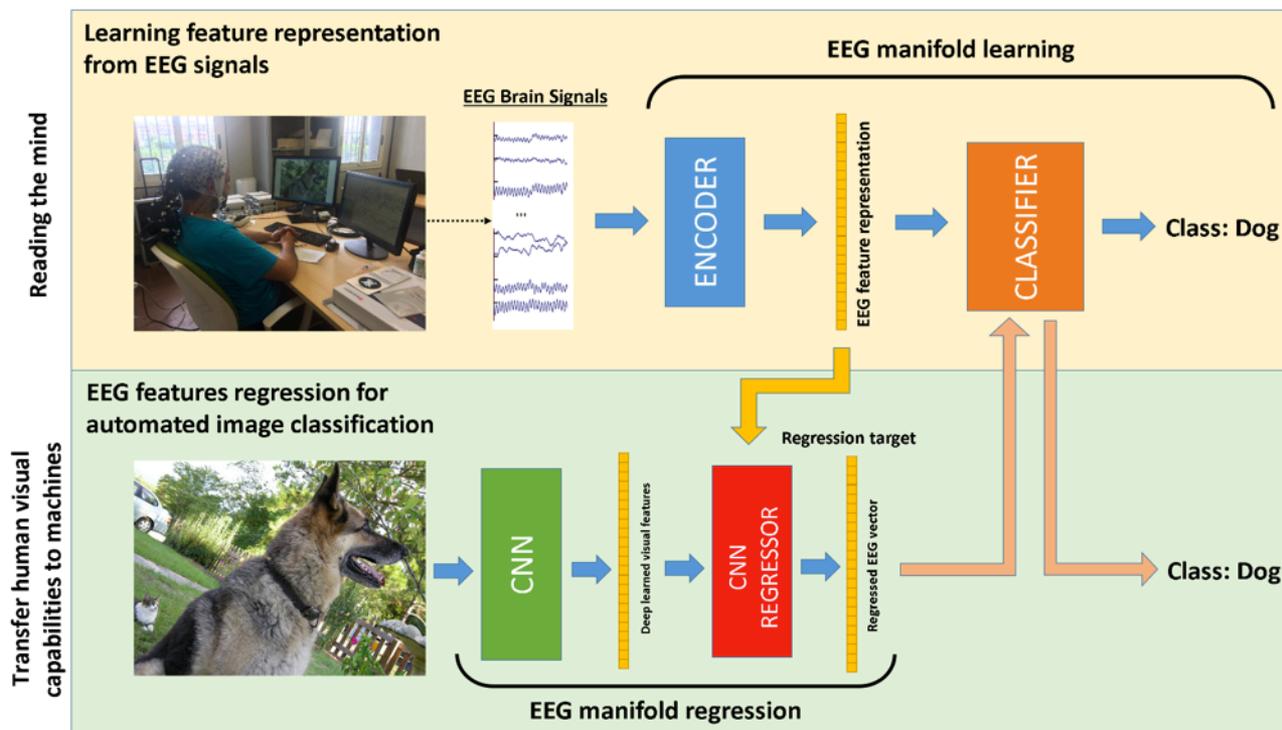
- データベースの大規模化はある程度収束
 - 「ただ大きい！」だけのデータは必要ない？
 - 意義のあるデータ/ クリーンなデータが重要と分かってきた
 - 弱いラベル/ ラベル無し データでの手法を考案も評価高

- 教師を自動・半自動で獲得する
 - 弱いラベル（画像単位, クリックなど）による学習
 - CGなどから人工的な学習データを生成（静止画, 動画問わず）
 - より実データに近いデータをGANなどにより生成

CVPR2017の動向・気付き

– DNNの解明に向けた研究

- 中間層の挙動解明 (e.g. Network Dissection)
- クラス間の境界を探る (e.g. Perturbation)
- 脳の信号から画像識別する研究も見られた (Deep Learning Human Mind; 下図)



CVPR2017の動向・気付き

– 従来課題をベースに問題を複雑化

- 基礎的な問題をベースにして問題を拡張
- ImageNetの学習済みモデルを使用した識別がほぼ全ての礎となり、物体検出・セグメンテーションなどへ応用
- より多様（姿勢，環境，キャプション，動画）で実世界に近くなる
- 複数のタスクを単一のフレームワークで同時実行

新しいデータセットやアノテーションを用意して
少しでも問題を拡張/他の研究と差別化する工夫

CVPR2017の動向・気付き

– ボリュームデータへの移行

- ボリュームデータとは，画像xy空間から新しい次元を追加
- 奥行Zを含めた実空間（XYZ），時間軸Tを含めた動画像（XYT）等
- 基礎的なデータや手法の整備をまずは進めている



Figure 1. Example reconstructed spaces in ScanNet annotated with instance-level object category labels through our crowdsourced annotation framework.

 **ACTIVITYNET**

Kinetics

SPORTS 1M

YouTube | 8M

6766 / 13220

28108

306245

1M / 8M

51 / 101

200

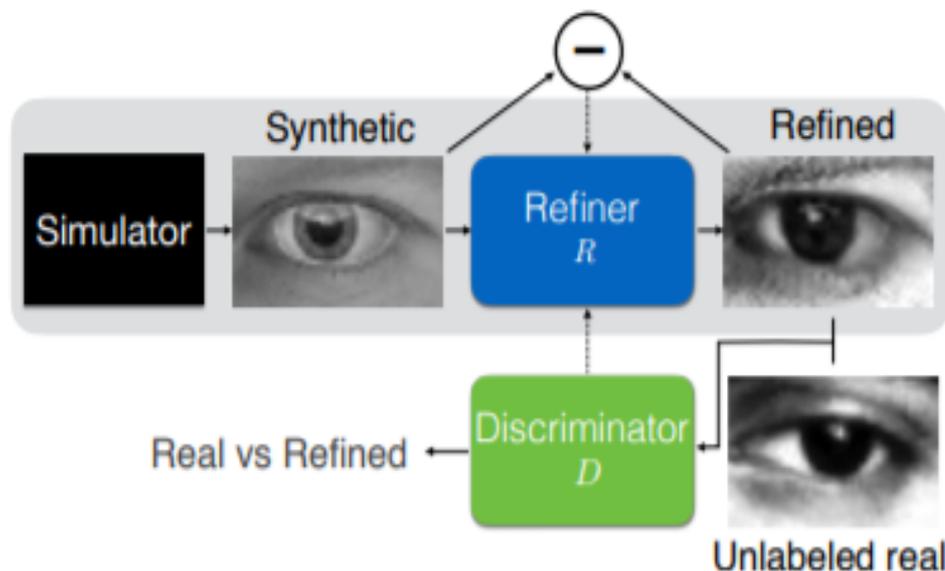
400

487 / 4716

CVPR2017の動向・気付き

– GANの効果的な利用

- CVの国際会議で画像生成が流行るとは思っていなかったが、これからは「識別器 (D) を向上させる」側に注目が集まりそう？
- Apple: 合成データ (Synthetic Data) をリアルにする仕組みを導入してリアル画像と見分けがつかない画像へと変換 (SimGAN; Best Paper)
=> (世界中で) パターン空間を広げる取り組みが進行中？
=> データを作るのみならず, 学習で同時に識別器を鍛える？



SimGAN [Shrivastava+, CVPR17]
CVPR Best Paper

CVPR2017の動向・気付き

– 根強く残る分野へ攻める

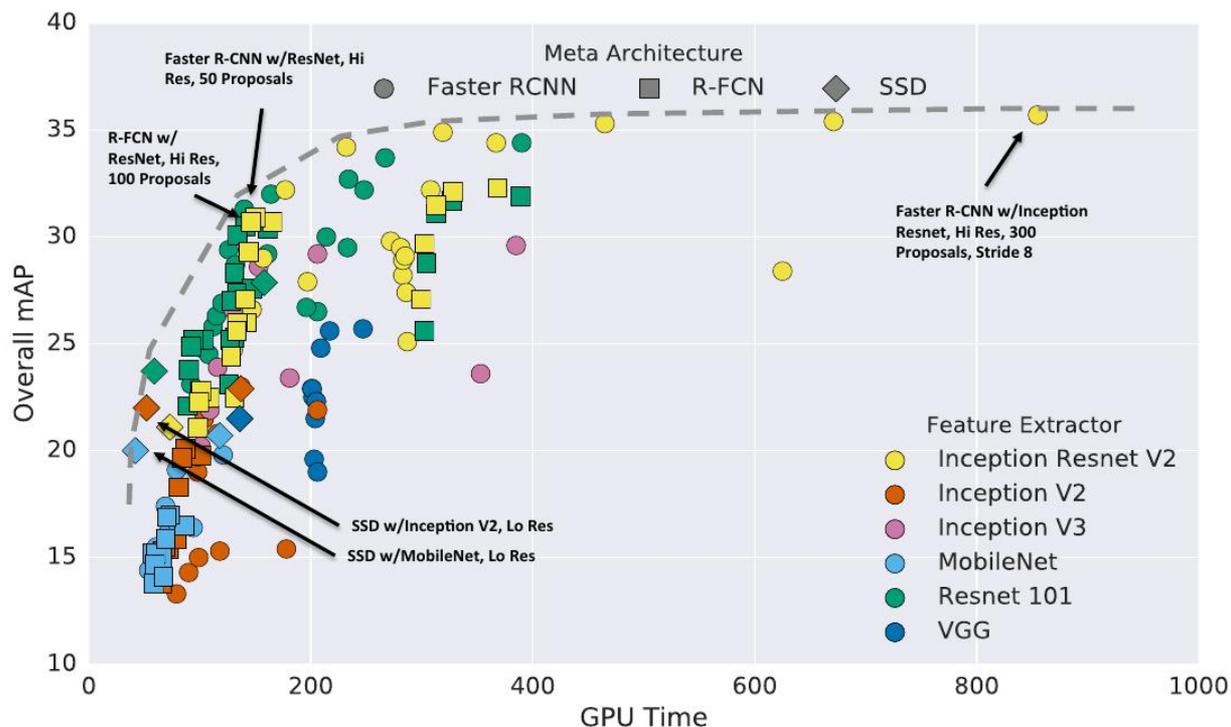
- Computational Photography, カメラ幾何等の純CV
- まだDeep Learningが入り込んでいない（というか入らなくても良い分野ももちろんある！）
- Machine Learningのセッションと比較して人数が少ない（= 査読を突破する可能性も上がる？）

今の時代、（逆に）DNNをやらない研究の方が通しやすい？

CVPR2017の動向・気付き

– 徹底的な比較論文は通る

- Speed/Accuracy Trade-off (Google)
- HPPatch: 局所特徴量のベンチマークと徹底比較
- 比較による分野への貢献 (=新規性)



CVPR2017の動向・気付き

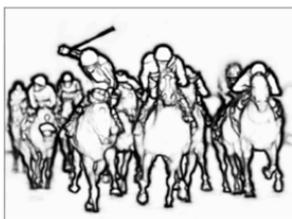
– ある意味万能・多機能なネットワークを構成

- UNet, pix2pix

Input



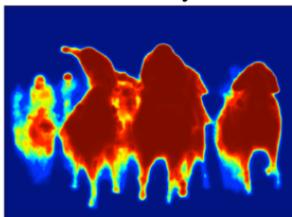
Boundaries



Surface Normals



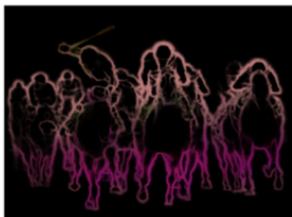
Saliency



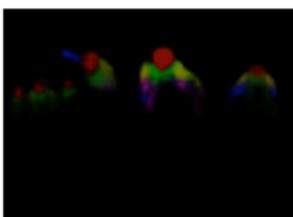
Semantic Segmentation



Semantic Boundaries



Human Parts



Detection



Labels to Street Scene

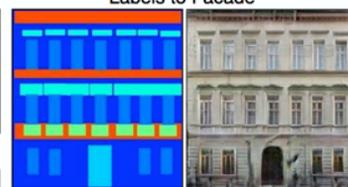


input



output

Labels to Facade



input



output

BW to Color



input



output

Aerial to Map



input

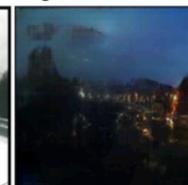


output

Day to Night



input



output

Edges to Photo



input



output

CVPR2017の動向・気付き

– 何と言っても本質はデータ！

- データを活かすDNN構造の考案が基本
- 問題を的確に捉え理論的に考えること

何をしてもある程度うまくいってしまうが、
目的からの逆算で手法/データの組み替えが最大効果

CVPR2017の動向・気付き

- 論文の覇権争いは国際会議前に終わっている
 - 事前に宣伝 (arXiv, GitHub, 公開データ)
 - 会議前：見たいと思わせる, 会議後：やっぱり凄かったと思わせる
 - 今回, YOLO9000やpix2pixの戦略が参考になった (真似できるかどうかは置いて。.)
 - 何れも宣伝はもちろん, Twitterでも効果的に拡散
 - 特に, YOLO9000の圧倒的な開発力は評価に価するし, ポスター (下図) はズルい



CVPR2017の動向・気付き

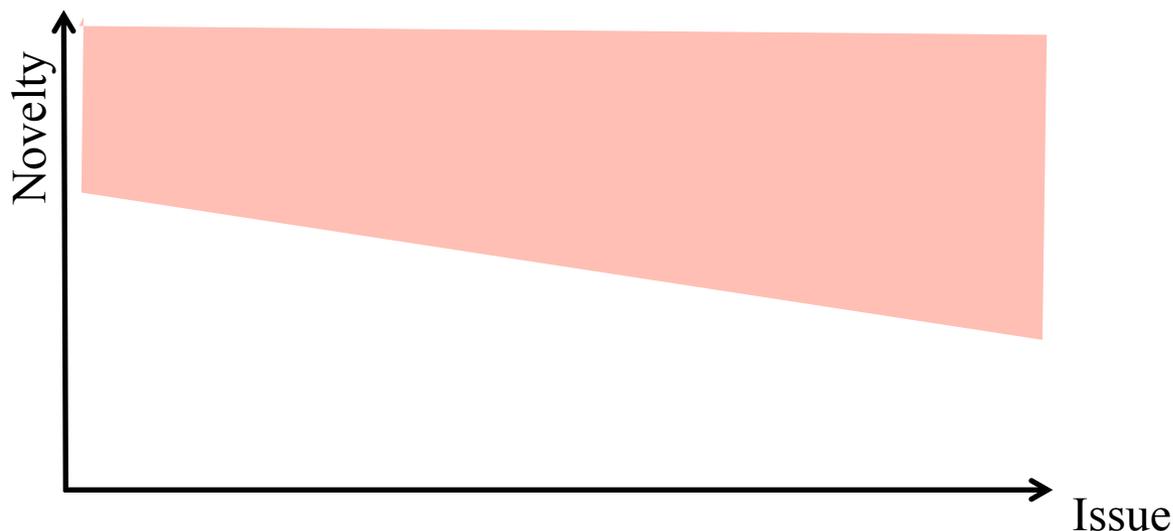
- アイディアや研究テーマが似通ってきた？
 - 同データ (ImageNet/PascalVOC等) で同手法 (ResNets/R-CNN/FCN等) をベースにすると必然的にアイディアは重なる
 - (重要1) 完成度を高めないと取り組み自体消される
 - (重要2) 問題設定・手法ともにBrave Newなアイディアが重宝
- 思いついたアイディアはその年のうちにやる
 - 次の年に同じような研究が出てくる (可能性が高い)
 - 我々のグループ内の (ブレストにより出てきた) アイディアが今年のCVPRでも散見された
 - 誰でも思いつきそうはことは世界中で同時多発的に研究される

「せっかく思いついていたのに」は意味なくて、実際にやって見せないと意味がない

CVPR2017の動向・気付き

– どんな論文が通っている？

- (当たり前だが) 手法として新規性のある手法が通る
- データセットだけを提案する論文はよほど面白くないと通らない(らしい)
- 問題設定をIssue, 手法をNoveltyとすると (直感的に) 下記エリア
 - Novelty重視



CVPR2017の動向・気付き

– 全部読むより現場に行こう！

- 当グループは2015年，CVPR論文を全て読んだが，

「現場に来て情報収集」 する方が良い！

- さらに言うと複数人で来て議論を行う
 - 情報収集を共有（複数人で来て自然に議論している研究グループは強い）
 - 論文の著者と直接話す
 - 欲を言えばその場で研究戦略やテーマを考案

CVPR2017の動向・気付き

– 分野間がシームレスになる傾向？

- これまでよりもアイデアの重なりが大きい（CVで戦うツラさ）
- DNNがツールとして整い他分野の研究者がCVPRに投稿
=> 逆を言うとDNNを武器に他分野で発表するチャンス？

CVPR2017の動向・気付き

– そろそろ次の時代へ？

- 生の声を聞いているとDNNそろそろ飽きたという声もちらほら
 - パラメータ調整問題
 - 複雑ネットワーク構築問題
 - 大規模データセットアノテーション問題 等
- （去年くらいから）Post-DNNを探している研究グループもあるのでは？

今後目指すべき方向

CVPR2017の動向・気付き (Bonus slide)

– 世界トップ研究室の戦略

- 問題意識を先に論文にする
- DBを提案したラボを例にとると。。。



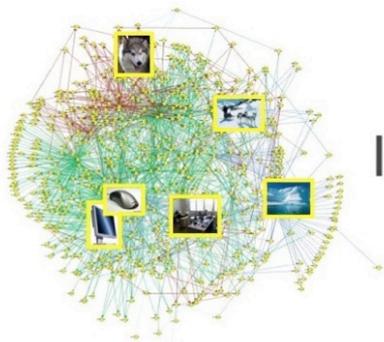
Fei-Fei Li
(Stanford Univ.)

<http://vision.stanford.edu/img/feifei2.jpeg>



Antonio Torralba
(MIT)

<http://web.mit.edu/torralba/www/assets/images/antonioTorralbaS.jpg>



IMAGENET



CVPR2017の動向・気付き (Bonus slide)

– データセットを提供した後に何をするか？

- コンペティションを開催する（問題を解く方ではなく、出す方に回る）
 - 与えられた問題の精度を高くできる人はいくらでもいる
 - 出てきた手法はツールとして使っている
- 自分たちはその間データを拡張および次の問題を創る研究
 - イケメン研究者になろう！

問題を創るために（徹底して）論文を読む，アイデアを考える，手を動かす



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

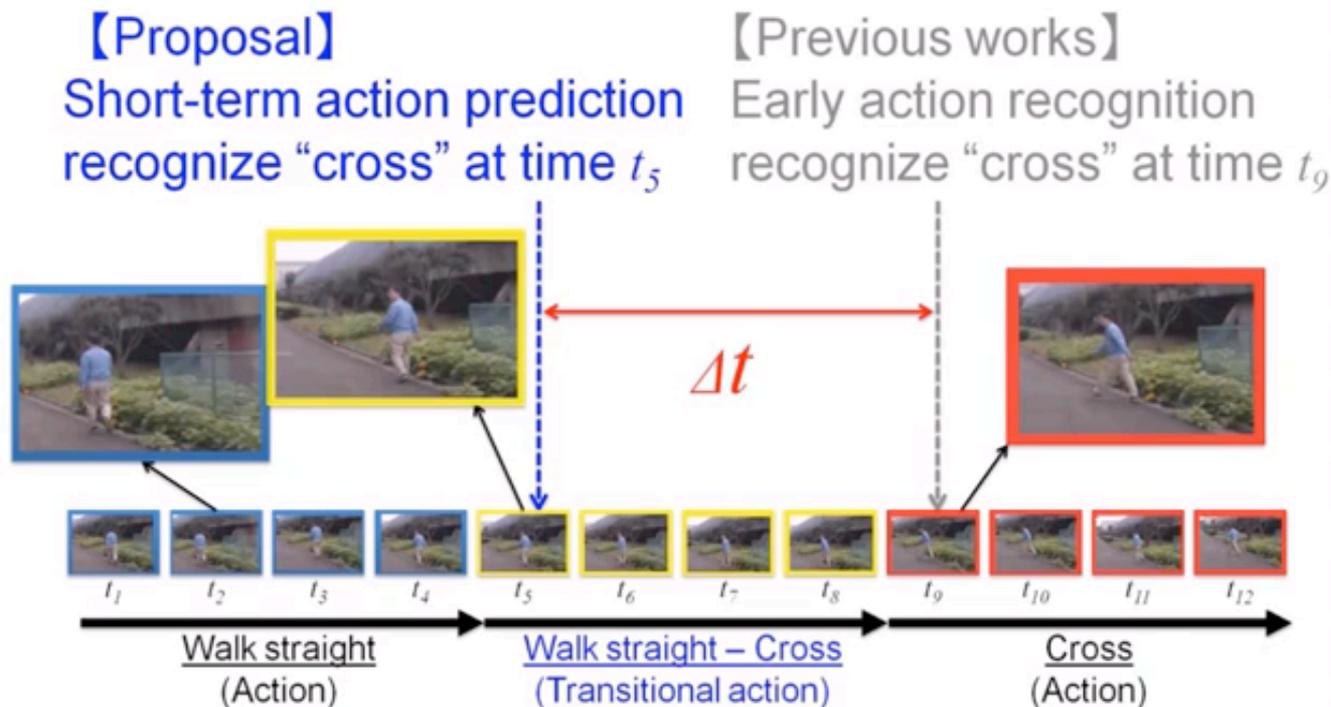
cvpaper.challengeの目指すもの

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

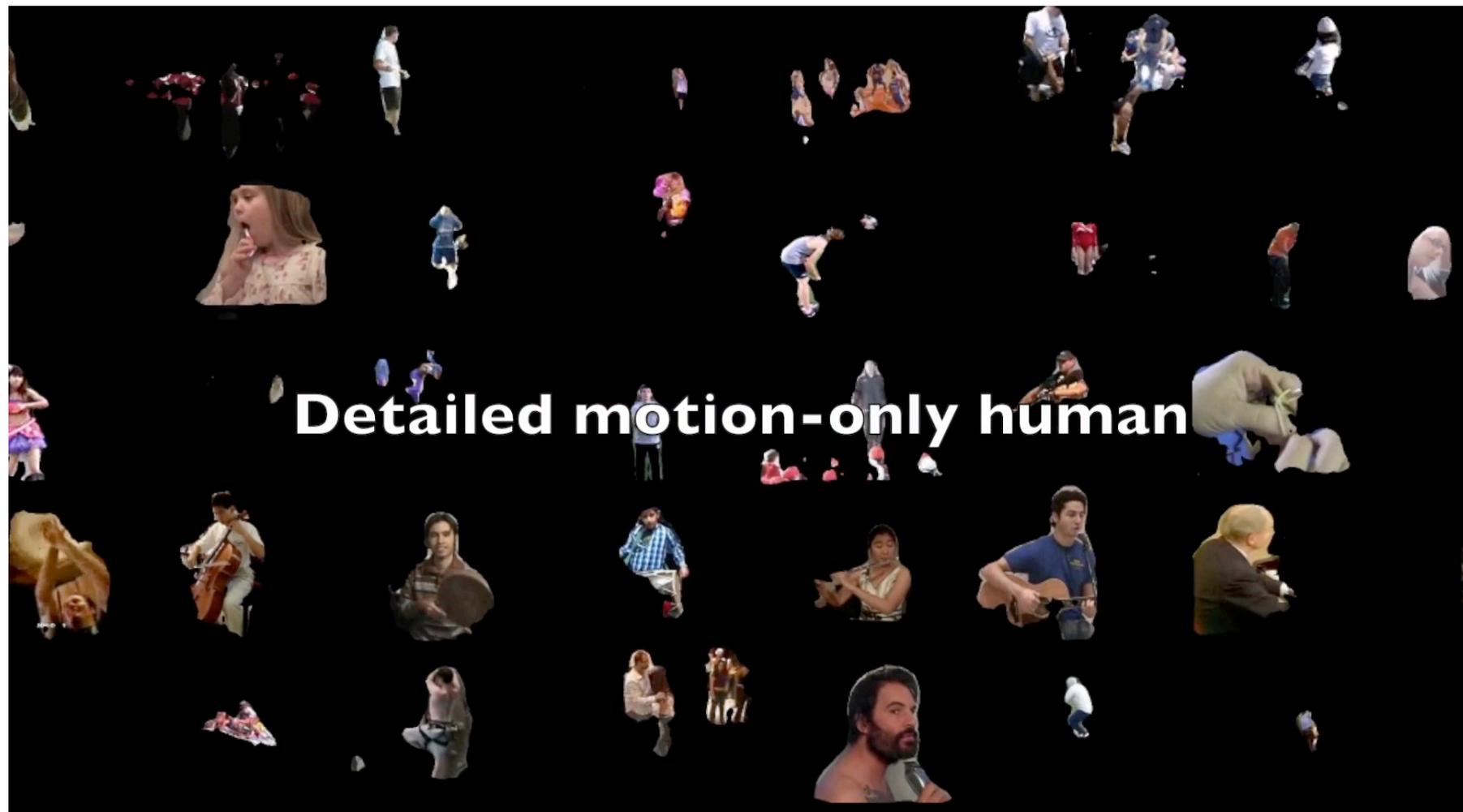
遷移状態を把握して行動予測



The cross action can be predicted

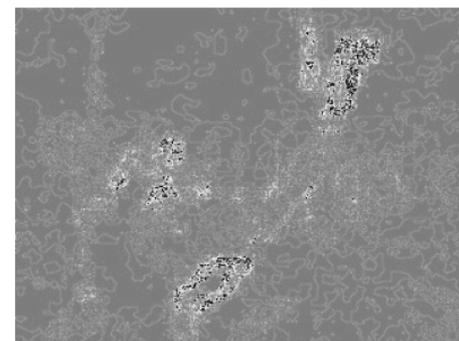
- at time t_5 (our proposal)
- at time t_9 (previous works)

背景領域が人物行動認識に与える影響を評価



加速度を2D画像に投影して特徴抽出

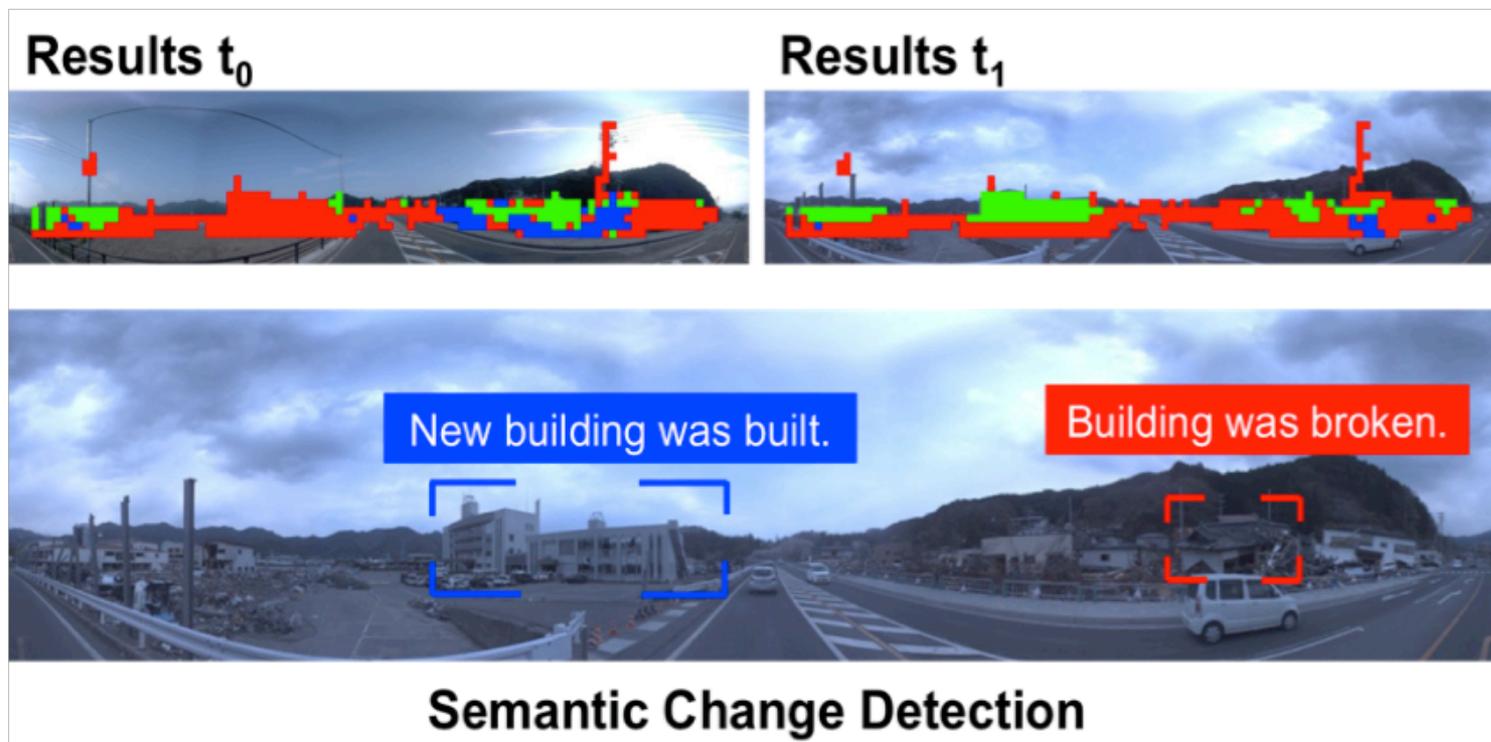
- 位置 (RGB), 速度 (Flow)に対して加速度を追加
- 特徴抽出はDNNに任せる



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

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



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?

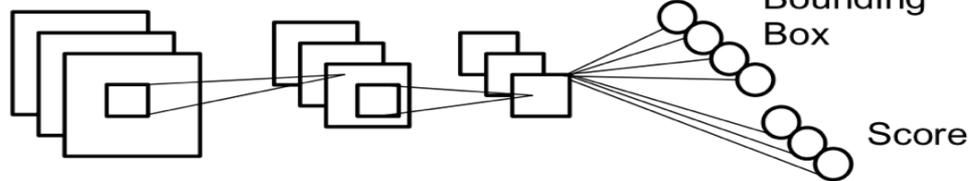


歩行者のニアミス状態を検出

- 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



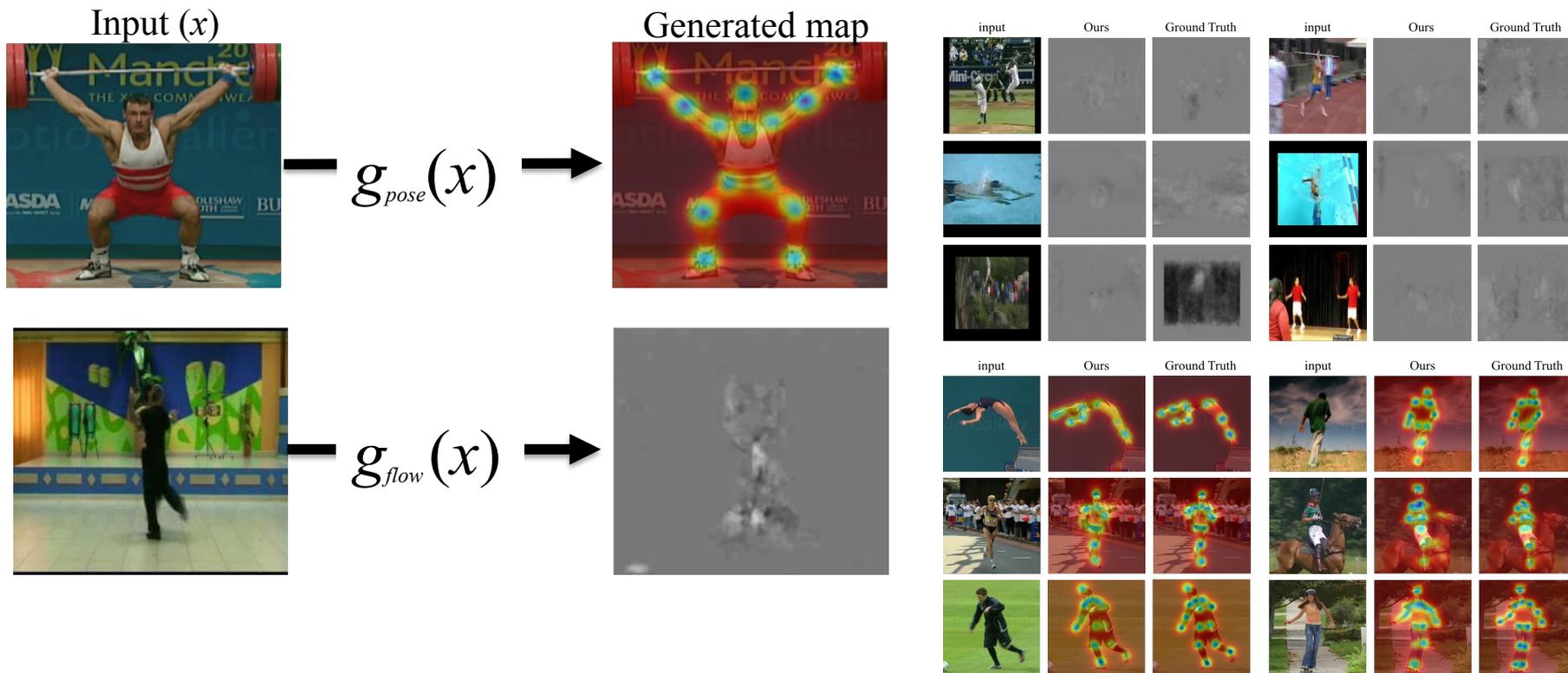
Risk high



Risk low

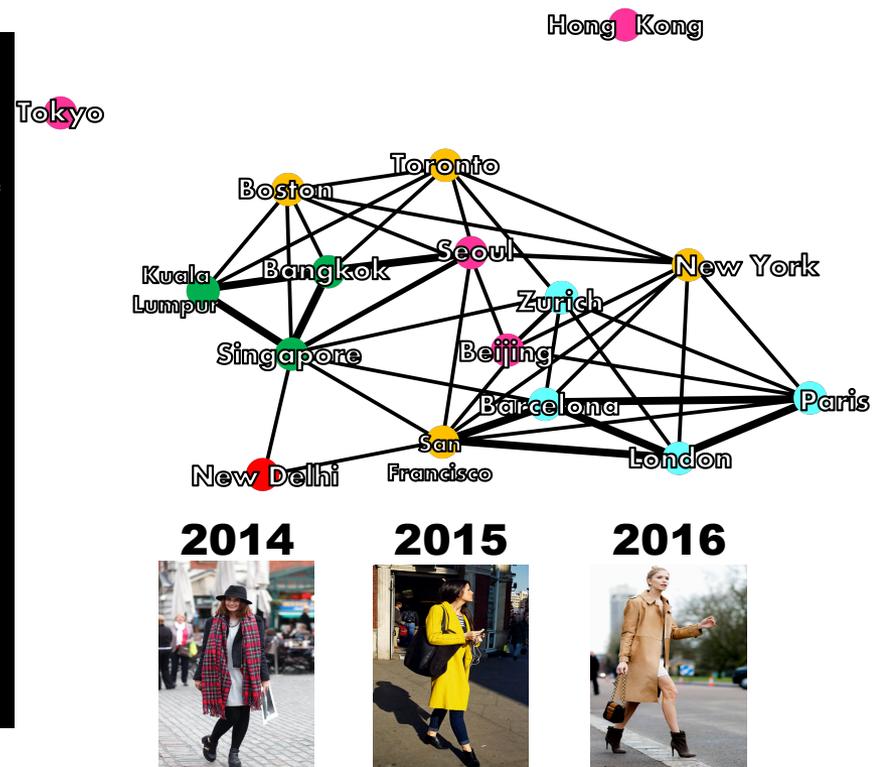
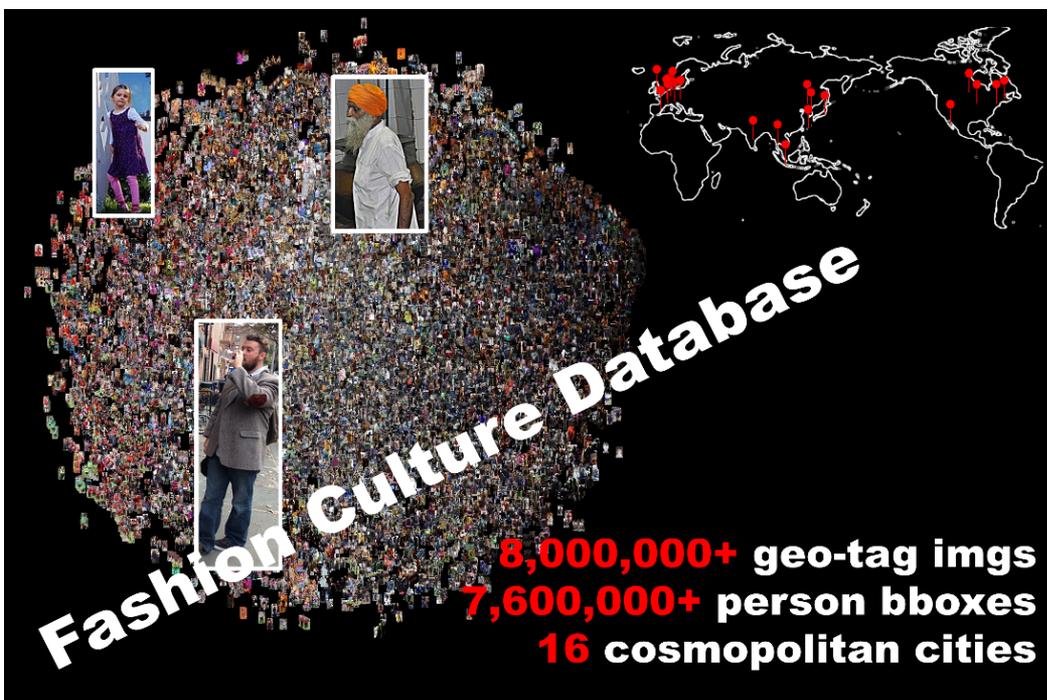
入力から直接 動作解析に必要なマップに変換

- 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



大規模DB (76M) からファッショントレンドを発見

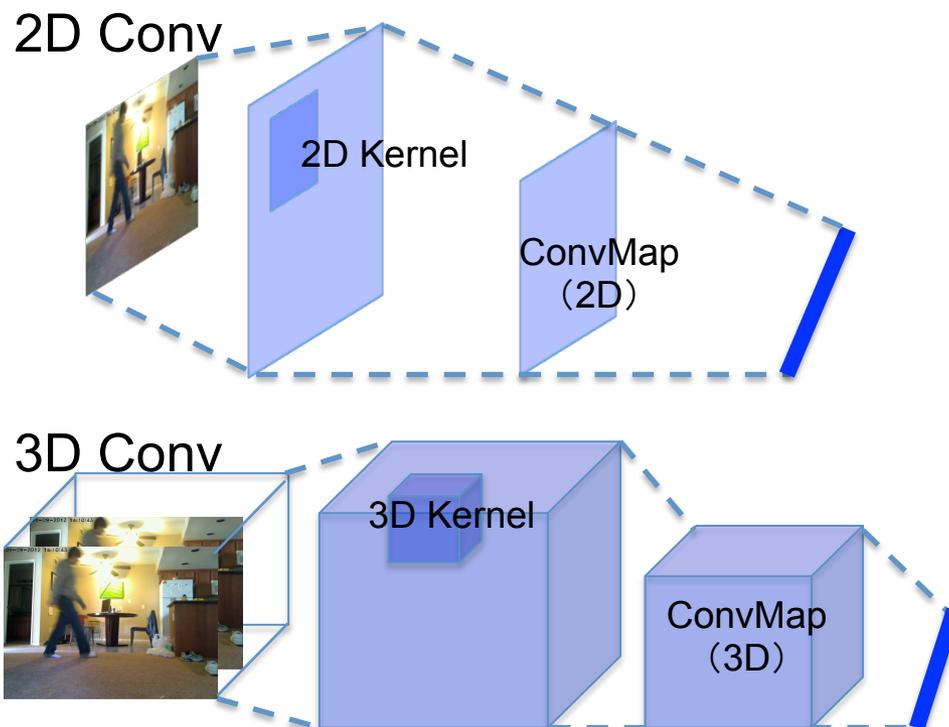
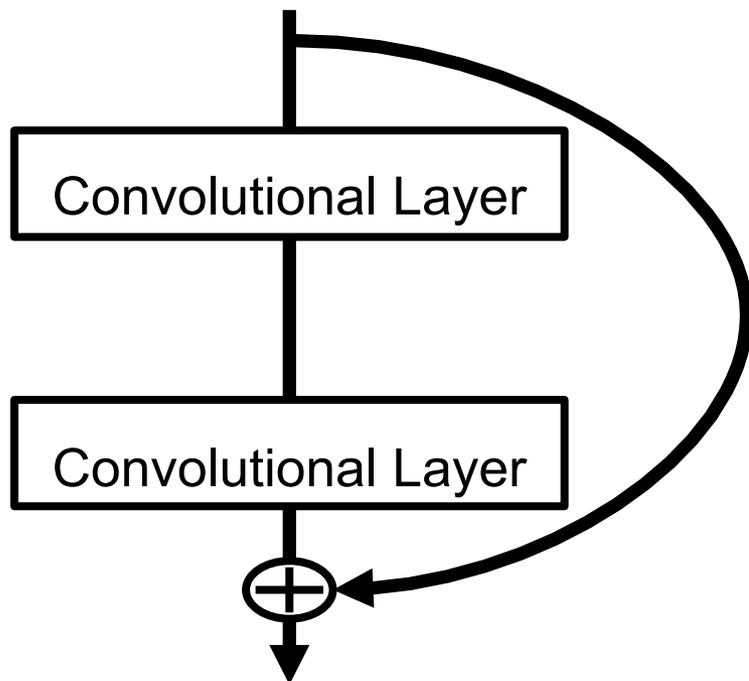
- 動的に変化するファッション (服装) を捉える
- 世界主要都市の動向をSNS画像から推定



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>



コンピュータビジョンの現状

独創的な問題 and/or 強い手法がないと記録・
記憶に残らない

- トップ会議の論文として例外でない (主要2会議で1400本)
- 中途半端に分割した複数論文よりもパーフェクトな1本
- 動画やスライド公開・コード共有・DBリリースなども (できる限り) 徹底して揃える

他の追隨を許さぬ強い手法を作る！

独創的な問題！強い手法！

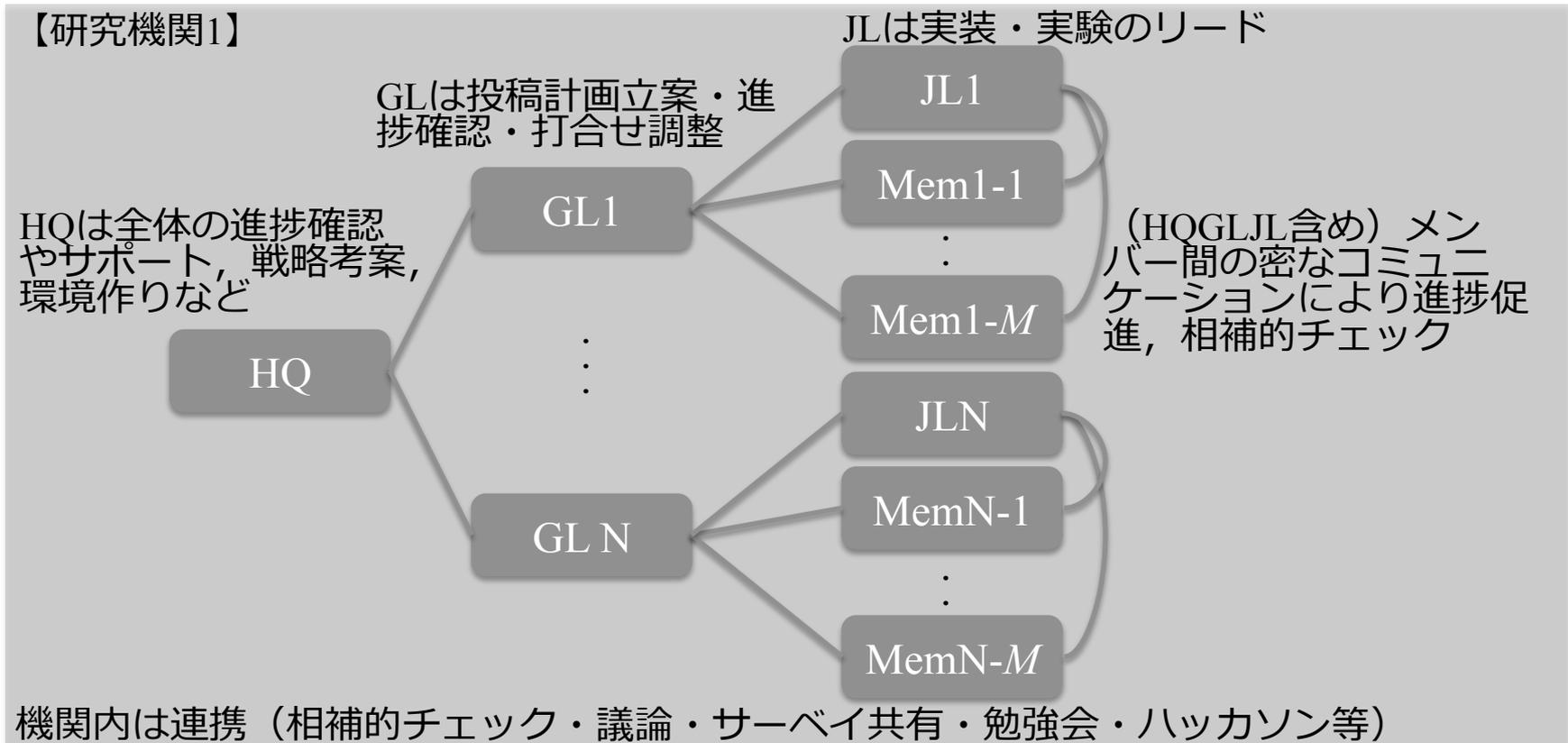
- 受賞論文や注目論文にあるように「極めて高速かつ高精度」を実現, コードをリリースして分野に貢献が理想
- 問題設定に対しての強い手法でも構わない
- 作れたらいいですね！ => e.g. DenseNets, YOLO9000, OpenPose, PSPNet...

研究体制への考え方

今まで以上にチームの力が重要

- 高品質論文 (前スライドより) には1人のパワーでは不十分？
- 当グループでは仕組みを再考
 - 通常の学生：1人1テーマ3年間継続 (学部~修士を想定)
 - cvpaper.challenge：2~4人1テーマ1年でテーマ拡張/変更

どうやって研究しているか？



- 知識：論文・Web・講演資料・サーベイテンプレート等から把握
- テーマ：メンバー間の議論により洗練
- 進捗：打ち合わせにより把握
- コミュニケーション：ITツールで補助 (e.g. Slack, Twitter, FB, LINE)
- 論文：多重チェックが重要, 即ち早め (テーマ立案・実験と同時) に書く

研究方針

CV系に限らず他分野へ投稿する

- （当然だが）CVに拘らない
- 分野の成熟は他の分野の手助けになるということ

外の世界に出て行こう！

研究方針

ロストアイデアを復活できないか

- (一時的に)消されたアイデアを復活
- 可能性があるにも関わらず, 消えたアイデアという意味
- もちろん, 今の技術を効果的にアレンジ

コンピュータビジョンの今を映し未来を創るチャレンジ,
一緒に挑戦しませんか？



<https://sites.google.com/site/cvpaperchallenge/recruit>