CVPR2017/ICCV2017から見た研究動向

@名古屋CV・PRML勉強会

片岡 裕雄

http://hirokatsukataoka.net/

この２年半に渡り，
チーム体制をとって
サーベイや研究活動を実施

2,843

169,390

17

cvpaper.challengeの２年半を簡単に

- 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）

情報量が加速！

783 papers!

621 papers!

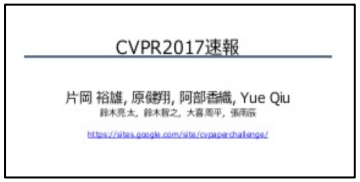
出てくる情報量から，一人の限界を感じた。。

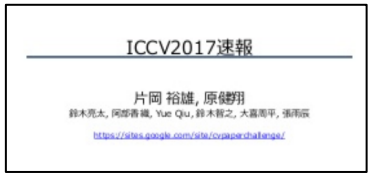
チーム力の重要性

捉えられる情報量も加速させるべき！

– サーベイ量はもちろん，ディスカッション，方針，資料の質に至るまで個人よりもチーム

現在までに参加した会議録





CVPR 2016 速報		CVPR 2017 速報	ICCV 2017 速報
6,284 views		19,883 views	7,354 views
17 likes	<	290 likes	24 likes
57 DLs		413 DLs	146 DLs
16 months		4 months	1 months

単独で書いた結果

チーム力の結集

本資料は，CVPR2017 / ICCV2017の速報をベースにしています

その前に（かなり大雑把な）DNNの概要

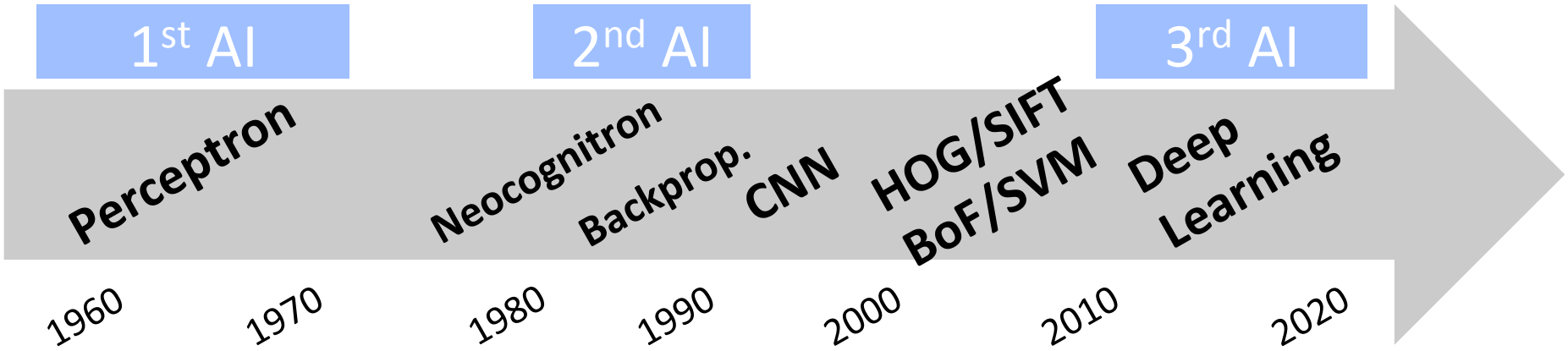
かなり大雑把なDNNの紹介

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

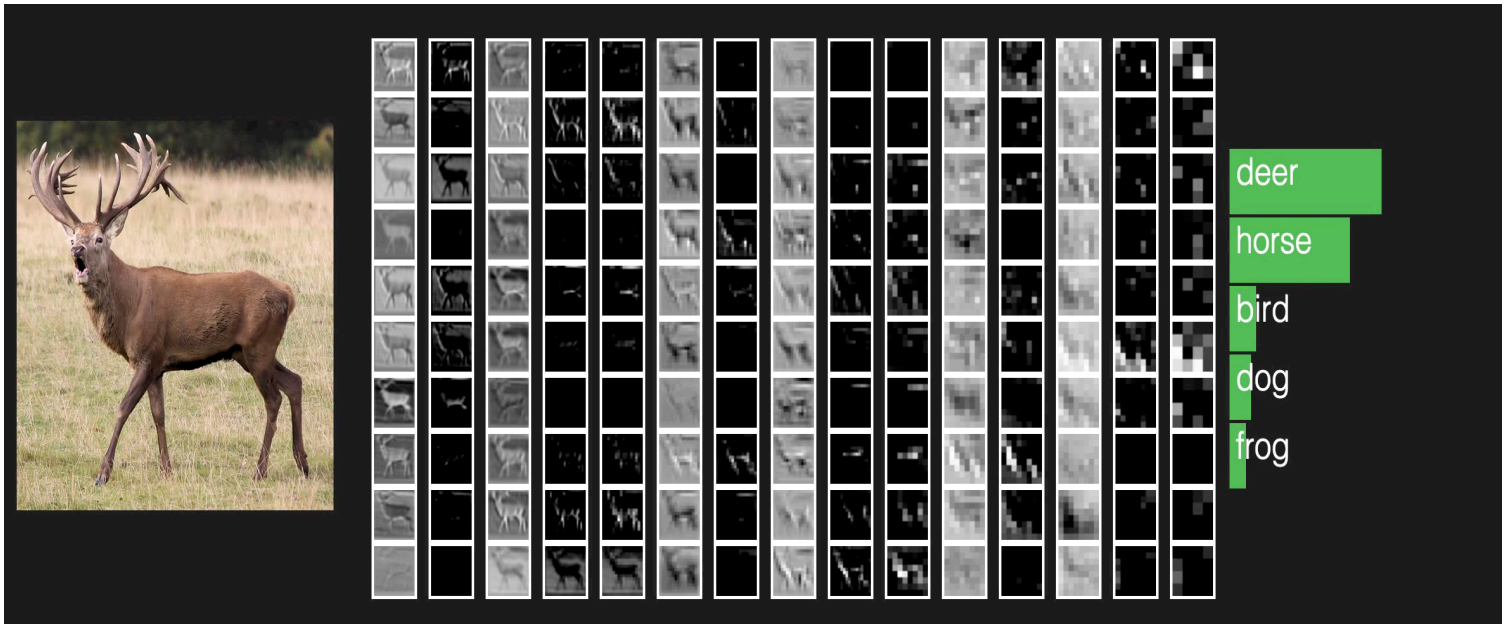


F. Rosenblatt et al. “Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms” in 1961.
Rumelhart et al. “Learning representations by back-propagating errors” in Nature 1986.
K. Fukushima, “Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition 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



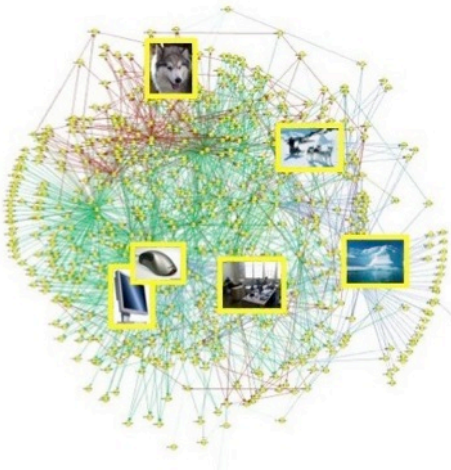
https://www.youtube.com/watch?v=zj_JlVqWK1M

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）



IMAGENET

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

×

NVIDIA®

http://cvpr2017.thecvf.com/

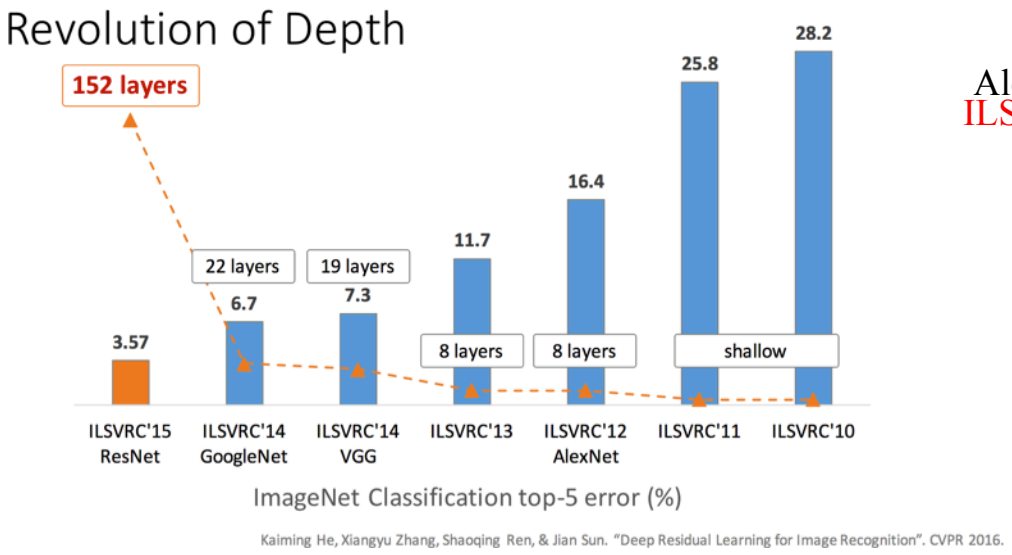
DNNの動向（4/8）

構造の深化

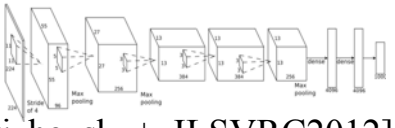
– 2014年頃から「構造をより深くする」ための知見が整う

– 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]

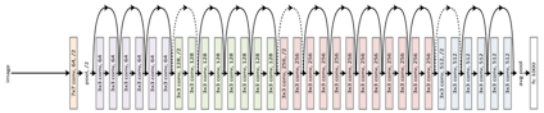
ILSVRC2012 winner，DLの火付け役

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）

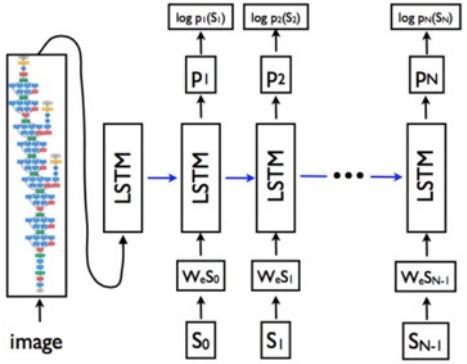
他タスクへの応用（画像認識・動画認識）

- 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

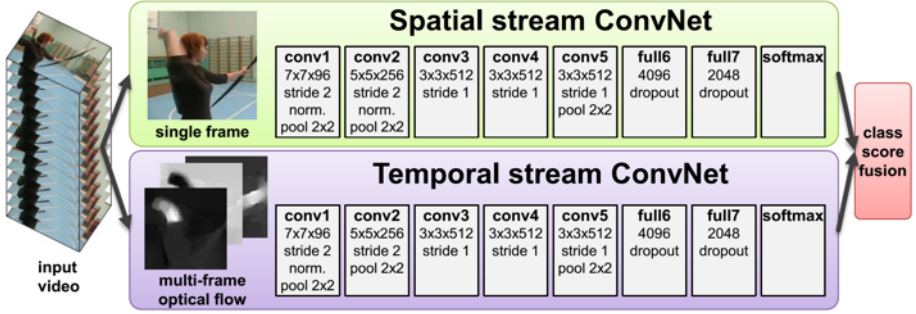
ResNet + FRCNN
[He+, CVPR16]

https://
www.youtube.com/
watch?
v=WZmSMkK9VuA

PSPNet [CVPR17]



Show and Tell [Vinyals+, CVPR15]



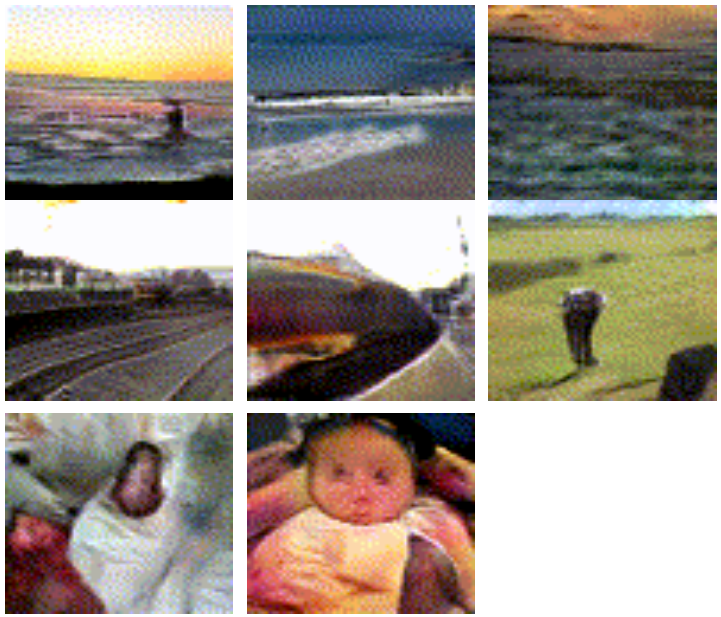
Two-Stream CNN [Simonyan+, NIPS14]

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

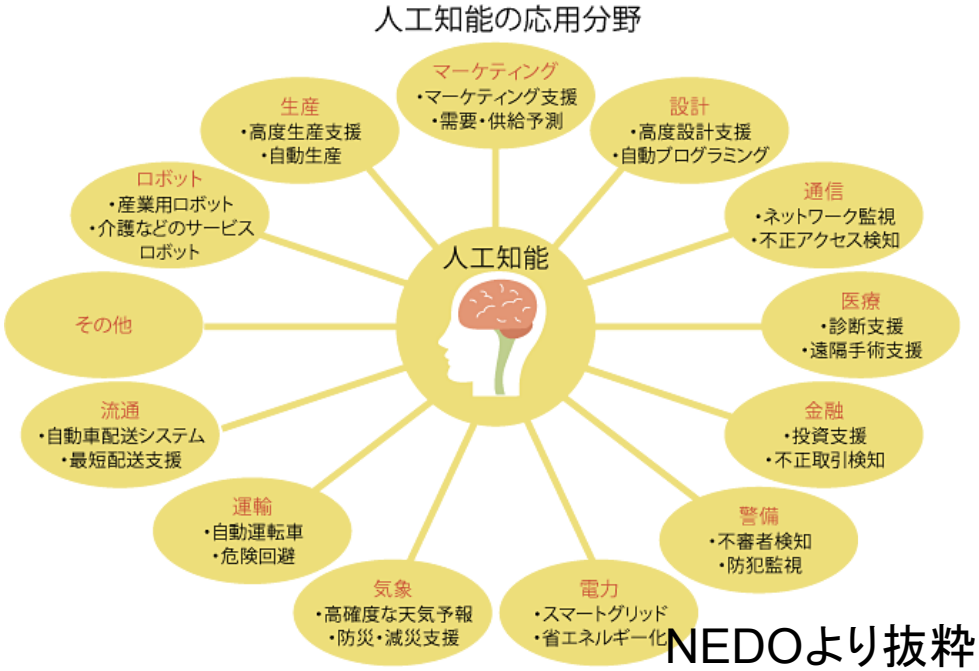
http://pytorch.org/docs/master/_static/pytorch-logo-dark.svg

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

CVの現状

CVPR/ICCVのみならずCV全体の現状も含めて

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：ICCV2017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

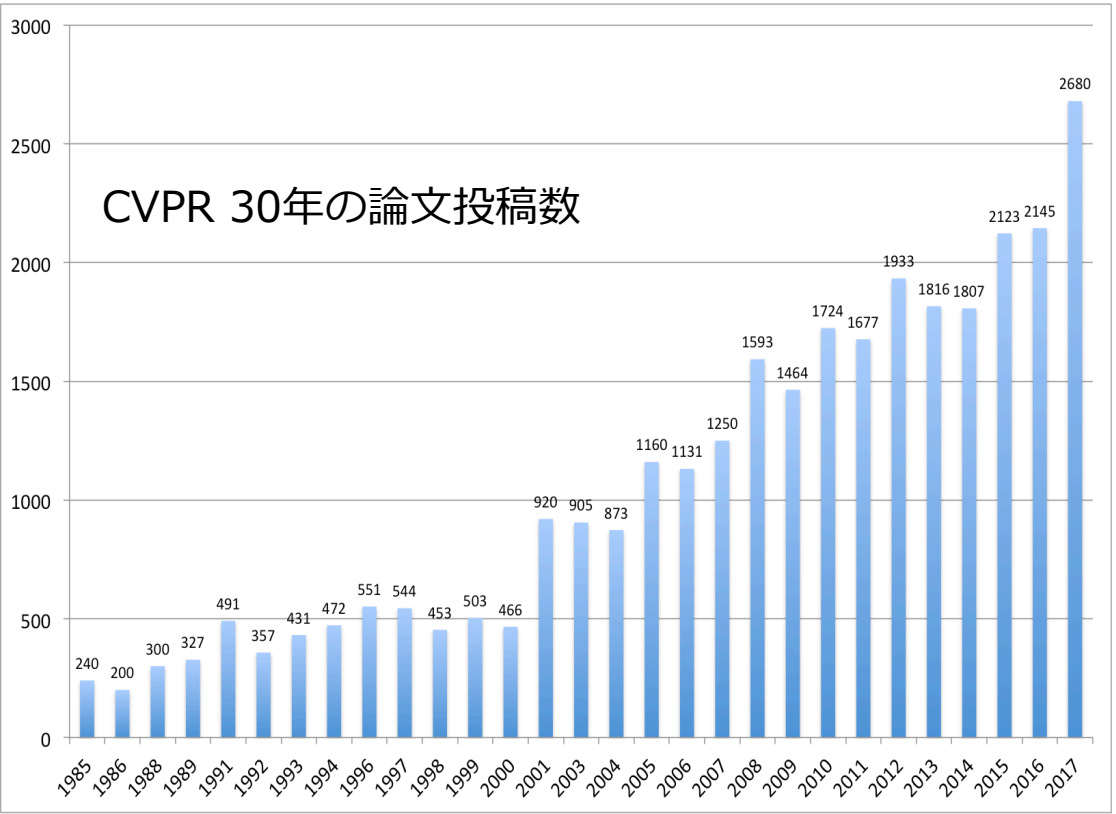
- Google Scholar (CV&PR)の統計にてCVPRに次ぐ国際会議
- 採択率は25~30%
- 2017年はイタリア・ベネチアにて開催

http://mytabi-italy.com/venice-lido/

CVの現状：会議の統計

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@CVPR17，2,143論文@ICCV17を記録
- 参加者数 約5,000人（CVPR17）約3,000人（ICCV17）
- スポンサー数127

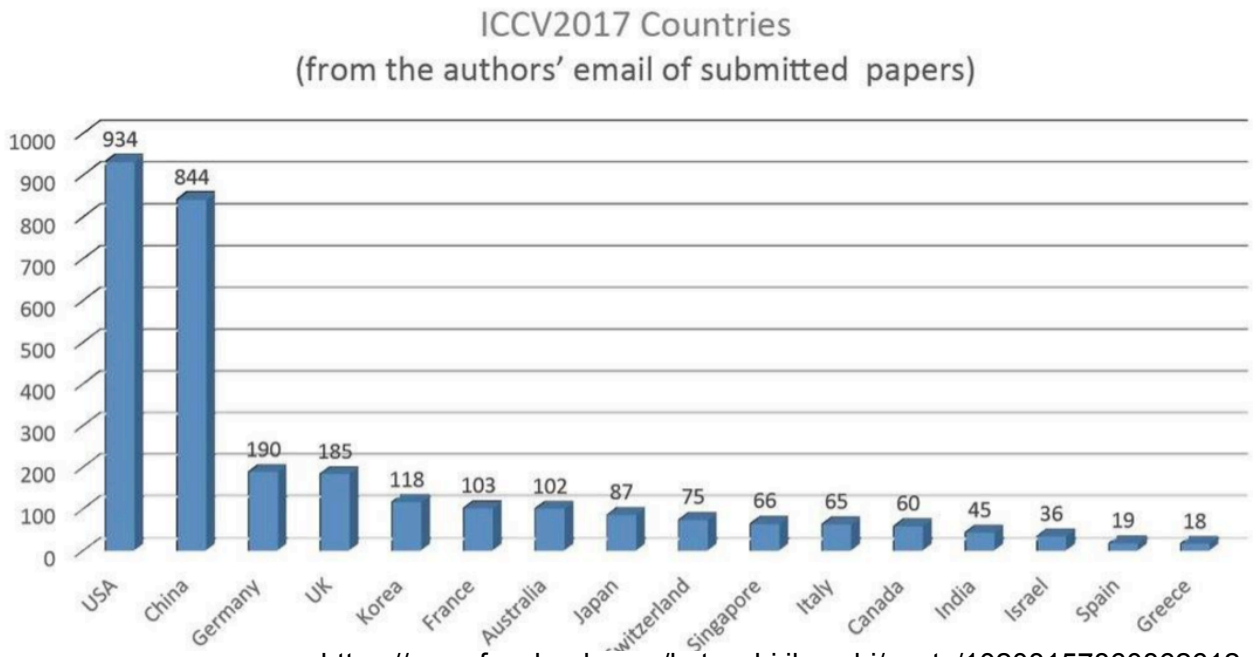


http://img.mp.itc.cn/upload/20170727/a94c7cfda6c34d2fb3f52825c3c6b928_th.jpg

CVの現状：会議の統計

中国の勢いがすごい！

- 今に始まった話ではないが，やはりすごい（下図USAにも中国勢がいるはずなので実質的にはトップ？）
- 人口が多い以上に，勢いがある
- 「とにかくみんな論文出そう」とか「研究資金，人材，データなどを集めて戦えるような体勢に」など見習わないといけない



https://www.facebook.com/katsushi.ikeuchi/posts/10208157860862612

arXiv.org

30 - 50/day @CS:CV

7,500+ papers/year？

https://upload.wikimedia.org/wikipedia/commons/thumb/a/a8/ArXiv_web.svg/2000px-ArXiv_web.svg.png

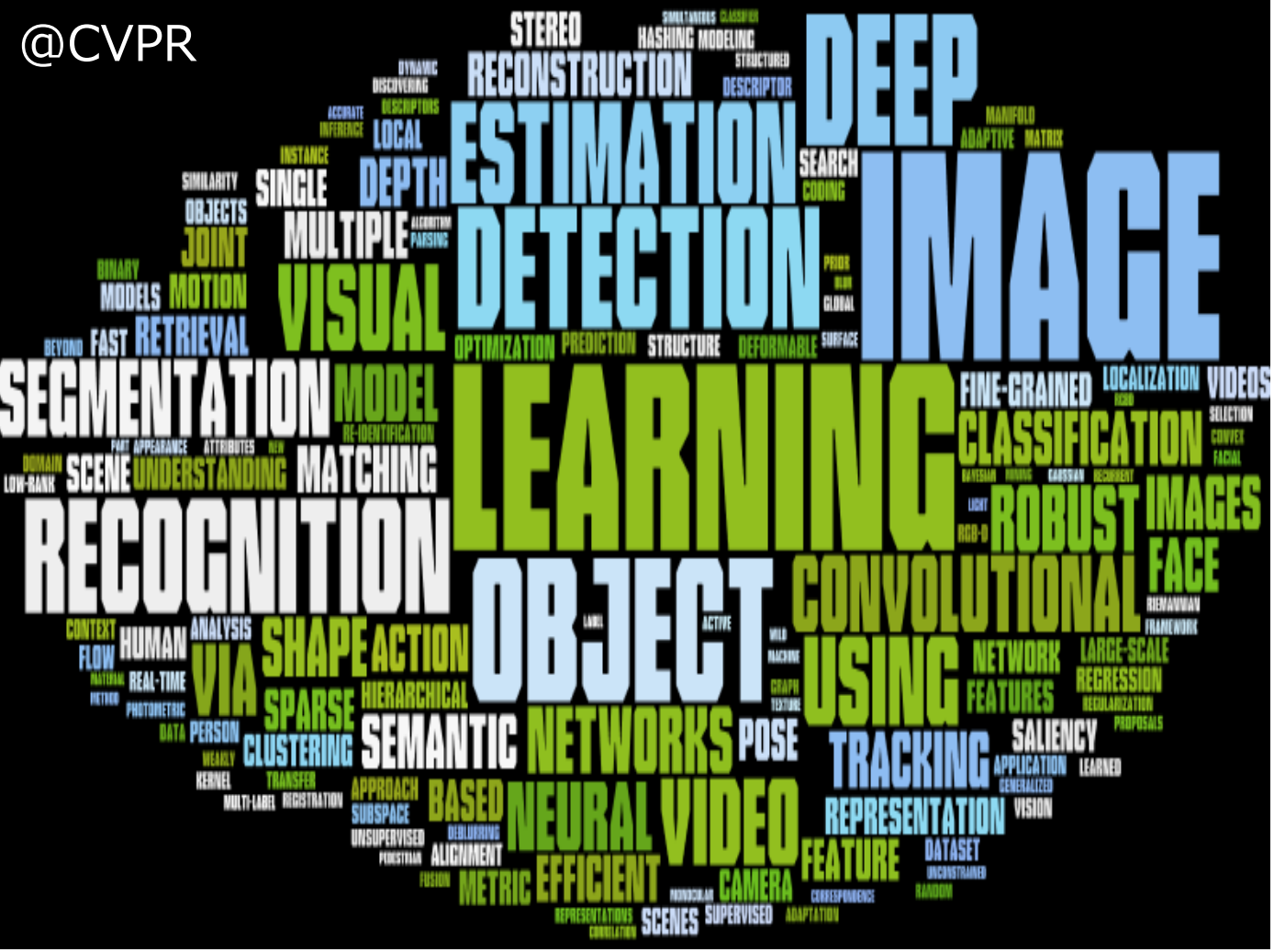
学会	速報
密な議論、人脈	とにかく速い （信憑性はコードで測定）
学会は査読を突破した という承認	認められてから査読を突破するルート?
ライブに行く感覚	速報はライブの広告・宣伝

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！

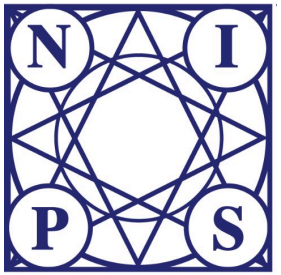
参考：北本朝展先生（NII）「ディープラーニングとオープンサイエンス」
http://agora.ex.nii.ac.jp/~kitamoto/research/publications/sparc16-ppt.pdf

@CVPR



CVの現状：シームレス化

http://labs.criteo.com/wp-content/uploads/2017/09/ICMLlogo17.jpg



https://pbs.twimg.com/profile_images/785902424242348032/94Wht4ER.jpg

https://pbs.twimg.com/profile_images/832295325469536257/Ooqjbgzc.jpg

成熟したCV技術のアドバンテージを享受出来る数少ないチャンス？

https://pbs.twimg.com/profile_images/756021057522311168/LUW1ha4-.jpg

http://iv2017.org/wp-content/uploads/2016/11/Ska%CC%88rmavbild-2016-11-27-kl.-15.42.40.png

https://web.fe.up.pt/~ieeeitsc2016/wp-content/uploads/2016/01/logotipo_small.png

https://web.fe.up.pt/~ieeeitsc2016/wp-content/uploads/2016/01/logotipo_small.png

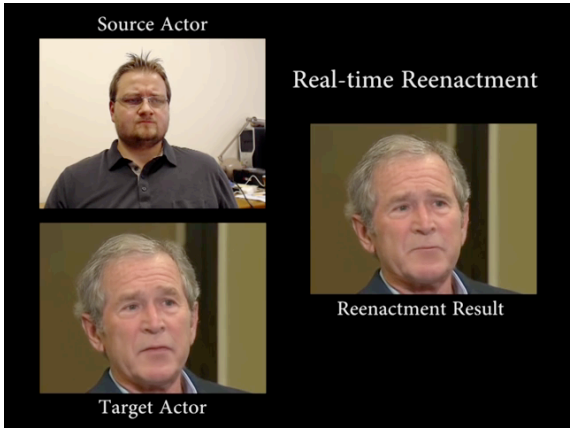
CVの現状: こんなことができている

YOLO_v2（物体検出）

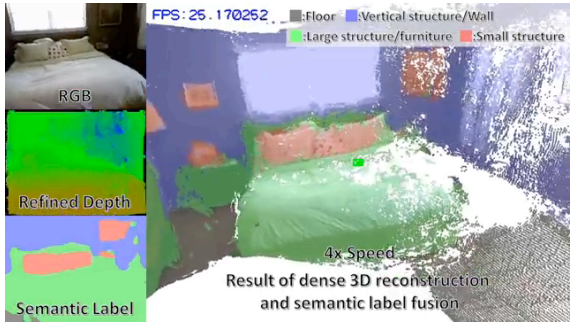
PSPNet（意味ラベル）

OpenPose（姿勢推定）

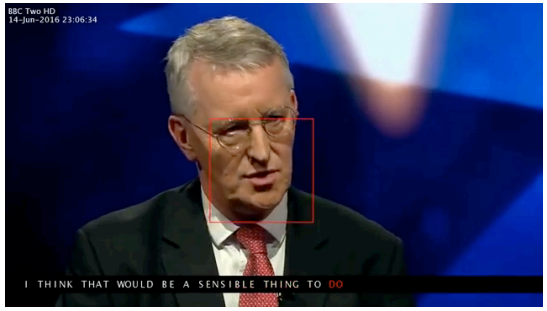
Face2Face（顔器官追跡）



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？

CVPR/ICCVの動向・気付き

CVPR/ICCV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

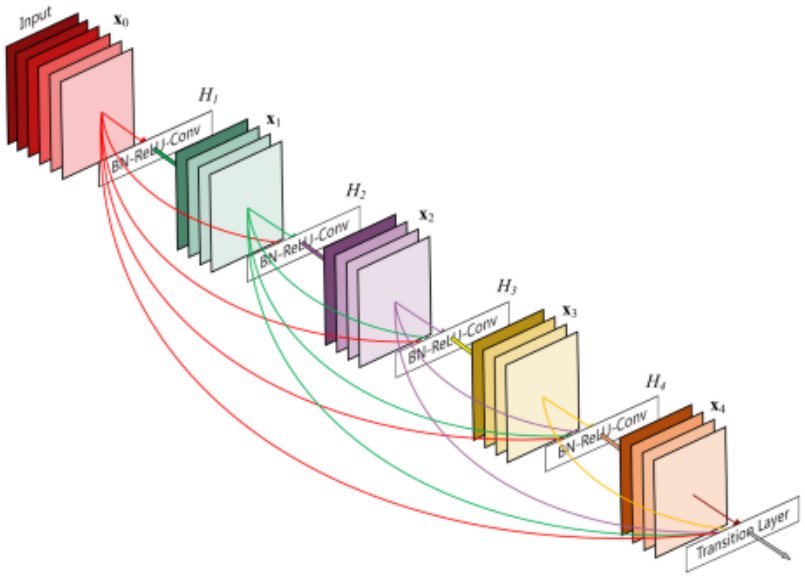
CVPR/ICCVの動向・気付き

- 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

CVPR/ICCV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

CVPR/ICCV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

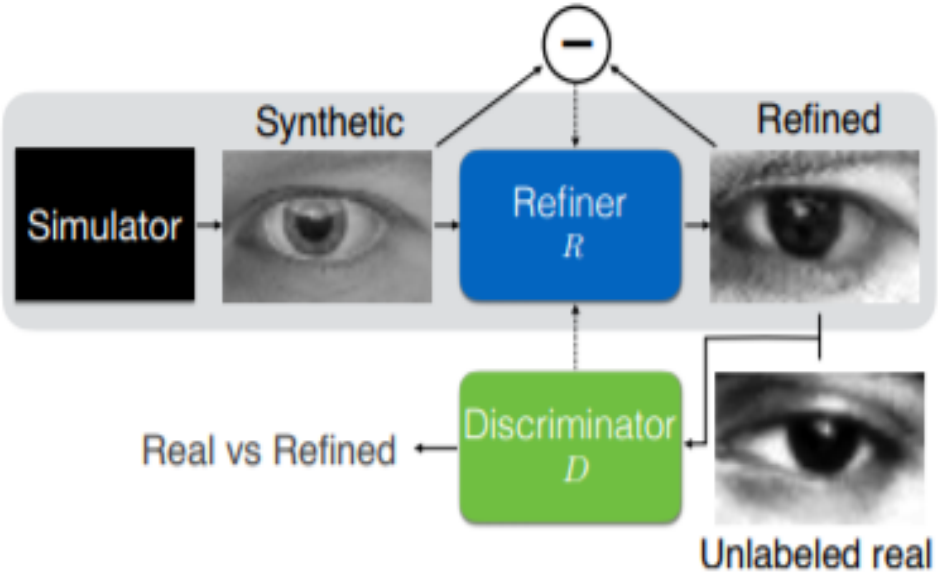
CVPR/ICCVの動向・気付き

– 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

CVPR/ICCV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？

CVPR/ICCVの動向・気付き

- 3D Convolution（xyt/xyz/xy+α）が加速しそう
  - もともとはパラメータが多く小規模データでは過学習気味
  - データスケール/コードレベルで枠組みが整い，誰でも再現可
- 論文例
  - Pseudo-3D Conv
  - 3D-ResNets（コードあり）
    - https://github.com/kenshohara/3D-ResNets-PyTorch

CVPR/ICCVの動向・気付き

- 高次なラベルを返却する手法が登場
  - 識別（e.g. 犬）に対して意味（e.g. 危険，異常）に踏み込む
  - 人間の直感による正解ラベルが必要
- クラウドワーカーの力が必須？
  - 人間の直感，さらにはクラウドワーカーの集合知を結集
  - 手元では到底処理しきれないレベルのアノテーションを解決
  - 同時にN人に発注，クロスチェックによりラベルの質を保証

CVPR/ICCVの動向・気付き

- 合成データの重要性が説かれる
  - （時間さえかければ）大規模データを生成できる
  - データバリエーションを用意に拡張可能（スケール/ビューポイント/ノイズ/オクルージョン/背景など）
  - アノテーションの質はコントロールできる
- 特に人物系のデータの場合，プライバシーを保証
  - 昨今では個人情報保護法など法令の遵守が重要
  - 「存在しないと思われる人」のデータを合成する
  - ファッション/人物行動など特に人の研究は合成データが必要

CVPR/ICCVの動向・気付き

- 特徴表現学習
  - 教師なしで特徴表現を学習してしまう
  - ImageNetやPlacesを置き換えるような学習の仕組みが今後登場？
  - 上記２つのデータセットは画像識別用であるが，問題に合わせて特徴表現を学習できたら良い

CVPR/ICCVの動向・気付き

- 徹底比較により研究分野を整理する
  - DNN研究が「こうしたら良くなった」という実験的向上の繰返し
  - 誰かが体型的にまとめたマニュアル的論文が重宝
- 新しい論文の書き方？
  - （会議の性質にもよるが）Top-tierの会議に比較論文が通ることは稀であった
  - 成功事例集/失敗事例集をまとめた論文は新しい知見を与え，分野の加速に寄与することから採択している（と解釈できる）

CVPR/ICCVの動向・気付き

- 画像生成ではなく，より高精度にするためにGANを使用
  - 戦いの中で強くなる識別器を提案
  - 画像生成（ラベル付きデータ数）と画像識別器（生成を高度にしながら自ら賢くなる）のせめぎ合いで精度向上
- タスクの複雑化
  - 画像識別 => 物体検出 => セマンティックセグメンテーション
  - 動画，RGB-D，3Dポイントクラウド?

CVPR/ICCVの動向・気付き

- 説明性を与えるDNNモデルや構造改善
  - 精度が出ていても「なぜ？」の部分ができていないとアプリケーションにはできない
  - 活性部分を表示（Grad-CAM），視覚的理由付け（Visual Reasoning）

CVPR/ICCVの動向・気付き

- ImageNetの学習済みモデルに頼るのはそろそろ限界？
  - タスクが複雑になり画像識別から乖離しているにもかかわらずImageNetの特徴に頼るのはナンセンス！（と偉い人が言っていた）
  - タスク別の洗練された特徴が必要だが，データを大量に収集，かつリッチなラベルを付与するのは限界がある => 教師なしで実世界の特徴を獲得する手法が必要？
  - もちろん，タスクによってはImageNetは多大なる効果を発揮する

CVPR/ICCVの動向・気付き

- 予測という高次な情報を推定
  - 時間 t の状態が高度に観測できるようになったので t+n の状態を予測する
  - 少し調べただけでも下記のような論文が見つかる
    - Predicting Human Activities Using Stochastic Grammar
    - First-Person Activity Forecasting With Online Inverse Reinforcement Learning
    - Visual Forecasting by Imitating Dynamics in Natural Sequences
    - Fashion Forward: Forecasting Visual Style in Fashion（ファッションスタイルの予測）
    - The Pose Knows: Video Forecasting by Generating Pose Futures
    - What Will Happen Next? Forecasting Player Moves in Sports Videos
    - Encouraging LSTMs to Anticipate Actions Very Early
    - Anticipating Daily Intention Using On-Wrist Motion Triggered Sensing

CVPR/ICCVの動向・気付き

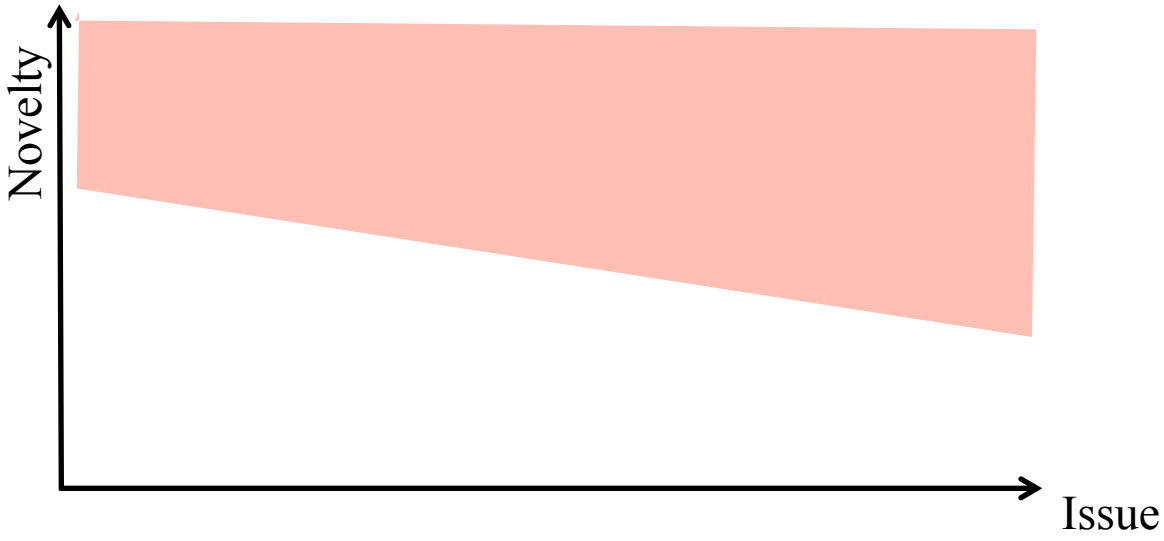
- 技術が成熟してきたことで産業応用が加速する！
- 例：葉っぱセグメンテーション/カウントワークショップ
  - Link: https://www.plant-phenotyping.org/CVPPP2017
- 日本は産業に力を入れている印象であり，より産業応用に近い研究でプレゼンスを取るべき？
- SSII/ViEWなどはまさにそのフィールド
  - SSII: https://confit.atlas.jp/guide/event/ssii2017/top
  - ViEW: http://www.tc-iaip.org/view2017/

CVPR/ICCVの動向・気付き

- アプリケーション寄りの研究が増加
  - Vision for Xというセッションが存在
    - Fashion Concept Discoveryや
      Predicting Privacy Risksなど内容は様々
  - それ以外のセッションでもファッション関連の研究など
    出口が近い研究がいくつも見られる
    - 出口が近くなると評価の仕方も変わってくる
    - 単純な教師付き学習での評価尺度から離れていくことも
      今後進んでいく？

CVPR/ICCV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みがないと通らない（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



CVPR/ICCV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

CVPR/ICCV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ィ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？

CVPR/ICCV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といて。。）
    - 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

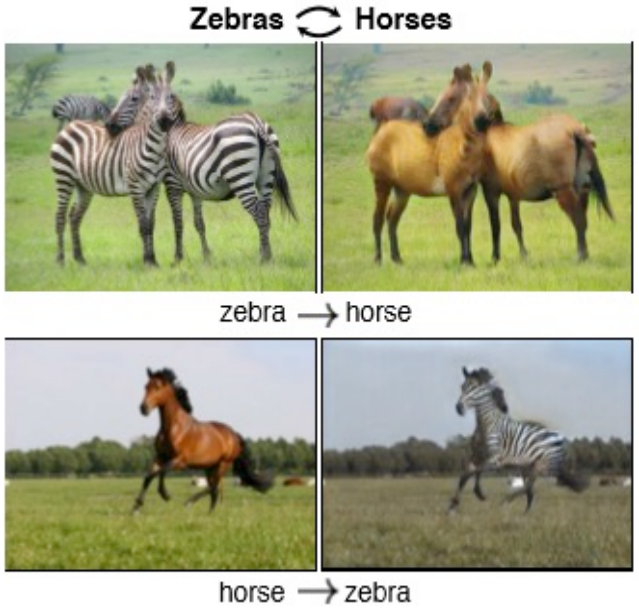
https://pbs.twimg.com/media/DFnnqn0VwAExuiU.jpg

CVPR/ICCVの動向・気付き

– 馬をシマウマに変換しよう！

- CycleGANより
- 「だからどうした！？」だけど、みんなが言うとものすごく大きな宣伝効果になる
- 馬をシマウマに変換する画像はおそらくWeb用

最も効果的なプレゼンを適切なチャンネルで行う



https://junyanz.github.io/CycleGAN/

CVPR/ICCVの動向・気付き

- トップの研究者は難しいことを理路整然と説明
  - 難しい理論もわかった気になってしまう！
  - 論文は難しい，がプレゼンは非常にわかりやすいという人もいる（発表を聞きに行こう！）

CVPR/ICCVの動向・気付き

- 研究者の動きが激しい
  - 特に，企業に移る人が多くなった
  - トップ会議やジャーナルに突破している人も移る（世界的に有名な先生も完全移籍する噂も出るくらい）
  - 海外でも10年後には企業から大学の先生になる人も多くなる？

CVPR/ICCVの動向・気付き

- 日本のプレゼンスは確実に落ちている
  - 池内先生（General Chair）の投稿 からも読める
    - https://www.facebook.com/katsushi.ikeuchi/posts/10208157860862612
  - 参加者・投稿数から減っている
  - 投稿数：ICCVは（特別）難しいという印象がある？（ICCVだっていち国際会議である）
  - 参加数：論文がオンラインで読めるから完結？（原因は他にも）

継続的に参加・投稿して査読を突破しよう！

まずは一人一本，投稿しよう！！

CVPR/ICCVの動向・気付き

- 独国大学の卒業条件
  - 会場で聞いたリアル
  - メジャーな国際会議に２本以上論文を通すこと
    - この分野だとCVPR/ICCV/ECCV/NIPS/ICML等
  - 世界的に知名度の高いジャーナルに１本以上論文を通すこと
    - この分野だとTPAMI/IJCV/TIP/PatternRecognition/CVIU等
  - プレッシャーと戦いながらも楽しそうに研究している！
    - 実際にドロップアウトする人ももちろんいる
  - 人に見てもらえる会議でないと意味がないという考え？
    - 実際に使ってもらえる技術もその中で磨かれる

CVPR/ICCVの動向・気付き

- CVPR2017と比較して「進展した！」というほどではない
  - CVPR2017は7月末，ICCV2017は10月末
    - ３ヶ月しか経ってない
  - CVPRでリジェクトされた論文も相当数ICCVに投稿される？
    - 研究されたのはほとんど2016年？
    - とすると劇的な進展があるのは2018年の会議？

2018年の動向に注目！？（今後進展するか硬直するか）

CVPR/ICCVの動向・気付き

- プロジェクト化しよう
  - 論文を連続的に出せる仕組みを構築
  - グループで研究しよう！
  - 国際会議１：データセットを考案
  - 国際会議２：発展手法を提案
  - 国際会議３：別視点（別問題）で論文を執筆

CVPR/ICCV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？

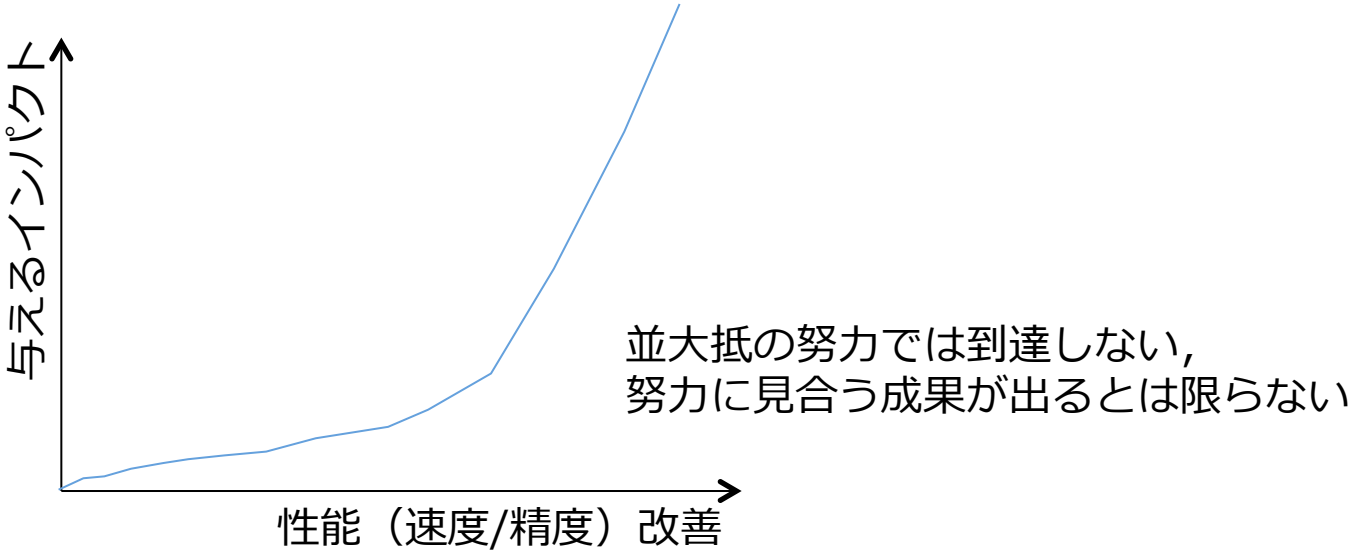
ところで，YOLO9000がでたら
YOLO9001を作ろうと思いますか？

ところで，YOLO9000がでたら
YOLO9001を作ろうと思いますか？

- あまり学術的発展がない

劇的な性能向上は（非常に）評価される

- みんなが使いたい技術で精度/速度ともに劇的な改善
- YOLO（2016）とYOLOv2（2017）はまるで違う
  - 精度：63.4 => 76.8(v2) @PascalVOC07 better
  - 速度：67FPS faster
  - 強さ：9,000 カテゴリ stronger
- AlexNet（2012）
  - 言うまでもなく。。

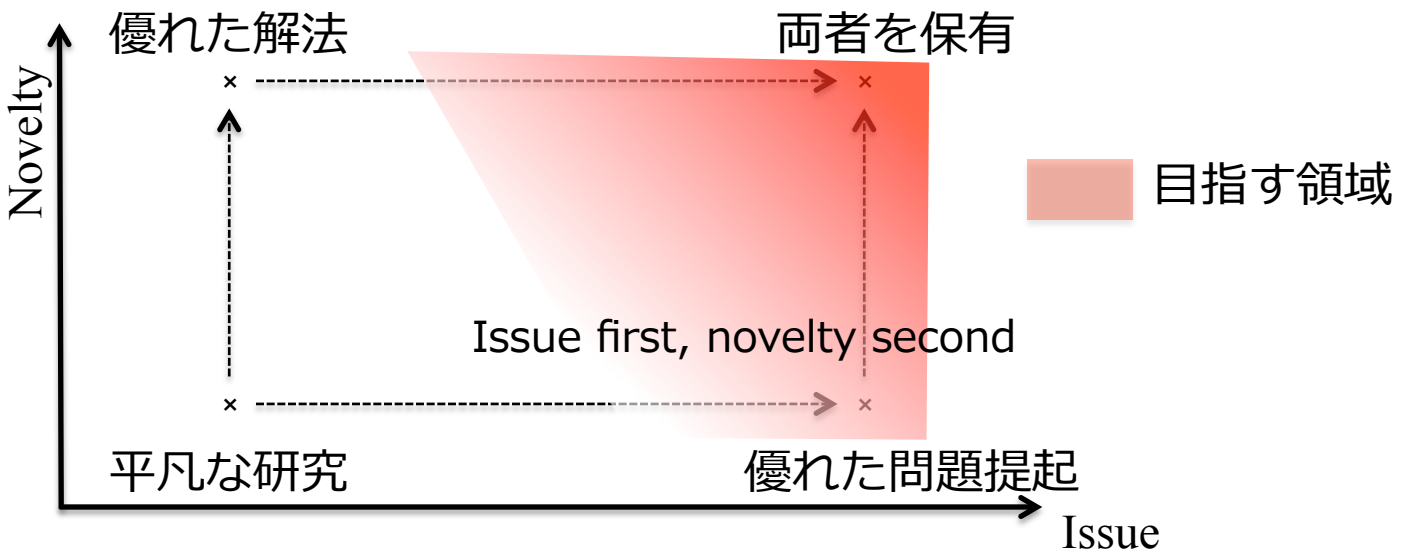


研究に対する意識の転換

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

この２年半，サーベイ・発想・実装法自体を研究



サーベイ・発想・グループ実装

短く紹介

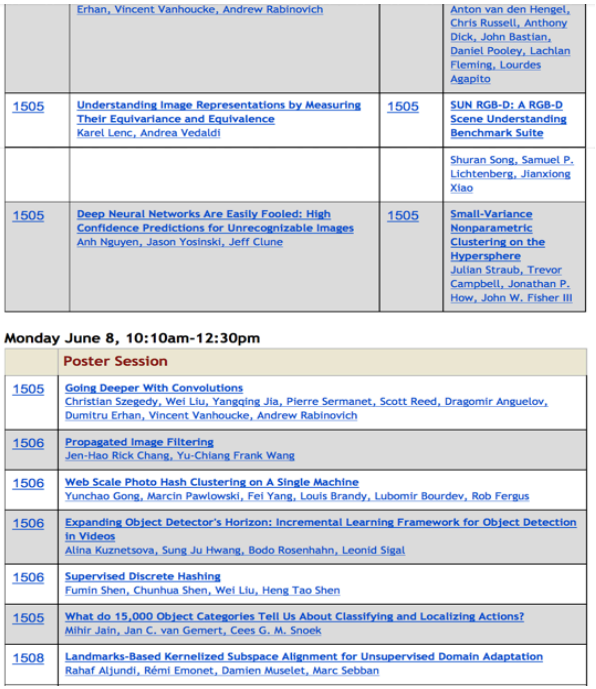
サーベイ

チームの力を駆使して精密かつ網羅的に

- 年間1,000本くらいは読む（さらに資料化する）力が備わった
- サーベイを組織化したら効率が良くなった

「誰もやらない」ことをやるのと効率化したのがよかった？

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

発想

サーベイ～研究テーマの移行を手順化

- キーワードは「集合知」，再帰的な分割と統合
- 考案（個人単位）議論（ブレスト）洗練化（追加調査）のループ
- メンバー全員でアイディアを創り上げる，学生も研究員も無関係

例：Dynamic Fashion Cultures（MIRU17学生賞）

– スタバの自撮り画像を収集すれば意識高い解析？

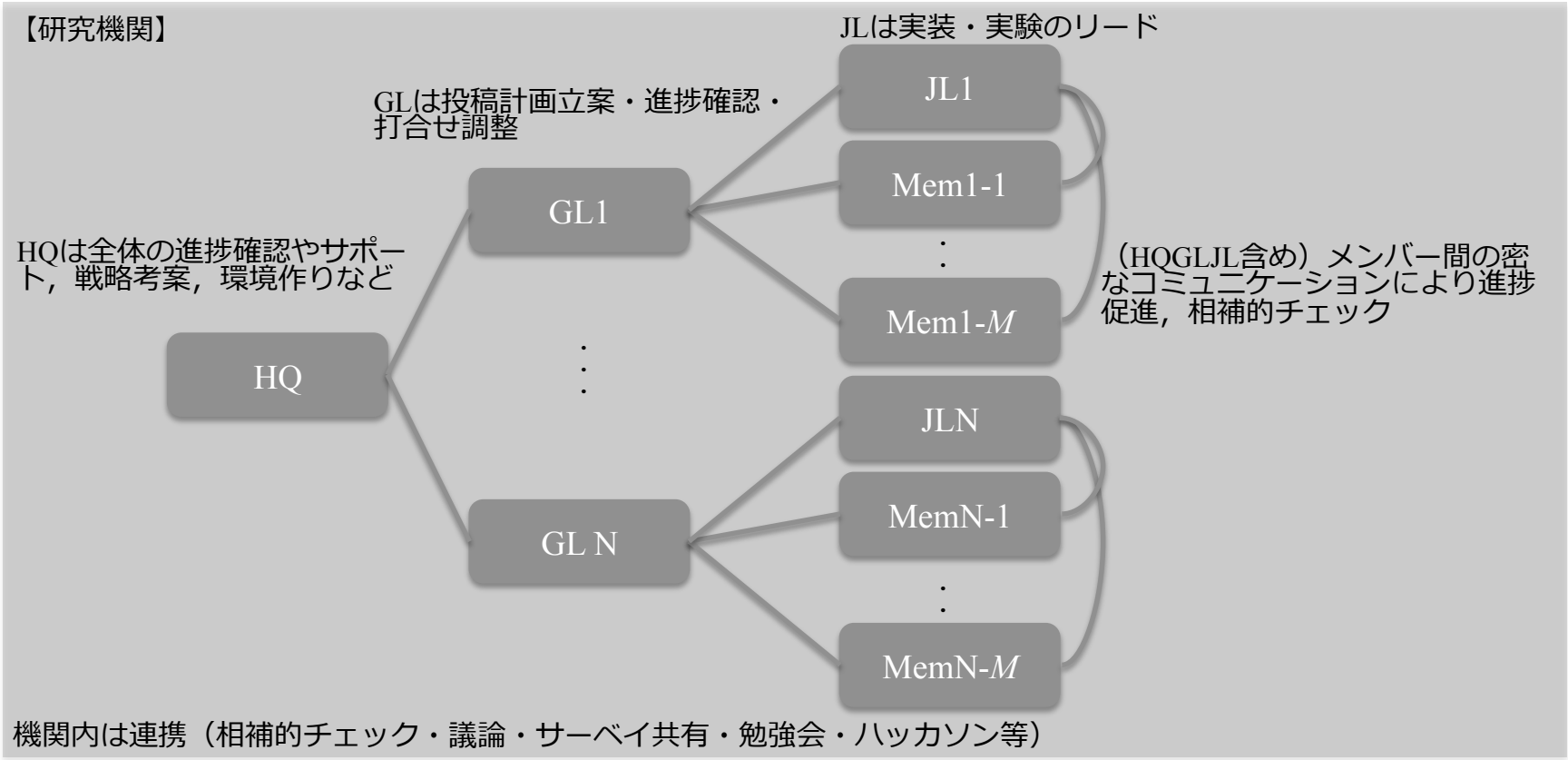
– City Perception（ECCV14）：世界各地の都市解析

・・・

– SNS画像収集して世界の動的ファッション解析！

グループ実装

個人作業と，Grp打ち合わせ/チャットにより進行



- 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

目指す研究姿勢

技術で面白いことをやる・見せる

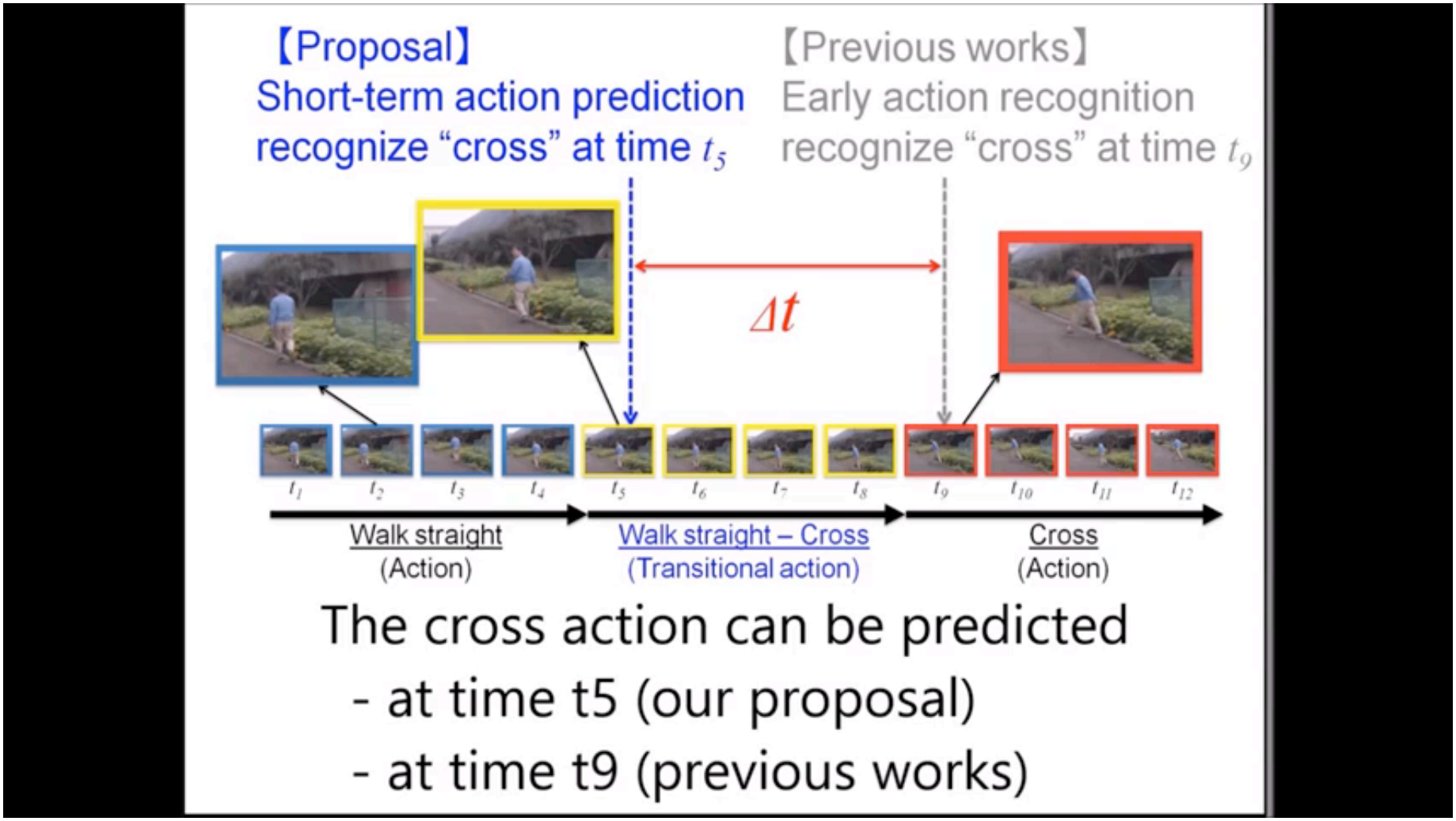
- こんな面白いことができる！と未来を見せる論文
- 今の技術を「使い，組み合わせ，洗練させ」ギリギリ実現可能な将来ビジョンを創り出す
- 新しい技術はどんどん使うが，その代わりそれでどんな未来ができるのかを想像して研究する

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

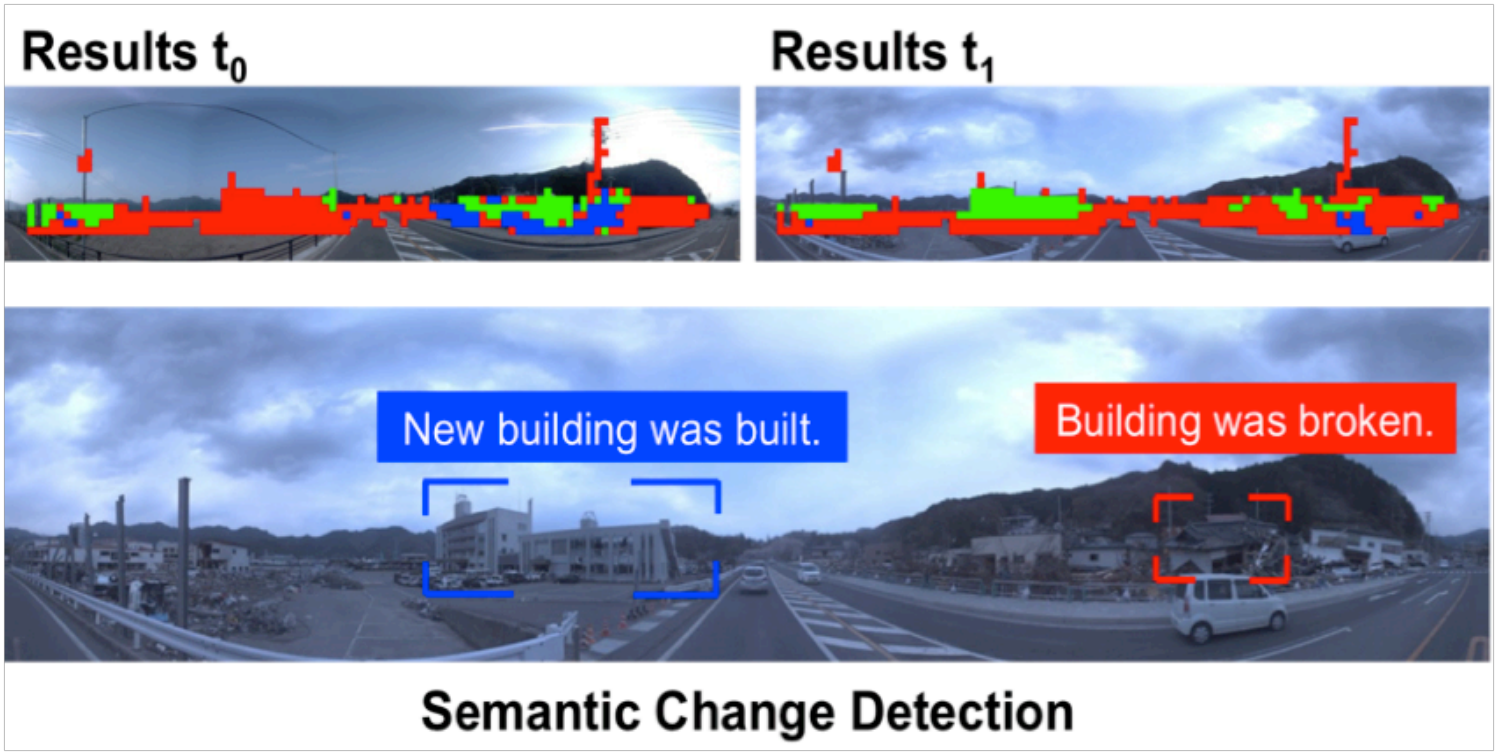
H. Kataoka et al. "Motion Representation with Acceleration Images" in ECCV 2016 Workshop, 2016.

Semantic Change Detection

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?

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

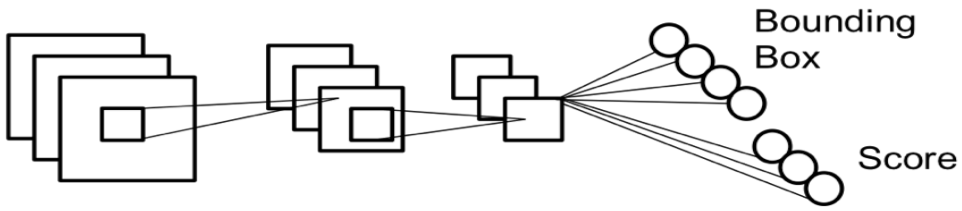
Pedestrian Near-Miss Detection

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

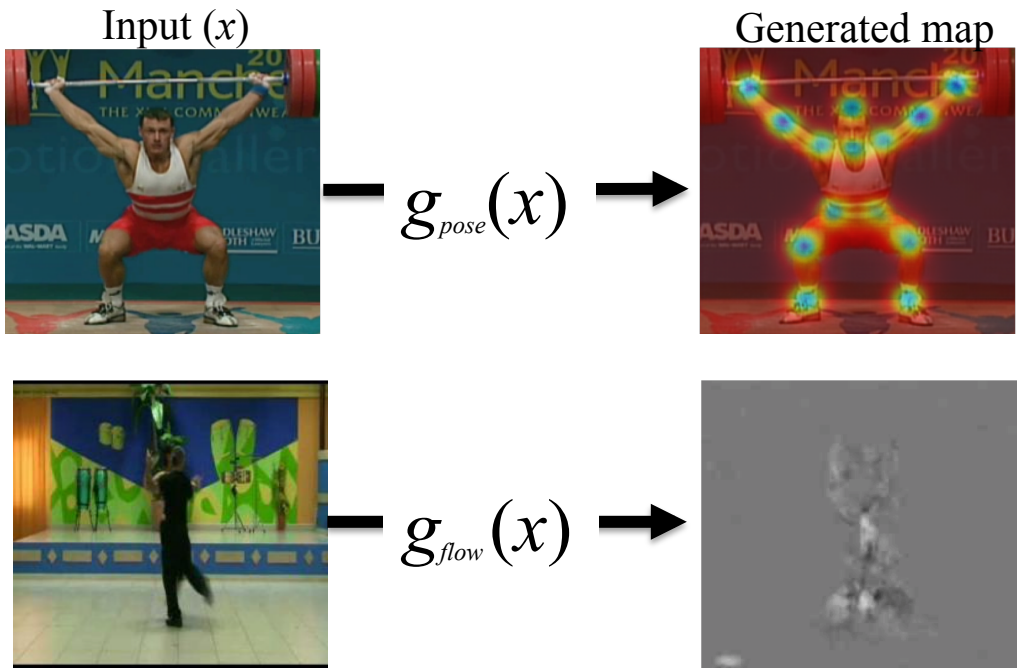
Risk low

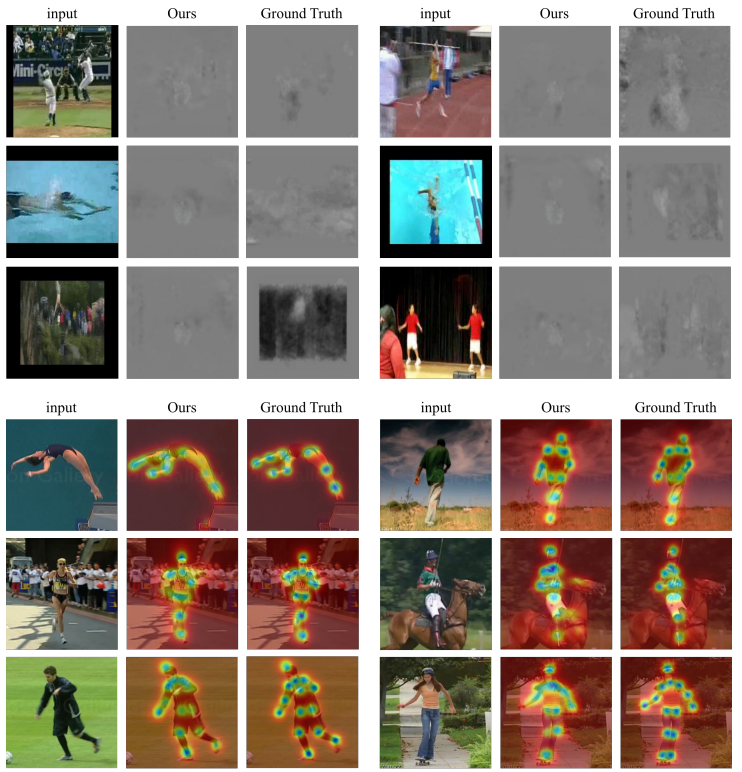
T. Suzuki et al. “Pedestrian Near-Miss Analysis on Vehicle-Mounted Driving Recorders” in MVA, 2017.

Generated Motion Maps

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



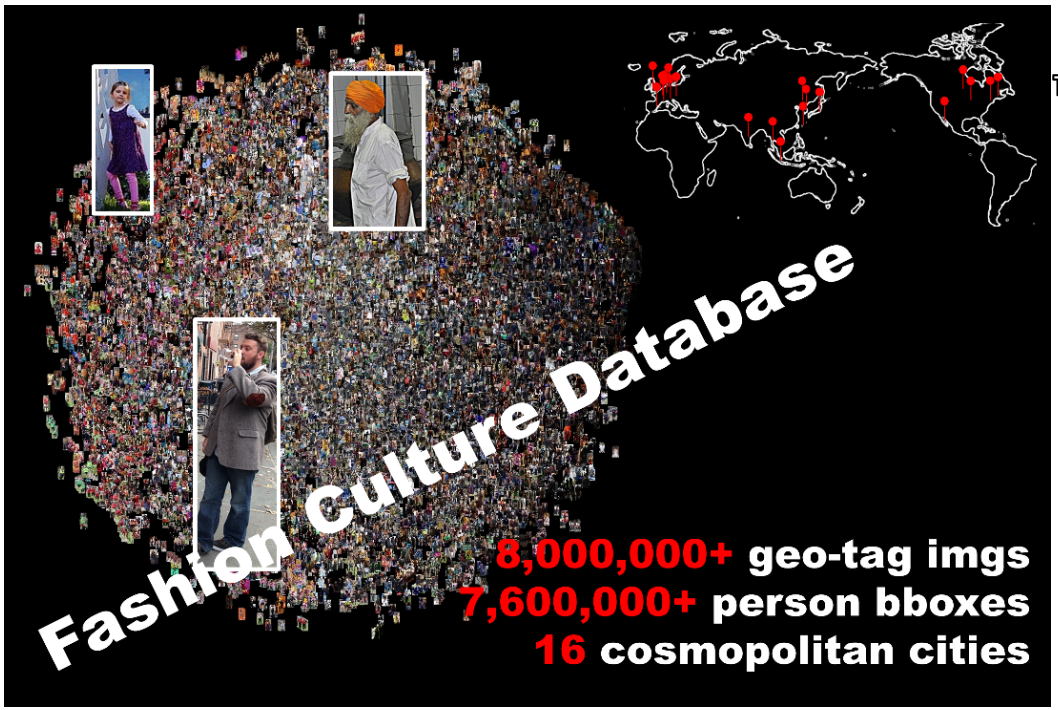


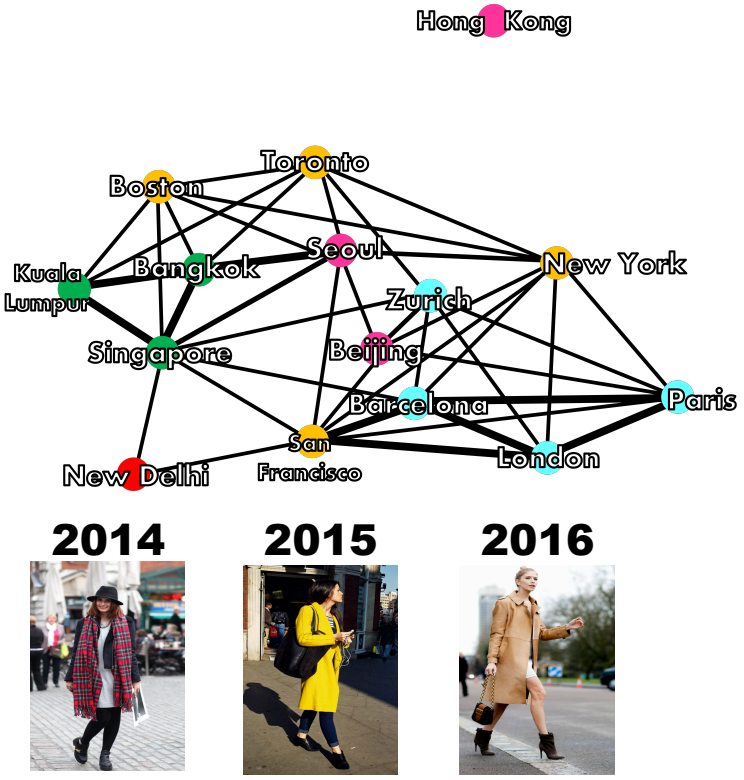
Y. Matsuzaki et al. “Generated Motion Maps” in CVPRW, 2017.

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





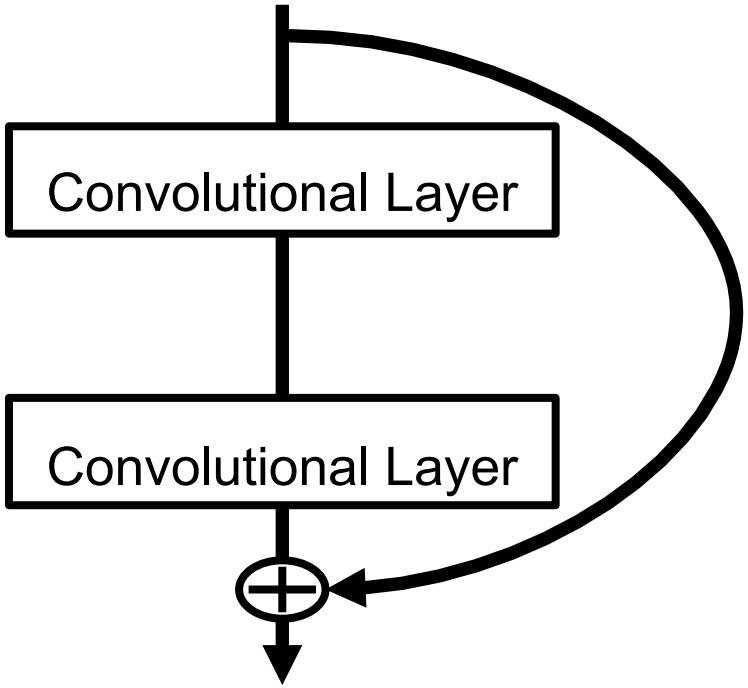
K. Abe et al. “Dynamic Fashion Cultures”, arXiv, 2017 & MIRU 2017 Oral. (Best Student Paper)

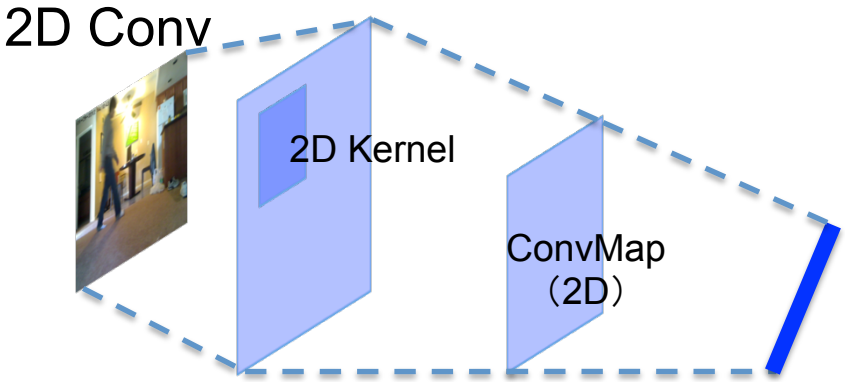
3D ResNets

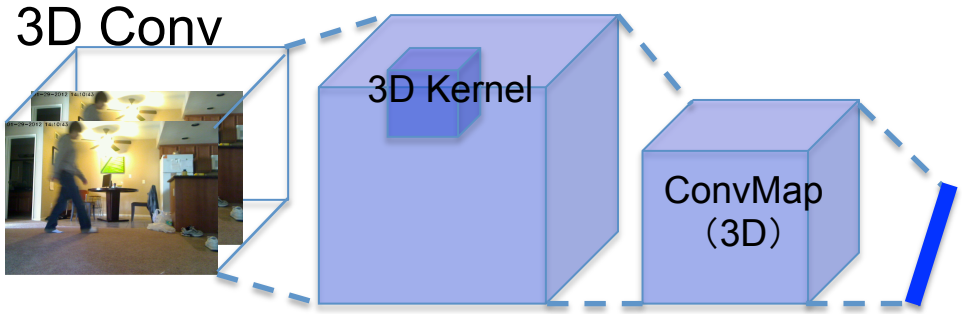
ResNetsのカーネルに3D Convを使用

- 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）

次に何をすべきか？

Beyond 2017

今後フォーカスすべき研究分野

より少ない教師で特徴表現学習

- 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！

今後フォーカスすべき研究分野

徹底評価により知見を獲得

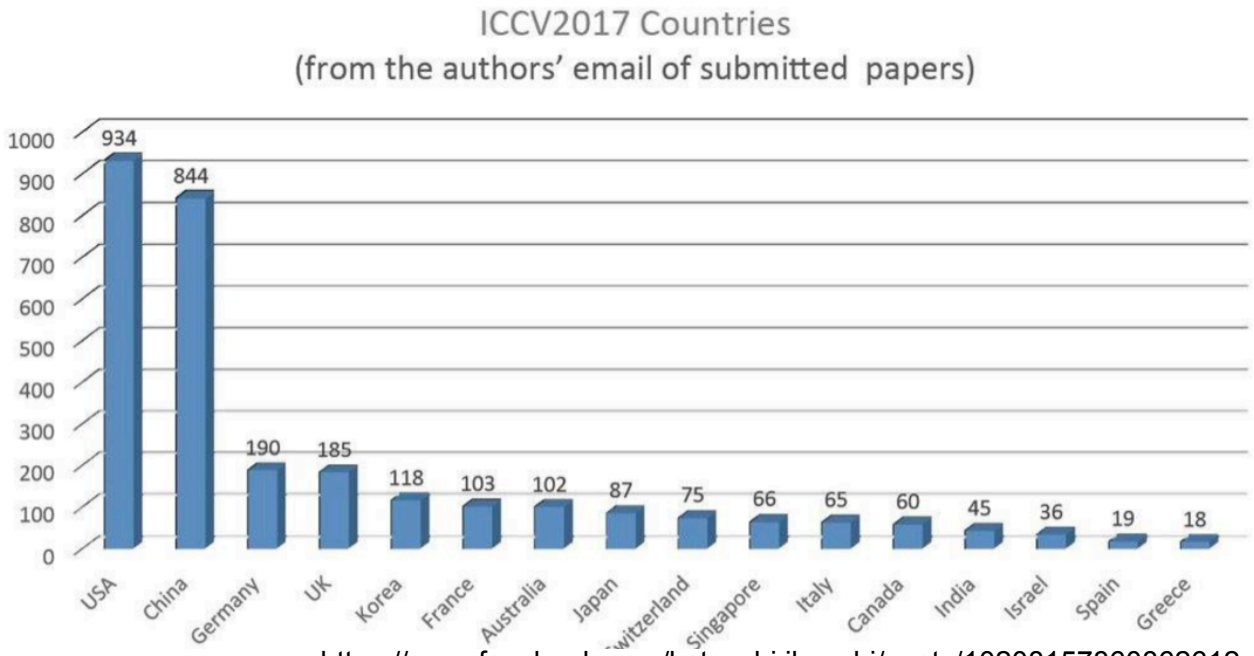
- 「やってみたらできた」という論文が多いので整理する論文は定期的に必要
- 次世代の手法になりうるか？何をどう使えば良い？

今後フォーカスすべき研究分野

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

今後，日本のCVが浮上するシナリオは？



https://www.facebook.com/katsushi.ikeuchi/posts/10208157860862612

「かつて，日本のCVは凄かった」らしい

投稿「数」を圧倒的に増やす？

修士の学生が中心的に投稿するのが近道か
投稿数を稼げないのに質なんて向上できるわけがない

B/Mの学生が活躍できるエコシステム

教員を介さなくても相互補完/自己組織的に研究の考案や質を高める
名古屋CV・PRML（特にアルコン）はそのいい実例だと思う

名古屋CVPRML勉強会から日本のCVを動かそう！

ポテンシャル，経験ともに十分
あとはチャレンジするのみ