MIRU2018若手プログラム

グループサーベイ法

片岡 裕雄, *Ph.D.*

http://www.hirokatsukataoka.net/

片岡 裕雄（かたおか ひろかつ）

産業技術総合研究所 CV研究グループ/AL連携研究室/AIセンター 研究員
東京電機大学客員研究員

Hirokatsu KATAOKA, *Ph.D.*
片岡 裕雄
Research Scientist, AIST, Japan.

2014年慶應義塾大学大学院理工学研究科修了，博士（工学）．2013，2014年 ミュンヘン工科大学 Visiting Scientist，2014年東京大学博士研究員，2015年産総研特別研究員．2016年4月より現職．画像認識，動画解析，人物行動解析に興味を持つ．cvpaper.challenge主宰．2011年ViEW小田原賞，2013年電気学会誌論文奨励賞，2014年藤原賞，2016年ECCV WS Brave New Idea Award.

mypage: http://www.hirokatsukataoka.net/

cvpaper.challenge: http://hirokatsukataoka.net/project/cc/index_cvpaperchallenge.html

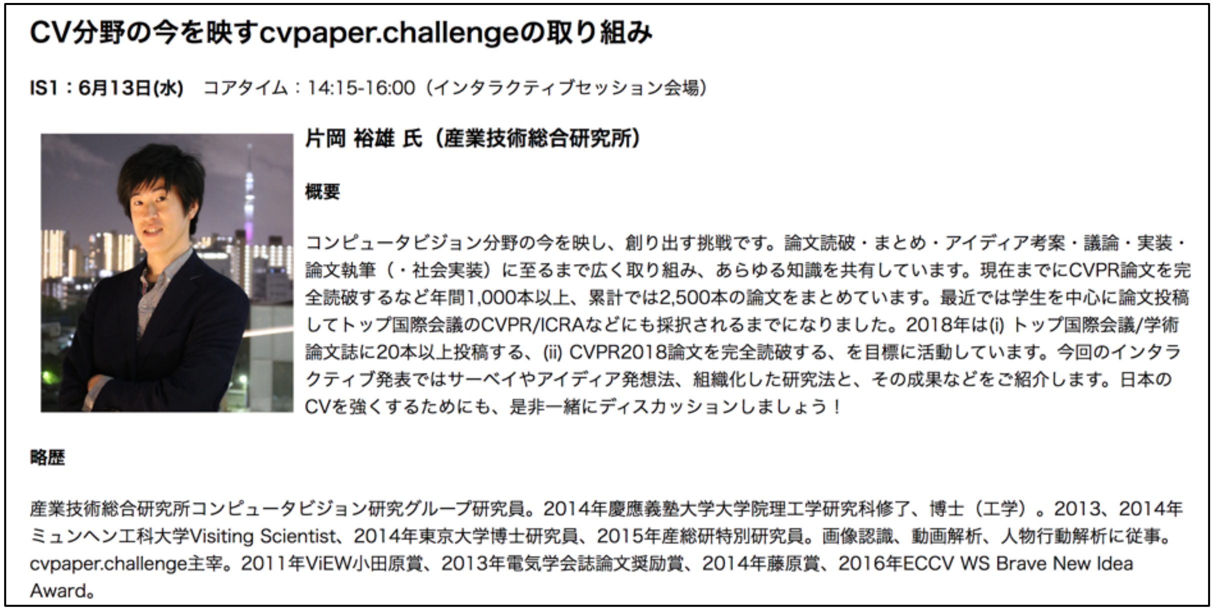
片岡の主宰するcvpaper.challenge

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/電機大/慶應大/早大/東大による約30名
- BraveNewなアイディアをトップ国際会議*に投稿

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

年間1,000本以上，累計2,500本以上のスライドを作成
本取り組みの結果10本以上の論文（含CVPRx2, ICRA, BMVC, ICPRx2, CVPRWx6, ECCVWx2, ICCVW）が採択
8件の招待講演，3件の国内外での受賞



CV分野の今を映すcvpaper.challengeの取り組み

IS1：6月13日(水)　コアタイム：14:15-16:00（インタラクティブセッション会場）

片岡 裕雄 氏（産業技術総合研究所）

概要

コンピュータビジョン分野の今を映し、創り出す挑戦です。論文読破・まとめ・アイディア考案・議論・実装・論文執筆（・社会実装）に至るまで広く取り組み、あらゆる知識を共有しています。現在までにCVPR論文を完全読破するなど年間1,000本以上、累計では2,500本の論文をまとめています。最近では学生を中心に論文投稿してトップ国際会議のCVPR/ICRAなどにも採択されるまでになりました。2018年は(i) トップ国際会議/学術論文誌に20本以上投稿する、(ii) CVPR2018論文を完全読破する、を目標に活動しています。今回のインタラクティブ発表ではサーベイやアイディア発想法、組織化した研究法と、その成果などをご紹介します。日本のCVを強くするためにも、是非一緒にディスカッションしましょう！

略歴

産業技術総合研究所コンピュータビジョン研究グループ研究員。2014年慶應義塾大学大学院理工学研究科修了、博士（工学）。2013、2014年ミュンヘン工科大学Visiting Scientist、2014年東京大学博士研究員、2015年産総研特別研究員。画像認識、動画解析、人物行動解析に従事。cvpaper.challenge主宰。2011年ViEW小田原賞、2013年電気学会誌論文奨励賞、2014年藤原賞、2016年ECCV WS Brave New Idea Award。

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
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: http://hirokatsukataoka.net/project/cc/index_cvpaperchallenge.html
Twitter: @CVpaperChalleng
SlideShare: @cvpaperchallenge

サーベイチュートリアル

- 前半: 米谷
  - 特にサーベイを論文化するうえでどういう点に留意すべきか
- 後半: 片岡
  - グループでサーベイをするうえでどういう点に留意すべきか

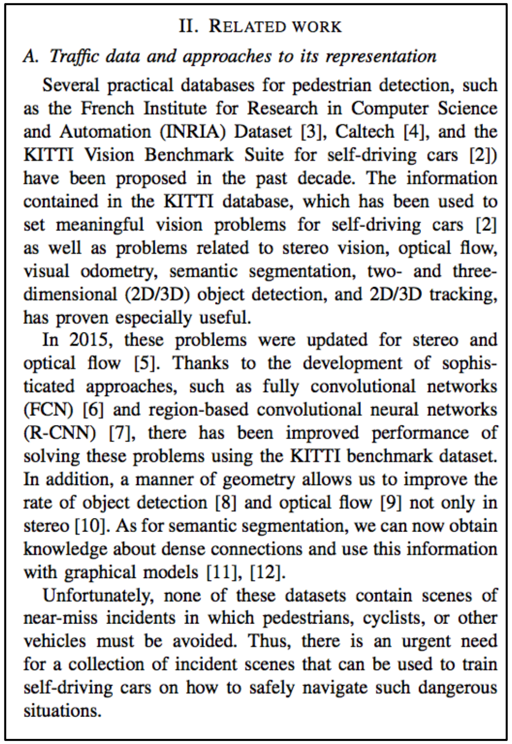
- 本講演の様子およびスライドは後日若手Pウェブサイト上で公開します

サーベイについて

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，サーベイをするのか？

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（ここ重要）

- 分野の方向性を自ら定める
- より良く，正しい方向へ導く

どのように，サーベイをするのか？

「網羅的に調べる」ことで分野の状況を高精度に概観

– サーベイの質や量が研究テーマ考案・選定に直結

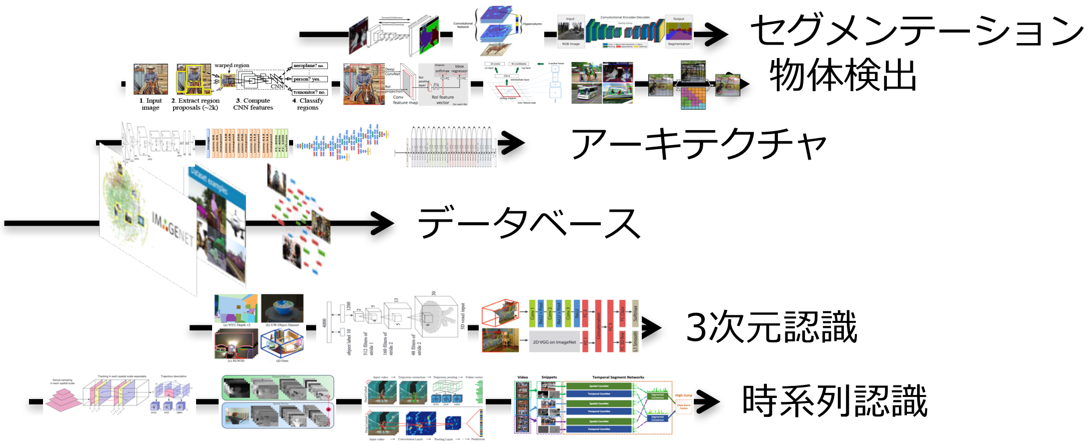
– （研究室単位で）1,000^{+}本/年は論文を読む力をつけよう（速読/精読が重要）

「歴史/トレンド」のストリームを多数把握

– 点より線/面で技術を捉える（他分野からも学ぶ）

– 次に何をすれば良いか/その技術の本質や行間が，肌感覚でわかる

「解かれた問題」が犇めく空間を埋めつつ「解かれていない問題を探索」

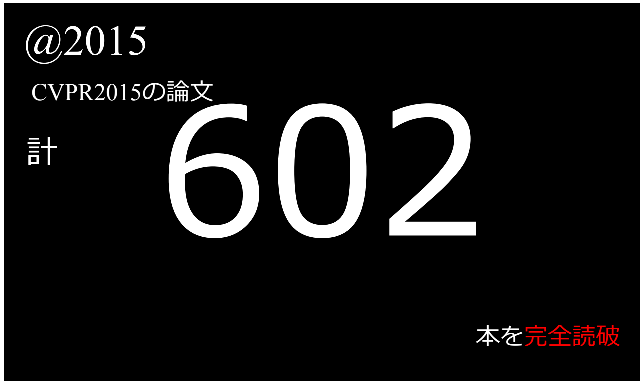

セグメンテーション
物体検出
アーキテクチャ
データベース
3次元認識
時系列認識

どれくらいサーベイをしてるの？

個人/グループとしてサーベイを推進

- 「個人」で達成
  - 2015年度 615本，2016年度 400+本
  - うち速読900本，精読100本くらい？
- 「グループ」で達成
  - 2015年，CVPR 2015 完全読破（約10名）
  - 2016年，1,000本読破達成（約20名）
  - 2018年，CVPR 2018 完全読破チャレンジ実行中（約30名）

http://hirokatsukataoka.net/project/cc/cvpr2018survey.html

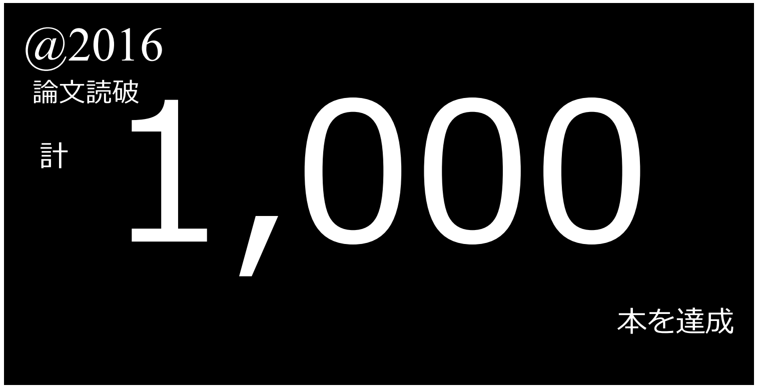


@2015

CVPR2015の論文

計 602

本を完全読破



@2016

論文読破

計 1,000

本を達成

意識的にサーベイして何が変わった？

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

組織的にサーベイしてから…

- 論文調査：網羅的かつ流れを把握
- 研究：物事の本質に迫るような問いを意識
- サーベイ/テーマ設定/実験/論文執筆に至るまで「質・効率を高める努力」を徹底
  - 探索は終わらない！

優れた問いを見つけよう！

思いつき（単純拡張）研究から離脱しよう

- To invent, you need a good imagination and a pile of junk
  - 良い発想とガラクタの山（膨大な知識量）が必要だ
- 知識を詰め込む/整理することで視野が繋がる
  - 各個人で頑張る（量増加），集団で整理して体系化（質向上）
  - 個人でモチベーションを保ち，チーム力を発揮できる体制を整える

個人とグループの知識獲得がキモ

個人のサーベイ

グループのサーベイ

個人のサーベイ
（分割）

グループのサーベイ
（統合）

「分割と統合」の組み合わせが重要

個人サーベイ

個人サーベイで意識すること

速読と精読の組み合わせ

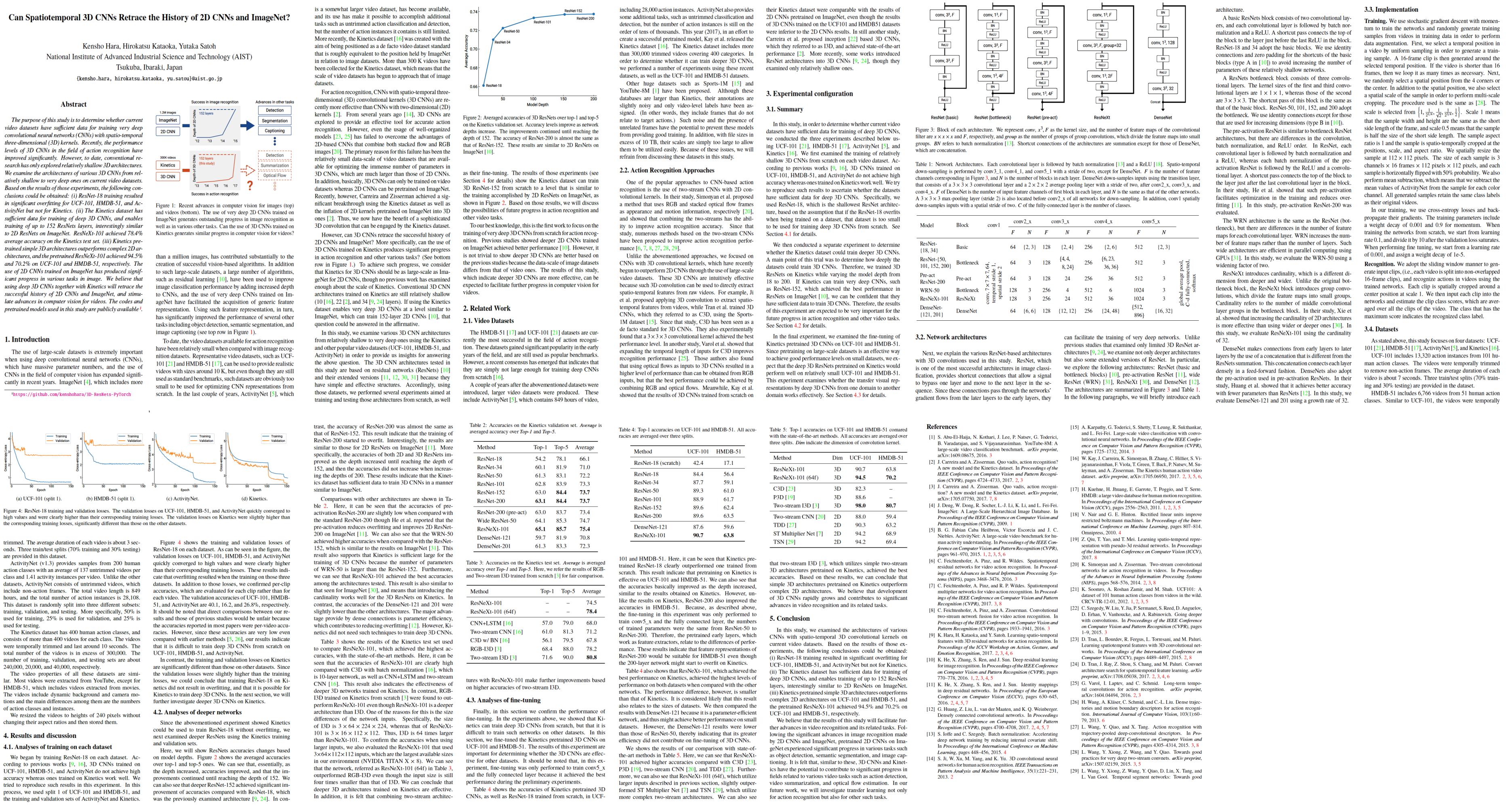
- 速読（50^{+}本/月）
  - とにかく「広く浅く」
  - 研究テーマやアイディア考案の時に行うサーベイ法
- 精読（10^{-}本/月）
  - 実装レベルで「狭く深く」
  - 具体的なテーマが決まっている際に

尺度を変えた読み方の統合で研究の効率化

論文を読もう！

論文は目的別に読むために体系化されている

– CVの論文を参考に解説します！



Can Spatiotemporal 3D CNNs Retrace the History of 2D CNNs and ImageNet?

Kensho Hara, Hirokatsu Kataoka, Yutaka Satoh

National Institute of Advanced Industrial Science and Technology (AIST)

Tsukuba, Ibaraki, Japan

K. Hara, H. Kataoka, Y. Satoh, “Can Spatiotemporal 3D CNNs Retrace the History of 2D CNNs and ImageNet?”, in CVPR 2018.

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

気をつけていること

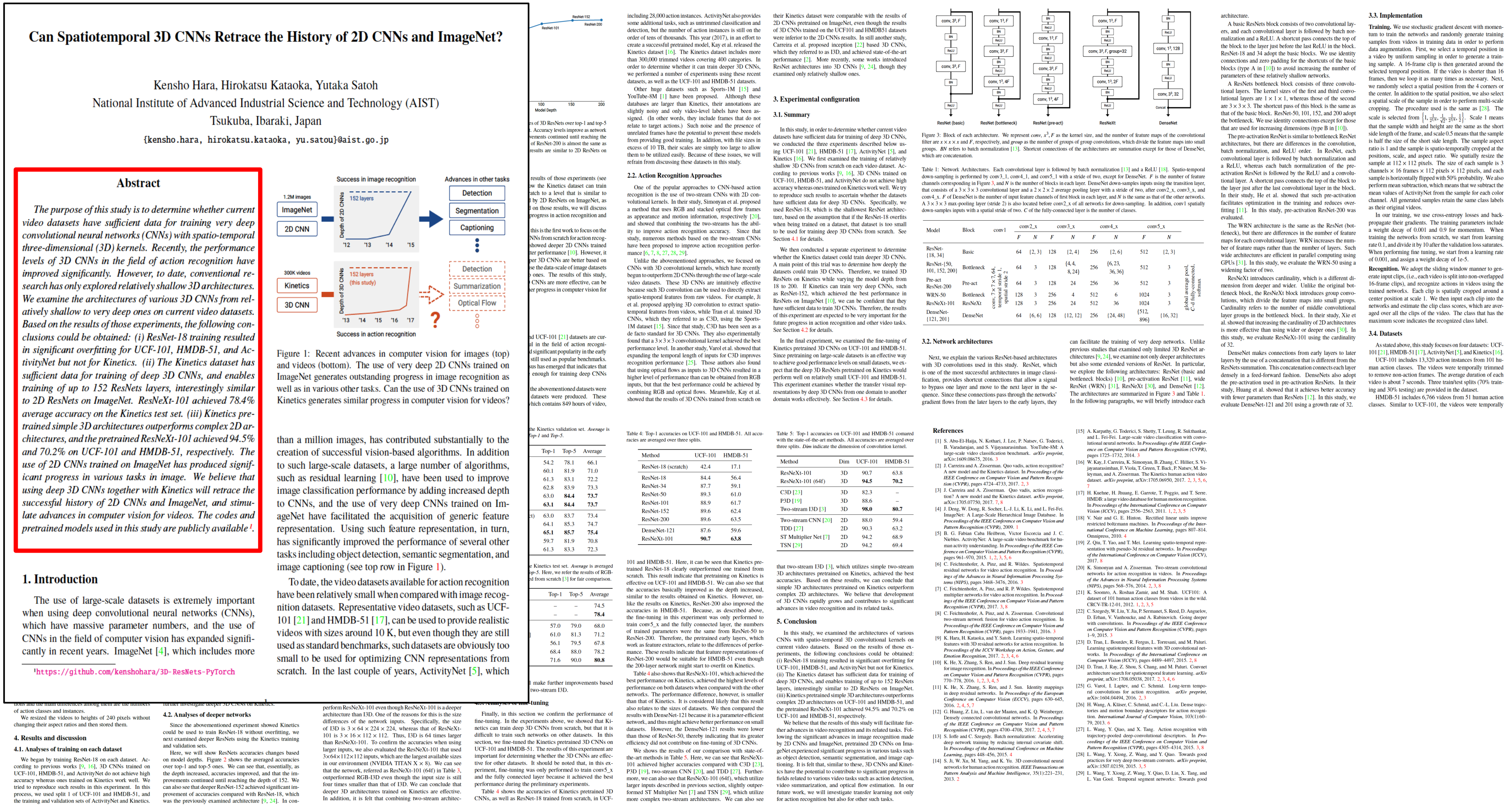
森を見る，木を見る

- 森：ざっと全体を通して見る
  - 速読ならこれで十分
  - 背景，イントロ，図表，結果，結論を中心に
- 木：細かいところまで目を通す（但し，目的を見失わない）
  - どんな情報が欲しいかを明確にして読む
  - 実装したい？輪講資料を作りたい？

森を見る（速読）

アブストラクトを見る

– 良いアブストラクトだとなんとなく全体がわかる



Can Spatiotemporal 3D CNNs Retrace the History of 2D CNNs and ImageNet?

Kensho Hara, Hirokatsu Kataoka, Yutaka Satoh
National Institute of Advanced Industrial Science and Technology (AIST)
Tsukuba, Ibaraki, Japan
{kensho.hara, hirokatsu.kataoka, yu.satou}@aist.go.jp

Abstract

The purpose of this study is to determine whether current video datasets have sufficient data for training very deep convolutional neural networks (CNNs) with spatio-temporal three-dimensional (3D) kernels. Recently, the performance levels of 3D CNNs in the field of action recognition have improved significantly. However, to date, conventional research has only explored relatively shallow 3D architectures. We examine the architectures of various 3D CNNs from relatively shallow to very deep ones on current video datasets. Based on the results of those experiments, the following conclusions could be obtained: (i) ResNet-18 training resulted in significant overfitting for UCF-101, HMDB-51, and ActivityNet but not for Kinetics. (ii) The Kinetics dataset has sufficient data for training of deep 3D CNNs, and enables training of up to 152 ResNets layers, interestingly similar to 2D ResNets on ImageNet. ResNeXt-101 achieved 78.4% average accuracy on the Kinetics test set. (iii) Kinetics pretrained simple 3D architectures outperforms complex 2D architectures, and the pretrained ResNeXt-101 achieved 94.5% and 70.2% on UCF-101 and HMDB-51, respectively. The use of 2D CNNs trained on ImageNet has produced significant progress in various tasks in image. We believe that using deep 3D CNNs together with Kinetics will retrace the successful history of 2D CNNs and ImageNet, and stimulate advances in computer vision for videos. The codes and pretrained models used in this study are publicly available[1].

1. Introduction

The use of large-scale datasets is extremely important when using deep convolutional neural networks (CNNs), which have massive parameter numbers, and the use of CNNs in the field of computer vision has expanded significantly in recent years. ImageNet [4], which includes more than a million images, has contributed substantially to the creation of successful vision-based algorithms. In addition to such large-scale datasets, a large number of algorithms, such as residual learning [10], have been used to improve image classification performance by adding increased depth to CNNs, and the use of very deep CNNs trained on ImageNet have facilitated the acquisition of generic feature representation. Using such feature representation, in turn, has significantly improved the performance of several other tasks including object detection, semantic segmentation, and image captioning (see top row in Figure 1).

To date, the video datasets available for action recognition have been relatively small when compared with image recognition datasets. Representative video datasets, such as UCF-101 [21] and HMDB-51 [17], can be used to provide realistic videos with sizes around 10 K, but even though they are still used as standard benchmarks, such datasets are obviously too small to be used for optimizing CNN representations from scratch. In the last couple of years, ActivityNet [5], which

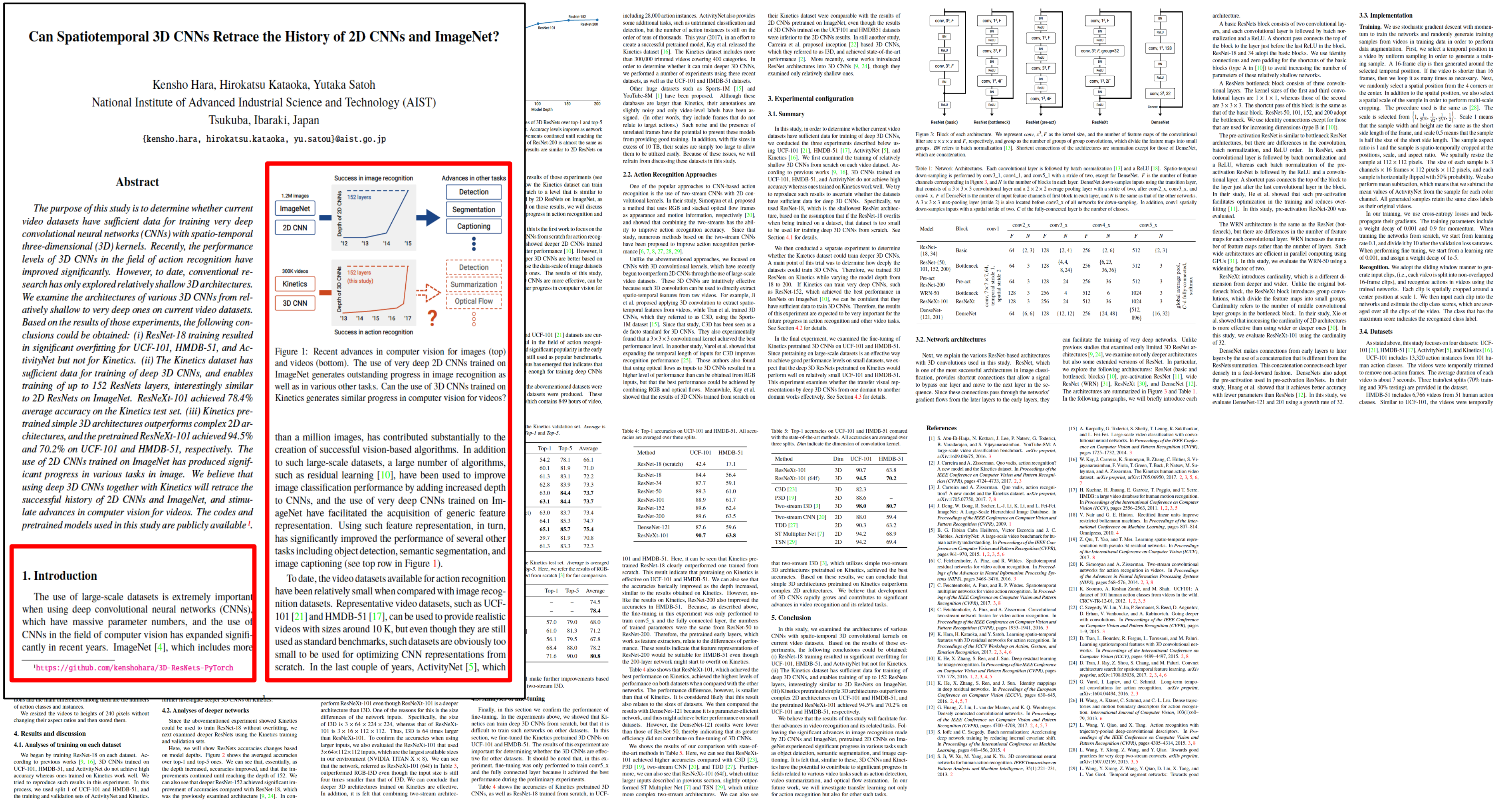
Figure 1: Recent advances in computer vision for images (top) and videos (bottom). The use of very deep 2D CNNs trained on ImageNet generates outstanding progress in image recognition as well as in various other tasks. Can the use of 3D CNNs trained on Kinetics generates similar progress in computer vision for videos?

[1] https://github.com/kenshohara/3D-ResNets-PyTorch

森を見る（速読）

イントロ，概念図を見る

– どんな問題設定？その研究の新規性は？



Can Spatiotemporal 3D CNNs Retrace the History of 2D CNNs and ImageNet?

Kensho Hara, Hirokatsu Kataoka, Yutaka Satoh
National Institute of Advanced Industrial Science and Technology (AIST)
Tsukuba, Ibaraki, Japan
{kensho.hara, hirokatsu.kataoka, yu.satou}@aist.go.jp

Abstract

The purpose of this study is to determine whether current video datasets have sufficient data for training very deep convolutional neural networks (CNNs) with spatio-temporal three-dimensional (3D) kernels. Recently, the performance levels of 3D CNNs in the field of action recognition have improved significantly. However, to date, conventional research has only explored relatively shallow 3D architectures. We examine the architectures of various 3D CNNs from relatively shallow to very deep ones on current video datasets. Based on the results of those experiments, the following conclusions could be obtained: (i) ResNet-18 training resulted in significant overfitting for UCF-101, HMDB-51, and ActivityNet but not for Kinetics. (ii) The Kinetics dataset has sufficient data for training of deep 3D CNNs, and enables training of up to 152 ResNets layers, interestingly similar to 2D ResNets on ImageNet. ResNeXt-101 achieved 78.4% average accuracy on the Kinetics test set. (iii) Kinetics pretrained simple 3D architectures outperforms complex 2D architectures, and the pretrained ResNeXt-101 achieved 94.5% and 70.2% on UCF-101 and HMDB-51, respectively. The use of 2D CNNs trained on ImageNet has produced significant progress in various tasks in image. We believe that using deep 3D CNNs together with Kinetics will retrace the successful history of 2D CNNs and ImageNet, and stimulate advances in computer vision for videos. The codes and pretrained models used in this study are publicly available[1].

1. Introduction

The use of large-scale datasets is extremely important when using deep convolutional neural networks (CNNs), which have massive parameter numbers, and the use of CNNs in the field of computer vision has expanded significantly in recent years. ImageNet [4], which includes more than a million images, has contributed substantially to the creation of successful vision-based algorithms. In addition to such large-scale datasets, a large number of algorithms, such as residual learning [10], have been used to improve image classification performance by adding increased depth to CNNs, and the use of very deep CNNs trained on ImageNet have facilitated the acquisition of generic feature representation. Using such feature representation, in turn, has significantly improved the performance of several other tasks including object detection, semantic segmentation, and image captioning (see top row in Figure 1).

To date, the video datasets available for action recognition have been relatively small when compared with image recognition datasets. Representative video datasets, such as UCF-101 [21] and HMDB-51 [17], can be used to provide realistic videos with sizes around 10 K, but even though they are still used as standard benchmarks, such datasets are obviously too small to be used for optimizing CNN representations from scratch. In the last couple of years, ActivityNet [5], which

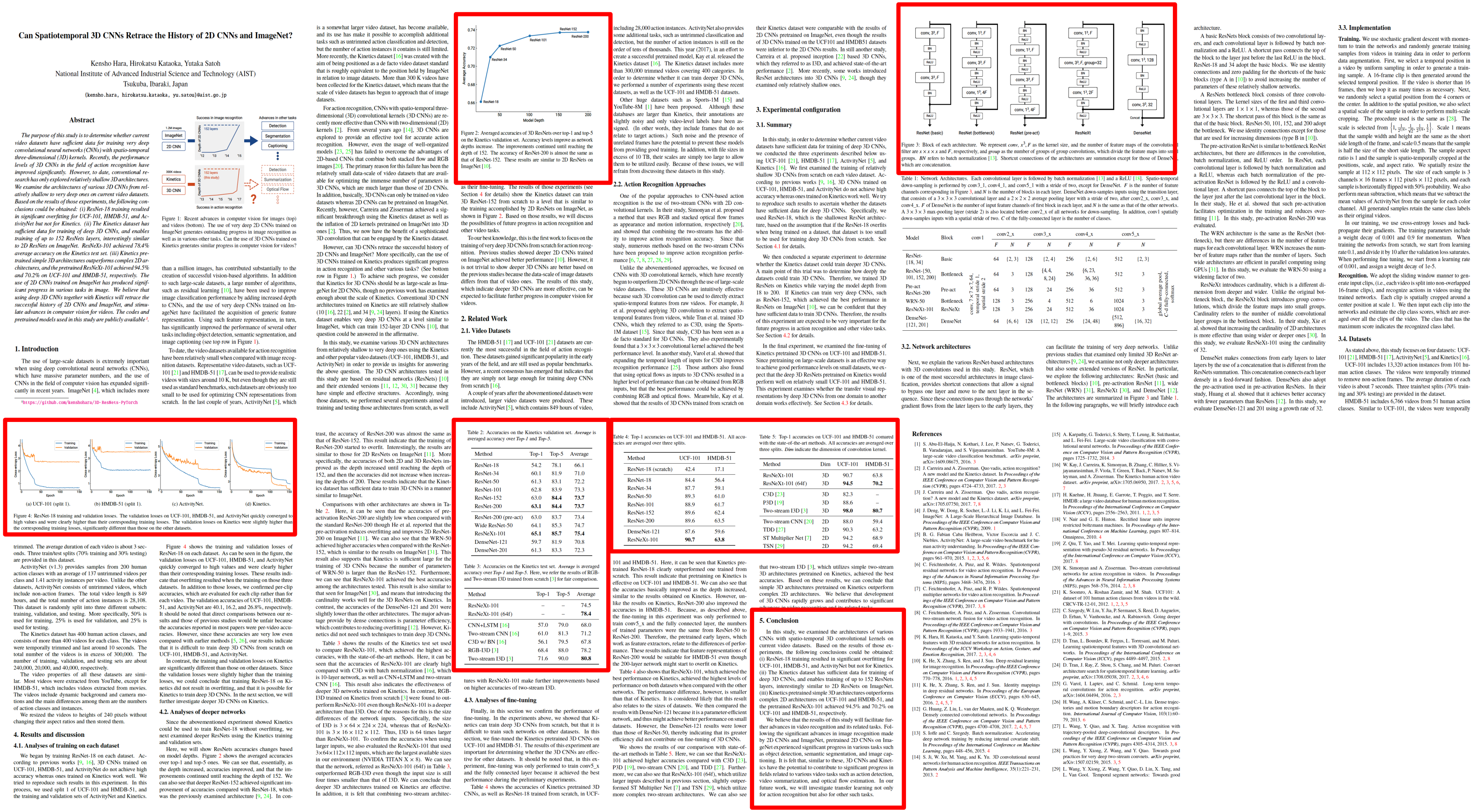
Figure 1: Recent advances in computer vision for images (top) and videos (bottom). The use of very deep 2D CNNs trained on ImageNet generates outstanding progress in image recognition as well as in various other tasks. Can the use of 3D CNNs trained on Kinetics generates similar progress in computer vision for videos?

[1] https://github.com/kenshohara/3D-ResNets-PyTorch

森を見る（速読）

その後は重要そうな部分に目を通す

- 手法，結果，結論など図表を中心に情報量の多い部分から目を通す



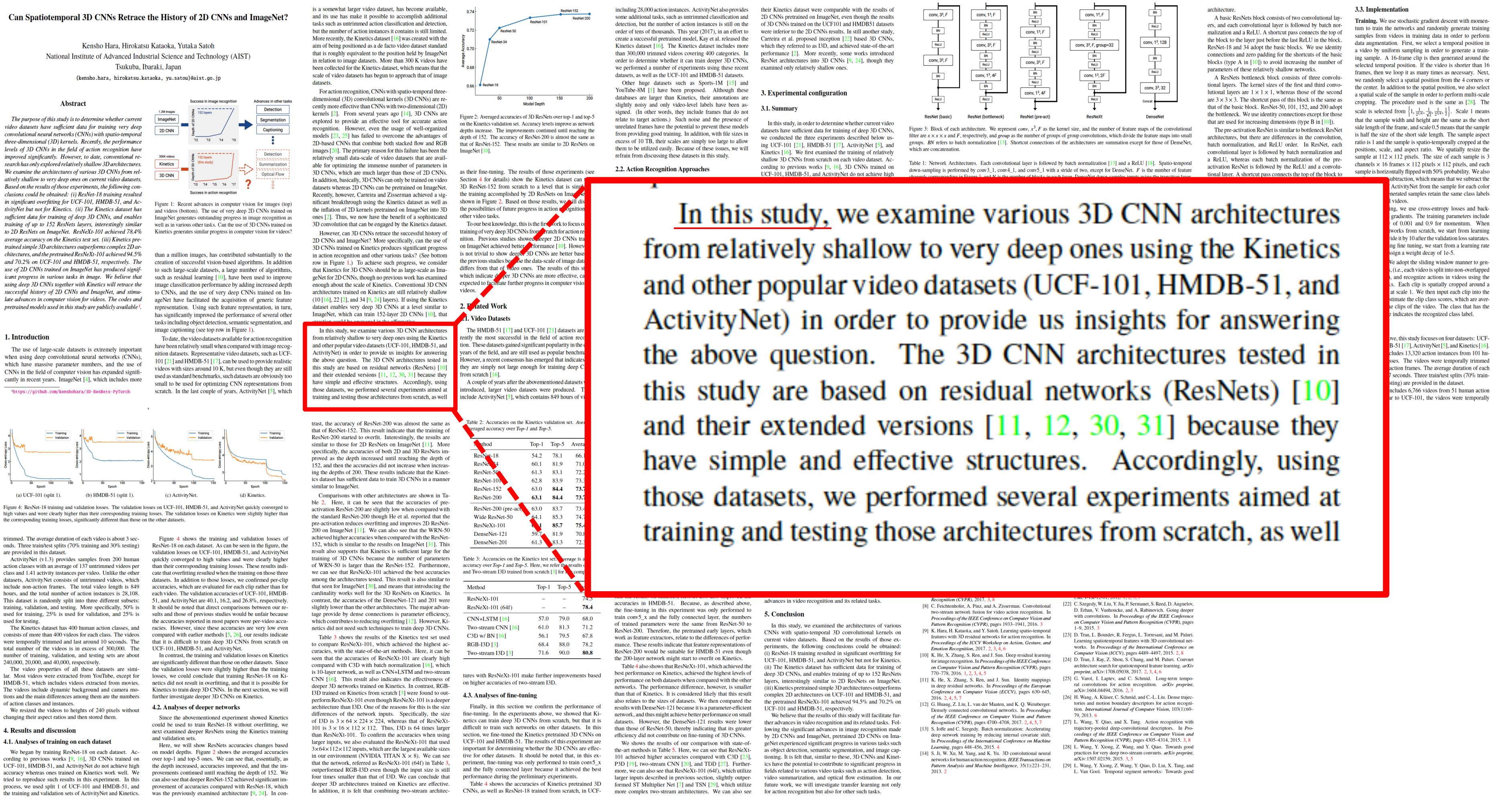
Can Spatiotemporal 3D CNNs Retrace the History of 2D CNNs and ImageNet?

Kensho Hara, Hirokatsu Kataoka, Yutaka Satoh
National Institute of Advanced Industrial Science and Technology (AIST)
Tsukuba, Ibaraki, Japan

森を見る（速読）

ポイントは意外と限られる？

– In this study (paper/research), ~ や Our contribution(s)~ など序盤に現れる



Can Spatiotemporal 3D CNNs Retrace the History of 2D CNNs and ImageNet?

Kensho Hara, Hirokatsu Kataoka, Yutaka Satoh

National Institute of Advanced Industrial Science and Technology (AIST)

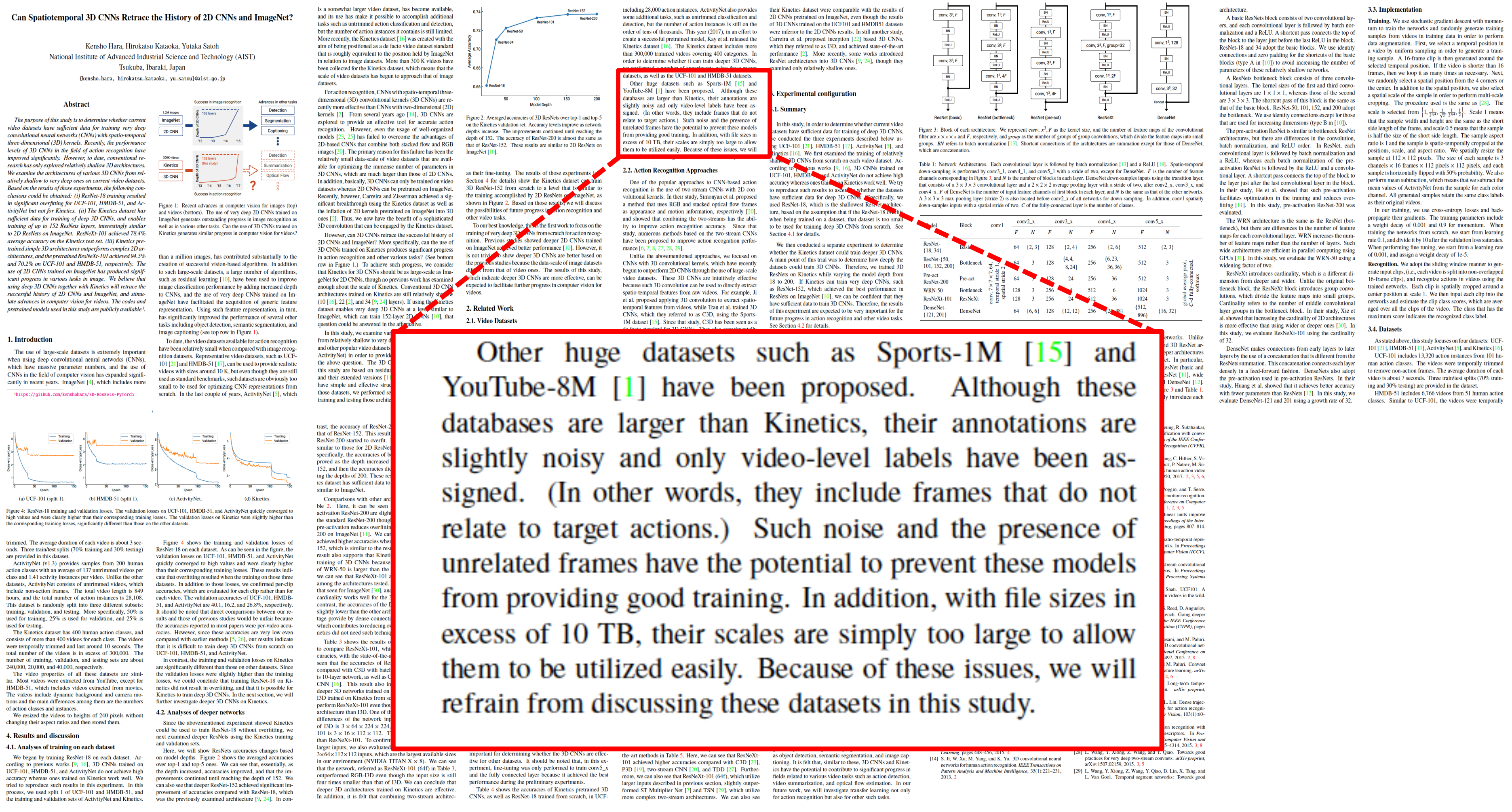
Tsukuba, Ibaraki, Japan

In this study, we examine various 3D CNN architectures from relatively shallow to very deep ones using the Kinetics and other popular video datasets (UCF-101, HMDB-51, and ActivityNet) in order to provide us insights for answering the above question. The 3D CNN architectures tested in this study are based on residual networks (ResNets) [10] and their extended versions [11, 12, 30, 31] because they have simple and effective structures. Accordingly, using those datasets, we performed several experiments aimed at training and testing those architectures from scratch, as well

森を見る（速読）

ポイントは意外と限られる？

- 関連研究の最後に従来との比較・差分が書かれる（ことが多い）

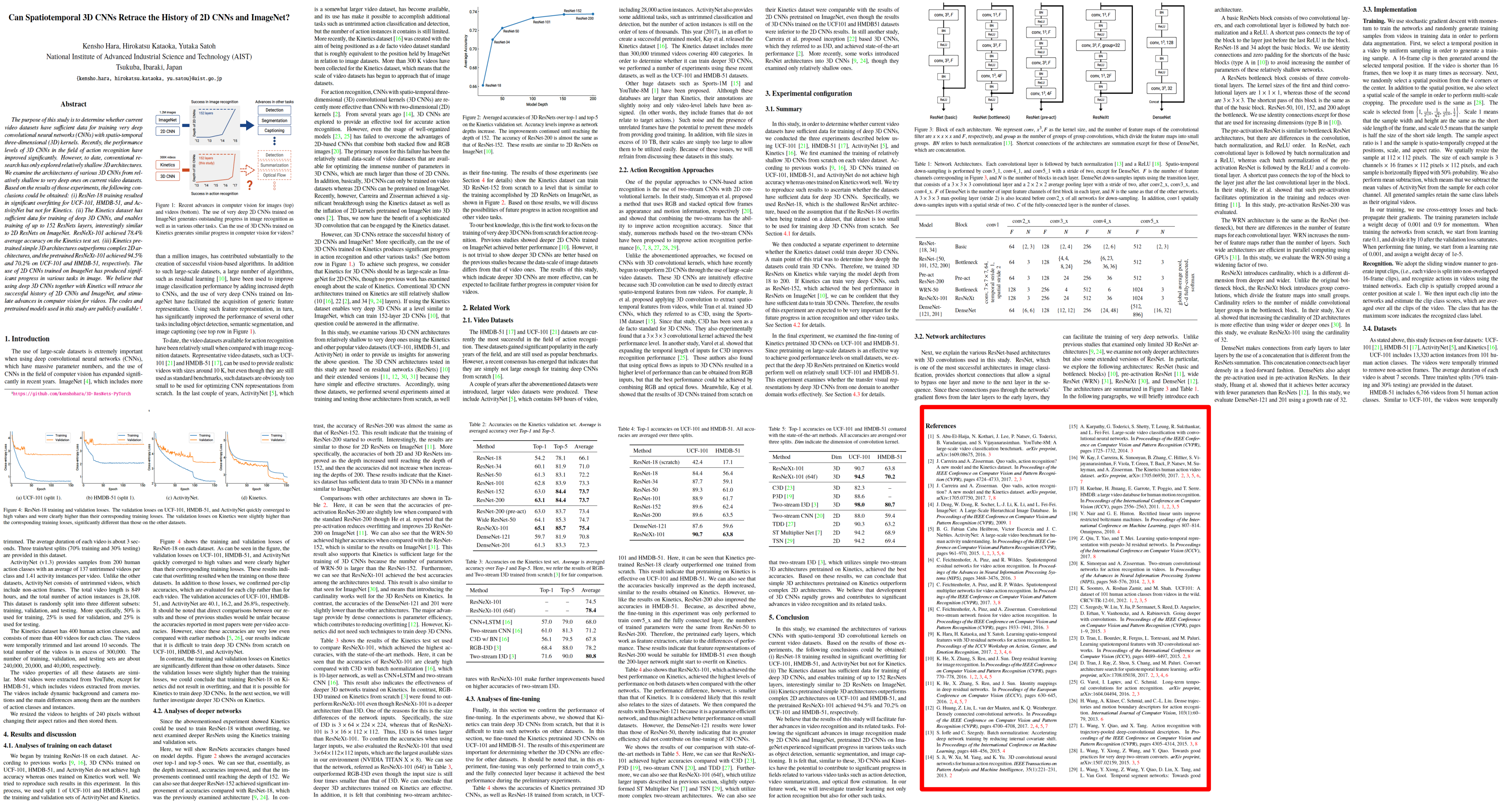


Other huge datasets such as Sports-1M [15] and YouTube-8M [1] have been proposed. Although these databases are larger than Kinetics, their annotations are slightly noisy and only video-level labels have been assigned. (In other words, they include frames that do not relate to target actions.) Such noise and the presence of unrelated frames have the potential to prevent these models from providing good training. In addition, with file sizes in excess of 10 TB, their scales are simply too large to allow them to be utilized easily. Because of these issues, we will refrain from discussing these datasets in this study.

森を見る（速読）

その分野に長けていれば…

– 引用されている論文や場所を見ればなんとなく内容が分かる？



Can Spatiotemporal 3D CNNs Retrace the History of 2D CNNs and ImageNet?

Kensho Hara, Hirokatsu Kataoka, Yutaka Satoh
National Institute of Advanced Industrial Science and Technology (AIST)
Tsukuba, Ibaraki, Japan
{kensho.hara, hirokatsu.kataoka, yu.satou}@aist.go.jp

Abstract

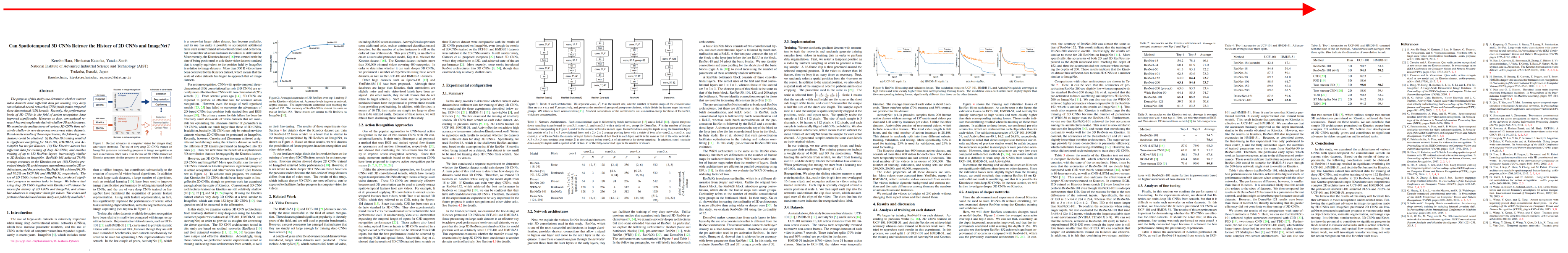
The purpose of this study is to determine whether current video datasets have sufficient data for training very deep convolutional neural networks (CNNs) with spatio-temporal three-dimensional (3D) kernels. Recently, the performance levels of 3D CNNs in the field of action recognition have improved significantly. However, to date, conventional research has only explored relatively shallow 3D architectures. We examine the architectures of various 3D CNNs from relatively shallow to very deep ones on current video datasets. Based on the results of those experiments, the following conclusions could be obtained: (i) ResNet-18 training resulted in significant overfitting for UCF-101, HMDB-51, and ActivityNet but not for Kinetics. (ii) The Kinetics dataset has sufficient data for training of deep 3D CNNs, and enables training of up to 152 ResNets layers, interestingly similar to 2D ResNets on ImageNet. ResNeXt-101 achieved 78.4% average accuracy on the Kinetics test set. (iii) Kinetics pre-

Figure 1: Recent advances in computer vision for images (top) and videos (bottom). The use of very deep 2D CNNs trained on ImageNet generates outstanding progress in image recognition as well as in various other tasks. Can the use of 3D CNNs trained on Kinetics generate similar progress in computer vision for videos?

木を見る（精読）

速読＋目的に合わせた読み方を

- 基本的には速読と，部分を詳細に読み込んでいく
- ここから先は何をしたいかに依存する



Can Spatiotemporal 3D CNNs Retrace the History of 2D CNNs and ImageNet?

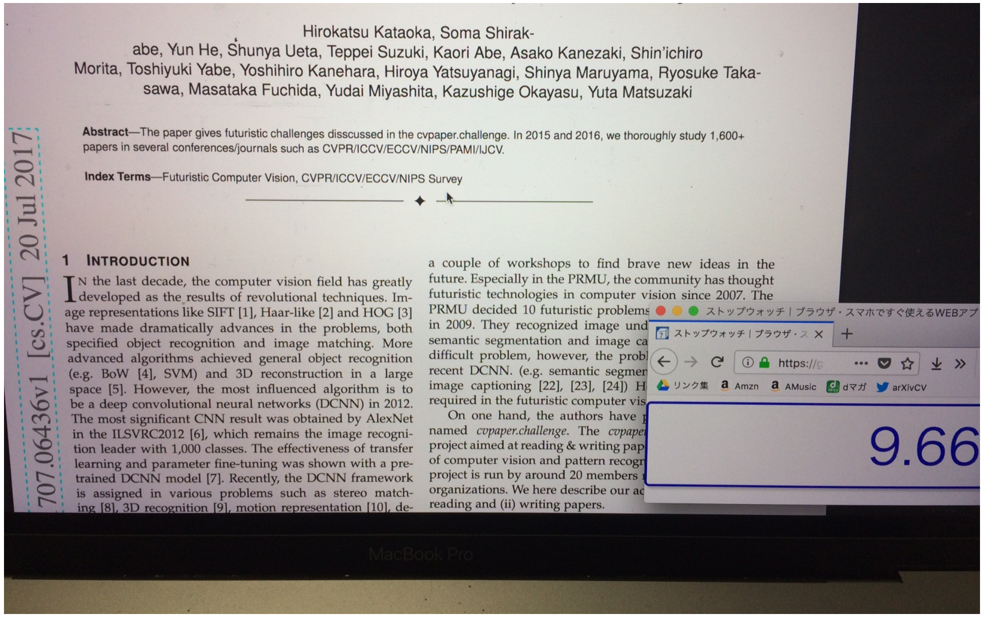
Kensho Hara, Hirokatsu Kataoka, Yutaka Satoh
National Institute of Advanced Industrial Science and Technology (AIST)
Tsukuba, Ibaraki, Japan
{kensho.hara, hirokatsu.kataoka, yu.satou}@aist.go.jp

分からない！という場合には最初から最後まで読んで見てもよいかも？

気をつけていること（個人サーベイ）

タイムトライアル

- 時間を気にして，締まりある読みにする
  - 実際ストップウォッチおいて論文読んでます！
- 目安時間
  - 速読（15～30分）
  - 精読（1時間～理解できるまで）



Hirokatsu Kataoka, Soma Shirakabe, Yun He, Shunya Ueta, Teppei Suzuki, Kaori Abe, Asako Kanezaki, Shin'ichiro Morita, Toshiyuki Yabe, Yoshihiro Kanehara, Hiroya Yatsuyanagi, Shinya Maruyama, Ryosuke Takasawa, Masataka Fuchida, Yudai Miyashita, Kazushige Okayasu, Yuta Matsuzaki

Abstract—The paper gives futuristic challenges disscussed in the cvpaper.challenge. In 2015 and 2016, we thoroughly study 1,600+ papers in several conferences/journals such as CVPR/ICCV/ECCV/NIPS/PAMI/IJCV.

Index Terms—Futuristic Computer Vision, CVPR/ICCV/ECCV/NIPS Survey

✦

707.06436v1 [cs.CV] 20 Jul 2017

1 INTRODUCTION

IN the last decade, the computer vision field has greatly developed as the results of revolutional techniques. Image representations like SIFT [1], Haar-like [2] and HOG [3] have made dramatically advances in the problems, both specified object recognition and image matching. More advanced algorithms achieved general object recognition (e.g. BoW [4], SVM) and 3D reconstruction in a large space [5]. However, the most influenced algorithm is to be a deep convolutional neural networks (DCNN) in 2012. The most significant CNN result was obtained by AlexNet in the ILSVRC2012 [6], which remains the image recognition leader with 1,000 classes. The effectiveness of transfer learning and parameter fine-tuning was shown with a pre-trained DCNN model [7]. Recently, the DCNN framework is assigned in various problems such as stereo matching [8], 3D recognition [9], motion representation [10], de- a couple of workshops to find brave new ideas in the future. Especially in the PRMU, the community has thought futuristic technologies in computer vision since 2007. The PRMU decided 10 futuristic problems
in 2009. They recognized image und
semantic segmentation and image ca
difficult problem, however, the probl
recent DCNN. (e.g. semantic segmen
image captioning [22], [23], [24]) H
required in the futuristic computer vis

On one hand, the authors have
named *cvpaper.challenge*. The *cvpaper*
project aimed at reading & writing pap
of computer vision and pattern recogn
project is run by around 20 members
organizations. We here describe our ac
reading and (ii) writing papers.

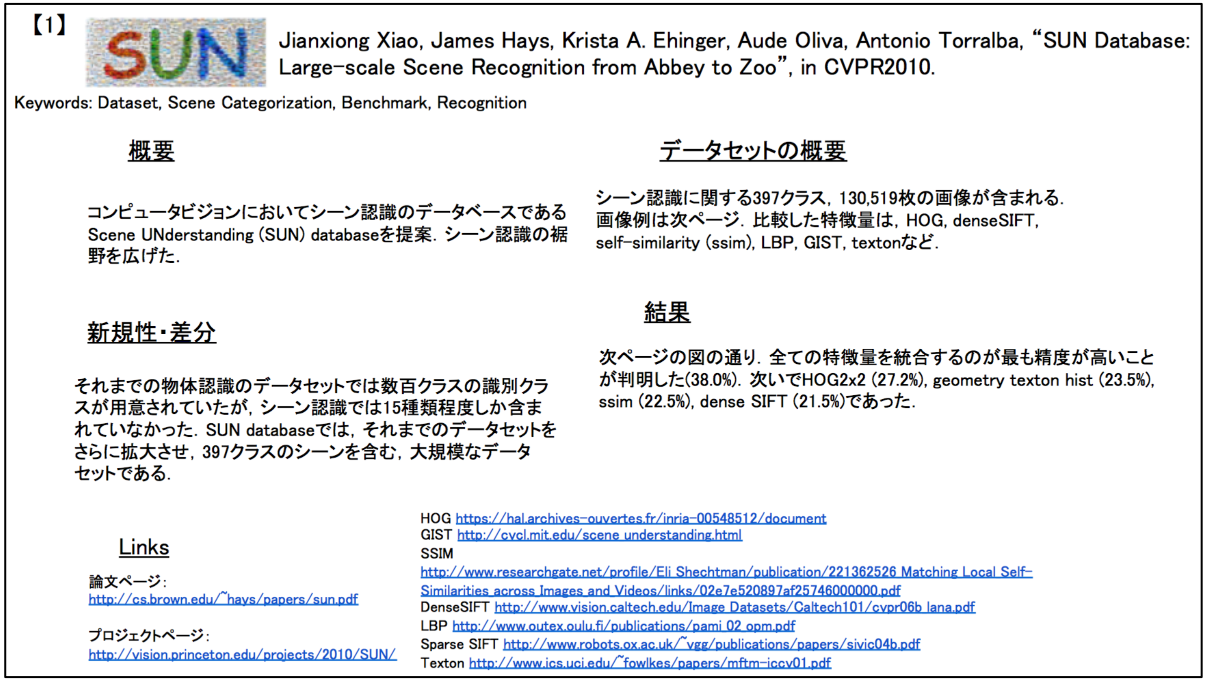
ストップウォッチ | ブラウザ・スマホですぐ使えるWEBアプ
ストップウォッチ | ブラウザ・ス
https://g
リンク集
Amzn
AMusic
dマガ
arXivCV
9.66
MacBook Pro

気をつけていること（個人サーベイ）

論文サマリを作成しよう

– （自分だけでなく他の人が）素早く研究の肝をつかむ

– まとめることで記憶の定着を早くする



【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

まとめの例：概要，新規性，手法，結果，リンク等があるのが望ましい

グループサーベイ

グループサーベイで意識すること

集団の力をうまく利用する

「モチベーション維持」

「集合知」

「作業分割と統合」

等を発揮するためグループで活動するメリットは多い

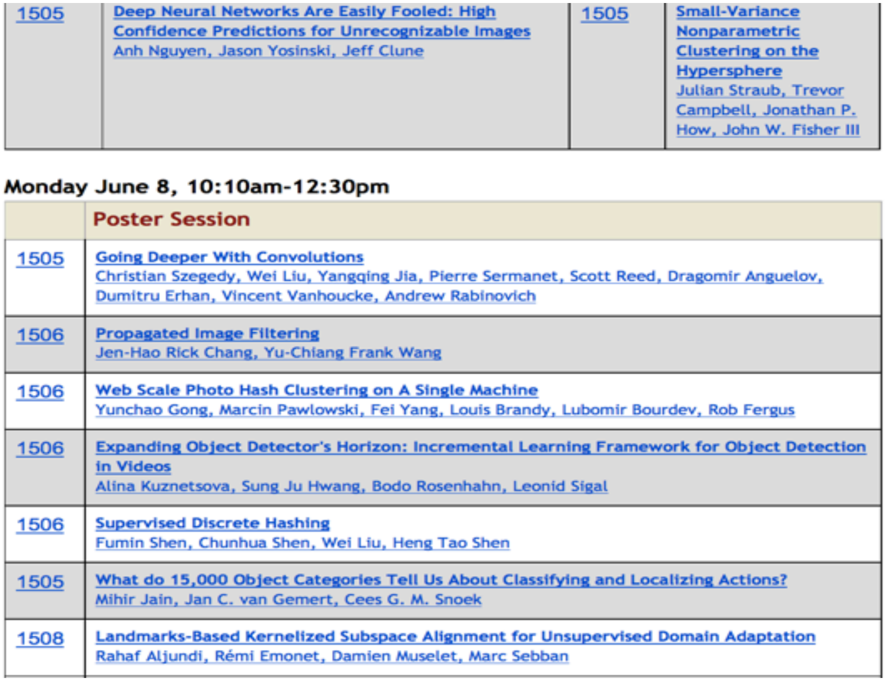
最近の重要論文は本数が多いので協力して知識を獲得する

(メジャー会議$_{\text{のみ}}$で1,400^{+}本/年，arXivは5,000^{+}本/年？)

論文リストを作成

タスク/進捗を可視化する

- {全体数, 残りの本数} がどれくらいか見えるように
- ノルマはゆるーく決めるが，チェックは確実に
  - タイトに個人を決めると精神的に辛くなるのでカバーし合う



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

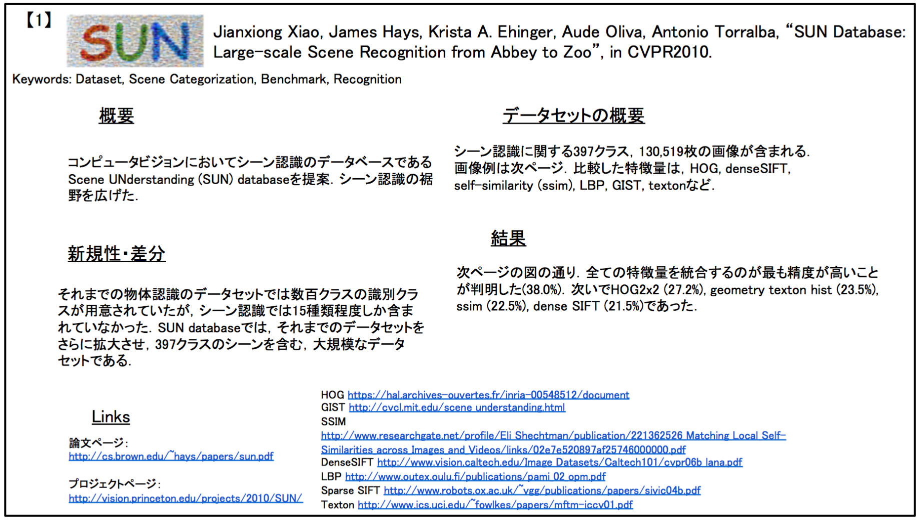
	Poster Session
1505	Going Deeper With Convolutions Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich
1506	Propagated Image Filtering Jen-Hao Rick Chang, Yu-Chiang Frank Wang
1506	Web Scale Photo Hash Clustering 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

CVPR 2015 完全読破チャレンジより

分担して資料を作成してお互いに参照

論文サマリを上手に活用

- 論文を確実にまとめる，を徹底すると後に読む人が楽
  - リストにチェックされている論文はサマリがある
- 素早くポイントをつかんで原論文を読み始める



【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．

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/

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

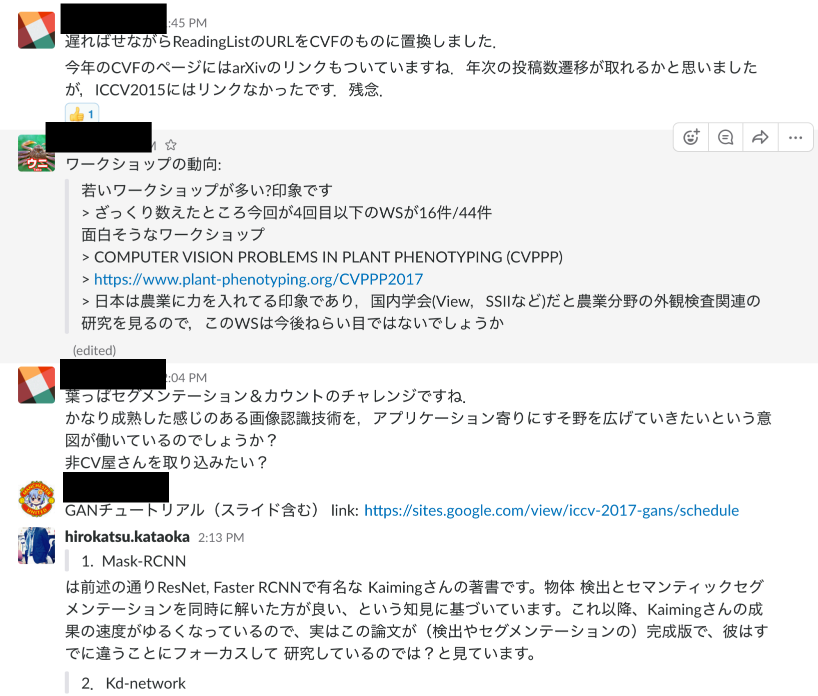
次ページの図の通り．全ての特徴量を統合するのが最も精度が高いことが判明した(38.0%)．次いでHOG2x2 (27.2%), geometry texton hist (23.5%), ssim (22.5%), dense SIFT (21.5%)であった．

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

ITツールを活用してディスカッション

一人で読むよりも，全員で協力して進めるのが現代流

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



[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。

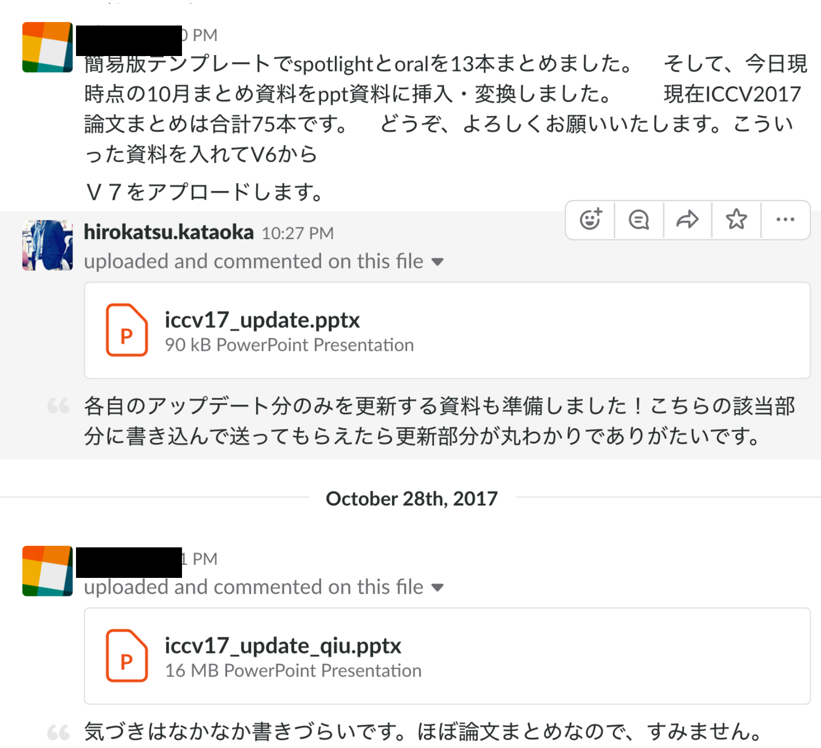
> 2. Kd-network

ICCV 2017 速報の資料作成

スライドを共有

資料をみんなで作り上げていく

- Ver.を上げていくごとに自分と他人の知見を混ぜていく
- 間接的に読んで，議論を重ねるうちに自分にも知識が定着



[redacted] PM

簡易版テンプレートでspotlightとoralを13本まとめました。　そして、今日現時点の10月まとめ資料をppt資料に挿入・変換しました。　　現在ICCV2017論文まとめは合計75本です。　どうぞ、よろしくお願いいたします。こういった資料を入れてV6から

Ｖ７をアプロードします。

hirokatsu.kataoka 10:27 PM

uploaded and commented on this file ▾

iccv17_update.pptx
90 kB PowerPoint Presentation

各自のアップデート分のみを更新する資料も準備しました！こちらの該当部分に書き込んで送ってもらえたら更新部分が丸わかりでありがたいです。

October 28th, 2017

[redacted] PM

uploaded and commented on this file ▾

iccv17_update_qiu.pptx
16 MB PowerPoint Presentation

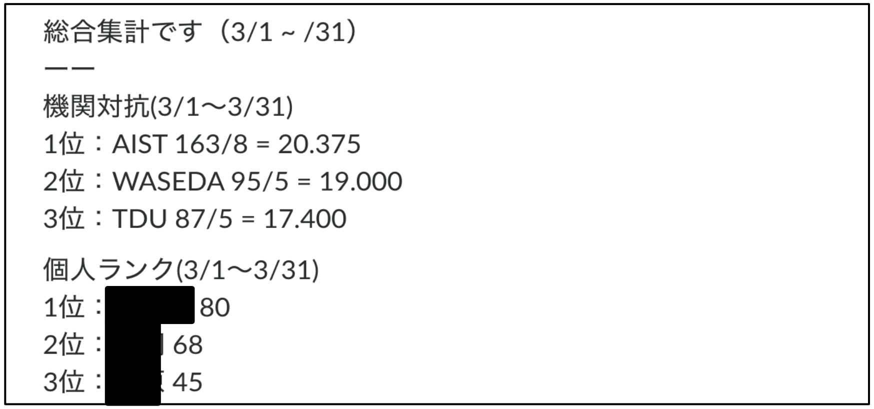
気づきはなかなか書きづらいです。ほぼ論文まとめなので、すみません。

CVPR/ICCV 2017 速報の資料作成

可視化，競争

可視化すると意識して読むようになる

- 週間，月間ランクなど
- 機関総合 (2018/3/1〜2018/3/31)　1位：163本，2位：95本，3位：87本
- 個人総合 (2018/3/1〜2018/3/31)　1位：80本，2位：68本，3位：45本



総合集計です（3/1 ~ /31）
ーー

機関対抗(3/1〜3/31)
1位：AIST 163/8 = 20.375
2位：WASEDA 95/5 = 19.000
3位：TDU 87/5 = 17.400

個人ランク(3/1〜3/31)
1位：[redacted] 80
2位：[redacted] 68
3位：[redacted] 45

気をつけていること（グループサーベイ）

集団はできる限り仲良くなるように

- リアルだとお茶会/懇親会をする
- リモートだとおしゃべりチャンネルみたいのを作る？

個人へのケアを大事にする

- 活動量が減っている個人と話す（責めずに状況を聞く）
- 活躍したら讃える

理想は全員が自由に発言できる集団

資料公開のススメ

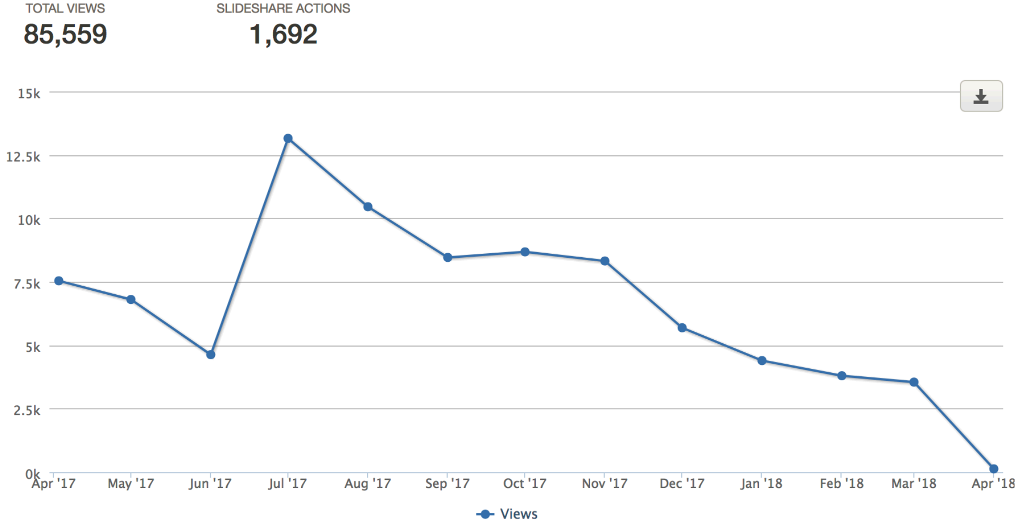
人目に触れて叩き上げる，プレゼンスを高める

– 評価を見て資料の出来栄えを判断（面白いかどうかくらいはわかる）

– 学会などでリアルに会うと反応をもらえる

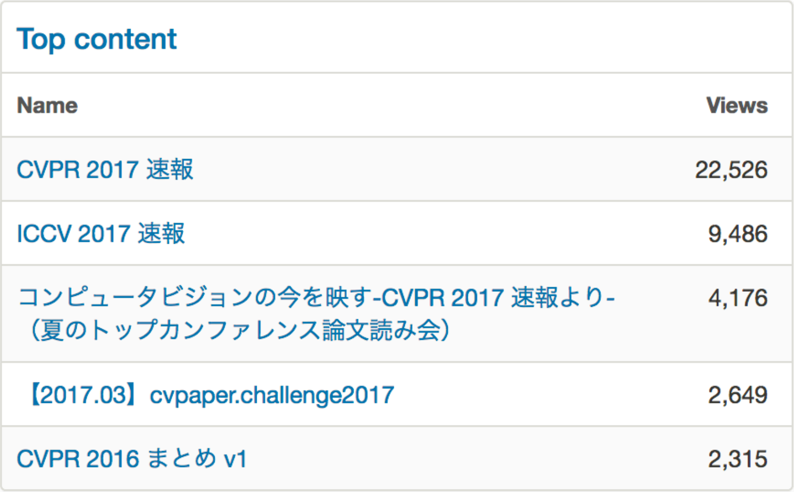
論文が通った時だけ宣伝，ではなく

普段やっていることでプレゼンスを上げる



TOTAL VIEWS 85,559

SLIDESHARE ACTIONS 1,692



Top content

Name	Views
CVPR 2017 速報	22,526
ICCV 2017 速報	9,486
コンピュータビジョンの今を映す-CVPR 2017 速報より-（夏のトップカンファレンス論文読み会）	4,176
【2017.03】cvpaper.challenge2017	2,649
CVPR 2016 まとめ v1	2,315

サーベイ法まとめ

サーベイの方法論について，個人/グループという単位で説明

- 速読と精読を組み合わせた個人サーベイ
- 組織的サーベイで早く/確実に知識を作り上げていく

もっと大事なこと

サーベイ（に限らず研究）は楽しんでやるもの！こんな楽しいことができるようになった，分からなかったことを知識として明らかにした，を知る贅沢