Mailto: cvpaper.challenge@gmail.com
Twitter: @CVpaperChalleng
SlideShare: @cvpaperchallenge

cvpaper.challenge in CVPR2015

- CVPR2015のまとめ-

片岡 裕雄, *Ph.D.*
産業技術総合研究所
知能システム研究部門 コンピュータビジョン研究グループ (CVRG)

http://www.hirokatsukataoka.net/

宮下侑大，山辺智晃（東京電機大），白壁奏馬（筑波大）
佐藤晋一，星野浩範，加藤遼，阿部香織，今成隆了，
小林直道，森田慎一郎，中村明生（東京電機大）

産総研コンピュータビジョン研究グループでは，
インターン・大学院生や博士研究員，常勤職員を募集しております

cvpaper.challengeについて

CV分野の論文サーベイチャレンジ

- 産総研，電大，筑波大による合同プロジェクト
- 2015年の取組み
  - 多読：CVPR2015論文602本約5,000ページ読破！
  - 体系化：グループ内の密な議論
  - 共有：全論文のまとめを共有中！※

※ Twitter, SlideShareもご覧ください
Twitter: @CVpaperChalleng
SlideShare: @cvpaperchallenge

なぜ，そこまでサーベイするか？

世界水準の研究のために

- 著名な研究者は流れを網羅してテーマ設定
- 歴史や最近の動向を知る
- 日常的に論文のアップデートを議論し共有

最先端技術を把握しよう！

- 知識がないと既存研究の劣化版を作りかねない
- 動向を知らないと(天才でない限り)最先端はできない
- 密なサーベイで最先端 + アップデートのサイクルに突入

なぜ，そこまでサーベイするか？

関連の取り組み CHI勉強会2015

CHI勉強会2015 http://hci.tokyo/seminar/chi2015/

The ACM CHI Conference on Human Factors in Computing Systemsの論文（Proceedingsに入っているPapers&Notes）、全485本を5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10年間続く取り組み
- 日本のHCI分野の好調を支えている

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

サーベイから議論・アイディア発想！

インプットのみならず，アウトプットまで

– 従来：サーベイのみ

従来の勉強会のカバーする範囲

– cvpaper.challenge：グループで知識を共有してアイディア発想まで

cvpaper.challengeのカバーする範囲

資料共有により，すぐに最先端を把握可能？

高効率化のための体系化・資料化

資料共有により作業効率や論文カバー率向上

- クラウドの時代なので作業を分担
- インプットとアウトプットの回転効率を飛躍的に高める

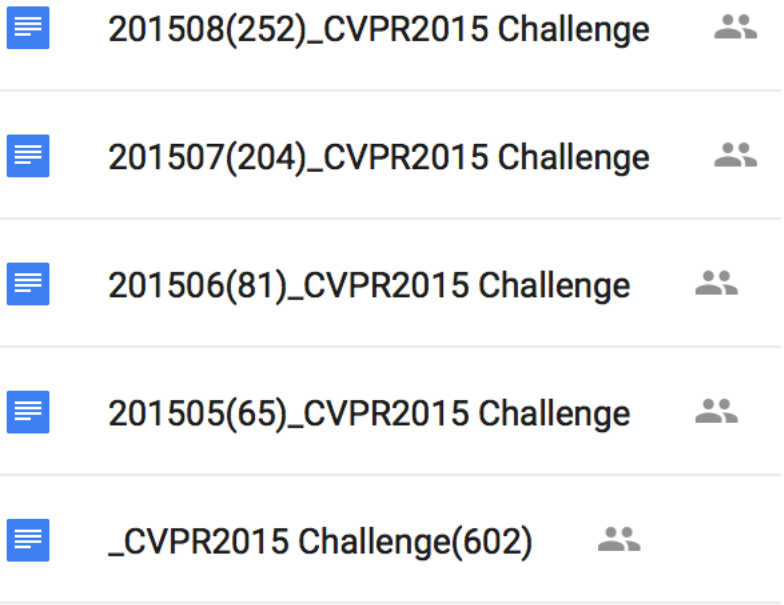
特に学部生・大学院生にこそ，最先端を知ってほしい

- 自分の後悔
  - 修士論文執筆までトップ会議の存在すら知らない状態
  - 博士課程に入るまで英語論文もほとんど読んでいない
- 単純に，スゴイ技術は面白い！

グループ内での共有方法

クラウド

- Google Driveをメンバーで同期
- 読んだ/読んでない 論文を可視化
- 成果・進捗を月別に管理



	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

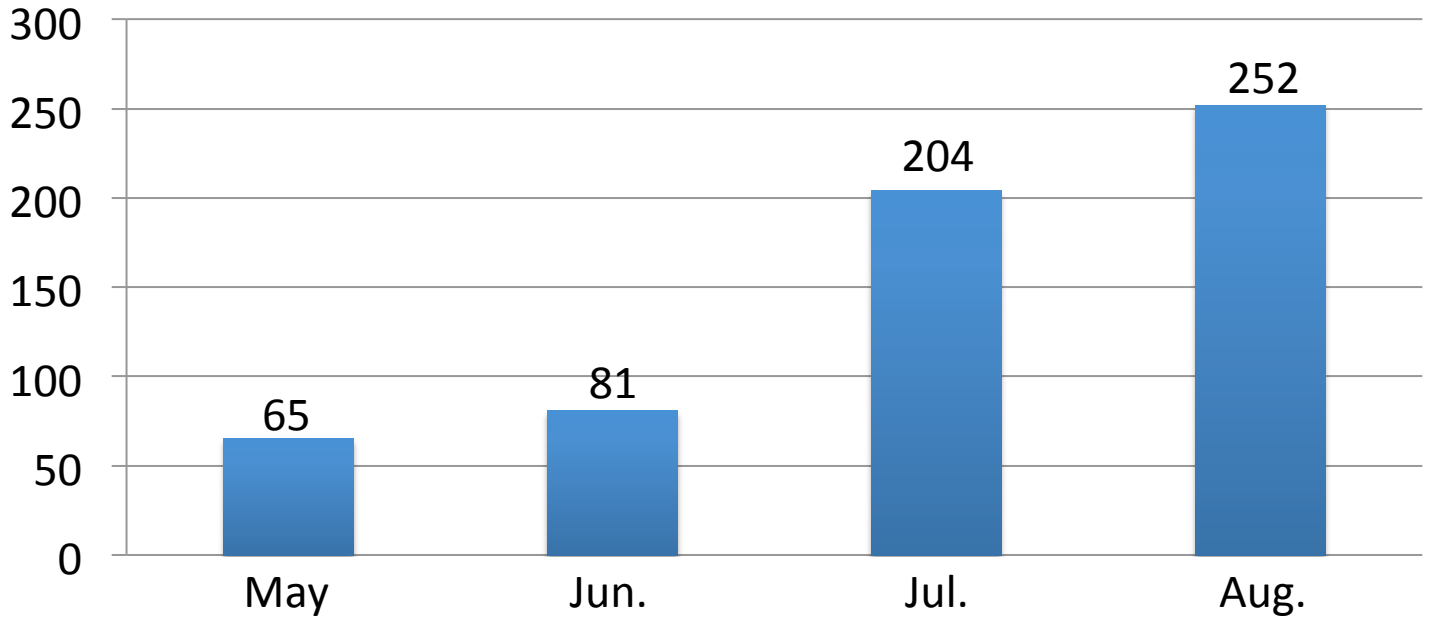
_CVPR Challenge (602)

さらに知りたい方は直接お声がけください！

論文の統計

CVPR2015を4ヶ月弱で読破！

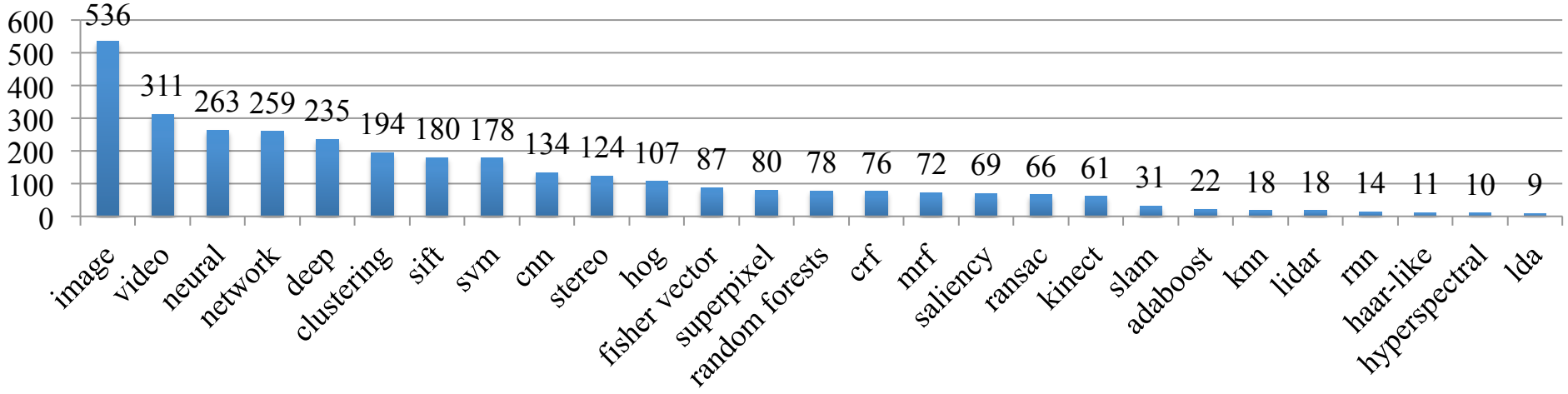
- 2015年5月7日～2015年8月25日 (110日間)
- 全602本の論文
- 5.47本/日
- 5月：65本　6月：81本　7月：204本　8月：252本



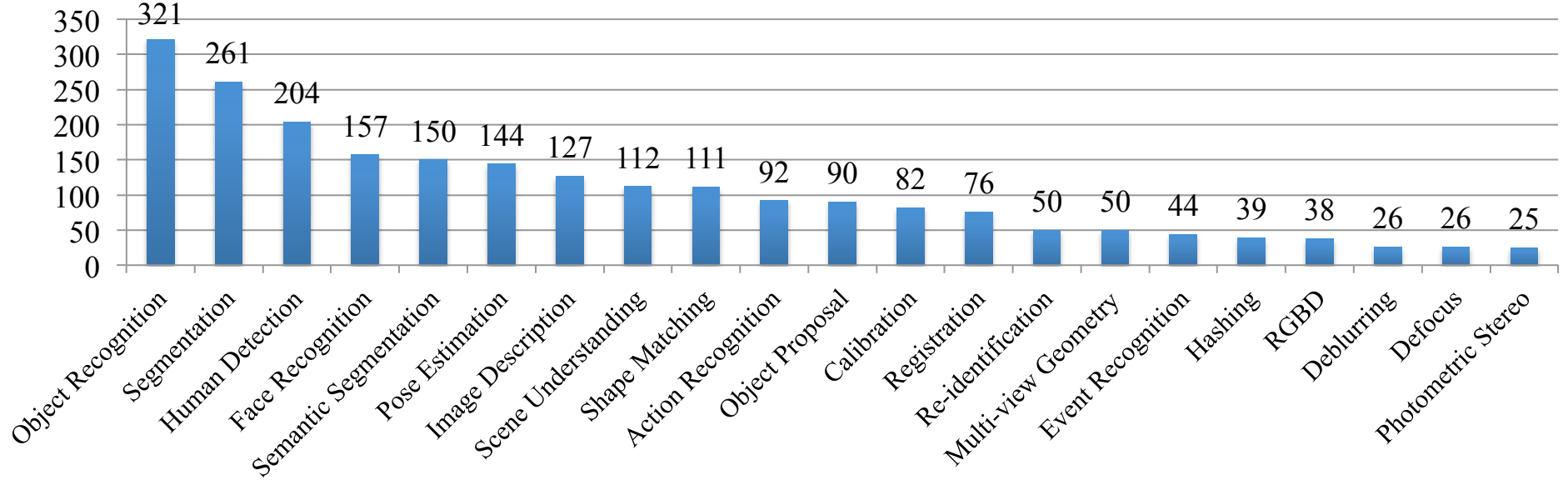

STEREO
RECONSTRUCTION
HASHING
MODELING
DEEP
ESTIMATION
DETECTION
IMAGE
LEARNING
OBJECT
RECOGNITION
SEGMENTATION
VISUAL
MODEL
MATCHING
UNDERSTANDING
SCENE
RETRIEVAL
MOTION
JOINT
MULTIPLE
SINGLE
DEPTH
LOCAL
DESCRIPTORS
DYNAMIC
DISCOVERING
INSTANCE
SIMILARITY
OBJECTS
BINARY
MODELS
FAST
BEYOND
OPTIMIZATION
PREDICTION
STRUCTURE
DEFORMABLE
SURFACE
SEARCH
CODING
MANIFOLD
ADAPTIVE
MATRIX
PRIOR
GLOBAL
FINE-GRAINED
CLASSIFICATION
LOCALIZATION
VIDEOS
IMAGES
ROBUST
FACE
CONVOLUTIONAL
USING
NETWORK
FEATURES
LARGE-SCALE
REGRESSION
TRACKING
SALIENCY
NETWORKS
SEMANTIC
POSE
NEURAL
VIDEO
BASED
REPRESENTATION
FEATURE
DATASET
EFFICIENT
METRIC
CAMERA
SCENES
SUPERVISED
ALIGNMENT
UNSUPERVISED
SUBSPACE
APPROACH
CLUSTERING
SPARSE
HIERARCHICAL
SHAPE
ACTION
VIA
CONTEXT
FLOW
HUMAN
ANALYSIS
REAL-TIME
PERSON
KERNEL
TRANSFER
REGISTRATION
MULTI-LABEL
PROPOSALS
LEARNED
VISION
UNCONSTRAINED
RANDOM
CORRESPONDENCE
ADAPTATION
REPRESENTATIONS
CORRELATION
FUSION
PEDESTRIAN
DEBLURRING
PHOTOMETRIC
WEAKLY
DATA
LOW-RANK
DOMAIN
PART
APPEARANCE
ATTRIBUTES
RE-IDENTIFICATION
BEYOND
ACCURATE
INFERENCE
SIMULTANEOUS
CLASSIFIER
STRUCTURED
DESCRIPTOR

CVPR2015の統計

ワードでヒットした論文数

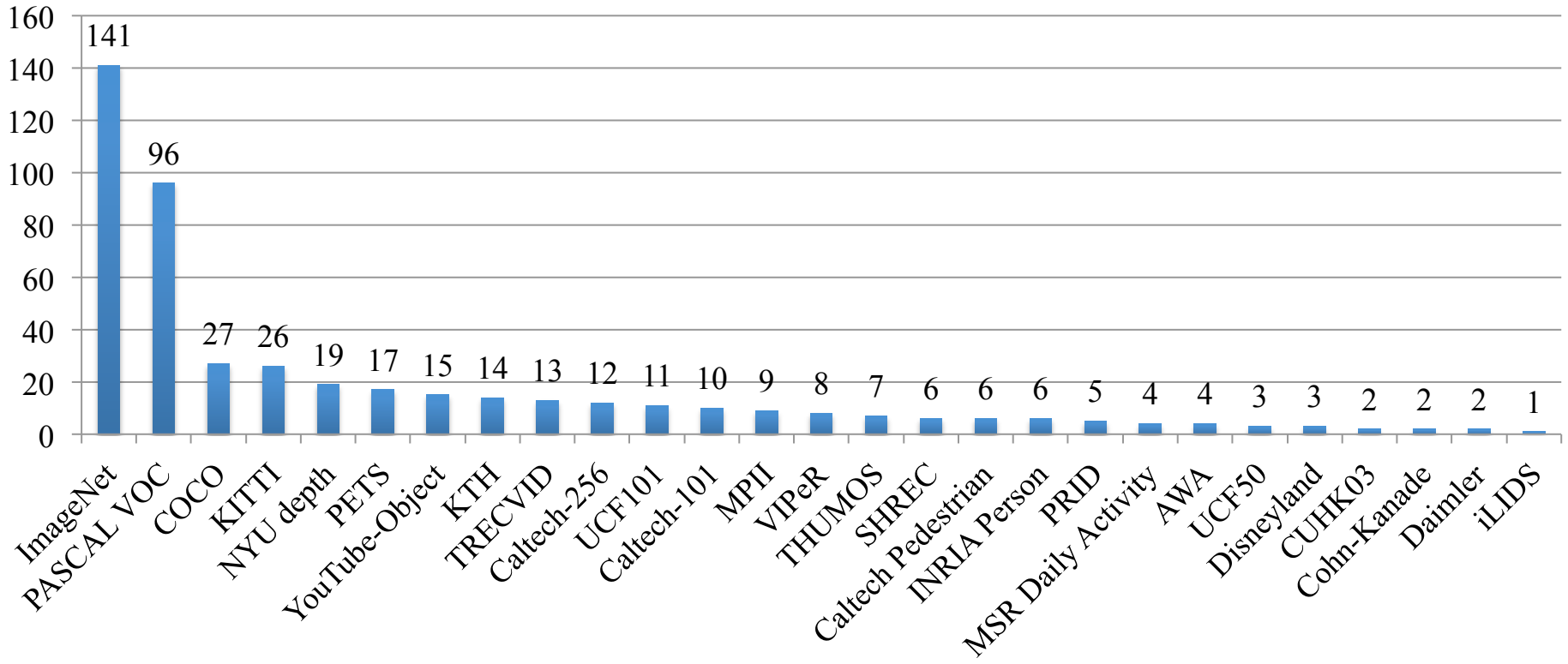


研究部門でヒットした論文数



CVPR2015の統計

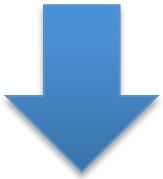
データセットでヒットした論文数



CVPR2015 最大の波

Convolutional Neural Network (CNN)の隆盛！

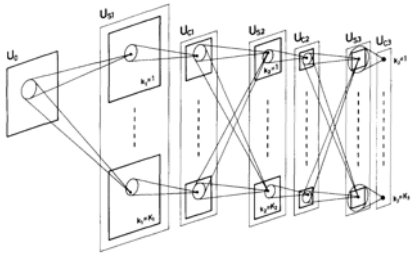
- 約250件 (全体の約40%)が深層学習の研究に関連



- 精度を出すための知識が整った
  - パラメータを抑えて，より深い構造に
  - 学習のためのデータ数を膨大に
- ライブラリ/Big Data/計算リソースが(個人レベルで)整った
  - Caffe/Torch/Pylearn2/Theano/Chainer/TensorFlow
  - ImageNet, Microsoft COCO, …

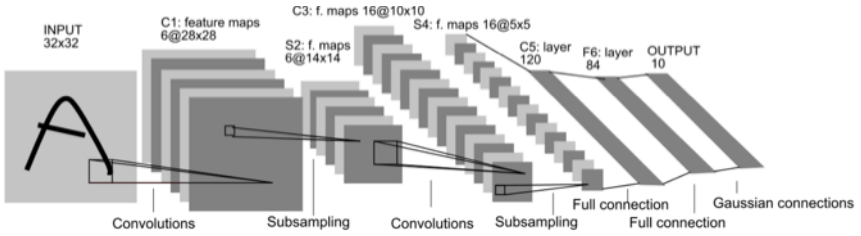
CNN発展の歴史

ネットワークアーキテクチャ



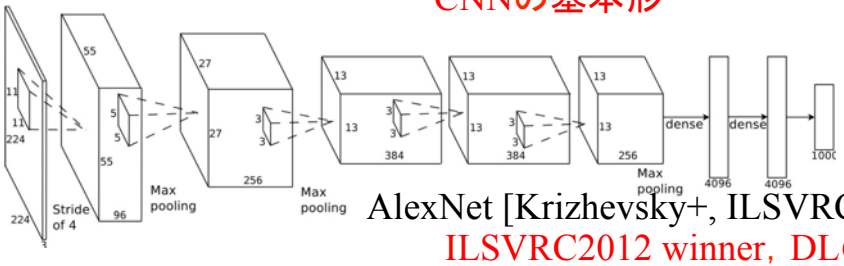
Neocognitron [Fukushima, 1980]

CNNのベースモデル



LeNet-5 [LeCun+, 1998]

CNNの基本形

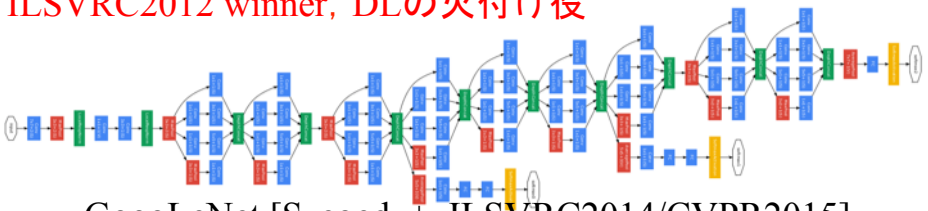


AlexNet [Krizhevsky+, ILSVRC2012/NIPS2012]

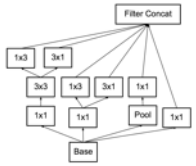
ILSVRC2012 winner，DLの火付け役

VGGNet [Simonyan+, ILSVRC2014/ICLR2015]

16/19層ネット，deeperモデルの知識

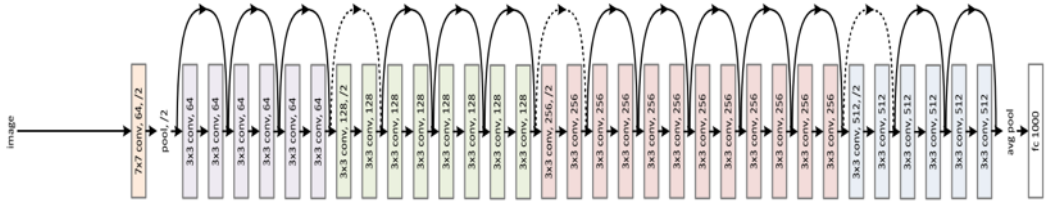


GoogLeNet [Szegedy+, ILSVRC2014/CVPR2015]



ILSVRC2014 winner，22層モデル

Inception-v3 [Szegedy+, arXiv1512]



ResNet [He+, ILSVRC2015]

ILSVRC2015 winner，152層！(実験では10^3+層も)

CVPR2015におけるCNN

CNNアーキテクチャ

GoogLeNet [Szegedy et al.]
Easily Fooled [Nguyen et al.]
Epitomic [Papandreou et al.]
Equivariance and Equivalence [Lenc et al.]
DPMs are CNNs [Girshick et al.]
DeepID-Net [Ouyang et al.]
FVs meet NN [Perronnin et al.]
CNN-DPM-NMS [Wan et al.]
Sparse CNN [Liu et al.]
CNN at Constrained Time Cost [He et al.]
Deep Image Representation [Mahendran et al.]

物体検出

Bayesian Optimization and Structured Prediction [Zhang et al.]
Weakly-supervised learning [Oquab et al.]
Adaptive Region Pooling [Tsai et al.]

人物応用

TA-CNN [Tian et al.]
FaceNet [Schroff et al.]
MPIIGaze [Zhang et al.]
Finding Action Tubes [Gkioxari et al.]
TDD [Wang et al.]
RNN for Skeleton [Du et al.]
Deeply learned face [Sun et al.]
DevNet [Gan et al.]
Deeply Learned Attributes [Shao et al.]
Deeper Look at Pedestrian [Hosang et al.]
Fusing Deep Channels [Xiong et al.]
Markerless MoCap [Elhayek et al.]

CVPR2015におけるCNN

DNN+α

Deep Pattern Mining [Li et al.]
Deep Domain Adaptation [Chen et al.]
Deep Hashing [Liong et al.]
Deep Representation for Geolocalization [Lin et al.]

セグメンテーション

Hypercolumns [Hariharan et al.]
Saliency based Multiscale Deep Features [Li et al.]

Fine-grained

Two-level Attention Models [Xiao et al.]
Hyper-class [Xie et al.]
Deep LAD [Lin et al.]

3次元

3D Deep Shape Descriptor [Fang et al.]
DeepShape [Xie et al.]

画像処理

CNN for Motion Blur [Sun et al.]
Shadow Optimization [Shen et al.]

等多数．．．
ここでは一部を紹介

認識分野の流れ

- 深層学習
  - アーキテクチャ改善，物体認識精度の向上
  - 物体検出・セグメンテーション・映像認識への移行
  - その他，各分野への拡がり
    - 属性推定，人物解析
    - ブラー除去，影領域推定，エッジ検出，
    - 3次元特徴量，特徴統合，階層的処理
    - など
- データセットの拡大

Christian Szegedy, Wei Liu, Yanqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru, “Going Deeper with Convolutions”, in CVPR, 2015.

Keywords: GoogLeNet, Convolutional Neural Networks (CNN)

概要

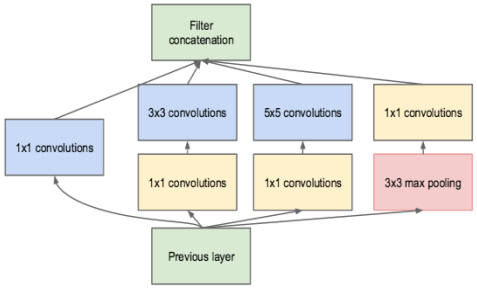
Googleの提案するILSVRC2014の識別モデル．22層から構成される．

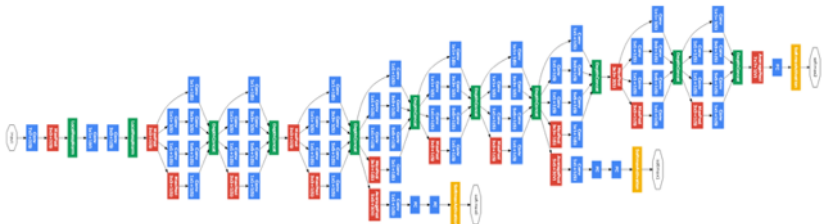
新規性・差分

ILSVRC2014にてtop-5のエラー率が6.67%で識別部門で1位になった．CPUでも動くようにパラメータ数を削減．

Links

論文
http://www.cv-foundation.org/openaccess/content_cvpr_2015/papers/Szegedy_Going_Deeper_With_2015_CVPR_paper.pdf
Caffeモデル
https://github.com/BVLC/caffe/tree/master/models/bvlc_googlenet





手法

Inceptionは前の層からの入力から”1x1”, ”1x1=>3x3”, “1x1=>5x5”, “3x3max-pool=>1x1”の畳み込みを行うモジュールで，これを再帰的に繰り返すことにより，ディープな構造を実現し，識別精度を高める．

結果

ImageNetチャレンジにて，1000クラスの識別の精度top-5のエラー率が6.67%と，参加チーム中1位の識別精度を実現した．

Anh Nguyen, Jason Yosinski, Jeff Clune, “Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images”, in CVPR, 2015.

Keywords: Convolutional Neural Networks (CNN), Compositional Pattern-Producing Network (CPPN)

概要

CNNは人間と認識の構造において異なる点を主張した．CNNの弱点に関して検証.

新規性・差分

CNNの弱点を知ることで，次の戦略を立てるために必要な準備をするという提案.

手法

CNNは画像を入力した場合には必ず出力を出すので，人間にとってはよく分からないパターンに関しても確率分布として出力してしまうため，意図していないものに関しては識別ができない．また，CNNが誤る最悪のパターンを生成して画像に埋め込むことにより，見た目に反して識別の誤りを出力してしまう．下図はCPPNにより生成したパターンや誤りの例である.

Links

論文 http://arxiv.org/abs/1412.1897
プロジェクト http://www.evolvingai.org/fooling

結果

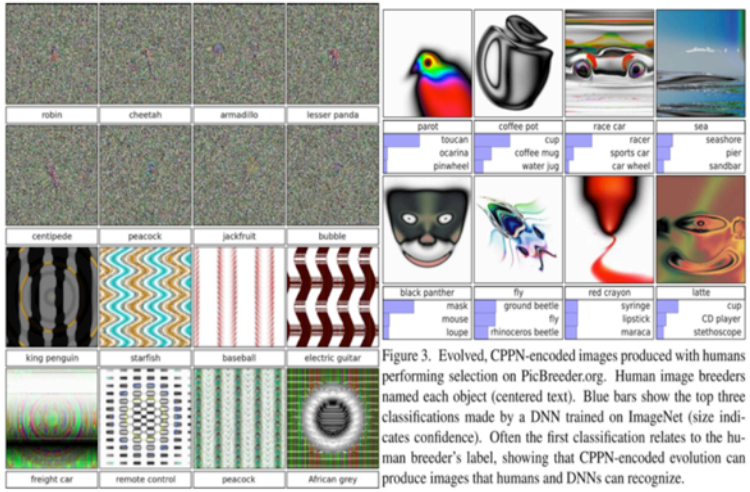
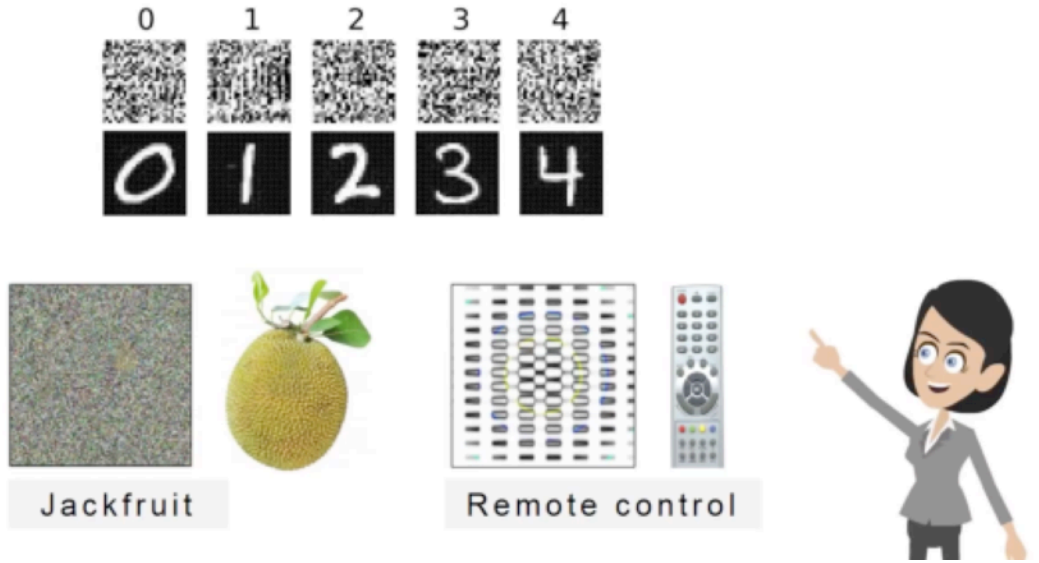


Figure 3. Evolved, CPPN-encoded images produced with humans performing selection on PicBreeder.org. Human image breeders named each object (centered text). Blue bars show the top three classifications made by a DNN trained on ImageNet (size indicates confidence). Often the first classification relates to the human breeder's label, showing that CPPN-encoded evolution can produce images that humans and DNNs can recognize.



L. Wan, D. Eigen, R. Fergus, “End-to-End Integration of a Convolutional Network, Deformable Parts Model and Non-Maximum Suppression”, in CVPR2015.

Keywords: Convolutional Neural Networks (CNN), Deformable Parts Model (DPM), Non-Maximum Suppression (NMS)

概要

DPMやCNN両者の特性を組み合わせる.

新規性・差分

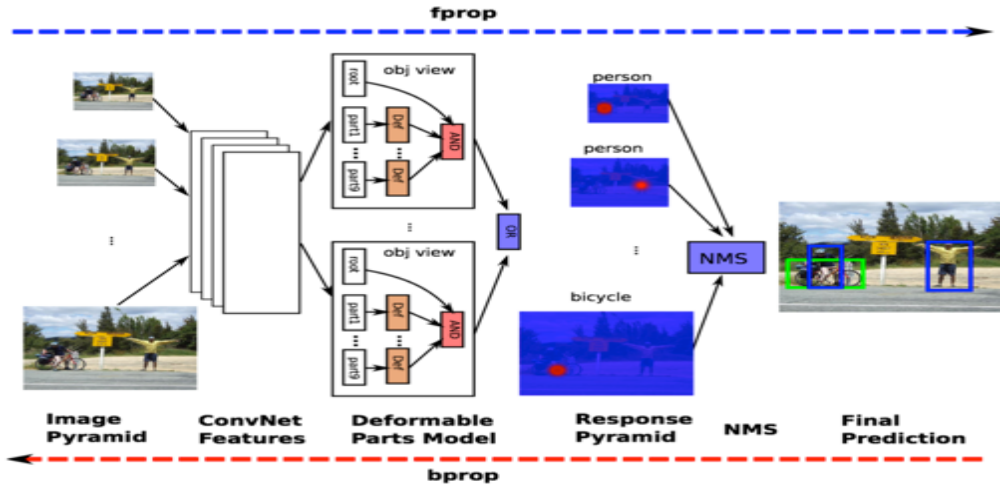
CNN, DPM, NMSの効果的な組み合わせを検討し，精度を向上させたことにある．また，Backpropagationにより精度や位置ずれのエラーを修正可能であることが分かった.

Links

論文 http://arxiv.org/pdf/1411.5309.pdf

手法

DPMは潜在変数にてパーツとその位置を保持する手法であり，CNNはニューラルネットの自動学習特徴量により非常に高度な特徴抽出を実現できる．組み合わせてバウンディングボックスの位置ずれのエラーを最小化するために構造化された損失関数を定義する．これは，Non-maximum suppression (NMS)を用いることでモデル化できることが判明した．下図は提案手法のフローであり，ピラミッド表現された画像からCNNの自動学習特徴量を取得してDPMへの入力とする．別視点からキャプチャした応答を返却し，NMSにより最終的な出力を得る．Backpropagationにより検出のエラー地は各工程にフィードバックされる.



結果

実験ではVOC2007, 2011, 2012を用いており，非常に高い精度を達成しHOG-DPM, HSC-DPM, DP-DPMや初期のR-CNNを上回る検出率を達成したが，最新のR-CNN(v4)には及ばなかった.

A. Mahendran, A. Vedaldi, “Understanding Deep Image Representations by Inverting Them”, in CVPR2015.

Keywords: CNN, Visualization

概要

CNNの可視化表現方法に関する解析.

新規性・差分

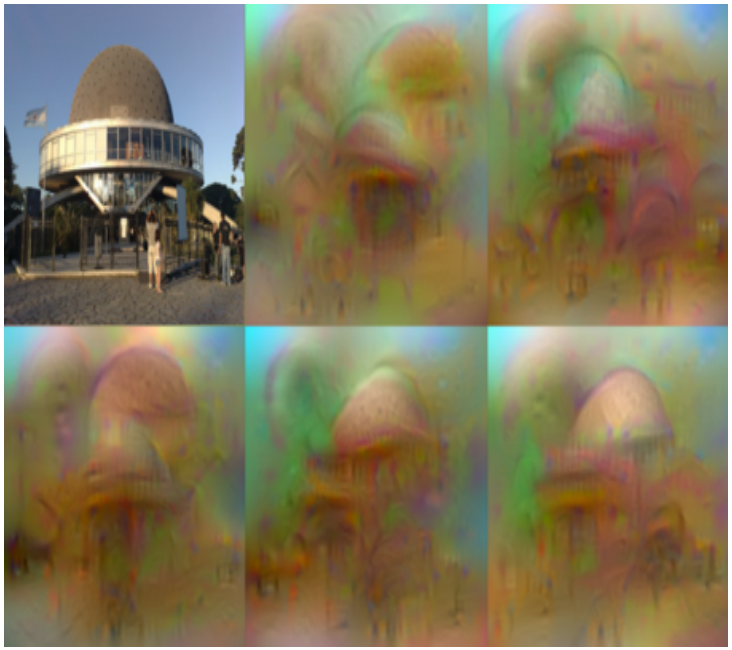
CNN特徴のみならず，従来のhand-crafted特徴に関しても可視化できるフレームワークを提案した.

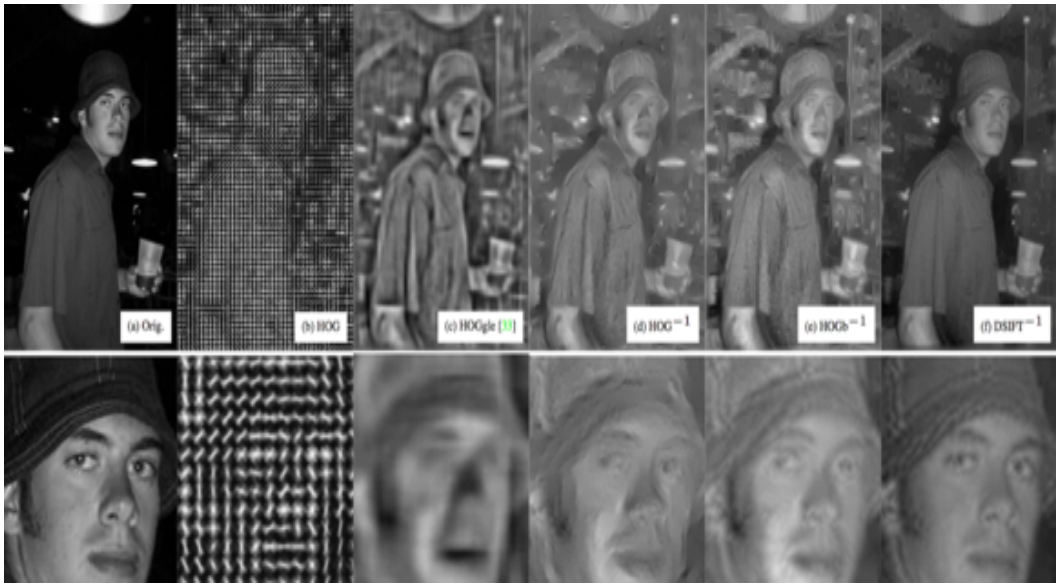
Links

論文 http://www.robots.ox.ac.uk/~vedaldi/assets/pubs/mahendran15understanding.pdf
プロジェクト https://github.com/aravindhm/deep-goggle

手法

CNNの各層の特徴を可視化する．ここでの提案は，CNN特徴量だけでなくSIFTやHOGを含めた特徴量の逆変換的表現である．画像の事前確率に従って正規化や損失に関しても評価．また，HOGglesよりも簡潔かつ効果的に可視化手法もできることを示した．特徴量からの復元も実施しており，復元率の面でもエラー率が減少している.





Florian Schroff, Dmitry Kalenichenko, James Phillbin, “FaceNet: A Unified Embedding for Face Recognition and Clustering”, in CVPR, 2015.

Keywords: FaceNet, Face Recognition, Convolutional Neural Networks (CNN)

概要

ほぼ100%の精度で顔認識できることが判明した．(99.6%)

新規性・差分

コンパクトな特徴量により，横顔や陰影を含む顔画像に対しても頑健な顔の認識が可能である．

Links

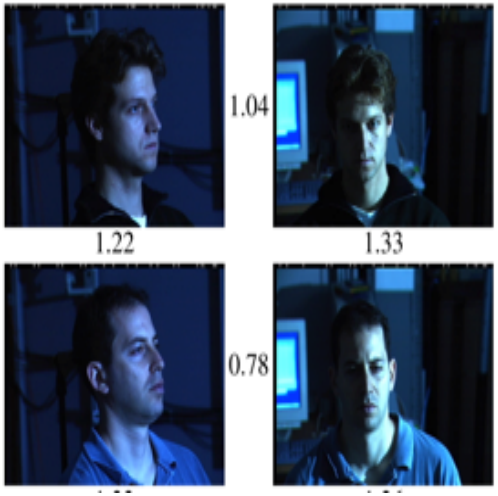
論文
http://www.cvfoundation.org/openaccess/content_cvpr_2015/papers/Schroff_FaceNet_A_Unified_2015_CVPR_paper.pdf

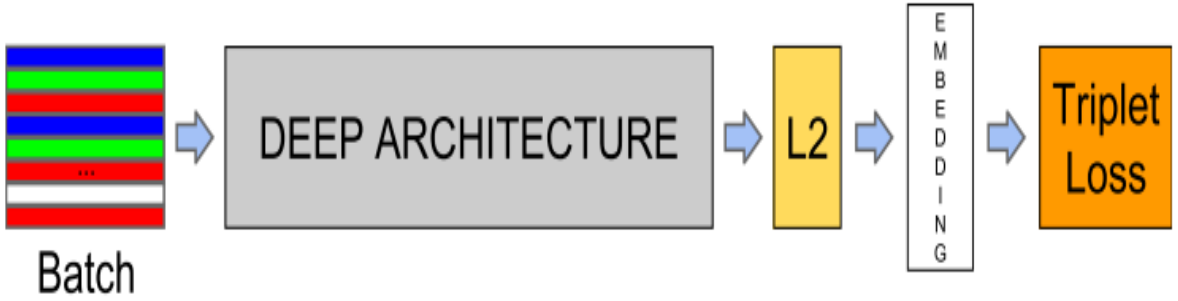
手法

FaceNetはGoogLeNetのアーキテクチャから特徴量を取り出し，L2正規化やLMNNによる損失関数により最適化して最終的な特徴量を得ている．ネットワークの構造はGoogLeNetのinception moduleの5bまでをくぐり抜けた特徴量をaverage pooling, 全結合，L2正規化により出力とする．

結果

Labeled Faces in the Wild (LFW)やYoutube Faces DBを用いて実験を行う．LFWでは99.6%の精度で，Youtube Faces DBでは95.12%の精度で顔を認識できることが判明した．DeepFace [Taigman+, CVPR2014]よりも高い精度で顔認識を実現している．





Xucong Zhang, Yusuke Sugano, Mario Fritz, Andreas Bulling, “Appearance-Based Gaze Estimation in the Wild”, in CVPR, 2015.

Keywords: Convolutional Neural Networks (CNN), MPIIGaze dataset, Gaze Estimation

概要

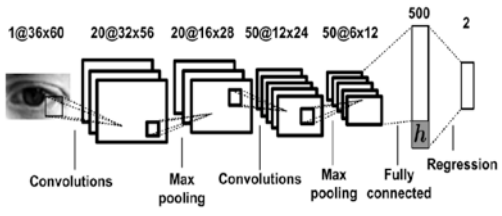
視線推定のための目領域検出にCNNを用い，さらにはMPIIGaze datasetを提案.

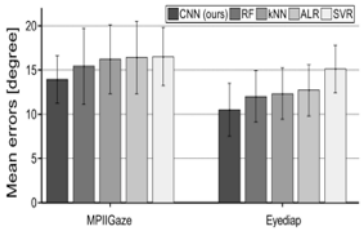
新規性・差分

視線推定のデータセットに，長期観測を導入する．さらには，目領域の検出にCNN識別器を導入した.

Links

論文
http://perceptual.mpi-inf.mpg.de/files/2015/04/Zhang15_arxiv.pdf
プロジェクト (データセットあり)
https://www.mpi-inf.mpg.de/departments/computer-vision-and-multimodal-computing/research/gaze-based-human-computer-interaction/appearance-based-gaze-estimation-in-the-wild/



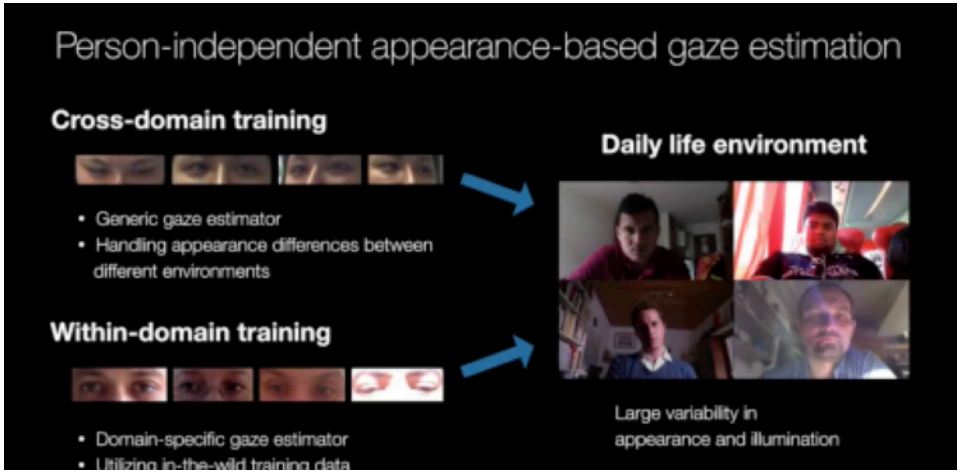


手法

視線推定を非制約(in the wild)の状況で行う．ラップトップPCに備え付けのカメラから人物の顔を撮影しデータセットを構築するが，他のデータセットと異なるのは，45日間に渡り撮影し続けるという提案．照明や姿勢の変動などに対する視線推定は，下図のCNN構造(Multimodal Convolutional Neural Networks)が非常に有効であった.

結果

中央図はMPIIGaze datasetやEyediap datasetに対する結果である.



Georgia Gkioxari, Jitendra Malik, “Finding Action Tubes”, in CVPR, 2015.

Keywords: Action Detection, Convolutional Neural Networks (CNN)

概要

行動検出問題のためのCNN特徴の統合. 可視画像・モーション画像からCNN特徴を取り出し，組み合わせることで位置まで含めた行動識別が実現.

新規性・差分

R-CNNにインスパイアされ，人物行動においても候補領域抽出や識別を行った.

Links

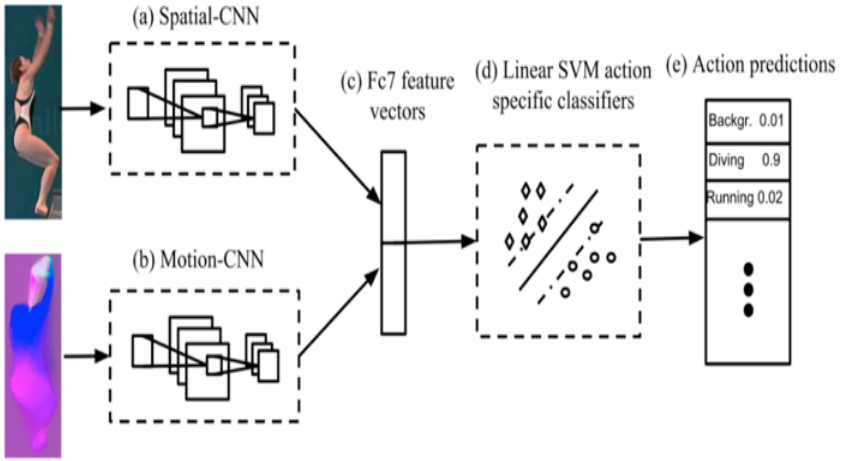
論文 http://www.cs.berkeley.edu/~gkioxari/ActionTubes/action_tubes.pdf
プロジェクト http://www.cs.berkeley.edu/~gkioxari/ActionTubes/
コード https://github.com/gkioxari/ActionTubes

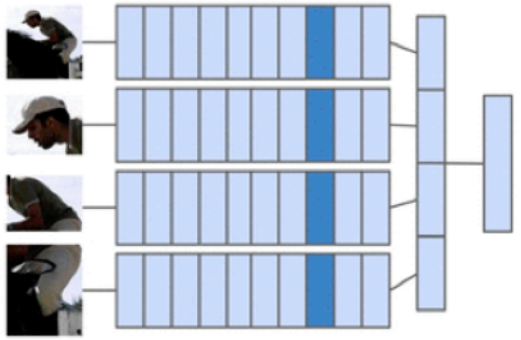
手法

候補領域の抽出とCNN特徴による行動検出を行った. 候補領域抽出にはMotion Saliencyを用い，オプティカルフロー強度の正規化により候補を抽出した．CNNによる特徴抽出のために可視画像やフロー空間を画像化して入力とした．CNNの第7層(fc7)から特徴を抽出して組み合わせることで行動検出を実現．識別器にはSVMを適用した.

結果

UCF Sportsにおけるframe-APは68.1%，映像内でタグ付けするvideo-APでは75.8%の精度で行動検出に成功した.





Yonglong Tian, Ping Luo, Xiaogang Wang, Xiaoou Tang, “Pedestrian Detection aided by Deep Learning Semantic Tasks”, in CVPR, 2015.

Keywords: Pedestrian Detection, Task-Assistant CNN (TA-CNN)

概要

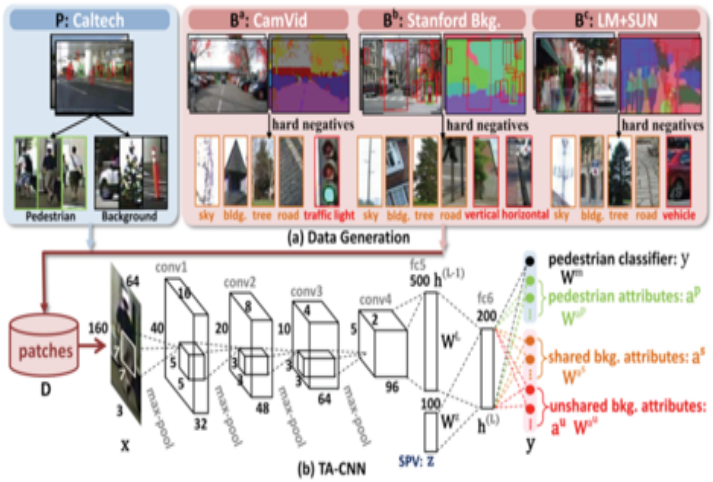
CNN特徴量と，歩行者検出のための属性特徴を組み合わせることにより，精度を高める．

新規性・差分

Deep Learningの特徴量と回帰モデルにより歩行者の属性特徴量を組み合わせることにより，精度を向上させた．

Links

論文
http://www.cv-foundation.org/openaccess/content_cvpr_2015/papers/Tian_Pedestrian_Detection_Aided_2015_CVPR_paper.pdf
プロジェクト http://mmlab.ie.cuhk.edu.hk/projects/TA-CNN/

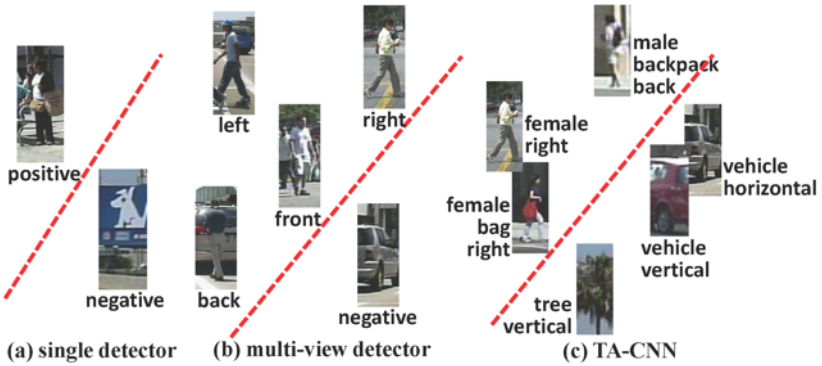


手法

TaskAssisted-CNNでは，歩行者と背景の学習以外にも属性ラベルを与える．人物は帽子やバッグ，性別，オクルージョンなど，シーンには空・木・建物・道路など．CNNの構造は6層．出力の200次元には歩行者やそのアトリビュートが含まれる．Structural Projection Vector (SPV)では，トップダウンで特徴を与え，第6層へと結合している．

結果

Caltech datasetにて実験を実施．具体的には，エラー率が22.49%のkatamari [Benenson+, ECCVW2014]や21.89%のSpatialPooling+よりも高精度な20.86%のエラー率を実現．クロスデータセットにおいても，INRIAのデータ学習でETH datasetにて34.99%のエラー率を達成．



Limin Wang, Yu Qiao, Xiaoou Tang, “Action Recognition with Trajectory-Pooled Deep-Convolutional Descriptors”, in CVPR, 2015.

Keywords: Dense Trajectories, CNN, Trajectory-pooled deep convolutional descriptor (TDD)

概要

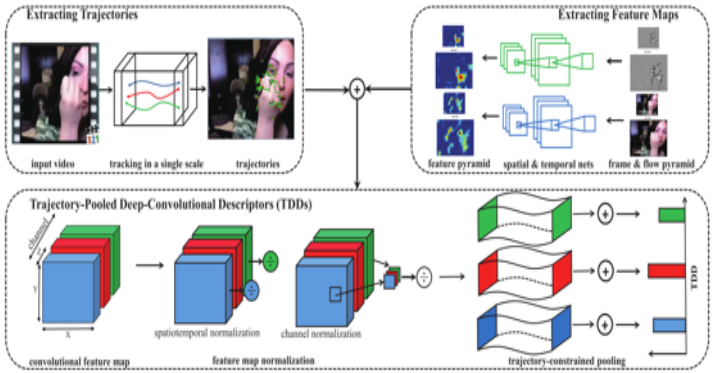
Trajectoryベースの特徴量とCNNの特徴抽出を組み合わせた行動記述子であるTrajectory-pooled deep-convolutional descriptor (TDD)を提案.

新規性・差分

行動認識のデータセットに対して最先端の精度を達成した. UCF101にて91.5%, HMDB51にて65.9%.

Links

論文
http://www.cv-foundation.org/openaccess/content_cvpr_2015/papers/Wang_Action_Recognition_With_2015_CVPR_paper.pdf
プロジェクト http://wanglimin.github.io/tdd/index.html



手法

IDTによるHand-crafted features (HOG, HOF, MBH)やDeep-learned features (TDD; trajectory-pooled deep convolutional descriptor)を組み合わせて, 最先端の精度を実現する行動認識手法を実装. CNNベースの特徴には, 画像空間とオプティカルフロー空間の入力を組み合わせたTwo-stream ConvNet [Simonyan+, NIPS2014]を適用する. 特徴を取り出す層から以降の処理を省略すること, zero-paddingを行うことがTwo-stream ConvNetに対してのチューニングである. さらに, 特徴マップの正規化や特徴のプーリングにより適切な統合方法を実現した.

結果

右の表が各特徴量とその精度である. HMDB51とUCF101を用いており, どちらも最先端の精度を実現した. また, IDTやTwo-stream ConvNetsからの精度向上が見られ, 組み合わせによりさらに良好な精度を実現.

Algorithm	HMDB51	UCF101
HOG [31, 32]	40.2%	72.4%
HOF [31, 32]	48.9%	76.0%
MBH [31, 32]	52.1%	80.8%
HOF+MBH [31, 32]	54.7%	82.2%
iDT [31, 32]	**57.2%**	**84.7%**
Spatial net [24]	40.5%	73.0%
Temporal net [24]	54.6%	83.7%
Two-stream ConvNets [24]	**59.4%**	**88.0%**
Spatial conv4	48.5%	81.9%
Spatial conv5	47.2%	80.9%
Spatial conv4 and conv5	**50.0%**	**82.8%**
Temporal conv3	54.5%	81.7%
Temporal conv4	51.2%	80.1%
Temporal conv3 and conv4	**54.9%**	**82.2%**
TDD	63.2%	90.3%
TDD and iDT	**65.9%**	**91.5%**

Yao Li, Lingqiao Liu, Chunhua Shen, Anton van den Hengel, “Mid-level Deep Pattern Mining”, in CVPR, 2015.

Keywords: Convolutional Neural Networks(CNN), Pattern Mining, Association Rules

概要

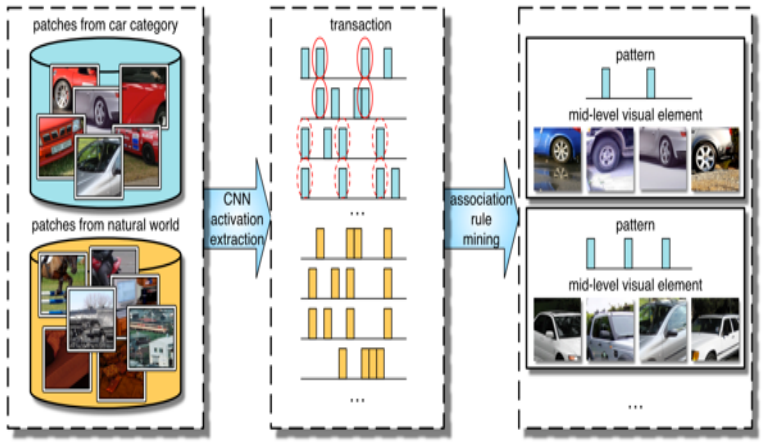
Deep Learningによる特徴量をマイニングすることにより，有効な特徴量を特徴空間中から抽出できる.

新規性・差分

CNNの特徴量をマイニングによりサブセットに落とし込むことができる．次元の削減を実行しつつ高精度な認識ができる.

Links

論文
http://www.cv-foundation.org/openaccess/content_cvpr_2015/papers/Li_Mid-Level_Deep_Pattern_2015_CVPR_paper.pdf
プロジェクト https://github.com/yaoliUoA/MDPM



手法

Caffeによる実装，モデルはAlexNetを用いており，第6層の全結合層から特徴量を取り出す．マイニングにはAssociation Rulesを用いており，探索空間は上位N次元の数値を用いる．Association Rulesでは次元数が多いと探索が非常に困難になるため上位N次元と限定している．同クラスの物体であれば特徴の強度分布も類似するという戦略でサブセットを作成.

結果

表はMIT Indoor Datasetに対する提案法と従来法の比較．提案法の50の要素を用いた場合では69.69%の精度であり，従来法のいずれの結果よりも高精度.

Method	#elements	Acc(%)
D-patch [1]	210	38.10
BoP [3]	50	46.10
miSVM [5]	20	46.40
MMDL [6]	11	50.15
D-Parts [4]	73	51.40
RFDC [7]	50	54.40
DMS [2]	200	64.03
LDA-Retrained	20	58.78
LDA-Retrained	50	62.30
LDA-KNN	20	59.14
LDA-KNN	20	63.93
Ours	20	68.24
Ours	50	**69.69**

V. E. Liong, J. Lu, G. Wang, P. Moulin, J. Zhou, “Deep Hashing for Compact Binary Codes Learning”, in CVPR2015.

Keywords: Deep Hashing, Binary Coding

概要

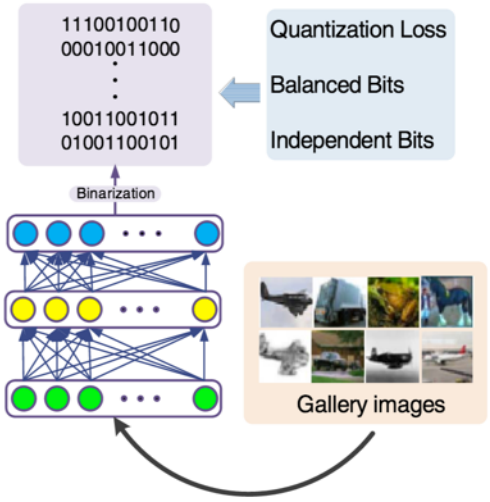
バイナリハッシング(Binary Hashing)にディープラーニングを適用したDeep Hashing(DH)を提案.

新規性・差分

CNNの物体に対する特徴類似性を，バイナリハッシングに導入してハッシングの性能を高めることに成功した.

Links

論文
http://www.cv-foundation.org/openaccess/content_cvpr_2015/papers/Liong_Deep_Hashing_for_2015_CVPR_paper.pdf

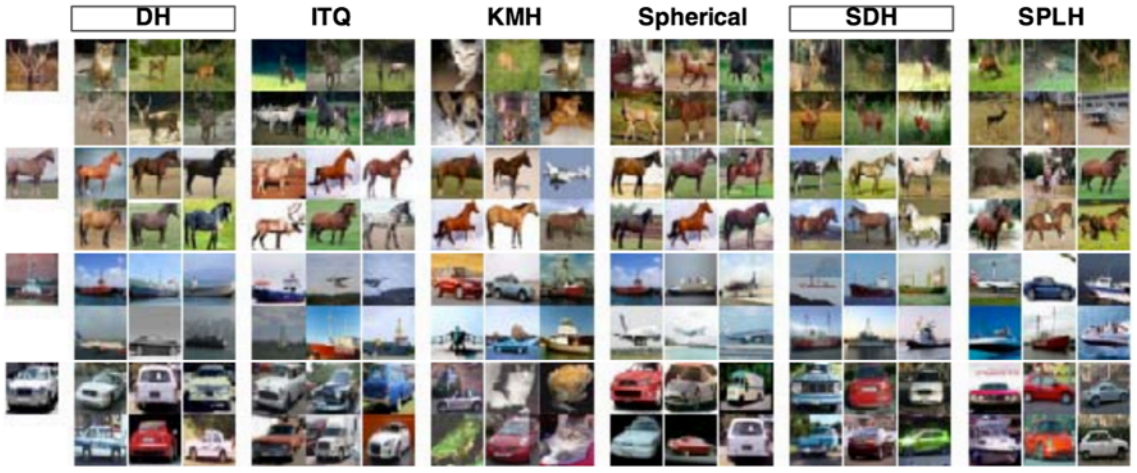


手法

CNNの特徴類似性を用いて画像の主要な成分を取り出して類似度の高いハッシング値にしようというトライ.
各畳み込み層ではProjection matrixとバイアスベクトルを表現，最終層を取りだしてバイナリ出力.
また，supervised deep hashing (SDH)も同時に提案する.
各画像から特徴抽出してペアを作成し，positive(対象と同じクラス)は距離が近くなるように，negative(対象とは異なるクラス)は距離が遠くなるようにハッシング.

結果

Unsupervised, supervisedそれぞれ評価した結果が下の表である．データにはCIFAR-10を用いており，unsupervisedな手法においてスタンダードな手法であるPCA-ITQ[Gong+, CVPR2011]からの精度向上が見られる．Supervisedな手法においてはいずれも最高性能を実現した.



Bharath Hariharan, Pablo Arbelaez, Ross Girshick, Jitendra Malik, “Hypercolumns for Object Segmentation and Fine-grained Localization”, in CVPR, 2015.

Keywords: Segmentation, Fine-grained Localization, Convolutional Neural Networks (CNN)

概要

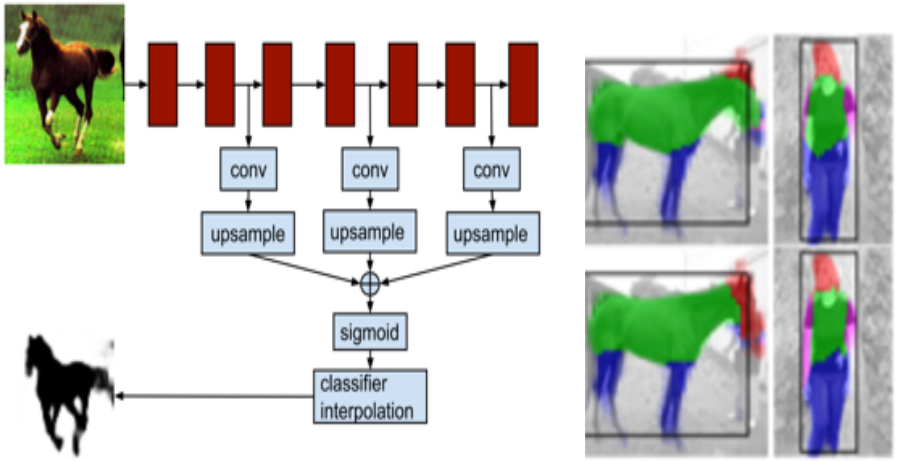
CNNにより詳細な(Fine-grained)物体のセグメンテーションと位置合わせに取り組む.

新規性・差分

R-CNNはCNNの特徴量を取り出しているが，物体検出やセグメンテーションのための特徴としては荒いので，途中の層の情報も用いることで物体検出からセグメンテーションにまで応用を広げている.

Links

論文
http://www.cs.berkeley.edu/~rbg/papers/cvpr15/hypercolumn.pdf



手法

Hypercolumnsによる特徴表現を実現した．CNNの特徴取得としては，第2プーリング層(pool2)，第4畳み込み層(conv4)，第7層の全結合層(fc7)である．これらの組み合わせにより，位置情報の正確さと属性情報推定の精度を確保することができる．左図はHypercolumnsによる表現であり，このフレームワーク内にてpool2+conv4+fc7の特徴統合を行いセグメンテーションとローカライズを同時に行う.

結果

右図はセグメンテーションの結果を表した図である．上の行が従来手法によるセグメンテーション，下の行が提案手法によるセグメンテーションの結果である.

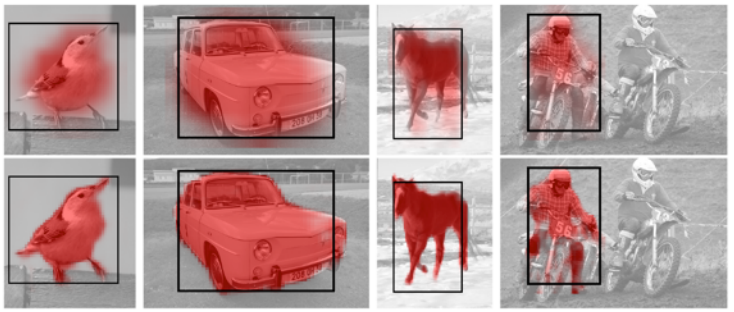


Figure 4. Figure ground segmentations starting from bounding box detections. Top row: baseline using *fc7*, bottom row: Ours.

J. Sun, W. Cao, Z. Xu, J. Ponce, “Learning a Convolutional Neural Network for Non-uniform Motion Blur Removal”, in CVPR2015.

Keywords: CNN, Motion Blur Removal

概要

カメラ撮影時の手ぶれなどにより発生する不規則なモーションブラーを，CNNを用いて解析し，補正する研究.

新規性・差分

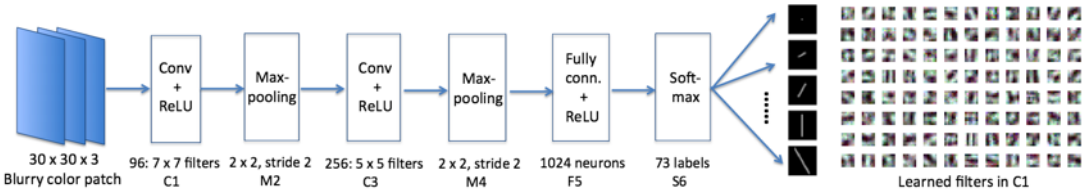
モーションブラーの推定のためにネットワークアーキテクチャを新規に考案したことや，ペア学習を行い，モーション場を推定する手法を考案した.

Links

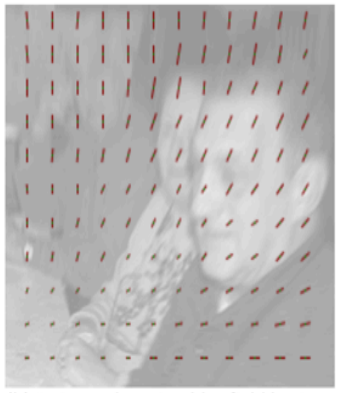
論文 http://arxiv.org/pdf/1503.00593.pdf

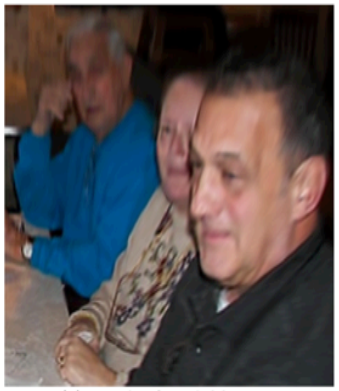
手法

提案手法ではパッチを設定して，CNN特徴によりモーションカーネルの確率を計算する．さらに，Markov Random Field (MRF)も用いることでパッチ内の確率場の計算をデンスに行うことができる．モーションフィールドを推定するために，パッチサイズはオーバーラップを含む30x30pixelsに設定する．このモーションを学習するために，140万のブラーを含むパッチとそのブラーの方向・強度をペアとしてCNNの学習に与える．CNNは6層構成.




(a) Input image


(b) Estimated motion blur field by CNN


(c) Result after deblurring

結果

入力のブラーを含む画像と出力のデブラーした結果が表示されている．中央にモーションブラーが推定された画像が表示されている.

Soonmin Hwang, Jaesik Park, Namil Kim, Yukyung Choi, In So Kweon, “Multispectral Pedestrian Detection: Benchmark Dataset and Baseline”, in CVPR2015.

Keywords: CNN, Motion Blur Removal

概要

歩行者検出のデータセットを，マルチスペクトルカメラを用いて構築した.

新規性・差分

従来の可視画像のみならず，チャネルを増やすことにより夜間の認識や認識率向上を実現.

Links

論文 http://rcv.kaist.ac.kr/ykchoi/papers/CVPR2015_MutispectalPedestrian.pdf
プロジェクト https://sites.google.com/site/pedestrianbenchmark/

データセット

車両から撮影した640×480，20Hzのマルチスペクトル画像で構成密なアノテーション，1182の対象物に対して人，人々，cyclistと手動でパラメータを付加

Fabian Caba, Victor Escorcia, Bernard Ghanem, Juan Carlos Niebles, “ActivityNet: A Large-Scale Video Benchmark for Human Activity Understanding”, in CVPR, 2015.

Keywords: CNN, Motion Blur Removal

概要

行動認識に関しても，ImageNetのように大規模化を図り，ActivityNetを構築した.

新規性・差分

現在までの行動認識データセットでは，あるドメインに限定していたが，ここではデータや行動のバリエーションを格段に増加させた.

Links

論文
http://www.cv-foundation.org/openaccess/content_cvpr_2015/papers/Heilbron_ActivityNet_A_Large-Scale_2015_CVPR_paper.pdf
プロジェクト http://activity-net.org/

データセット

従来までの行動認識のデータセットは単純な行動に限定されていたが，ここではさらにバリエーションやトリミングされていない行動データを拡張することで，行動認識の認識率や行動位置まで含めた認識(行動検出)の研究を加速させようとする試み．さらに難しい問題設定もできるよう，膨大なデータを準備した．データは階層的に構成されており，より上位の行動クラスの方がより長い行動(イベントに近い)を示す傾向にある．ここで，トリミングされたデータは203クラス，トリミングされていないデータは137クラス，合計849時間もの行動データを含んでいる.

3次元画像処理研究

DynamicFusionをはじめとした3D分野のアップデート！

- CVPR Best Paper！非剛体の物体をリアルタイムで3次元再構成
- データベースの更新
- 3次元特徴量にCNNを適用
- 距離画像の生成
- 3次元セマンティックセグメンテーション
- SLAM

等多数．．．
3Dは ほんの少しを紹介

Richard Newcombe, Dieter Fox, Steve Seitz, “DynamicFusion: Reconstruction and Tracking of Non-rigid Scenes in Real-Time”, in CVPR2015.

Keywords: Dynamic Fusion

概要

Kinectなどのセンサから時系列距離画像を入力として，リアルタイムに3次元再構成を実現する．6Dのモーションフィールドをリアルタイムで復元可能.

新規性・差分

非剛体の変換や統合を可能にした．また，Volumetric warpを効率化しリアルタイムに計算できるようにした.

手法

潜在幾何学的表面に対して非固定的に変形シーンを分解することによって剛性正準空間$S \subseteq R^3$内に再構築し，フレーム毎にライブ映像の物体表面を体積ワープ空間に変換する.

・体積モデルに変換するワープ空間パラメータの推定

・推定したワープ空間を介して，ライブ映像の深度を正空間へ融合

・ワープフィールド構造に新たにキャプチャした形状を適宜追加

Links

論文：
http://grail.cs.washington.edu/projects/dynamicfusion/papers/DynamicFusion.pdf

YouTube動画：
https://www.youtube.com/watch?v=i1eZekcc_lM&feature=youtu.be

プロジェクトページ：
http://grail.cs.washington.edu/projects/dynamicfusion/

Jared Heinly, et al., “Reconstructing the World* in Six Days *(As Captured by the Yahoo 100 Million Image Dataset)”, in CVPR2015.

Keywords: 3D Reconstruction, Yahoo 100 Million Image Dataset

概要

単一のコンピュータを用いて都市規模のモデリングから世界規模のモデリングまで大規模なモーションフレームワークの構造をデータ拡張し最先端の技術で作成する

新規性・差分

クラウドにより収集された1億枚を超える画像群から世界中に点在する大規模な構造物を3次元復元する問題.

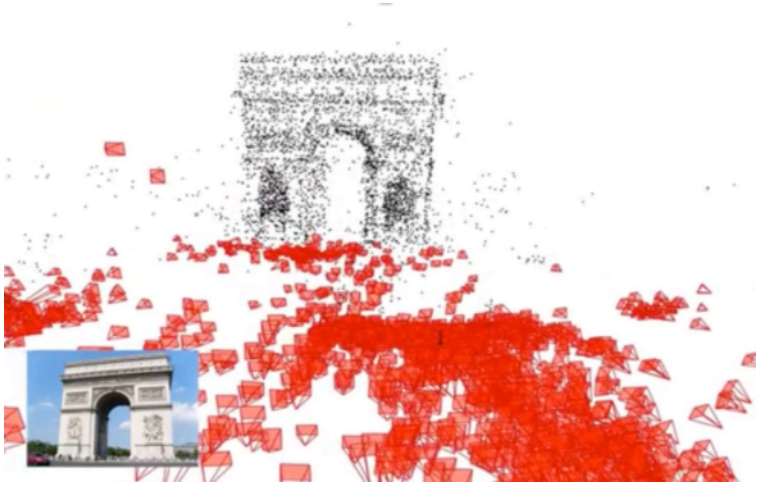
手法

・データの関連付け(頑健性，拡張性，完全性を考慮)
・ディスクからの画像を連続して一度だけ画像読み込む
・必要に限りメモリ内画像を保管する

Links

プロジェクトページ：
http://jheinly.web.unc.edu/research/reconstructing-the-world-in-six-days/

Red = Geolocations of Registered Images



Shuran Song, Samuel P. Lichtenberg, Jianxiong Xiao, “SUN RGB-D: A RGB-D Scene Understanding Benchmark Suite”, in CVPR2015.

Keywords: Scene Understanding, SUN RGB-D

概要

シーン認識SUN397をRGB-Dデータに拡張したSUN RGB-Dを提案．SUN RGB-Dでは屋内環境のシーン認識やセグメンテーションなど複数のチャレンジを設定している．

新規性・差分

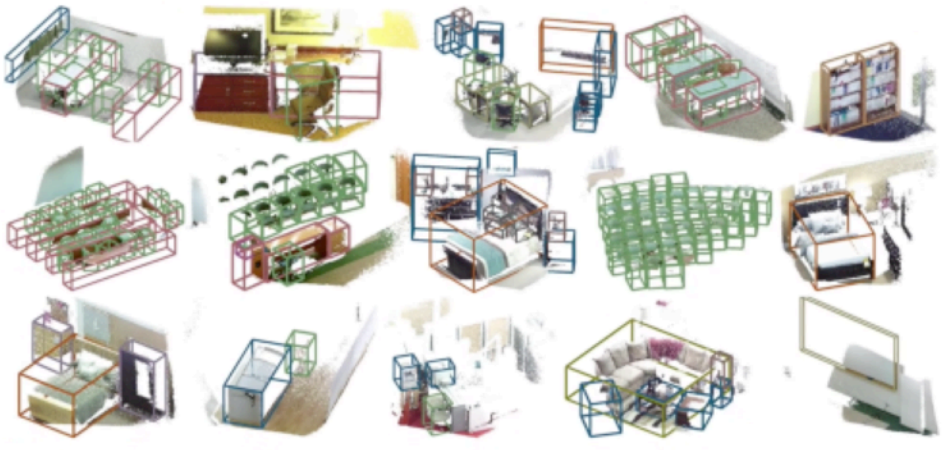
屋内環境データセットにて，大規模なデータを構築した．

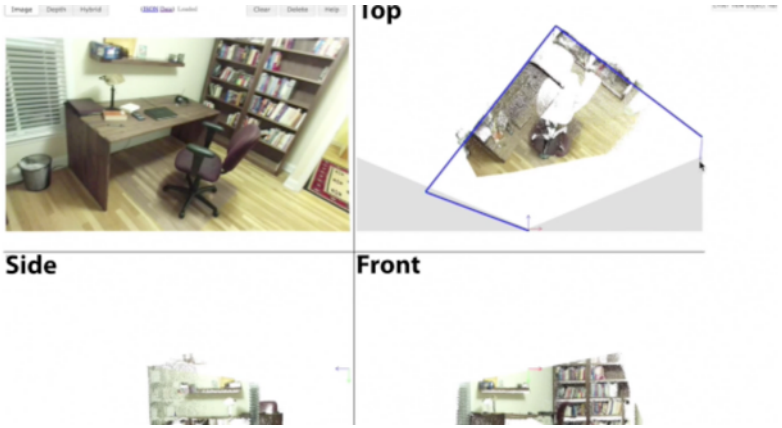
Links

論文 http://rgbd.cs.princeton.edu/paper.pdf
プロジェクト (動画，データセットなどあり) http://rgbd.cs.princeton.edu/

手法

屋内環境を3次元で捉えた大規模データセットとして提案した．総計で10,335枚ものRGB-D画像を取得しており，Scene Categorization, Semantic Segmentation, Object Detection, Room Layout Estimation, Total Scene Understanding といった3次元シーン認識における重要な課題を含んでいる．





Srinash. Sridhar, Franziska Mueller, Antti Oulasvirta, Christian Theobalt, “Fast and Robust Hand Tracking Using Detection-Guided Optimization”, in CVPR2015.

Keywords:

概要

デプスの入力，Quad-treesによるダウンサンプリング，Part-labelsによる各関節位置表現により3次元ハンド追跡.

新規性・差分

Gaussian Mixtureによるハンドモデルの効率的な表現.

Links

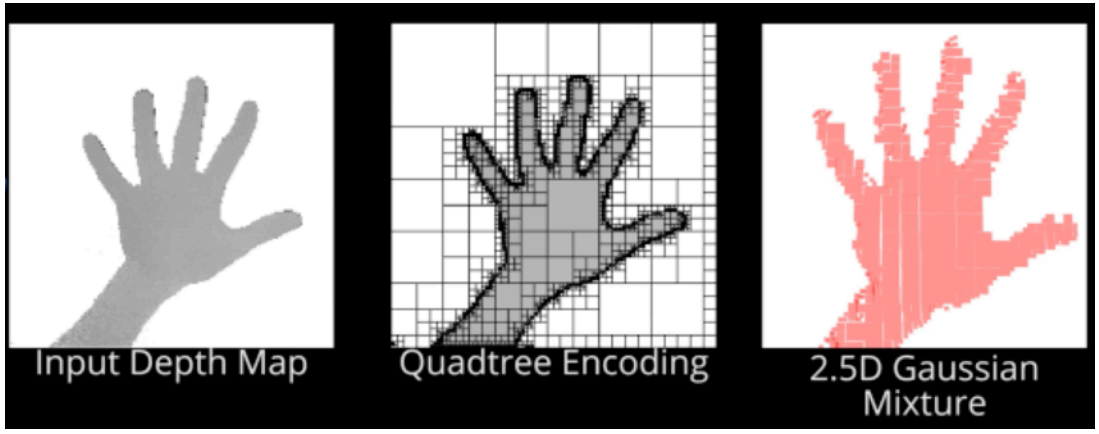
論文
http://handtracker.mpi-inf.mpg.de/projects/FastHandTracker/content/FastHandTracker_CVPR2015.pdf
Project
http://handtracker.mpi-inf.mpg.de/projects/FastHandTracker/

手法

Depth を入力としてQuadtree Encodingにより，手の分岐点やエッジ位置の重要度を高くし，2.5D画像におけるGaussian Mixtureモデルを生成.

結果

Depth のみを使用してエネルギー最適化するよりも，検出情報を事前情報として加えた方が精度が高くなりリアルタイムで手の追跡が可能となる．GPUを使うことなく50FPSで動作する.



CVPR2015の論文を読んで

所感

- キツかった，とにかくキツかった
- 5月，先が見えない (一日中読んでも4本)
- 6月，まだまだ長い
- 7月，確実に読めるようになってくる (一日最大20本)
- 8月，視野が広がる感覚と読破の自信
- 結局読んで得るものが非常に大きかった，自分も含めメンバーの成長

> 最近の中村研の状況

- 次もやりましょう！と言われた時は素直に嬉しかった

CVPR2015の論文を読んで

一番強く感じたのは．．．

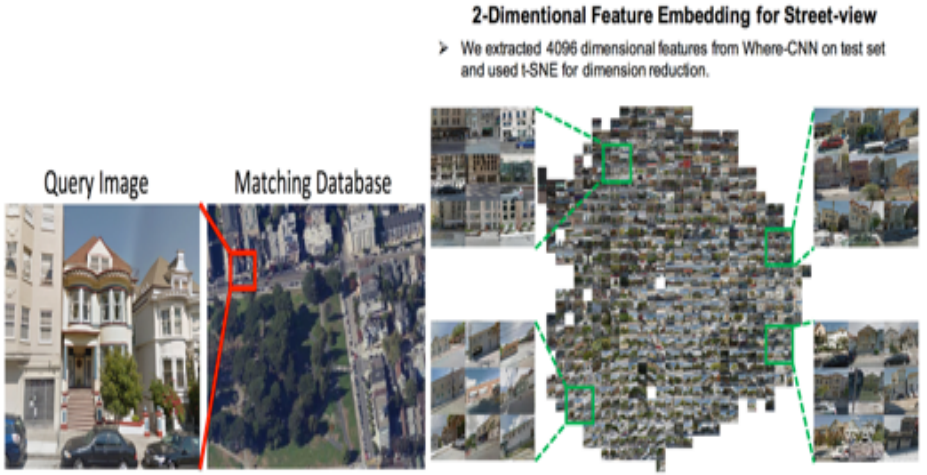
問題設定が出来るようになること！

問題設定をした論文 (1)

Deeply Learned Attributes[Shao+, CVPR2015]

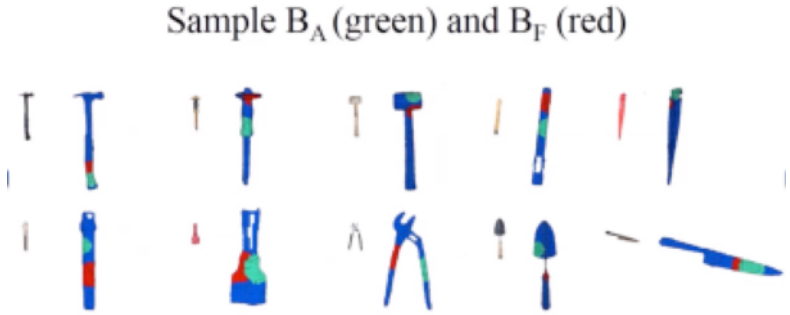
混雑環境解析の複数属性認識
CNNやデータセットも考案

Ground-Aerial Images[Lin+, CVPR2015]



地上画像から航空画像の位置推定

Understanding Tool Use[Zhu+, CVPR2015]



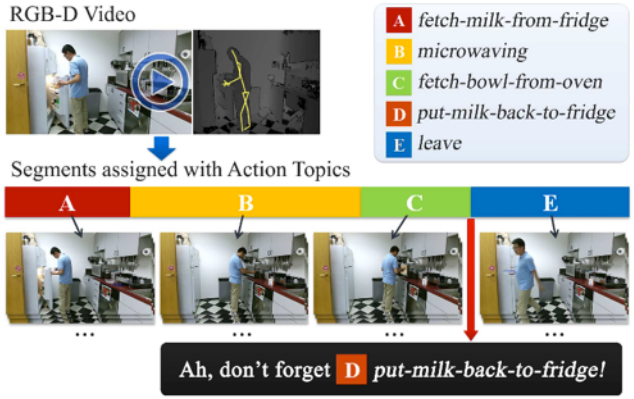
道具の使い方(Function)認識

Ped. Detection without Real-data[Zhu+,CVPR2015]

リアルデータなしの学習により歩行者検出の精度を向上

問題設定をした論文 (2)

Watch-n-Patch[Wu+, CVPR2015]

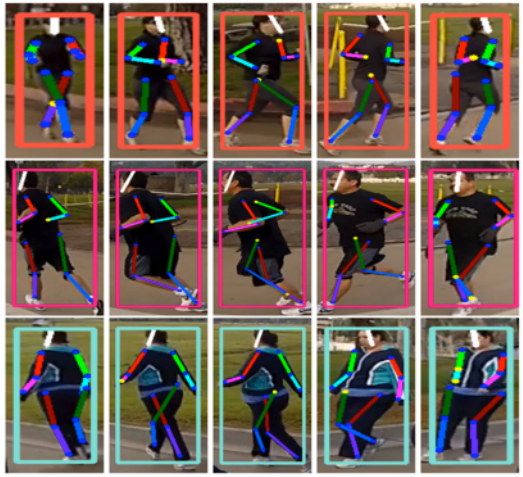


行動間の共起性を確認

3D ShapeNet[Wu+, CVPR2015]

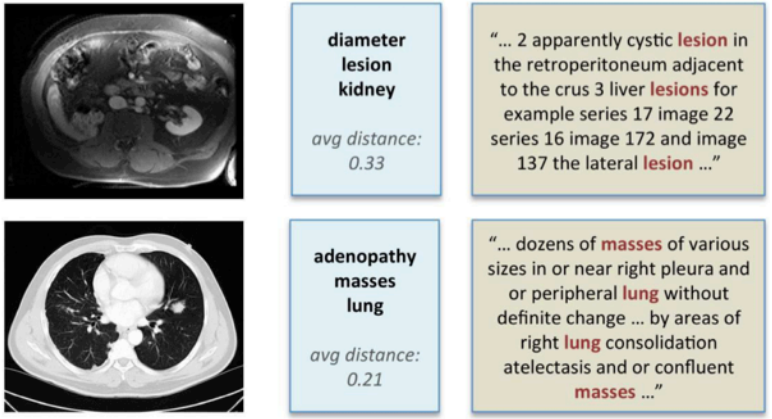
3次元特徴量とデータセットを提案

Fine-grained Ped. Dataset[Hall+, CVPR2015]



歩行者の詳細な属性を推定

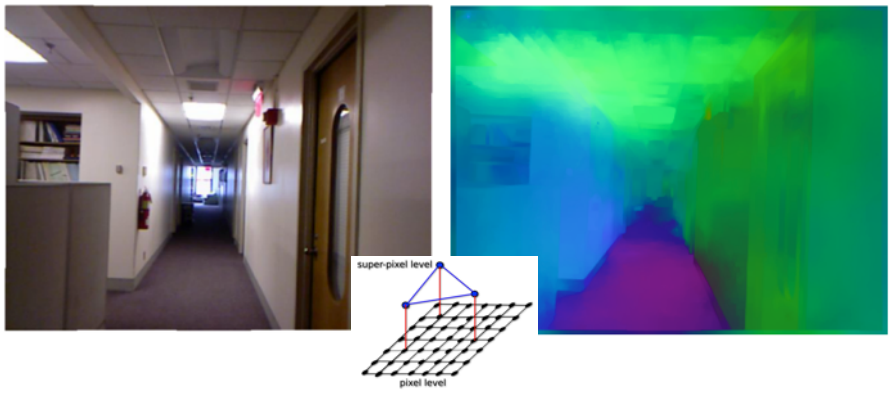
Large-scale Radiology Data[Shin+, CVPR2015]



医用画像の説明文により専門家の知識を再現

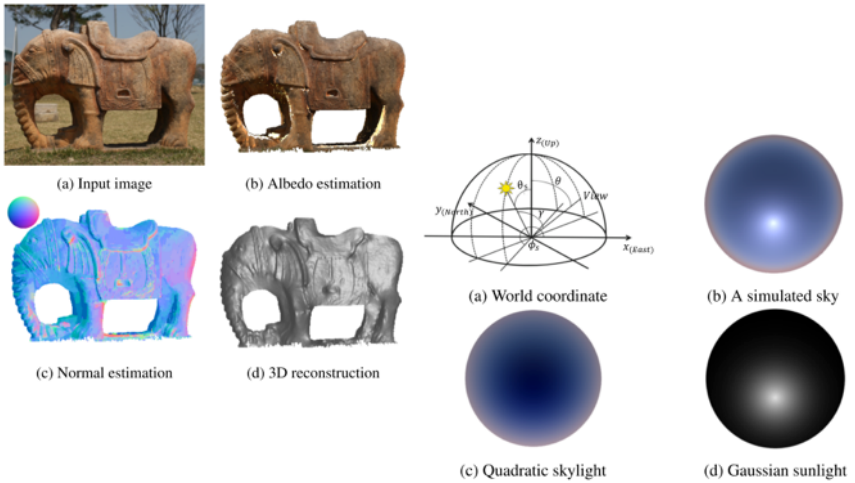
問題設定をした論文 (3)

Depth & Surface Prediction[Li+, CVPR2015]



単眼カメラによる3D距離推定

Outdoor Photometric Stereo[Jung+, CVPR2015]

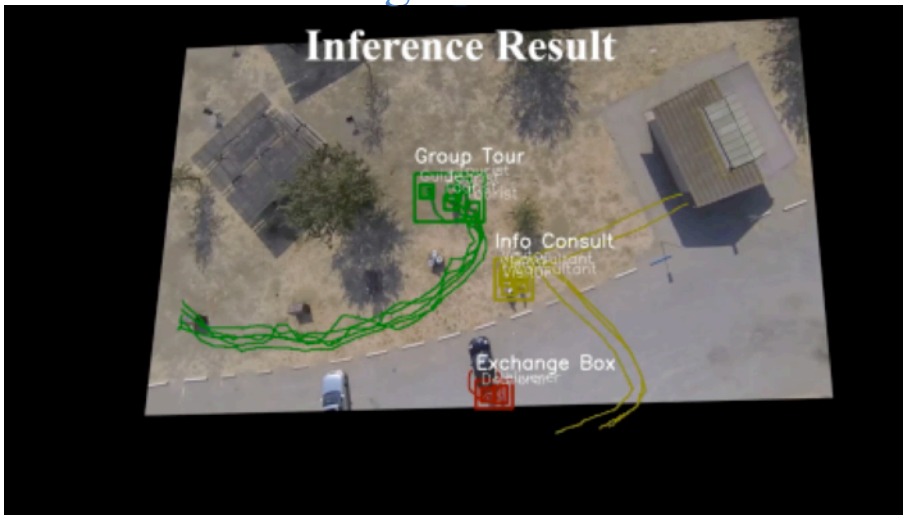


太陽光からPhotometric stereo

Social Saliency Prediction[Park+, CVPR2015]

人物の注目度をSaliencyと定義

UAV Event Recognition[Shu+, CVPR2015]

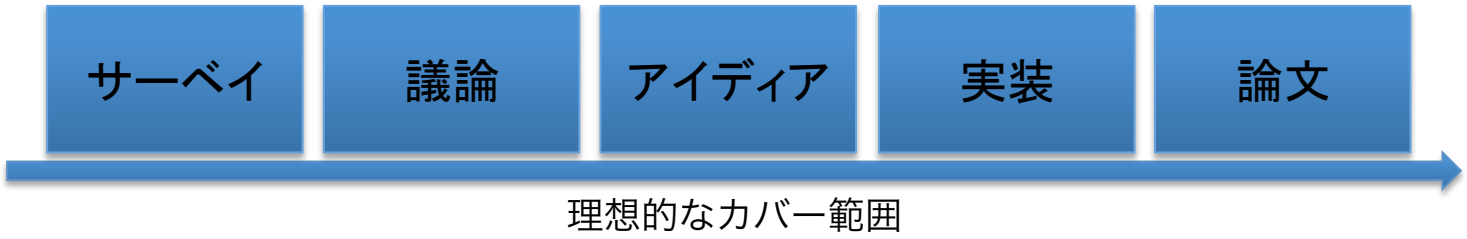


ドローンからイベントを認識

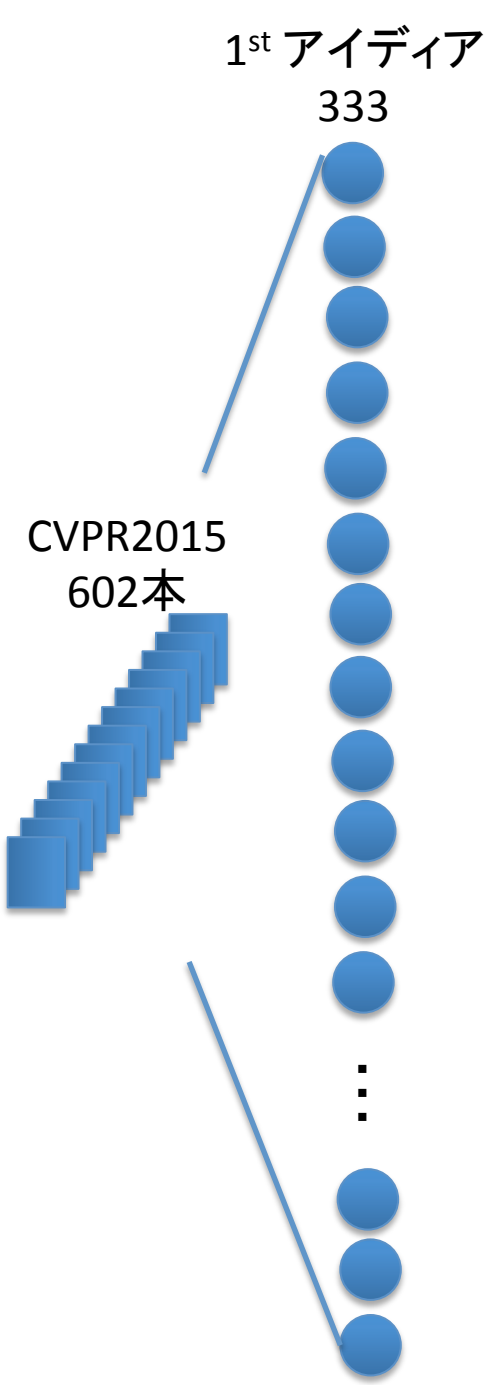
新規問題設定へ

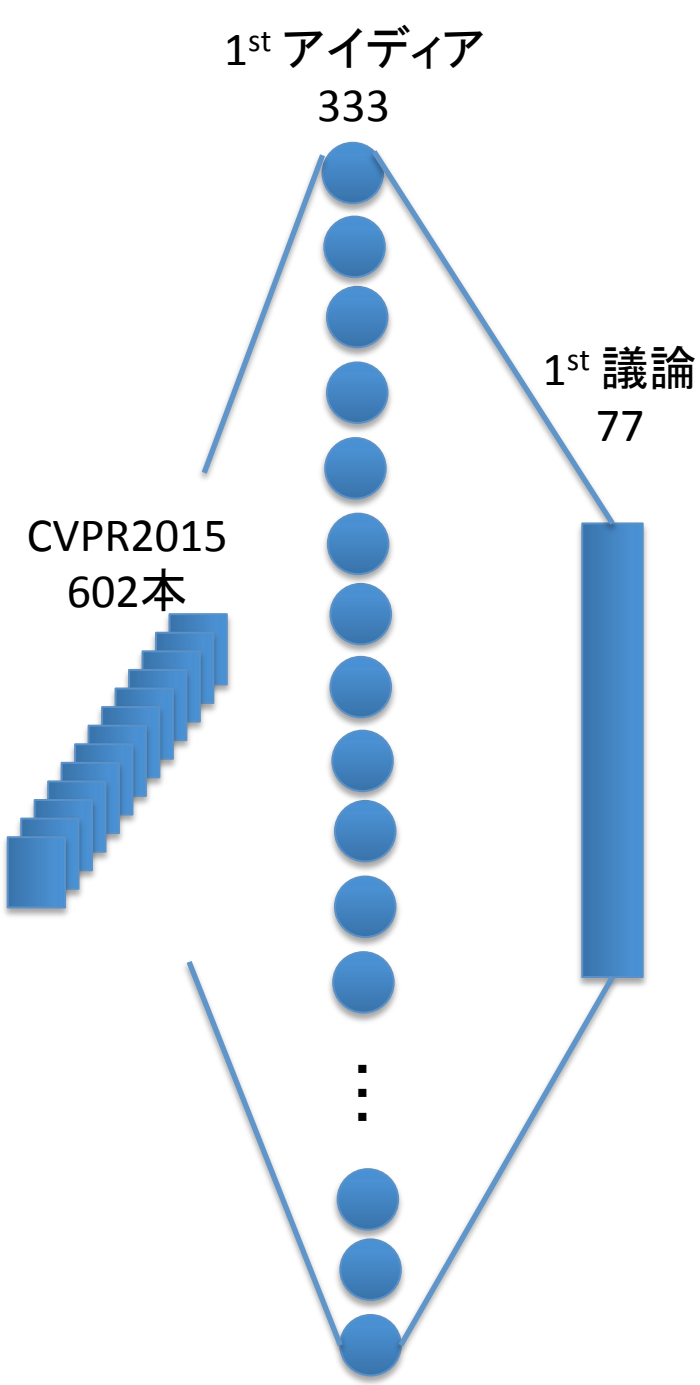
論文の多読・議論・アイディア考案・実装．．．

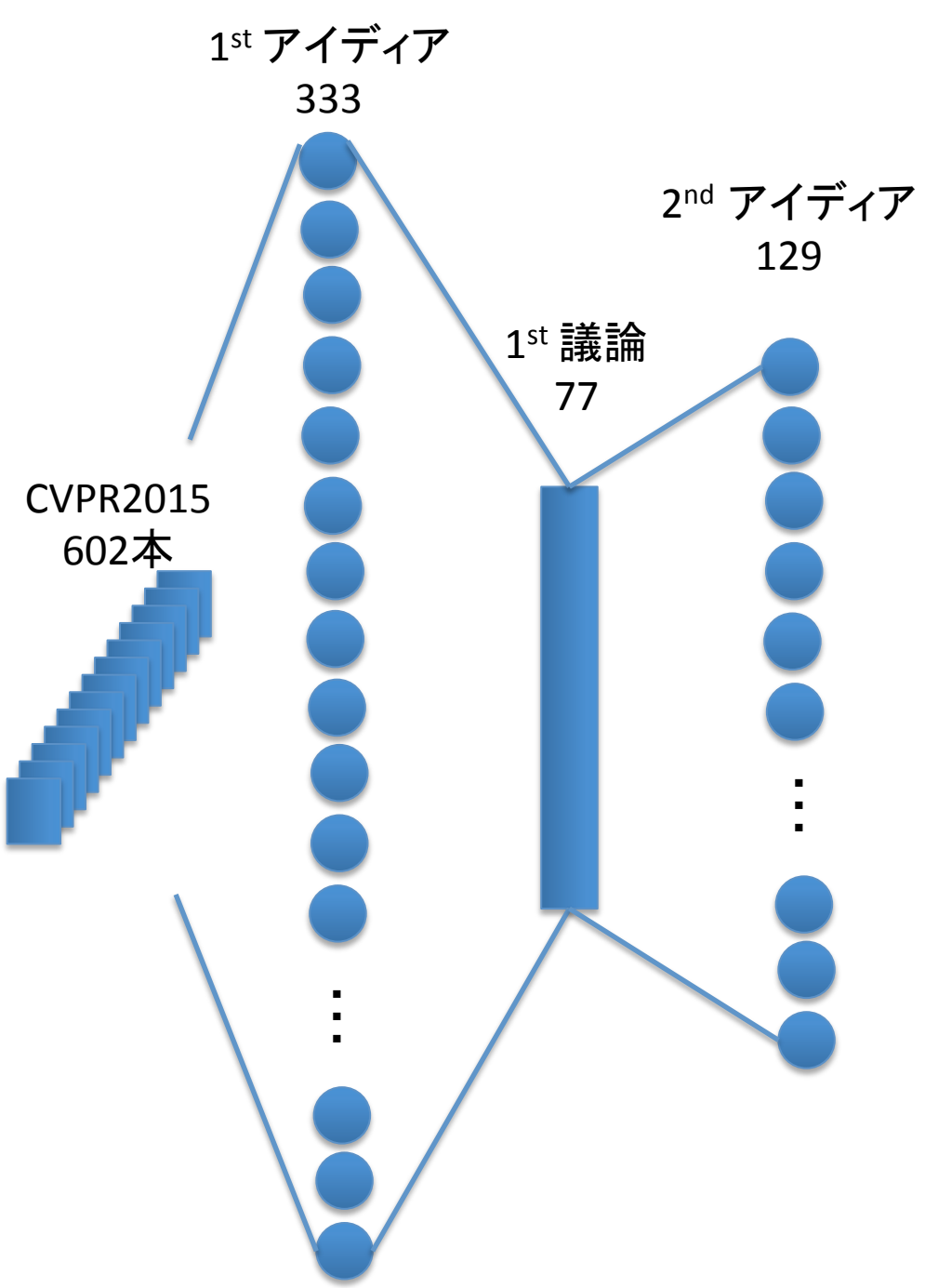
- CVPR2015論文+αをインプットとする
- まずはアイディアを各自考案してメモ
- 考えたら議論
- アイディア考案・議論の繰り返し
- 十分に練られたら実装工程へ

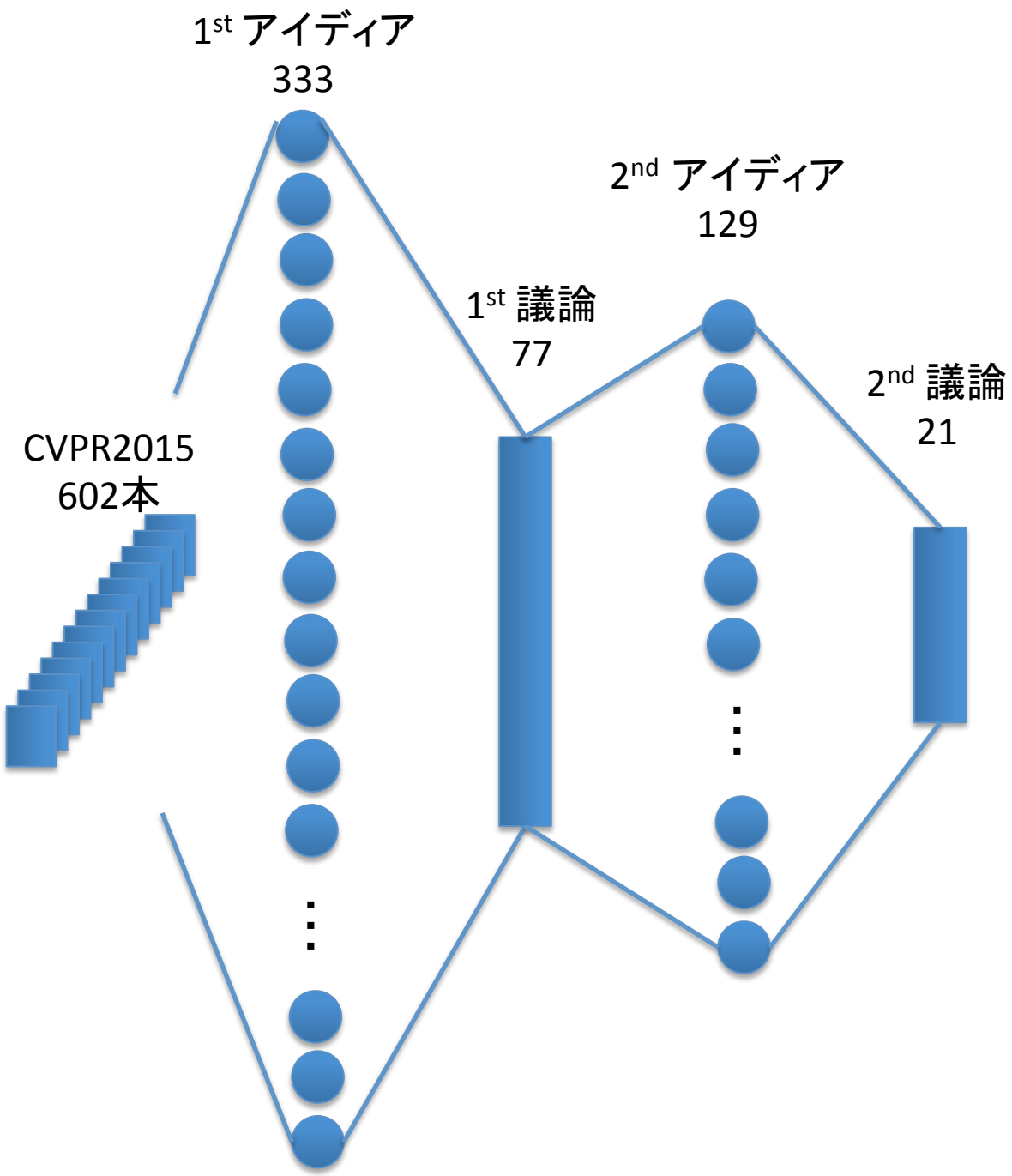


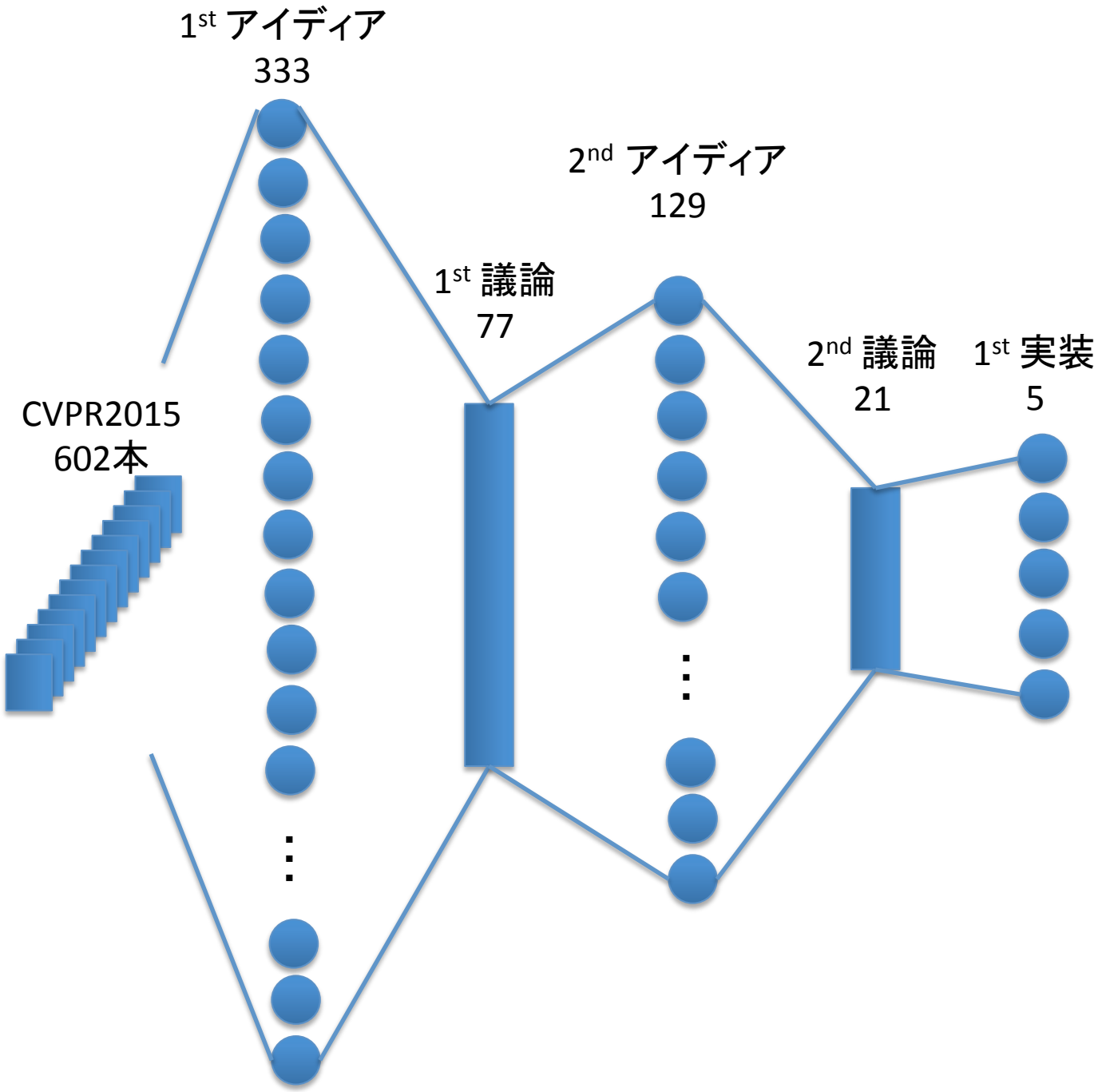
CVPR2015
602本

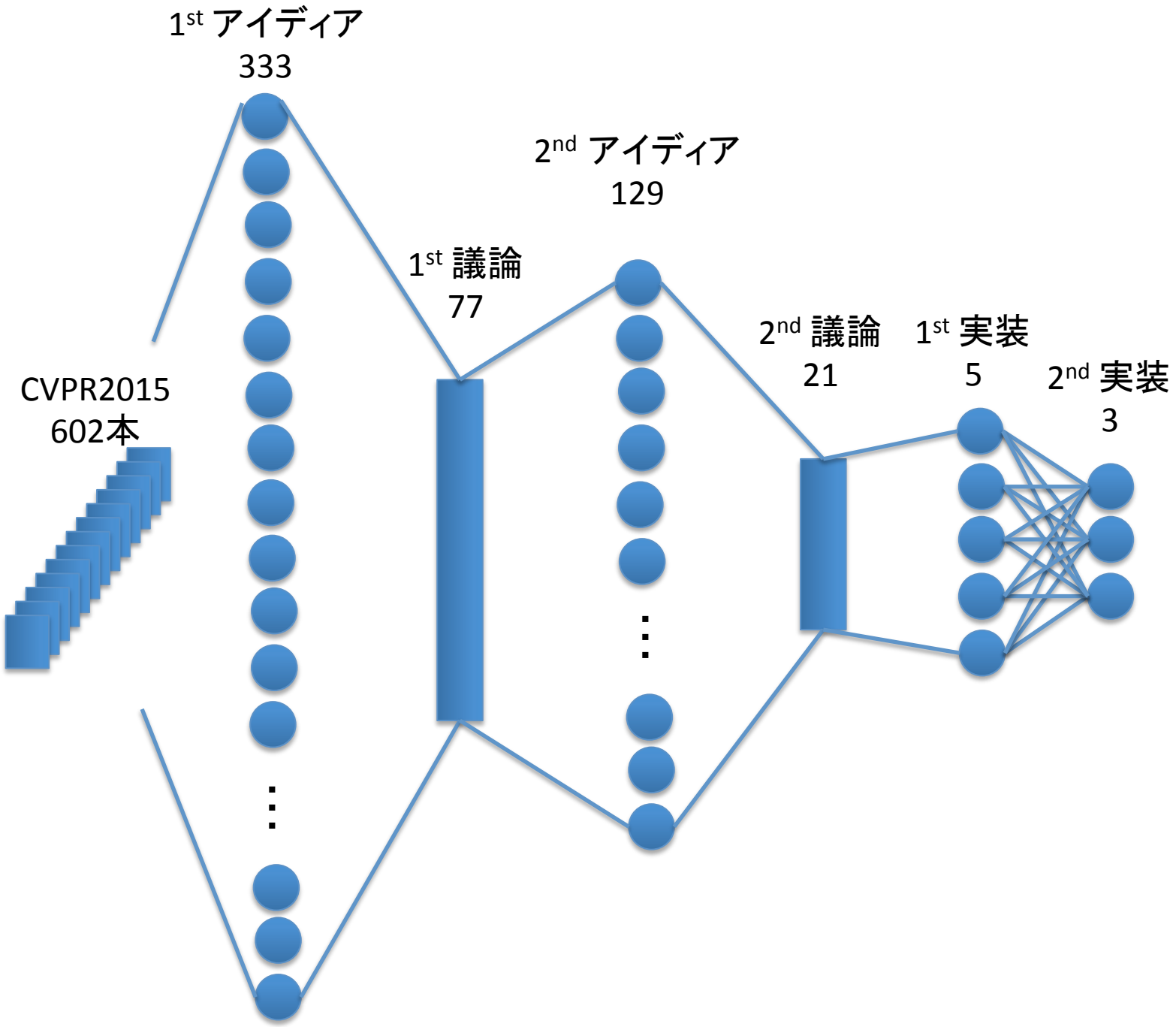

1st アイディア
333
CVPR2015
602本

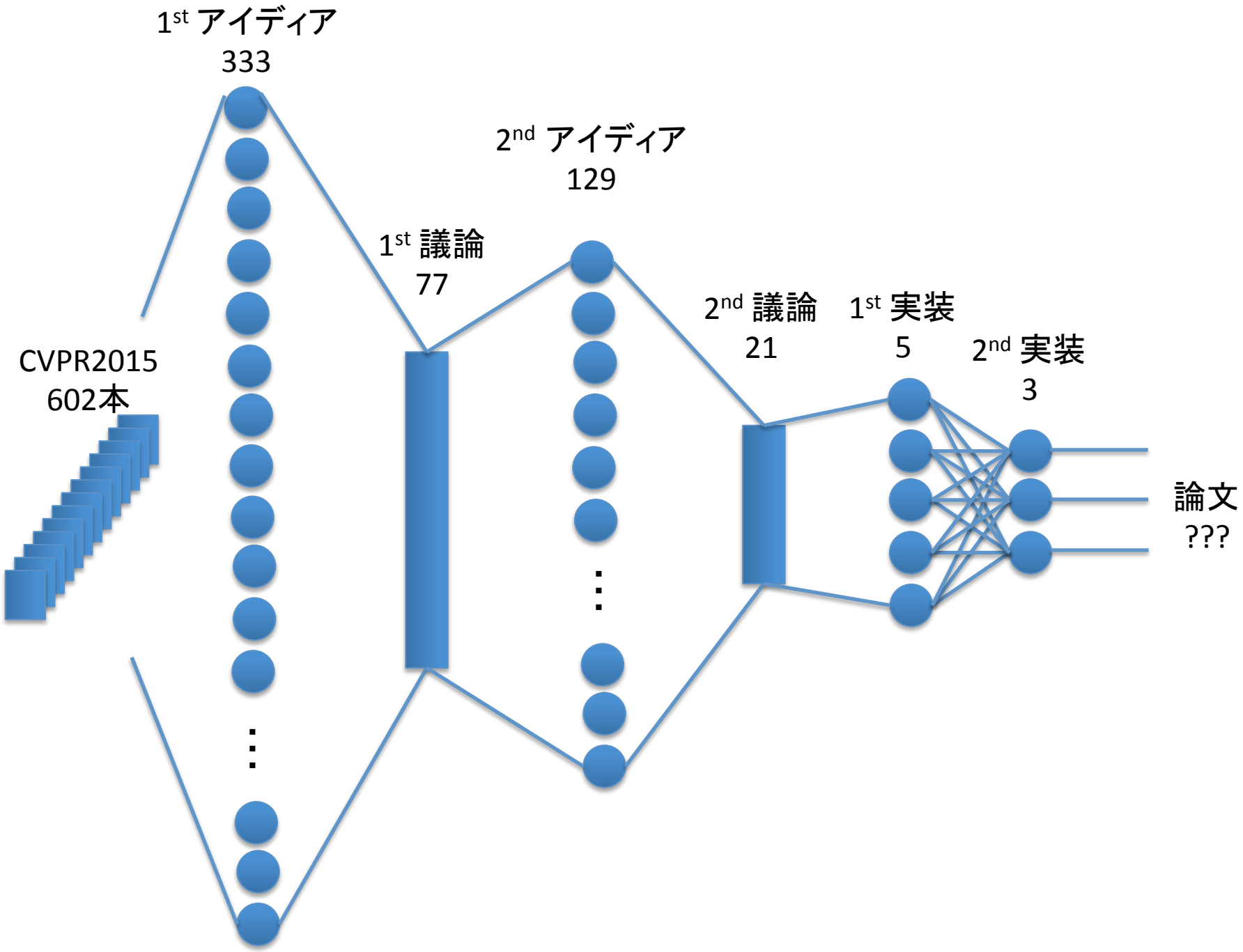

1st アイディア
333
1st 議論
77
CVPR2015
602本

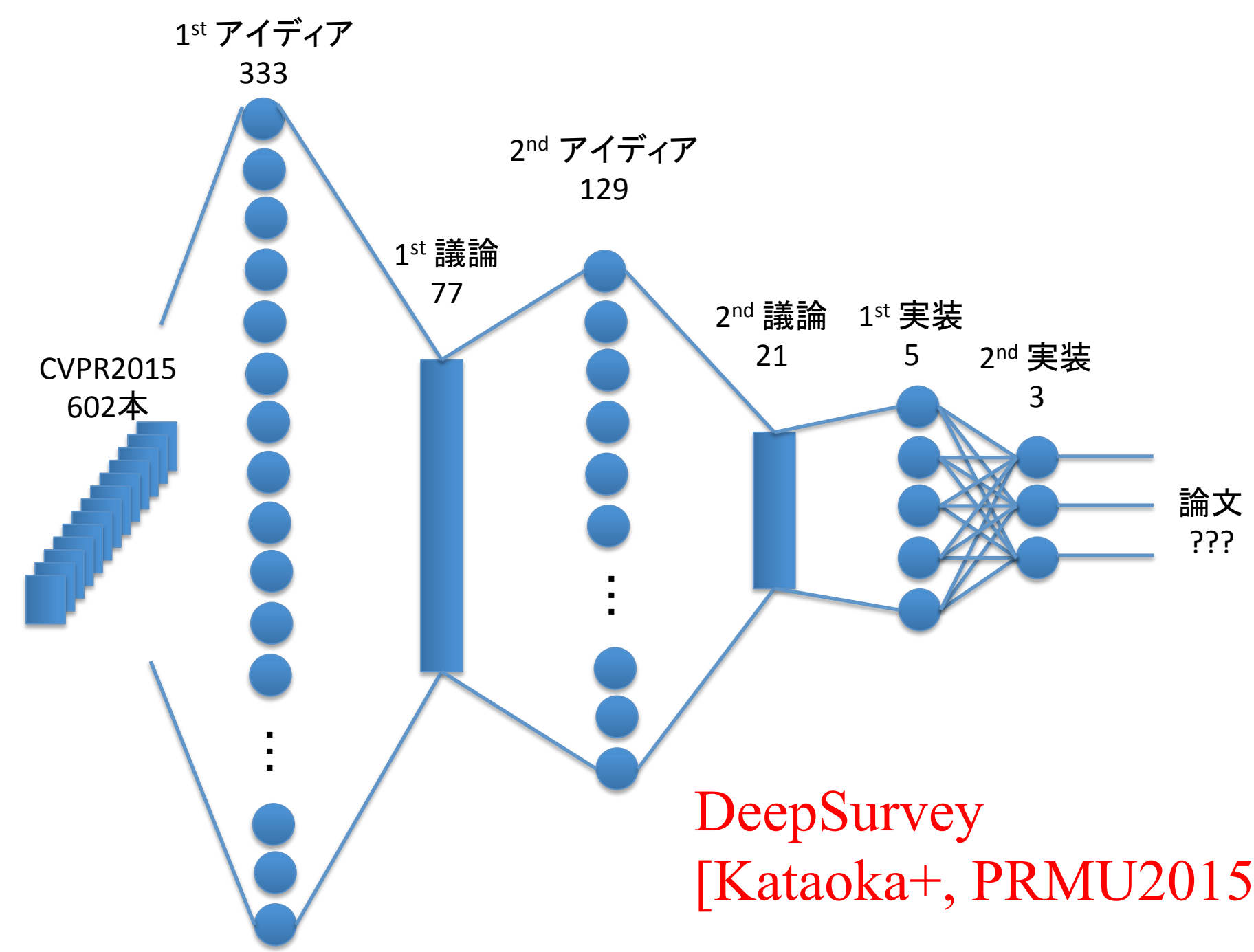

1st アイディア
333
2nd アイディア
129
1st 議論
77
CVPR2015
602本


1st アイディア
333
2nd アイディア
129
1st 議論
77
2nd 議論
21
CVPR2015
602本


1st アイディア
333
2nd アイディア
129
1st 議論
77
2nd 議論
21
1st 実装
5
CVPR2015
602本

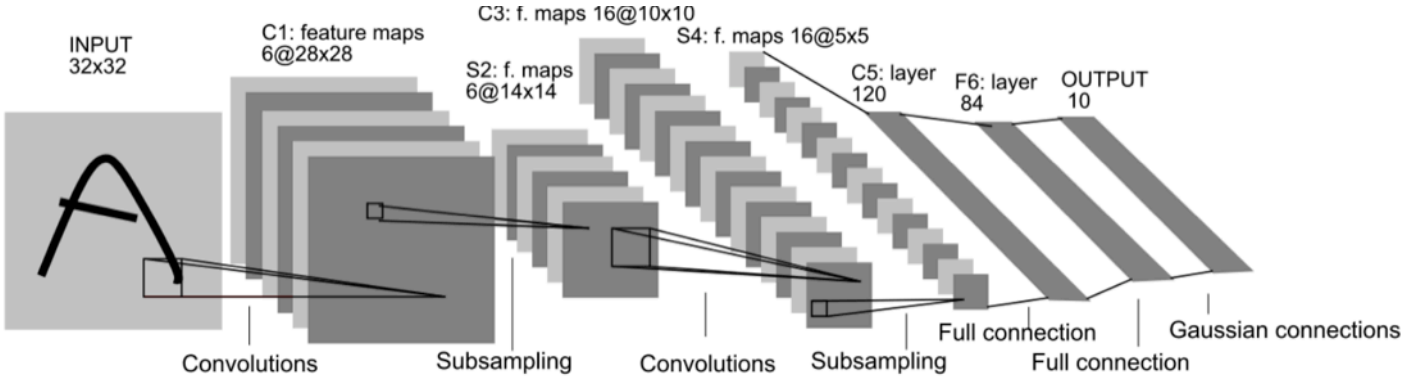

1st アイディア
333
2nd アイディア
129
1st 議論
77
2nd 議論
21
1st 実装
5
2nd 実装
3
CVPR2015
602本

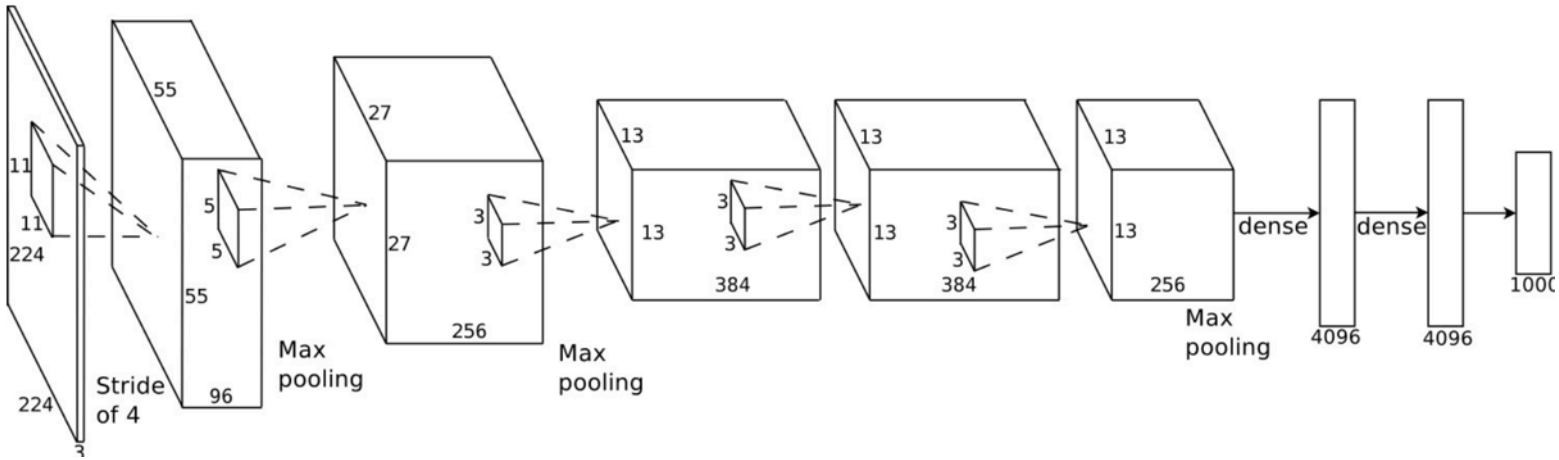

1st アイディア
333
2nd アイディア
129
1st 議論
77
2nd 議論
21
1st 実装
5
2nd 実装
3
CVPR2015
602本
論文
???


1st アイディア
333
2nd アイディア
129
1st 議論
77
2nd 議論
21
1st 実装
5
2nd 実装
3
CVPR2015
602本
論文
???
DeepSurvey
[Kataoka+, PRMU2015]

従来の深層学習と比較して．．．

- 自らがニューロンの一部となるという画期的な手法
- 出力が実物で出てくるという画期的な手法
- 一人だけではなくてみんなの力を合わせるという画期的な手法
- 学習を繰り返すことでニューロン自体が成長するという画期的な手法





cvpaper.challengeによって感じた成長

話者：宮下 侑大 (東京電機大学)

- 論文・読む編
- 論文・書く編
- 研究編

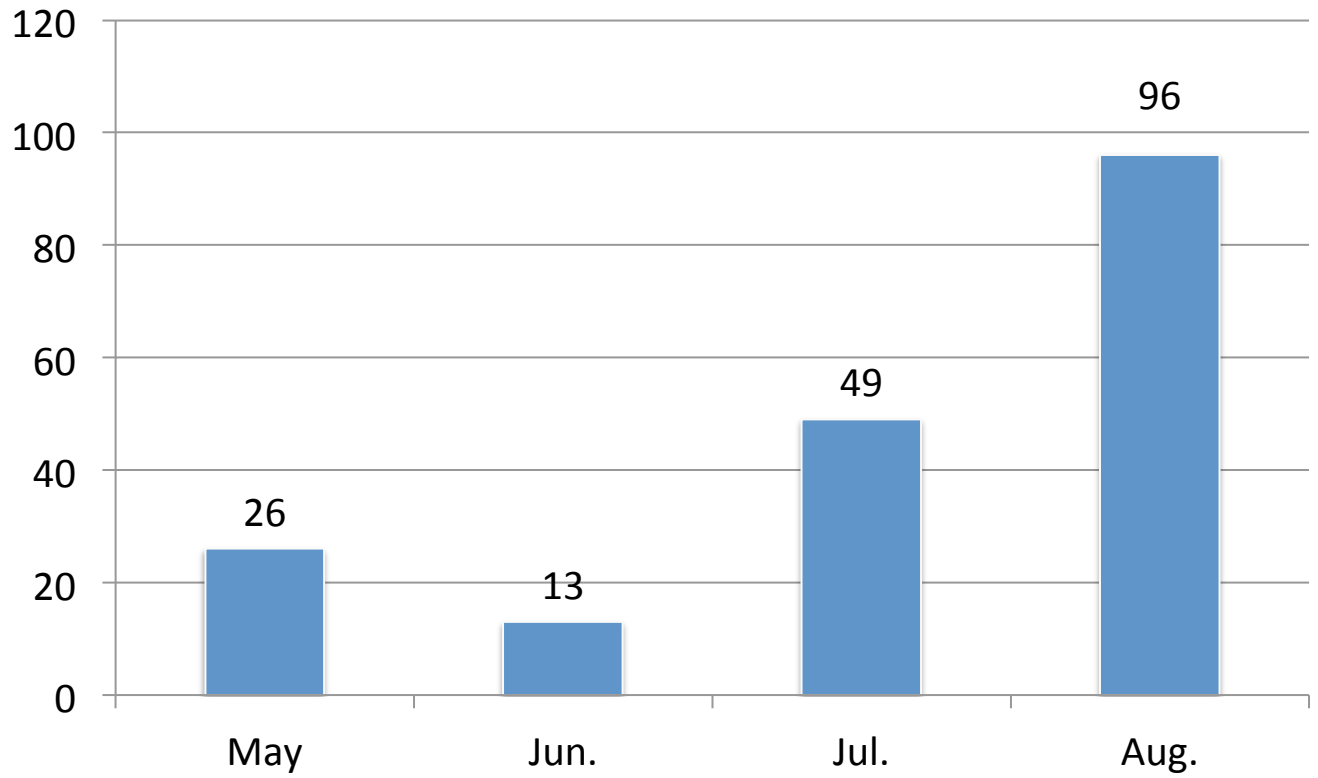
論文・読む編

学年別サーベイ論文数

– B4：20本程度,　M1：30本程度,　M2：180本程度

M2月別サーベイ論文数

– 5月：26本,　6月：13本,　7月：49本,　8月：96本



論文・読む編

サーベイテーマの広がり

- Person Re-ID → モーション解析 → 3次元認識

分野を限定すれば体系的にまとめることも可能

- PRMU資料のPerson Re-IDを担当
  - 5ヶ月でエキスパートに!?
- 1論文にかかる時間も大幅に短縮
  - 2時間 → 1時間 → 40分 → 20分

論文・書く編

学年別論文執筆数(共著含む)

- B4：1本, M1：1本, M2：12本
- 最先端を知っているからこそ, どのように主張して書けば良いのかが分かる
- 単純に参考文献にCVPRの論文があると説得力が増す

研究の質も向上した

- B4・M1: エッジ抽出, Hough変換, Graph Cut
- M2： Dense Trajectories, SVM, Random Forests, CNN

研究編

明らかに研究の進め方が違う

- B4・M1: 独りよがりのアイディアを実装 → 既に提案されていた
- M2: 最先端の手法と比較しながらアイディア考案

純粋に英語に対する抵抗感が無くなった

- 英語論文のサーベイ
- プログラムの英語リファレンス
- 英語メール

現在稼働中の3グループ

Semantic Change Detection (SCD) Grp.

- 画像の変化検出+意味付け
- 「どこがどのように変化したか」を認識

Movie Attribute Recognition (MAR) Grp.

- 映画や動画に対して複数ラベルやその尤度を自動解析
- キーフレーム検出や動画中の索引付けが課題

Nature Recognition (NR) Grp.

- 自然に対して一般的な認識を実行
- 炎の検出，風速計測，気候の自動認識

まとめと展望

2015年のサーベイの取り組みと論文化システム

- 論文の多読・体系化・共有
  - グループでの論文管理システム
  - CV分野(主に認識)のまとめ
- 論文化システム(DeepSurvey)の説明
  - インプットからアイディア，論文化システム

2016年のチャレンジ

- 2015年からカバー範囲・本数を増やす
- より質の高い論文へ

Mailto: cvpaper.challenge@gmail.com
Twitter: https://twitter.com/CVpaperChalleng
SlideShare: http://www.slideshare.net/cvpaperchallenge

Join us?

産総研コンピュータビジョン研究グループでは,
インターン・大学院生や博士研究員, 常勤職員を募集しております