

優れた問いを見つける

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/>

Concept in the cvpaper.challenge

充分に進化したコンピュータ視覚 (CV) は人間に新しい知見をもたらす。CVが飛躍的に進んだ先に、人間に対して拡張した知能や技能を教示できる。 (*Def. 拡張知能; Extended Intelligence*)

コンセプトを具現化する手段

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

- 人員：産総研，東京電機大，筑波大，慶應大による約20名
- 読む：多読・体系化・共有
- 書く：洗練された問題設定を論文化・社会実装



Diamond in your heart

cvpaper.challenge
@CVpaperChalleng

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

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



cvpaper.challenge

Edit profile

16 SlideShares
23 Followers
0 Clipboards

👤 Researchers, Graduate Students

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

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

※ HP, Twitter, SlideShareもご覧ください
HP: <https://sites.google.com/site/cvpaperchallenge/>
Twitter: @CVpaperChalleng
SlideShare: @cvpaperchallenge

@2016

論文1,000本読破 + 上位会議への投稿

計

1,000

本を達成

@2016

論文1,000本読破 + 上位会議への投稿

計

1000



2 papers@ECCV 2016 Workshop
Oral presentation at BNMW

本を達成

@2016
論文1,000本読破 + 上位会議への投稿

計

1

Accepted 4 papers
Submitted 4 papers
(2 top conferences)
Preparing 3 papers

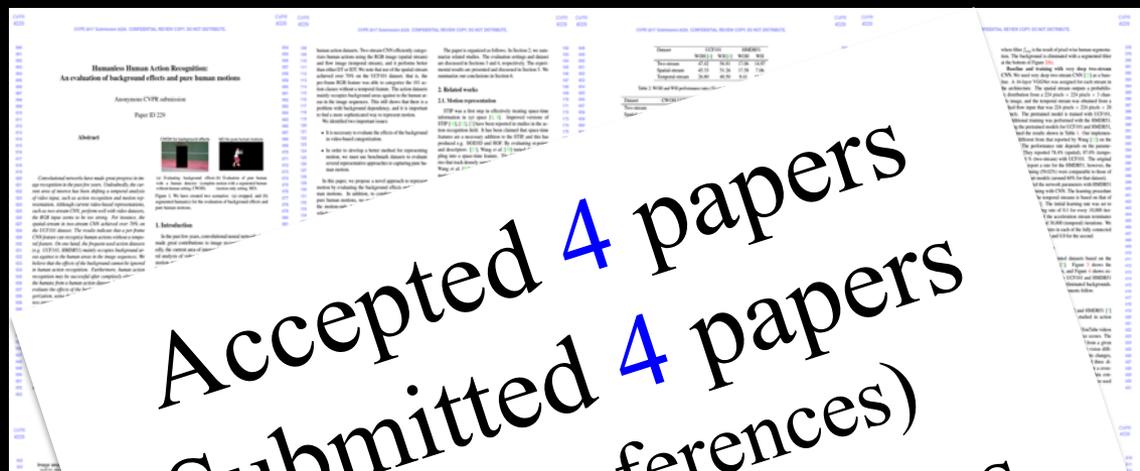
2 papers@ECCV 2016 Workshop
Oral presentation at BNMW



本を達成

@2016
論文1,000本読破 + 上位会議への投稿

計



Vision Engineering Workshop 2016

ビジョン技術の実利用ワークショップ

ViEW若手奨励賞

ECCV'16

EUROPEAN CONFERENCE ON COMPUTER VISION

Best Paper Award@BNMW
October 8 - 16, 2016 | Amsterdam | the Netherlands

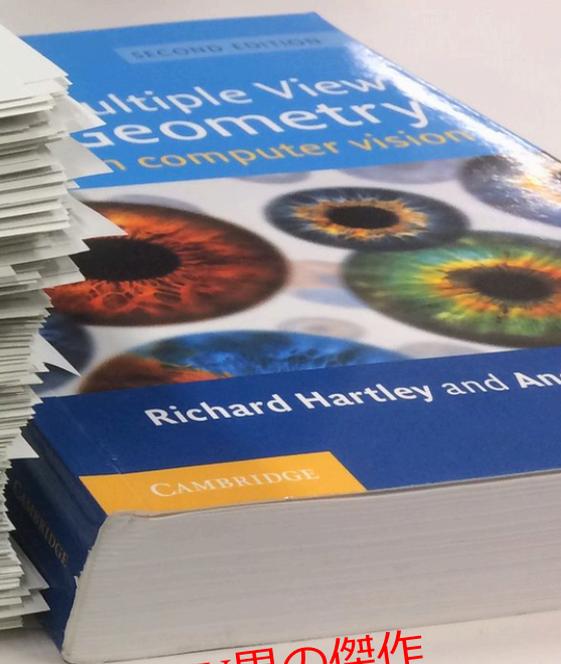


本を達成

@2016
論文1,000本読破 + 上位会議への投稿



(左) まとめ資料



(右) 700pあるというCV界の傑作

arXivにて週間 世界第一位を達成！

- 取り組みが世界的に話題 (参考文献数も)
- CambridgeやStanfordからも問い合わせあり

The screenshot shows the Arxiv Sanity Preserver interface. At the top, there's a search bar and navigation options like 'most recent', 'top recent', 'recommended', and 'library'. Below that, there are filters for 'Only show v1' and time-based filters like 'Last day', 'Last 3 days', 'Last week', 'Last month', 'Last year', and 'All time'. The main content area displays a list of papers. The first paper, 'cvpaper.challenge in 2015 - A review of CVPR2015 and DeepSurvey', is highlighted with a red border. It lists authors: Hirokatsu Kataoka, Yudai Miyashita, Tomoaki Yamabe, Soma Shirakabe, Shin'ichi Sato, Hironori Hoshino, Ryo Kato, Kaori Abe, Takaaki Imanari, Naomichi Kobayashi, Shinichiro Morita, and Akio Nakamura. The date is 5/26/2016, and it's categorized as 'cs.CV | cs.LG | cs.MM | cs.RO'. Below the title is a 'Survey Paper' label and a grid of thumbnail images representing the paper's content. A short abstract follows, describing the 'cvpaper.challenge' group and their goal of summarizing CVPR2015 papers. The second paper, 'Control of Memory, Active Perception, and Action in Minecraft', is partially visible below.

related fields. As the first step of this endeavor, we undertook to read all the 602 papers accepted during the CVPR2015 [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602]. This review summarizes all the papers of CVPR2015 we read as the first project of

Publications

【2017】

- Yuta Matsuzaki*, Kazushige Okayasu*, Takaaki Imanari, Naomichi Kobayashi, Yoshihiro Kanehara, Ryousuke Takasawa, Akio Nakamura, Hirokatsu Kataoka, "Could you guess an interesting movie from the posters?: An evaluation of vision-based features on movie poster database", IAPR Conference on Machine Vision Applications (MVA2017), May 2017. (* equal contribution) **[New!]**
- Teppei Suzuki, Yoshimitsu Aoki, Hirokatsu Kataoka, "Pedestrian Near-Miss Analysis on Vehicle-Mounted Driving Recorders", IAPR Conference on Machine Vision Applications (MVA2017), May 2017. **[New!]**

【2016】

- 賀雲, 白壁奏馬, 佐藤雄隆, 片岡裕雄, "人を見ない人物行動認識", ViEW, 2016. **(オーラル, ViEW若手奨励賞)**
- 鈴木哲平, 片岡裕雄, 青木義満, "車載動画におけるニアミス状況下での歩行者解析", ViEW, 2016. **(ハイブリッドオーラル)**
- 片岡裕雄, 白壁奏馬, 賀雲, 阿部香織, 上田隼也, 鈴木哲平, 松崎優太, 岡安寿繁, 矢部俊之, 金原慶拓, 八柳光哉, 丸山慎也, 高澤良輔, 淵田正隆, 森田慎一郎, "これからのコンピュータビジョン技術 - cvpaper.challenge in PRMU Grand Challenge 2016 ", PRMU グランドチャレンジ, 2016. [\[Link\]](#) [\[Slide\]](#)
- Yun He, Soma Shirakabe, Yutaka Satoh, Hirokatsu Kataoka, "Human Action Recognition without Human", ECCV 2016 Workshop on Brave New Ideas for Motion Representations in Videos (BNMW), Oct. 2016. [\[PDF\]](#) [\[Project\]](#) **(Oral, Best Paper)**
- Hirokatsu Kataoka, Yun He, Soma Shirakabe, Yutaka Satoh, "Motion Representation with Acceleration Images", ECCV 2016 Workshop on Brave New Ideas for Motion Representations in Videos (BNMW), Oct. 2016. [\[PDF\]](#) [\[Project\]](#)
- 片岡裕雄, "cvpaper.challenge – サーベイの共有と可能性について", 画像応用技術専門委員会定例研究会, Jul. 2016. [\[Link\]](#)
- Hirokatsu Kataoka, Yudai Miyashita, Tomoaki Yamabe, Soma Shirakabe, Shin'ichi Sato, Hironori Hoshino, Ryo Kato, Kaori Abe, Takaaki Imanari, Naomichi Kobayashi, Shinichiro Morita, Akio Nakamura, "cvpaper.challenge in 2015 - A review of CVPR2015 and DeepSurvey", arXiv pre-print 1605.08247, May 2016. [\[PDF\]](#) [\[arXiv\]](#) **(Top-1 paper in arXiv sanity (weekly))**
- Hirokatsu Kataoka, Soma Shirakabe, Yudai Miyashita, Akio Nakamura, Kenji Iwata, Yutaka Satoh, "Semantic Change Detection with Hypermaps", arXiv preprint arXiv:1604.07513, Apr. 2016. [\[arXiv\]](#)

【2015】

- 片岡裕雄, 宮下侑大, 山辺智晃, 白壁奏馬, 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, "cvpaper.challenge in CVPR2015 -CVPR2015 のまとめ-", パターン認識・メディア理解研究会 (PRMU) , Dec. 2015. [\[PDF\]](#) [\[Slide\]](#)

なぜ、サーベイするのか？

世界水準の研究のため

トレンドの把握

- 知識がないと既存研究の劣化版を作りかねない
- ティンドを知らないと(天才でない限り)最先端の研究を生ま出すことは難しい

自身の研究の立ち位置を確認

- 何が違う？なぜやる？どこが良いのか？という哲学

究極的には次のトレンドを作るため (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

CHI勉強会2016

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

- CHIはユーザインタフェース界のトップ会議
- 2006年から11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CV分野をさらに盛り上げるための起爆剤になれるか？

「優れた問い」とは？

「優れた問い」とは？

【定義】次世代の研究トレンドを作れる研究および問題設定

- それまでの流れを変える、若くは促進させる重要研究
- 最初、どんなに小さなムーブメントでも良いから挑戦
 - 複数の問いの連鎖により大きな流れにもできる！

【再掲 & 追記】

(サーベイの目的は) 究極的には次のトレンドを作るため

- 分野の方向性を自ら定める
 - 研究の醍醐味であり、面白さを感じる部分（自ら道を切り開く）
- より良く、正しい方向へ導く
 - 本質とは何か？を追求する（ただ面白い、以上の価値を見出す）

「優れた問い」の例

現在のトレンド「深層学習」

- DeepLearning（特に画像におけるCNN）の手法
 - Neocognitron (1980)：人間のような認識
 - CNN (1989)：畳み込み/プーリング/全結合によるニューラルネットの原型
 - AlexNet (2012)：構造をディープにしてみたら多クラス分類に適合
- 一般物体認識データの発展
 - Caltech101、Caltech256 (2004, 2005)：データを集める重要性
 - ImageNet (2009)：膨大な画像数、網羅性も考慮して圧倒的に収集
 - ILSVRC (2012)：コンペティションを行ってみたらどうだろう
- 各々の連鎖&両者の融合にて2012年に革命が起きた

画像認識分野における研究動向

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

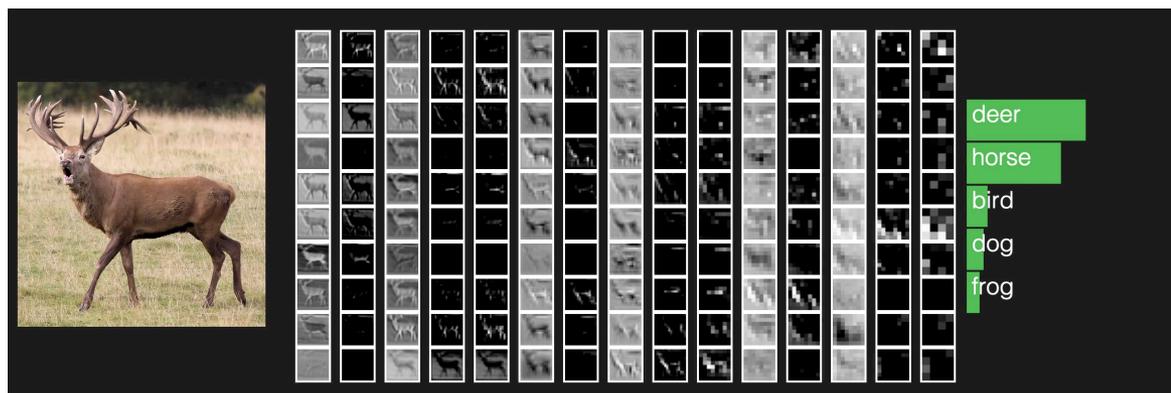
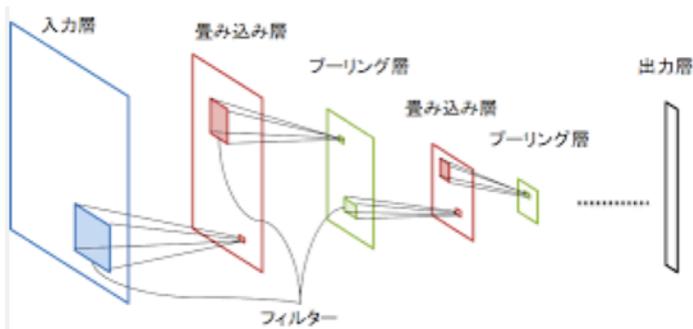


- 精度を出すための知識が整った
 - 勾配を保存しつつ, より深い構造に
 - 学習のためのデータ数を膨大に
- ライブラリ/Big Data/計算リソースが整った
 - Caffe/Torch/Pylearn2/Theano/Chainer/TensorFlow
 - ImageNet, Microsoft COCO, ...

深層学習

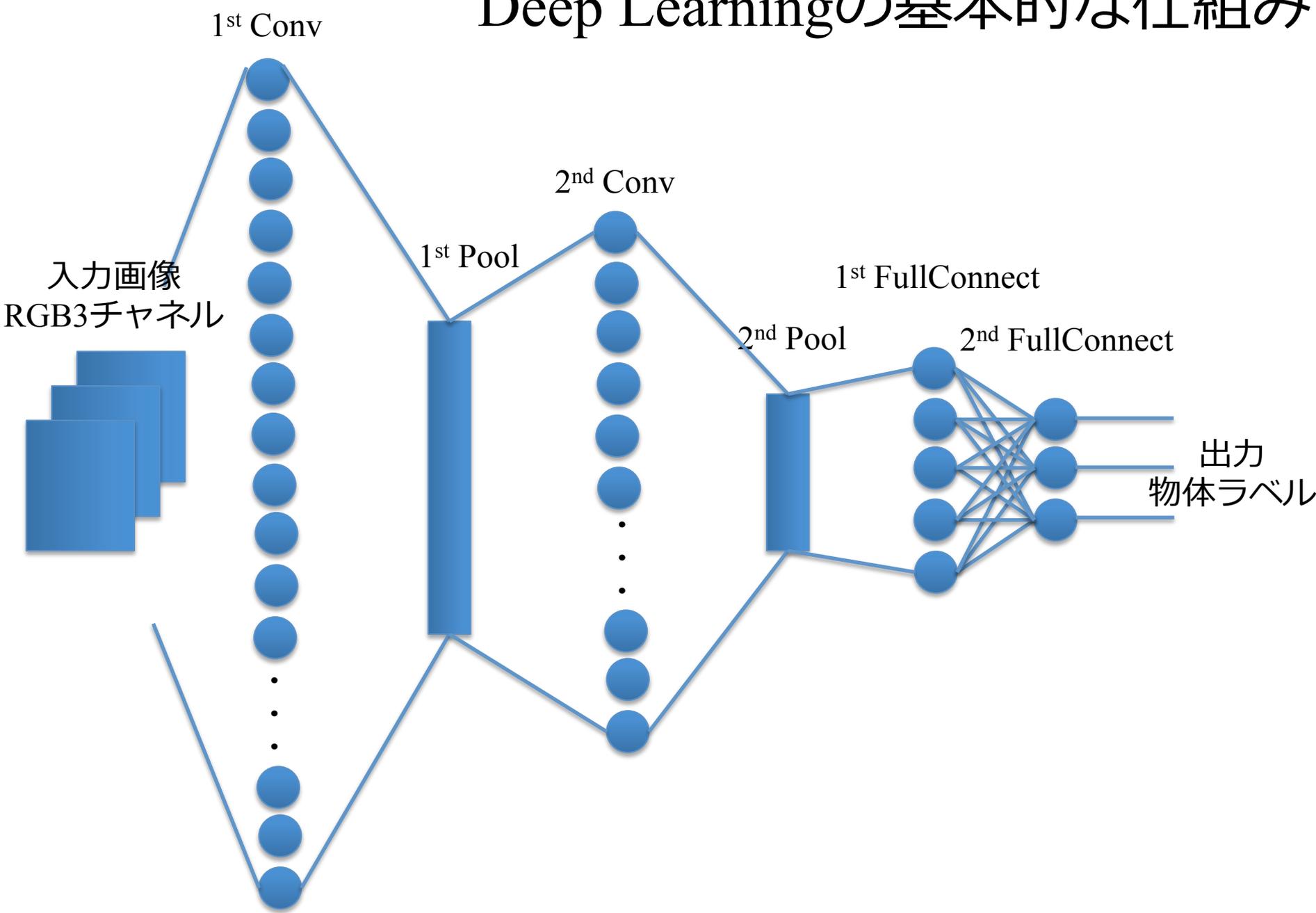
深層畳み込みニューラルネット (CNN)

- 畳み込みと呼ばれるフィルタリングにより特徴抽出
- プーリングにより位置ずれの影響を最小限に抑える
- 畳み込み・プーリングの繰り返しにより画像から概念へ
- 全結合層により特徴を高次化



https://www.youtube.com/watch?v=zj_JIVqWK1M

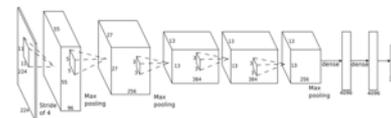
Deep Learningの基本的な仕組み



深層学習の衝撃

画像認識の精度が格段に向上

- 2012年のILSVRCにてDeep Learningが大勝利
- 2016年現在1,000クラス分類にて3.08%のエラー率
- 画像認識のみならず, 物体検出/ セグメンテーション/ 時系列解析/ 画像生成/ 可視化/ 3次元/ カメラ幾何などに応用



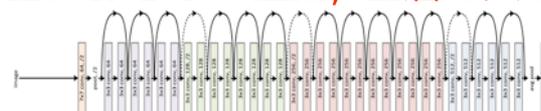
AlexNet [Krizhevsky+, ILSVRC2012]
ILSVRC2012 winner, DLの火付け役



VGGNet [Simonyan+, ILSVRC2014]
16/19層ネット, deeperモデルの知識

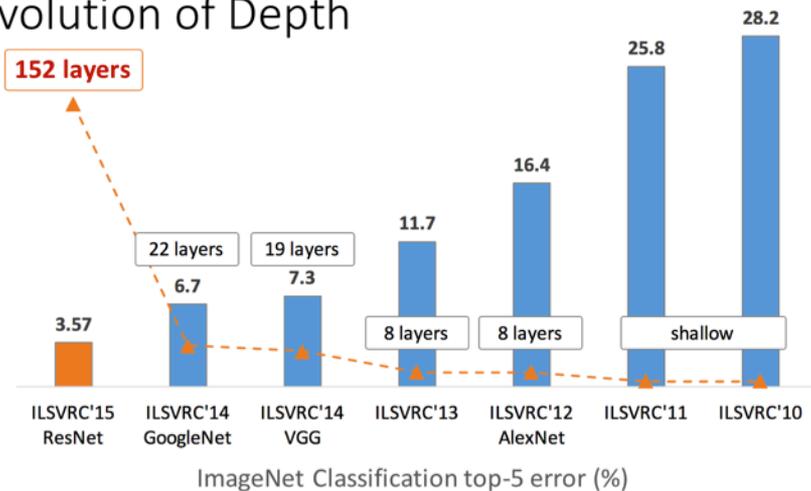


GoogLeNet [Szegedy+, ILSVRC2014/CVPR2015]
ILSVRC2014 winner, 22層モデル



ResNet [He+, ILSVRC2015/CVPR2016]
ILSVRC2015 winner, 152層! (実験では10³+層も)

Revolution of Depth

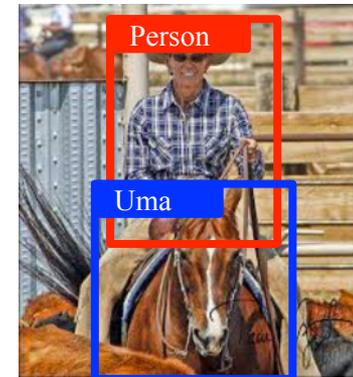
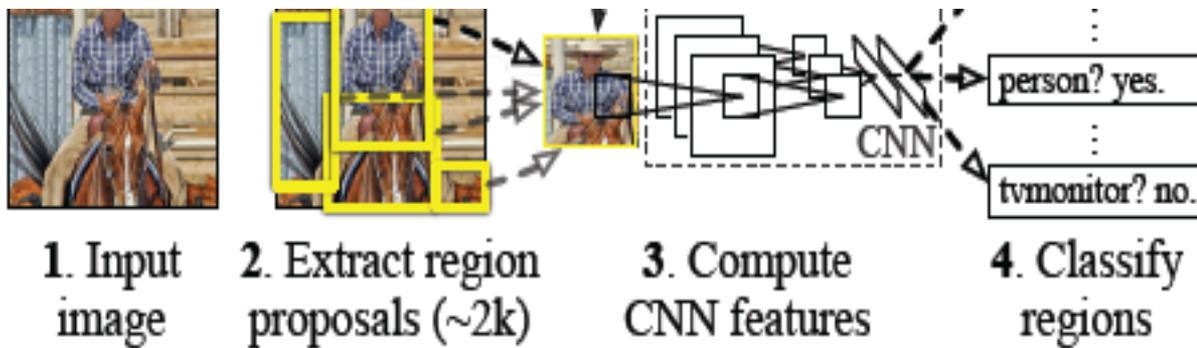


Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. "Deep Residual Learning for Image Recognition". CVPR 2016.

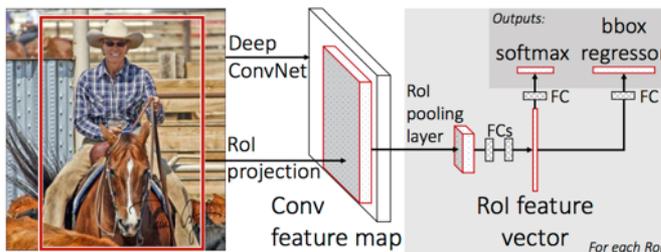
物体検出の仕組みと高速化

Region with CNN (R-CNN)が物体検出の先駆け

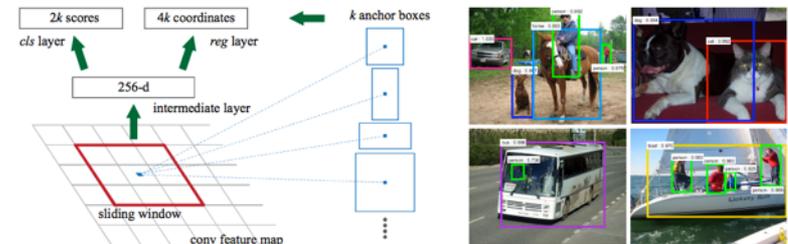
- 物体候補領域を任意の手法で抽出, 各領域にてCNNの識別
- Fast R-CNNでは畳み込みマップを共有して識別
- Faster R-CNNでは候補領域抽出自体もCNN内にて実行



R-CNN [Girshick+, CVPR14] 0.021fps



Fast R-CNN [Girshick, ICCV15] 3.1fps



Faster R-CNN [Ren+, NIPS15] 5.0fps
候補領域をニューラルネット内で実行

驚くべきことに . . .

高速化と同時に高精度化も実現！

- @PASCAL VOC
- 動画は運転シーン



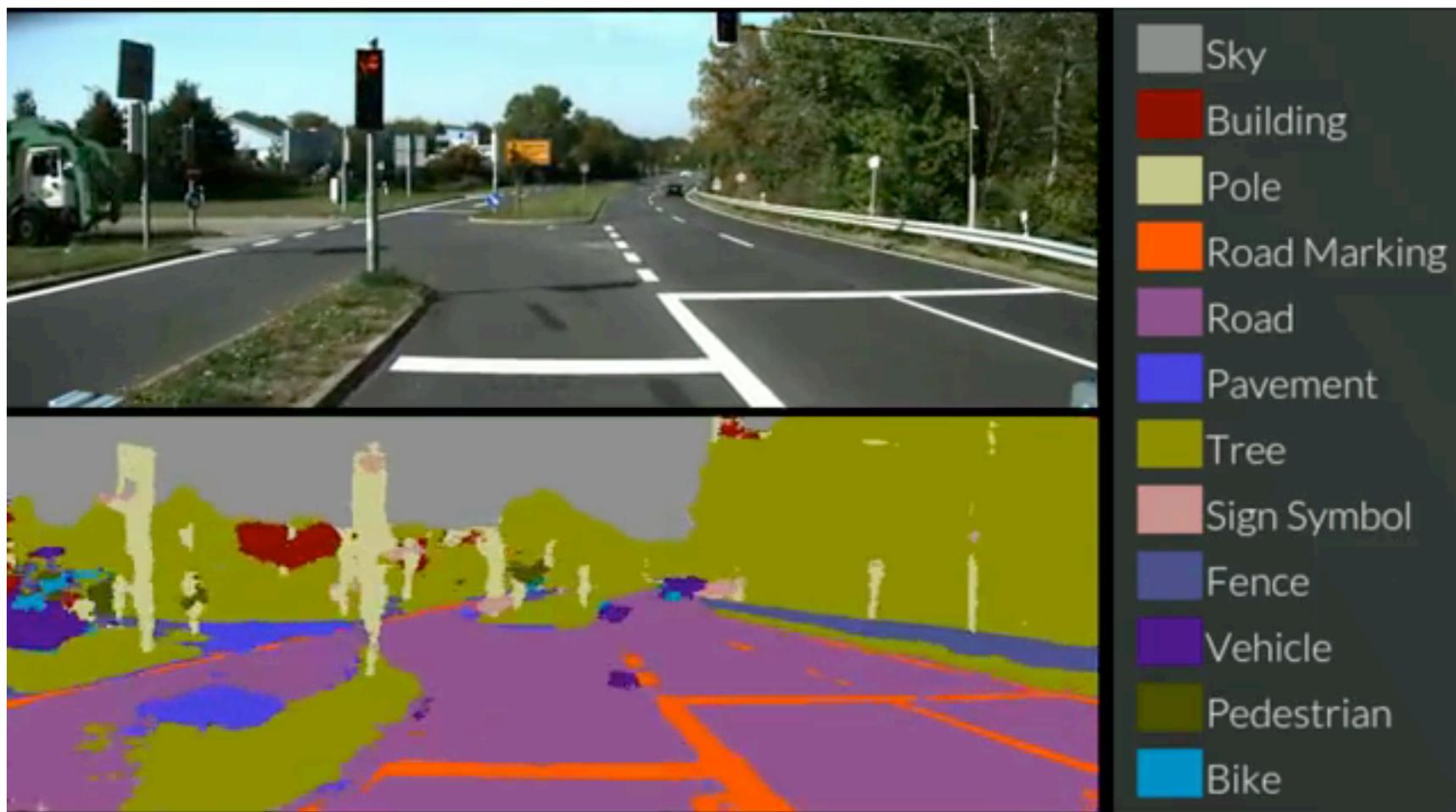
<https://www.youtube.com/watch?v=WZmSMkK9VuA>

	2007	2010	2012
R-CNN [CVPR14] 0.021 fps	58.8	53.7	53.3
R-CNN+ [PAMI15]	66.0	62.9	62.4
Fast R-CNN [ICCV15] 3.1 fps	70.0	68.8	68.4
Faster R-CNN (VGGNet) [NIPS15] 5.0 fps	73.2	-	70.4
Faster R-CNN (Resnet) [CVPR16] 5.0 fps?	85.6	-	83.8
YOLO (Model Fusion) [CVPR16] 30 ~ 150 fps	75.0	-	70.7

セマンティックセグメンテーション

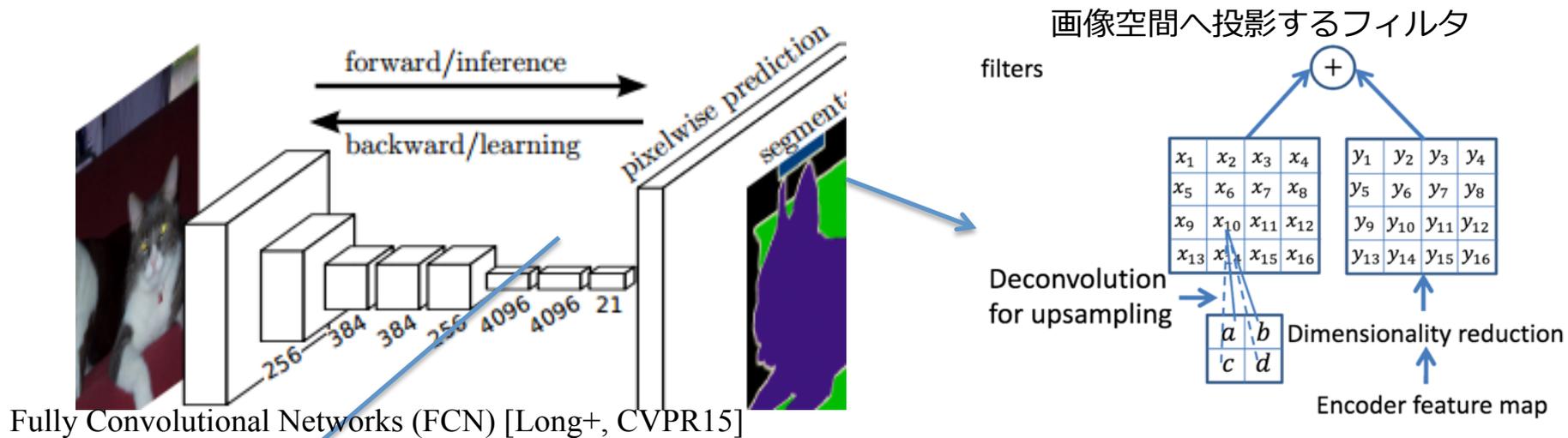
クラス分類と領域セグメンテーションを同時実行

– 近年最も急激に進んだ技術である

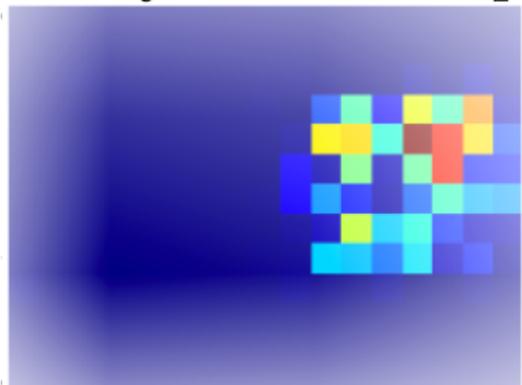


セマンティックセグメンテーションの仕組み

概念レベルの特徴からセマンティック化, 画像位置に投影

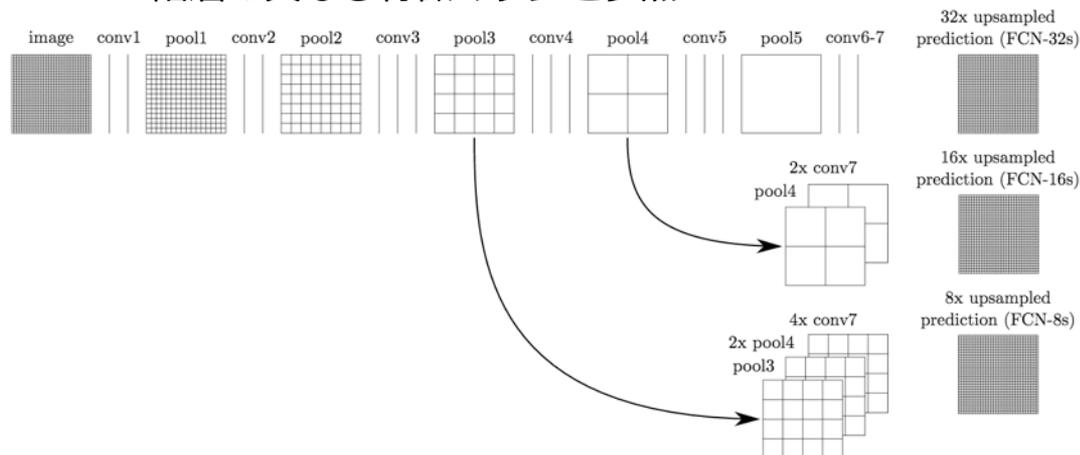


tabby cat heatmap



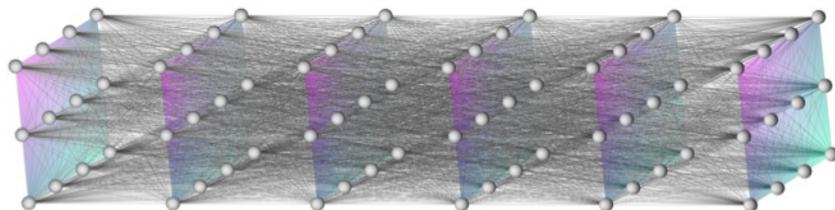
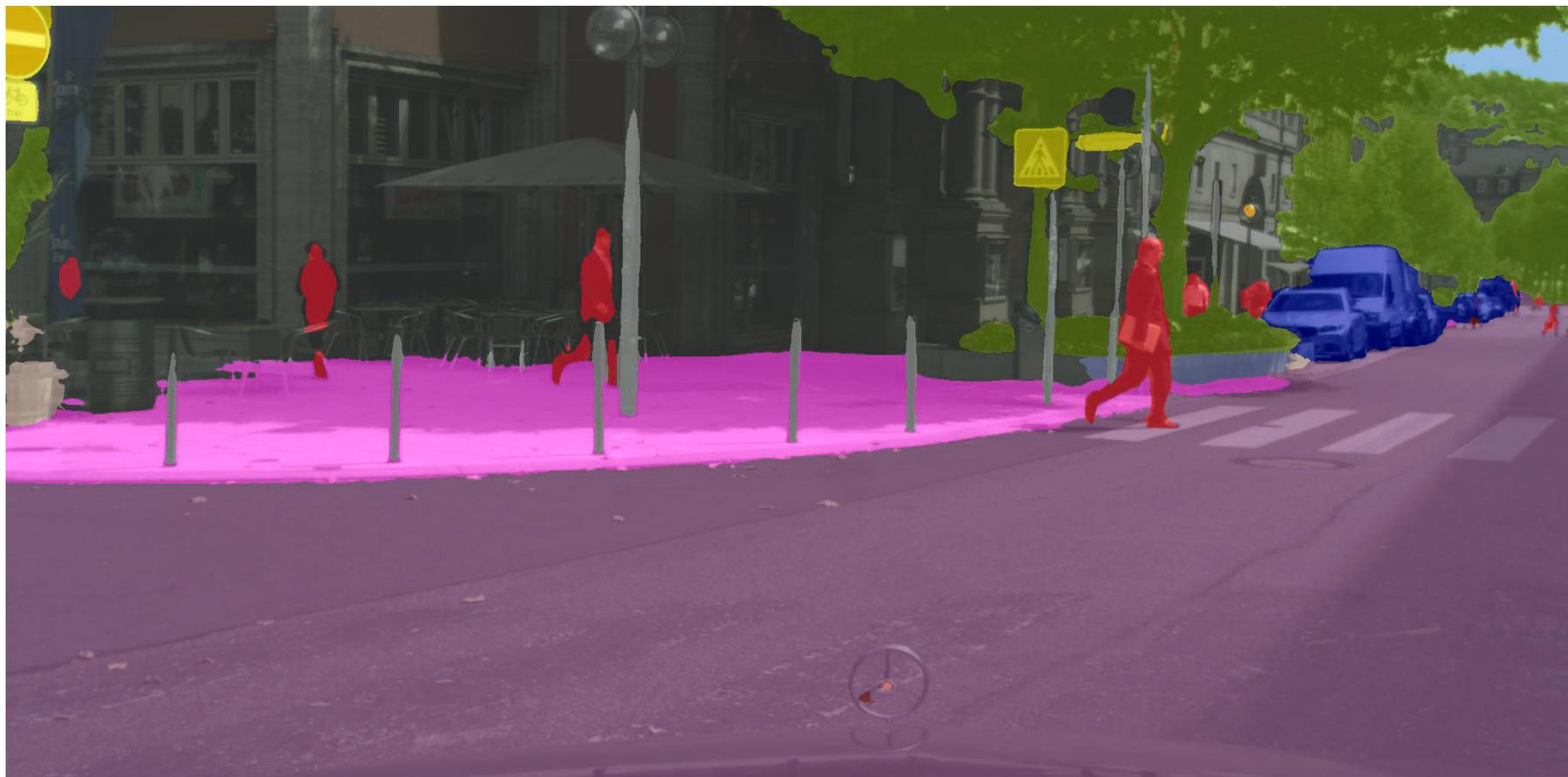
物体の概念レベルの特徴マップ

階層の異なる特徴マップを参照



セマンティックセグメンテーション

自動運転の実用レベル？



3D Dense CRF [Kundu+, CVPR16]
空間的・時間的に密な結合による領域評価

2D画像認識は終わったか？

研究レベルでは収束しているように見える

- 少なくとも、データセットレベルではPerfectに近い精度を実現
- ImageNet (画像識別), Pascal VOC (物体検出) 等
- 「過去の問い」 になっていく？

応用面ではやり残したことが数多く残される

- 実空間ではまだうまく動かないことも多い
- 「過去の問い」 から「優れた問い」 に転換するセンスが問われる

2D画像認識の先へ

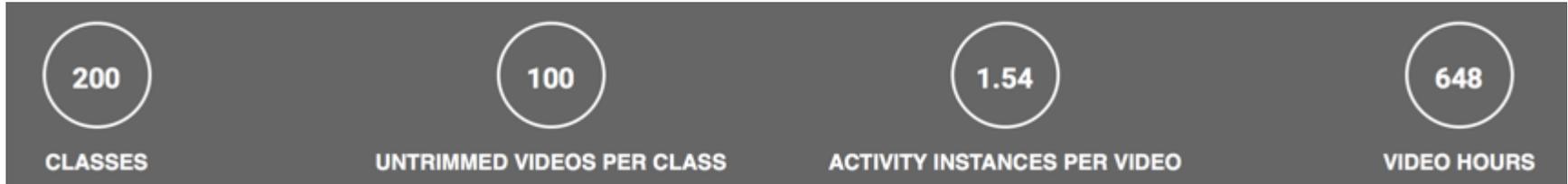
時系列(XYT)解析

メディア解析のための時系列特徴抽出

- YouTubeなど動画共有サイトの動画像を解析
- 1動画に対して1ラベルを付与する問題設定が主流

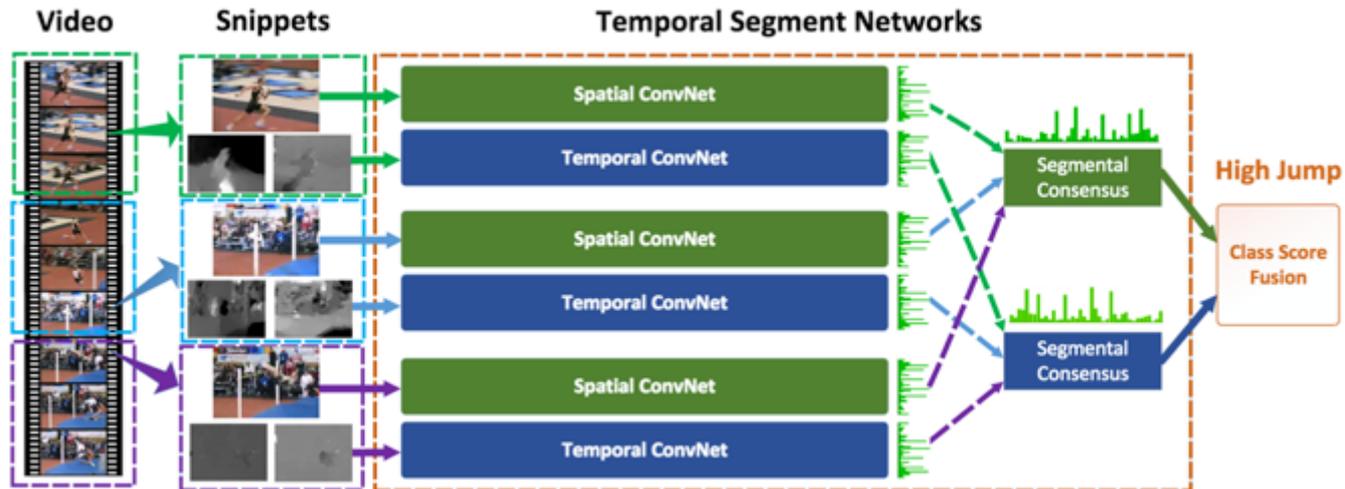


ActivityNet Challenge



動画認識の世界選手権

- 200クラス分類問題にてTop-1が88%, Top-3が96%
- Temporal Segmental Networks [Wang+, ECCV2016]により特徴抽出



行動予測



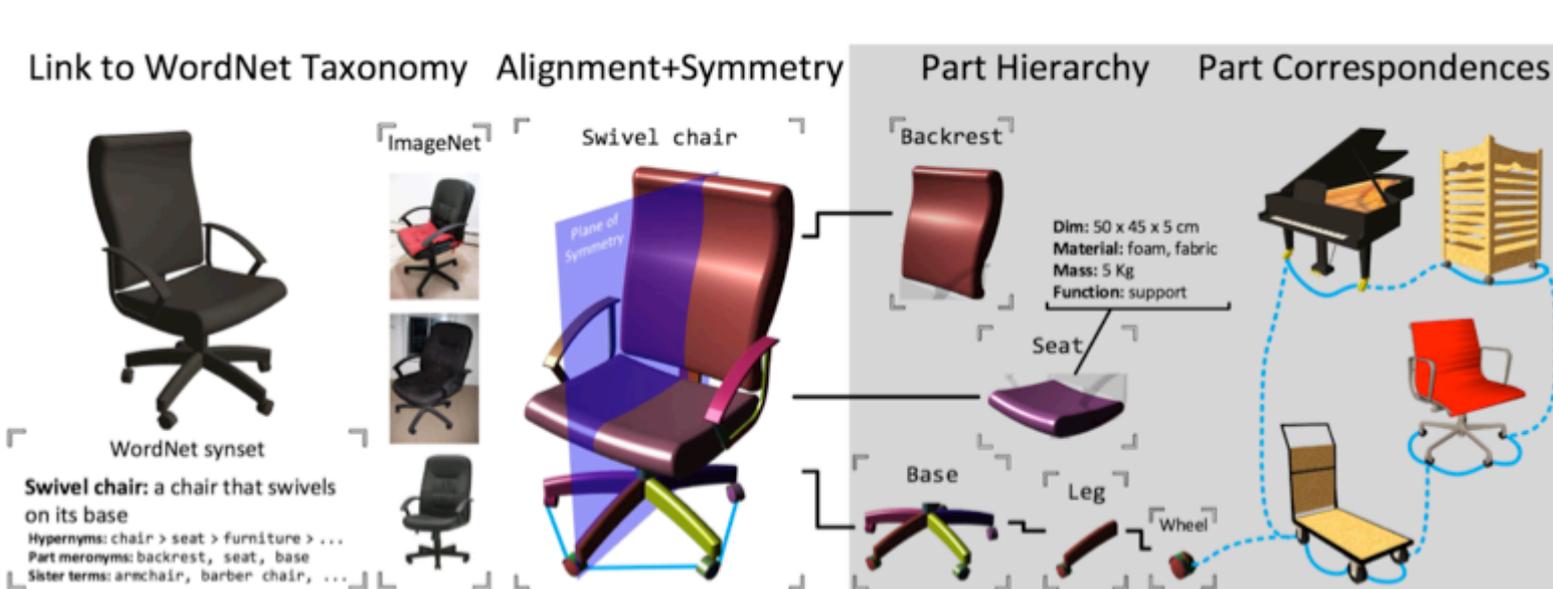
Sudden motion changes that are recognized as abnormal or dangerous should be detected early in traffic and surveillance scene.

3D(XYZ)解析

3次元空間の物体を認識



- 大規模DBであるShapeNet (ImageNetの3D版)の例
- 3D CAD Modelが3M, うち220Kが3,135カテゴリに分類
- 物体ごとの位置合わせ、アフォーダンス、重量などのラベル有
- 55カテゴリ(ShapeNet-Core55)におけるコンペティションも開催



Amazon Picking Challenge (APC)

橋本先生・藤吉先生・山下先生の前で私が言うまでもありません。。。



DynamicFusion



Live Input Depth Map



Live Model Output



Live RGB Image (unused)



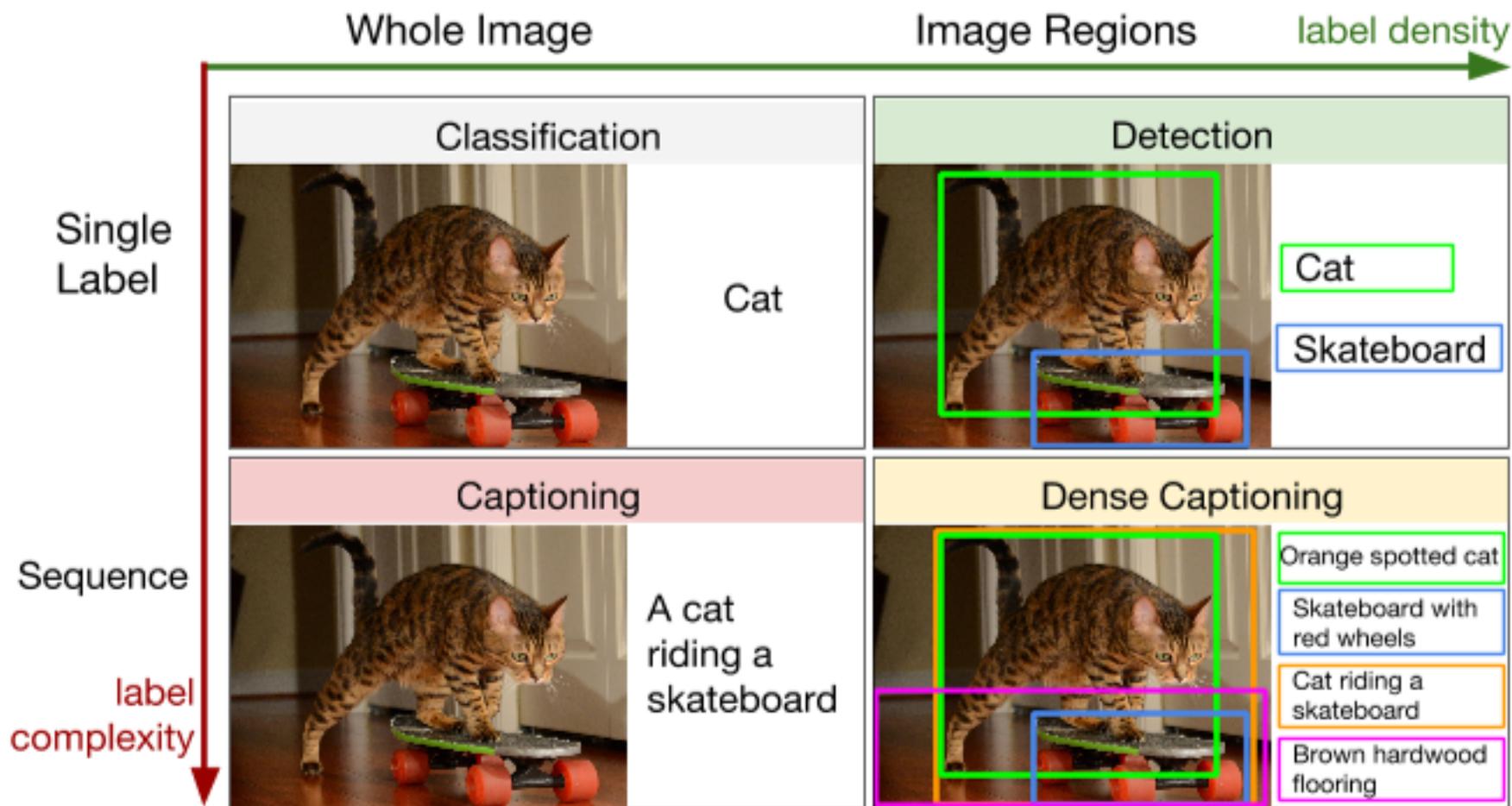
Canonical Model Reconstruction



Warped Model

Dense Captioning

画像中のあらゆる領域を説明する文章を生成



Face2Face

Real-time Facial Reenactment

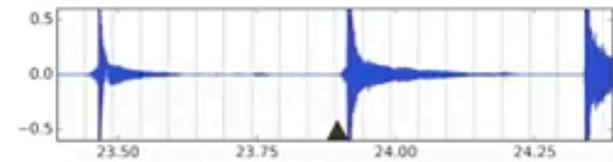


Live capture using a commodity webcam

Visually Indicated Sounds



Silent video



Predicted soundtrack

Lip Reading in the Wild

BBC Two HD
16-May-2016 23:11:35



WE HAVE TO LOOK AT WHETHER IT WORKS FOR THE UK OR NOT

こんなに面白いことが出てきている！(&これからもできそう)

– ぜひ当事者になりましょう！

いかに「優れた問い」を見つけるか？

結論から

サーベイだけではダメ、実装だけではダメ

– 両者を繋げるのはアイデア考案と徹底的な議論

– 論文読破・まとめ・アイデア考案・

議論・実装・論文執筆が一連の流れに！

論文を「読破」「まとめ」

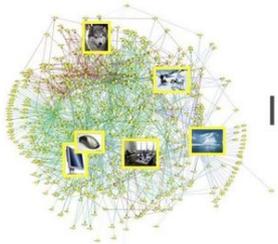
「多読」と「精読」のバランスが重要

- （注力分野に関しては）論文を網羅的に調査し歴史・トレンド把握
- 多読：概要レベルで「問題設定」を読む
- 精読：実装することを目的に「徹底的に」読む

（注）読みながらアイデア考案を同時に行うと流れが良くなります

入力のイメージ

点 (1論文)



IMAGENET

入力のイメージ

点 (1論文) => 線 (1分野)



まとめの例

テンプレートに従って1論文1ページ

- 素早く研究の肝をつかむ
- 【概要】 【新規性・差分】 【手法】 【結果】 【Links】

【1】



Jianxiong Xiao, James Hays, Krista A. Ehinger, Aude Oliva, Antonio Torralba, “SUN Database: Large-scale Scene Recognition from Abbey to Zoo”, in CVPR2010.

Keywords: Dataset, Scene Categorization, Benchmark, Recognition

概要

コンピュータビジョンにおいてシーン認識のデータベースである Scene Understanding (SUN) databaseを提案. シーン認識の裾野を広げた.

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.

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.

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_oopm.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>

アイデアの「考案」「議論」

「思いつき」から「研究テーマ」レベルへ

1. 個人のアイデア考案・グループ内でのブレスト
2. グループ内の議論

上記の繰り返し（3～4回繰り返せば研究テーマレベルに昇華）

個人ではなく、チームの力でテーマ設定

- テーマ設定に学生を参加させる
- バイアスを排除し思考を拡張

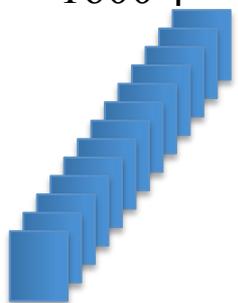
論文を「書く」

新しい問題設定を作る！

- 論文を網羅的に調査して歴史やトレンドを把握
- 膨大な知識から「洗練された」問題設定を世に送り出す
- アイディア考案・議論・実装の繰り返し

次のスライドから図示します！

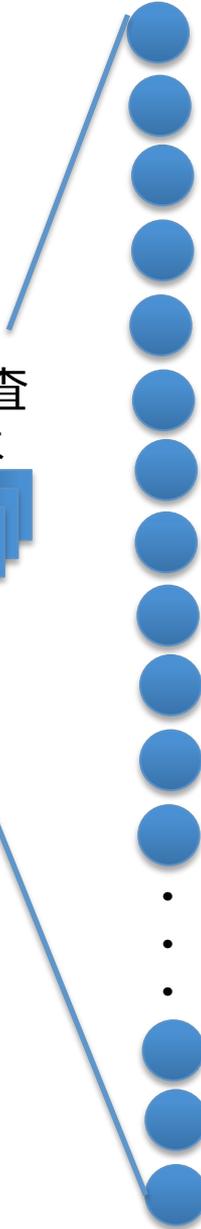
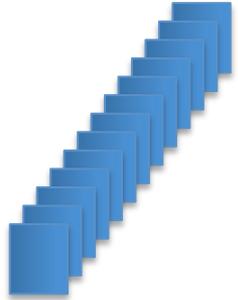
論文調査
1600本



1st アイディア

333

論文調査
1600本



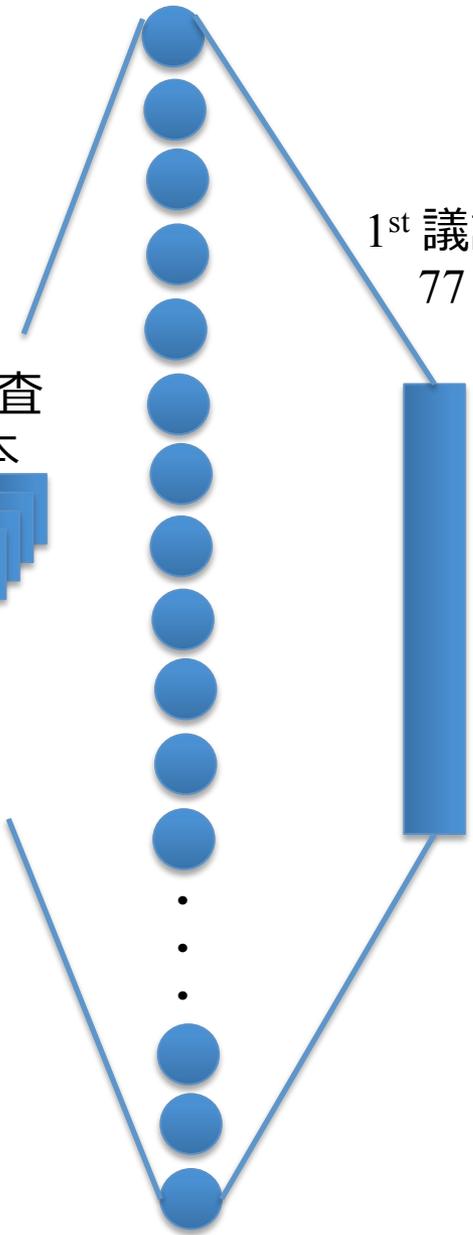
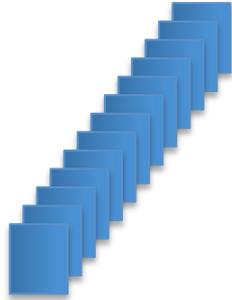
1st アイディア

333

1st 議論

77

論文調査
1600本



1st アイディア

333

2nd アイディア

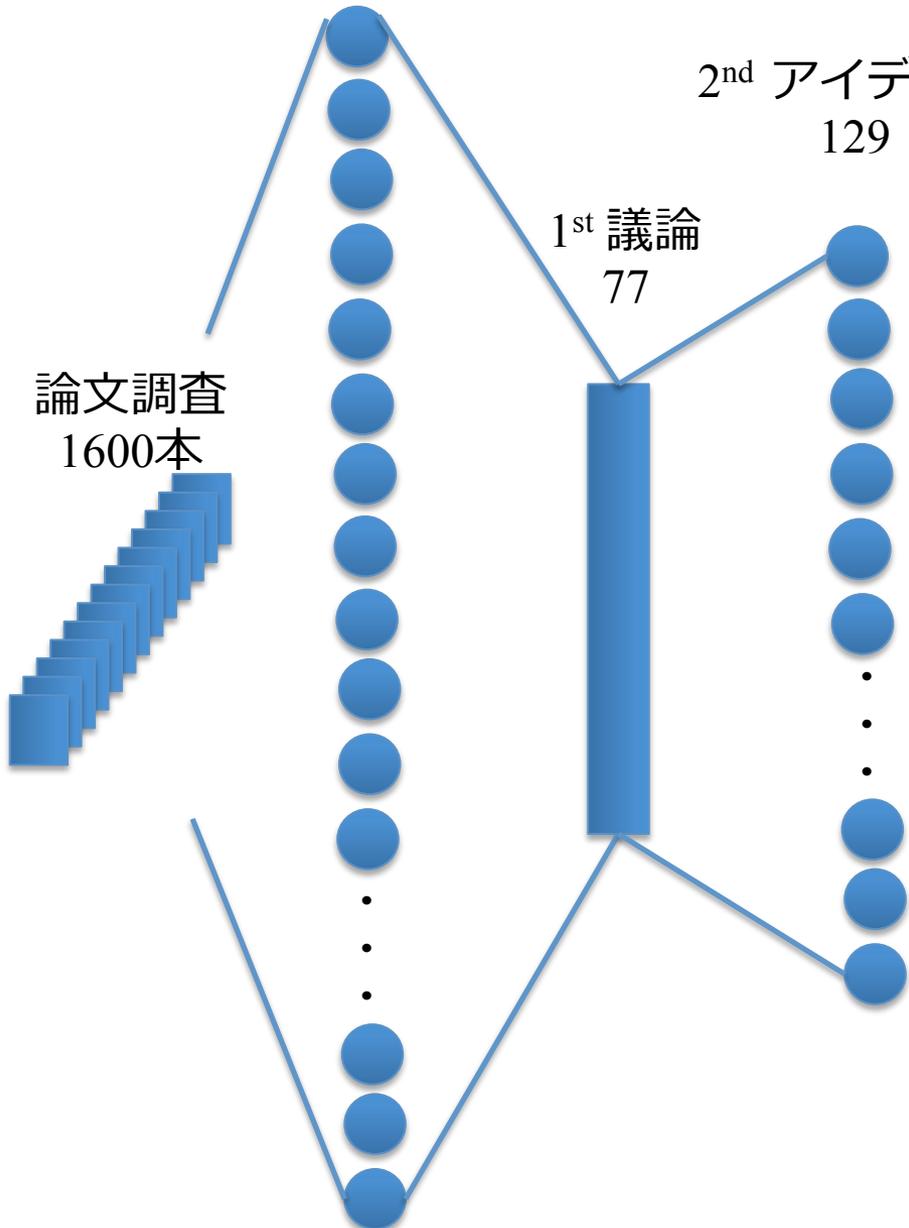
129

1st 議論

77

論文調査

1600本



1st アイディア

333

2nd アイディア

129

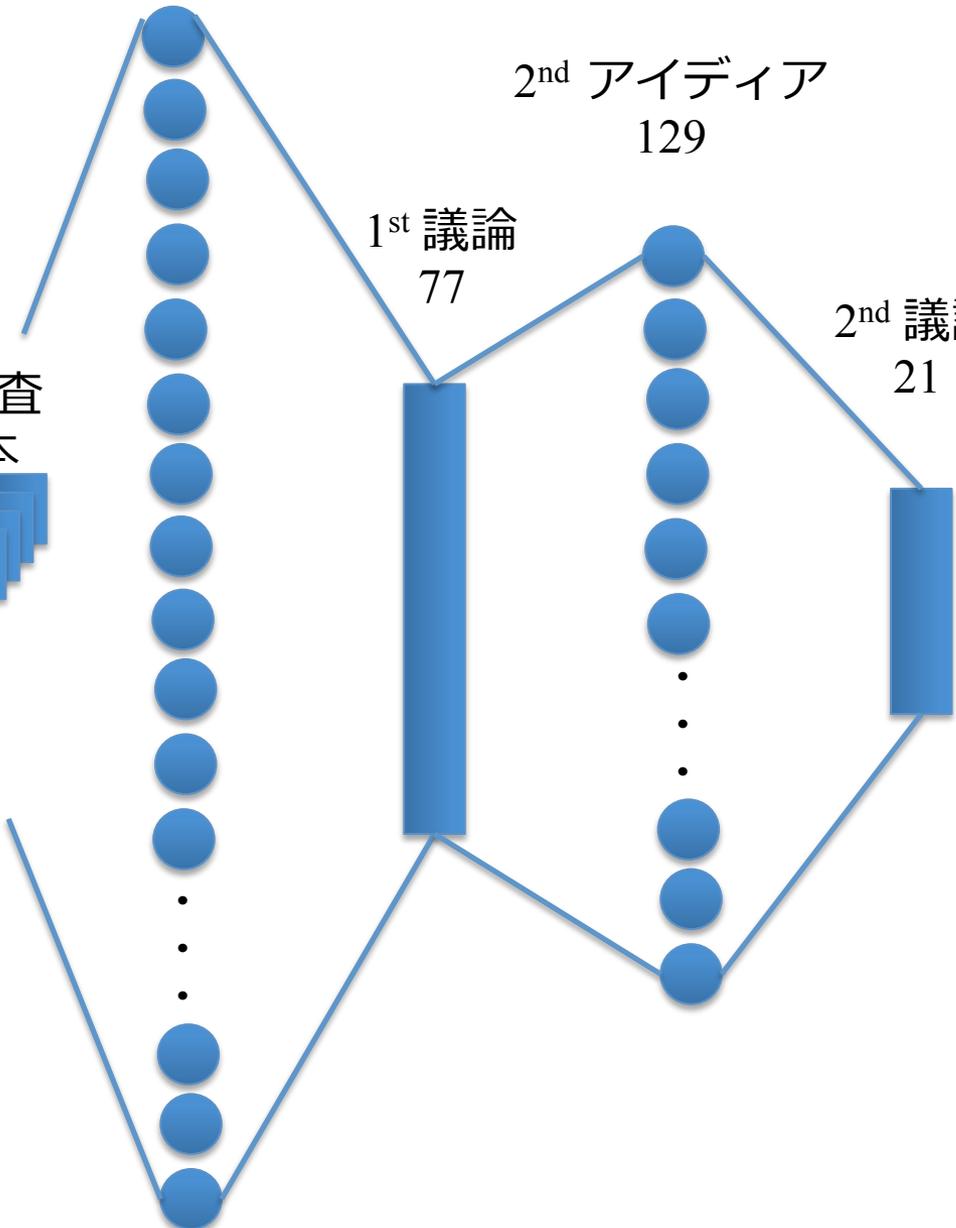
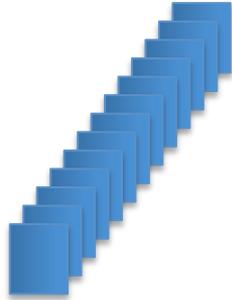
1st 議論

77

2nd 議論

21

論文調査
1600本



1st アイディア

333

2nd アイディア

129

1st 議論

77

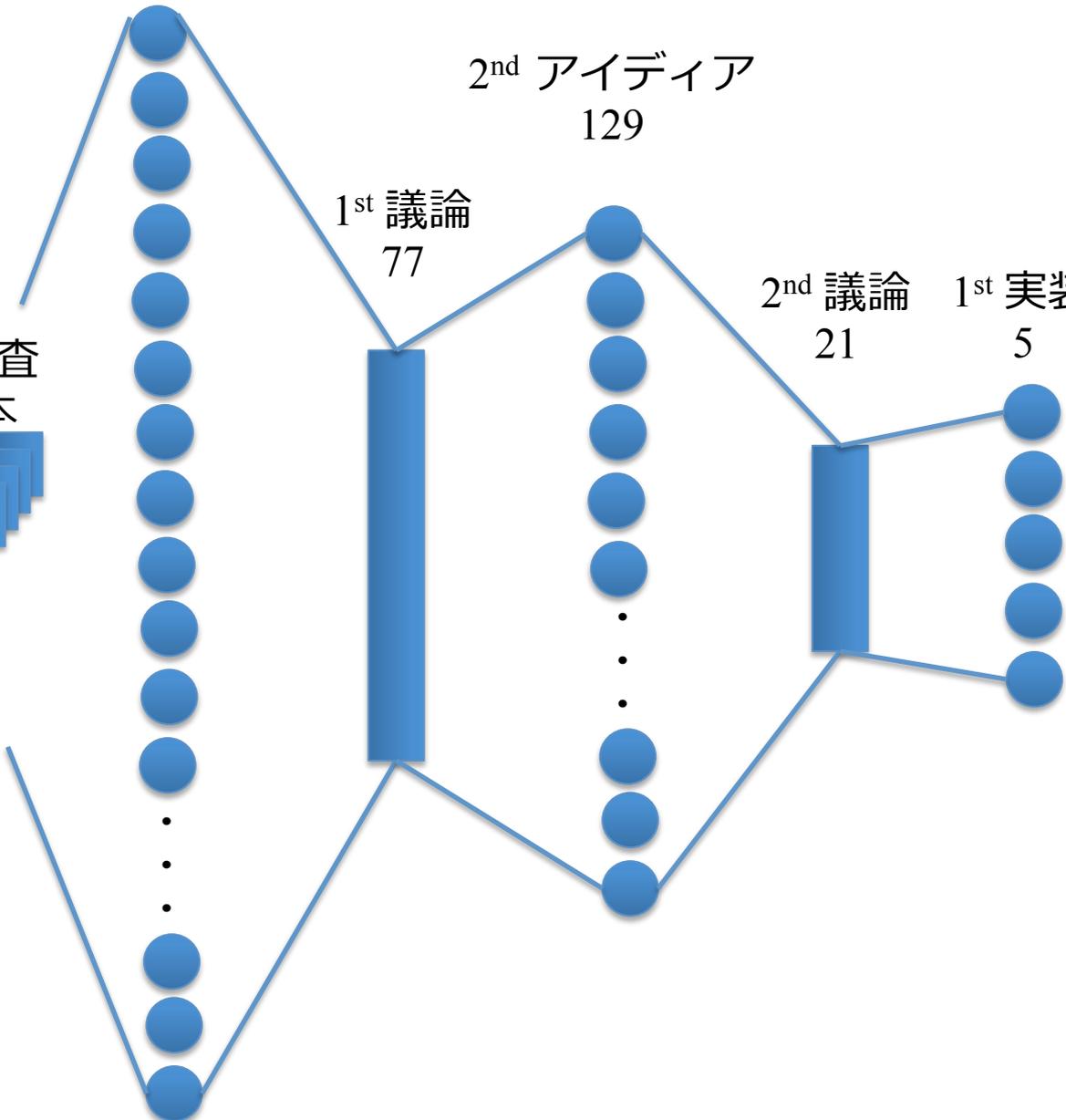
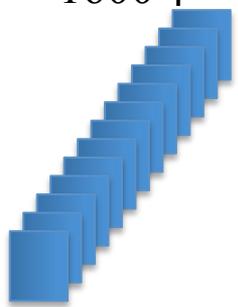
2nd 議論

21

1st 実装

5

論文調査
1600本



1st アイディア

333

2nd アイディア

129

1st 議論

77

2nd 議論

21

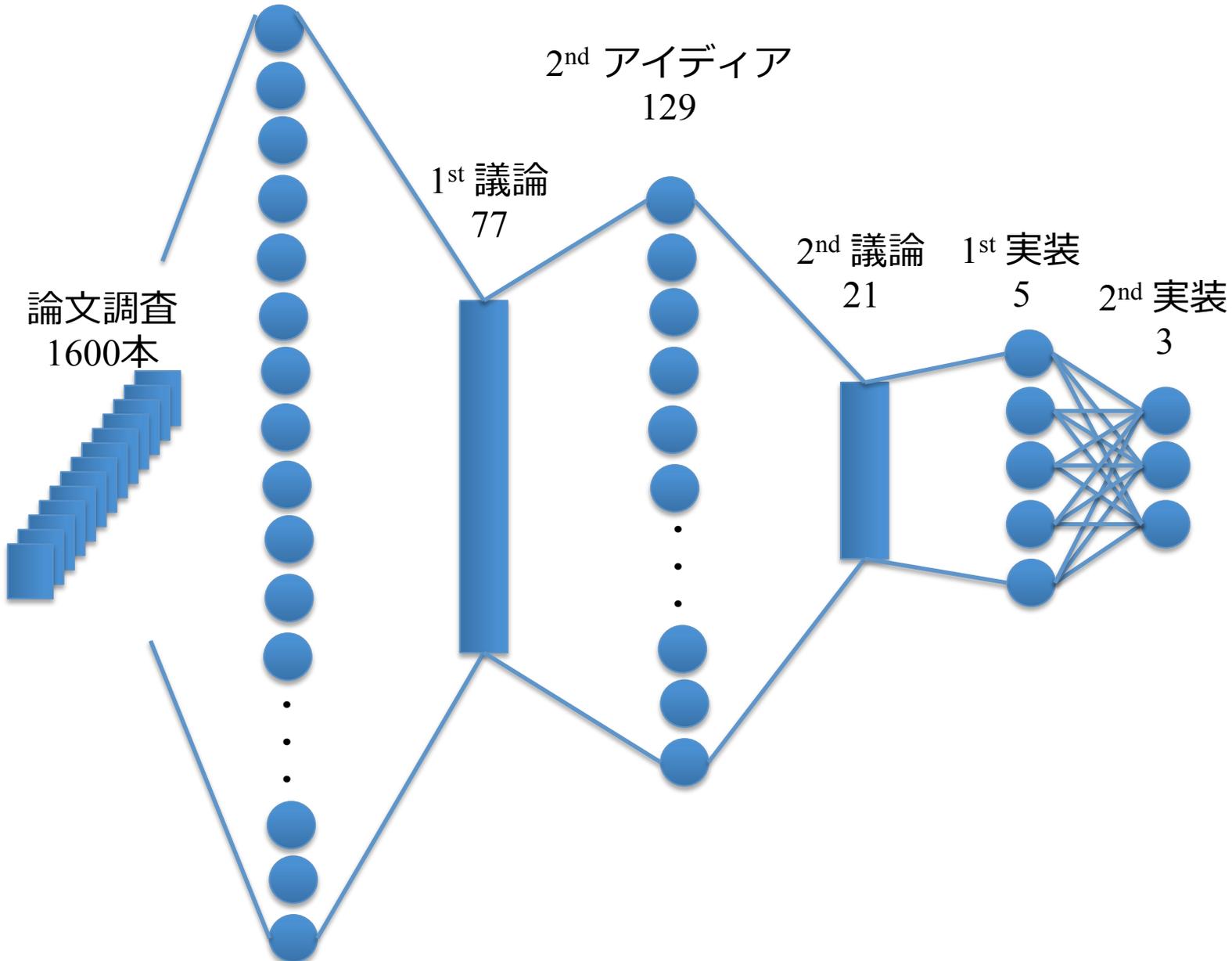
1st 実装

5

2nd 実装

3

論文調査
1600本



1st アイディア

333

2nd アイディア

129

1st 議論

77

2nd 議論

21

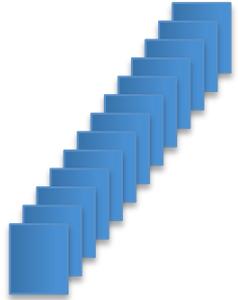
1st 実装

5

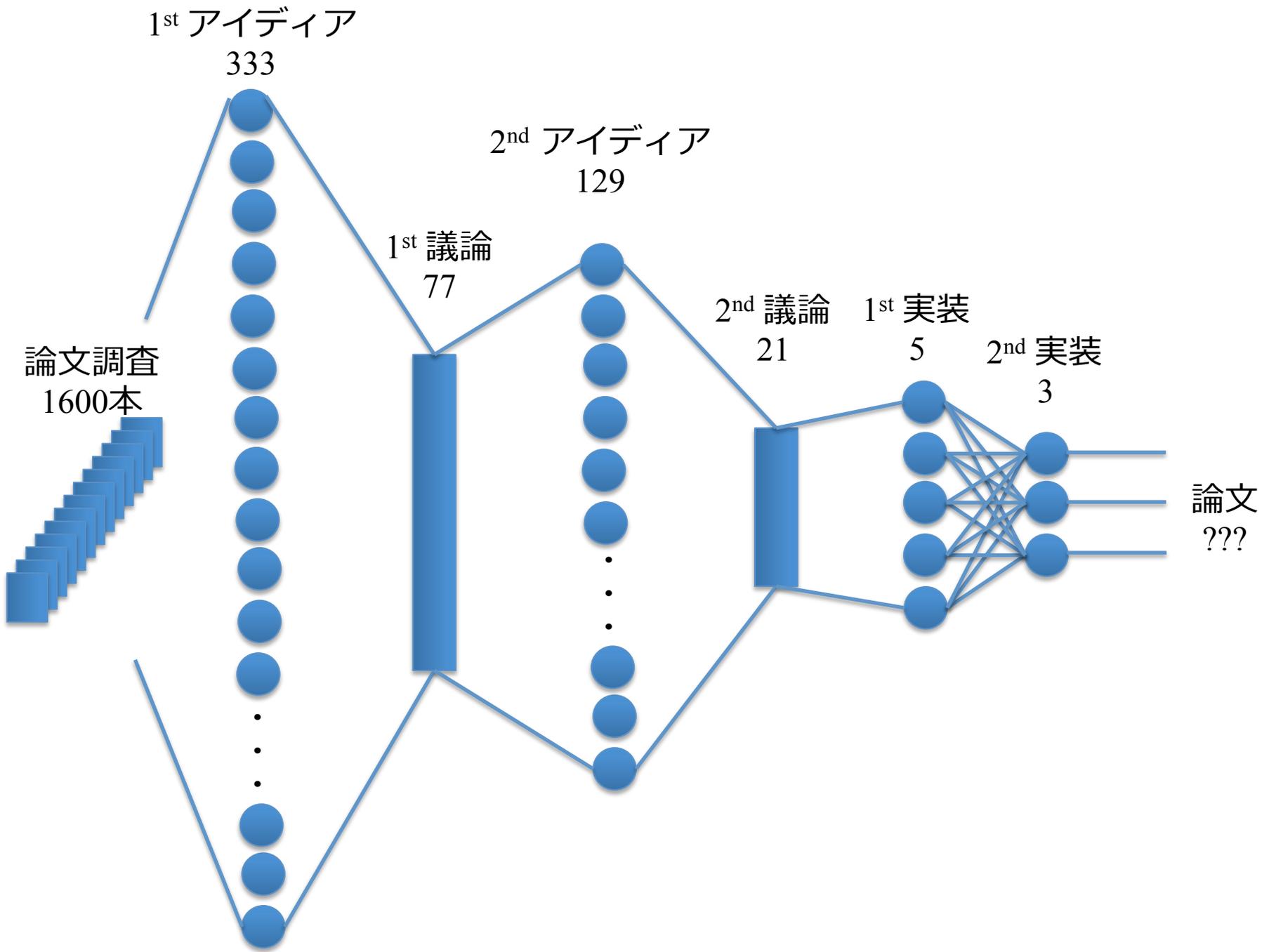
2nd 実装

3

論文調査
1600本



論文
???



これ、どこかで見た事ありませんか？

1st アイディア
333

2nd アイディア
129

1st 議論
77

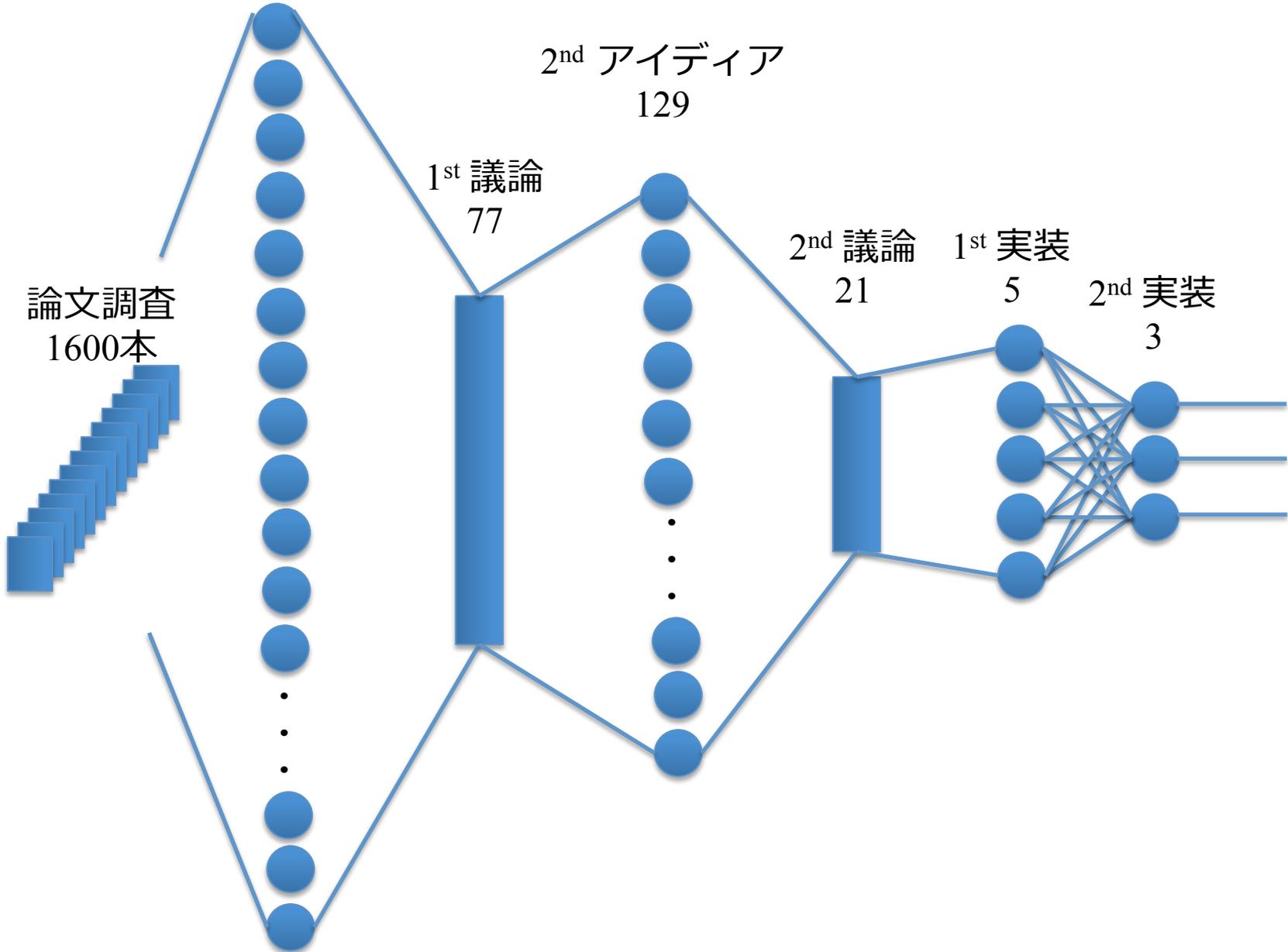
2nd 議論
21

1st 実装
5

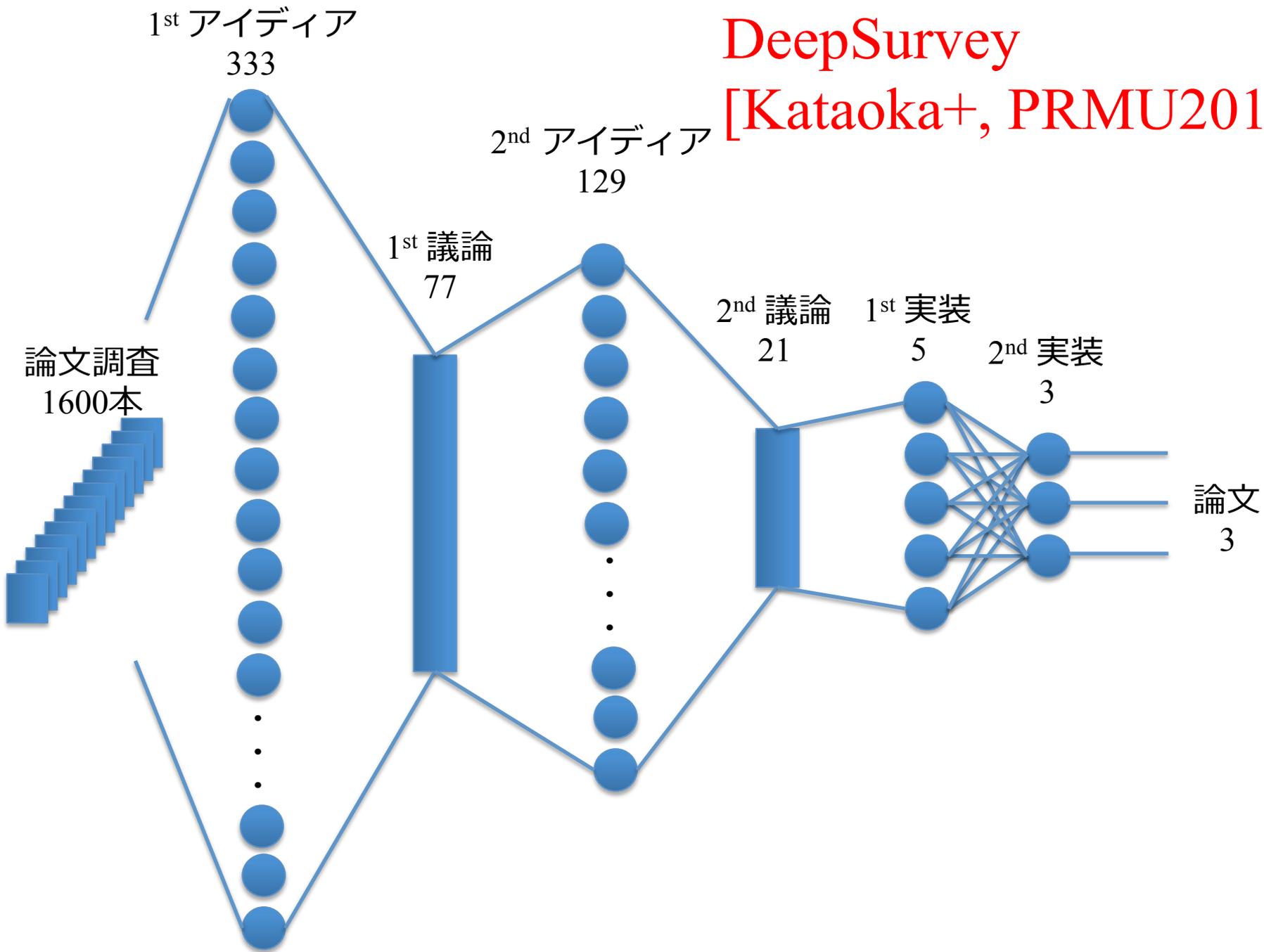
2nd 実装
3

論文調査
1600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になる」という画期的な手法
- 「出力が論文という形式で出てくる」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
- 「学習によりニューロン自体が成長する」という画期的な手法

目指すはVGG (16, 19層) ? ResNet (50, 101, 152, 200層) ?

DeepSurvey Model-2016は7層構造 (4-conv/pool, 2-fc, 1-output)



http://www.robots.ox.ac.uk/~karen/pdf/ILSVRC_2014.pdf

http://kaiminghe.com/icml16tutorial/icml2016_tutorial_deep_residual_networks_kaiminghe.pdf

良い出力のためには？

実施期間は重要（オリジナルの概念）

- 畳み込むタイミングは非常に重要（ニューロンは忙しい）

ニューロン同士、仲良くする（オリジナルの概念）

- コミュニケーションが研究の進度に直結

読む論文数が多いほうが良い（1Dより2Dの入力）

- 点 (Pixel), 線 (Line)より面 (Image)の入力

キーとなる論文の理解は重要（勾配消失防止）

- 局所的に良い勾配を持ち，領域は大きいほうが良い

ブレストと議論の数は多いほうが良い（Conv/Pool Layer）

- 畳み込み，プーリング数を増やす

実装はグループごとに密に行う（Multi-stream Networks）

- 全結合は複数のストリームに分ける

良い出力のためには？ その2 [new!]

メンバーの入れ替わりは重要（ドロップアウト）

- 欠席・日替わりグループ等によるメンバー入れ替えは異なる視点を生む

フィードバックは極めて重要（誤差逆伝播法）

- 早いペースでの論文投稿・誤差逆伝播、即ち繰り返し回数が良い論文の執筆には必要

アイディアの復活（Residual Learning/Skip Connection）

- 昔のアイディアの復活と現在のアイディアへの統合は良いアイディアになることが多い

DeepSurveyは、今も進化中。。。。

本プロジェクトでの問題設定と成果

Human Action Recognition without Human

(人を見ない人物行動認識; ViEW若手奨励賞、ECCV16WS Oral&BestPaper)



Oscar Prediction



score

0.82

0.65

0.45

0.43

0.37

0.34

0.28

0.18

Prediction

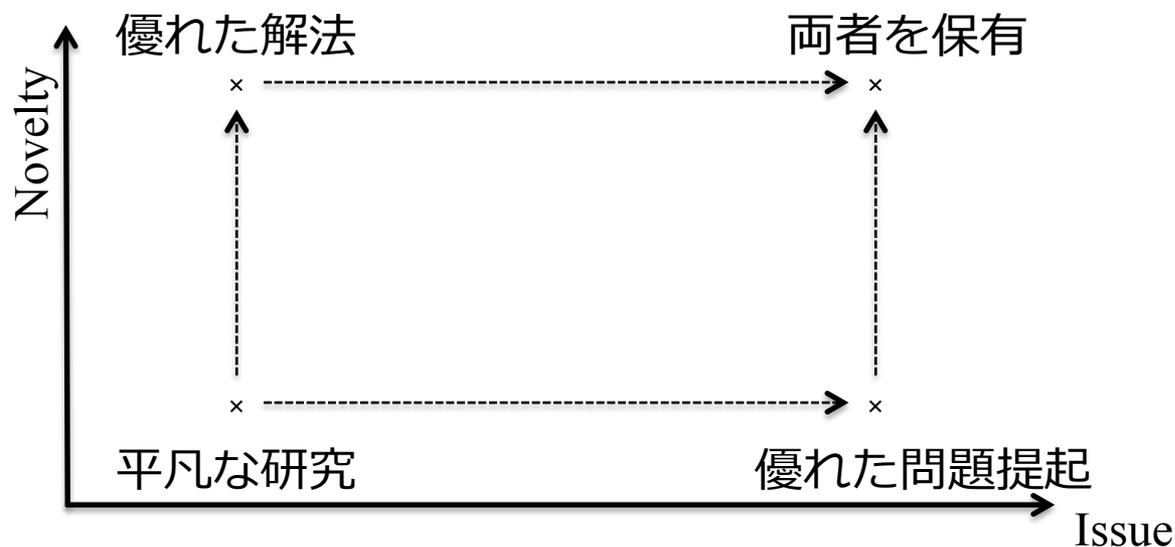
Y. Matsuzaki & K. Okayasu et al. "Can you guess an interesting movie from the posters?" MVA 2017 (accepted)

今後の取り組み

2017年の研究戦略について

最先端の研究を行うために

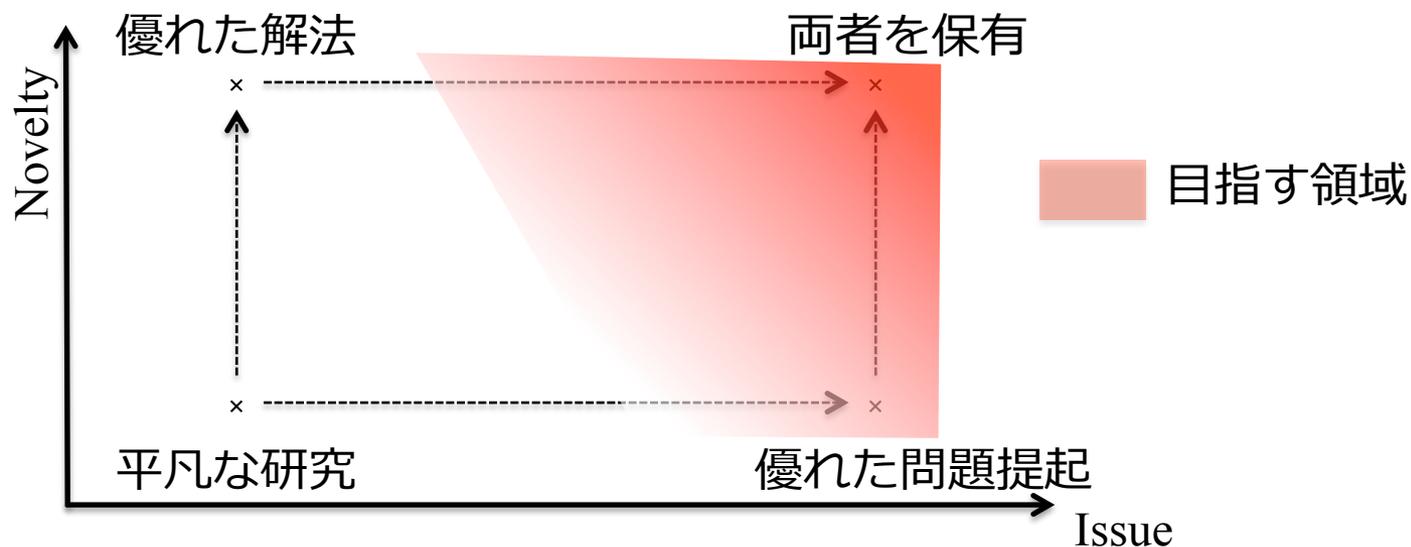
1. 【問題設定：Issue】
2. 【解の質：Novelty】
3. 上記1, 2のいずれかを保持するもしくは両方同時に持つ研究を実行



2017年の研究戦略について

最先端の研究を行うために

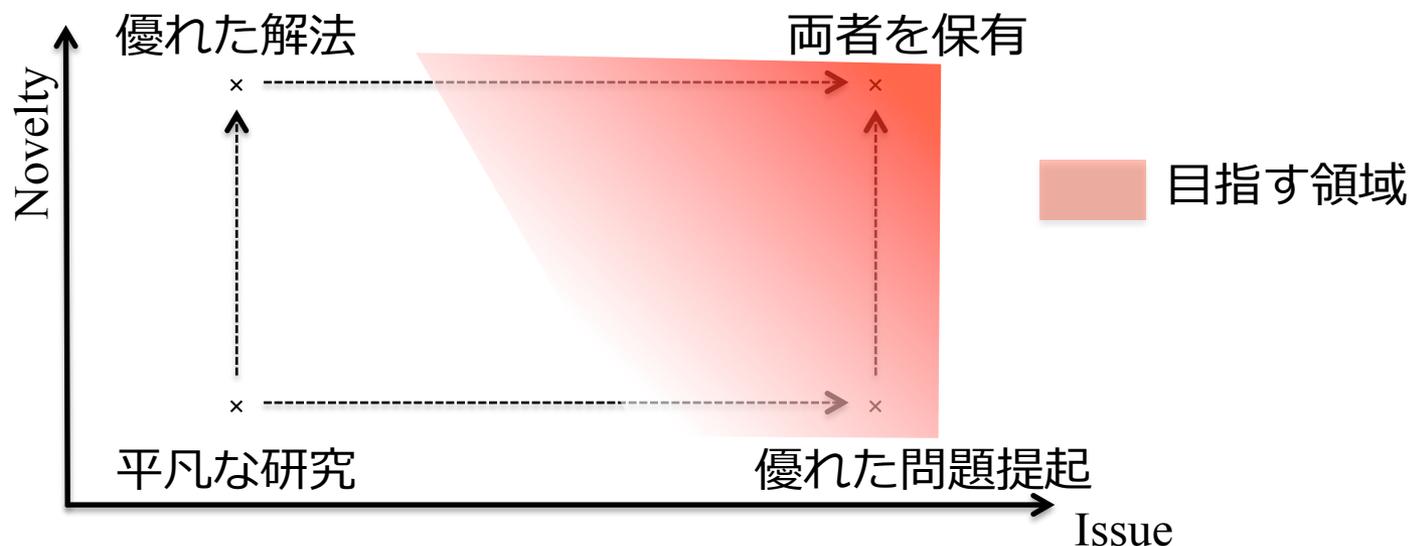
1. 【問題設定：Issue】
2. 【解の質：Novelty】
3. 上記1, 2のいずれかを保持するもしくは両方同時に持つ研究を実行



2017年の研究戦略について

最先端の研究を行うために

1. 【問題設定：Issue】
2. 【解の質：Novelty】
3. 上記1, 2のいずれかを保持するもしくは両方同時に持つ研究を実行



指針として掲げたのが「cv-top10」

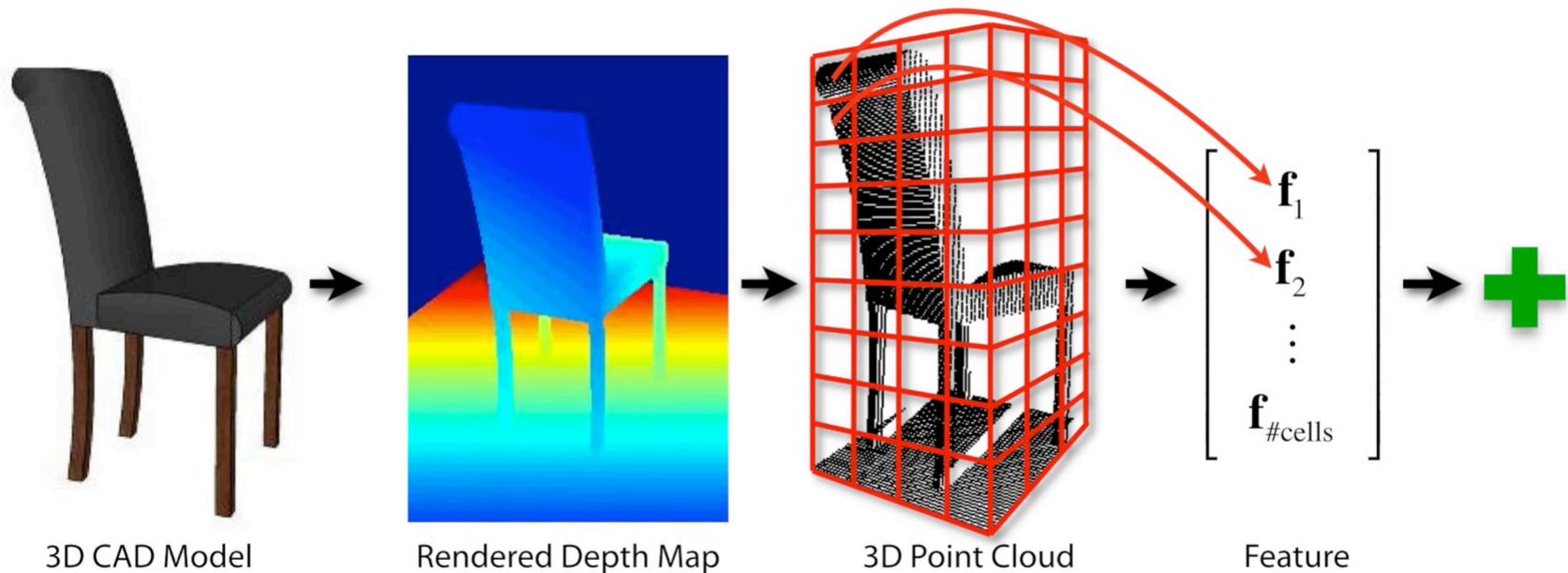
=> 詳細はcvpaper.challenge in PRMU Grand Challenge 2016参照

注力分野：*cv-top10*

1. 動画の純粹なモーシヨソ推定
2. 3次元特徴の普遍化
3. 世界規模の認識や復元
4. WebやCG, オートメーションを用いた完全自動データ生成
5. 物理量の変換や理解
6. 画像に映らない背景知識を見る
7. 自己学習
8. より少ないデータで汎用性を出す学習技術
9. 歴史は繰り返す？
10. ピクセルの再定義

2. 3次元特徴の普遍化

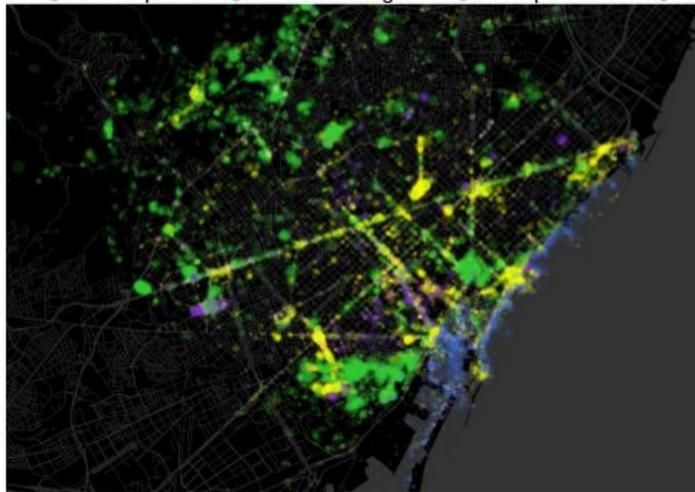
Training: Positive Exemplar



The object is represented by a 3D grid, and its feature is used as a positive example.

3. 世界規模の認識や復元

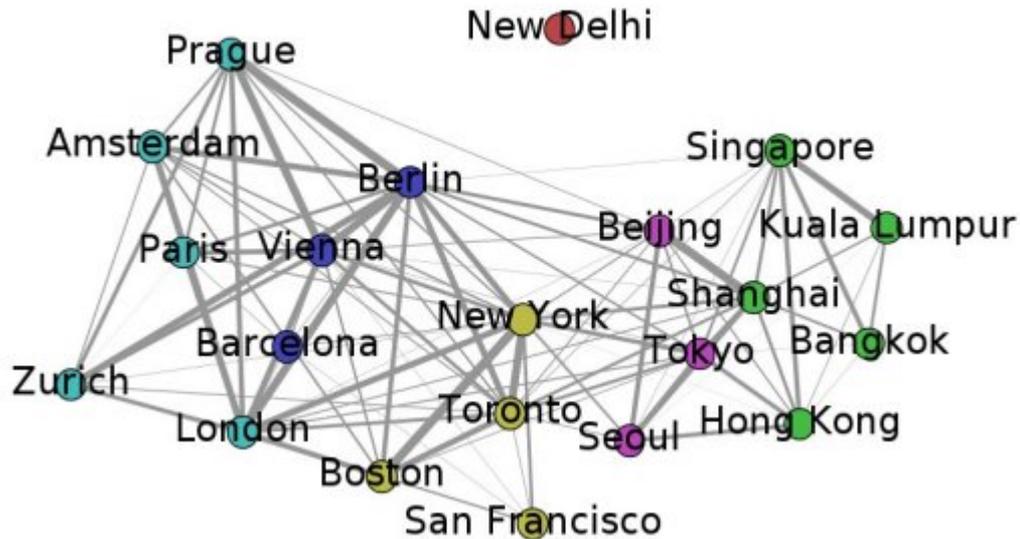
● Green space
 ● Water coverage
 ○ Transportation
 ● Architecture
 ● Vertical building
 ● Athletic activity
 ● Social activity



Barcelona

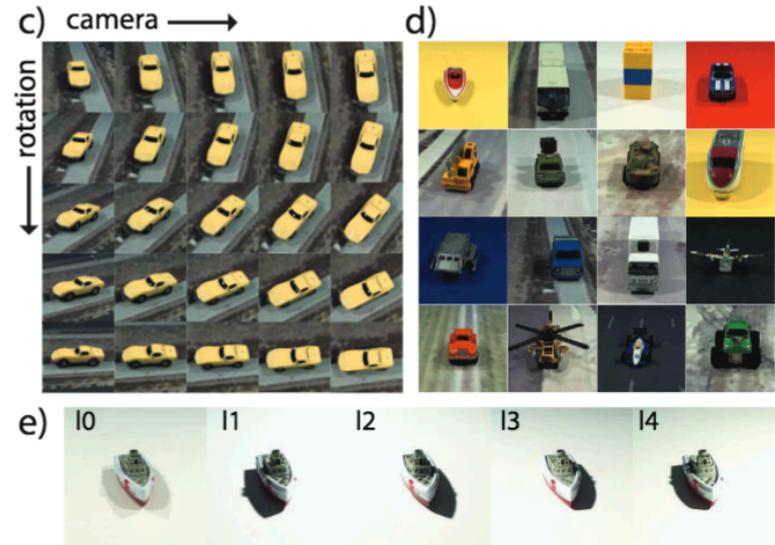
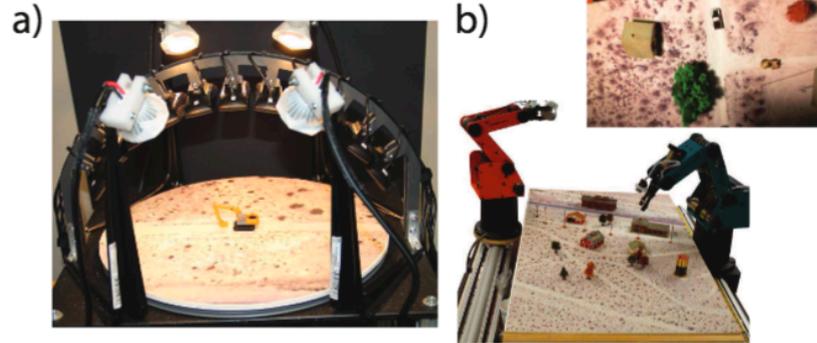


New York City



Bolei Zhou et al. "Attribute Analysis of Geo-tagged Images for City Perception" in ECCV14.

4. WebやCG, オートメーションを用いた完全自動データ生成

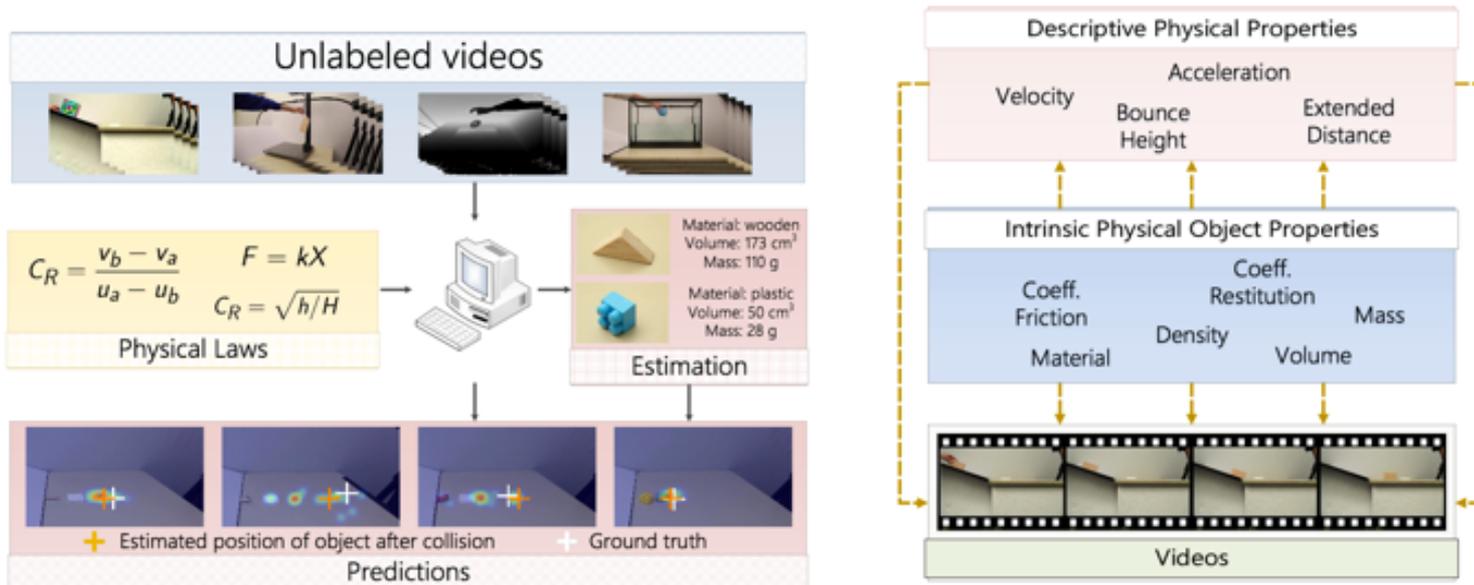


Dataset	Ref	Domain	Object Classes	Objects per Class	Backgrd per obj	Views per obj+bg	Bounding Box?	Object Contours?	Total Images
COIL	[41]	Handheld	100	1	1	72	Implicit	No	7,200
SOIL-47	[29]	Handheld	—	47	1	42	Implicit	No	1,974
Pascal	[14]	Misc	20	790-10,129	1	1	Yes	Partial	11,540
Caltech-101	[15]	Google	102	31-800 ($\mu = 90$)	1	1	No	No	9,144
Caltech-256	[22]	Google	257	80-827 ($\mu = 119$)	1	1	No	No	30,607
LabelMe	[48]	Misc	900	?	~ 1	~ 1	Partial	Partial	62,197 (a)
NORB	[35]	Toys	5	10	1 (b)	1,944	Implicit	No	48,600 (b)
FERET	[44]	Faces	1	1,199	1	1-24	Yes	No	14,051
MNIST	[34]	Digits	10	6,000	1	1	Implicit	No	60,000
ETHZ	[17]	Natural	5	32-87	1	1	Yes	Yes	255 (c)
TINY	[61]	Web	75,062	?	1 (?)	1	Implicit	No	79,302,017 (d)
CIFAR-100	[30]	Web	100	600	1	1	Implicit	No	60,000 (d)
ALOI	[19]	Handheld	1,000 (e)	~ 1	1	108	Implicit	No	110,250
GRAZ	[42]	Photographs	4	311-420	1	1	No	Partial	1,476
CoPhIR	[3]	Flickr	? (f)	?	1 (?)	1 (?)	No	No (f)	106,000,000
ImageNet	[8]	Misc	21,841	~ 1	~ 1	~ 1	Yes	No	14,197,122
SUN	[64]	Misc	3,819	(g)	1	1	Yes	Yes	131,067
MS COCO	[38]	Misc	91	$\sim 5,000$	1	1	Yes	Yes	328,000 (a)
RGB-D	[33]	Household	51	~ 6	1	250	Yes	No	250,000
Big-BIRD	[55]	Household	100	1	1	600	Yes	No	250,000
iLab-20M	—	Toy vehicles	15	25-160	14-40	1,320	Implicit	No	21,798,480

5. 物理量の変換や理解



A. Owens et al. "Visually Indicated Sounds" in CVPR16.



J. Wu et al. "Physics 101: Learning Physical Object Properties from Unlabeled Videos" in BMVC16.

6. 画像に映らない背景知識を見る

Powered by H. Kataoka

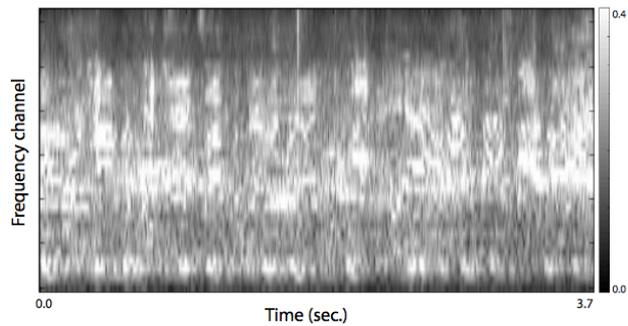


In a traffic scene, short-term action predictions are particularly crucial for avoiding accidents between humans and vehicles.

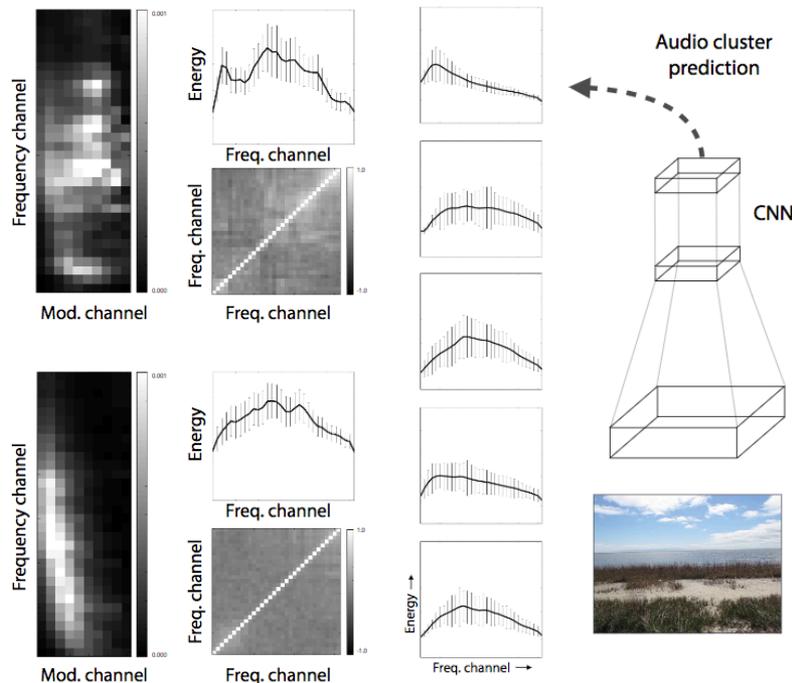
7. 自己学習



(a) Video frame



(b) Cochleagram



(c) Summary statistics

A. Owens et al. "Ambient Sound Provides Supervision for Visual Learning" in ECCV16.

8. より少ないデータで汎用性を出す学習技術

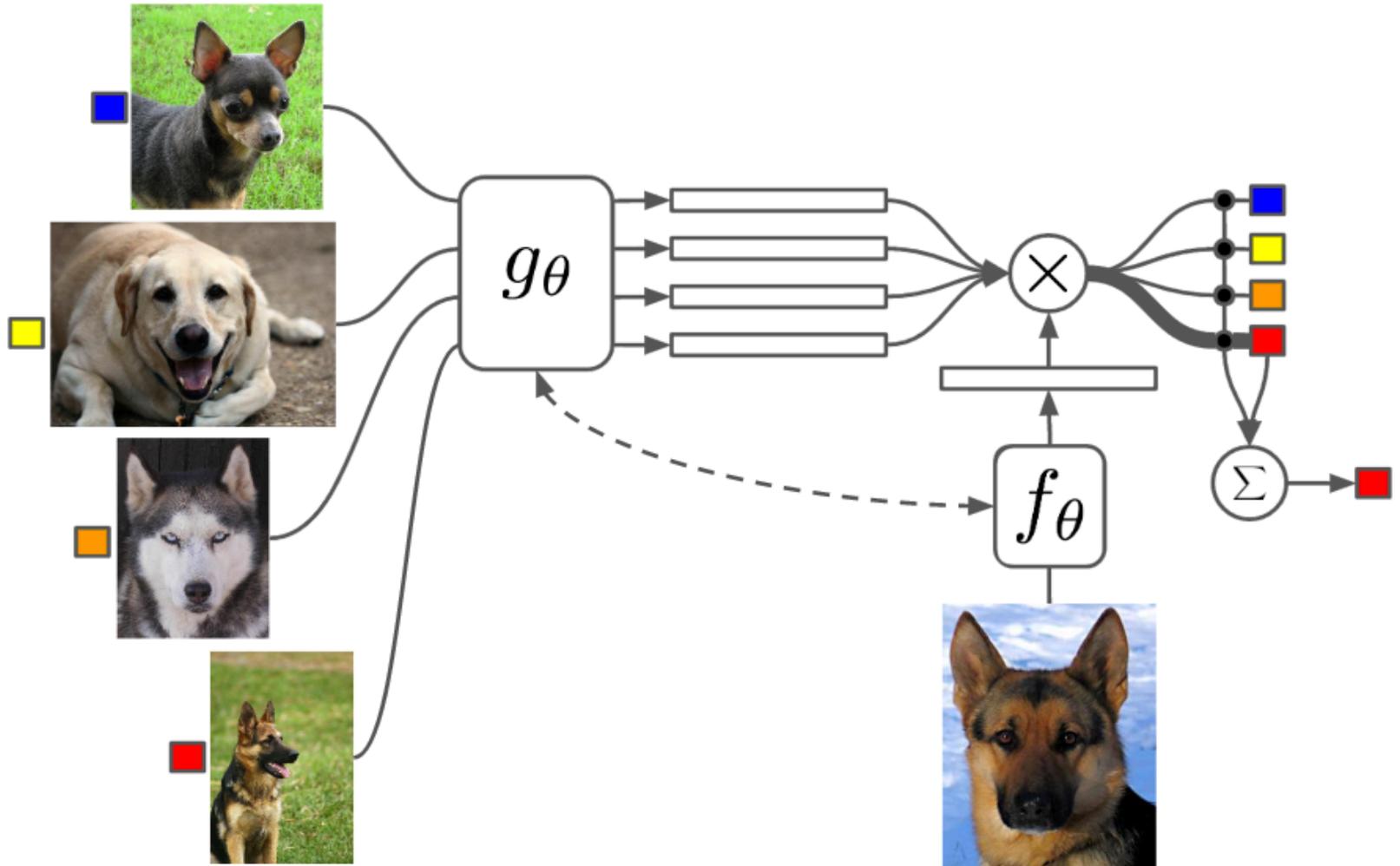
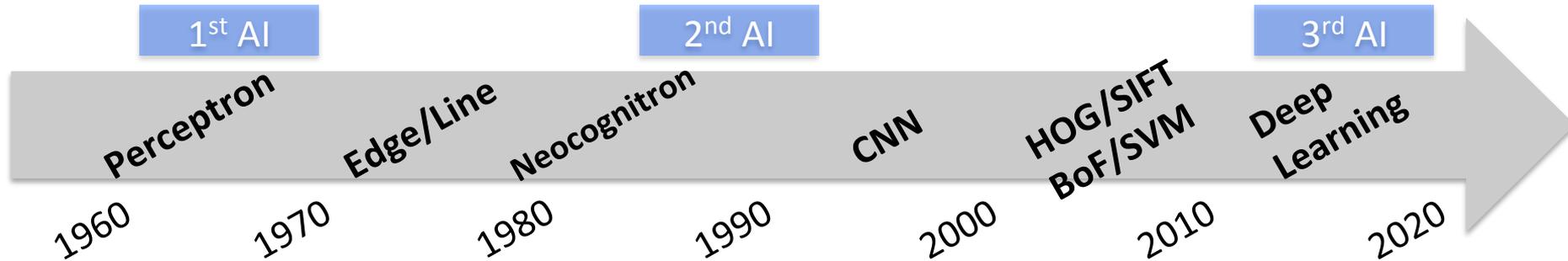


Figure 1: Matching Networks architecture

9. 歴史は繰り返す？



F. Rosenblatt et al. "Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms" in 1961.

J. F. Canny "Finding Edges and Lines in Images" in 1983.

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.

J. H. Holland "Adaptation in Natural and Artificial Systems" in MIT Press 1975.

L. J. Eshelman "The CHC Adaptive Search Algorithm: How to Have Safe Search When Engaging in Nontraditional Genetic Recombination," in Foundations of Genetic Algorithms, 1991.

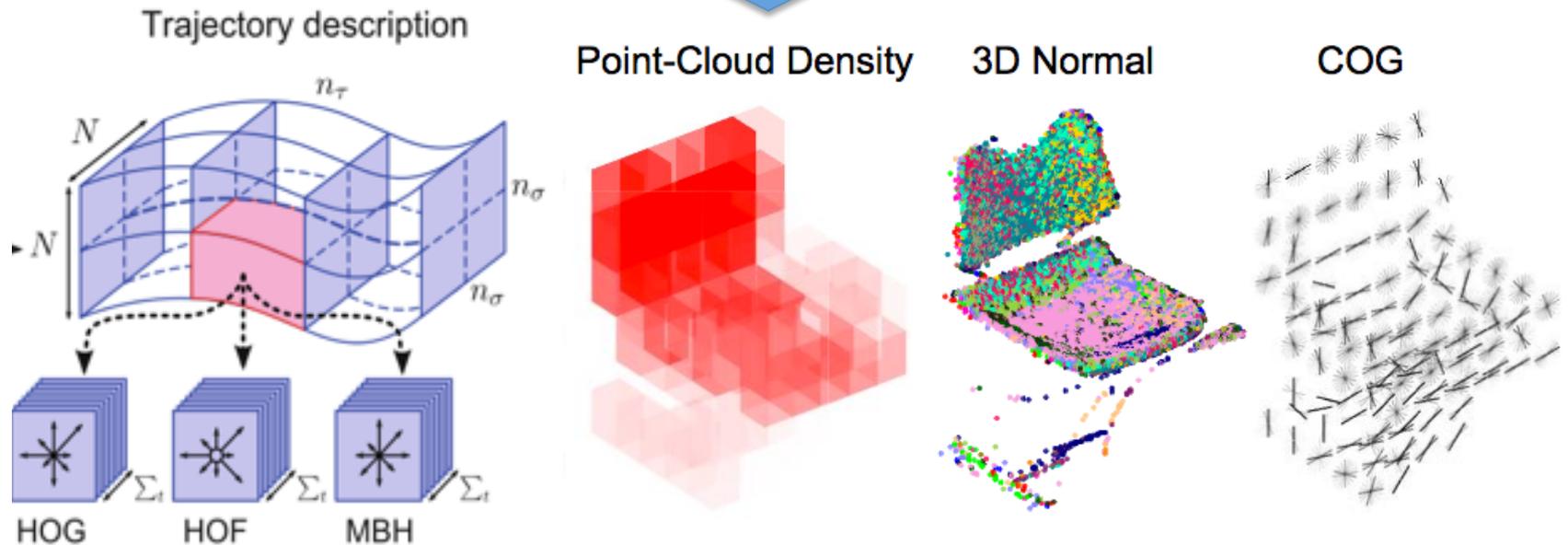
G. Huang et al. "Deep Networks with Stochastic Depth," in arXiv pre-print, 2016.

T. Yamasaki et al. "Efficient Optimization of Convolutional Neural Networks Using Particle Swarm Optimization," in MIRU, 2016.

ニューラルネット・勾配特徴の繰り返し, その間に遺伝的アルゴリズム



ボリュームデータ (e.g. 動画, 3D) への転移

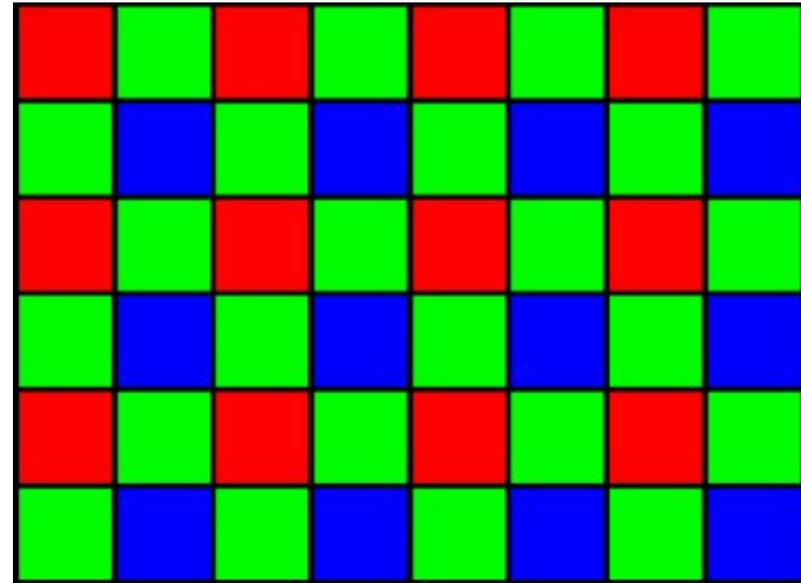
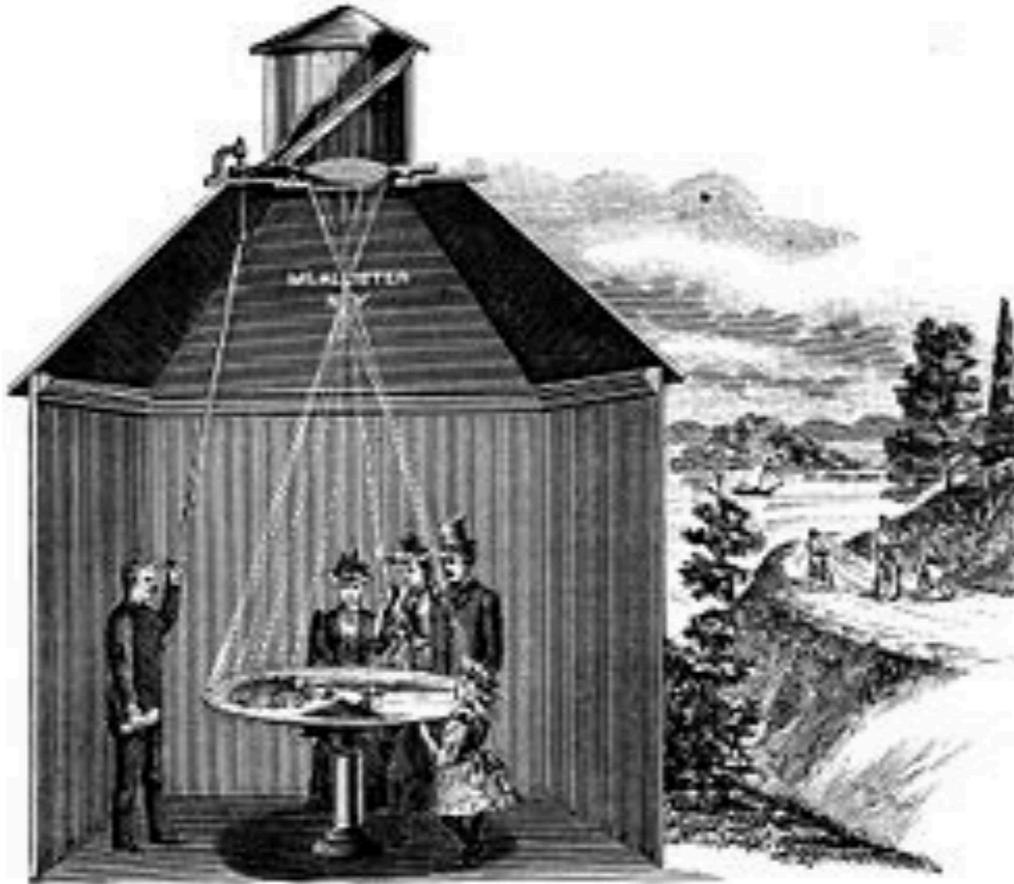


H. Wang et al. "Action Recognition by Dense Trajectories" in CVPR11.

Z. Ren et al. "Three-Dimensional Object Detection and Layout Prediction using Clouds of Oriented Gradients" in CVPR16.

10. ピクセルの再定義

RGBの枠組みで良いのか？ピクセルの表現は？



H. P. Gage. "Optic projection, principles, installation, and use of the magic lantern, projection microscope, reflecting lantern, moving picture machine," in 1914.

State-of-the-art Computational Photography



6,000,000,000,000 (6 TRILLION) FPS? "Ultra High-Speed Camera FEMTO" <https://www.youtube.com/watch?v=ZHW0fIOEasc>

6兆FPSのカメラが登場している時代にRGBセンサでいいの？

撮影した瞬間に認識が終わっている
or 認識しやすい画素形態



<https://researchweb.iiit.ac.in/~dineshreddy.n/zerotype/assets/sms.png>

特殊なものが見えるセンサ

光：前スライド

素粒子：下図

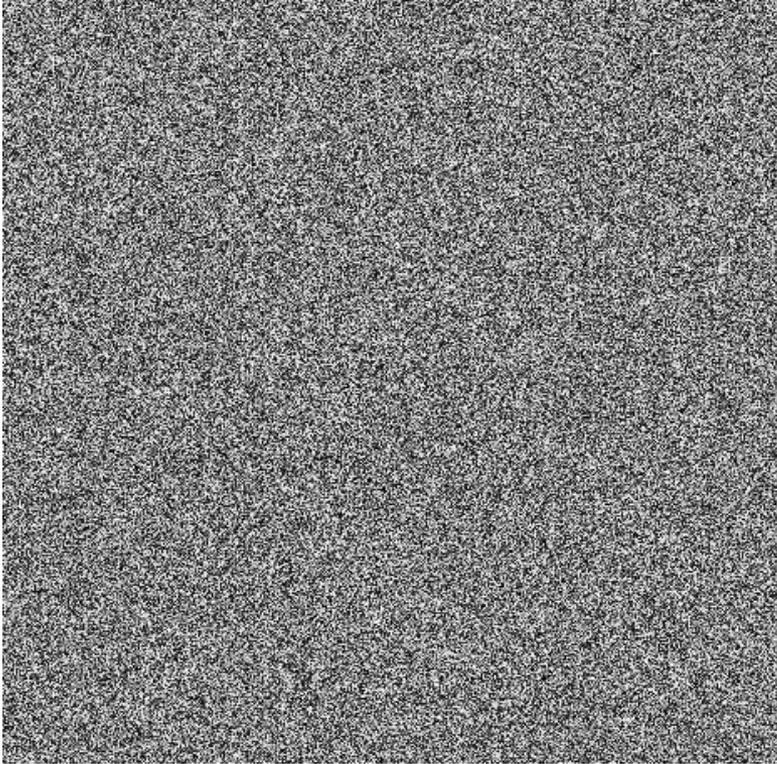
9 or 10 次元の世界：宇宙の始まり

		フェルミオン			ボソン	
クォーク	u アップ	c チャーム	t トップ	γ 光子		
	d ダウン	s ストレンジ	b ボトム	g グルーオン		
	ν_e 電子 ν	ν_μ ミュー ν	ν_τ タウ ν	W W ボソン		
レプトン	e 電子	μ ミューオン	τ タウ	Z Z ボソン	H ヒッグス	

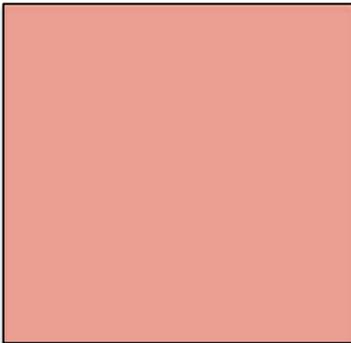
<https://upload.wikimedia.org/wikipedia/commons/thumb/f/f9/%E6%A8%99%E6%BA%96%E7%90%86%E8%AB%96%E7%B4%A0%E7%B2%92%E5%AD%90%E8%A1%A8.svg/1211px-%E6%A8%99%E6%BA%96%E7%90%86%E8%AB%96%E7%B4%A0%E7%B2%92%E5%AD%90%E8%A1%A8.svg.png>

Pixel Space

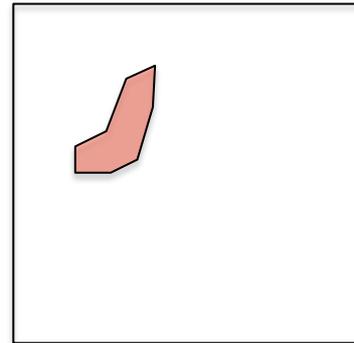
Random



Natural image



Pixel space (ALL)

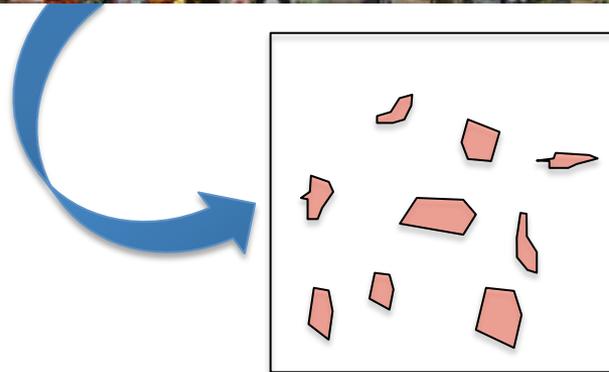


Pixel space (Dog)

自然画像を構成するピクセル空間を推定 物体やシーン，人間など自然な空間というものをモデル化



<http://karpathy.github.io/assets/cnntsne.jpeg>

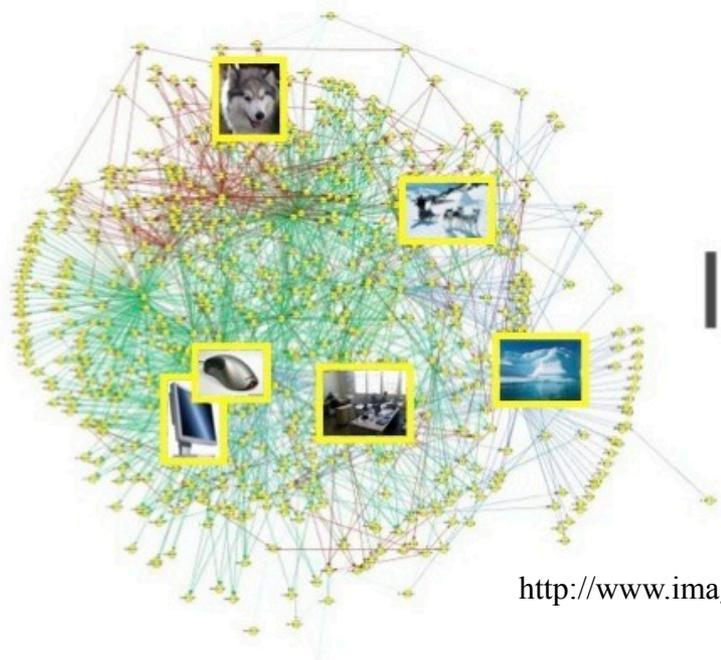


Pixel space (Natural Images)

今後の方針

問題設定と同時にデータセットを作る

- 問題と同時にデータセットを自分で作る
- 覚悟する！
- データの作成が新規性のひとつとして論文に書ける



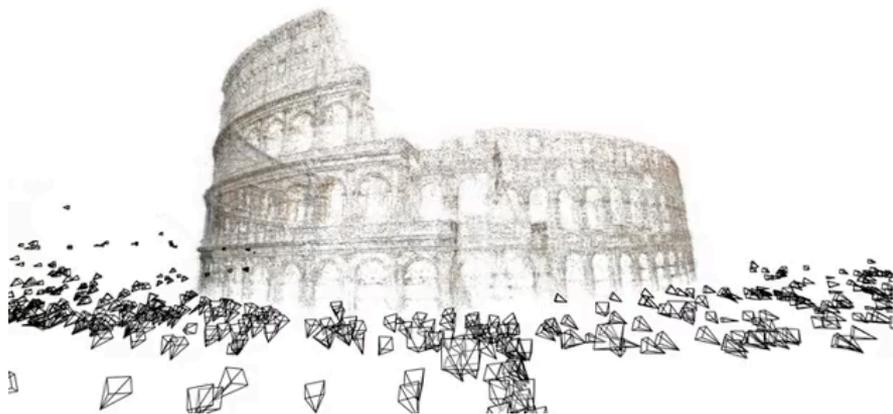
IM  GENET

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

今後の方針

新規性をうまくアプリケーションとして見せる

- 昔のBuilding Rome in a Day (ICCV2009)のイメージ
- 2016年だとFace2Faceなど
- 少しの差分でもその違いで見える未来を（わかりやすく）説明
- 分かりやすく，というところが重要



<https://www.youtube.com/watch?v=kxtQqYLRaSQ>

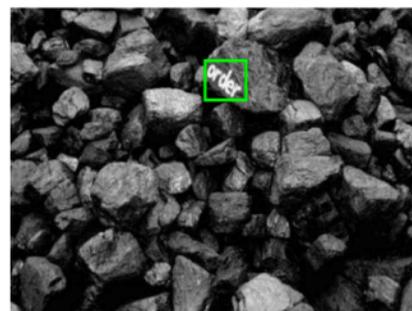


<https://www.youtube.com/watch?v=ohmajJTcpNk>

今後の方針

ロストアイディアを復活できないか？

- (一時的に) 消されたアイディアという意味
- もちろん, 今風にアレンジ
- 可能性があるにも関わらず, 消えたアイディアという意味
- e.g. ニューラルネット, 文字検出



研究って楽しい！

なぜ、研究者になりたかった？

新たな概念（コンセプト）を世に出したい

- 今あるものだけでなく、ないものを生み出す
 - 100年以上生き続けるコンセプトってカッコイイ！
 - e.g. 電話の概念は19世紀後半以降、現在も受け継がれている
- 計算機視覚が人の視覚を超えるところにそれ（拡張知能）がある？

人の視覚を超えるCV技術を実装し続ける

- （大きく言うと）人類のブレークスルーがありそう
- 昨今のAIブームがそれを象徴している？

修士課程・博士課程・ポスドクに進んだ経緯

修士

- 研究者になることは決めていたので修士は最低条件、M進学
- 修士のうちに10本、国内外の会議に投稿&発表

博士

- M2で内定を取ったが評価は研究者ではなかった、D進学
- 在外研究（UC, AIST, TUM）を2.5年間、COE/学振DC等

ポスドク

- 修行する時間が欲しかった、PD就任
- AISTでプロ研究者と議論、論文を年600本読んだ

「自身の成長」「論文への反映」「社会実装」の繰り返し？

cvpaper.challengeの役割

自分が辿ってきた道の凝縮版を後輩たちに伝えている

- 研究の「現在の動向」と「進め方」、「効果のある研究」
- 根底にある原動力は研究の「楽しさ」

産総研では10名、東京電機大では10名の学生やポスドクが活動

もし、本日の講演に感じるところがあるならば、
一緒に「優れた問い」を見つけて問題を解いていきませんか？

– 実は「講演」と同時に「勧誘」に来ました（笑）

cvpaper.challenge 2017 開幕

cvpaper.challenge 2017の目標

トップ会議※に年間10本以上投稿する

- 問題提起を狙って論文を書くが追従型研究もOK
 - 前年の成果含め5本は通す & 国内で賞を獲得
- コード・デモ・DB・動画を積極的に作成・公開
- (もちろん) 常に論文インプット
 - 精読やコード実装を多くする

※ Top-20 (subcategory) at

Google Scholar

3年目はアウトプット重視

2015, 2016は計1,600本のまとめ論文を作成

- サーベイテンプレートによるインプットの蓄積
- 高速に追読可能な体制を整備

2017はInput/Outputの劇的な効率化

- 読破数と同様「圧倒的な」アウトプット
- Outputの種類
 - 論文、サーベイ論文、arXiv、デモ、動画、データ、チュートリアル資料、サーベイ資料、まとめ資料、コンペティション、コード公開 等
 - 勉強会・ハッカソンなど面白そうなことは自由に提案してください!
 - 学習した技術を使って遊ぶ! というくらいの気持ちが良い
- 分割したグループ単位のリーダー(GL)の指示や自発的な行動が必要

各工程を経るごとに論文の質が自然に向上

- メンバーのモチベーションが極めて高い
 - 問題設定が面白いという意見が多い
 - 2年間続けてこられたのは学部・大学院生のおかげ

支えているのは個人作業とチームとしての議論

- 簡単には真似できない論文調査量とまとめ
- 徹底的な議論によりアイデアを研究テーマレベルに向上

仲良くなることが成果に直結

- 昼食会
 - 輪講・技術紹介・ブレインストーミング
- お茶会
 - お互いをよりよく知る

【Issue: 問題設定】と【Novelty: (手法としての)解の質】による研究レベル向上

- 問題設定：論文読み込み，議論，アイデアの選択と統合
- 解の質：利用可能なテクノロジーの最先端を駆使，およびグループ員の得意技術を活用

Target in the cvpaper.challenge 2017

目指すのは「クールな精鋭集団」

IssueとNoveltyを高いレベルで実現
して上位会議に突破・社会実装まで
到達しましょう