



会議レポート

CVPR 2015 参加報告

CVPR (IEEE Conference on Computer Vision and Pattern Recognition) は, ICCV, ECCV とともにコンピュータビジョン (CV) 分野のトップカンファレンスに位置づけられる国際会議である。今年 (2015 年) は 6 月 7 ~ 12 日に, マサチューセッツ州ボストンのハインズコンベンションセンターで行われた。本稿では, 2015 年の CVPR に関する概要と全体的な統計について述べた後に, 特に重要であると考えられる複数の論文と, 3 日目に行われた招待講演について説明する。紙面の都合上, CVPR の論文について言及する際, 著者名だけ本文中に示し参考文献には当該文献を記載しない。また, 2013 年以降で CVPR に採択されたすべての論文は原則無償でインターネット上に公開されており, 誰でも自由にダウンロードできる。詳細は CVPR のプログラムページ^{☆1}を参照されたい。

会議の概要と統計

ほかの主要な国際会議と同様に, 初日の CVPR は CV の研究に関するチュートリアルが開催された。2 日目から 4 日目の 3 日間は本会議があり, 今年の CVPR に採択された論文の口頭発表およびポスター発表が行われた。5 日目と 6 日目の 2 日間はワークショップが開催され, 各分野の研究者が熱心に議論を行っていた。

近年, CV 分野における国際会議への投稿件数は年々増加しており, 特にその傾向は CVPR のようなトップ会議において顕著である (表 -1)。統計的には, 今年は計 2,123 本の論文が投稿され, その中から 602 本の論文が採択 (採択率 28.4%), うち 71 本の論文が口頭発表論文として採択された。投稿論文数は昨年約 1.17 倍に相当する。このような CV の人気に呼応する形で, 2013 年には 32 社であったスポンサー企業の数も, 2014 年には 39 社, そして今年は 50 社 (特に今年はベンチャー企業のスポンサーが多い) と増加傾向にある。

	投稿論文数	うち採択数	採択率
2013	1,798	472	26.2 %
2014	1,807	540	29.9 %
2015	2,123	602	28.4 %

表 -1 近年の投稿件数と採択数の推移

ディープラーニングの隆盛

今年の CVPR で最も印象的であった事柄は, ディープラーニングに関する論文がこれまでの CVPR と比較して多数発表された点である。ディープラーニングは多層のニューラルネットワークを用いた機械学習の方法論の総称であり, 特に CV の分野では畳み込みニューラルネットワーク (CNN) と呼ばれるモデルを指す場合が多い。2012 年の一般物体認識コンテスト (ILSVRC) において Krizhevsky らが提案した CNN が既存手法を大幅に上回る認識性能を達成した。それ以降, 多くの研究者がディープラーニングに関心を示していたものの, 実際の論文数はこれまであまり多いとは言えないものであった。この増加は今年に入ってからのはじめてのことである。具体的には, 論文のタイトルに "Deep" を含む論文数が 2013 年には 3 本, 2014 年は 14 本と比較的少数であった一方, 今年は 68 本と以前より大幅に増加していた。以降の章で, 特に重要であると考えられる複数の研究分野とその成果について述べる。

物体検出とセグメンテーション

与えられた画像がどのカテゴリ (例: violin, air-plane) に属しているかを推定する画像分類問題に関しては, すでに CNN が人間と比肩するレベルの精度を達成したことが知られている。そのため, 今年の CVPR では画像分類問題を直接解く研究はあまり見かけず, 代わりに画像分類問題の発展形である, 物体検出問題と画像のセグメンテーション問題に関する研究が多く見られた。

与えられた画像から物体の矩形領域を検出する物体検出問題は産業分野の需要が高く, 以前から盛んに研究が行われている分野である。これまでの研究によって, CNN が画像分類だけでなく物体検出問題に関しても比較的有効であることはよく知られている。特に現在では, 画像から物体であろうと考えられる複数の矩形候補を抽出し, そのすべての候補について特定のカテゴリかそうでないかを CNN によって逐次判定する R-CNN と呼ばれる手法が有名である。G. Papandreou らはこの方法をさらに発展させ, CNN で矩形候補のカテゴリを判定した後, その矩形位置をサイズ最適化で微調整し, より正しい矩形領域に修正する方法論を新しく提案した。

セグメンテーション問題は画像の各画素がどのカテゴリに属しているかを推定する問題であり, 一般に画像分類, 物体検出問題よりも難しい問題とされる。セグメン

☆1 <http://www.pamitc.org/cvpr15/program.php>

ーションは一般に出力解の次元数が入力画像の次元数よりも多いため、プーリング層で中間層の情報量を逐次的に削減していくことで少数の出力解を求める CNN とは相性が悪い。この問題に関しては J. Long らが提案する、プーリング層で削減した中間レイヤをアップサンプリングによって元の解像度まで復元し、それをセグメンテーションに利用する手法によって部分的に解決された。なお、この論文は今年の CVPR の Honorable mention に選出され、すでにこれを発展させたいいくつかの研究論文が arXiv 上で公開されている。

画像から説明文を生成

画像認識への関心が高まりつつあると同時に、与えられた画像の説明文を生成する研究も注目を集めている。説明文生成に関する研究は 2010 年頃から行われていたものの、特に今年の研究成果は生成された説明文が自然な文体であったことから、一般のメディアも取り上げるほど大きなインパクトがあった。

O. Vinyals ら、A. Karpathy ら、J. Donahue らが同時期に提案した 3 種類の手法は、詳細は異なるものの基本的なアプローチはすべて共通している。はじめに、彼らは画像分類問題を解くために学習された学習済みの CNN を利用して、入力画像から画像の識別に有効な特徴ベクトルを抽出する。次に、その特徴ベクトルをリカレントニューラルネットワーク (RNN) と呼ばれる時系列データを扱うニューラルネットワークモデルに入力することで、自然な説明文を出力する。自然言語処理 (NLP) の分野において、RNN が既存手法と同等程度の精度を達成していることはすでに知られている。CV の分野で有効性が実証されている CNN と、NLP の分野で現在精力的に研究されている RNN を組み合わせることは自然な流れであることから、CV と NLP の融合領域においてこのような組合せを用いた研究は今後も増えていくものと期待される。

大規模データセットの公開

機械学習の研究において特に重要であるのが大規模かつ良質なデータセットの存在である。たとえば、近年の画像分類問題における CNN の成功は、自然画像とそのラベルを大量に収集した Imagenet プロジェクト^{☆2}の貢献が大きい。また、説明文生成の分野においては、近年になって提案された自然画像とそれを記述した複数の説明文を含む MS COCO データセット^{☆3}の貢献が大きい。今年の CVPR は特に、歩行者検出、映画の説明文、群衆など、大規模なデータセットに関する研究が多数発表された。本章では口頭発表で採択された 2 件の研究に焦点を当てて解説する。

RGB 色情報に深度 (Depth) を加えた RGB-D 画像に関する物体認識の研究は Kinect に代表される安価な RGB-D

センサの登場以降活発に行われている。しかしながら、CNN にも利用できる規模のデータセットはこれまで存在しなかった。S. Song らが公開した SUN RGB-D データセットは、RGB 色空間に加えて深度情報も含む約 1 万枚の室内の RGB-D 画像と、椅子やソファなどの物体の 3 次元位置がラベル情報として含まれる巨大なデータセットである^{☆4}。また、J. Shao らは WWW Crowd Dataset と呼ばれる、群衆が写っている 1 万個もの動画とその動画を記述するタグ (outdoor, walk, audience など) のペアが含まれるデータセットを新しく公開した^{☆5}。著者らは、本データセットを群衆の行動解析や人物の追跡等へ応用することを想定しているようである。

招待講演

3 日目の午後は著名な研究者による 2 件の招待講演が行われた。その 2 件ともがディープラーニングに関するものであり、今年の CVPR でいかにディープラーニングが注目されていたかがうかがえる。特に印象的であった講演は CNN の提案者として知られる Y. LeCun の、ディープラーニングの現状とその課題についての講演であった。

LeCun は、自身のこれまでの取り組みとその研究成果をふまえながら、現状のディープラーニングには大きく 4 つの課題が残されていると述べた。

1 つ目はニューラルネットワークの理論的解析の必要性、2 つ目は人体のポーズ推定や物体検出など、推定解がベクトルや木構造などの構造を持つ問題に関する研究、3 つ目は機械翻訳などの問題を解決するために必要な短期記憶モデルの開発、最後が無教師学習を上手く扱うモデルの開発である。今後のディープラーニングに関する研究は、これらの課題を軸に進展するものと期待される。

全体の印象

以上、CVPR 2015 の統計と複数の論文の概要について説明した。全体的に、今年の CVPR は以前流行していたランダムフォレストに焦点を当てた論文はあまり見られず、代わりに CNN が大きく取り上げられている印象を受けた。今後も論文の題名に "Deep" を含めるような傾向が続くかは予想できないものの、これまでの手法で解決できなかった問題を CNN で解こうとする試みはしばらく続くものと思われる。次回の CVPR は 2016 年 6 月 26 日から 7 月 1 日に、米国ネバダ州ラスベガスで開催される予定である。

(齋藤真樹/東北大学)

☆2 <http://image-net.org>

☆3 <http://mscoco.org>

☆4 <http://rgbd.cs.princeton.edu>

☆5 <http://www.ee.cuhk.edu.hk/~jshao/WWWCrowdDataset.html>