



会議レポート

CVPR 2019 参加報告

CVPR とは

International Conference on Computer Vision and Pattern Recognition (CVPR) は、IEEE Computer Society と Computer Vision Foundation が主催しているコンピュータビジョンとパターン認識に関する国際会議である。今年の開催場所はロサンゼルス近郊のロングビーチで、2019年6月16日から20日にかけて行われた。その熱狂的な状況をどのように表現したら読者諸氏にお伝えできるだろうか。たとえば、2019年の Google Scholar Metrics によれば、CVPR の H5-index^{☆1} は現在 240 であり、これは情報処理に関する国際会議の中で 1 位、科学雑誌を含む全分野のランキングでも 10 位に

☆1 H5-index とは過去 5 年でその会議や雑誌で出版された論文の h-index である。h-index とは、h 回以上引用された論文が、h 本以上ある、という指数である。

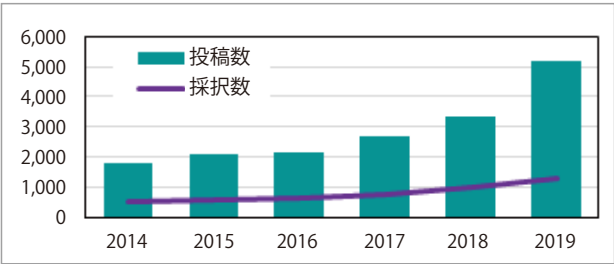


図-1 過去6年間のCVPRの投稿数と採択数 (CVPR 2019 Opening Slides¹⁾ より引用)

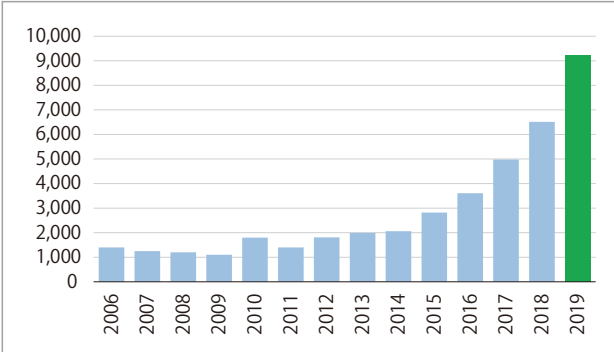


図-2 CVPR参加者数の推移 (CVPR 2019 Opening Slides¹⁾ より引用)

位置している。参加者は1万人弱、スポンサーは284社、団体展示は104あり、企業などからの寄付金総額は約3億3千万円に上る。論文の投稿数は5,160本に達し、採択されたのは1,294本、すなわち、4,000本弱は不採択という巨大かつ競争の激しい状況になっている。

図-1と図-2に投稿件数と参加者数の推移を示す。いずれも加速的に増加しており、今年の参加者数は9,227人に達した。図-3はオープニングの際の会場の様子である。このホールは同時進行する3つのセッションのうちの1つを行う会場であり、CVPRに投稿して優秀と認められた論文はポスター発表に加えて、この聴衆の前で自身の研究を口頭発表するという機会を得ることができる。なお、CVPRの論文投稿の締切は例年11月中旬頃であり、その後、査読コメントの通知と1回限りの反論の機会である rebuttal を経て、3月頃に最終の査読結果が通知される。図-4, 5に示すように、日本からは参加者数の割に論文投稿数で存在感が少なく、中国とは対照的な状況にある。



図-3 オープニング会場

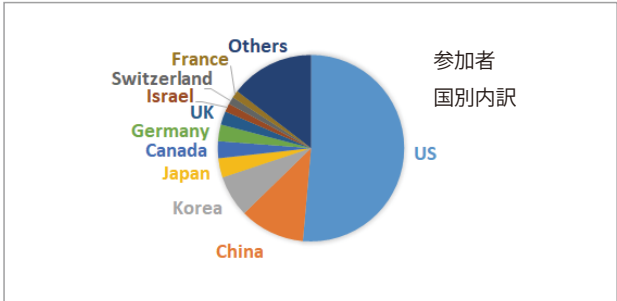


図-4 CVPR 2019の参加者の国別内訳 (CVPR 2019 Opening Slides¹⁾ より引用)

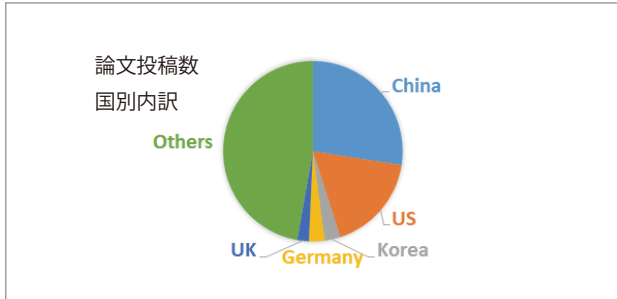


図-5 CVPR 2019の投稿論文の国別内訳 (CVPR 2019 Opening Slides¹⁾ より引用)

参加者数の増加に表れているこの狂騒的な状況は、自動運転やAIブームに後押しされている。ポスターブースの周りには企業展示が大手、中小その他を問わず、多数ある(図-6)。特に今年は、昨年度よりも多くの自動運転車が展示され(図-7)、投資・競争が激化している様子が垣間見られた。また、本会議の直前2日間には25個のチュートリアルと88個ものワークショップ^{☆2}が開催され、多くの未解決問題と広範な応用に関する議論が行われた。また、会議内でのリクルーティング活動も盛んであり、各国の企業がCVPRで発表を行うような優秀な学生・研究者を呼び込もうとして開催期間中の夕方などは会場周辺でさまざまな企業のプライベートイベントも行われた。

トピックつまみ食い

1,300本弱の研究が発表されたCVPRには実にさまざまなテーマの発表が行われていた。この報告の中で、それらすべてを網羅することは不可能ながら、本報告の著者らが興味を持ったいくつかのトピックについて紹介したい。

Monocular Depth Estimation

CVPRの企業展示では、Light Detection and Ranging(LiDAR)などのセンサを大量に搭載した自動運転車が展示されており、見るものの目を奪っていた。しかし、このような大量のセンサを搭載した自動運転車は聴衆を惹きつける一方で、価格等の面から一般に広く普及するとは言いがたい。

今年のCVPRでは、このようなLiDARを大量に用いた企業展示とは対照的に、カメラ画像だけを用いて車両周囲の環境を3次元的に把握するための研究が数多く発表されていた。

特に、単眼の車載カメラ映像のみを用いて周囲の距離マップ(Depth map)を推定する単眼距離推定(Monocular Depth Estimation)と呼ばれる問題設定に関する研究が数多く見られた。

単眼距離推定は、従来のステレオカメラを用いた距離

^{☆2} チュートリアルは、最近注目を集めつつある手法などについて、そのトピックの有名論文の著者などが解説を行う形式の集会。ワークショップは、本会議よりも具体的なテーマを設定することで、興味ある研究者が相互交流・情報交換をするための小規模な会議である。



図-6 ポスターブースと企業展示

推定とは異なり、両眼視差などの3次元的な情報を用いることなく距離を推定する必要があるため、非常に難しい問題である。多くの単眼距離推定に関する発表がある中、1枚の車載カメラ画像から距離だけではなく、画像中に存在する物体の3次元検出と車両の形状復元(shape recovery)を同時に実現した研究²⁾が発表されている。

まだまだ課題も多い問題だが、実用化されれば高価な機器を導入することなく、安価なセンサのみで自動運転や自律ロボット制御を実現することができるため、今後の技術発展が注目される。

Non-Line-Of-Sight イメージング

近年、画像認識等のタスクは学習ベースの手法により、ほぼ習熟したといえるほどの精度を達成している。しかしながら、多くの手法はあくまで「見える」ものを対象とした場合である。今年のCVPRのBest paperには、「見えない」ものをセンシングする技術が採択された³⁾。Non-Line-Of-Sight(NLOS)イメージングと呼ばれるこの夢の技術は、人類の可能性を大きく飛躍させる可能性を秘めている。たとえば、自動運転車に応用すれば、車から見えない位置から飛び出してくる人を、事前に認識することができるかもしれない。

NLOSイメージングの多くの手法では、シーンに対してパルスレーザを照射し、特殊なセンサを用いてピコスケールやナノスケールの時間解像度で各時刻にセンサに受光した光子の数を観測する。基本的なアイデアは、この時系列データから、レーザ→レーザを照射したセンサから見える点→そこから反射し到達したセンサから見えない点→さらに反射して到達したセンサから見える点→センサ、の経路を見つけることである。今年のBest paperは、Fermat pathと呼ばれる概念を提唱し、それが時系列データ上で不連続となる性質を理論的に証明した。そして、この性質を用いることで、センサから見えない場所にある物体の3次元形状の復元に成功した。

こうした特殊なセンサを用いた技術以外にも、通常のカメラを用いた研究もまだまだ少数ながら行われており、今後の発展が期待される。

少数データからの学習

深層学習を使ってみようとしたものの、学習データが



図-7 展示されていた自動運転車

十分に集まらず断念してしまった経験はないだろうか。深層学習がパフォーマンスを発揮するためには1カテゴリにつき1,000個ほどの学習データが必要と言われており、解きたいタスクに合わせたデータセットを作るコストの大きさは現在でも問題視されている。今年のCVPRでも、通常の学習方法ではデータ不足のため難しいような問題設定を扱う研究が多く発表された。

少数ショット学習 (Few-shot learning) は1～20程度のごく限られたサンプルから認識モデルを学習することを目指すタスクである。近年は畳み込みニューラルネットワーク (CNN) ベースの画像分類での少数ショット学習が盛んに試みられており、今年度のCVPRではグラフニューラルネットワークを用いてクラス間の構造を学習し、少数サンプルのクラスで不足する知識を補うといったアプローチが複数見られた^{4), 5)}。

ゼロショット学習 (Zero-shot learning) ではより極端に、識別したいクラスの学習画像が1枚も存在しないような状況を扱う。この場合はクラスの視覚的特徴を記述した補助情報 (たとえば鳥なら羽の色、くちばしの形、模様など) と分類器の関係を学習するアプローチが有力である。これに関して今回特に興味深かったのは、既存のベンチマークの問題点を指摘した研究である⁶⁾。既存のベンチマークでは学習用クラス (学習用データが存在する) とテスト用クラス (学習用データが存在しない) に似たものが多すぎ、ほとんどの既存手法よりシンプルな“自明な”手法によってより良い精度が得られてしまうことが指摘された。新しいタスクに取り組む上では、データセットと評価方法から慎重に検討する必要があるということだろう。

GAFAなどのビッグデータ企業がAI利用を進める一方、日本に強みのある製造業などでは利用できるデータに限りがありAI利用が難しいことがよくある。限られたデータからの学習にはまだまだ困難な点が多いが、期待は大きく引き続き目が離せない分野であろう。



図-8 CVPR 2019の投稿論文から作成したワードクラウド

これからこの分野の勉強をする人へ

図-8は今年のCVPRに採択された論文から得たワードクラウドとなる。上記では取り上げられなかったが、多様な環境で動作する機械学習モデルを得るためのドメイン適応^{☆3}に関するさまざまな発表や、教師なしでさまざまな画像処理の問題を解くUnsupervised Learning、あるいは画像と音声やテキストといった複数のモダリティを同じ特徴空間に埋め込むなどといったCross Media/Cross Modalの発表も多くあったように思われる。

一方で、ここ2,3年の間に非常に多くの研究発表がなされた画像を生成する敵対的生成ネットワーク (Generative Adversarial Network : GAN^{☆4}) は、やや存在感を減じている。ワードクラウド (図-8) の中に小さく入っている「GAN」という単語が見つけれられるだろうか? どうやら、これだけ1つの会議で発表件数が多いということは、主要な問題は世界中で一斉に、ヨーイドン!で研究がなされ、恐るべき速度で解かれていくということを意味するようだ。GANにより成熟した敵対的学習は、人間に解釈可能な特徴量を得るための技術として応用され、「敵対的 (Adversarial)」という単語はGANより大きくなっている。それ以外にも、自動運転1つとっても、解くべき問題はまだまだ数多く残されているし、新しい手法が呼び水となって応用の幅は広がり続けている。AIは幻滅期に入ったといわれることもあるが、少なくとも視覚情報処理の分野では、若い技術者の活躍の場は、まだまだこれからも広がっていくだろうということ、今年のCVPRに参加して改めて感じた。

参考文献

- 1) <http://cvpr2019.thecvf.com/>
- 2) Manhardt, F. et al. : ROI-10D: Monocular Lifting of 2D Detection to 6D Pose and Metric Shape, CVPR 2019.
- 3) Xin, S. et al. : A Theory of Fermat Paths for Non-Line-of-Sight Measurements, CVPR 2019.
- 4) Gidaris, S. and Komodakis, N. : Generating Classification Weights With GNN Denoising Autoencoders for Few-Shot Learning, CVPR 2019.
- 5) Kim, J. et al. : Edge-Labeling Graph Neural Network for Few-Shot Learning, CVPR 2019.
- 6) Hascoet, T. et al. : On zero-shot Recognition of Generic Objects, CVPR 2019.

(橋本敦史/オムロンサイニックス(株), 平川 翼/中部大学 中部高等学術研究所, 吉橋亮太/ヤフー(株), 藤村友貴/京都大学大学院 情報学研究所)

☆3 データの収集を行った環境が異なることに起因する機械学習モデルの精度低下を防ぐための技術の総称。

☆4 乱数を入力として、本物のデータと区別できなくなるような出力を得るよう学習する手法・およびそのネットワークのこと。最近、GANで生成された絵画が高額で落札されるなどして話題となった。