



会議レポート

ICCV 2019 参加報告

ICCV について

International Conference on Computer Vision (ICCV) 2019 は、2019 年 10 月 27 日から 11 月 2 日にかけて、韓国ソウル市内に位置する COEX Convention Center にて開催された。ICCV は IEEE Computer Society と Computer Vision Foundation が主催するコンピュータビジョンに関する主要な国際会議の 1 つであり隔年開催されている。ICCV の h5-index^{☆1} は 2019 年 11 月現在で 129、CORE^{☆2} のカンファレンスランキングによると A* となっており、コンピュータビジョンに関する国際会議では Computer Vision and Pattern Recognition (CVPR)、European Conference on Computer Vision (ECCV) に並びトップカンファレンスの 1 つである。

オープニングセッションで今回の ICCV についてさまざまな数値データが公開されたのでいくつか紹介する。まず発表数について、論文投稿数 4,303 本のうち採択数は 1,075 本であり、そのうち口頭発表数は 200 本であった。選定にあたり、2,500 名以上の査読者が携わっていることが明らかにされた。今年は 17 回目の開催となるが、参加者は 7,500 人を超え、前回の ICCV 2017 と比較しても 2.4 倍以上となり (図-1)、世界的にもコ

☆1 過去 5 年でその会議や雑誌で出版された論文の h-index (被引用数が h 回以上であるものが h 本以上ある)。

☆2 The Computing Research and Education Association of Australasia. カンファレンスランキングは、被引用率や採択率、主催者や会議にかかわる主要な人物の研究実績等の組合せによって評価される。

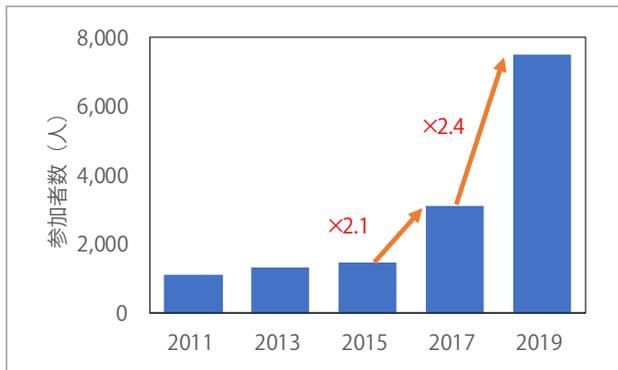


図-1 近年開催分の参加者数の推移 (ICCV 2019 正式公開データから筆者が作成)

ンピュータビジョンへの注目度が急速に上昇していることがうかがえる。開催地韓国からの参加者が最も多く、次いで中国やアメリカからの参加が多かった。日本は 4 番目に多かったが中国からの参加者の半数程度と割合としてはやや少なく感じた。また、企業等 56 団体がスポンサーとなっており、ポスター会場に併設の EXPO では NAVER 社など韓国企業を中心に、さまざまな企業等から 72 件の展示が行われた。

ICCV 2019 のプログラムを簡単に紹介する。開催期間のうち 3 日間は Tutorial や Workshop に割り当てられており、基調講演や招待講演もこの期間に行われた。また、本会議は開催期間のうち 4 日間を占めており、口頭発表は 2 会場の平行形式の全 13 セッションで構成された。発表の様子は、2 会場それぞれの入り口の大きなモニターと、別の部屋のスクリーンに同時中継され、各自が好きなように聴講することができた (図-2)。ポスターセッションは期間中 7 回開催されたが 1 回あたり約 130 件もの発表が行われ、どのポスターにも多くの参加者が集まり活発な議論が行われていた。ポスターセッションと同時にデモセッションも設けられており、インタラクティブに動作する技術展示が行われた。また、ポスター会場に併設された企業ブースでは採用のためのスタッフやテーブルも用意されていることが多く、デモや展示の説明の傍ら、学生と企業従業員が対話している様子もよく見られた。

トピック紹介

図-3 は、論文タイトルに使用されたキーワードから得たワードクラウドである。近年の傾向として、少ない教師データ、あるいは教師データなしで問題を解く Weakly supervised や Unsupervised Learning などのキーワードが比較的多くなったように見受けられる。3 次元データを扱うものも多く、さらに視野を共有しないカメラ間での人物や物体の照合問題に関する



図-2 会場の様子
上図：口頭発表会場内、下図：口頭発表会場入り口付近、右図：ポスターセッション会場の様子

Re-Identification といったキーワードなども見られた。GAN (Generative Adversarial Network) に関してもまだ熱が冷めておらず、ポスターセッションで特に活発な議論が行われていた。そのほか、画像や映像に加え、音声やテキスト、センサデータなどの別のチャネルを用いたものもいくつか見られた。デモセッションでは、CVPR や ECCV など他会議で発表された論文を発展させたデモ等も行われた。たとえば、Re-Identification と表情認識技術を活用し、空港を訪れる人々のジャーニー(動きやその時々感情など)を分析可能にすることで顧客満足度を高める取り組み¹⁾ 等も紹介された。

ICCV 2019 では、Best Paper Award が 1 本²⁾、Best Student Paper Award が 1 本³⁾ 選定された。Best Paper を受賞した論文について簡単に紹介する。

“SinGAN: Learning a Generative Model from a Single Natural Image”²⁾ が Best Paper Award として選定された。近年のディープラーニングを用いたコンピュータビジョン分野における研究の中で、GAN は非常に注目度が高く、これまでもさまざまな論文が発表されてきた。従来の GAN は、高品質な画像生成を行うためにはモデルの学習に多量の画像を用いる必要があったが、SinGAN は学習用画像を 1 枚のみ用いることを実現している。SinGAN のポイントは階層ピラミッド構造を持つネットワークと、さまざまな解像度に対応し得る生成器ネットワークであり、学習画像からさまざまな情報を引き出すことができるため、わずか 1 枚の学習画像で画像から画像への多様な変換を可能にした。GAN を用いたさまざまなタスクに応用可能とし、発表では低解像度画像から高解像度画像への変換や、画像調和 (図-4)、画像の修正、

入力画像に類似の画像が生成できることを活かしたアニメーションなどが紹介された。選定された論文は口頭発表に加えオープニングセッションの後にも内容を説明する機会が与えられており、SinGAN の応用例紹介のシーンでは、会場から驚きの声も聞かれるなど現地でのインパクトも大きかった。この論文だけでなく、画像 1 枚のみを用いて学習する GAN の発表がいくつかあり (InGAN⁴⁾ など)、GAN 研究の流れの変化を感じた。

先述の通り、ほかにも幅広い観点で数多くの発表があったため、それぞれの論文を参照されたい。

ICCV 2021

次回の ICCV は、2 年後にカナダのモントリオールで開催されることが決定している (開催日程は本報告執筆時点で未定)。コンピュータビジョン研究のトレンドや世界の研究者の熱を直接感じることができるため、コンピュータビジョンに興味を持たれている方や研究を始めたばかりの皆様にもぜひ参加していただきたい。

参考文献

- 1) Sugianto, N. and Tjondronegoro, D. : Deep Residual Learning for Analyzing Customer Satisfaction using Video Surveillance, 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) (2018).
- 2) Shaham, T. R. et al. : SinGAN : Learning a Generative Model from a Single Natural Image, The IEEE International Conference on Computer Vision (ICCV), pp.4570-4580 (2019).
- 3) Duff, T. et al. : PLMP - Point-Line Minimal Problems in Complete Multi-View Visibility, The IEEE International Conference on Computer Vision (ICCV), pp.1675-1684 (2019).
- 4) Shocher, A. et al. : InGAN : Capturing and Retargeting the “DNA” of a Natural Image, The IEEE International Conference on Computer Vision (ICCV), pp.4492-4501 (2019).

(吉岡隆宏／富士通研究所)

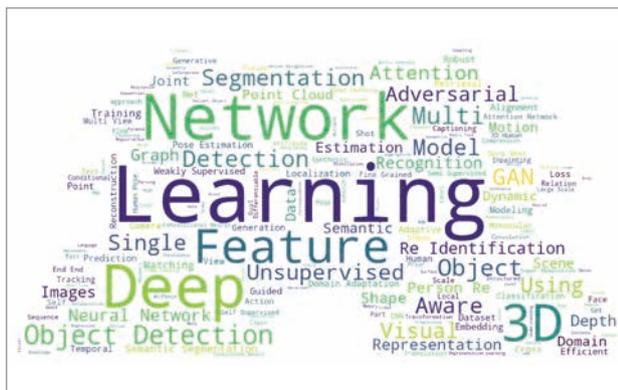


図-3 ICCV 2019 ワードクラウド (発表論文タイトル一覧から筆者が作成)

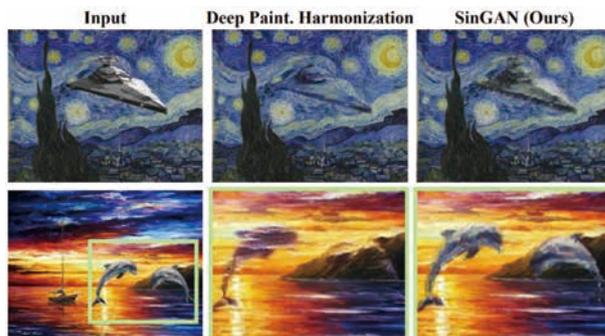


図-4 SinGAN による画像調和の結果例 (文献 2) 中から引用 (図-13)。図右が SinGAN の出力。絵画中の詳細な描写を残しつつ、追加した部分にテクスチャの特徴などの大域的な情報を反映して調和しているのが分かる