画像応用技術専門委員会

人工知能の実用化とこれからの ViEW*

The Practical Applications of Artificial Intelligence and the Future of ViEW

加藤邦人** 西山正志*** 川西亮輔† 片岡裕雄††
Kunihito KATO, Masashi NISHIYAMA, Ryosuke KAWANISHI and Hirokatsu KATAOKA

Key words Vision Engineering Workshop（ViEW), artificial intelligence（AI), AI ethics, robot manipulation

1. 緒　　言

精密工学会画像応用技術研究会（Technical Committee on Industrial Application of Image Processing : IAIP）は，毎年「ビジョン技術の実利用ワークショップ ViEW」を開催している．例年，横浜市のパシフィコ横浜にて開催し，600 名を超える参加者を集めている．2020 年は，COVID-19 感染症対策のためオンライン開催となったが，過去最高の参加者を集め盛況に終わった．

ViEW の歴史は長く，1988 年に「ハンダ目視検査の自動化技術」としてスタートし，1989 年に広く外観検査の自動化を目指した「外観検査の自動化ワークショップ」に改名．そして 2003 年に「ビジョン技術の実利用ワークショップ Vision Engineering Workshop（ViEW2003)」となり，外観検査にとどまらず，画像認識技術の実利用に関するワークショップとして大きく発展を遂げた．

画像認識技術は産業界からの期待が大きく，外観検査，FA，マシンビジョンなど製造現場だけにとどまらず，ADAS（先進運転支援システム)，セキュリティ，防災，インフラ管理，農林水産，映像技術，エンターテインメントなどに実利用の幅を広げている．

この 10 年の参加者の推移を**図 1** に示す．2020 年はオンライン開催にあわせ，学生参加費を無料としたため学生比率が非常に高くなっているが，参加者は 10 年で 400 名から 600 名まで順調に増え，画像認識技術への期待が非常に高まっていることがわかる．また，企業参加者が 50% を大きく超えることも ViEW の特徴となっている．**図 2** に業種別参加者割合の推移を見ると，製造業からの参加者が非常に多く，実利用を強く意識したワークショップであることがここからもわかる．

さて，過去のデータ分析より，産業界から強く期待されている ViEW であるが，産業界は ViEW の何に期待しているのだろうか．一番は，2010 年代後半のディープラーニングをはじめとする AI 技術の発展が大いに貢献していると思われる．ディープラーニングによる画像認識技術は，この数年で驚くべき発展を遂げ，かつては不可能と思われていた高度な認識を可能とした．そして，ディープラーニングは，応用のステージに入り始めていると多くの技術者が考え，ViEW に大きな期待を寄せていると思われる．

ディープラーニングが応用のステージに入ると，さまざ

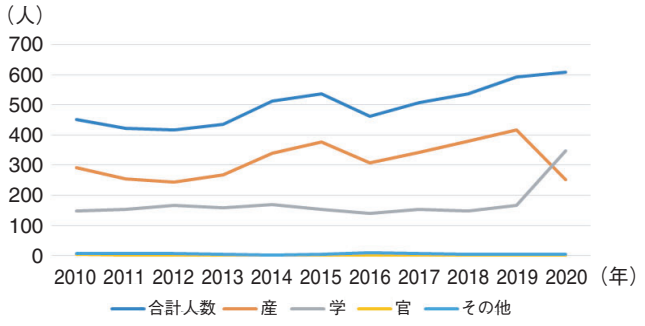


Fig. 1　Transitive graph of number of participants

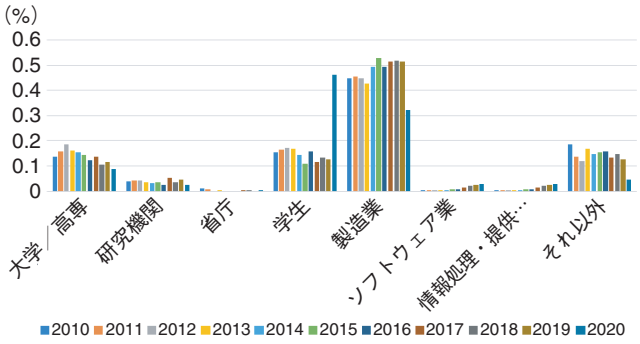


Fig. 2　Transitive graph of the percentage of participants' affiliations

*原稿受付　令和 3 年 7 月 21 日
**正 会 員　岐阜大学工学部電気電子・情報工学科情報コース（岐阜県岐阜市柳戸 1-1)
***鳥取大学工学部（鳥取県鳥取市湖山町南 4-101)
†正 会 員　東京ロボティクス(株)（東京都新宿区山吹町 347 藤和江戸川橋ビル 5F)
††産業技術総合研究所（茨城県つくば市梅園 1-1-1 中央第一)

まな問題が明らかになってきた．そこで，2章および3章ではそれぞれ「実利用を阻む人工知能の現状」「産業現場における人工知能活用」と題して AI 開発を取り巻く状況について紹介する．

2. 実利用を阻む人工知能の現状

特に 2010 年代の人工知能分野の隆盛は目覚ましく，過去数十年間にわたり画像認識のみではできないといわれてきた問題が続々と解決されている．今後しばらくはこの動向は続き，次々に新しい分野の開拓や認識率向上という文脈で性能が向上し続けるものと思われる．

しかし，光があれば影があることもまた事実である．本章では特に深層学習と画像データセットを用いた場合の問題を紹介する．AI 開発や実利用化の際に大きな障壁となり得るため，分野全体で解消に向けた基盤技術の構築と開発を要すると考える．ここでは，特に画像データセット側の問題である「公平性」「透明性」「プライバシー」の項目に分けて紹介する．

2.1 画像データセット関連の問題

2.1.1 公平性

画像認識における公平性ではデータセットの偏りやラベルの誤りが学習したモデルにそのまま現れてしまうことが問題視されている．例として画像に付与されたカテゴリに含まれる画像の割合により性別・人種・年齢などの面で不公平な出力が生じてしまうことがわかっている．プログラマでは男性が，看護師では女性の出力をしやすくしてしまうことが明らかになっているだけでなく，画像中に映されている環境や物体にも起因して同様の問題が発生することが報告されている．コンピュータビジョン分野で公平性問題が広く知れ渡ったのは間違いなくデファクトスタンダードである ImageNet[1]にデータの偏りや攻撃的なラベルが含まれることが明らかになってからである[2]．各社の公開していた顔認識システムや公開データセットも相次いで削除・公開停止されるなど，公平性問題は根深い問題となりそうである．ここで，性別や人種などデータセット中の偏りを排除するのみでは，カテゴリの認識のしやすさなどから完全な公平性を担保することは難しいようであり，今後も引き続き議論が続きそうである．

2.1.2 透明性

画像認識に用いられるメジャーなデータセットは基本的に人手によりラベル付けされたものである．画像識別タスクの場合には画像収集～画像ラベル定義～画像ラベル付け～画像ラベルのクロスチェックを経ることがスタンダードな流れである．しかし，データセットが公開されるまではデータセットの詳細を知ることはできず，仮に公開データセットになったとしても各画像のラベルについて，アノテータの詳細情報や意図などを知ることは困難である．このように，究極的にはデータセットが公開されていても完全な透明性というのは担保されていない状態であり，人間がラベルを付与してクロスチェックを経ても数％はラベル誤りが含まれてしまうという問題も報告されている[3]．画像に対するラベル誤りを学習したモデルが重大な問題を引き起こす可能性もないわけではなく，透明性の向上も課題として扱われようとしている．

2.1.3 プライバシー

前の項目「透明性」でも画像データセット構築の代表的な流れが示されていたが，その中でも画像収集の時点で集められたデータに個人情報が混ざってしまう可能性もある．自ら積極的に撮影して SNS などに公開したわけでなくとも，背景の人物としてたまたま映り込んでしまうなどあり，データの取り扱いはこれまで以上に注意を要する．このような側面から，現在の主要なデータセット（含：ImageNet）は学術および教育目的に限定されており，商用利用は想定されていない状態である．クラウドソーシングにより，不特定多数のクラウドワーカーが Web から画像を収集する方式ではどうしても画像のプライバシーについて不明瞭になりがちである．クラウド上で画像や動画像を研究者に引き渡す際に権利についても委譲する取り組み[4]も取り入れられており，改善に向けた取り組みは進んでいる．

2.2 データセット問題解決に向けて

データセットの公平性・透明性・プライバシーの問題は，AI 開発や実利用化のための大きな障壁であり，場合によっては昨今の AI ブームを終わらせかねない人類の共通課題である．特に問題視されている文脈が，学習済みパラメータを用いた転移学習である．産業技術総合研究所（以降，産総研）でもこの問題の根本解決に取り組んでいる．従来の画像データは多様な物体をあらゆるシーンにて撮影した実画像に対して，人間がペアとなる教師ラベルを付与することでデータセットを構成，モデルが学習することで基礎的な視覚能力を獲得できることが既成事実となっていた．産総研では実画像の利用および人間の教師こそが実は公平性や透明性，プライバシーを害する原因であると仮説を立て，実画像および人間の教師を一切用いず，自然の法則に基づいた数式から画像パターンを生成し，基礎的な視覚表現を獲得させる「数式ドリブン教師あり学習」を確立した[5]．事前学習の文脈においては実画像や人間の教師を用いないため，人間関連のラベルは含まれず公平性やプライバシーの問題については問題なく，数式から画像パターンとラベルが生成されているため透明性があることは自明である．さらに，たび重なる実験の結果，画像識別の対象次第ではデファクトスタンダードな手法に匹敵し，実利用化に向けた精度を議論できるレベルにまで到達している．ここで，**図 3** に数式ドリブン教師あり学習において現在最も高い性能を叩き出している事前学習データセットである FractalDB を示す．図中のような画像群を用いて事前学習を行い，実画像を認識するために十分な特徴表現を獲得できているという点で驚きの結果を得ている．数式ドリブン教師あり学習のレベルを高めていくことが，従来の事前学習を置き換え，画像データセット関連の問題を解



Fig. 3　Fractal DataBase（FractalDB）

決するための一手法となりうると期待している．

3. 産業現場における人工知能活用

製造業，医療，マーケティングなどさまざまな産業現場で人工知能の活用が進められており，実用化されているケースも少なくない．本章では筆者の関心領域であるロボット，特にロボットアームによって物体把持をするピッキングロボットに関して，研究事例や筆者の取り組みを紹介する．これを通じてピッキングロボットへの AI の適用事例や課題，学術界に対する今後の期待について述べたい．

3.1　ピッキングロボットの従来手法

ロボットアームによって乱雑に積み上げられた物体を把持して取り出すタスク，いわゆるバラ積みピッキングは，昨今の AI ブーム以前から難易度の高いチャレンジングな課題として研究開発が進められていた（文献 6）など）．バラ積みピッキングを実現させるアプローチはさまざまであるが，大きく切り分けるとすれば対象物の位置姿勢を認識する方式か，対象物の姿勢を考慮せず把持位置を検出する方式の 2 つである．

従来手法として前者では例えば Point Pair Feature[7] などのハンドクラフトの特徴量が用いられ，これまでさまざまな改良手法が提案されている（文献 8）など）．特徴量は開発者が設計するものの，あらかじめ対象物が有する特徴量を抽出して学習するというフェーズがあることから，前者の方式は広義の AI といえるだろう．一方，後者の手法はバラ積みシーンに対してどの位置姿勢でハンドを差し込めば把持が可能かどうかを判定する（文献 9）など）．ヒューリスティックな手法ではあるが，対象物の位置姿勢を推定するという難しい過程を経ることなく把持が可能な点で扱いやすいという特徴がある．どちらの手法が適しているかは対象物やアプリケーションによって異なるため，ケースバイケースで使い分けられてきた．

3.2　ピッキングロボットの AI 研究事例

上記のような従来手法はピッキングロボット向けの 3D センサ製品の認識機能としてすでに実用化されており，一定の成果を上げている．しかしながら，バラ積みピッキングはシーンが複雑でかつ対象物も多岐にわたる．そのため，各々の手法の適用範囲が狭いことや，十分な認識性能を得るために試行錯誤的にパラメータ調整をしなければならないという問題が残る．



Fig. 4　Cluttered scene and detection results of fried chickens

そこで，さまざまな対象物に汎用的に適用可能で，かつ調整不要なピッキングを実現するため，深層学習を用いた手法が盛んに研究されている．初期の研究例としては前述の把持位置を検出するアプローチに類するものが比較的多い（文献 10）など）．従来手法との違いは，多数の物体に対してあらかじめ好ましい把持状態を教師データとして与えて学習させておくことで，対象物ごとにパラメータ調整などをすることなく把持位置の検出ができることである．また，深層学習においては単一タスクを学習させるよりも関連するタスクを同時に学習したほうが認識性能は向上するという報告もあり，把持位置の検出だけでなく，物体検出やセグメンテーションなどのピッキングに必要な認識タスクを同時に解くアプローチも研究されている（文献 11）など）．

3.3　東京ロボティクス株式会社の取り組み

従来手法の適用が難しい対象物やタスクはさまざまあるが，その 1 つとして形状やアピアランスに個体差のある物体のピッキングが挙げられる．例として唐揚げのバラ積みシーン例を**図 4** に示す．

このような対象物には 3D モデルがないことが多く，そのアピアランスも不定で物体どうしの境界もあいまいであることが多いため，ルールベースの検出手法は適用が難しい．そこで筆者は深層学習ベースのセグメンテーション手法を用いて解決している．

セグメンテーションのアーキテクチャとしては Mask R-CNN[12] をベースとし，商用可能な公開画像データベースである Open Images Dataset[13] からラベル情報を抽出し，バラ積みシーンを自動生成することで独自のデータセットを作成して学習する．生成したバラ積みシーンの一例を**図 5** に示す．

複数クラスの自然物のラベルデータと，多様なバラ積みシーンを用いて学習することで，1 つの学習モデルでアピアランスの異なる未学習の対象物に対しても幅広くピッキング可能な精度のセグメンテーションが実現できることを確認している（文献 14））．

3.4　残る課題と今後の期待

3.3 節で紹介したように，自然物に対しては汎用性高く検出可能な学習モデルが得られたが，人工物に対する汎用的な検出手法は未解決課題である．その理由としては，利用可能なデータセットが少ないことが 1 つの大きな要因で



Fig. 5　Generated cluttered scene（class : tomato）

ある．日本では情報の機密性が高いため，実案件に至るまで対象物に関する具体的な情報が得られないことも多い．一方で現場適用が可能な見込みがなくては実案件に結び付きづらく，実機で検証するコストを回収できるかが不透明な状態で開発をせざるを得ないというジレンマがある．

この問題を解決する方法の１つとして期待するのはシミュレーション技術の発展である．すでに DexNet[15] をはじめとしてシミュレーション環境で物体検出やロボット動作を学習する取り組みがなされているが，シミュレーションとしては研究用途の性格が強いものが多く，産業界で広く活用されているとはいいがたい状況であった．近年では Isaac sim[16] のようなツールが公開され，実機検証の時間的・費用的コストを最小限にして開発を進められる環境が整ってきているといえる．

また，そのほかの課題としてはソースコードやデータセットのライセンスが挙げられる．ソースコードの公開が一般的になり，最新研究を手元ですぐに試すことができるようになってきているが，公的機関の研究であっても非商用のライセンスが付与されているものが少なくない．産業応用に結び付きやすい公開方法を，学術界全体として検討していただければ幸いである．

参　考　文　献

1）J. Deng et al. : ImageNet : A large-scale hierarchical image database, IEEE Conference on Computer Vision and Pattern Recognition（2009）.
2）K. Yang et al. : Towards fairer datasets : filtering and balancing the distribution of the people subtree in the ImageNet hierarchy, Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency（FAT）,（2020）.
3）C.G. Northcutt et al. : Pervasive Label Errors in Test Sets Destabilize Machine Learning Benchmarks, ICLR 2021 RobustML and Weakly Supervised Learning Workshops（2021）.
4）G.A. Sigurdsson et al. : Hollywood in Homes : Crowdsourcing Data Collection for Activity Understanding, European Conferene on Computer Vision（ECCV）,（2016）.
5）H. Kataoka et al. : Pre-training without Natural Images, Asian Conference on Computer Vision（ACCV）,（2020）.
6）M-Y. Liu et al. : Fast Object Localization and Pose Estimation in Heavy Clutter for Robotic Bin Picking, The International Journal of Robotics Research, **31**, 8（2012）951.
7）B. Drost et al. : Model Globally, Match Locally : Efficient and Robust 3D Object Recognition, 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition,（2010）998.
8）S. Akizuki et al. : Stable Position and Pose Estimation of Industrial Parts Using Evaluation of Observability of 3D Vector Pairs, Journal of Robotics and Mechatronics, **27**, 2（2015）174.
9）D. Yukiyasu et al. : Fast Graspability Evaluation on Single Depth Maps for Bin Picking with General Grippers, 2014 IEEE International Conference on Robotics and Automation,（2014）1997.
10）I. Lenz et al. : Deep Learning for Detecting Robotic Grasps, The International Journal of Robotics Research, **34**, 4-5（2015）705.
11）R. Araki et al. : MT-DSSD : Deconvolutional Single Shot Detector Using Multi Task Learning for Object Detection, Segmentation, and Grasping Detection, 2020 IEEE International Conference on Robotics and Automation,（2020）10487.
12）K. He et al. : Mask R-CNN, Proceedings of the IEEE International Conference on Computer Vision,（2017）2961.
13）A. Kuznetsova et al. : The Open Images Dataset v4 : Unified Image Classification, Object Detection, and Visual Relationship Detection at Scale, arXiv preprint（2018）arXiv : 1811.00982.
14）川西，ほか：不定形物体ピッキングのためのセグメンテーションの学習方法における検討，ROBOMECH2021,（2021）.
15）J. Mahler et al. : Learning Deep Policies for Robot Bin Picking by Simulating Robust Grasping Sequences, Conference on robot learning, PMLR,（2017）.
16）Isaac Sim : Omniverse Robotics App, https://developer.nvidia.com/isaac-sim.



加藤邦人
1996 年中京大学大学院情報科学研究科修士課程修了．同年，同大学院博士課程入学．現在，岐阜大学工学部教授．岐阜大学高等研究院人工知能研究推進センターセンター長．2011 年メリーランド大学 Faculty staff．コンピュータビジョン，特に深層学習の基礎，ならびに産業応用に関する研究に従事．電子情報通信学会，精密工学会会員．博士（情報認知科学）．



西山正志
2002 年岡山大学大学院自然科学研究科博士前期課程了．同年（株）東芝入社．同社研究開発センターを経て，現在鳥取大学大学院工学研究科准教授．2011 年東京大学大学院学際情報学府にて博士（学際情報学）を取得．カメラを用いた人物認識をはじめとするパターン認識およびインタラクションの研究に従事．



川西亮輔
2012 年，静岡大学創造科学技術大学院博士後期課程修了（博士（工学））．同年，三菱電機（株）に入社．先端技術総合研究所にてロボットビジョン，3D センシング，物体認識などの研究開発に従事．2020 年より東京ロボティクス（株）に入社し，3 次元カメラを用いたロボット知能化に関する研究開発の企画と実務を担当．IROS2009 にて IEEE Robotics and Automation Society Japan Chapter Young Award, R & D100Awards，FA 財団論文賞などを受賞．



片岡裕雄
2014 年慶應義塾大学大学院理工学研究科後期博士課程修了，博士（工学）．2020 年 10 月より（国研）産業技術総合研究所主任研究員．画像認識，動画解析，人物行動解析に従事．2011/2020 年 ViEW 小田原賞，2019 年度産総研論文賞，2020 年 ACCV 2020 Best Paper Honorable Mention Award 受賞．