



Go beyond Artificial Intelligence, towards Augmented Intelligence

澤田友哉[†]

まえがき

こんにちは！三菱電機でAIの研究開発をしている澤田友哉です。このコラムでは、AIの研究開発における難しさや面白さを、企業目線で“中のヒト”の声として紹介したいと思います。

私自身は、大学ではコンピュータサイエンスを専攻しており、山梨大学で茅暁陽教授、豊浦正広准教授に師事し、博士号を取得してメーカーに入社しました。大学に入るまでほとんどパソコンを使っておらず、書道をしていたアナログな人間でしたが、どうせーから勉強するなら今後必須になるコンピュータサイエンスを学ぼうと思って専攻を決めました。大学では留学生の先輩や同級生の勉学に対する姿勢に刺激を受けて研究が好きになり、指導していただいた先生方のサポートのおかげで研究活動を楽しみながら進めることができました。国際会議発表や他大学との勉強会、国際交流会など研究室で活発にイベントがあったことがモチベーション維持につながったと思います。

就職をするときにメーカーを選定したのは、研究室で作られた素晴らしい技術も誰にも知られず使われないのは残念だと考えたためです。当時から、三

菱電機米国研究所 (Mitsubishi Electric Research Laboratory) の研究はコンピュータグラフィックスやコンピュータビジョンの分野ではとても有名で、MERLと共同研究や研究成果の製品化、サービス展開ができるとういと考えて三菱電機に入社しました。

研究所での開発

2015年に入社して、当時から話題だったディープラーニングの研究のキャッチアップに努めました。黎明期ということもあって、3ヵ月に一度の頻度で技術革新が起こりすぐに次の技術が主流になるというサイクルの早さに驚きました。先輩たちと読んだ論文の議論をしたり、実装して動作検証したりと、ディープラーニングの新しい考え方が新鮮でとても高揚したのを覚えています。ディープラーニングの利用はConvolutional Neural Networks (CNNs)をはじめとするコンピュータビジョン分野での研究例が多く、とりわけObject Detection (一般物体認識) に代表されるような産業界の事業に直接つながる応用研究が盛んです。

当時、研究所では物体認識技術の演算量削減や、車載基板やマイコン上でのリアルタイム動作に取り組んでおり、**図1**のような「どこに、何があるのか」を画像1枚で認識するAI技術の開発を行っていました。特に自動運転技術に代表される車載システムでの利用を念頭に置いて開発を進めていたのですが、全天候対応で遠くの物体まで認識できる技術は最新研究を用いても

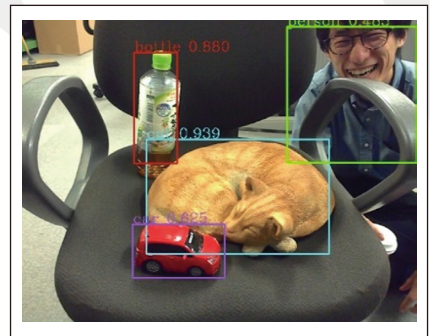


図1 基板上的物体認識の推論結果

なかなか実現できず、大きな挑戦となりました。

そこで、大学で研究していた視覚情報処理の知見を取り入れ、目立つ領域に優先的に注目する人間の視覚システムを模倣したSaliency Map (顕著性) を用いて物体らしさや誘目性の推定を行い、CNNsの画像認識で推論する技術を開発しました。学習済みCNNsの高次元特徴やタスクの概念情報を元に認識するアプローチをTop-downと呼ぶのに対し、人間の認識システムは視覚情報の色・輝度・方向性などの低次元特徴を手掛かりに統合させて推論していることが知られており、これを模倣した手法をBottom-upアプローチと呼びます。Bottom-upの利点として高速な演算が可能で、車載用組み込み機器上で100m以上遠方の小さな物体をリアルタイムで認識することに成功しました。開発した技術は産業広告や広報発表を行い国内外30以上のメディアから取材を受けました。また、研究開発本部成果披露会の展示や株主総会展

[†]三菱電機株式会社 ビジネスイノベーション本部
ビジネスイノベーション・DX戦略室 Maisart
共創センター

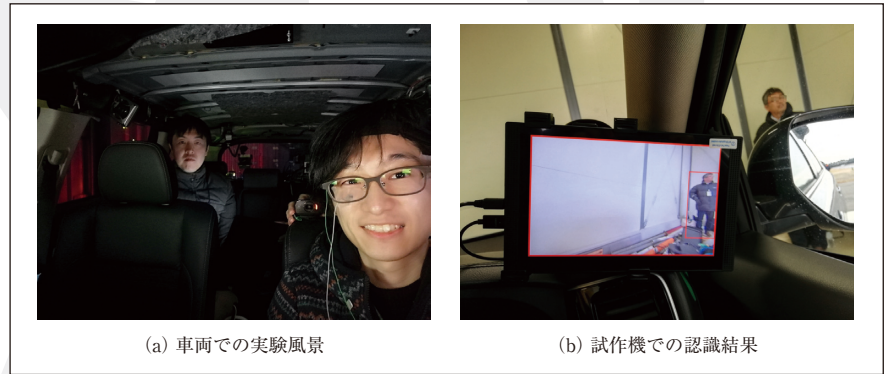
"Go beyond Artificial Intelligence, towards Augmented Intelligence" by Tomoya Sawada (Maisart Co-Innovation Center, Business Innovation & DX Strategy Div., MITSUBISHI ELECTRIC Corp., Tokyo)



図2 展示会での技術展示

示, CEATEC 2018の展示などを行いました(図2)。

一方で, Bottom-upアプローチは全天候に対応する必要がある車載系システムではまだまだ改善の余地がありました。そこでCNNsの高次元特徴に対し, Bottom-up特徴をAttention mechanisms(注意機構)として組み込む手法を開発しました。CNNsは層を深くするときchannelを増やして特徴量を高次元にするため, 空間解像度をpoolingによって削減し, 演算量を削減しています。しかし, 遠方の物体は画像では小さく映るので, 空間解像度を削減してしまうとCNNsの前段の浅い層で情報が喪失してしまい, 多量のデータを学習させても認識率が上がりにくいという問題がありました。人の視覚系を模したSaliency Map(顕著性)であれば遠方物体の情報をより強



(a) 車両での実験風景

(b) 試作機での認識結果

図4 実証実験による価値検証

く残しているため, CNNs内部で類似度が高いFeature Mapほど高い正の報酬を与えることで, 小さな物体の特徴を獲得しやすくして認識率の大幅な向上を達成しました¹⁾(図3)。

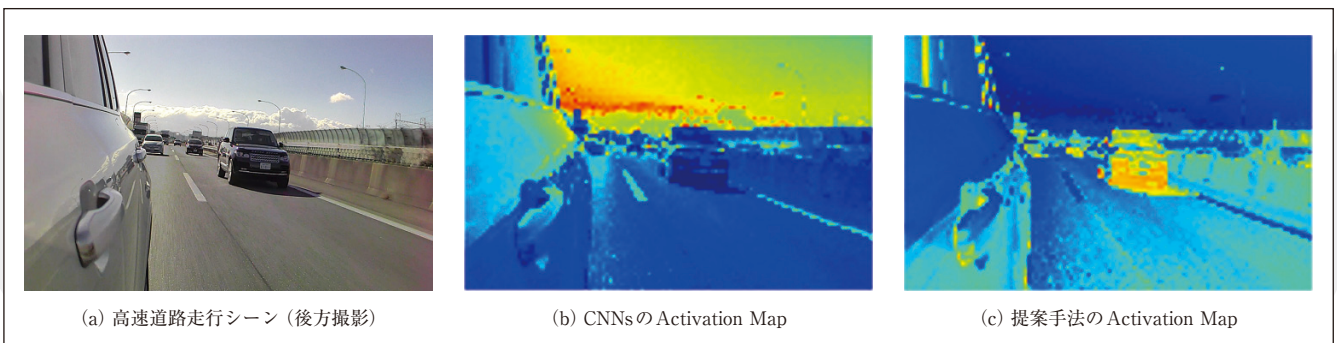
同時に, 開発した技術がユーザー目線ではどのような体験を生み出すのか, 価値検証も行っています(図4)。体にセンサをたくさんつけて実際の車両を改造し, 実際に運転してみてAIによる認識結果の通知がどういうメリットやデメリットがあるのかを明らかにします²⁾。

開発した技術は, CES 2020で展示し, 国際会議(ICIP 2021, ICME 2022)で発表しました。展示会は多くの人に関わって各企業のブースを作り上げ, 多くのお客様に技術やサービスを紹介して, みんなで成功を祝います。普段, 研究所の中において直接お客様とお話をする機会がなかった私にとっては, 質

問を受けたりビジネスの商談につながったりと刺激になりました。また, 国際会議は最近ではオンラインで開催されることが多く以前より安く参加できるメリットがあります。個別の発表だけでなく, 技術分野を俯瞰して紹介する講義も併設されており, 学生がキャッチアップしやすいようによく配慮された内容になっていますので, ぜひ参加してみてください。

海外大学や研究機関との共同研究

2016年から, 南カリフォルニア大学(USC)と共同研究を行い, Antonio Ortega教授に師事し, Graph Signal Processingの研究に従事しました。人間の行動認識や複雑な手の動きの認識などでは, Recurrent Neural Networks(RNNs)を一般的に使いま



(a) 高速道路走行シーン(後方撮影)

(b) CNNsのActivation Map

(c) 提案手法のActivation Map

図3 CNNs内部推論の可視化結果

CNNsの反応が強いところを可視化してみると,従来手法では, (b)のように空などの重要でない箇所が強く発火していて車などは反応していません。提案手法は, (c)のようにそのシーンに合わせた重要物体に着目して発火できるようになっています。

すが、毎フレームで高次元特徴を取得し過去の情報を使いながら推論を行うため学習データや計算機リソースを多く必要とし、ネットワークも複雑になりがちです。そこで、OpenPose³⁾で抽出したキーポイントを入力としてグラフを作成し、体のパーツや手のキーポイントで関連のある箇所をグラフで連結して、動きの情報をうまく記述できるように設計しました。

このグラフを、工場などの組み立て工程のタスクに適用したところ、作業者が変わっても手の動きの特徴的なパターンを抽出できるようになり、教師なしで作業者の動作分類が自動化できるようになりました⁴⁾。

2018年からは、マサチューセッツ工科大学(MIT)と共同研究を行いました。Antonio Torralba教授に師事し、Generative Adversarial Networks(GANs)やAnnotationの研究に従事しました。GANsは元々、Generatorというノイズから画像を生成するネットワークと、Discriminatorという生成された画像と正解の画像を見分けるネットワークの二つから成り、本物のように高精細な画像を自動生成したり、Image TranslationやStyle Transferといった画像の見かけを自然に変化させたりするタスクに用いられました。最近では、画像1枚からStyle Transferを行う技術⁵⁾(図5)や、ドメイン変換をGANsで行う研究があります。可視光を赤外光に変換するようなドメイン

を変化させる技術はデータが集めにくい環境を模擬して画像生成が可能なので企業向きと言えます⁶⁾。

企業で教師あり機械学習を行う場合、ターゲット環境でデータを集めて正解データを付与するAnnotationが必要になりますが、ここには無視できないほどのコストがかかります。そこで、GANsを用いた高精細な学習データが自動生成できればコスト削減が狙えると考え開発を進めました。同時に、Annotationを行う際にわずかな教師で自動的にデータを集められる、効率化を目指した技術開発も行いました⁷⁾。

2019年にはMERLとの共同研究で研究訪問し、Michael Jones, Tim Marks, Anoop Cherian, Alan Sullivanらのコンピュータビジョンチームの研究員に師事し、Object Detection, Semantic Segmentation, Face Recognition, Video Anomaly Detectionなどの研究に従事しました。世界で初めて顔検出技術の開発に成功したのはMERLで、Viola-Jones法として教科書でも取り上げられています。Michael Jones博士はその開発者です。近年では、Face Recognitionではコロナ禍でマスクをつけていても正確に顔のキーポイントが推定できる技術を研究しています⁸⁾。また、赤外線カメラで顔を撮影して動画から非接触でリアルタイムに脈波を推定する技術も開発しています⁹⁾。

Video Anomaly Detectionではあるシーンの正常状態を学習してそこから逸脱する動きがあった場合に異常として検出することができます。これまで、異常状態や不審行動などあいまいで定義することが困難な動きに対して機械学習を適用することはデータが集まらないために難しかったのですが、この技術はあいまいな検出対象の学習データを必要としないので製造現場や地域の防犯用途での運用が期待できます¹⁰⁾。

DX・AI・CV

研究所ではAI技術の習得や実課題の解決に取り組んできました。データが入手しづらかったり、安価な機器の上

で動作しなければいけなかったりと、実際に適用する環境への依存が大きい機械学習は事業での運用が難しいと感じていました。しかし最近では、クラウド運用が普及し演算処理ではなく通信速度がボトルネックになるくらい、高度なAIがリアルタイムサービスとして受けられるようになってきました。

2021年より三菱電機のビジネスイノベーション本部に加わり、Digital Transformation(DX)戦略室で顧客視点に立ってAIの利活用を促進するプロジェクトに参加しています。ここで主眼が置かれているのは、AIの精度や機能ではなく、Customer Value(CV)です。AIを使ったサービスでは、お客様のデータを集めて仮説を立て、適用するAIを選定したり作ったりして実証実験を行い、継続的にお客様に価値が提供できるかを見極めます。このサイクルを回す上では、常に変化する環境やお客様の要望に適応する共創活動が重要になります。これまでの機能や精度を追求する研究開発の知見を活かしつつ、実際にお客様と一緒にAIを使ったサービスを展開していきたいです。

ビジネスイノベーション本部は2020年に新設された組織で、三菱電機がこれまでに取り組んでこなかったビジネス形態や事業領域に切り込んで行く挑戦的な部隊です。オフィスはフリーアドレスなこともあり、組織のトップを含めたメンバとの物理的・精神的距離がとても近く、いつでも雑談できる雰囲気が好きです(図6)。コロナ禍で入社制限もあり、リモートワークが大半ですが、そんな中でも雑談会が週一以上で催されるので、職場の風通しはとてもよいと思っています。また、いち早くリモートワークで働ける仕組みを会社が提供してくれたおかげで安心して働ける側面もありましたし、Teamsなどのアプリを使った情報共有が進んで密に連絡を取るので、業務効率の改善につながっていると感じています。

大きな組織になると意思決定に時間がかかる傾向があり、スピード感が求



図5 GANsを用いたStyle Transfer⁴⁾



図6 職場風景 (岩橋亜沙子撮影)

められるIT分野において不利になることがあります。そんな時トップ層が話を速やかにまとめてくれるのがこの組織のよいところであり、若手のやりがいにもつながるところがあると思っています。バックアップが心強いです。

これからの取り組みとして、せっかく研究所でMERLとよい関係構築ができたので国際連携して、世界的な研究成果である最新のAIを三菱電機のサービスや製品に順次展開していきたいです。学生時代からの夢が叶いつつある幸運に感謝していますし、共に研究やプロジェクトに取り組むチームメンバーにも感謝しています！

振り返ってみると、国内外を問わず多才な方々と研究やプロジェクトを一緒に進めることができ、自分が大きく成長できたこともあって、メーカーに入社してよかったと思っています。

むすび

一般的なAIの定義は、Artificial Intelligence (人工知能) のことで、主に人の代わりに自動で推論するシステムを構築することを目指しています。

一方で、技術の成熟とともに現在の

定義だけでは解決できない課題も明らかになってきていて、AIの位置づけが見直され始めています。その一つが、Augmented Intelligence (拡張知能) で、人の手を離れて自律推論するのではなく、あくまで人の機能を拡張するツールとしてのAIを指しています。

2019年に行われたG20 Osaka Summitで、“AI原則の提言”が行われ、各国の目指すべきAIの姿として、AIを人間の道具として使うHuman-centered AI (人間中心型AI) という考え方が採択されました¹¹⁾。同じAIでも、自動化や機能の拡大や性能の向上だけを指すArtificial Intelligenceから、より成熟し顧客視点で製品やサービスを考えられるAugmented Intelligenceにシフトして、世の中に変革を起こせるとよいと思って、コラムのタイトルを決めました。この観点は社会実装に重きを置く、企業ならではの価値観のように感じるかもしれませんが、成熟した社会でAIとどのように向き合うか、AI倫理についても産業界だけでなく国際会議でも重要なトピックとして取り扱われています。

われわれメーカーも、人々の生活を豊

かにするために常に顧客視点に立って、Augmented Intelligence (拡張知能) としてのAIによるサービスや製品展開を目指します！ (2022年3月29日受付)

〔文 献〕

- 1) T. Sawada, T.-Y. Lee and M. Mizuno: "Bottom-Up Saliency Meets Top-Down Semantics for Object Detection", 2021 IEEE International Conference on Image Processing (ICIP), pp.729-733 (2021)
- 2) T. Sawada and M. Nakamura: "Intelligent Warning System Monitoring Vehicle Surrounding and Driver's Behavior", 2022 IEEE International Conference on Multimedia and Expo Workshops (ICMEW) (2022)
- 3) Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei and Y.A. Sheikh: "OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields", IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) (2019)
- 4) P. Das, J. Kao, A. Ortega, T. Sawada, H. Mansour, A. Vetro and A. Minezawa: "Hand Graph Representations for Unsupervised Segmentation of Complex Activities", International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp.4075-4079 (2019)
- 5) M.J. Chong and D. Forsyth: "JoJoGAN: One Shot Face Stylization", arXiv preprint (arXiv:2112.11641) (2021)
- 6) V.V. Kniaz, V.A. Knyaz, J. Hladůvka, W.G. Kropatsch and V.A. Mizginov: "ThermalGAN: Multimodal Color-to-Thermal Image Translation for Person Re-Identification in Multispectral Dataset", Computer Vision -- ECCV 2018 Workshops, Springer (2019)
- 7) T. Sawada, T. -Y. Lee and M. Mizuno: "Video Object Segmentation with Online Mask Refinement", 2022 IEEE International Conference on Multimedia and Expo Workshops (ICMEW) (2022)
- 8) T.K. Marks, A. Kumar, W. Mou, C. Feng and X. Liu: "UGLLI Face Alignment: Estimating Uncertainty with Gaussian Log-Likelihood Loss", IEEE International Conference on Computer Vision (ICCV) Workshop on Statistical Deep Learning for Computer Vision (SDL-CV), pp.778-782 (2019)
- 9) A. Comas, T.K. Marks, H. Mansour, S. Lohit, Y. Ma and X. Liu: "TURNIP: Time-series U-NET with Recurrence for NIR Imaging PPG", IEEE International Conference on Image Processing (ICIP) (2021)
- 10) B. Ramachandra and M.J. Jones: "Street Scene: A new dataset and evaluation protocol for video anomaly detection", IEEE Winter Conference on Applications of Computer Vision (WACV), pp.2569-2578 (2020)
- 11) Ministry of Foreign Affairs of Japan: "Ministerial Statement on Trade and Digital Economy" (2019)