

メディア工学の研究動向

村松正吾^{†1}, 渡邊 修^{†2}, 奥田正浩^{†3}, 青木義満^{†4}, 伊藤康一^{†5}, 東海彰吾^{†6}

1. まえがき

ここ数年の間、深層学習(DL: Deep Learning)技術に代表される人工知能(AI: Artificial Intelligence)技術が社会に浸透し、データ駆動によって設計されるシステムが、個人々の生活のみならず社会のありようを変える威力を備え始めた¹⁾。技術面では、代表的なネットワークアーキテクチャであった畳込みニューラルネットワーク(CNN: Convolutional Neural Network)や長短期記憶(LSTM: Long Short Term Memory)を利用した再帰型ネットワーク(RNN: Recurrent Neural Network)などから自己注意(SA: Self-Attention)機構やそのSA機構を利用したTransformerに研究開発の主流が移行した²⁾。自然言語処理(NLP: Natural Language Processing)分野における機械翻訳技術に端を発したTransformerは、文章の要約や校正のほか、自然な対話や流暢な長文生成、コード生成までも実現する大規模言語モデル(LLM: Large Language Model)の核技術となり、世界に大きなインパクトを与えている³⁾⁴⁾。

LLMの登場は、自然言語以外のメディア、例えば音声や画像、映像などにも影響が波及し、音声認識や音声合成、映像の自動キャプションなどに飛躍的な性能向上をもたらしている。Transformerのアイデアは、Vision Transformer(ViT)に引き継がれ、画像のコンテンツそのものを扱えるように展開し、既存のあらゆる画像処理タスクへの導入が試みられている⁵⁾。Audio Spectrogram Transformer(AST)

はViTのアイデアを音声・音響のメルスペクトログラムを扱えるよう修正されている⁶⁾。これらメディアの大規模モデル、特に生成系AIの研究開発により、人間に代わって機械に任せられるタスクが増え便利な生活が期待される。

一方、人間が創造した著作物の権利を脅かす存在、人間の仕事、思考力を奪う存在、社会を煽動する偽情報を生成する存在として懸念され、生成系AIが社会的に無条件に受け入れられる状況にはないことも事実である。DL技術は、高い性能が得られる反面、それらが導く結果の根拠がわかり難いという課題もある。特に、医療や自動運転車、防災への応用では、安全性を担保するために説明可能性や解釈可能性への追及も欠かせない⁷⁾。安心安全な技術は、稀にしか生じない事象、偶発的条件の組み合わせへの対応が必須である。しかしながら、応用によっては少量のデータしか入手できない。これは汎用的な大規模モデルが苦手とする領域である。その領域(ドメイン)で専門家が有する事前知識や理論を活用し、結果に対する説明責任を果たせる仕組みが重要である。

このような背景から、モデルベースによるメディア処理についても関心が寄せられている。事前知識を積極的に利用する手法としてスパースモデリングがある^{8)~10)}。この解法をデータ駆動と組み合わせるとDLの枠組みで実現する深層展開技術は、事前知識をニューラルネットワークと融合させる手段として注目されている^{11)~15)}。内閣府が提唱する日本の未来社会の姿Society 5.0では、サイバーフィジカルシステム(CPS: Cyber-Physical System)の活用が期待されている。CPSでは、さまざまな物理現象の状態の把握、予測、制御が求められる。データ駆動による動的モデリングが重要な役割を担う¹⁶⁾。したがって、物理法則や制約を事前知識として利用するDLやスパースモデリングは今後重要性が増すだろう^{17)~22)}。

メディア工学の分野においても生成系AIや事前知識の活用といった技術動向の影響を受け、日々進化している。本稿では、このような動向に関連し、限られた紙面の上でいくつかの注目すべき技術を取り上げ、その概要および今後の展望を紹介していく。第2章では圧縮・符号化技術、第3章では計測・再構成技術、第4章では認識・生成技術、

†1 新潟大学 工学部 工学科 電子情報通信プログラム

†2 拓殖大学 工学部 電子システム工学科

†3 同志社大学 理工学部

†4 慶應義塾大学 理工学部 電気情報工学科

†5 東北大学 大学院情報科学研究科

†6 福井大学 学術研究院 工学系部門 情報・メディア工学講座

"Recent Researches on Media Engineering" by Shogo Muramatsu (Electronics, Information and Communication Engineering Program, Faculty of Engineering, Niigata University, Niigata), Osamu Watanabe (Department of Electronics and Computer Systems, Takushoku University, Tokyo), Masahiro Okuda (Faculty of Science and Engineering, Doshisha University, Kyoto), Yoshimitsu Aoki (Department of Electronics and Electrical Engineering, Faculty of Science and Technology, Keio University, Yokohama), Koichi Ito (Graduate School of Information Sciences, Tohoku University, Sendai) and Shogo Tokai (Information Science, Faculty of Engineering, University of Fukui, Fukui)

第5章ではセキュリティ技術、第6章では可視化・拡張技術の動向について紹介する。第7章ではメディア工学技術の展開について話題を提供し、最後に、第8章で今後の研究動向についてまとめる。

2. 静止画像符号化の標準化動向

本章では、静止画像の符号化技術の国際標準化を担うISO/IEC JTC 1/SC 29/WG 1(以下、JPEG委員会と略記)における標準化プロジェクトの動向を紹介する。現在、JPEG委員会ではJPEG, JPEG 2000, JPEG XT, JPEG XS, JPEG XLといった多値2次元画像の変換符号化方式や、ライトフィールド、ホログラフィ、ポイントクラウド画像の符号化プロジェクトであるJPEG Pleno、深層学習を要素技術とするJPEG AIなど、多数のプロジェクトがアクティブに議論されている。本稿では、多値2次元画像変換符号化、JPEG Pleno, JPEG AIの動向に焦点を当てる。

2.1 多値2次元画像の変換符号化プロジェクト

1992年に誕生したJPEG(以下JPEG-1と記述)²³⁾²⁴⁾は、インターネットやデジタルカメラの普及とタイミングが重なった背景から、幅広い分野で採用された。現在においてもJPEG-1は現役の符号化方式といえよう。JPEG-1は、空間画像を周波数変換し、周波数領域において人間の視覚特性を考慮した量子化を行い、量子化係数をエントロピー符号化するという変換符号化方式の一つである。後発の規格においてもこの基本構造は変わっていない。JPEG-1では、周波数変換に2次元ブロック離散コサイン変換(DCT: Discrete Cosine Transform)、量子化には量子化マトリクスを用いた線形量子化、エントロピー符号化にDPCM(Differential Pulse-Code Modulation)およびハフマン符号化がそれぞれ採用されている。基本方式であるベースライン方式では、入力画像のビット深度が8 bit/pixelまでという制限が存在する他、高圧縮時に生じるブロック歪みや、モスキート雑音等の問題が指摘されてきた。

これを受けて、8 bit/pixelを超えるビット深度に対応し、より良い符号化効率を持つ符号化方式としてJPEG 2000の標準化が開始され、2000年に基本方式であるPart 1が出版された²⁵⁾。周波数変換に離散ウェーブレット変換(DWT: Discrete Wavelet Transform)、量子化にはサブバンド毎に異なるステップサイズを用いる線形量子化、エントロピー符号化にはコンテキストベースの適応2値算術符号化が採用されている。JPEG 2000は、柔軟なスケラビリティ機能を持つことに加えて、1パスレート制御(所望のビットレートを量子化を繰り返すことなく達成すること)を行うことが可能である。これらの機能を実現たらしめているのは、エントロピー符号化を担うEBCOT(Embedded Block Coding with Optimized Truncation)アルゴリズムである。JPEG 2000はその符号化効率や豊富な機能と引き換えに、EBCOTに起因する膨大な演算量が要求されるた

め、採用が広まったのはデジタルシネマや医療・衛星画像、アーカイブといったプロフェッショナル分野であった。しかしながら、2019年に出版されたHTJ2K(High-Throughput JPEG 2000)²⁶⁾²⁷⁾によってEBCOTを高速なブロックコードに置き換えることが可能となり、JPEG 2000の欠点であった高い計算コストとそれに伴う低いスループットという問題点は解決されている。

JPEG-1やJPEG 2000は、用いられるアプリケーションを限定せず、幅広い符号化レートで良好な性能を発揮することを目的としているが、2019年に規格化が完了したJPEG XS²⁸⁾²⁹⁾では、低圧縮・低遅延なメザニン符号化方式を実現することを目的に開発された。JPEG XSのターゲットは、1:2~1:6程度の圧縮率、許容されるレイテンシがサブフレーム(1フレーム未満)といったアプリケーションである。周波数変換にはラインベースのDWT、量子化には線形量子化およびTCQ(Trellis Coded Quantization)、エントロピー符号化にはBit plane count codingと呼ばれる、ビットプレーン係数の有意性に基づく方式が採用されている。JPEG XSはFPGA(Field Programmable Gate Array)等を用いて効率的な実装が可能となるよう、ハードウェアによる実装を強く意識した標準となっている。

JPEG-1を置き換えるべく、いくつかの符号化方式が提案されてきたが、いずれもコンシューマ用途においては成功を収めたとは言い難い。JPEG-1にはHDR(High Dynamic Range)画像への対応や符号化効率等の問題が存在し、より良い方式が望まれる一方、JPEG-1との互換性を持たない方式ではその代替となることは難しい。2022年に規格化が完了したJPEG XL^{30)~32)}はJPEG-1とのトランスコーディング機能によってJPEG-1との後方互換性を持たせた標準である。JPEG XLの符号化性能についてはWebサイト*1で確認することができる。規格化完了以降、さまざまなアプリケーションで急速に採用が進んでおり、直近ではAdobe社や、Apple社も、自社の製品におけるJPEG XLのサポートを表明した*2*3。特に、Apple社はOSレベルでのサポートを表明しており、今後、モバイルにおけるWeb用途を中心にJPEG XLによって符号化された画像の流通量が増大することが予想される。JPEG XLの主な要素技術は、XYB色空間、可変ブロックサイズDCT、適応量子化、ANS(Asymmetric Numeral Systems)符号化である。多くの画像符号化方式で、RGB画像を輝度色差空間であるYCbCr色空間あるいはその派生に変換しているが、JPEG XLでは、より人間の視覚特性を考慮した色空間であるXYB色空間を採用した。DCTのブロックサイズは4×4~256×256と可変となっており、画像の局所的性質に応じ

*1 https://jon-cld.s3.amazonaws.com/test/ahall_of_fshame_SSIMULACRA_2_modelD.html

*2 <https://developer.apple.com/documentation/safari-release-notes/safari-17-release-notes>

*3 <https://helpx.adobe.com/camera-raw/using/hdr-output.html>

て最適なサイズを選択することが可能となっている。JPEG XLにおける量子化は、JPEG-1と異なり、画像の局所的性質に応じて量子化マトリクスをスケーリングすることが可能である。エントロピー符号化として採用されたANS³³⁾ 34) は、2009年に発表された比較的新しいエントロピー符号化アルゴリズムである。ANSの特徴は、算術符号化と同等の符号化効率を持つことと、ハフマン符号化に匹敵する処理速度を達成可能であることである。

2.2 JPEG Pleno

JPEG Plenoプロジェクト^{35)~37)}は、従来の平面的な2次元画像にとらわれず、ライトフィールド、ポイントクラウド、ホログラフィといった3次元的な画像情報をスコープにした標準ファミリーである。Part 1では、ライトフィールド、ホログラフィ、ポイントクラウドの三つのモダリティ間のインタオペラビリティを定義するフレームワークが規定されている。ライトフィールド画像の符号化については、Part 2で規定されており、カメラのビューポイント座標 (u, v) とそのビューポイントにおける画像の各画素位置 (x, y) を4次元信号として捉え、4次元DCTを行う符号化方式が採用されている³⁸⁾。ホログラフィ画像の符号化方式はPart 5において審議中であり、現在DIS (Draft International Standard) 段階にある。ホログラフィ画像信号は、通常の画像信号と大きく異なる性質を持つ。具体的には、光の干渉パターンを表す、2次元の複素信号となる。Part 5で審議中の方式では、入力信号に対して短時間フーリエ変換 (STFT: Short-time Fourier transform) を施し、その後局所ブロック毎に、RD最適化処理に基づく量子化ステップサイズを求め、Fixed-point Arithmetic Encoding³⁹⁾を行うことが検討されている。ポイントクラウドの符号化についてはPart 6において審議中であり、DLに基づく方式の提案があり、現在、その性能を検討中である⁴⁰⁾。

2.3 JPEG AI

DLを要素技術とした画像符号化方式の国際標準を目指して、2021年初頭にISO/IEC 6048としてJPEG AIが新たなプロジェクトとして承認された⁴¹⁾。産業界、学術界から高い関心を集め、Final Call for Proposalには6カ国、11のチームより13の方式提案があった⁴²⁾。JPEG AIの主な要求仕様は、既存の標準と同等程度の主観画質をより低いビットレートで達成すること、圧縮領域での画像処理やコンピュータビジョン関連のタスクを実行可能にすること、の二つである。ただし、これらの要求仕様は段階的に達成される見込みであり、現在は、圧縮効率の観点を中心に標準化作業が進められている。JPEG AIへの貢献は中国企業が中心となっている。

2023年7月会合終了時に、CD (Committee Draft) が発行され、技術的仕様はほぼ固まったといえる。現段階で、VVC (Versatile Video Coding, ITU-T H.266|ISO/IEC 23090-3) イントラモードと比較して最大40%ほどのビット

レート削減が確認されている⁴³⁾。その一方で、CPUベースの演算量の評価を見ると、実行時間ベースの比較で同じくCPUベースのVVC実装と比較して4倍程度の時間がかかっており、今後は演算量の削減、あるいはGPU・NPU (Neural Processing Unit) へのオフロードによってどの程度実行時間が短縮できるのかが焦点となると予想される。今後は、画質をできる限り保ったままデコードに要する演算量をいかに削減するかが課題となるが、7月会合中にスマートフォンで動作するデコーダのデモ実装が披露され、演算量削減という課題はまもなく解決されることが予想される。

2.4 その他のプロジェクト

生成AIの登場によって、フェイクメディアや悪意ある画像の捏造など、画像データの真正性 (Provenance) を担保する枠組みの構築が課題となっている。また、SNS (Social Networking Service) 等に投稿した画像がプラットフォームによって学習データとして利用されることに抵抗を感じるユーザも多く、自身の持つデータの学習利用へ許諾をどう実現するかという問題も存在する。2023年4月にISO/IEC 21617-1として正式承認されたJPEG Trustプロジェクトは、これらの諸課題を解決するためのフレームワークを提供すべく、現在標準化作業が活発に進められている。

3. 計測・再構成技術の動向

イメージセンサと信号再構築技術は相互に強く依存した関係性を有している。近年では、イベントセンサ、マルチプルカメラ、デュアルピクセルセンサ、偏光フィルタなどのセンシング技術と高度な画像再構築手法を組み合わせることによって、デブラー、ヘイズ除去、反射除去、高ダイナミックレンジの実現、深度推定などの多岐にわたる画像処理・ビジョンタスクに効果的に適用されている。本章では、その中でも特に近年積極的に研究されている計測技術であるイベントセンサと、自己教師を利用した画像再構築手法に焦点を当てる。

3.1 イベントセンサ

イベントセンサは、ピクセルごとの明るさの変化を非同期に測定し、イベントストリームとして輝度変化を出力する新しい種類のセンサである。イベントセンサは、従来のセンサでは捉えにくい高速に変動するシーンを捉えるのに適している。主な応用例としては、オブジェクトトラッキング、モニタリング、深度推定、オプティカルフロー推定、HDR画像再構築、SLAM (Simultaneous Localization and Mapping) などが挙げられ、自動運転、ロボティクス、産業用検査、医療、セキュリティ、スポーツ分析など、多岐にわたる分野での応用が見込まれている。

イベントセンサは各ピクセルが独立に、輝度変化の有無のみを測定して出力する特性を有しており、標準的なフ

レームベースのカメラのように、全画素で統一した露光時間を待つ必要がなく、低いレイテンシーで高速物体を捉えることが可能になる。輝度変化のみを出力するため低消費電力であり、また120 dB以上の高いダイナミックレンジを実現できるため、室内暗所から日中のシーンまでの情報取得が可能となる。

イベントセンサが出力する非同期の輝度変化情報を一般的な画像形式に変換する技術は、多くのアプリケーションで不可欠となり、これまでにさまざまな手法が提案されている^{44)~50)}。高画質を達成している手法の多くは、低フレームレートのカメラとイベントカメラを同期させることで、高解像度かつ高フレームレートの実現を図っている。イベントセンサのみから動画を再構築する研究も多く行われているが、現状では画質向上が主要な課題とされている。

3.2 画像の自己教師あり復元

劣化した画像から高解像度の画像への復元技術は半世紀以上の歴史を持つ一方、近年、深層ニューラルネットワーク (DNN: Deep Neural Network) を活用した教師付き学習手法が多く提案され、多様なタスク (ノイズ除去、超解像度、デブラー、画素補間等) で優れた成果を実現している。これらの教師付き学習手法は、劣化画像とそのオリジナルの画像がペアとして提供されることが通常である。しかしながら、実世界の応用例において、劣化のない画像が利用できない場合には、これらの方法が適用できないという制約が存在する。

DL技術の発展に伴い、教師データとしてオリジナル画像を必要としない、自己教師付き学習およびゼロショット学習を用いた画像復元法が提案されている^{51)~53)}。これらの手法は主にCNNの潜在的バイアスを用いている。つまり、CNNは畳込み処理を通じて、対象の画素と周辺画素との関係性を解析する特性を持つ、この特性を利用し、教師データなしでも低周波成分にエネルギーが集中する画像の復元が可能となる。

そのなかでもNoise2Void⁵²⁾に代表されるブラインドスポット戦略を用いた手法が高い性能を示すことがわかっている。これは、学習フェーズで周囲のピクセルから意図的に一部のピクセルを隠し (ブラインドスポット)、それを予測することでノイズ除去モデルを構築するアプローチを指す。最近の自己教師付きノイズ除去方法であるBlind2Unblind⁵³⁾も、ブラインドスポット戦略を導入している。このアプローチでは、ブラインドスポットが 2×2 の領域に配置され、異なるブラインドスポットパターンを含む多数の入力が供給され、それぞれの入力に対応する復元画像を得る。出力は、得られた復元画像からそれぞれのブラインドピクセルに対応するピクセルのみを抽出し、組み合わせることによって得られる。

これらの手法の多くは、広範な劣化画像データセットを必要とする。よって大量のデータを準備するコストが高い

ドメインには不向きである。具体的にはハイパースペクトル画像 (HSI: Hyper-Spectral Image) がこれに該当する。このようなドメインには、単一の劣化画像のみを用いるゼロショット学習手法が効果的である。Deep HS Prior⁵⁴⁾は、Deep Image Priorの一種の拡張であり、観測された劣化HSIに適用可能である。しかしながら学習プロセスが進んだ場合、観測された劣化HSIをそのまま出力する欠点がある。文献⁵⁵⁾では、劣化プロセスの情報を活用して上記の問題に取り組んでいる。このモデルは、観測された劣化HSIを目標データとし、それに沿ったさらなる劣化HSIを入力データとして性能を向上させている。文献⁵⁶⁾ではブラインドスポット戦略を用いたHSI復元手法を提案し、この手法は劣化プロセスの情報を必要とせずにゼロショットで高品質な復元が可能であることを示している。現時点ではこれらのゼロショット復元はノイズ除去などのタスクに偏っているが、今後他のタスクにも展開していくことが予想される。

4. 認識・生成技術の動向

画像認識においては、CNNの進化と活用が進み、物体検出、人物姿勢推定や行動認識、意味的領域分割、動画画像認識など、さまざまなタスクへの有効性が示されてきた。2020年以降、NLP分野において利用されていたTransformerベースのアーキテクチャが画像認識タスクにも適用されるようになり、CNNベースの認識性能を上回るパフォーマンスを発揮している。特に、2021年に登場したViTは、画像認識におけるモデルアーキテクチャやアプローチの多様性を生み出し、DLの研究における重要なターニングポイントとなっている。ViTは、画像を一連のパッチに分割し、それらをトークンとして扱い、Transformer構造を使用してこれらのトークン間の関係を学習する方法を導入した。また、Transformerのアテンションメカニズムは、局所的な情報だけでなく、画像全体のグローバルな関係を捉えることができるため、従来のCNNが難しいとされていた画像の大域的な構造や関係を効果的に捉えることが可能となった。ViTの学習においては、大量のデータと計算リソースでの事前学習が効果的であることが示され、特に大規模なデータセット上でViTを事前学習することで、その後のタスク固有のファインチューニング時の性能が大幅に向上することが示された。

CNN, ViTを問わず、大量の画像を有するデータセットで事前にモデルを学習し、その結果をもとに別の小規模なデータセットで学習を行うという2段階の学習が主流となっている。この1段階が事前学習、2段階が転移学習であり、ターゲットドメインの小規模なデータセットしかない場合でも、事前学習のデータセットを適切に準備すれば、高性能なモデルを構築できる点が注目されている。しかし、大規模な事前学習用の画像データセットについては、ImageNet等のオープンデータがあるが、実社会のさまざま

まな課題に対して有効な事前学習ができるとは限らない。また、近年では、データセットの公平性や個人情報、著作権保護の観点から、実世界のデータを用いる際には十分な配慮が必要となっている。これらの問題を解決する方法として、正解ラベルを用いない、もしくは実世界データを利用しない事前学習法が発展してきている。ここでは、ViTにおける自己教師あり学習、および数式ドリブン教師あり学習について述べる。

ViTにおける自己教師あり学習としては、CLIP (Contrastive Language - Image Pre-training)⁵⁷⁾ という手法がある。テキストと画像のペアを自己教師ありの方法で学習することを目的としている。特定のテキストとそれに関連する画像を近づけ、関連しないものを遠ざけるという方法で、画像とテキストの関連性をコントラスト学習している。アーキテクチャとしては、ViTを画像処理に、Transformerアーキテクチャをテキストの処理に使用しており、物体認識から芸術的な内容の生成まで、多岐にわたる視覚的タスクに適用可能であることが示されている。CLIPの学習は、インターネットから収集された大量のテキスト-画像ペアに基づいており、この大規模なデータセットを用いることで、多様なシナリオやコンテキストでの性能が向上している。一方で、Webからの学習データには著作物や有害コンテンツが含まれる可能性がある点に注意が必要である。テキストと画像という異なるモダリティを含むマルチモーダルな基盤モデル (Foundation Model) として、現在さまざまな改良や派生モデルが登場している。

次に、まったく実世界データを利用しない事前学習手法である、数式ドリブン教師あり学習⁵⁸⁾ について紹介する。これは、ある法則により画像パターンとそのペアになる画像カテゴリーを自動生成することで大規模画像データセットを構築し、DLモデルを事前学習する枠組みである。大規模データセットによる事前学習の重要性については前述したとおりであるが、近年、データセットの持つ公平性などのAI倫理の問題が大きく注目を集め、国際的に法規制を含め検討されている状況である。数式ドリブン教師あり学習では、数式に基づいて生成されるフラクタル幾何画像をデータセットとして用いることで (FractalDB⁵⁸⁾)、自然画像を用いずに画像認識に必要な視覚特徴表現を獲得することができる。数式を用いて画像パターンを生成するので、著作権やプライバシーといった画像使用に関する権利問題に抵触することがないことから、AI倫理に関する心配もなく学習済みモデルを構築することができる点でさらに注目を集めている。シンプルな画像識別から始まり、現在では難易度の高い画像領域分割⁵⁹⁾ や画像超解像タスク等、さまざまな応用が期待されている。

画像生成については、敵対的生成ネットワーク (GAN: Generative Adversarial Network)⁶⁰⁾ や変分自己符号化器 (VAE: Variational Autoencoder) が主流となっていた中、

新たなモデルとして拡散モデル (Diffusion Models) が注目されるようになった。拡散モデルベースの画像生成は、GANやVAEを上回る高精細でバリエーションに富んだ画像を生成できる手法として進化し、DALL-E2⁶¹⁾ やStable Diffusion⁶²⁾ などのテキストからの高品質な画像生成手法が特に注目を集めている。

以上、ここ数年で、画像認識、画像生成ともに、基盤モデルをベースとしながら、テキストを含むマルチモーダルな学習を行う手法が大きく発展し、注目を集めている。大規模データセットの構築コストや法的規制の課題もあることから、どちらのタスクについても、事前学習した基盤モデルをいかにリアルドメインの課題に、省コストで適応させるかが大きな研究開発課題となっている。

5. セキュリティ技術の動向

これまでに述べてきたように、DLやCNNが登場して以来、画像認識をはじめとする各種画像処理の性能が著しく向上し、今までは解くことが難しいとされていた多くの問題が解かれつつある。一方で、その高い性能が故に、大量のデータで学習された機械学習モデルには、安全性に関する特有の問題がある。そのため、最近では、CNNなどを用いた画像認識モデルの安全性に関する研究が盛んに行われている。本章では、筆者 (伊藤) が主として行っている生体認証 (バイオメトリクス)⁶³⁾ から顔認証⁶⁴⁾ に絞ってセキュリティ技術の動向について紹介する。

5.1 Adversarial Examples

敵対的サンプル (AE: Adversarial Examples) は、視認できないような微小な摂動を画像に加えることで誤認識させる手法である⁶⁵⁾。AEでは、GANなどを用いて、入力画像を別のクラスに誤認識させつつ、入力画像の見た目を変えないようなノイズを画像に付与する。一般的な画像認識モデルを誤認識させる手法であるが、顔画像に対して誤認識させる摂動を加えることで顔認証モデルを誤認識させることができる。画像中の1画素を置き換えるだけで誤認識を誘発できることも報告されている⁶⁶⁾。顔認証においては、眼鏡のフレームに摂動を加えることで誤認識させる手法が提案されている⁶⁷⁾。AEは、画像認識モデルを誤認識させる攻撃としてだけでなく、防御にも用いられている。例えば、顔認証において、AEを用いてプライバシー保護を行う手法が提案されている⁶⁸⁾⁶⁹⁾。SNSなどの利用が進み、インターネット上から顔画像を容易に収集することができる。例えば、顔認証を用いて特定の人物の顔画像を集めたり、顔認証システムへの攻撃に使用したりすることができる。このような不正利用を防ぐために、顔画像の非識別化 (de-identification) が検討されている。AEを利用することで顔画像の見た目をほとんど変えずに顔画像の人物として認証させないようにすることで顔画像のプライバシーを保護することができる。AEを用いた手法⁶⁸⁾⁶⁹⁾ では、非識別

化性能(摂動の大きさ)と画質がトレードオフになる。これに対して、筆者(伊藤)は、特徴埋め込みによる顔画像の非識別化手法を提案している⁷⁰⁾。公開する顔画像に他人の顔特徴量を埋め込むことで、人間には顔画像を認識させつつ、顔認証モデルには別人として認証させることができる。

5.2 DeepFake

DL, 特にGANの登場により、画像生成技術が爆発的に進展した。それに伴って、画像生成により本物と見分けがつかない画像を作る攻撃手法が提案されている。その中でもDeepFakeは、現実的な脅威となり得る。例えば、有名人の口元の画像を差し替えることで異なることを話しているような動画画像を生成することができる。そのため、DeepFakeによる偽物の生成とその検知に関する研究が盛んに行われている⁷¹⁾。Meta AIが大規模なDeepFakeデータセットを構築し、それをういたDeepfake Detection Challenge *4をKaggleにて開催した。DeepFakeは、顔画像だけではなく、音声や姿勢などさまざまな個人性に適用される。現在では、DeepFakeが一つの大きな研究分野になりつつある。

5.3 なりすまし

なりすまし(spoofing)は、顔認証システムにおける重要な課題の一つである⁷²⁾。顔認証は、撮影環境の変化に対して高いロバスト性を有していなければ、実用性が低下する。一方で、顔の向き、明るさ、表情などの変化にロバストであるため、顔認証システムのユーザの顔写真でも認証されてしまう可能性がある。顔認証システムとしてのロバスト性を高めることで、なりすまし攻撃に対応できなくなるため、撮影された顔画像が本物かなりすましかを認証する前に検知する必要がある。なりすましは、印刷された顔画像、タブレットで再生された顔動画、3次元マスクをかぶった別人などさまざまなアプローチがある。これらに対応するためには、大規模なデータセットで学習されたCNNが有効である。例えば、顔動画からrPPG(Remote Photoplethysmography)を検出し、CNNで抽出されるテクスチャ特徴量と組み合わせるなりすましを検知する手法⁷³⁾や、CNNの畳込み演算を中心差分をとることで微小な変化を捉え、なりすましを検知する手法⁷⁴⁾が提案されている。多くの手法は顔に注目しているが、印刷された顔画像やタブレットで再生された顔動画は、背景にもなりすましの要素が含まれる。顔以外の領域も考慮してなりすましを検知するために、筆者(伊藤)らは、ViTに基づくパッチベースの検知手法を提案している⁷⁵⁾。それでも未知のなりすまし攻撃を検知することは困難であり、現在でも数多く手法が提案されている。

以上のように、DLの進展に伴って困難とされてきた多くの課題が解決されたり、画像認識などの性能が大幅に向上してきた。一方で、本章で述べたように、画像認識モデルの安全性を保つために、数多くの研究が進められている。DLに

よる性能向上だけではなく、DLによって新たに生まれた安全性に関する懸念が多く生じている。コンピュータビジョンのトップカンファレンスであるIEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)においてセキュリティ技術が一つの分野になりつつあるように、DLならびに機械学習に関するセキュリティは、大きな研究の潮流となりつつある。

6. 可視化・拡張技術の動向

視点自由度の高いシーン状況の可視化は自由視点画像・映像の生成としてメディア工学分野の重要な研究トピックの一つであり、光線情報の獲得・提示・処理技術を中心に20年を超える歴史がある。近年、自由視点像を生成する技術としてNeRF (Neural Radiance Field) と呼ばれる技術が注目されている。

NeRFは複数視点から撮影された被写体の静止画像群から、ニューラルネットを利用して任意視点からの静止画像を生成する技術である⁷⁶⁾。被写体を取り囲むように撮影された30~50枚の画像を入力として、その被写体シーン状況に対する多層パーセプトロン(MLP: Multi-Layer Perceptron)構造の256ノード×12段の全結合ネットワークを学習する。このMLPは、空間の点座標と観察方向を5次元のベクトル (x, y, z, θ, ϕ) として入力し、その点における色と透過度 (r, g, b, σ) を出力する関数 F_θ を学習する。得られた F_θ を用いて、任意視点から見たシーンの像を、ボリュームレンダリングの技術を用いて画像化する方法である。

F_θ の学習では、入力される実撮影視点画像の再現に加え、被写体のエッジ部分の見えの再現を扱う損失関数を設定し、シーンに対して F_θ をオーバフィッティングすることで、対象とするシーンの光線情報を再現できる F_θ を得ている。この手法の特徴として、生成される自由視点からの像の品質が高いことがあげられ、従来法に比べて、局所的にシャープなエッジの見えの再現や、鏡面反射や透明物体の見えの再現が可能である(図1)。

一方で、入力となる画像群について、その撮影位置や撮影方向が十分な精度で獲得できることが求められること、および F_θ の学習に相当の時間を要すること、 F_θ が単一の静的空間のみしか扱えないこと等が問題点である。これらの問題の解決策の提案も活発に行われている。例えば、学習処理の高速化に関する手法⁷⁷⁾、動的シーンへの適用⁷⁸⁾、シーンの見え方を事後的に修正するエディティング手法⁷⁹⁾等が提案されており、今後もNeRFをベースとした技術動向が注目される。

NeRFに対してある程度の高速化手法が適用できるようになったことから、NeRFをベースとしたサービスも提供されつつある。LumaAI社が提供するアプリケーションでは、スマートフォン用の専用アプリで撮影した被写体画像群をアップロードさせ、そのシーンの可視化モデルとして

*4 <https://ai.meta.com/datasets/dfdc/>

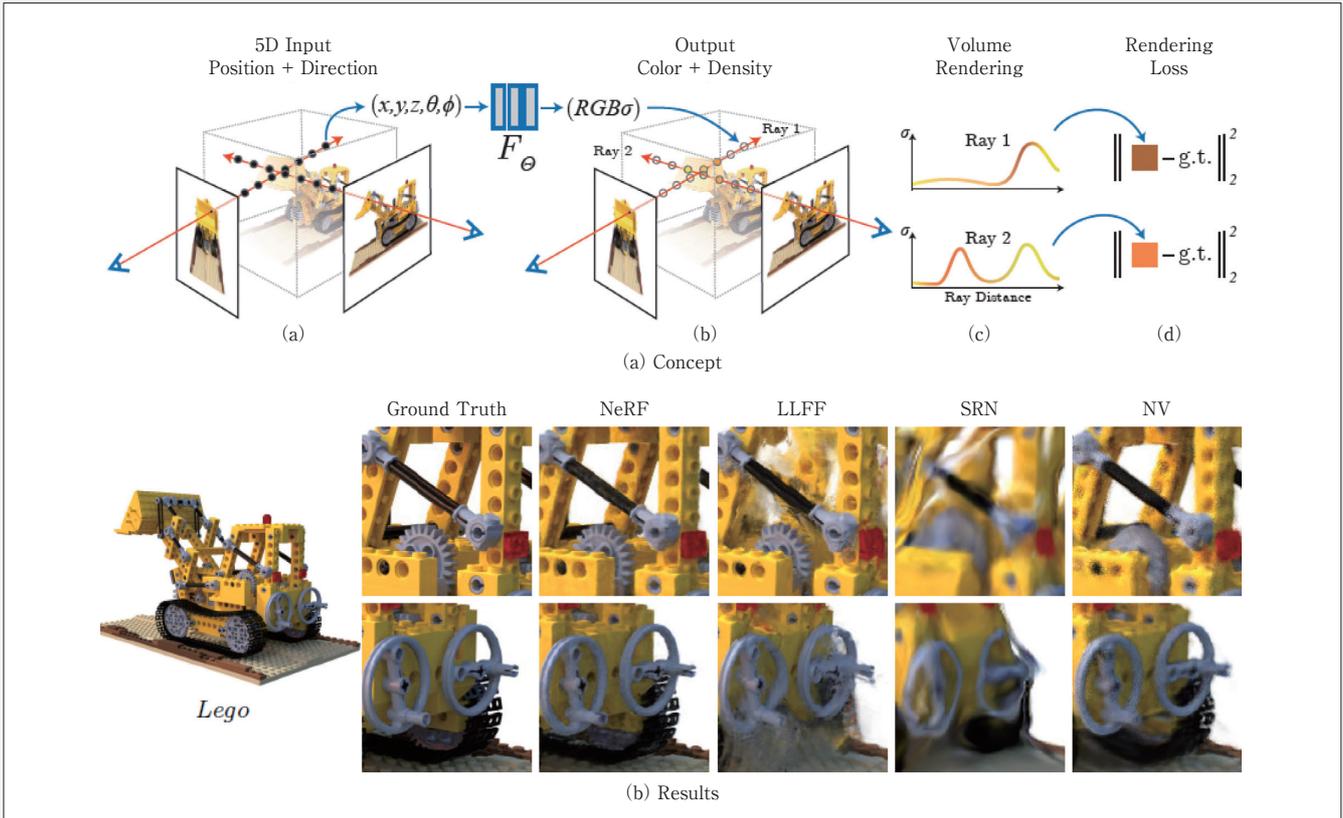


図1 NeRFの原理と結果の比較（文献⁷⁶より）

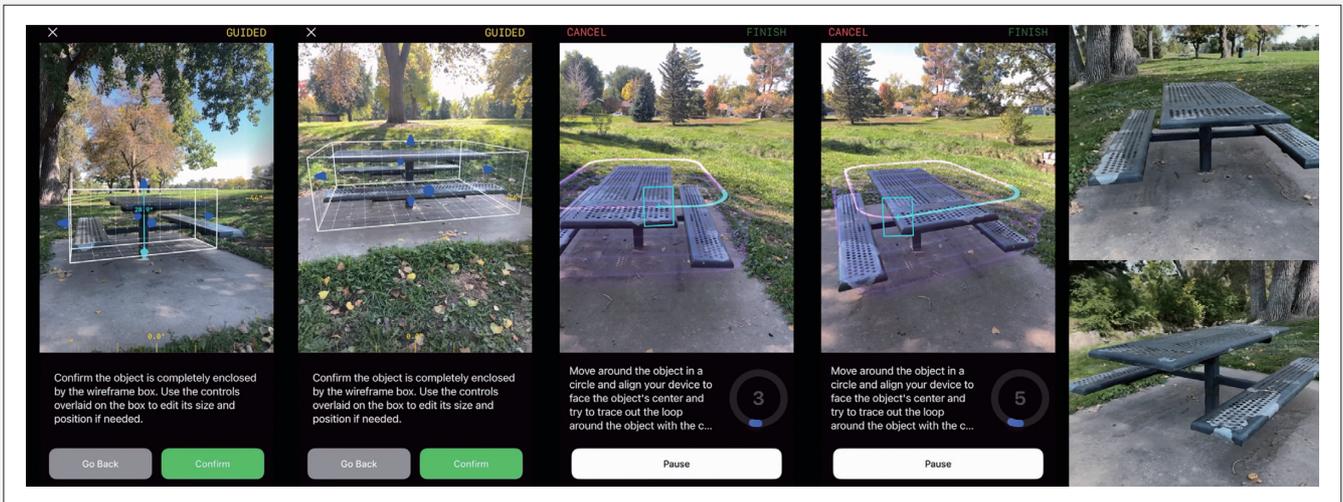


図2 Luma AI社アプリと生成例（PR映像⁸¹より）

の F_{θ} を算出して提供する⁸⁰。現状はまだ処理時間を有するが、任意視点の像を可視化できるモデルが提供され、ゲームコンテンツなどに利用できる。このアプリケーションでは、従来からのSfM (Structure from Motion) による被写体の大まかな形状獲得とスマートフォンの持つAR (Augmented Reality) 機能を用い、NeRFによって理想的な品質の結果を得られるように、ユーザにメインの被写体を選択させ、それに応じた撮影視点や撮影方位をARを併

用しながらユーザに指示する機能がある（図2）。今後も、ユーザによる撮影操作を伴うさまざまなスマートフォンアプリケーションにおいて、結果の品質を向上させるための指示やインタラクションの手法が適用されることが予想され、メディア工学の一つの切り口として注目される。

このような、スマートフォンをベースとしたメディア処理の一つのアプローチとして、多視点撮影時の撮影支援技術が挙げられる⁸²。複数の撮影者が場を共有する状況にお

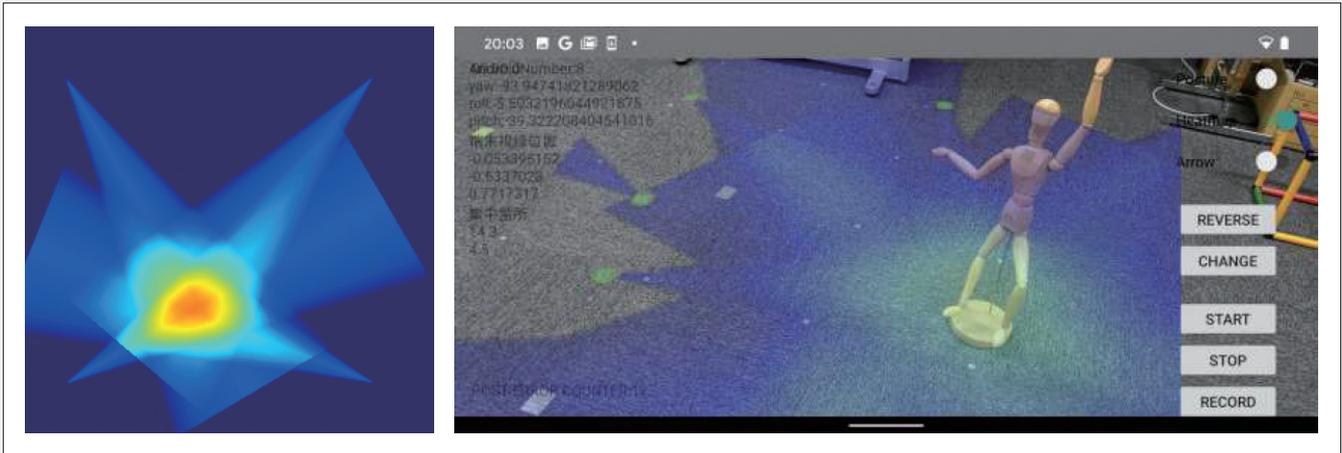


図3 撮影支援システム(左:視野集中解析結果, 右:重畳表示例)(文献⁸²⁾より)

いて、撮影者らがどのような場所を撮影しているかを共有することで、個々の撮影者の撮影を支援する(図3)。具体的には、それぞれの撮影者の位置や撮影方位を時々刻々収集し、その重なりの様子を各撮影者にフィードバック、さらにその情報を撮影像に重畳表示することで、シーンにどの部分を撮影すればよいかのリコメンドとして利用する。現状、スケーラビリティや3次元化に課題があり、研究が進められている。

7. メディア工学技術の展開

未来の技術を見極めることは困難なため、本章は筆者(村松)が関心を持つ内容に偏る点にご容赦いただきたい。

計測技術の発展により、音や画像に限らず多様かつ膨大なデータの取得が可能となった。同時に、複雑な物理現象の理解、予測、制御への要求が高まっている¹⁶⁾。近年は、地震のほか大雨や火災などの大規模な災害が頻発し、人命と社会資産の損失が繰り返されている。防災の観点からも、複雑な物理現象の理解と予測、CPSによる制御が希求されている。道路や橋など社会基盤の老朽化が進む一方、土木技術者の年齢構成が偏り、熟練者から若手への技術と知識の継承ができず、メンテナンスすら困難な状況である。超高齢化や気候変動、国際情勢、経済格差などに依存する多くの社会的課題に、不足する人材で対応していかなければならない。

例えば、防災を目的としたCPSを構築するためには、対象となる物理現象を的確に表現できるプラントモデルもしくはデジタルツインが必要である。動的システムは、状態ベクトル \mathbf{x} と時刻 t に依存するベクトル場 \mathbf{f} により、 $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t)$ と記述される。状態ベクトル \mathbf{x} の時間発展はベクトル場 \mathbf{f} に支配される。複雑な物理現象を対象とする防災科学やロボットなどの分野では、 \mathbf{f} をデータ駆動により与える方法が主流となりつつある。データ駆動型モデリングでは、学習データを用いた機械学習によってベクトル場 \mathbf{f} を近似する。固

有直交分解(POD: Proper Orthogonal Decomposition)、動的モード分解(DMD: Dynamic Mode Decomposition)¹⁶⁾⁸³⁾、線形回帰自己符号化ネットワーク(LRAN: Linearly-Recurrent Autoencoder Network)⁸⁴⁾、多重解像度畳込み自己符号化器(MrCAE: Multi-Resolution Convolutional Autoencoder)⁸⁵⁾などの例がある。自己符号化器は、画像などメディア処理のタスクでもしばしば利用されるが、動的システムの時間発展の特徴を効果的に捉えることにも利用できる⁸⁶⁾。

ここで筆者(村松)らの研究事例を紹介したい。文献⁸⁷⁾では、畳込み合成辞書をDLの枠組みで設計し、この辞書とDMDと組み合わせて高次元状態ベクトル \mathbf{x} の動的モデリングをデータ駆動で実現している。同手法はスパースモデリングを利用した画像復元手法の適用例であり、地震動や河川流などの動力学解析に適用できる。畳込み合成辞書を構築する際に、状態ベクトル \mathbf{x} の時空間分布に関する事前知識を活用できる。

2018年、Chenらはニューラルネットワークの学習に物理法則を事前知識として用いることを提案した⁸⁸⁾。その後、この論文は、学習の際の損失関数が偏微分方程式(PDE: Partial-Differential Equation)を満たすように強制し、解を見つけようとする物理情報ニューラルネットワーク(PINN: Physics-Informed Neural Network)として発展し、一つの分野を切り開いた^{89)~91)}。PINNは、与えられたデータセットを支配するあらゆる物理法則を表現できる普遍的な関数近似の新しい手法を提供している。ただし、ニューラルネットワークを利用しているため、ネットワーク構成のための試行錯誤、説明可能性や解釈可能性の課題は依然として残る²¹⁾。

PINNにおいて、高階・高次元の問題に対する系統的・戦略的なネットワーク構成法は未だ確立されていない。画像やボリュームデータ、ポイントクラウドデータなどを扱う際には、効果的なネットワーク構成の構築に配慮が必要で

ある。スパースモデリングにおいても事前知識としての物理情報に関心が寄せられている。2023年になり、IEEE Signal Processing Magazineにおいて二つの号にまたがり物理駆動機械学習の特集が組まれたことは注目に値する⁹²⁾⁹³⁾。筆者(村松)らも文献⁹⁴⁾において光の屈折率と反射率の関係を事前知識として利用した光干渉断層撮像データの復元手法を提案している。

科学的な興味を理由に、学習結果から第一原理を推論し、データからPDEを構築するようなリバースエンジニアリングも望まれる。CPS応用では、センサで取得されたデータの解析、認識、圧縮など、音や画像などに対して培われ発展してきた多くのメディア処理が役立つことは注目に値する。メディア工学にとっても新しい研究課題を見いだせる場となっている。

8. むすび

本稿では、これまでの変遷も含め、近年のメディア工学に関する技術の発展について紹介した。特に生成系AI技術については黎明期を脱し、爆発的な応用展開を迎える過渡期に突入している。一方で、解決すべき社会的な課題も山積し、AI技術がそれらの解決に役立つのか、あるいはそれ自身が新たな課題をもたらすのか注視が必要な状況といえる。いずれにせよメディア工学が学問として社会に貢献できる分野であることに変わりはなく、その動向には引き続き着目していきたい。

(2023年10月10日受付)

〔文 献〕

- 1) I. Goodfellow, Y. Bengio and A. Courville: "Deep Learning", The MIT Press (2016)
- 2) A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, U. Kaiser and I. Polosukhin: "Attention is All You Need", Advances in Neural Information Processing Systems, 2017-December, 6, pp.5999-6009 (2017)
- 3) J. Devlin, M.W. Chang, K. Lee and K. Toutanova: "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding", NAACL HLT 2019-2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies Proceedings of the Conference, 1, 10, pp.4171-4186 (2018)
- 4) T. Eloundou, S. Manning, P. Mishkin and D. Rock: "GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models", 3 (2023)
- 5) A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit and N. Houlsby: "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale", In ICLR 2021 - 9th International Conference on Learning Representations, International Conference on Learning Representations, ICLR (2021)
- 6) Y. Gong, Y.-A. Chung and J. Glass: "AST: Audio Spectrogram Transformer", In Proc. Interspeech 2021, 1, 4, pp.571-575, International Speech Communication Association (2021)
- 7) E. Tjoa and C. Guan: "A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI", IEEE Transactions on Neural Networks and Learning Systems, 32, 11, pp.4793-4813 (2021)
- 8) M. Elad: "Sparse and redundant representations: from theory to applications in signal and image processing", Springer New York (2010)

- 9) 永原正章: "スパースモデリング-基礎から動的システムへの応用", コロナ社, 10 (2017)
- 10) 村松正吾: "フィルタバンク理論に基づく畳込み辞書学習-構造制約を利用した畳込みネットワーク構築-", 信学基礎・境界ソサイエティ Fundamentals Review, 17, 2, pp.116-125 (Oct. 2023)
- 11) K. Gregor and Y. Lecun: "Learning Fast Approximations of Sparse Coding", In Proceedings of the 27th International Conference on International Conference on Machine Learning, 152, pp.399-406 (2010)
- 12) 和田山正, 高遠賢史: "深層展開に基づく信号処理アルゴリズムの設計-収束加速とその理論的解釈-", 信学基礎・境界ソサイエティ Fundamentals Review, 14, 1, pp.60-72 (July 2020)
- 13) V. Monga, Y. Li and Y.C. Eldar: "Algorithm Unrolling: Interpretable, Efficient Deep Learning for Signal and Image Processing", IEEE Signal Processing Magazine, 38, 2, pp.18-44 (Mar. 2021)
- 14) 村松正吾: "画像復元における分析・合成システム", 信学誌, 106, 1, pp.2-9 (Jan. 2023)
- 15) 和田山正: "モデルベース深層学習と深層展開", 森北出版 (June 2023)
- 16) S.L. Brunton and J.N. Kutz: "Data-Driven Science and Engineering: Machine Learning, Dynamical Systems and Control", Cambridge University Press, (May 2022)
- 17) T. Matsubara, A. Ishikawa and T. Yaguchi: "Deep Energy-Based Modeling of Discrete-Time Physics", In NeurIPS (2020)
- 18) 松原崇, 陳 涵, 谷口隆晴: "幾何学的深層学習による力学系のグレーボックスモデル化", 人工知能, 38, 3, pp.308-317 (May 2023)
- 19) 田中佑典: 銦ガウス過程と物理現象のモデル化", 人工知能, 38, 3, pp.318-325 (May 2023)
- 20) 堀江正信, 三目直登: "グラフニューラルネットワークによる物理シミュレーションの機械学習とその最新動向", 人工知能, 38, 3, pp.326-334 (May 2023)
- 21) 白鳥英, 中村耀, 杉原伊織: "Physics-Informed Neural Networkの研究動向と液膜流れの問題への適用例", 人工知能, 38, 3, pp.335-344 (May 2023)
- 22) K. Azzizadenesheli, N. Kovachki, Z. Li, M.L.-Schiaffini, J. Kossaifi and A. Anandkumar: "Neural Operators for Accelerating Scientific Simulations and Design", arXiv (2023)
- 23) Information technology - Digital compression and coding of continuous-tone still images: Requirements and guidelines, ITU-T T.81 - ISO/IEC 10918-1 (June 1994)
- 24) G. Hudson, A. Lger, B. Niss and I. Sebestyn: "Jpeg at 25: Still going strong", IEEE Multi-Media, 24, 2, pp.96-103 (2017)
- 25) Information technology - JPEG 2000 image coding system - Part 1: Core coding system, ITU-T T.800 - ISO/IEC 15444-1 (June 2019)
- 26) Information technology - JPEG 2000 image coding system: High-throughput JPEG 2000, ITU-T T.814 - ISO/IEC 15444-15 (June 2019)
- 27) D. Taubman, A. Naman and R. Mathew: "High throughput block coding in the htj2K compression standard", In 2019 IEEE International Conference on Image Processing (ICIP), pp.1079-1083 (2019)
- 28) Information technology - JPEG XS low-latency lightweight image coding system - Part 1: Core coding system, ISO/IEC 21122-1 (June 2022)
- 29) A. Descampe, T. Richter, T. Ebrahimi, S. Foessel, J. Keinert, T. Bruylants, P. Pellegrin, C. Buyschaert and G. Rouvroy: "Jpeg xs-a new standard for visually lossless low-latency lightweight image coding", Proceedings of the IEEE, 109, 9, pp.1559-1577 (2021)
- 30) Information technology - JPEG XL image coding system - Part 1: Core coding system, ISO/IEC 18181-1 (June 2022)
- 31) J. Alakuijala, R.v. Asselton, S. Boukourt, M.n Bruse, I.-M. Coma, M. Firsching, T. Fischbacher, E. Kliuchnikov, S. Gomez, R. Obryk, K. Potempa, A. Rhatushnyak, J. Sneyers, Z. Szabadka, L. Vandevenne, L. Versari and J. Wassenberg: "JPEG XL next-generation image compression architecture and coding tools", In Andrew G. Tescher and Touradj Ebrahimi, editors, Applications of Digital Image Processing XLII, 11137, p.111370K, International Society for Optics and Photonics, SPIE (2019)
- 32) J. Alakuijala, S. Boukourt, T. Ebrahimi, E. Kliuchnikov, J. Sneyers, E. Upenik, L. Vandevenne, L. Versari and J. Wassenberg: "Benchmarking JPEG XL image compression", In Peter Schelkens and Tomasz Kozacki, editors, Optics, Photonics and Digital

- Technologies for Imaging Applications VI, 11353, p.113530X, International Society for Optics and Photonics, SPIE (2020)
- 33) J. Duda: Asymmetric numeral systems (2009)
- 34) J. Duda: Asymmetric numeral systems: entropy coding combining speed of Huffman coding with compression rate of arithmetic coding (2014)
- 35) T. Ebrahimi, S. Foessel, F. Pereira and P. Schelkens: "Jpeg pleno: Toward an efficient representation of visual reality", *IEEE MultiMedia*, 23, 4, pp.14-20 (2016)
- 36) P. Schelkens, T. Ebrahimi, A. Gilles, P. Gioia, K.-J. Oh, F. Pereira, C. Perra and A.M.G. Pinheiro: "Jpeg pleno: Providing representation interoperability for holographic applications and devices", *ETRI Journal*, 41, 1, pp.93-108 (2019)
- 37) P. Astola, L. Cruz, E.A.B. da Silva, T. Ebrahimi, A.P.G. Freitas, Gilles, K.-J. Oh, C. Pagliari, F. Pereira, C. Perra, S. Perry, A. Pinheiro, P. Schelkens, I. Seidel and I. Tabus: "JPEG Pleno: Standardizing a coding framework and tools for plenoptic imaging modalities", *ITU Journal: ICT Discoveries, Special issue on the Future of Video and Immersive Media*, 3, 1
- 38) G.D.O. Alves, M.B.D. Carvalho, C.L. Pagliari, P.G. Freitas, I. Seidel, M.P. Pereira, C.F.S. Vieira, V. Testoni, F. Pereira and E.A.B.D. Silva: "The jpeg pleno light field coding standard 4d-transform mode: How to design an efficient 4d-native coDec", *IEEE Access*, 8, pp.170807-170829 (2020)
- 39) L.H. Witten, R.M. Neal and J.G. Cleary: "Arithmetic coding for data compression", *Commun. ACM*, 30, 6, p.520-540 (June 1987)
- 40) Verification Model Description for JPEG Pleno Learning-based Point Cloud Coding v3.0, ISO/IEC JTC 1/SC29/WG1 N100566 (July 2023)
- 41) J. Ascenso, P. Akyazi, F. Pereira and T. Ebrahimi: "Learning-based image coding: early solutions reviewing and subjective quality evaluation", In Peter Schelkens and Tomasz Kozacki, editors, *Optics, Photonics and Digital Technologies for Imaging Applications VI, 11353*, p.113530S. International Society for Optics and Photonics, SPIE (2020)
- 42) Report on the JPEG AI Call for Proposals Results, ISO/IEC JTC 1/SC29/WG1 N100250 (Oct. 2022)
- 43) J. Ascenso, E. Alshina and T. Ebrahimi: "The jpeg ai standard: Providing efficient human and machine visual data consumption", *IEEE MultiMedia*, 30, 01, pp.100-111 (Jan. 2023)
- 44) W. He, K. You, Z. Qiao, X. Jia, Z. Zhang, W. Wang, H. Lu, Y. Wang and J. Liao: "Timereplayer: Unlocking the potential of event cameras for video interpolation", In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.17804-17813 (2022)
- 45) S. Wu, K. You, W. He, C. Yang, Y. Tian, Y. Wang, Z. Zhang and J. Liao: "Video interpolation by event-driven anisotropic adjustment of optical flow", In *European Conference on Computer Vision*, pp.267-283. Springer (2022)
- 46) Z. Yu, Y. Zhang, D. Liu, D. Zou, X. Chen, Y. Liu and J.S. Ren: "Training weakly supervised video frame interpolation with events", In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp.14589-14598 (2021)
- 47) T. Kim, Y. Chae, H.-K. Jang and K.-J. Yoon: "Event-based video frame interpolation with cross-modal asymmetric bidirectional motion fields", In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.18032-18042 (2023)
- 48) S. Lin, J. Zhang, J. Pan, Z. Jiang, D. Zou, Y. Wang, J. Chen and J. Ren: "Learning event-driven video deblurring and interpolation", In *Computer Vision-ECCV 2020: 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part VIII 16*, pp.695-710, Springer (2020)
- 49) S. Tulyakov, D. Gehrig, S. Georgoulis, J. Erbach, M. Gehrig, Y. Li and D. Scaramuzza: "Time lens: Event-based video frame interpolation", In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp.16155-16164 (2021)
- 50) S. Tulyakov, A. Bochicchio, D. Gehrig, S. Georgoulis, Y. Li and D. Scaramuzza: "Time lens++: Event-based frame interpolation with parametric non-linear flow and multi-scale fusion", In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.17755-17764 (2022)
- 51) V. Lempitsky, A. Vedaldi and . Ulyanov: "Deep image prior", In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.9446-9454. IEEE (June 2018)
- 52) A. Krull, T.-O. Buchholz and F. Jug: "Noise2void-learning denoising from single noisy images", In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp.2129-2137 (2019)
- 53) Z. Wang, J. Liu, G. Li and H. Han: "Blind2Unblind: Self-supervised image denoising with visible blind spots", In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.2027-2036 (2022)
- 54) O. Sidorov and J.Y. Hardeberg: "Deep hyperspectral prior: Single-image denoising, inpainting, super-resolution", In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops* (2019)
- 55) R. Imamura, T. Itasaka and M. Okuda: "Zero-shot hyperspectral image denoising with separable image prior", In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops* (2019)
- 56) T. Itasaka and M. Okuda: "Zero-shot hyperspectral image denoising with self-completion with patterned masks", In 2023 IEEE International Conference on Image Processing (ICIP), IEEE, pp.1340-1344 (2023)
- 57) A. Radford, J.W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger and I. Sutskever: "Learning Transferable Visual Models from Natural Language Supervision", pp. 8748-8763. PMLR (July 2021)
- 58) H. Kataoka, K. Okayasu, A. Matsumoto, E. Yamagata, R. Yamada, N. Inoue, A. Nakamura and Y. Satoh: "Pre-Training Without Natural Images", *International Journal of Computer Vision*, 130, 4, pp.990-1007 (Apr. 2022)
- 59) R. Shinoda, R. Hayamizu, K. Nakashima, N. Inoue, R. Yokota and H. Kataoka: "SegRCDB: Semantic Segmentation via Formula-Driven Supervised Learning", In *ICCV* (2023)
- 60) I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio: "Generative adversarial networks", *Proc. Conf. Neural Information Processing Systems*, pp.2672-2680 (Dec. 2014)
- 61) A. Ramesh, P. Dhariwal, A. Nichol, C. Chu and M. Chen: "Hierarchical Text-Conditional Image Generation with CLIP Latents" (Apr. 2022)
- 62) R. Rombach, A. Blattmann, D. Lorenz, P. Esser and B. Ommer: "High-Resolution Image Synthesis with Latent Diffusion Models", In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2022-June*, pp.10674-10685. IEEE Computer Society (2022)
- 63) A.K. Jain, P. Flynn and A.A. Ross: "Handbook of Biometrics", Springer US (2008)
- 64) S.Z. Li and A.K. Jain: "Handbook of Face Recognition", Springer (2011)
- 65) I. Goodfellow, J. Shlens and C. Szegedy: "Explaining and harnessing adversarial examples", *Proc. Int'l Conf. Learning Representations*, pp.1-11 (May 2015)
- 66) J. Su, D.V. Vargas and K. Sakurai: "One pixel attack for fooling deep neural networks", *IEEE Trans. Evolutionary Computation*, 23, 5, pp.828-841 (Oct. 2019)
- 67) M. Sharif, L. Bhagavatula, S. and Bauer and M.K. Reiter: "Accessorize to a crime: Real and stealthy attacks on state-of-the-art face recognition", *Proc. ACM SIGSAC Conf. Computer and Communications Security*, pp.1528-1540 (Oct. 2016)
- 68) X. Yang, Y. Dong, T. Pang, H. Su, J. Zhu, Y. Chen and H. Xue: "Towards face encryption by generating adversarial identity masks", *Proc. Int'l Conf. Computer Vision*, pp.3897-3907 (Mar. 2021)
- 69) S. Shan, E. Wenger, J. Zhang, H. Li, H. Zheng and B. Zhao: "Fawkes: Protecting privacy against unauthorized deep learning models", *Proc. USENIX Security Symp.*, pp.1589-1604 (Aug. 2020)

70) G. Hanawa, K. Ito and T. Aoki: "Face image de-identification based on feature embedding for privacy protection", Proc. Int'l Conf. Biometrics Special Interest Group, pp.1-10 (Sep. 2023)

71) A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies and M. Niener: "FaceForensics++: Learning to detect manipulated facial images", Proc. Int'l Conf. Computer Vision, pp.1-11 (Oct. 2019)

72) S. Marcel, M.S. Nixon, J. Fierrez and N. Evans: "Handbook of Biometric Anti-Spoofing", Springer (2019)

73) Y. Liu, A. Jourabloo and X. Liu: "Learning deep models for face anti-spoofing: Binary or auxiliary supervision", Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition, pp.389-398 (June 2018)

74) Z. Yu, C. Zhao, Z. Wang, Y. Qin, Z. Su, X. Li, F. Zhou and G. Zhao: "Searching central difference convolutional networks for face anti-spoofing", Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition, pp.5295-5305 (June 2020)

75) K. Watanabe, K. Ito and T. Aoki: "Spoofing attack detection in face recognition system using Vision Transformer with patch-wise data augmentation", Proc. Asia-Pacific Signal and Information Processing Association Annual Summit and Conf., pp.1558-1562 (Nov. 2022)

76) B. Mildenhall, P.P. Srinivasan, M. Tancik, J.T. Barron, R. Ramamoorthi and R. Ng: "Nerf: Representing scenes as neural radiance fields for view synthesis", In ECCV (2020)

77) Q. Xu, Z. Xu, J. Philip, S. Bi, Z. Shu, K. Sunkavalli and U. Neumann: "Point-nerf: Point-based neural radiance fields", 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp.5428-5438 (2022)

78) A. Pumarola, E. Corona, G.P.-Moll and F.M.-Noguer: "D-nerf: Neural radiance fields for dynamic scenes", 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp.10313-10322 (2020)

79) Y.-J. Yuan, Y.-T. Sun, Y.-K. Lai, Y. Ma, R. Jia and L. Gao: "Nerf-editing: Geometry editing of neural radiance fields", 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp.18332-18343 (2022)

80) Luma ai, <http://lumalabs.ai/> (Accessed: 2023-09-22)

81) Getting started with luma ai: Guided capture mode, <https://www.youtube.com/watch?v=JXba3ZIGnI> (Accessed: 2023-09-22)

82) 小西稜也, 竹下優輝, 張潮, 東海彰吾: "AR機能を利用した携帯端末群による多視点撮影支援システム", 映像学技報, 46, 4, pp.75-78 (Feb. 2022)

83) J.H. Tu, C.W. Rowley, D.M. Luchtenburg, S.L. Brunton, J.N. Kutz: "On dynamic mode decomposition: Theory and applications", Journal of Computational Dynamics, 1, 2, pp.391-421 (Dec. 2014)

84) S.E. Otto and C.W. Rowley: "Linearly Recurrent Autoencoder Networks for Learning Dynamics", 18, 1, pp.558-593, (Mar. 2019), <https://doi.org/10.1137/18M1177846>

85) Y. Liu, C. Ponce, S.L. Brunton and J.N. Kutz: "Multiresolution convolutional autoencoders", Journal of Computational Physics, 474, p.111801 (Feb. 2023)

86) K.P. Murphy: "Probabilistic Machine Learning", The MIT Press (Mar. 2022)

87) E. Kobayashi, H. Yasuda, K. Hayasaka, Y. Otake, S. Ono and S. Muramatsu: "Multi-Resolution Convolutional Dictionary Learning for Riverbed Dynamics Modeling", IEEE, pp.1-5 (May 2023)

88) R.T.Q. Chen, Y. Rubanova, J. Bettencourt and D. Duvenaud: "Neural Ordinary Differential Equations", In Proceedings of the 32nd International Conference on Neural Information Processing Systems (NIPS'18), pp.6572-6583 (2018)

89) M. Raissi, P. Perdikaris and G.E. Karniadakis: "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations", Journal of Computational Physics, 378, pp.686-707 (Feb. 2019)

90) G.E. Karniadakis, I.G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang and L. Yang: "Physics-informed machine learning", Nature Reviews Physics 2021 3:6, 3, 6, pp.422-440 (May 2021)

91) A.S. Krishnapriyan, A. Gholami, S. Zhe, R.M. Kirby and M.W. Mahoney: "Characterizing possible failure modes in physics-informed neural networks", Advances in Neural Information Processing Systems, 34, pp.26548-26560 (Dec. 2021)

92) B. Wen, S. Ravishankar, Z. Zhao, . Giryes and J.C. Ye: "Physics-Driven Machine Learning for Computational Imaging [From the Guest Editor]", IEEE Signal Processing Magazine, 40, 1, pp.28-30 (Jan. 2023)

93) B. Wen, S. Ravishankar, Z. Zhao, R. Giryes and J.C. Ye: "Physics-Driven Machine Learning for Computational Imaging: Part 2 [From the Guest Editors]", IEEE Signal Processing Magazine, 40, 2, pp.13-15 (Mar. 2023)

94) R. Kobayashi, G. Fujii, Y. Yoshida, T. Ota, F. Nin, H. Hibino, S. Choi, S. Ono and S. Muramatsu: "Sparsity-Aware OCT Volumetric Data Restoration Using Optical Synthesis Model", IEEE Transactions on Computational Imaging, 8, pp.505-520 (2022)



村松 正吾 1995年, 都立大学大学院修士課程修了。現在, 新潟大学工学部教授。多次元信号処理, 画像復元の研究に従事。博士(工学)。IEEE, IEICE各会員。正会員。



渡邊 修 2004年, 都立大学大学院博士課程修了。現在, 拓殖大学工学部教授。画像符号化の研究に従事。ISO/IEC JTC 1/SC 29/WG 1 (JPEG) エキスパート。情報規格調査会JPEG小委員会主査。博士(工学)。IEEE, IEICE各会員。正会員。



奥田 正浩 1998年, 慶應義塾大学大学院博士課程修了。現在, 同志社大学理工学部教授。多次元信号処理, 画像復元の研究に従事。博士(工学)。IEEE, IEICE各会員。正会員。



青木 義満 2001年, 早稲田大学博士課程修了。現在, 慶應義塾大学理工学部教授。コンピュータビジョン, パターン認識の研究に従事。博士(工学)。画像センシング技術研究会会長, 日本顔学会理事。IEEE, IEICE, IPSJ, SICE各会員。正会員。



伊藤 康一 2005年, 東北大学博士課程修了。現在, 東北大学情報科学研究科准教授。画像処理, バイオメトリクス, 医用画像処理などの研究に従事。博士(情報科学)。IEEE, IEICE, IPSJ各会員。正会員。



東海 彰吾 1996年, 名古屋大学博士課程修了。現在, 福井大学工学系部門教授。多視点撮影画像・映像処理の研究に従事。博士(工学)。IEICE, IPSJ, IEEEJ, 芸術科学会各会員。正会員。