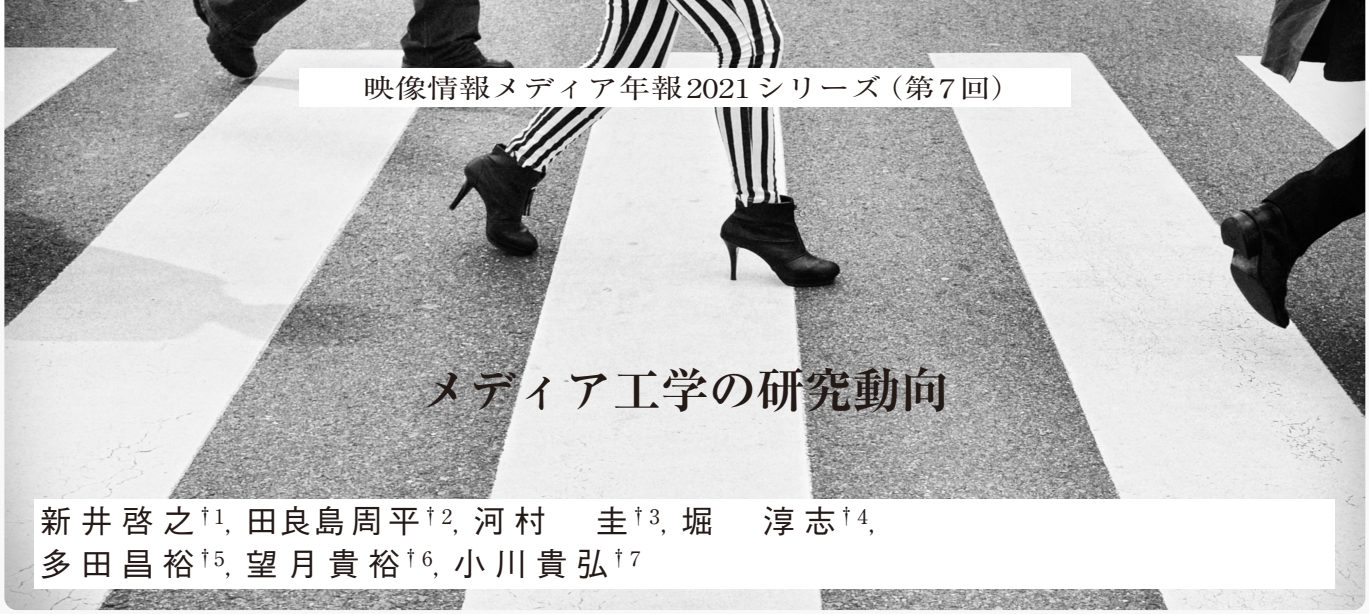

メディア工学の研究動向

新井啓之[†1]，田良島周平[†2]，河村　圭[†3]，堀　淳志[†4]，
多田昌裕[†5]，望月貴裕[†6]，小川貴弘[†7]

1．まえがき

メディア工学に関わる近年の研究においては，深層学習（DeepLearning）技術を始めとする革新的なAI技術の登場が先導する形で，その社会実装が進みつつある．本稿では，こういった動きの中から，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．

AI技術については，画像・映像を対象とし，俯瞰した視点からこれまでの研究の流れを踏まえつつ，今後の研究の方向性，可能性について分析する．特に，これまで成功を収めてきたCNN（Convolutional Neural Network：畳込みニューラルネットワーク）を超えていこうとするさまざまな試みについて紹介する．

AR（Augmented Reality）およびVR（Virtual Reality）技術については，これらを利用するためのデバイス等の環境が揃いつつあり，今後一層の実用化が期待される．今後の普及においては，データの形式，符号化等の標準化が重要となるが，一言でAR/VRと言っても，想定するメディアの形態や，想定されるアプリケーションは多様であり，さまざまな観点からの標準化が進められている．またAR/VRの可能性を拓くさまざまなシステム，応用も続々と提案されてきている．本稿では，AR/VRに関する標準化とシステム・応用の二つの観点から最近の動向を紹介していく．

自動運転技術はメディア工学技術に大きな期待が寄せられている応用分野である．本稿では自動運転技術の実用化に向けた流れを踏まえつつ，運転者状態のモニタリング，自動車と歩行者等の周囲の交通参加とのコミュニケーションを軸として最近の動向を紹介する．

2021年夏には東京オリンピック・パラリンピックが開催されたが，開会式・閉会式をはじめ，さまざまな競技のテレビ視聴，会場等での視聴においてさまざまなメディア工学技術が応用された．COVID19の問題のために残念ながら日の目を見なかったものも含め，さまざまな技術が確立されたが，その中から本稿では，ドローンを使ったパフォーマンス，プロジェクションマッピング，超高臨場感通信技術を取り上げ，その概要を紹介する．

最後に，AI技術の進展により，さまざまな場面において実用化に向けた取り組みが行われつつあるが，その中で見えてきた共通的な課題，また，利用場面に固有な問題について，実際の取り組みの中から見えてきた知見，考え方を紹介する．

2．画像・映像AI技術の動向

画像・映像AI研究コミュニティの規模は増加の一途を辿っており[1)2)]，本学会の年次・冬季大会でも関連技術の発表は恒常的に多くなされている．メディア工学研究会では，特に画像・映像AIに関連のある今後10年のグランドチャレンジとして

(1) 巨大なメディアデータの認識理解
(2) メディア生成とその真正性判定
(3) 自律的に進化するメディアAIの実現
(4) 教師データの定義・生成が困難な課題への対応
(5) 技術進歩と社会への適合

を掲げ，学会誌[3)]，70周年記念大会[4)]，国際会議[5)]といった場での情報発信を行ってきた．また，AI実運用の観点で近年急速な発展を見せるMLOps（機械学習技術と運用の連携）に着目したイベント[6)]を2021年冬季大会で実施するなど，拡大・変化を続けるメディア工学関連技術トピックの情報共有を行う場の創出にも取り組んでいる．

画像・映像AIの最近の研究動向について，すべてをこの場で網羅することは叶わない．以下では，特に直近1年で注目を集めた，畳込みニューラルネットワーク（CNN）に

†1 日本工業大学 先進工学部
†2 NTTコミュニケーションズ
†3 株式会社KDDI総合研究所
†4 三菱電機株式会社 情報技術総合研究所
†5 近畿大学 理工学部
†6 NHK 放送技術研究所
†7 北海道大学 大学院情報科学研究院
"Recent Researches on Media Engineering" by Hiroyuki Arai (Nippon Institute of Technology, Saitama), Shuhei Tarashima (NTT Communications, Tokyo), Kei Kawamura (KDDI Research, Inc., Saitama), Atsushi Hori (Mitsubishi Electric Corporation), Masahiro Tada (KINDAI University, Osaka), Takahiro Mochizuki (Science & Technology Research Laboratories, NHK, Tokyo) and Takahiro Ogawa (Faculty of Information Science and Technology, Hokkaido University, Sapporo)

映像情報メディア学会誌 Vol. 76, No. 1, pp. 78～85（2022）

依存しない画像・映像AIモデルの話題を紹介したい．

2.1 CNN非依存の画像・映像AIモデルの萌芽

AlexNet[7]以降，CNNは画像・映像AIのデファクトスタンダードなアプローチとなった．CNNには演算の局所性やシフト不変性などの帰納バイアスが構造として組み込まれており，ILSVRC-2012 Challenge[8]規模（約130万枚の画像を1000クラスに分類）のタスクではモデルの汎化性能向上に寄与しているとされる．一方で自然言語処理分野では，任意の集合データに対し適用可能なTransformer[9]（Self-Attentionを要素モジュールとするEncoder-Decoder型ニューラルネットワーク）をベースとする巨大なモデルがさまざま提案され[10][11]，大規模データセットを用いたそれらの教師なし学習の成功事例が相次いで報告されている．これらの背景をふまえ最近では，画像・映像AIモデルの畳込み処理をSelf-Attentionベースのモジュールに置き換え，構造由来の帰納バイアスを緩和させようとする試みが活発である．

例えば，SASA[12]やHaloNet[13]では，ResNet[14]内の畳込み処理をLocal Self-Attentionで置換することで入力に対し動的なフィルタが実現され，画像分類や物体検出タスクに成功裏に適用されている．Axial-DeepLab[15]では，Self-Attentionを画像の縦方向と横方向とに連続して適用するモジュールを畳込み処理の代わりに導入することで，計算量を抑えつつ画像全体を考慮した特徴抽出を行い，画像分類やパノプティックセグメンテーションの精度向上につなげている．Vision Transformer（ViT）[16]では，入力画像を固定サイズのパッチに分割し，各パッチから抽出される埋込みベクトルをトークンとみなしてTransformer Encoderを適用することで画像分類する方法が提案され，特にImageNet-21k[17]やJFT-300M[18]といった大規模データセットを用いて学習した場合に，ピーク性能でCNNを上回ることが示されている．

ViTは，その潜在的なピーク性能の高さがモデルのシンプルさと相まって特に大きな注目を集めており，発表から1年と経たずにすでに数多くの発展研究が存在する．例えば，ViTの精緻な理解を試みる研究として，位置情報埋込み[19]や敵対的攻撃に対する挙動[20]を分析する研究が存在する．また，教師なし/自己教師あり学習について論じている研究[21][22]では，得られたViTモデルの画像検索や弱教師あり学習へ転用可能性が示唆されている．さらに，ViTへのマルチスケールな特徴抽出機構導入に取り組んだ研究群として，SwinTransformer[23]，PVT[24]，ViL[25]，MViT[26]，PiT[27]，DPT[28]，ViViT[29]，HVT[30]，CrossViT[31]がほぼ同時期に提案されており，中規模データセット[8]におけるViTの有効性が確認されているほか，物体検出[23]～[25][27]，セグメンテーション[23]～[25][28]，深度推定[28]，映像分類[26][29]への適用もすでに実現されている．

現時点では，CNN非依存のこれらのアプローチに，CNNで培われてきたあらゆるトリック（例えば，逆畳込み等）の代替手段が確立されているわけではない．また，物体検出等のダウンストリームタスクを解くためのヘッドモジュールは，現状CNNをバックボーンとする技術[32][33]からそのまま持ち込まれているのみである．ViTをはじめとするCNN非依存の画像・映像AIモデルには，依然多くの改良の余地が残されていると考えられる．今後も急速な発展が見込まれ，その影響は異なるモダリティを含むより広い領域へと波及していくのではないかと予想する．

3．AR/VR技術の動向

3.1 AR/VR技術の標準化動向

AR/VRに代表されるImmersive mediaは，自由度という観点から分類されている[34]．ヘッドマウントディスプレイなどを想定し，頭の向きを自由に動かしてさまざまな方向を視聴できる場合は3DoF（3自由度）と呼ぶ．次に，椅子に座っていて，頭の位置を少しだけ動かし，何かの影に隠れているものも見えるような視聴体験を3DoF+と呼ぶ（運動視差）．さらに，身体の位置も自由に動かして，あらゆる視点から視聴可能な体験を6DoFと呼ぶ．それぞれに対応するメディアフォーマットとして，360°映像/VR映像，ステレオ映像や奥行き情報のある映像，密な点群情報が挙げられる．以下では，それぞれのImmersive mediaに対応する標準化動向を紹介する．

VR映像を含む2次元映像に対する最新の符号化方式として，VVC（Versatile Video Coding）が2020年10月に標準化された．これまでのHEVC（High Efficiency Video Coding）と比較して，主観画質で約2倍の符号化効率を実現している[35]．また，VR映像に特化した符号化ツールや領域分割方式が内包されており，これまで以上に使い勝手の良い規格となっている．

複数カメラで取得し，奥行き情報のある映像に対しては，MIV（MPEG Immersive Video）符号化が標準化されている．複数カメラ間の冗長な領域をプレ処理で除去し，アトラスと呼ばれる重複のないテクスチャ画像，奥行き画像，占有画像を生成する．これらを通常の映像符号化方式により符号化する．実際に標準化されているのは，アトラスに関するメタデータのみである．プレ処理方法には自由度があり，その手順や性能はエンコーダ実装に依存する．また，仮想視点における視点合成も仮想参照レンダラーとして検討が進められており，合成された仮想視点の画質が評価されている[36]．

V-PCC（Video-based Point Cloud Compression）は，点群を複数の適当な平面に射影し，1枚の大きなテクスチャ画像とデプス画像に統合する．これらを通常の映像符号化方式により符号化する[37]．具体的には，オブジェクトを包含する立方体を想定し，各面に点群を射影すれば大部分の点群を平面のテクスチャ画像とデプス画像に変換できる．実際に標準化されているのは，テクスチャ画像とデプス画

像から点群を再構成する部分，すなわち復号方式のみである．変換方法には自由度があり，その手順や性能はエンコーダ実装に依存する．モバイル端末でもリアルタイム復号・視聴が可能であり，実装例がオープンソースソフトウェアとして公開されている[38]．MIVとV-PCCはいずれも映像符号化方式を圧縮技術として利用し，そのプレ処理・ポスト処理に特徴があるという共通点がある．さらに，プレ処理により得られた中間形式（映像符号化の入力形式）も類似していることから，MIVとV-PCCで共通部分はV3C（Visual Volumetric Video-based Coding）として標準化されている．

G-PCC（Geometry-based Point Cloud Compression）は点群の座標（Geometry）を最初に符号化し，次にその座標における属性（Attribute）を符号化する[39]．まず，座標の符号化方法には2種類ある．Octree codingでは，立方体を再帰的に八分木で分割し，小立方体内部の点の個数や分布の偏りを用いて存在確率を推定し，実際の点の有無を推定確率に基づいて算術符号化する．Geometry predictive tree codingでは，低遅延符号化を想定し，入力された点群から任意の木構造を構築し，符号化済みのノードから当該ノードの位置情報を予測して残差を符号化する．次に，属性の符号化は，複数の点を対象とした変換が3種類あり，領域適応階層変換（Region Adaptive Hierarchical Transform; RAHT）と補間的階層最近傍予測変換（Predicting Transform），リフティング付き補間的階層最近傍予測変換（Lifting Transform）である．G-PCCは標準化の最終段階にあり，規格文書の洗練化が進められている（2022年標準化予定）．

標準化における寄書（入力文書）は基本的に非公開となっているが，一部の文書は公開されている[40]．また，必要なソフトウェアも合わせて公開されている．関心のある読者はこれらを確認されたい．

3.2　AR/VRシステム・応用

AR/VRのシステム適用，応用方法については，近年目覚ましく発達している状況である．本節ではシミュレーション，広告，ディジタルアートにかかわる適用事例について紹介する．

まず，シミュレーションへの適用事例として，リクルートによる「AR Room Simulator」[41]を紹介する．本アプリはARによる部屋の模様替えを目的としている．近年スマートフォンは最新のデバイスをサポートすることで継続的に性能向上が進んでいる．特に本アプリではLiDAR（Light Detection and Ranging）を活用したリアリティの高いAR表示により，高臨場感を体験できることを特徴としている．次に，ネット広告適用事例を紹介する．広告媒体としてのネットサイトを拡張現実やバーチャル空間に応用したものが，VR広告となる．近年，各種イベントの開催を見据えたVR広告の実証実験が行われ始めている．一例としてNTT，電通により，バーチャル空間を通じた体験型広告[42]を提供する取り組みが行われている．本取り組みでは，新型コロナウイルスの感染拡大による人々の移動が制限される状況の下，バーチャル空間を活用したイベントの開催や参加へのニーズが高まっていることが背景として挙げられており，その場にいるような体験や臨場感など，VRによる公告が単純な現実空間の置き換えだけでないことを特徴としている．最後に，ディジタルアートについてのVR応用事例として，Misoshitaらによるクリプトアートプロジェクト「Metaani」[43]を紹介する．本プロジェクトでは，利用者がNFT（Non-Fungible Token）を活用してメタバース（インターネット上の3D空間）内にただ1種のみしか存在しないアバターを発行し，VR空間を楽しめる．また，3Dモデルには汎用フォーマットを採用していることから，他のメタバースプロジェクトへ持ち出して楽しむことが可能となっている．さらにデータフォーマットとしてIPFS（InterPlanetary File System）[44]を採用することで耐障害性・負荷分散・耐検閲性を有している．本プロジェクトでは利用者はディジタル通貨を利用して決済することが可能となっている．また，AR応用事例として，カナダNextech AR Solution社によるNFTプラットフォーム構築計画[45]を紹介する．今回発表されたプラットフォームは，段階的な展開を計画している．最初の段階において，利用者がサードパーティのNFTマーケットプレイスで購入したARヒューマンホログラムを楽しむARアプリの提供を行う．次の段階において，同プラットフォームでNFTの作成と，ヒューマンホログラムの購入・販売を可能とする．

以上，最近のシステム・応用動向をいくつか実例を挙げた．AR/VR技術は，スマートフォンなどのデバイスの進化によるコンテンツ臨場感の向上，またコロナ禍の影響による新たな広告形態への適用，さらにNFTなどブロックチェーン技術との組み合わせによる活用方法が生み出されていることが興味深い．今後さらなる新しい分野への適用が進むと期待される．

4．自動運転を支えるメディア技術

自動走行システムは，機械が担う運転タスクの程度によってSAEレベル0から5までに分類されており[46]，わが国では速度制御・車間距離維持制御や操舵制御を車両側で代替するSAEレベル1，2相当の部分自動運転車両の普及が進んでいる．2020年4月には道路交通法が改正され，自動運行装置を使用して車両を運行することも運転行為に含まれる旨の規定が追加された[47]．これを受け，2021年3月には限定された条件下ではあるものの，従来人が担ってきた周辺監視も車両が担うSAEレベル3相当の車両が発売されるなど，自動走行システムを搭載した車両が普及した交通社会の到来が間近に迫っている．内閣府の戦略的イノベーション創造プログラムでは，2025年を目処にSAEレベル4

の自動運転車両の市場化を可能にすることが目標に掲げられており[48]，社会実装に向けた自動運転車両の実証実験が各地で実施されている[49]．

国土交通省の自動運転車の安全技術ガイドラインでは，個々の運転自動化システムの性能および用途に応じた運行設計領域（ODD: Operational Design Domain）を設定して走行環境や運用方法を制限し，合理的に予見される防止可能な事故が生じないことを確保するとされている[50]．併せて道路交通法では，自動運行装置に係る使用条件を満たさない場合（ODDから外れる場合）には，運転者は直ちにそのことを認知しシステムに代わって運転可能な状態にあることを，自動走行システム使用の条件としている．

以上の背景のもと，自動走行システム使用中のドライバが，運転を代替できる状態にあるか把握するためのビジョンベースのドライバモニタリング技術の研究開発が盛んに行われている[51)〜53)]．これらビジョンベースのドライバモニタリング技術は，ドライバの覚醒状態の低下に着目し，単位時間当たりの閉眼時間割合（PERCLOS）や瞬目回数などを用いてドライバの眠気の推定を試みるものである．一方で，多田らはドライビングシミュレータを用いた検討の結果，自動走行システム使用時には，ドライバが眠気の症状を訴えていない場合であっても，わき見の増加や周辺への注意が低下し注意散漫状態を誘発するリスクがあることを指摘している[54]．近年は，ドライバが疲労や携帯電話の使用など運転以外のタスクをしていることにより運転に集中できていない状態（Driver Distraction）をビジョンベースで検知する仕組み検討のためのデータベースが整備されるなど[55]，自動走行システムを安全に運用するためのドライバモニタリング技術の開発は今後も盛んになされていくと考えられる．

運転では，一般に，自車両周辺の他の交通参加者（車両，自転車，歩行者）とヘッドライトやウィンカー，アイコンタクト，ジェスチャなどを用いつつコミュニケーションすることで円滑な運行が実現されている．一方で，自動走行システムではアイコンタクトやジェスチャなど人が用いるコミュニケーション手段を利用することができない．自動走行システムの実証実験では，1,740 kmの走行距離に対し，自動走行が継続できず人による介入を必要とした事案が1,046回発生したことが報告されている[49]．このうち，路上駐車の回避（17%）に続いて多かったのが他車両とのすれ違い（7%），歩行者や自転車の回避（7%）であり，自動走行システムの普及を図る上で他の交通参加者とコミュニケーションする手段確立の重要性を示唆する結果となっている．そこで近年は自車両内部で利用するためのHMI（Human-Machine Interface）だけでなく，車外に情報呈示するためのexternal Human-Machine Interfaces（eHMI）の研究開発が国内外で盛んになされている[56]．eHMIで用いられる車外への情報呈示方法としては，信号を模したLEDによる情報呈示や，文字メッセージをディスプレイに呈示するもの，音による情報呈示などが提案されているほか[57]，LEDヘッドライトを用いて道路上に情報を投影する仕組み[12]なども開発されており，今後ますますの研究開発が期待される分野と言える．

5．東京2020におけるメディア工学技術

5.1　ドローンによるパフォーマンス

開会式では，数々の魅力的な演出のひとつとして，大量のドローンで形成されたエンブレムや地球が空中に浮かび上がるパフォーマンスが大きな話題を呼んだ．他のイベントなどでも同様の技術が披露されている，飛行台数は多くとも500台程度である．今回の開会式では1824台のドローンを使用しており，より高度な仕組みが必要となる．

使用されたドローンは，平昌冬季五輪でも活躍したIntelの「Shooting Star」であり，重さ330グラム，ローターの直径が15センチの小型クアッドコプターである．LEDライトの組み合わせで40億色以上の光を表現することができ，1台のPCで数千台のドローン群をコントロール可能である．Shooting Starには2タイプのドローンがあり，開会式で使われたのは，より高性能な「Premi-um Drone」である[58]．

大量のドローンをどのように運用するのか．Intel社が公開している概要を述べる[59]．社内のクリエイターチームが専用ツールで作成したアニメーション動画を取り込んで，ドローンの飛行をシステム上でシミュレーションする．Intel独自の技術により，空中で表現できるデザインの制約はほぼない．また，屋外でのパフォーマンスには天候や風速などの気象条件が大きく影響するため，シミュレーションの際に，気象庁の予報，過去の気象データ，および同社独自に予測した当日の天候を，パラメータとして採用した．飛行ルートを決める際には，離陸時に発生する風の影響も考慮している．さらに，ドローンにはGPSとAIが搭載されており，突発的な強風などで飛行ルートから外れた場合の自動修正や，位置補正技術による機体間距離の維持など，さまざまなアクシデントの回避が可能である．

5.2　プロジェクションマッピング

開閉会式や陸上競技などにおいて，芸術的かつ近未来的なプロジェクションマッピングが高い注目を浴びた．使用されたプロジェクタは，パナソニックのハイエンド機種「PT-RQ50 KJ」である[60]．世界初の最高輝度5万ルーメンを実現した4K対応プロジェクタであり，リオ五輪で使用した同社のプロジェクタと比較して，投射面積が1.7倍，照度が1.2倍に進化した．また，冷却機構の効率化によるコンパクト化により，設置台数もリオ五輪のおよそ3/4（約60台）に削減するこができた[61]．さらに，「2波長青色レーザ＋赤色レーザ＋蛍光体」の光源システムの採用が，青色や赤色の発色性能を高め，印象的な演出の実現に大きく貢献した．開会式において，白い衣装をまとったダンサーが

赤いロープでつながりながら，赤い糸の映像が投影されたフィールドを歩き回る場面があったが，赤色の発色性が高いPT-RQ50 KJの特徴が活かされていたといえる[62]．

5.3 超高臨場感通信技術

NTTは，遠隔観戦の実現に向けて開発を続けてきた超高臨場感通信技術「Kirari！（キラリ）」[63]を，五輪期間中に特設会場（日本科学未来館）で展示した[64]（新型コロナの影響により，一般への公開は中止）．

Kirari！は，現場においてカメラ，マイク，センサによって抽出された情報を，ネットワークを介してメディア制御，メディア処理，同期して視聴会場に伝送する技術である．メディア制御は，カメラで撮像された人物とセンサによって得られた位置情報，および照明情報を関連付ける空間情報と，人物の配信時刻を絶対時刻で制御するための時間制御から構成される．メディア制御と並行して，人物領域の抽出，波面合成音響技術による音声の高臨場化，複数カメラ映像のワイド合成および3Dフィールド再構成，および符号化といったメディア処理を施す．そして，メディア伝送規格MMT（MPEG Media Transport）に独自の拡張を行ったAdvanced MMTにて同期伝送する．絶対時間で同期できることから，世界中の任意の地点に同時刻に伝送可能である．今後，新たなビューイングスタイルの確立に向けて，さまざまなイベントでの利用が期待される．

6．メディア工学技術の展開

近年のマルチメディアAI技術に関する関心の高まりから，実データを対象とした課題，具体的にさまざまな分野（医療や社会インフラ維持管理等）に存在する課題を解決する試みが活発に進められている[65)～68)]．しかしながら，実データを対象とした場合，近年の研究で用いられているオープンデータセットと性質が大きく異なることから，さまざまな問題に直面する．例えば，大量データの収集が困難であること，実利用に向けた解釈性の向上が必要であることが挙げられる．また，適用対象のそれぞれの分野に存在する固有の制約等を乗り越える方法論も検討しなければならない．一方で，これらの問題の解決を発端として，マルチメディアAI技術に関する新たな研究課題が生み出されることも期待されている．以下では，それらの具体的な研究動向について説明する．

まず，オープンデータセットの共有を前提としたこれまでの基礎的研究と実データを対象とした応用的研究では，収集可能なデータ量に大きな隔たりが存在する．特に，医療，社会インフラ維持管理等の分野では，個人情報や秘匿性の観点から，機関を横断したデータの共有が容易ではなく，充分な量の学習データを準備することが困難な場合が多い．したがって，限られたデータ量からの学習を可能にするための方法論の実現が必要となる．これまでに，半教師あり学習等に基づいて，この問題を解決する方法[69)]に加え，近年では，データ量の不足を，同時に利用するモダリティの数を増やすことで解決する試み[70)]も存在する．一方で，データの共有が困難となる根本的な問題を解決する試みも実施されている．具体的に，近年注目されているデータ蒸留のアルゴリズムを導入することで，学習画像群を1枚の特性が大きく異なる画像に圧縮し，この画像のみから深層学習モデルの構築を可能とする研究が提案されている[71)]．データ蒸留前の医用画像等には個人情報が含まれているのに対し，蒸留後の画像には個人情報が含まれないことから，複数の機関を横断したデータ共有の障壁を取り除くことが可能な画期的な技術として注目されている[72)]．

以上の取り組みにより，実データを対象とした高精度なAI技術の実現が可能になっている．一方，実際の現場でこれらの技術を利用する際，極めて高い精度が実現されたとしても，それらの結果に対して最終判断を行うのは人間である．したがって，推定結果と人間の判断が異なる場合に備えて，推定結果に対する解釈性を与えるような情報を提示できることが望ましい．このような問題に対し，近年，推定結果の解釈性向上を目指したExplainable AIに関する研究が活発に行われている．具体的に，医用画像分類において，画像中に撮像されている各領域がどの程度分類に貢献しているのかを可視化する手法[73)]が存在し，解釈性の向上が図られている．また，社会インフラ維持管理においては，点検時に変状を撮影した画像と同時に入力されるテキスト情報を協調的に用いることで，推定に貢献する領域を明らかにするとともに，それらの領域に基づいて最終的な推定精度の向上を実現する取り組みが行われている[74)]．医療の分野と同様に，社会インフラ維持管理の分野においても，マルチメディアAI技術による推定結果の解釈性向上は必要不可欠である．

以上に示した問題解決の取り組みは，複数の分野において共通するものであり，多くの研究者によって研究が進められている．一方で，各分野に存在する固有の制約を乗り越えるための研究については，特に異分野連携研究として実施されている場合が多い．以下では，社会インフラ維持管理と情報科学の連携による取り組みについて紹介する．構造物の維持管理業務では，点検箇所をさまざまな場所から撮像する．これは，変状に関する情報に加えて，その原因となる情報や構造物全体に対する位置等を正確に把握するためである．このように，各点検箇所で撮像される複数枚の画像の中には変状が撮像されていないものが多い．したがって，それらすべての画像をそのまま学習データとした場合，正確な推定モデルの構築は困難となる．これに対し，文献[75)]では，それらの画像の中から推定に有効な画像を同時に選択可能とする手法を構築した．これは，弱教師あり学習，自己教師あり学習に基づいた研究とみなすこと

ができ，独自の新たな研究課題が生み出された一例である．

7．むすび

本稿で紹介した通り，近年のメディア工学技術の発展は目覚ましく，特にAI技術については精力的な研究開発が続けられており，今後も進展していくものと思われる．またその社会実装が新たな課題を提示し，それを技術が解決していく流れもできつつある．技術開発と社会実装が両輪となる形で，メディア工学技術が社会を変えていくものと期待される．

（2021年11月1日受付）

〔文　献〕

1) https://github.com/hoya012/cvpr-2020-paper-statistics（2021年10月10日参照）
2) https://github.com/hoya012/cvpr-2021-paper-statistics（2021年10月10日参照）
3) 村上和人，長谷山美紀，田川憲男，田良島周平，石井大祐：“映像情報メディア学会の研究会活動～過去から未来へ～1-7メディア工学研究委員会”，映情学誌（2020）
4) 田良島周平，石井大祐：“［創立70周年記念講演会］映像情報メディアのグランドチャレンジ講演5メディア工学研究委員会”，映情学創立70周年記念大（2020）
5) S. Tarashima and D. Ishii: "Future trends in image information and media technology (3) grand challenges of media engineering over next decade", IDW'21 Special Events (2021)
6) “MLOpsに基づくAI/ML実運用最前線”，映情学冬季大2021特別イベント（2021）
7) A. Krizhevsky, I. Sutskever and G.E. Hinton: "Imagenet classification with deep convolutional neural networks", In F. Pereira, C.J. C. Burges, L. Bottou and K.Q. Weinberger, editors, Advances in Neural Information Processing Systems, **25**, Curran Associates, Inc. (2012)
8) O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg and L. Fei-Fei: "ImageNet Large Scale Visual Recognition Challenge", International Journal of Computer Vision (IJCV), **115**, 3, pp.211-252 (2015)
9) A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser and I. Polosukhin: "Attention is all you need", In I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett, editors, Advances in Neural Information Processing Systems, **30**, Curran Associates, Inc. (2017)
10) J. Devlin, M.-W. Chang, K. Lee and K. Toutanova: "BERT: Pre-training of deep bidirectional transformers for language understanding", In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, **1** (Long and Short Papers), pp.4171-4186, Minneapolis, Minnesota (June 2019, Association for Computational Linguistics
11) A. Radford, J. Wu, R. Child, D. Luan, D. Amodei and I. Sutskever: "Language models are unsupervised multitask learners" (2019)
12) P. Ramachandran, N. Parmar, A. Vaswani, I. Bello, A. Levskaya and J. Shlens: "Standalone self-attention in vision models", In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox and R. Garnett, editors, Advances in Neural Information Processing Systems, **32**. Curran Associates, Inc. (2019)
13) A. Vaswani, P. Ramachandran, A. Srinivas, N. Parmar, B. Hechtman and Jo. Shlens: "Scaling local self-attention for parameter efficient visual backbones", In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp.12894-12904 (June 2021)
14) K. He, X. Zhang, S. Ren and J. Sun: "Deep residual learning for image recognition", In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2016)
15) H. Wang, Y. Zhu, B. Green, H. Adam, A. Yuille and L.-C. Chen: "Axial-deeplab: Stand-alone axial-attention for panoptic segmentation", In European Conference on Computer Vision (ECCV) (2020)
16) 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 International Conference on Learning Representations (2021)
17) J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei: "Imagenet: A large-scale hierarchical image database", In 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp.248-255 (2009)
18) C. Sun, A. Shrivastava, S. Singh and A. Gupta: "Revisiting unreasonable effectiveness of data in deep learning era", In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Oct. 2017)
19) K. Wu, H. Peng, M. Chen, J. Fu and H. Chao: "Rethinking and improving relative position encoding for vision transformer", In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp.10033-10041 (Oct. 2021)
20) K. Mahmood, R. Mahmood and M. van Dijk: "On the robustness of vision transformers to adversarial examples", In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp.7838-7847 (Oct. 2021)
21) M. Caron, H. Touvron, I. Misra, H. Jegou, J. Mairal, P. Bojanowski and A. Joulin: "Emerging properties in self-supervised vision transformers", In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp.9650-9660 (Oct. 2021)
22) X. Chen, S. Xie and K. He: "An empirical study of training self-supervised vision transformers", In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp.9640-9649 (Oct. 2021)
23) Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin and B. Guo: "Swin transformer: Hierarchical vision transformer using shifted windows", In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp.10012- 10022 (Oct. 2021)
24) W. Wang, E. Xie, X. Li, D.-P. Fan, K. Song, D. Liang, T. Lu, P. Luo and L. Shao: "Pyramid vision transformer: A versatile backbone for dense prediction without convolutions", In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp.568-578 (Oct. 2021)
25) P. Zhang, X. Dai, J. Yang, B. Xiao, L. Yuan, L. Zhang and J. Gao: "Multi-scale vision longformer: A new vision transformer for high-resolution image encoding", In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp.2998-3008 (Oct. 2021)
26) H. Fan, B. Xiong, K. Mangalam, Y. Li, Z. Yan, J. Malik and C. Feichtenhofer: "Multiscale vision transformers", In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp.6824-6835 (Oct. 2021)
27) B. Heo, S. Yun, D. Han, S. Chun, J. Choe and S.J. Oh: "Rethinking spatial dimensions of vision transformers", In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp.11936-11945 (Oct. 2021)
28) R. Ranftl, A. Bochkovskiy and V. Koltun: "Vision transformers for dense prediction", In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp.12179-12188 (Oct. 2021)
29) A. Arnab, M. Dehghani, G. Heigold, C. Sun, M. Lucic and C. Schmid: "Vivit: A video vision transformer", In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp.6836-6846 (Oct. 2021)
30) Z. Pan, B. Zhuang, J. Liu, H. He, and J. Cai: "Scalable vision transformers with hierarchical pooling", In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp.377-386 (Oct. 2021)
31) C.-F. (R.) Chen, Q. Fan and R. Panda: "Crossvit: Cross-attention multi-scale vision transformer for image classification", In Proceedings of

the IEEE/CVF International Conference on Computer Vision (ICCV), pp.357-366 (Oct. 2021)
32) K. He, G. Gkioxari, P. Dollar and R. Girshick: "Mask r-cnn", In Proceedings of the IEEE International Conference on Computer Vision (ICCV) (Oct. 2017)
33) Z. Dai, B. Cai, Y. Lin and J. Chen: "Up-detr: Unsupervised pre-training for object detection with transformers", In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp.1601-1610 (June 2021)
34) MPEG: "Point Cloud Coding", https://mpeg.chiariglione.org/standards/mpeg-i/point-cloud-compression
35) B. Bross, et al.: "Overview of the Versatile Video Coding (VVC) Standard and its Applications", IEEE Trans. on CSVT, 31, 10, pp.3736-3764 (Oct. 2021)
36) J.M. Boyce, et al.: "MPEG Immersive Video Coding Standard", in Proceedings of the IEEE, 109, 9, pp.1521-1536 (Sep. 2021)
37) E.S. Jang, et al.: "Video-Based Point-Cloud-Compression Standard in MPEG: from Evidence Collection to Committee Draft", IEEE Signal Processing Mag., 36, 3 (2019)
38) Nokia: "Video Point Cloud Coding (V-PCC) AR Demo", https://github.com/nokiatech/vpcc
39) M. Preda, et al.: "G-PCC codec description v12", ISO/IEC JTC1/SC29/WG7 N00151 (2021)
40) MPEG-3DG: "MPEG Point Cloud Compression", http://www.mpeg-pcc.org/
41) リクルート：“AR Room Simulator”，https://prtimes.jp/main/html/rd/p/000001236.000011414.html（2021）
42) 電通：“NTTと電通は共同で大規模なVR空間における広告モデル実証を始動”，https://www.dentsu.co.jp/news /release/2021/0929-010443.html（2021/9/29）
43) Misoshita, Mekezzo, Nandemo Token: "クリプトアートプロジェクト「Metaani」", https://conata.world/metaani/gen (2021)
44) J. Benet: "Content Addressed, Versioned, P2P File System", https://ipfs.io/ipfs/QmR7GSQM93Cx5eAg6a6yRzNde1FQv7uL6X1o4k7zrJa3LX/ipfs.draft3.pdf (2014)
45) Nextech AR Solutions Corp.: "Nextech to Launch Augmented Reality NFT Hologram Creator Platform: nextech-to-launch-augmented-reality-nft-hologram-creator-platform" (2021/7/26)
46) SAE International: "Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles", J3016_201806 (2018)
47) 警察庁：“改正道路交通法（自動運転関係）の概要”，https://www.npa.go.jp/bureau/traffic/selfdriving/trafficact.pdf（2021年10月31日参照）
48) 内閣府：“戦略的イノベーション創造プログラム（SIP）自動運転（システムとサービスの拡張）研究開発計画”（2020）
49) 井坪慎二，馬渡真吾，岩里泰幸，関谷浩孝，澤井聡志：“実証実験を通じた中山間地の自動運転の課題と対応についての分析”，第60回土木計画学研究発表会・講演集（2019）
50) 国土交通省自動車局：“自動運転車の安全技術ガイドライン”（2018）
51) 大見拓寛：“画像センサによる眠気状態推定とドライバステータスモニタの開発”，Denso technical review，21，pp.93-102（2016）
52) 式井愼一，砂川未佳，楠亀弘一，望月誠，北島洋樹，下村義弘：“眠気検知・予測技術に基づくドライバモニタシステム”，パナソニック技報，64, 2, pp.69-73（2018）
53) 日向匡史，木下航一，西行健太，長谷川友紀：“自動運転時代におけるドライバモニタリング技術”，OMRON TECHNICS，50，1，pp.36-41（2018）
54) 多田昌裕，小出瑞，西田将之，飯田克弘，澤田英郎：“生理指標と行動指標を用いた部分自動運転機能使用時のドライバの運転特性把握”，土木学会論文集D3（土木計画学），76，5，pp.I_739-I_746（2021）
55) I. Jegham, A.B. Khalifa, I. Alouani, M.A. Mahjoub: "A novel public dataset for multimodal multiview and multispectral driver distraction analysis: 3MDAD", Signal Processing: Image Communication, 88, 115960 (2020)
56) J. Carmona, C. Guindel, F. Garcia, A. Escalera: "eHMI: Review and Guidelines for Deployment on Autonomous Vehicles", Sensors, 21, 9, 2912 (2021)
57) Daimler: "DIGITAL LIGHT", https://www.daimler.com/innovation/specials/geneva-2018/digital-light.html (2021年10月31日参照)
58) https://www.itmedia.co.jp/news/articles/2107/24/news021.html (2021年10月28日参照)
59) https://www.itmedia.co.jp/news/articles/2107/30/news142.html (2021年10月28日参照)
60) https://biz.panasonic.com/jp-ja/tokyo2020_lp_projector (2021年10月28日参照)
61) https://av.watch.impress.co.jp/docs/news/1356136.html (2021年10月28日参照)
62) https://project.nikkeibp.co.jp/mirakoto/atcl/mirai/h_vol90/ (2021年10月28日参照)
63) “超高臨場感通信技術Kirari! Beyond 2020”，NTT技術ジャーナル 2018年10月号
64) https://www.asahi.com/articles/ASP826VYFP82ULFA00Q.html (2021年10月28日参照)
65) Z. Guo, X. Li, H. Huang, N. Guo and Q. Li: "Deep learning-based image segmentation on multimodal medical imaging", IEEE Transactions on Radiation and Plasma Medical Sciences, 3, 2, pp.162-169 (2019)
66) J. Ker, L. Wang, J. Rao and T. Lim: "Deep learning applications in medical image analysis", IEEE Access, 6, pp.9375-9389 (2018)
67) H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyama and H. Omata: "Road damage detection and classification using deep neural networks with smartphone images", Computer-Aided Civil and Infrastructure Engineering, 33, 12, pp.1127-1141 (2018)
68) M.T. Cao, Q.V. Tran, N.M. Nguyen and K.T. Chang: "Survey on performance of deep learning models for detecting road damages using multiple dashcam image resources", Advanced Engineering Informatics, 46, 101182 (2020)
69) Z. Li, R. Togo, T. Ogawa and M. Haseyama: "Chronic gastritis classification using gastric X-ray images with a semi-supervised learning method based on tri-training", Medical & Biological Engineering & Computing, 58, 6, pp.1239-1250 (2020)
70) K. Maeda, S. Takahashi, T. Ogawa and M. Haseyama: "Estimation of deterioration levels of transmission towers via deep learning maximizing canonical correlation between heterogeneous features", IEEE Journal of Selected Topics in Signal Processing, 12, 4, pp.633-644 (2018)
71) G. Li, R. Togo, T. Ogawa and M. Haseyama: "Soft-label anonymous gastric X-ray image distillation", 2020 IEEE International Conference on Image Processing (IEEE ICIP 2020), pp.305-309 (2020)
72) R. Khurana: "How to make artificial intelligence more democratic", Scientific American, https://www.scientificamerican.com/article/how-to-make-artificial-intelligence-more-democratic/
73) R. Togo, K. Ishihara, T. Ogawa, M. Haseyama: "Estimation of salient regions related to chronic gastritis using gastric X-ray images", Computers in Biology and Medicine, no.77, pp.9-15 (2016)
74) N. Ogawa, K. Maeda, T. Ogawa, M. Haseyama: "Correlation-aware attention branch network using multi-modal data for deterioration level estimation of infrastructures", 2021 IEEE International Conference on Image Processing (ICIP 2021), pp.1014-1018 (2021)
75) N. Ogawa, K. Maeda, T. Ogawa and M. Haseyama: "Distress image retrieval for infrastructure maintenance via self-trained deep metric learning using experts' knowledge", IEEE Access, 9, pp.65234-65245 (2021)

新井　啓之　1989年，東京理科大学理工学部物理学科卒業．1991年，北海道大学理学研究科物理学専攻修士課程修了．同年，日本電信電話（株）研究所に入所．2001年～2006年，情報通信研究機構ナチュラルビジョンプロジェクト特別研究員．2017年より，日本工業大学工学部情報工学科（現，先進工学部情報メディア工学科）教授．画像処理，画像認識技術の研究開発に従事．正会員．

田良島周平　2009年，東京大学工学部卒業．2011年，同大学大学院新領域創成科学研究科修士課程修了．同年，NTT入社．現在，NTTコミュニケーションズ（株）イノベーションセンター所属．画像認識に関する研究開発に従事．正会員．

河村　圭　2004年，早稲田大学理工学部電子・情報通信学科卒業．2005年，同大学大学院国際情報通信研究科修士課程修了．2010年，同大学大学院国際情報通信研究科博士課程修了．同年，KDDI（株）入社．現在，（株）KDDI総合研究所超臨場感通信グループリーダー．2017年，本学会鈴木記念奨励賞等を受賞．主に，動画像符号化方式の研究・開発および国際標準化に従事．博士（国際情報通信学）．正会員．

堀　淳志　1997年，芝浦工業大学大学院修士課程修了．同年，三菱電機（株）入社．現在，同社情報技術総合研究所情報表現技術部HMSソリューショングループマネージャ．主にグラフィックスシステムおよびヒューマンマシンシステムの研究開発に従事．正会員．

多田　昌裕　2005年，中央大学大学院理工学研究科博士後期課程修了．同年，（株）国際電気通信基礎技術研究所（ATR）入所．2013年，近畿大学理工学部講師，2018年，同大学准教授となり，現在に至る．ドライバ行動解析，運転支援システム等の研究開発に従事．博士（工学）．正会員．

望月　貴裕　1994年，東京工業大学工学部情報工学科卒業，1996年，同大学大学院物理情報工学専攻修士課程修了．同年，NHK入局．報道技術センター中継制作部を経て，1998年より，放送技術研究所に勤務．画像検索や映像自動要約などの画像解析技術に関する研究に従事．現在，同研究所スマートプロダクション研究部上級研究員．博士（工学）．正会員．

小川　貴弘　2003年，北海道大学工学部卒業．2007年，同大学大学院情報科学研究科博士後期課程修了．2007年～2008年，同大学博士研究員．2008年～2016年，同大学助教．2016年より，同大学准教授．マルチメディアデータを対象としたAI・IoT・ビッグデータ解析に関する研究に従事．博士（情報科学）．正会員．

新刊図書

教養としてのデータサイエンス

北川源四郎・竹村彰通　編
内田誠一・川崎能典・孝忠大輔・佐久間淳・
椎名　洋・中川裕志・樋口知之・丸山　宏　著

データサイエンスといえば，最近ではいくつかの大学に新しい学部ができるなど，学問の新しい分野としての盛り上がりを見せています．それでも多くの人にとっては，実際にデータサイエンスがどんなものなのか，まだまだ未知のものなのではないでしょうか．私もデータサイエンスに関連する分野を専門とするなかでさまざまなバックグラウンドを持つ方とお話をする機会がありますが，「人工知能は人よりも賢いんでしょ」などの声を聞くこともあります．もちろん，人工知能だけがデータサイエンスではありません．例えば，いろいろなメディアで世間の年収の平均値などが話題になることがありますが，思った以上に高くて驚くかもしれません．しかし，平均値がどういうものなのかを知っていれば，その裏にある年収の分布に考えを巡らせることができます．いろいろなデータや統計が身の回りに溢れている現代では，これらをどう解釈するかを知っていることが大きな強みになります．

本書は，データサイエンスに含まれる多くのキーワードを広くカバーする入門書で，タイトルにもあるように，教養としてこの分野を俯瞰したい方におすすめの一冊です．三章構成となっており，第一章では導入として，データがどのように活用されているのか，また人工知能とは一体どのようなものなのかを具体的な例とともに紹介しています．この章のキーワードから興味のあるトピックを掘り下げていくこともできるでしょう．第二章は，どのようにデータを扱うのか，統計からデータの可視化まで，データサイエンスで基本となるツールについて学ぶことができます．専門用語も平易な言葉で丁寧に解説されており，さらに具体的な例も多く紹介されているので，今まで統計を敬遠していた方にも親しみやすいのではないでしょうか．第三章では，データサイエンスに関わる際に注意すべき個人データの取り扱いや人工知能の負の事例が紹介されています．人工知能の差別的な振る舞いにもつながるデータセットバイアスなど，新しい話題にも触れられています．

すべてのページがカラーで印刷された本書は，多数の具体的な事例を図や写真を含めてわかりやすく解説しています．教科書としてだけではなく読み物としても面白く，データサイエンスに踏み込む第一歩に適した良書だと思います．

紹介　中島悠太（阪大）

講談社刊（2021年6月17日発行），A5版，240頁，定価1,800円＋税