STI Horizon 2024 Vol.10 No.1 https://doi.org/10.15108/stih.00363

ほらいずん

海外技術情報「コンピュータビジョンとパターン認識に関する国際会議 2023　CVPR2023」参加報告
－コンピュータビジョンと人工知能分野のトップカンファレンス－

科学技術予測・政策基盤調査研究センター 研究員 鎌田 久美

【概　要】

「コンピュータビジョンとパターン認識に関する国際会議 2023（CVPR2023)」は、2023 年 6 月 18 日から 22 日まで、カナダのバンクーバーで開催された。

CVPR はコンピュータビジョンと人工知能分野で採択率が約 25%前後の最難関の国際会議であり、当該分野の最先端の研究が発表された。発表内容は、マルチビューの 3 次元（3D）構成画像に関する研究、機械学習を用いた画像・動画の生成に関する研究、人の顔やポーズの画像に関する研究等の発表が実施された。

今回の投稿数は 9155 件、採択数は 2359 件、採択率は 25.8%であった。国別の参加者数は、第 1 位が米国の 2779 件、第 2 位が中国の 1413 件、第 3 位が韓国の 815 件、第 4 位がカナダの 672 件、第 5 位がドイツの 372 件、第 6 位が日本の 304 件であった。

コロナ禍以来、2020 年から 2021 年まではバーチャル開催であったが、2022 年以降はハイブリッド開催となり、現地参加者が戻りつつあり、研究者同士が対面で会って議論することが再び増えた国際会議となった。

キーワード：コンピュータビジョン，ジェネレーティブ AI，3D 構成画像，自動運転

1. 投稿数と採択数の推移（2006-2023）

CVPR の投稿数、採択数及び採択率の推移が発表された。

CVPR2023 の投稿数 (全体) は 9155 件、採択数は 2359 件、採択率は 25.8%となっている。投稿数は 2017 年以降急激に増加しており、採択数も同様に増加している。採択率は約 25%～30%であり、CVPR はコンピュータビジョン分野のトップカンファレンスであり、最難関とされている（図表 1)。

2. CVPR2023 の国・地域別の参加者数の内訳

CVPR2023 の国別の参加者数の内訳が発表された（図表 2)。

図表 1　CVPR の投稿数と採択数の推移（2006-2023）

採択年	投稿数	採択数	採択率	採択年	投稿数	採択数	採択率
2006	1131	318	28.1	**2015**	2123	602	28.3
2007	1250	353	28.2	**2016**	2145	643	29.9
2008	1593	508	31.9	**2017**	2680	783	29.2
2009	1464	383	26.2	**2018**	3309	979	29.6
2010	1724	462	26.8	**2019**	5165	1294	25.1
2011	1677	440	26.2	**2020**	6424	1467	22.8
2012	1993	466	23.4	**2021**	7093	1661	23.4
2013	1798	472	26.3	**2022**	8161	2063	25.3
2014	1807	540	29.9	**2023**	9155	2359	25.8

参照：CVPR2023 オープニング発表より作成

（2024.3.21 公開）

図表 2　CVPR2023 の国別の参加者数の内訳

順位	国名	件数	順位	国名	件数
1	米国	2779	11	香港	119
2	中国	1413	12	イスラエル	118
3	韓国	815	13	オーストラリア	115
4	カナダ	672	14	インド	101
5	ドイツ	372	15	イタリア	85
6	日本	304	16	スペイン	52
7	英国	303	17	スウェーデン	50
8	スイス	173	18	オランダ	47
9	シンガポール	148	19	デンマーク	30
10	フランス	141	20	ベルギー	27

参照：CVPR2023 オープニング発表より作成

第 1 位は米国の 2779 件、第 2 位は中国の 1413 件、第 3 位は韓国の 815 件、第 4 位はカナダの 672 件、第 5 位はドイツの 372 件、第 6 位は日本の 304 件等となっている。米国と中国の参加者が多数を占めていた。

3. CVPR2023 の分野別の発表件数

CVPR2023 の分野別の発表件数（著者数による順位）が発表された（図表 3）。

第 1 位の「マルチビューとセンサからの 3 次元構成」は、マルチビュー（多視点）を利用して、対象物を複数の画像からとらえて 3 次元構成画像を生成する研究であり、著者数が 1090 名、論文数が 246 件と最も発表が多かった。

第 2 位の「画像と動画の合成と生成」は、Stable Diffusion（ステイブル・ディフュージョン）などの深層学習を用いた画像と動画の生成に関する研究が多数発表されており、著者数 889 名、論文数が 185 件と多くを占めていた。

第 3 位の「人の画像」に関する研究は、顔、身体、ポーズ、ジェスチャー動作等の立体画像を基にした、様々な表情や動作の創出に関する研究、人のポーズやジェスチャー動作の表現に関する研究が発表されており、著者数 813 名、論文数 166 件であった。

4. 受賞論文

CVPR2023 の論文受賞は、最優秀論文が 2 本、佳作（優秀）論文が 1 本、学生最優秀論文が 1 本、学生佳作（優秀）論文が 1 本であった。

①最優秀論文 -1

「Visual Programming（VISPROG）; Compositional Visual Reasoning Without Training：（ビジュアルプログラミング：訓練なしの構成的視覚的推論）」[1)]

Tanmay Gupta 氏、Aniruddha Kembhavi 氏（Allen Institute for AI）（米国）（図表 4）

この発表では、自然言語の指示を受けて、複雑で構成的な視覚的タスクを解決することが可能な、ビジュアルプログラミング（VISPROG）を提案している。1 枚又は複数画像と自然言語の命令を与えて、GPT-3 を利用して命令プログラムを生成し、そのプログラムを実行することで目的の出力を得るシステムとなっている。GPT-3 によってタスク固有の訓練が不要となり、少数の例からプログラムを作成できること、中間出力を確認することで間違いの理由や視覚的根拠を得ることが特徴となっている。

ビジュアルプログラミングは、大規模言語モデルのコンテキスト内学習機能を使用して、モジュール型プログラムを生成し、コンピュータビジョンモデル、画像処理サブルーチン、又は Python 関数を呼び出して、中間出力を生成する。これらの作業によって、解釈可能な理論的根拠と解決策を得ることを実現している。そして、視覚的質問への構成的な回答、画像ペアのゼロショット推論、事実知識オブジェクトのタグ付け、及び言語ガイド付き画像編集の 4 つのタスクにより、一連の作業の柔軟性を実証している。

今回紹介された、Neuro-Symbolic AI（ニューロシンボリック AI）[注 1] によるアプローチに基づくビジュアルプログラミングは、AI システムを簡単及び効果的に拡張して、利用者が実行したいと考えている複雑なタスクに対応ができる、非常に有用なシステムであ

注 1　Neuro-Symbolic AI（Neuro ＝ニューラルネットワーク、Symbolic ＝記号的表現に基づく）：深層ニューラルネットワークの強み、及びシンボリック（記号的表現に基づく）AI の強みを合わせ持った AI のことを指す。IBM などが中心となって進めている最新の AI 技術である。

図表 3 CVPR2023 の分野別の発表件数

順位	タイトル	著者数	論文数
1	マルチビューとセンサからの３次元構成 （3D from multi-view and sensors）	1090	246
2	画像と動画の合成と生成 （Image and video synthesis and generation）	889	185
3	人の画像：顔、身体、ポーズ、ジェスチャー動作 （Humans: Face, body, pose, gesture, movement）	813	166
4	転移学習、メタ学習、ローショット学習、継続学習、ロングテール学習 （Transfer, meta, low-shot,continual,or long-tall learning）	688	153
5	認識：分類、検出、検索 （Recognition : Categorization, detection, retrieval）	673	139
6	ビジョン・言語・推論 （Vision,language,and reasoning）	631	118
7	低解像度ビジョン （Low-level vision）	553	126
8	セグメンテーション. グルーピング・形状分析 （Segmentation,grouping and shape analysis）	524	113
9	ディープラーニングのアーキテクチャーと手法 （Deep learning architectures and techniques）	485	92
10	マルチモーダル学習 （Multi-modal-leaning）	450	89
11	1 枚の画像からの 3 次元画像の生成 （3D from single images）	431	91
12	医学的生物学的ビジョン、細胞顕微鏡 （Medical and biological vision, cell microscopy）	420	53
13	ビデオ：アクションとイベントの理解 （Video: Action and event understanding ）	373	63
14	自動運転 （Autonomous driving）	359	69
15	自己教師あり表現学習又は自己教師なし表現学習） （Self-supervised or unsupervised representation learning）	349	71
16	データセットと評価 （Datasets and evaluation）	344	54
17	シーンの分析と理解 （Scene analysis and understanding）	276	54
18	敵対的な攻撃と防御 （Adversarial attack and defense）	274	61
19	効率的でスケーラブルなビジョン （Efficient and scalable vision）	252	48
20	コンピュテーショナルイメージング （Computational Imaging）	226	53
21	ビデオ：低解像度分析、モーション、トラッキング （Video: Low-level analysis, motion and tracking）	215	46
22	ビジョンアプリケーション及びシステム （Vision applications and systems）	171	35
23	ビジョングラフィックス （Vision graphics）	155	32
24	ロボティクス （Robotics）	141	23
25	透明性、公平性、説明責任、プライバシー、倫理、ビジョン （Transparency,fairness,accountability,privacy,ethics and in vision）	129	30
26	説明可能なコンピュータビジョン （Explainable computer vision）	107	24
27	身体性ビジョン：アクティブエージェント、シミュレーション （Embodied vision: Active agents,simulation）	80	14
28	ドキュメントの分析と理解 （Document analysis and understanding）	72	23
29	機械学習（深層学習以外） （Machine learning (other than deep learning)）	65	14
30	物理ベースビジョン及び X からの形状 （Physics-based vision and shape-from-X）	55	12
31	バイオメトリクス （Biometrics）	51	11
32	その他 （Others）	47	12
33	最適化手法（深層学習以外） （Optimization methods (other than deep learning)）	46	12
34	写真測量とリモートセンシング （Photogrammetry and remote sensing）	38	6
35	コンピュータビジョン理論 （Computer vision theory）	33	5
36	ソーシャルグッドのためのコンピュータビジョン （Computer vision for social good）	25	5

参照：CVPR2023 オープニング発表より作成

図表 4　画像編集（上）と知識タグ付けタスク（下）

例 1）画像編集タスク（左上の画像）：言語指示を受けて、デカプリオ氏の画像（上）を、サングラスをかけたデカプリオ氏の画像（下）に編集している。
例 2）画像編集タスク（右上の画像）：言語指示を受けて、緑色のソファ（上）を青色のソファ（下）に編集している。
例 3）知識タグ付けタスク（左下の画像）：ドイツ、台湾、ニュージーランドの女性指導者のタグ付けを行っている。
出典：https://openaccess.thecvf.com/content/CVPR2023/papers/Gupta_Visual_Programming_Compositional_Visual_Reasoning_Without_Training_CVPR_2023_paper.pdf

ると紹介されている。

②最優秀論文 -2

「Planning-oriented Autonomous Driving：（計画的自動運転）」[2)]

Yihan Hu 氏、Jiazhi Yang 氏、Li Chen 氏、ほか

（Shanghai AI Laboratory, Wuhan University, Sense Time Research）（中国）（図表 5）

自動運転は非常に複雑な技術から構成されていて、モジュール化（単体で特定の機能を発揮することができる単位）したタスクである認識、予測、計画を順番に実行することが特徴であるが、これまではフレームワーク（枠組み）に関する研究は余り行われてこなかった。

自動運転の実現には、エラーや調整不足の発生に備えて、有利なフレームワークを考案し最適化すること、認識と予測の主要素を再検討し、全てのタスクが自動運転の計画に貢献するようにタスクに優先順位をつけることが必要となる。最近のアプローチは、高度なインテリジェンスを持つスタンドアロンモデル（機器やシステムが外部に接続、又は、依存せず単独で機能している状態のこと）、及びマルチタスクパラダイム（システムが同時に複数のタスクを平行して実行するという理論的な枠組み）を設計することが挙げられる。

今回の論文では、統合自動運転 Unified Autonomous Driving（UniAD）が提案された。これは、認識タスクにおける物体の検出、追跡、走路のマッピング、予測タスクにおける行動予測や占有状態予測など、フルスタックの運転タスクを 1 つのネットワークに組み込んだ、最新の包括的なフレームワークである[注 2]。このフレームワークによる革新的なアプローチは、自動運転技術研究における重要なブレークスルーとなり、受賞となった。

③佳作（優秀）論文

「DynlBaR: Neural Dynamic Image-Based Rendering：（ニューラルダイナミックイメージベースレンダリング）」

Zhengqi Li 氏、Qianqian Wang 氏、Forrester Cole 氏、ほか

（Google Research, Cornell Tech）（米国）

注 2　大規模なアブレーション（機械学習の予測モデルにおいて構成要素の一部分を取り除いて実験を行い、結果を比較すること）では、これまでの最先端技術を大幅に上回るパフォーマンスを示している。

図表 5　統合自動運転（UniAD）のパイプライン（一連のプロセスの自動化システム）

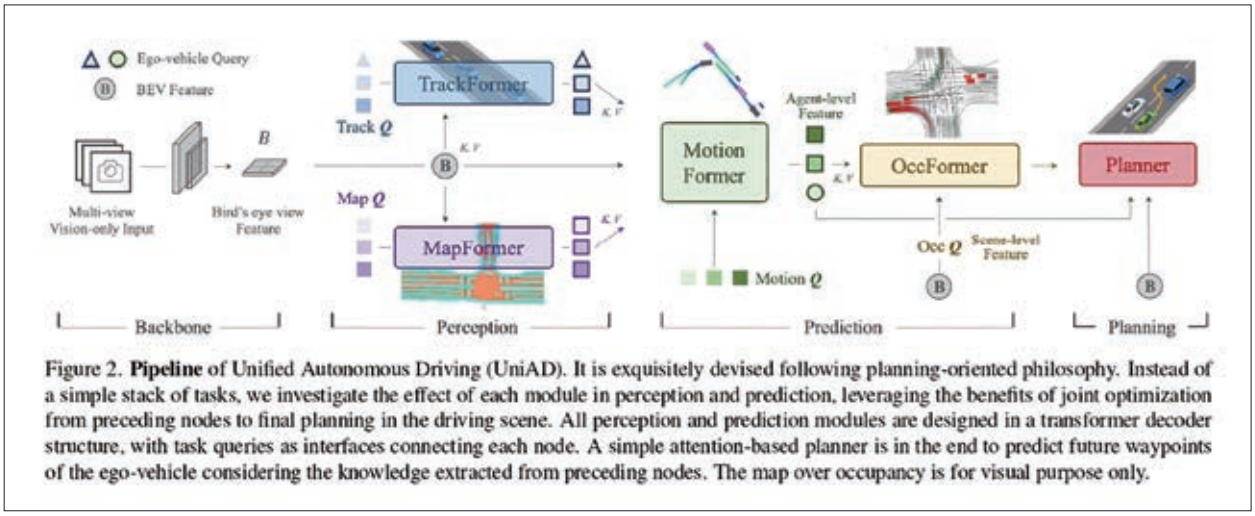

Figure 2. **Pipeline** of Unified Autonomous Driving (UniAD). It is exquisitely devised following planning-oriented philosophy. Instead of a simple stack of tasks, we investigate the effect of each module in perception and prediction, leveraging the benefits of joint optimization from preceding nodes to final planning in the driving scene. All perception and prediction modules are designed in a transformer decoder structure, with task queries as interfaces connecting each node. A simple attention-based planner is in the end to predict future waypoints of the ego-vehicle considering the knowledge extracted from preceding nodes. The map over occupancy is for visual purpose only.

単純なタスクの積み重ねではなく、認識と予測の各モジュールの効果を調査し、先行ノードから運転シーンにおける最終計画ノードに至るまでの全体最適化の利点を活用する。全ての認識及び予測のモジュールは、各ノードを接続するインタフェースとして、タスククエリを備えたトランスデコーダー構造で設計されている。
出典：https://openaccess.thecvf.com/content/CVPR2023/papers/Hu_Planning-Oriented_Autonomous_Driving_CVPR_2023_paper.pdf

単眼カメラで撮影された動画から、全く別の視点から見たときの映像を違和感なく再構成する。動きの激しい物体でも消えたりボケたりすることなく、任意視点でシーンを再構成できている。
出典：https://openaccess.thecvf.com/content/CVPR2023/papers/Li_DynIBaR_Neural_Dynamic_Image-Based_Rendering_CVPR_2023_paper.pdf

④学生最優秀論文
「3D Registration with Maximal Cliques：(最大限排他的 3 次元レジストレーション)」
Xiyu Zhang 氏、Jiaqi Yang 氏、Shikun Zhang 氏、ほか
(Northwestern Polytechnical University)（中国）
コンピュータビジョンで課題となっている、3 次元点群処理における点群ペアをそろえるための最適手法として、3 次元レジストレーション法とディープラーニング手法を組み合わせることで、性能が向上されたことが示された。
出典：https://openaccess.thecvf.com/content/CVPR2023/papers/Zhang_3D_Registration_With_Maximal_Cliques_CVPR_2023_paper.pdf

⑤学生佳作（優秀）論文
「DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation (被写体とテキストからのイメージのディフュージョンモデル)」
Nataniel Ruiz 氏、Yuanzhen Li 氏、Varun Jampani 氏、ほか
(Google Research, Boston University)（米国）
異なるテキスト入力に従って、被写体の特徴を多く残しつつ、様々な状況やシーンでの被写体が含まれる画像を生成する研究が発表された。
出典：https://openaccess.thecvf.com/content/CVPR2023/papers/Ruiz_DreamBooth_Fine_Tuning_Text-to-Image_Diffusion_Models_for_Subject-Driven_Generation_CVPR_2023_paper.pdf

5. 企業展示

国際会議への出展は、学界や産業界の世界的リーダーに宣伝するための最も費用効果の高い方法であり、業界のパイオニアとのつながり、ターゲットとする潜在的な顧客を獲得する機会となる。又、優秀な学生をリクルートする絶好の機会となっている（図表 6）。

6. 招待講演

CVPR2023 の招待講演の中で注目を浴びていた講演を紹介する。

基調講演「Recycling Old Vision Ideas in a Modern Computational World」
Rodney Brook 氏（MIT）[3)]
マサチューセッツ工科大学のロドニーブルック氏

図表 6　展示会の様子

① TikTok の展示

②テスラの自動運転車

（会場より筆者撮影）

による「現代のコンピューティング世界における、古典ビジョンのアイデアのリサイクル」と題する発表があった。「アイデアは何度も繰り返しやって来て、時々より良い結果をもたらす。最新世代のニューラルネットワークは、新しいシリコンのアーキテクチャーをもたらした。」と述べ、アイデアの温故知新の必要性について指摘した。

また、コンピュータの急速な進化について、「1900 年から進化を続けて、1000 ドルのコンピュータは、2000 年には昆虫の頭脳に匹敵するようになり、2020 年には一人の人間の頭脳に近づき、2045 年には全人類の頭脳に匹敵するようになる（シンギュラリティ）。」と述べた。

7. まとめ

対面を中心として開催された国際会議であったため、活気が戻ってきた雰囲気であった。大会当局も、ポスターセッションや展示会等に重点をおいた構成にして、対話を重視したプログラム構成となっていた。

コンピュータビジョンの研究内容は、基礎研究から応用研究まで発表されていたが、自動運転やロボット、3D 動画像などの研究が多数発表されていた。生成 AI を用いた画像編集の研究が発表されて注目を浴びていた。

日本の発表も貢献しており、参加者数も第 6 位と多く、この分野の関心が大きいことが示されていた。

参考文献・資料

1) https://absolute-value.github.io/vision%20and%20language/2023/05/01/VisProg.html
2) https://www.sensetime.jp/information_detail/1407
3) https://www.planbox.com/wp-content/uploads/2018/01/Exponential-Growth-of-Computing.jpg