SciGA: 学術論文における Graphical Abstract 設計支援のための統合データセット

川田拓朗 [1] 根本颯汰 [2] 北田俊輔 [2] 彌冨仁 [1,2]
[1] 法政大学理工学部 [2] 法政大学大学院理工学研究科
{takuro.kawada.3g, sota.nemoto.5s}@stu.hosei.ac.jp
shunsuke.kitada.0831@gmail.com, iyatomi@hosei.ac.jp

概要

Graphical Abstract (GA) は論文の要点を視覚的に伝える重要な表現手段である. 効果的な GA の作成には高度なデザインスキルが求められ, 設計支援技術の実現が期待される. 本研究では, 約 14.5 万の論文と GA を含む 141 万枚の図からなるデータセット SciGA-140k を構築した. また, GA 設計支援の前段として, Abstract を基に論文内から GA として適切な図を検索するタスク Abst2GA Retrieval を提案する. 我々は CLIP を基盤とするベースラインを設計し, 提案タスクの有効性を示した. ベースラインは他の論文の GA を検索し, デザイン案を提示する支援機能も提供する. 我々のアプローチは GA 設計支援の新たな方向性を示し, AI for Science の発展に貢献する.

1 はじめに

科学的発見とその伝達は新たな知識の構築に不可欠であるが, その進展は研究者の限られた資源に依存している. この課題解決に向け, 科学的発見の自動化が注目され, AI for Science が幅広く議論されてきた [1, 2, 3]. 同様に, スライド [4] やポスター [5], 図の生成 [6] など, 研究成果の伝達支援も重要である.

Graphical Abstract（GA）は, 論文の提案手法や結果を要約する視覚的表現である. 論文の Introduction で参照される図 1 やティザー画像も同様の役割を担い, 読者に研究内容を効果的に伝える [7]. 一方, 効果的な GA の作成には研究内容を的確に視覚化するスキルが求められる [8, 9, 10]. 論文内の図を再利用・加工して作成されることも多く, 魅力的な GA の作成には適切な構成要素の選定が鍵となる.

本研究では, GA 設計支援を目的とした初の大規模論文データセット **SciGA-140k** を構築する. SciGA-140k は 144,883 件の論文の Full-Text とメタデータ, GA を含む 1,148,191 枚の関連画像を収録し, 論文の本文と図の統合的な解析の基盤を提供する. 図 1 に示すように, 論文誌から収集した GA には作成方法に基づくラベルが付与されている.

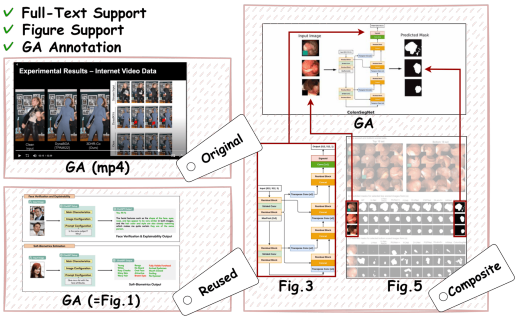


図 1: SciGA-140k が含む GA の例. [1)]

また, 本研究では, GA 設計プロセスの効率化に寄与する以下のタスクを定義する: 1) **intra-GA Recommendation**: 同一論文内から GA に適した図を推薦するタスク. 2) **inter-GA Recommendation**: ある論文の GA 作成において, 参考となる他の論文の GA を推薦するタスク. これらに対する直感的なアプローチは, 入力された図の評価値を直接推定することであるが, この古典的な手法は図の視覚的特徴のみに依存し, 論文の文脈を捉えられない.

そこで, 我々は intra-GA Recommendation に焦点を当て, Abstract をクエリとし, 論文内から GA 構成要素に適した図を検索するタスク **Abst2GA Retrieval** を提案する. そのベースラインでは, CLIP [11] を基盤に Abstract と図を同一空間に投影し, その類似度を GA 構成要素としての妥当性と解釈して intra-GA Recommendation を実現する. また, このベースラインの inter-GA Recommendation に対する応用可能性を示し, 将来的な GA 設計支援の統一フレームワークの基盤を提供する. 我々のアプローチは GA 設計を効率化し, AI for Science の新たな応用の幅を広げる.

2 SciGA-140k

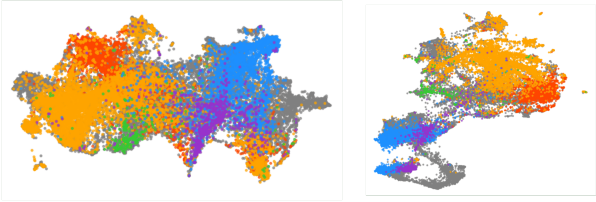
SciGA-140k は GA に関する注釈を提供する初のデータセットであり, 144,883 件の論文から収集したタイトル, 著者, Abstract, 本文, 投稿日, 研究分野, 著者によるコメント, DOI, 掲載論文誌, 採択会議に加え, 合計 1,148,191 枚の GA と図から構築される. 論文の本文と図を対象とした MMSci [12] などの従来のデータセットを大幅に上回る規模と多様性を備え, arXiv.org [2], ACM Computing Classification System (ACM-CCS) [3], Mathematics Subject Classification (MSC) [13] の階層的分類体系に基づき, 主要な 106 の研究分野と細分化された 5,656 の研究領域を網羅する. また, HTML 構造を基盤とした正確なデータ抽出により, 本文のセクションや図の subfigure の構造が正確に保持されている. 数式, 脚注, タグは特殊トークンで明確に識別可能であり, 文中の役割に応じた処理が容易である.

データ収集 SciGA-140k は, arXiv.org に投稿された論文を HTML 形式に変換した ar5iv:04:2024 dataset [14] から解析されたテキスト情報を基盤に構築された. 数式, 脚注, タグは特殊トークン<MATH>, <NOTE>, <TAG>で囲み, 識別可能とした. 図は ar5iv.org や arXiv の TeX ソースから収集し, 掲載論文誌から GA を画像や動画形式で収集した. また, arXiv API でメタデータを収集し, コメントから機械学習分野の国際会議の採択情報を抽出した.

アノテーション HTML 構造を基に特定したティザー画像に自動でラベルを付与した. また, 論文誌から収集した GA に対し, 以下の 3 種類のラベルを手動で付与した: 1) Original: 論文内の図を利用せず, に新規作成された GA. 2) Reuse: 論文内の図を再利用した GA. 3) Composite: 論文内の 1 つ以上の図を加工して作成された GA. GA として再利用, または加工された図は別途 GA 構成要素としてラベル付けした.

2.1 統計情報と分布分析

SciGA-140k から得られた統計情報は論文内の図や GA の特徴を示す. 各図のキャプションの平均トークン数は 48.11 ± 44.13, Abstract の平均トークン数は 164.72 ± 66.32 であった. また, 各論文の図の枚数は平均 6.16 ± 5.86 枚であり, subfigure を含めると平均 7.92 ± 10.45 枚, 最大 700 枚に達した. 特に, 天文物理学や実験核物理学で平均 10 枚を超える一方, 数学や数理物理学では 4 枚と少なく, 視覚的表現の重要性が分野ごとに異なることが確認された. GA の内訳は Original が 20.9%, Reused が 64.5%, Composite が 14.5%であった.



(a) GA. (b) Abstract.

図 2: 可視化した埋め込み. 分野ごとに着色している.

図 2 に GA と Abstract の分野ごとの分布を示す. 各点は CLIP で埋め込み, UMAP [15] で次元削減を行った GA, Abstract を示し, 分野ごとに着色されている. 分野ごとに分布の偏りが見られ, 特定分野に特化したモデルの適用が効果的である可能性を示唆する.

3 GA Recommendation

我々は, GA 設計支援を目的とした 2 種類の推薦するタスク intra-GA Recommendation, inter-GA Recommendation を定義する. 本研究では intra-GA Recommendation に着目し, 派生タスク Abst2GA Retrieval とベースラインを提案する.

3.1 問題定義

intra-GA Recommendation は「ある論文の GA あるいは同様の役割を持つ図 (Introduction で参照される図 1, ティザー画像) I_{GA} と論文に含まれるその他の n 枚の図 $I_1, I_2, \ldots, I_n$ からなる集合 $\mathcal{I}_{\mathrm{intra}} = \{I_i \mid i = 。\mathrm{GA}, 1, 2, \ldots, N\}$ の各図の GA としての妥当性を評価し, 上位の結果を推薦するタスク」と定義する. 各図の評価はその順序に依存せず独立に行う必要がある. I_{GA} の多くは図 1 であり, 図 1 を優先する単純なモデルは高評価を得るが, GA としての適性を適切に考慮できない.

inter-GA Recommendation は「N 枚の異なる論文の GA および同様の役割を持つ図からなる集合 $\mathcal{I}_{\mathrm{inter}} = \{I_{\mathrm{GA}}^{(i)} \mid i = 1, 2, \ldots, N\}$ の各図が, 別のある論文の GA 作成において, どれだけ参考になるかを評価し, 上位の結果を推薦するタスク」と定義する.

1) 左上: https://doi.org/10.1109/ACCESS.2023.3344658
左下: https://doi.org/10.1109/ACCESS.2021.3063716
右: https://doi.org/10.1109/ACCESS.2024.3370437
2) https://arxiv.org/category_taxonomy
3) https://dl.acm.org/ccs

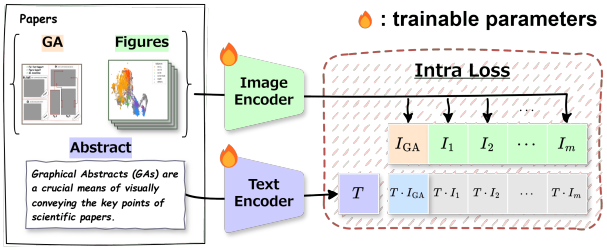


図 3: Abst2GA Retrieval のベースラインの事前学習.

3.2 Abst2GA Retrieval

本研究では intra-GA Recommendation の派生タスクとして, Abstract T をクエリとし, $\mathcal{I}_{\text{intra}}$ から I_{GA} とその構成要素 (加工して GA に利用された元の論文内の図) を検索する **Abst2GA Retrieval** を提案する. 各図 I_i は T との類似度で評価し, I_i が複数の subfigure から構成される場合は, 各 subfigure との類似度の最大値を評価値とする.

ベースライン Abst2GA Retrieval のベースラインとして CLIP を基盤とした対照学習モデルを構築し, T の埋め込み z^{T} と I_i の埋め込み z_i^{I} を同一空間に射影する. I_i は z^{T} と z_i^{I} のコサイン類似度 $\text{sim}(z^{\text{T}}, z_i^{\text{I}})$ で評価され, 類似度が大きいほど GA 構成要素として妥当であると判断される. 対照学習では, 一般的に次式の InfoNCE [16] で示される対照損失が用いられる:

$$\mathcal{L}_{\text{C}}(z^{\text{q}}, z^{+}, \{z_i^{-}\}) = -\log \frac{\exp\left(\frac{\text{sim}(z^{\text{q}}, z^{+})}{\tau}\right)}{\exp\left(\frac{\text{sim}(z^{\text{q}}, z^{+})}{\tau}\right) + \sum_i \exp\left(\frac{\text{sim}(z^{\text{q}}, z_i^{-})}{\tau}\right)} \quad (1)$$

ここで, z^{q} はクエリの埋め込み, z^{+} は正例の埋め込み, $\{z_i^{-}\}$ は負例の埋め込み集合, τ は温度パラメータを表す. $\mathcal{L}_{\text{C}}$ は $\text{sim}(z^{\text{q}}, z^{+})$ を大きく, $\text{sim}(z^{\text{q}}, z^{-})$ を小さくするようにモデルを最適化する. 我々は $sim(z^{\text{T}}, z_{\text{GA}}^{\text{I}})$ を高める損失 **Intra Loss** を $\mathcal{L}_{\text{intra}} = \mathcal{L}_{\text{C}}(z^{\text{T}}, z_{\text{GA}}^{\text{I}}, \{z_{i \neq \text{GA}}^{\text{I}}\})$ と定義し, 図 3 に示すように, 対照学習を行う. また, 論文ごとに図の枚数 n が異なるため, 学習時は m 枚の図を無作為抽出し, 不足分はゼロ埋めで補完する.

拡張戦略 ベースラインの検索能をさらに高めるため, 以下の戦略を提案する: 1) 各ミニバッチに含まれる B 枚の論文の各 Abstract の埋め込みを $z^{\text{T},(1)}, z^{\text{T},(2)}, \ldots, z^{\text{T},(i)}, \ldots, z^{\text{T},(B)}$, 各 GA の埋め込みを $z_{GA}^{\text{I},(1)}, z_{GA}^{\text{I},(2)}, \ldots, z_{GA}^{\text{I},(i)}, \ldots, z_{GA}^{\text{I},(B)}$ とする. ここで, 各論文の $z^{\text{T},(i)}$ と $z_{GA}^{\text{I},(i)}$ のペアを識別可能とする損失 **Inter Loss** を $\mathcal{L}_{\text{inter}} = (\mathcal{L}_{\text{C}}(z_i^{\text{T},(i)}, z_{GA}^{\text{I},(i)}, \{z_{GA}^{\text{I},(j \neq i)}\}) + \mathcal{L}_{\text{C}}(z_{GA}^{\text{I},(i)}, z^{\text{T},(i)}, \{z^{\text{T},(j \neq i)}\}))/2$ と定義する. 最終的な損失は, 係数 λ を用いて $\mathcal{L}_{\text{intra+inter}} = \mathcal{L}_{\text{intra}} + \lambda \mathcal{L}_{\text{inter}}$ と定義し, これを学習に用いる. 2) 各研究分野の論文を B 枚無作為抽出し, 同一分野の論文で各ミニバッチを構築する **D**omain-**C**onsistent Batch **S**ampling (**DCS**) を導入する. 3) I_i の埋め込み z_i^{I} と対応するキャプション C_i の埋め込み z_i^{C} を統合した埋め込み **C**aption **M**erged **E**mbedding (**CME**) を $z_i^{\text{CME}} = z_i^{\text{I}} \bigodot z_i^{\text{C}}$ と定義し, I_i を $\text{sim}(z^{\text{T}}, z_i^{\text{CME}})$ で評価して学習・推論を行う.

3.3 検索の曖昧さを考慮した評価指標

論文内には実際に注釈された Ground Truth (GT) の他にも GA として妥当な図が存在しうる. ゆえに, Recall@k (R@k) などの従来の評価指標のみではモデルを適切に評価できない. この課題に対し, 我々は次式で定義される **C**ertainty **A**djusted top1-GT **R**atio@k (**CAR@**k) を提案する:

$$\text{CAR@}k = \frac{p_{\text{GT}}}{p_1}\left[1 - \frac{1}{2}\max\left(0, \frac{2H(\boldsymbol{P}) - \log k}{\log k}\right)\right] \quad (2)$$

$\boldsymbol{P} \in \mathbb{R}^k$ は Top k 候補の評価値を z-score で標準化し, softmax 関数で確率に変換した系列, $p_1 \in \boldsymbol{P}$ は Top1 候補の確率, $p_{\text{GT}} \in \boldsymbol{P}$ は GT として注釈された候補の中で最上位にランク付けされた候補の確率, $H(P)$ は $\boldsymbol{P}$ の平均情報量を表す. 式 2 の角括弧内はモデルの確信度と解釈でき, 値域は $[0.5, 1.0]$ である. $(\log k)/2 < H(\boldsymbol{P}) \leq \log k$ の区間では, $H(\boldsymbol{P})$ が大きいほど (=特定候補に対する確信が弱いほど) 小さな値をとる. $0 \leq H(\boldsymbol{P}) \leq (\log k)/2$ の区間では 1.0 となる. CAR はこの確信度で Top1 候補と GT の確率比 p_{GT}/p_1 を調整した値であり, 強い確信を持ち GT を上位と判断した場合は 1.0, 下位と判断した場合は 0.0 に近づき, 確信が弱い場合は 0.5 付近の値をとる. また, 1 つのクエリに対する $\boldsymbol{P}$ のみに基づくため, クエリごとにモデルの振る舞いを定量評価できる.

4 評価実験

SciGA-140k の GA または同様の役割を持つ図を含む情報科学分野の論文 20,520 件を学習, 検証, 評価用に 8:1:1 で分割し, Abst2GA Retrieval ($\tau = 0.07, m = 7$) を実施した. バックボーンモデルには最大 248 トークンの入力長を持ち, Abstract 全体を入力可能な Long-CLIP [18] を用いた. $\mathcal{L}_{\text{intra+inter}}$ を損失とする拡張戦略では, $\lambda = 0.2$ とした. また, 提案タスクの有効性を確認するため, 以下の 2 つの手法と比較を行った: 1) T と C_i の BLEU ($N = 4$), ROUGE-2 を I_i の評

表 1: 各手法の推薦能と拡張戦略のアブレーション評価.

アプローチ	$\mathcal{L}_{\text{intra+inter}}$	DCS	CME	R@1	R@2	R@3	MRR	CAR@5	Accuracy	F1
単語一致率 (BLEU)	–	–	–	0.351	0.570	0.712	0.562	0.382	–	–
単語一致率 (ROUGE)	–	–	–	0.447	0.668	0.782	0.636	0.479	–	–
確率推定 (EfficientNetV2 [17])	–	–	–	0.449	0.674	0.797	0.643	0.486	0.639	0.561
確率推定 (CLIP 画像エンコーダ [11])	–	–	–	0.493	0.708	0.826	0.675	0.518	0.621	0.548
Abst2GA Retrieval (ours)				0.575	0.783	0.877	0.735	0.573	–	–
			✓	0.637	0.826	0.914	0.778	0.615	–	–
		✓		0.577	0.786	0.882	0.737	0.576	–	–
		✓	✓	0.635	0.827	0.914	0.778	0.616	–	–
	✓			0.557	0.769	0.870	0.723	0.562	–	–
	✓		✓	0.638	0.836	0.914	0.781	0.620	–	–
	✓	✓		0.563	0.774	0.874	0.727	0.564	–	–
	✓	✓	✓	**0.644**	**0.839**	**0.918**	**0.785**	**0.623**	–	–

価値とするとする手法 [19]. 2)EfficientNetV2 [17] および, CLIP 画像エンコーダに MLP を連結したモデルを用いて, I_i が GA か否かで分類する 2 クラス分類問題を解き, 評価値となる確率を推定する手法. 各モデルの推薦能は R@k, MRR, CAR@5 で評価し, 確率推定モデルの分類能は Accuracy, F1 で評価した.

inter-GA Recommendation への応用 次に, intra-GA Recommendation を目的に設計・学習された Abst2GA Retrieval のベースラインを用い, inter-GA Recommendation に対する性能を定性的に検証した. この実験では, 評価データの論文の Abstract をクエリとし, 学習データの論文の GA を検索対象とした.

4.1 結果と分析

各手法の効果 表 1 に示すように, Abst2GA Retrieval は, 単語一致率や確率推定をベースとする手法を大きく上回る推薦能を示し, ベースライン (拡張戦略なし) で R@1 が 0.575 を記録した. この結果は, 提案タスクが intra-GA Recommendation において有効なアプローチであることを示す.

3 つの拡張戦略を全て組み合わせた場合, 全指標において最良の結果を示し, R@1 は 0.644 に達した. CME はキャプションを用いて視覚的に類似した図の識別を補助し, DCS は各ミニバッチでモデルが捉える特徴の一貫性を高め, 学習の安定性を向上させたと考えられる. 一方, $L_{\text{intra+inter}}$ が持つ異なる論文間の GA を分離する特性は, 本来の目的である同一論文内での GA 検索には直接寄与せず, R@1 を 0.018 低下させた. しかし, CME と共に用いることで R@1 は 0.063 向上した. これは, CME が持つ文脈的情報が $L_{\text{intra+inter}}$ を最小化する過程で強調され, 本来の目的に適した埋め込みが得られたためと推察できる.

エラー分析 Top1 での検索の誤りを分析した結果, その特徴が明らかになった. 誤った推論の多くは, 提案手法の概要図など, GT ではないが GA として妥当な図を論文内に複数含む事例であった. これらは CAR が 0.45〜0.55 に集中し, モデルが強い確信を持てず候補間で迷っている様子が見られ, その振る舞いは妥当と評価できる. また, Introduction 内の図 1 を GT とした一部の論文では, GT が手法や成果ではなく研究背景の補足情報を示す例が確認された. これらの事例では CAR が 0.0 に近く, モデルが強い確信を持って GT を下位にランク付けする傾向にあったが, これも合理的な判断と考えられる. 一方, 実験機器の写真など, 特定分野で GA として効果的な図が低評価される事例も見られ, 分野横断的な特性に適応可能なモデル設計が求められる.

inter-GA Recommendation への応用 Abst2GA Retrieval のベースラインを他の論文の GA を検索対象として適用した結果, クエリと同一分野, 同一トピックの論文の GA が上位に検索される傾向が確認された. この結果は, 提案タスクのベースラインが inter-GA Recommendation への応用可能性を示し, 提案タスクが GA Recommendation の統一フレームワークの基盤となりうることを示唆する.

5 おわりに

本研究では, GA 設計支援を目的としたデータセット SciGA-140k と新タスク Abst2GA Retrieval を提案した. ベースラインや新評価指標 CAR で提案タスクの intra-GA Recommendation に対する有効性を確認すると共に, inter-GA Recommendation にも適用可能であることを示した. 本研究は GA 作成の効率化と科学的発見の伝達を促進する重要な一歩である.

参考文献

[1] Douglas B. Lenat. Automated Theory Formation in Mathematics. In **IJCAI**, 1977.

[2] Jeff Clune. AI-GAs: AI-generating algorithms, an alternate paradigm for producing general artificial intelligence, 2019. https://doi.org/10.48550/arXiv.1905.10985.

[3] Chris Lu, Cong Lu, Robert Lange, Jakob Foerste, Jeff Clune, and David Ha. The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery, 2024. https://doi.org/10.48550/arXiv.2408.06292.

[4] Tsu-Jui Fu, William Yang Wang, Daniel McDuff, and Yale Song. DOC2PPT: Automatic Presentation Slides Generation from Scientific Documents. In **AAAI**, 2022.

[5] Shohei Tanaka, Hao Wang, and Yoshitaka Ushiku. SciPostLayout: A Dataset for Layout Analysis and Layout Generation of Scientific Posters. In **BMVC**, 2024.

[6] Juan A. Rodriguez, David Vazquez, Issam Laradji, Marco Pedersoli, and Pau Rodriguez. FigGen: Text to Scientific Figure Generation. In **ICLR**, 2023.

[7] Jieun Lee and Jeong-Ju Yoo. The current state of graphical abstracts and how to create good graphical abstracts. **Science Editing**, Vol. 10, No. 1, pp. 19–26, 2023.

[8] Madhan Jeyaraman, Harish V. K. Ratna, Naveen Jeyaraman, Nicola Maffulli, Filippo Migliorini, Arulkumar Nallakumarasamy, and Sankalp Yadav. Graphical Abstract in Scientific Research. **Cureus**, Vol. 15, No. 9, 2023.

[9] Madhan Jeyaraman and Raju Vaishya. Attract readers with a graphical abstract – The latest clickbait. Journal of Orthopaedics. **Journal of Orthopaedics**, Vol. 38, No. 1, pp. 30–31, 2023.

[10] Ma Yuanyuan and Jiang Kevin. Verbal and visual resources in graphical abstracts: Analyzing patterns of knowledge presentation in digital genres. **Ibérica**, Vol. 46, pp. 129–154, 2023.

[11] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning Transferable Visual Models From Natural Language Supervision. 2022.

[12] Zekun Li, Xianjun Yang, Kyuri Choi, Wanrong Zhu, Ryan Hsieh, HyeonJung Kim, Jin Hyuk Lim, Sungyoung Ji, Byungju Lee, Xifeng Yan, Linda Ruth Petzold, Stephen D. Wilson, Woosang Lim, and William Yang Wang. MMSci: A Dataset for Graduate-Level Multi-Discipline Multimodal Scientific Understanding, 2024. https://doi.org/10.48550/arXiv.2407.04903.

[13] Edward Dunne and Klaus Hulek. Mathematics subject classification 2020. **EMS Newsl**, Vol. 115, pp. 5–6, 2020.

[14] Deyan Ginev. ar5iv:04.2024 dataset, an HTML5 conversion of arXiv.org, 2024. SIGMathLing – Special Interest Group on Math Linguistics.

[15] Leland McInnes, John Healy, Nathaniel Saul, and Lukas Grossberger. UMAP: Uniform Manifold Approximation and Projection. **The Journal of Open Source Software**, Vol. 3, No. 29, p. 861, 2018.

[16] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. **arXiv preprint arXiv:1807.03748**, 2018.

[17] Mingxing Tan and Quoc V. Le. Efficientnetv2: Smaller Models and Faster Training. 2021.

[18] Beichen Zhang, Pan Zhang, Xiaoyi Dong, Yuhang Zang, and Jiaqi Wang. Long-CLIP: Unlocking the Long-Text Capability of CLIP. In **ECCV**, 2024.

[19] Shintaro Yamamoto, Yoshihiro Fukuhara, Ryota Suzuki, Shigeo Morishima, and Hirokatsu Kataoka. Automatic Paper Summary Generation from Visual and Textual Information. In **ICMV**, 2018.

[20] "Kyle Lo, Lucy Lu Wang, Mark Neumann, Rodney Kinney, and Daniel Weld. S2ORC: The Semantic Scholar Open Research Corpus. In **ACL**, 2020.

[21] Tarek Saier, Johan Krause, and Michael Färber. unarXive 2022: All arXiv Publications Pre-Processed for NLP, Including Structured Full-Text and Citation Network. In **JCDL**, 2023.

[22] Juan A. Rodriguez, David Vazquez, Issam Laradji, Marco Pedersoli, and Pau Rodriguez. OCR-VQGAN: Taming Text-within-Image Generation. In **WACV**, 2023.

[23] Lei Li, Yuqi Wang, Runxin Xu, Peiyi Wang, Xiachong Feng, Lingpeng Kong, and Qi Liu. Multimodal ArXiv: A Dataset for Improving Scientific Comprehension of Large Vision-Language Models. In **ACL**, 2024.

表 2: 従来データセットとの比較

	Supports Full-Text	Supports Figures	GA Annotation	#SeedPapers	#Figures	#Subjects
S2ORC [20]	✗	✗	✗	81.1M	N/A	1
unarXiv2022 [21]	✓	✗	✗	1.9M	N/A	155
Paper2fig100k [22]	✗	✓	✗	69K	102K	4
ArxivCap [23]	✗	✓	✗	572K	6.4M	32
MMSci [12]	✓	✓	✗	131K	742K	72
SciGA-140k (ours)	✓	✓	✓	145K	1.1M	5,656

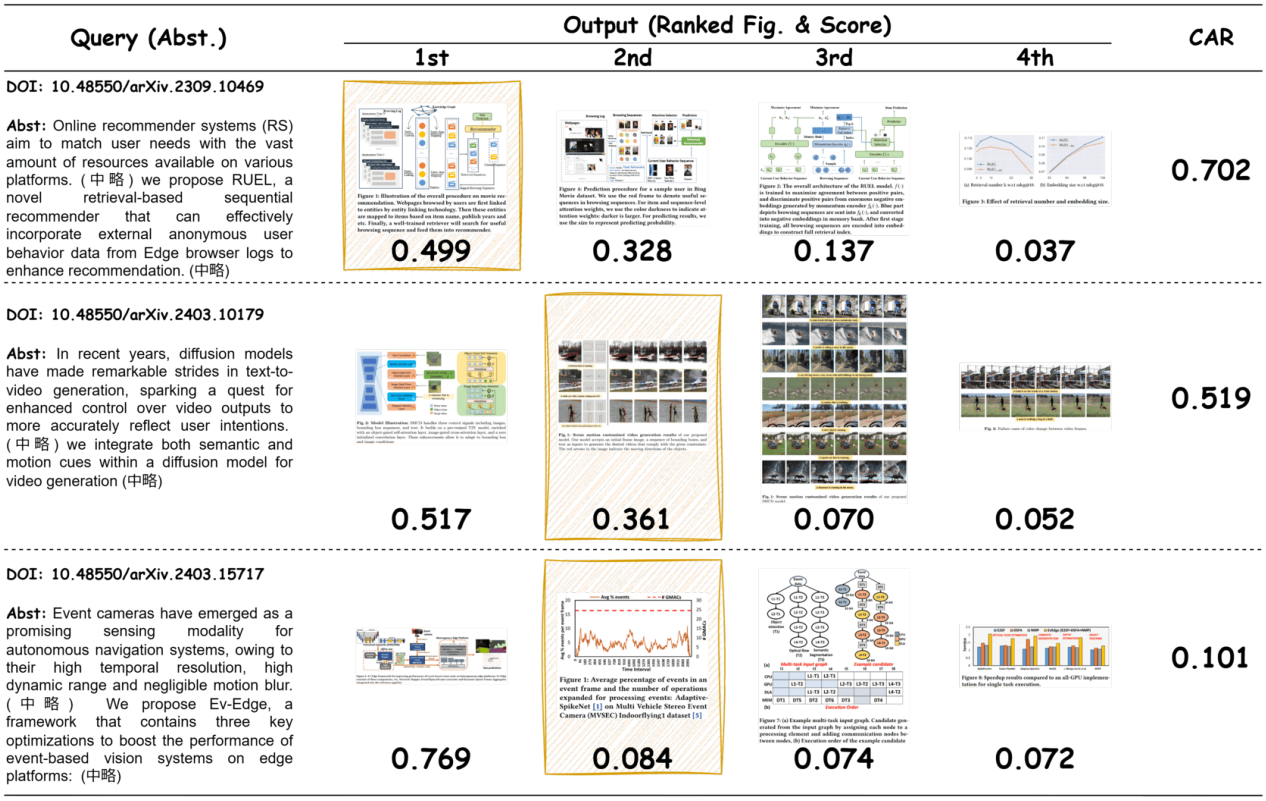


図 4: Abst2GA Retrieval の Top4 での検索結果と各図のスコアの例.[4] GT は黄枠で示す.

A 付録

従来のデータセットとの比較 表 2 に示すように, SciGA-140k は本文と図の両方をサポートするデータセットとして最大規模であり, GA に関する注釈が付与された初のデータセットである。また, 各論文は 5,656 種に細分化された研究分野や採択会議などの多様なメタデータを含み, 従来のデータセットに比較して詳細かつ多角的な分析を可能にしている.

Abst2GA Retrieval の例 図 4 に Abst2GA Retrieval の具体的な検索結果の例を示す. 上位に選ばれた図の多くは, モデルの概要図や従来手法との結果を比較する図であり, 一般的に GA として妥当な図と言える. 一方, グラフや詳細な実験結果を示す図などは低い評価値を得ていることが観察され, モデルが GA としての妥当性を適切に学習していることを示す. また, GT が 2 位で選ばれている場合でも, 1 位と GT のスコア差が小さく検索結果全体の中で上位に位置するならば, CAR が 0.5 程度となった. 一方, 同様に GT が 2 位で選ばれている場合でも, 1 位の評価値が極めて高く GT の評価値が低い場合, CAR は小さくなることが確認できた. この結果は, CAR が検索の曖昧さを定量的に評価する上で有効な評価指標であることを示す.

4) 上: https://doi.org/10.48550/arXiv.2309.10469
中: https://doi.org/10.48550/arXiv.2403.10179
下: https://doi.org/10.48550/arXiv.2403.15717