SciGA-Vec: 学術論文における ベクタ画像形式の Graphical Abstract データセット

川田拓朗 [1]　北田俊輔 [1]　彌冨仁 [1]
[1] 法政大学大学院理工学研究科
takuro.kawada.3g@stu.hosei.ac.jp, shunsuke.kitada.0831@gmail.com
iyatomi@hosei.ac.jp

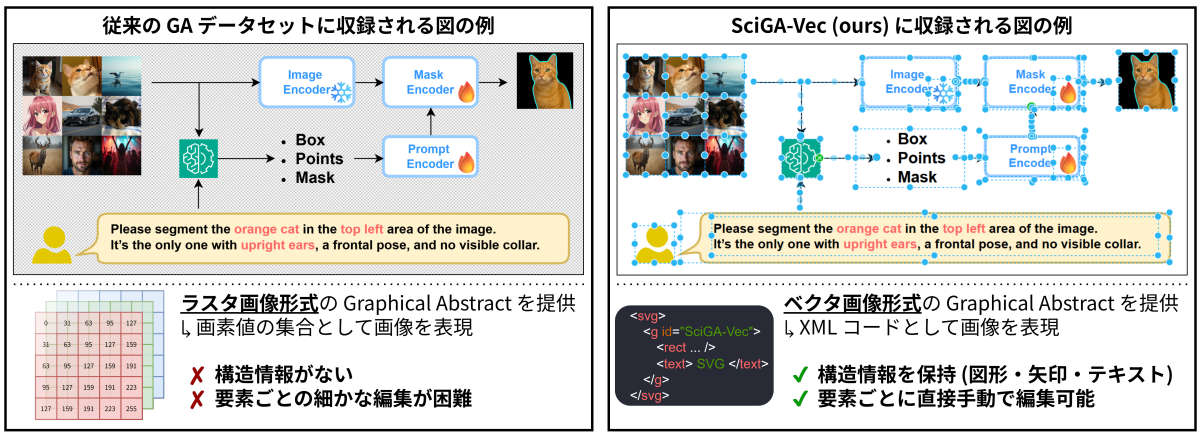


図 1: 従来のラスタ形式 GA データセット (左) と, 我々のベクタ形式 GA データセット SciGA-Vec (右) の比較.

概要

Graphical Abstract (GA) は，論文の主要な貢献を視覚的に伝える重要な表現である. 既存の GA データセットでは画素の集合として画像を表現するラスタ画像としてのみ提供されてきたため, 矩形や矢印, テキストといった構造情報を活用した GA の解析や編集, 生成に関する研究は制限されていた. 本研究では, ラスタ画像を構造情報を保持したベクタ画像へ変換するパイプラインを提案し, 既存の GA データセットに適用することでベクタ形式 GA データセット SciGA-Vec を構築する. SciGA-Vec は，GA の構造的解析や編集可能な生成モデルの学習, 評価を可能とし, GA を構造化インフォグラフィックとして扱う将来の研究の基盤を提供する.

1　はじめに

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. 矩形, 矢印, アイコン, テキストなどを組み合わせたインフォグラフィックであり, 論文の注目度を高めるだけでなく読者に概要を迅速に伝える [1, 2]. 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高度なデザインスキルが求められる [3, 4]. 近年では, 研究スライド [5] や研究ポスターの自動生成 [6, 7] など, 科学的伝達を効率化する支援技術が注目されている. GA の作成支援においても, ポスター生成モデルやレイヤー構造を考慮した画像生成モデルの応用が期待される.

GA のようなインフォグラフィックは一般的に画素の集合として表現されるラスタ画像ではなく編集可能なベクタ画像として作成, 編集される. ラスタ画像は Portable Network Graphics (PNG) などに代表され, 矩形要素, 矢印の向き, テキスト領域, 階層構造といった構造情報を明示的に保持することができない. 一方, ベクタ画像は Scalable Vector Graphics (SVG) などに代表され, 図形やテキストなどの要素を明示的に構造化した Extensible Markup Language (XML) などのテキストコードとして表現される. また, Microsoft PowerPoint や Adobe Illustrator といったドローイングツール上で, 図形の位置や色, テキストの内容やフォントといった構造や属性を直接編集可能である. しかし, GA が公開される段階では, 多くの場合ベクタ画像から不可逆的にラスタ画像へ変換され, 作成時に保持していた構造情報の大部分が失わ

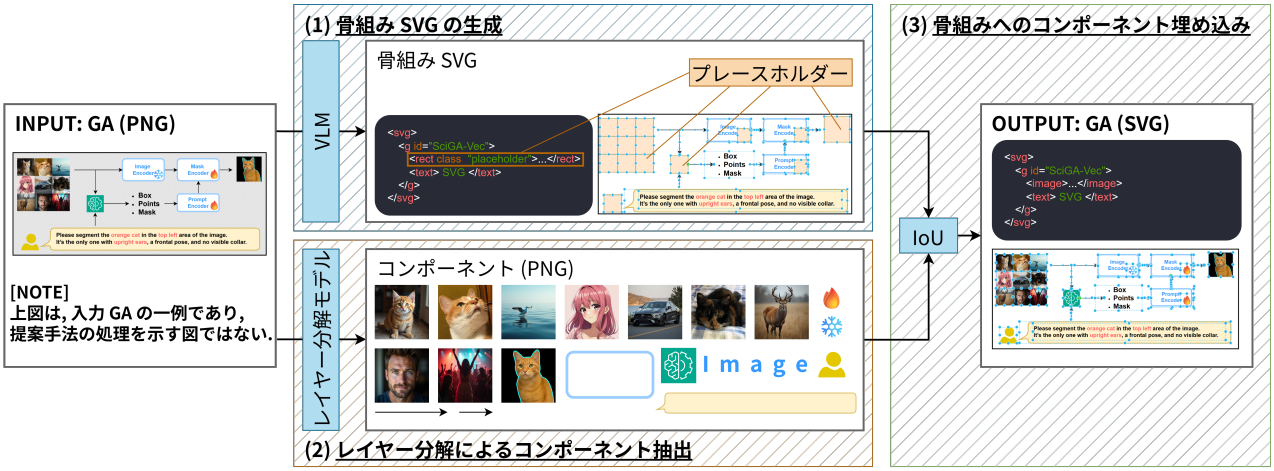


図 2: 提案するベクタライズパイプラインの概要.

れる. インフォグラフィック生成モデル [8, 9] や関連データセット [10, 11] の多くもラスタ画像を前提としており, ドローイングツールへの取り込みや画像の差し替え, 要素ごとの微調整は困難である. レイヤーやレイアウト, テキストを考慮したラスタ画像編集手法 [12] も存在するが, 失われた構造を厳密に制御することは困難であり，編集後の画像に境界のにじみや欠損が生じる場合が多い.

本研究ではラスタ画像をベクタ画像へ変換するパイプラインを提案し, このパイプラインを GA データセット SciGA-145k [10] に適用することで大規模ベクタ GA データセット SciGA-Vec を構築する (図 1). 本パイプラインでは, Vision-Language Model (VLM) を用いて GA の構造を推定し, SVG コードとして GA を再構成する. ここで, アイコンや科学的文脈において再構成すべきでない一部の画像などは元のラスタ画像から直接切り取ってコード内に埋め込む. また, 構築された SciGA-Vec は矩形, テキストなどの構造情報を保持する初のインフォグラフィックデータセットである. 我々は矩形や矢印などの構造パターン分析を行い, ラスタ画像では把握できなかったデザイン傾向を明らかにした. 本データセットは, 将来のベクタ画像生成モデル開発の基盤として利用できるだけでなく, 新規 GA 作成時のレイアウトテンプレートとしても活用可能である.

2 提案手法と提案データセット

2.1 ベクタライズパイプライン

本研究では, ラスタ形式の GA を構造を保持した SVG として再構成するために, 図 2 に示す 3 段階のベクタライズパイプラインを採用する: (1) 骨組み SVG の生成; (2) レイヤー分解によるコンポーネント抽出; (3) 骨組みへのコンポーネント埋め込み.

(1) 骨組み SVG の生成. まず, PNG 形式の GA から, 写真やアイコンなどの視覚的に複雑な要素を抽象化した SVG を生成する. 具体的には, VLM を画像理解に基づく変換器として用い, 入力されたラスタ画像をレイアウトや縦横比, 各要素の形状や色, テキストといった構造情報を保持した XML コードに変換する. ここで, ベクタ表現による再現が困難なアイコンや写真, 科学的忠実性の観点から再構成するべきでない生データ画像などの領域は, 矩形のプレースホルダー `<rect class="placeholder">...</rect>` として置き換える. この矩形領域には後段の処理で実画像が埋め込まれ, 本段階での出力は最終出力の骨組みとして働く.

(2) レイヤー分解によるコンポーネント抽出. 次に, GA に含まれる写真やアイコン, テキスト断片, 背景枠といった視覚的, 意味的に独立した要素を個別の画像コンポーネントとして取得する. 具体的には, ラスタ画像を透過情報を含む複数のレイヤーへ分離するレイヤー分解モデルに GA を入力し, 得られた各レイヤー画像に対してさらに連結成分を抽出する. 各コンポーネントについては, 元の GA 上の対応領域を切り出した RGBA 画像に加え, その位置情報およびバウンディングボックスを取得する.

(3) 骨組みへのコンポーネント埋め込み. 最後に, (1) で生成した骨組み SVG の各プレースホルダーの矩形領域と, (2) で抽出された各コンポーネントのバウンディングボックスとの Intersection over Union (IoU) [13] を計算する. 各プレースホルダー矩形に対

表 1: SciGA-Vec と既存の関連データセットの比較.

データセット	画像形式	インフォグラフィックを対象とするか
ChartGalaxy [11]	Raster	✓ (チャート図, 表)
ArxivCap [14]	Raster	✓ (科学論文の図)
SciGA-145k [10]	Raster	✓ (GA)
OmniSVG [15]	Vector	✗ (アイコン, イラスト)
UniSVG [16]	Vector	✗ (アイコン)
SciGA-Vec (ours)	Vector	✓ **(GA)**

表 2: SciGA-Vec に含まれる各 GA 1 枚あたりの主要な構成要素の数. 各値は平均値 ± 標準偏差を示す.

# 矩形	# 矢印	# テキスト	# 画像
10.26 ± 14.77	19.84 ± 20.55	2.40 ± 3.73	4.38 ± 6.72

し, IoU が最大となるコンポーネントを選択し, SVG 内の矩形 `<rect class="placeholder">...</rect>` を `<image>...</image>` タグへ置換することで, 当該領域に画像を埋め込む. 画像は Base64 形式でにエンコードされ, SVG コード内に直接埋め込まれる. これにより, 編集可能な構造表現と元の GA に対する視覚的忠実性を両立した最終的な SVG が得られる.

2.2 SciGA-Vec

SciGA-Vec は, 2.1 節で述べたベクタライズパイプラインを, 既存の GA データセット SciGA-145k [10] に含まれる 625 件の GA に適用することで構築したベクタ形式の GA データセットである. ここで, 骨組み SVG 生成のための VLM として Gemini-3 [17], コンポーネント抽出のためのレイヤー分解モデルとして LayerD [18] を用いた.

表 1 は SciGA-Vec と既存の関連データセットとの比較を示す. GA を含むインフォグラフィックを対象とした既存のデータセットの多くは画像をラスタ形式で提供しており, 図形要素やレイアウト, 編集可能な内部構造は明示的に保持されていない. また, 従来のベクタ画像データセットは, アイコンやイラストなどを対象とするものが中心である. これに対し SciGA-Vec は, GA をベクタ形式として提供する初のデータセットである. この特性により, GA をドローイングツール上で手動編集可能な構造化データとして解析・編集・生成の対象とすることが可能となる.

表 2 に, GA を構成する主要な要素である矩形 (`<rect>`), 矢印 (`<g class="arrow">`), テキスト (`<text>`), および埋め込み画像 (`<image>`), について, SciGA-Vec に含まれる各 GA 1 枚あたりの数を示す. 多くの GA は複数の矩形要素と矢印要素を含み, 処理の流れや概念間の関係を段階的に表現する構造を持つ. また, SciGA-Vec には一定数の画像要素が含まれており, 実験結果の生データ画像, アイコンなどのベクタ画像が GA の構成要素として用いられている.

3 評価実験

我々は, 提案するベクタライズパイプラインおよびそれにより構築された SciGA-Vec の品質を評価するため, GA が本来有していた (i) 視覚的情報および (ii) 意味的情報がラスタ画像からベクタ画像へ変換された後もどの程度保持されているかを検証する. 加えて, (iii) SVG コードの表現効率をトークン長の観点から評価する. また, 提案ベクタライズパイプラインを従来のベクタライズ手法 VTracer [19] と比較する. ここで, 公開データセットである SciGA-145k [10] および SciGA-Vec に含まれる論文集合を $\mathcal{D} = \{d_i \mid i \in \{1, 2, \ldots, N\}\}$ とし, 各論文 d_i 中の GA を x_i, 再構築された SVG 形式の GA をラスタ画像としてレンダリングしたものを x_i' と表す.

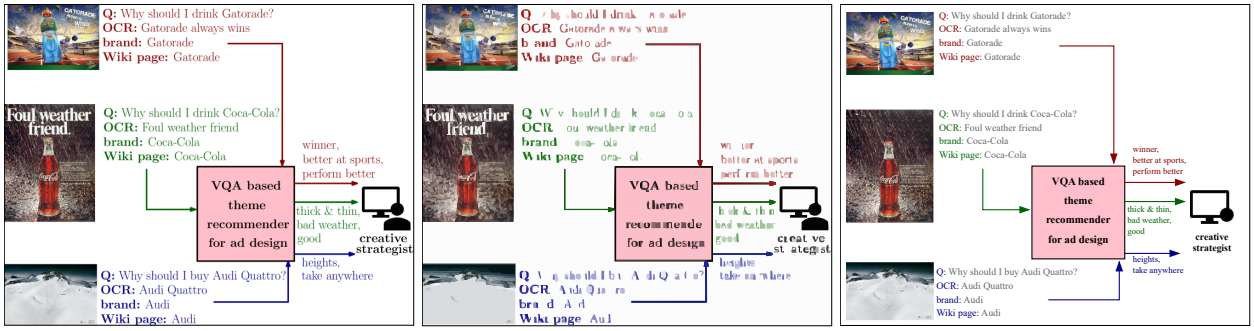
(i) 視覚的情報保持性の評価. GA のベクタライズ前後における視覚的情報の保持性を評価するため, 分布レベルの画質変化, 視覚的内容の整合性, および外観の細部差異の 3 つの観点から評価を行う. 具体的には, Fréchet Inception Distance (FID) [20], CLIP-Image (CLIP-I) [21], および PSNR [22], SSIM [23], LPIPS [24] を用いて測定する. FID は, x_i の分布と x_i' の分布の一致度を評価する. CLIP-I は, x_i と x_i' の視覚的内容の整合性を評価する. また, PSNR, SSIM, LPIPS は, x_i と x_i' の視覚的差異を画素値, 構造的特徴, 知覚的特徴で補完的に捉える. これらの指標を併用することで, 視覚的情報をどの程度保持した状態で再構築できているかを総合的に分析する.

(ii) 意味的情報保持性の評価. GA のベクタライズ前後における論文の主要貢献を伝達する能力の保持性を評価するため, GA のみを手がかりとして論文の主要貢献を答える Visual Question Answering (VQA) タスク Contribution Quiz (CQ) を定義し, その正解率に基づいた意味的情報保持性の評価指標 Semantic Retention Score (SRS) を導入する.

以下に, CQ タスクの構成手順を示す. まず, 各論文 d_i の全文を Gemini-3 [17] に入力し, 論文中の主要貢献を重要度順に抽出する. そのうち, 上位最大 M 件を選択し, 評価に用いる主要貢献文集合 $\mathcal{C}_i = \{c_i^{(j)} \mid j \in \{1, 2, \ldots, \min(M, m_i)\}\}$ を構成する.

表 3: 各ベクタライズ手法で再構成された GA の品質評価.

Method	(i) 視覚的情報保持性					(ii) 意味的情報保持性		(iii) コードのトークン長	
	FID ↓	CLIP-I ↑	PSNR ↑	SSIM ↑	LPIPS ↓	SRS (Gemini-3) ↑	SRS (GPT-5) ↑	gemma-2b ↓	o200k_base ↓
VTracer [19]	**31.654**	0.876	**33.912**	**0.878**	**0.172**	**0.936**	0.934	520,645	687,300
ours	44.210	**0.903**	33.324	0.573	0.348	0.926	**0.946**	**421,214**	**403,273**



(a) オリジナル　(b) VTracer [19]　(c) ours

図 3: 著者が作成した元の GA [1] と各ベクタライズ手法で再構成された GA.

ここで m_i は抽出された貢献文の総数であり, 論文ごとに異なる. 次に, 各主要貢献文 $c_i^{(j)}$ を正解文とする単一選択型の CQ 問題を構成するため, 論文 d_i 以外の全論文から抽出された貢献文集合 $\bigcup_{k\neq i}\mathcal{C}_k$ から無作為に K 個の文を抽出し, 負例文集合 $\tilde{\mathcal{C}}_i^{(j)}$ を構成する. これらを用い, 選択肢集合 $\mathbb{Q}_i^{(j)} = \{c_i^{(j)}\} \cup \tilde{\mathcal{C}}_i^{(j)}$ を作成する. CQ 問題では, VLM に GA x_i あるいは x_i' と $\mathbb{Q}_i^{(j)}$ を与え, GA が示す論文の貢献として最も適切な文を 1 つ選択させる.

本実験では $M = 3$, $K = 3$ とし, 各論文ごとに 3 問の 4 択 CQ 問題を構成する. VLM として Gemini-3 [17] および GPT-5 [25] を用い, このタスクを元の GA x_i および再構成 GA x_i' の双方に対して実行し, それぞれの正解率を $\mathrm{Acc}_{\mathrm{orig}}$, $\mathrm{Acc}_{\mathrm{recon}}$ とする. Semantic Retention Score (SRS) は, 再構築後の GA における意味理解性能を元の GA を入力した際の性能で正規化した指標であり, 次のように定義される:

$$\mathrm{SRS} = \frac{\mathrm{Acc}_{\mathrm{recon}}}{\mathrm{Acc}_{\mathrm{orig}}}.$$

SRS が 1.0 に近いほど, テキスト内容や矢印方向などの意味的手がかりが保持されていることを示す.

(iii) コードのトークン長の評価. 再構成された GA の表現効率, および Large Language Model (LLM) や VLM で処理することを想定した際の計算量を評価する指標として, SVG コードのトークン長を測定する. トークナイザとして, Gemini 系列のモデルで用いられる gemma-2b [26] と, GPT 系列のモデルで用いられる o200k_base [27] を採用した.

3.1 結果と分析

表 3 に, VTracer と提案パイプラインの比較結果を示す. CLIP-I, PSNR, 意味的情報保持性は両手法で同程度に達した. 特に, SRS は両手法とも概ね 1.0 となり, GA の貢献伝達能力がベクタライズ後も維持されていることが確認できる. 図 3 に示すように, VTracer は視覚的情報の保持を優先し, テキストや画像をストローク (<path>) の集合で近似する. これに対し, 我々はテキストをテキスト (<text>), 画像を画像 (<image>) として埋め込み, 構造情報の保持を優先する. したがって, 提案手法ではフォントや文字色の違い, 微小な埋め込み位置のズレが生じ, FID, SSIM, LPIPS では VTracer をやや下回る傾向が見られた. 一方, 提案手法では過剰なストローク分解を回避するため, SVG のトークン長が削減され, 可読性や編集可能性の高いベクタ表現を実現している.

4 おわりに

本研究では, ラスタ形式の GA をベクタ画像として再構築するパイプラインを提案し, ベクタ形式 GA データセット SciGA-Vec を構築した. 実験により, 提案手法は簡潔な SVG コードで意味的情報を保持しつつ, 可読性・編集可能性に優れた GA を再構成できることを示した. SciGA-Vec は, GA を構造化インフォグラフィックとして扱う基盤を提供し, 今後の編集可能な GA 自動生成への応用が期待される.

1) DOI: 10.48550/arXiv.2001.07194

参考文献

[1] Yohan Kim, Ji-Eun Lee, Jeong-Ju Yoo, Eun-Ae Jung, Sang Gyune Kim, and Young Seok Kim. Seeing Is Believing: The Effect of Graphical Abstracts on Citations and Social Media Exposure in Gastroenterology & Hepatology Journals. **Journal of Korean Medical Science**, Vol. 37, No. 45, e321, 2022.

[2] Hunter Bennett and Flynn Slattery. Graphical abstracts are associated with greater Altmetric attention scores, but not citations, in sport science. **Scientometrics**, Vol. 128, pp. 3793–3804, 2023.

[3] 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.

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

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

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

[7] Wei Pang, Kevin Qinghong Lin, Xiangru Jian, Xi He, and Philip Torr. Paper2Poster: Towards Multimodal Poster Automation from Scientific Papers. In **NeurIPS**, 2025.

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

[9] Naoto Inoue, Kento Masui, Wataru Shimoda, and Kota Yamaguchi. Opencole: Towards reproducible automatic graphic design generation. In **CVPR**, 2024.

[10] Takuro Kawada, Shunsuke Kitada, Sota Nemoto, and Hitoshi Iyatomi. SciGA: A Comprehensive Dataset for Designing Graphical Abstracts in Academic Papers, 2025. https://doi.org/10.48550/arXiv.2507.02212.

[11] Zhen Li, Duan Li, Yukai Guo, Xinyuan Guo, Bowen Li, Lanxi Xiao, Shenyu Qiao, Jiashu Chen, Zijian Wu, Hui Zhang, Xinhuan Shu, and Shixia Liu. Chartgalaxy: A dataset for infographic chart understanding and generation, 2025. https://doi.org/10.48550/arXiv.2505.18668.

[12] Shengming Yin, Zekai Zhang, Zecheng Tang, Kaiyuan Gao, Xiao Xu, Kun Yan, Jiahao Li, Yilei Chen, Yuxiang Chen, Heung-Yeung Shum, Lionel M. Ni, Jingren Zhou, Junyang Lin, and Chenfei Wu. Qwen-Image-Layered: Towards Inherent Editability via Layer Decomposition, 2025. https://doi.org/10.48550/arXiv.2512.15603.

[13] Paul Jaccard. Étude comparative de la distribution florale dans une portion des Alpes et du Jura. **Bulletin de la Société Vaudoise des Sciences Naturelles**, Vol. 37, No. 142, pp. 547–579, 1901.

[14] 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.

[15] Yiying Yang, Wei Cheng, Sijin Chen, Xianfang Zeng, Jiaxu Zhang, Liao Wang, Gang Yu, Xingjun Ma, and Yu-Gang Jiang. OmniSVG: A Unified Scalable Vector Graphics Generation Model. In **NeurIPS**, 2025.

[16] Jinke Li, Jiarui Yu, Chenxing Wei, Hande Dong, Qiang Lin, Liangjing Yang, Zhicai Wang, and Yanbin Hao. UniSVG: A Unified Dataset for Vector Graphic Understanding and Generation with Multimodal Large Language Models. In **ACM MM**, 2025.

[17] Google. Gemini-3, 2025. https://ai.google.dev/gemini-api/docs/gemini-3.

[18] Tomoyuki Suzuki, Kang-Jun Liu, Naoto Inoue, and Kota Yamaguchi. LayerD: Decomposing Raster Graphic Designs into Layers. In **ICCV**, 2025.

[19] VisionCortex. VTracer, 2020. https://www.visioncortex.org/vtracer-docs.

[20] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. In **NeurIPS**, 2017.

[21] Jack Hessel, Ari Holtzman, Maxwell Forbes, Ronan Le Bras, and Yejin Choi. CLIPScore: A Reference-free Evaluation Metric for Image Captioning. In **EMNLP**, 2021.

[22] Quan Huynh-Thu and Mohammed Ghanbari. Scope of validity of PSNR in image/video quality assessment. **Electronics Letters**, Vol. 44, No. 13, pp. 800–801, 2008.

[23] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. **IEEE Transactions on Image Processing**, Vol. 13, No. 4, pp. 600–612, 2004.

[24] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In **CVPR**, 2018.

[25] OpenAI. GPT-5, 2025. https://platform.openai.com/docs/models/gpt-5.

[26] Gemma Team, Morgane Riviere, Shreya Pathak, Pier Giuseppe Sessa, Cassidy Hardin, Surya Bhupatiraju, Léonard Hussenot, Thomas Mesnard, Bobak Shahriari, Alexandre Ramé, et al. Gemma 2: Improving Open Language Models at a Practical Size, 2024. https://doi.org/10.48550/arXiv.2408.00118.

[27] OpenAI. tiktoken, 2023. https://github.com/openai/tiktoken.