拡散モデルと極小領域再割当によるアニメ彩色

古賀 荘翠[1,a)]　藤村 友貴[1]　北野 和哉[1]　前島 謙宣[2]　舩冨 卓哉[3]　向川 康博[1]

概要

アニメ制作におけるバケツ塗り彩色は，線画領域に参照画像の色を割り当てるセグメント割当問題として定式化される．拡散モデルを本問題に適用する際には，パレット制約への非忠実性や小領域における不安定性が課題となる．本研究は拡散モデルを中核に据え，再割当ネットワークを導入し，極小領域を近傍色で置き換える自動彩色手法を提案する．One-shot 設定で最先端手法と同等，もしくは上回る精度を達成した．

1. はじめに

アニメ制作における **バケツ塗り彩色** は，労働集約的であり自動化の需要が高いタスクである．カラーパレットで定められたキャラクタ固有の配色に基づいて，線画で囲まれた閉領域に色を割り当てる工程である． 本工程の自動化に向けて，セグメント彩色問題として定式化され自動彩色モデルが提案されてきた．しかし，実作品では作画様式が作品ごとに大きく異なり，さらに遮蔽・視点差・ポーズ差により参照画像との対応が崩れるため，汎用的で安定したモデルの構築は困難な課題となっている

従来研究は主に，画像間の幾何的対応に基づき参照色を伝播する**マッチングベース** [13], [17], [21] により発展してきた．これらはパレット制約に忠実で制作フローと整合しやすい一方で，構図差に対する頑健性に欠けるため，誤対応が生じやすく精度が低下する．これに対し，線画や参照画像を条件として彩色結果を生成する**拡散モデルベース手法** [11], [18], [26] は，大域的な意味整合や構図差への頑健性を示しているが，画素単位の生成を本質とするため，セグメント彩色への適用については体系的に議論されてこなかった．一方，**セグメンテーションベース手法** [13], [17], [21] は一領域一色の彩色を実現できるが，One-shot 推論が難しくキャラクタごとの調整を要する．

本研究では，セグメント彩色に拡散モデルベースの彩色手法を適用する方法を提案する．拡散モデルは大域的な意味整合や構図差への頑健性を持つため，対応が不安定な状況でも安定した配色を生成できる可能性がある．しかし，拡散モデルをそのままセグメント彩色に適用すると，線画の改変，パレット非忠実，単領域内に複数色が出力されるなどの誤りが起こる．また，ハイライトなどの極小領域における色漏れが生じる．これは，これらの領域が小さく文脈情報が乏しいため，領域の役割を推定する手がかりが不足するためである．そこで本研究では，大域的な配色推定と小領域の補正を分離する 2 段階マルチスケール推論を導入する．第一段階では大域パッチに基づいてキャラクタ全体の配色を推定し，第二段階では局所パッチ推論により小領域の彩色を補正する．さらに，パレットへの非忠実性を投票処理で補正し，極小領域については近傍領域へ対応付ける軽量領域割当ネットワークを導入する．提案手法は，実アニメデータを用いた One-shot 設定でセグメント精度により汎化性能を評価した結果，最先端手法と同等，もしくは上回る精度を達成した．

2. 関連研究

2.1 セグメンテーションベース手法

画素単位の色分類と領域投票により彩色を行う手法が提案されている [13], [17], [20], [21]．これらの手法は高い精度を示す一方で，キャラクタ固有の学習や色クラス設計に依存する傾向があり，One-shot 推論ではなく追加学習やファインチューニングを必要とする．

2.2 マッチングベース手法

線画と参照画像間の領域同士を対応付けて伝播する手法は，セグメント彩色と親和性が高い．グラフベース手法では，領域をノードとしたグラフを構築し，形状・位置・隣接関係の整合性から対応を推定する [9], [10], [12], [19]．特徴マッチング手法では，領域ごとの特徴の類似度に基づいて対応関係を推定し，参照フレームから色を伝播する [1], [2], [3], [7], [8]．Kono *et al.* は，バケツ塗り後に生じる未塗り領域を対象とし，その領域色を U-Net により尤度を推定し，最も尤度の高い近傍領域の色を選択することで未塗り領域を補完する手法を提案した [6].

[1] 奈良先端科学技術大学院大学
[2] OLM Digital, Inc.
[3] 京都大学
[a)] koga.sosui.ko9@naist.ac.jp

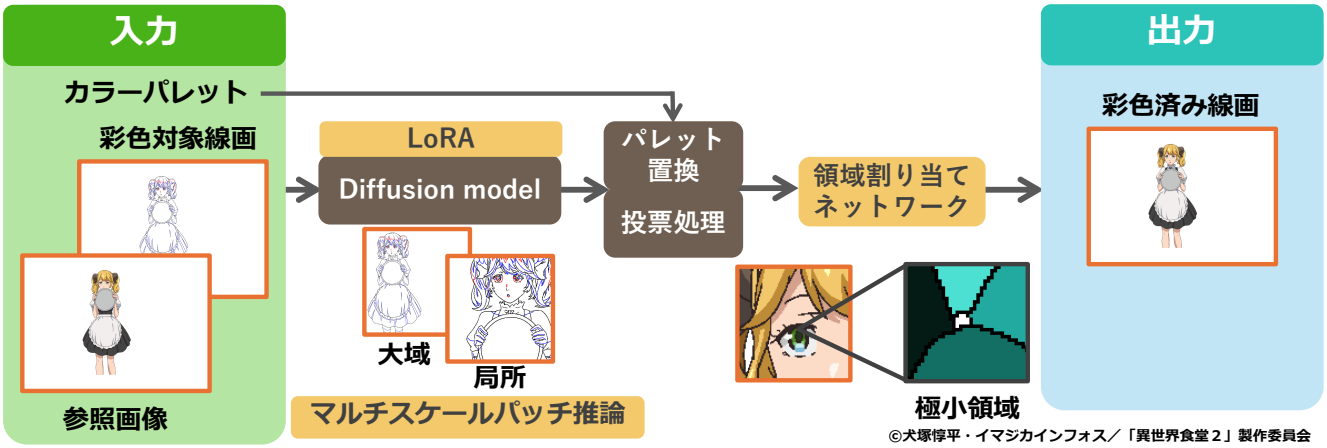

図 1: 推論パイプラインの概要

2.3 拡散モデルベース手法

参照画像および線画を条件として，拡散モデルを用いて彩色画像を生成する手法も提案されている [5], [11], [14], [16], [18], [23], [24], [25], [26]．この系統は大きな構図差にも対応可能である一方，参照画像に存在しない線画や色が生成され，領域単位の彩色が行われず，領域単位精度が低いことが報告されている [15]．

3. 提案手法

図 1 に提案する推論パイプラインの概要を示す．

3.1 問題設定

本研究の目的は，参照画像のキャラクタ配色を反映しつつ領域への色割り当てを決定することである．入力は，参照画像として彩色済みの線画画像 R 1 枚と，彩色の推論対象となる線画 L およびとカラーパレットである．本研究では，これらを用いた One-shot 設定で推論を行う．なお，推論時に対象キャラクタごとの追加学習を行わない．

3.2 拡散モデルを用いたマルチスケール推論

提案法では拡散モデルに LoRA[4] を適用する，参照画像のキャラクタ全体とそこから切り出された局所パッチ画像の複数枚入力による条件付き生成により，アニメドメインの彩色タスクを学習する．また，推論時には安定性のため固定解像度のパッチで推論を行い，2 種類のパッチサイズを用いた二段階のマルチスケール推論を行う．

第一段階: 大域パッチ推論

線画 L と参照画像 R から対象領域 L_c, R_c を抽出する．白背景を含めたまま入力すると有効信号が弱まるため，前景領域をクロップする．次に，クロップした線画 L_c を固定サイズのパッチ集合 $\{L_p\}_{p\in P}$ に分割する．ここで P はパッチのインデックス集合であり，L_p は p 番目の線画パッチを表す．各線画パッチ L_p に対して，クロップした参照画像 R_c 全体を条件として拡散モデル f を適用し，得られた出力を再構成することで第一段階の推論結果 $\hat{I}_1$ を得る．

$$\hat{I}_1 = \mathrm{Merge}\Big(\{\, f(L_p, R_c)\}_{p\in P}\Big)$$

第二段階: 局所パッチ推論

第一段階出力 $\hat{I}_1$ では小領域に色漏れや境界崩れが残る場合があるため，第二段階では，面積 A_{small} 未満の領域を対象して補正を行う．参照画像 R_c をパッチ集合 $\{P_i\}$ に分割し，各線画パッチ L_p に対して最も近い参照パッチ P_p^* を選ぶ．近さは LPIPS 距離により測る．

$$P_p^* = \arg\min_{P_i} \mathrm{LPIPS}(L_p, P_i)$$

線画パッチ L_p，選び出された参照パッチ P_p^*，およびクロップ済み参照画像 R_c を第二段階推論の入力とする．拡散モデルにより各パッチの出力を得て，合成して局所パッチ結果 $\hat{I}_2$ を生成する．

$$\hat{I}_2 = \mathrm{Merge}\Big(\{\, f(L_p, P_p^*, R_c)\,\}_{p\in P}\Big)$$

最終結果 $\hat{I}$ は，第一段階の結果 $\hat{I}_1$ のうち小領域に対応する部分を $\hat{I}_2$ により置き換えることで得る．

3.3 パレット置き換えと投票処理

拡散モデルの出力はパレット色に近いが完全には一致しないため，近傍パレット色への置換を行う．ただし単純な最近傍置換では一領域一色にならない場合があるため，領域ごとに投票処理を行って単一色を決定する．

図 2 に，拡散モデルの生出力，パレット置換後，および領域投票処理後の結果の比較を示す．投票処理により，領域内の色のばらつきが抑えられ，領域単位で一貫した彩色が得られることが分かる．

3.4 極小領域に対する領域割当ネットワーク

拡散モデルによる局所パッチ推論後も，極小領域では線

画色からの色漏れや変色の影響が残りやすい．そこで，多くの極小領域は周辺のいずれかの領域に対応するという経験的仮定に基づき，極小領域を周辺領域へ割り当て，割当先領域の代表色をコピーする後処理ネットワークを導入する．図 3 に提案する領域割当ネットワークのアーキテクチャを示す．U-Net 型の軽量 CNN g_θ により行う．

Kono *et al.* は各画素で極小領域と同色となる確率を推定する尤度マップベースの手法である．一方，本研究は極小領域の再割当を領域選択問題として扱い，領域特徴間の類似度に基づいて候補領域を選択する点で異なる [6]．

極小領域の抽出と候補領域集合の作成

画像の画素集合を Ω とする．線画から得られる領域マップを $\mathrm{seg} : \Omega \to \mathcal{S}$ とし，$\mathcal{S}$ を領域の集合とする．各領域 $s \in \mathcal{S}$ に属する画素集合を

$$\Omega_s = \{(x, y) \in \Omega \mid \mathrm{seg}(x, y) = s\}$$

と定義する．領域面積が閾値 A_{tiny} 以下の領域集合を $\mathcal{R}_{\mathrm{tiny}} \subset \mathcal{S}$ とする．各 $r \in \mathcal{R}_{\mathrm{tiny}}$ に隣接する領域から候補領域集合 $\mathcal{C}(r) \subset \mathcal{S}$ を構成し，候補数は最大 K に制限し，極小領域同士は候補から除外する．

領域特徴に基づく割り当て

サイズ $t \times t$ のパッチの RGB と線画グレースケールを連結したものを $X \in \mathbb{R}^{4\times t\times t}$ とし，$g_\theta(X)$ から得られる特徴マップ $\mathbf{F} \in \mathbb{R}^{D\times H\times W}$ とする．領域 s の特徴表現である領域特徴 $\mathbf{f}_s \in \mathbb{R}^D$ を

$$\mathbf{f}_s = \frac{1}{|\Omega_s|} \sum_{(x,y)\in\Omega_s} \mathbf{F}(x, y)$$

と計算する．次に，極小領域 r の特徴 $\mathbf{f}_r$ と候補領域 $c \in \mathcal{C}(r)$ の特徴 $\mathbf{f}_c$ のコサイン類似度に基づき，候補領域 c が極小領域 r と同一色となる条件付き確率を

$$p(c \mid r) = \frac{\exp\left(\alpha \frac{\mathbf{f}_r \cdot \mathbf{f}_c}{\|\mathbf{f}_r\| \|\mathbf{f}_c\|}\right)}{\sum_{c' \in \mathcal{C}(r)} \exp\left(\alpha \frac{\mathbf{f}_r \cdot \mathbf{f}_{c'}}{\|\mathbf{f}_r\| \|\mathbf{f}_{c'}\|}\right)}$$

と計算する．ここで α はスケーリング係数である．分類確率から最尤の候補領域をコピー元として選択する．

4. 実験

実装詳細

拡散モデルとして Qwen-Image[22] を用いて LoRA による追加学習を行った．学習データはアニメ『ぽんのみち』から抽出した 3660 枚の画像を用いた．最適化には AdamW を使用し，学習率は 5×10^{-5}，学習エポック数は 1 とした．

マルチスケール推論におけるパッチサイズを第一段階では 1024×1024 で推論し，第二段階では 384×384 とした．また，小領域の判定には面積閾値 $A_{\mathrm{small}} = 50$ を用いた．

再割当ネットワークの学習には，同データセットから抽出した極小領域を中心とする 128×128 のパッチを用いた．

図 2: 推論時の中間出力

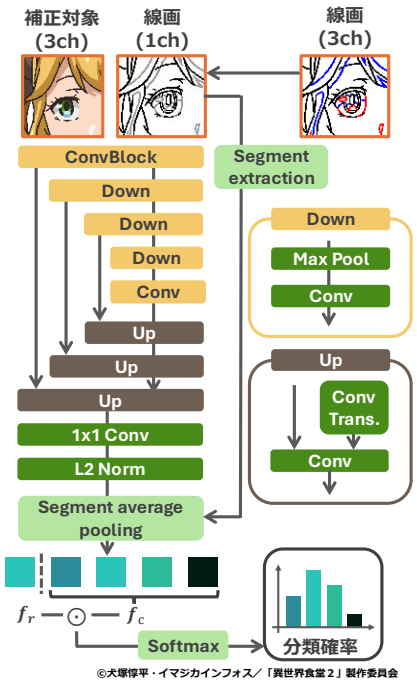

図 3: 領域割り当てネットワーク

損失関数にはクロスエントロピー損失を採用した．最適化には AdamW を用い，学習率は 1.0×10^{-4}，学習エポック数は 10 とした．極小領域のサイズ閾値は $A_{\mathrm{tiny}} = 5$ とし，各領域に対する候補数は最大 $K = 16$ に制限した．また，分類スコアにはスケール係数 $\alpha = 16$ を適用した．

検証データセット

実アニメデータセットにおける精度評価として，Maejima *et al.* [13] の評価用データセットと同様に，複数の TV アニメ作品由来のショット集合で構成し，各ショットは平均 11 フレーム程度の短い系列からなる．系列あたり数枚のキャラクタの主要パーツが視認可能な画像を参照画像とし，時間的に近い 1 枚を推論時に与える設定とした．学習には『ぽんのみち』を用いる一方，検証に用いる各ショットは別作品から構成することで作品単位で未見の設定で汎化性能を評価した．また，セグメント彩色ベンチマークとして使用される *Paint Bucket Dataset* (PBC) でも学習と評価を行う [2]．各キャラクターシーケンスを 1 枚の参照画像を用いて彩色を行う．PBC の学習・評価データはキャラクター単位で未見となるよう分離されている．括弧内は別データセットで学習した場合の結果を表す．用いる指標は領域単位の mIOU と Acc であり領域精度を測定するものである．

比較手法

セグメント彩色の One-shot 推論が可能な最先端手法である DACoN [15] と比較を行う．比較の公平性のため，各手法はそれぞれ評価対象データセットに対応する学習データで学習したモデルを用いる．

結果

表 1 および表 2 に結果を示す．結果の丸括弧内のデータ

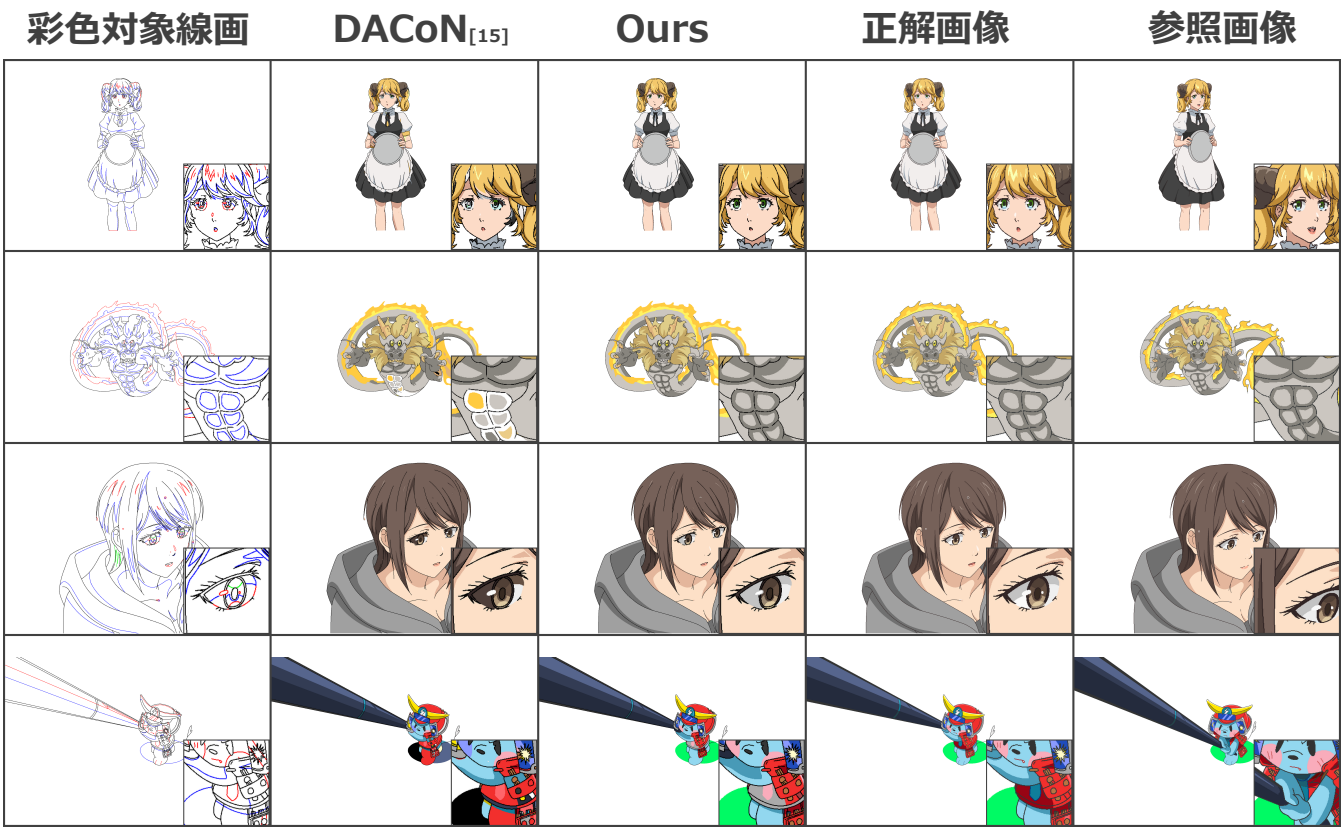

図 4: 実アニメデータセットにおける定性評価

表 1: 実アニメデータセットにおける精度評価

Method	mIOU	Acc
DACoN (PBC)	56.95	61.77
DACoN (実アニメ)	50.78	48.19
Ours w/o LoRA (実アニメ)	43.47	43.54
Ours w/o マルチスケール推論（実アニメ）	62.02	59.25
Ours w/o 再割当ネットワーク (実アニメ)	63.73	64.23
Ours (実アニメ)	**65.64**	**70.04**

表 2: PBC データセットにおける精度評価

Method	mIOU	Acc
DACoN (PBC)	56.16	**68.01**
Ours (実アニメ)	54.74	44.62
Ours w/o 再割当 (PBC)	58.30	49.48
Ours (PBC)	**60.23**	49.63

が学習に使用したデータである．提案手法は実アニメデータセットで DACoN を上回り，マルチスケール推論と再割当ネットワークの有効性も確認できる．また，PBC データセット評価においても，提案手法は PBC で学習した場合に DACoN を mIoU で上回っており，再割当ネットワークの導入により性能がさらに改善している．DACoN は PBC で高い性能を示す一方，実アニメデータセットでは性能低下が見られることから，PBC データセットの表現に比較的適合した性質を持つ可能性がある．以上より，提案手法は特定データセットへの適合に依存せず，異なるデータセットに対しても比較的安定した性能を示す点で優れている．また，実アニメデータセットにおける定性結果を図 1 に示す．右下に一部を拡大した画像を示している．小領域において安定した塗りを実現していることがわかる．

5. まとめ

本稿では，参照画像に基づく拡散モデルの大域パッチ推論と，小領域を補正する局所パッチ推論からなる二段階推論，ならびに極小領域に対する再割当ネットワークを提案した．提案手法は，作品分離された One-shot 設定でセグメント精度により汎化性能を評価した結果，最先端手法と同等，もしくは上回る精度を達成した．

謝辞

研究目的にイラストの使用許諾をいただいた犬塚惇平氏に感謝の意を表する．本研究は，経済産業省と国立研究開発法人新エネルギー・産業技術総合開発機構（NEDO）が実施する，国内の生成 AI の開発力強化を目的としたプロジェクト「GENIAC（Generative AI Accelerator Challenge）」の成果をもとに作成されました．また本研究は，JST，CRONOS，JPMJCS25K1 の支援を受けたものです．

参考文献

[1] Casey, E., Pérez, V. and Li, Z.: The Animation Transformer: Visual Correspondence via Segment Matching, *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 11303–11312 (2021).

[2] Dai, Y., Zhou, S., Li, Q., Li, C. and Loy, C. C.: Learning Inclusion Matching for Animation Paint Bucket Colorization, *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2024).

[3] Dang, T. D., Do, T., Nguyen, A., Pham, V., Nguyen, Q., Hoang, B. and Nguyen, G.: Correspondence Neural Network for Line Art Colorization, *ACM SIGGRAPH Posters* (2020).

[4] Hu, E. J., yelong shen, Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L. and Chen, W.: LoRA: Low-Rank Adaptation of Large Language Models, *International Conference on Learning Representations* (2022).

[5] Huang, Z., Zhang, M. and Liao, J.: LVCD: Reference-based Lineart Video Colorization with Diffusion Models, *ACM Transactions on Graphics (TOG)* (2024).

[6] Kono, M., Maejima, A., Koyama, Y. and Igarashi, T.: Predicting Colors in Unpainted Gaps for Anime-Style Illustration, New York, NY, USA, Association for Computing Machinery (2025).

[7] Lee, J., Kim, E., Lee, Y., Kim, D., Chang, J. and Choo, J.: Reference-Based Sketch Image Colorization Using Augmented-Self Reference and Dense Semantic Correspondence, *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5800–5809 (2020).

[8] Li, Y., Mao, J., Qiu, Q. and Matsui, Y.: Region-Wise Correspondence Prediction between Manga Line Art Images (2025).

[9] Liu, S., Wang, X., Liu, X., Wu, Z. and Seah, H. S.: Shape correspondence for cel animation based on a shape association graph and spectral matching, *Computational Visual Media* (2023).

[10] Liu, S., Wang, X., Wu, Z. and Seah, H. S.: Shape correspondence based on Kendall shape space and RAG for 2D animation, *The Visual Computer* (2020).

[11] Liu, Z., Cheng, K. L., Chen, X., Xiao, J., Ouyang, H., Zhu, K., Liu, Y., Shen, Y., Chen, Q. and Luo, P.: MangaNinja: Line Art Colorization with Precise Reference Following (2025).

[12] Maejima, A., Kubo, H., Funatomi, T., Yotsukura, T., Nakamura, S. and Mukaigawa, Y.: Graph matching based anime colorization with multiple references, *SIGGRAPH Asia Posters* (2019).

[13] Maejima, A., Shinagawa, S., Kubo, H., Funatomi, T., Yotsukura, T., Nakamura, S. and Mukaigawa, Y.: Continual few-shot patch-based learning for anime-style colorization, *Computational Visual Media* (2024).

[14] Meng, Y., Ouyang, H., Wang, H., Wang, Q., Wang, W., Cheng, K. L., Liu, Z., Shen, Y. and Qu, H.: AniDoc: Animation creation made easier (2024).

[15] Nagata, K. and Kaneko, N.: DACoN: DINO for Anime Paint Bucket Colorization with Any Number of Reference Images (2025).

[16] Ning, W., Muyao, N., Zhi, D., Zhihui, W., Zhiyong, W., Zhaoyan, M., Bin, L. and Haojie, L.: Coloring anime line art videos with transformation region enhancement network, *Pattern Recognition* (2023).

[17] Ramassamy, S., Kubo, H., Funatomi, T., Ishii, D., Maejima, A., Nakamura, S. and Mukaigawa, Y.: Pre- and post-processes for automatic colorization using a fully convolutional network, *SIGGRAPH Asia Posters* (2018).

[18] Rombach, R., Blattmann, A., Lorenz, D., Esser, P. and Ommer, B.: High-Resolution Image Synthesis with Latent Diffusion Models, *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2022).

[19] Sato, K., Matsui, Y., Yamasaki, T. and Aizawa, K.: Reference-based Manga Colorization by Graph Correspondence using Quadratic Programming, *SIGGRAPH Asia Technical Briefs* (2014).

[20] Takamura, R., Watanabe, T., Maejima, A., Shinagawa, S., Funatomi, T., Mukaigawa, Y., Morishima, S. and Kubo, H.: Fisheye Patch-Based Automatic Colorization Method for Anime Line-drawings, *Proceedings of the SIGGRAPH Asia 2025 Posters*, SA Posters '25, New York, NY, USA, Association for Computing Machinery (2025).

[21] Takano, Y., Maejima, A., Yamaguchi, S. and Morishima, S.: Anime Colorization Using Segment Matching with Candidate Colors, New York, NY, USA, Association for Computing Machinery (2025).

[22] Wu, C., Li, J., Zhou, J., Lin, J., Gao, K., Yan, K., ming Yin, S., Bai, S., Xu, X., Chen, Y., Chen, Y., Tang, Z., Zhang, Z., Wang, Z., Yang, A., Yu, B., Cheng, C., Liu, D., Li, D., Zhang, H., Meng, H., Wei, H., Ni, J., Chen, K., Cao, K., Peng, L., Qu, L., Wu, M., Wang, P., Yu, S., Wen, T., Feng, W., Xu, X., Wang, Y., Zhang, Y., Zhu, Y., Wu, Y., Cai, Y. and Liu, Z.: Qwen-Image Technical Report (2025).

[23] Yan, D., Yuan, L., Wu, E., Nishioka, Y., Fujishiro, I. and Saito, S.: ColorizeDiffusion: Improving Reference-Based Sketch Colorization with Latent Diffusion Model, *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 5092–5102 (2025).

[24] Yu, Y., Qian, J., Wang, C., Dong, Y. and Liu, B.: Animation line art colorization based on the optical flow method, *Computer Animation and Virtual Worlds* (2024).

[25] Zhang, Y., Wang, L., Wang, H., Wu, D., Lin, Z., Wang, F. and Song, L.: AnimeColor: Reference-based Animation Colorization with Diffusion Transformers, *Proceedings of the 33rd ACM International Conference on Multimedia*, MM '25, New York, NY, USA, Association for Computing Machinery, p. 6682–6690 (2025).

[26] Zhuang, J., Ju, X., Zhang, Z., Liu, Y., Zhang, S., Yuan, C. and Shan, Y.: ColorFlow: Retrieval-Augmented Image Sequence Colorization (2024).