階層的パッチによる周辺情報を考慮したアニメ線画への自動着色手法

高村 亮佑[1,a)]　渡邉 大起[1,b)]　前島 謙宣[2,c)]　品川 政太朗[3,d)]
舩冨 卓哉[3,e)]　向川 康博[3,f)]　森島 繁生[4,g)]　久保 尋之[1,h)]

概要

アニメ線画の着色は時間と手間を要する作業であり，これを自動化するために，パッチベースの手法が提案されてきた．パッチベースの手法は必要なメモリサイズが小さい一方で，従来手法のパッチは小さく限られていたため，画像全体の情報を十分に考慮することができず，着色精度の低下を招く要因となり得る．そこで本研究では，広域なコンテキスト情報を取り入れつつ，メモリ負荷を抑えた階層的パッチを用いた着色手法を提案する．提案手法では，線画を段階的にダウンサンプリングし，同一位置から同一サイズのパッチを抽出して階層構造を構成することで，局所情報と周辺情報の両方を保持する階層的パッチを生成する．評価実験の結果，提案手法は従来のパッチベース手法に比べ，特に複数のキャラクター間で視覚的に類似した領域や，線画の密度が低い領域において，より高精度かつ視覚的に一貫性のある着色を実現できることが明らかとなった．

1. はじめに

近年，日本の労働環境における課題を解決するため，業務効率化や働き方改革がさまざまな分野で進められている．アニメーション制作現場においても，アニメータの長時間労働や人材不足が深刻な課題となっており，制作工程の一部を自動化することで，作業負担の軽減や生産性の向上を図る試みが行われている．

アニメ制作工程において，線画への着色作業は集中力と手間を要する工程の一つであり，その自動化が実現すれば作業時間の短縮が大きく期待される．このような背景のもと，機械学習を用いたアニメ線画の自動着色に関する研究が進められている．

本研究では，アニメ線画に対する自動着色の精度向上を目的として，周辺情報を考慮可能なパッチ構造を導入する．高解像度の線画においては，パッチベース学習により効率的に着色処理を行える一方で，パッチ単体では周辺情報を十分に捉えることができず，文脈的な誤りが生じる可能性がある．このような課題に対し，本研究では，対象パッチに加えてその周囲の情報も活用することで，局所情報と周辺情報の双方を考慮した着色を実現する．

本論文では，その具体的な手法を提案し，従来のパッチベース手法との比較実験を通じて，精度および視覚的整合性の向上を検証する．

2. 関連研究

機械学習を用いた自動着色に関連する研究としては，アニメ作品の線画への自動着色 [1,2] 以外に，漫画やイラストへの自動着色 [3,4] や実写のグレースケール画像をカラー化する研究 [5,6] などが行われている．また，ユーザーが着色の補助となるヒントを与えるような半自動的な着色手法 [7] も提案されており，高精度で実用的な結果が得られている．

本研究で扱う着色対象であるアニメ線画については，品川ら [1] がこのタスクをセマンティックセグメンテーションとして定式化する手法を提案し，実用的な精度向上を達成している．しかし，高解像度のアニメ線画をそのまま直接ニューラルネットワークで処理することは，GPU のメモリ容量を圧迫するという課題があるため，前島ら [2] はアニメ線画の着色にパッチベース学習を導入した．パッチベース学習とは，線画を固定サイズの小領域（パッチ）に分割し，各パッチごとに個別に着色を学習させる手法である．一般的にアニメ線画は，陰影を表現する色トレス線を除けば輪郭線のみが黒，それ以外が白であるため，ランダムにパッチをサンプリングすると，パッチ内に線が全く存

[1] 千葉大学
[2] 株式会社オー・エル・エム・デジタル / 株式会社 IMAGICA GROUP
[3] 奈良先端科学技術大学院大学
[4] 早稲田大学
[a)] ryosuke3076@chiba-u.jp
[b)] twatanabe_qlab@chiba-u.jp
[c)] akinobu.maejima@olm.co.jp
[d)] sei.shinagawa@is.naist.jp
[e)] funatomi@is.naist.jp
[f)] mukaigawa@is.naist.jp
[g)] shigeo@waseda.jp
[h)] hkubo@chiba-u.jp

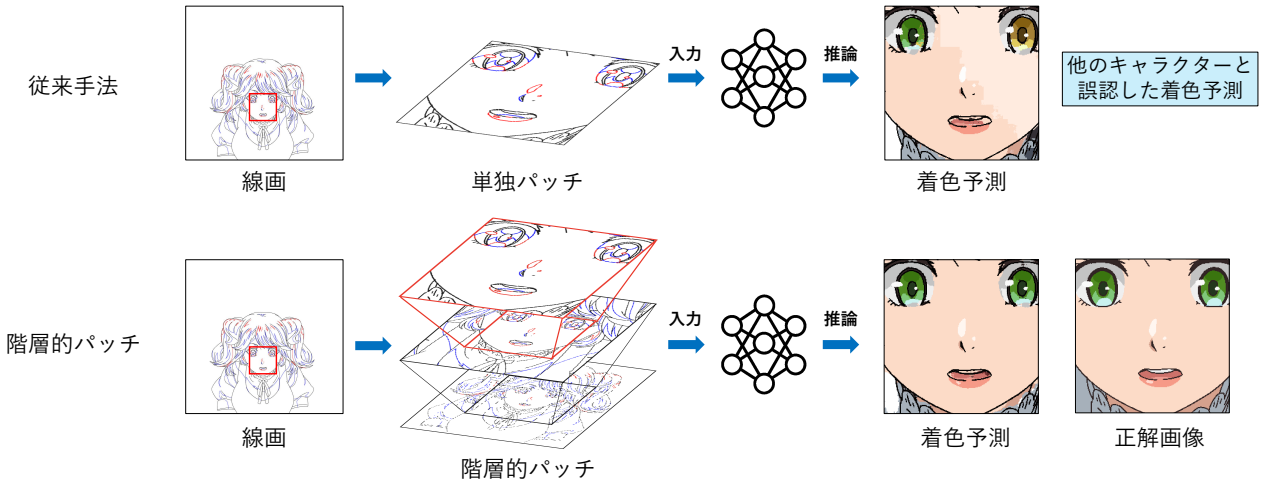

図 1　提案手法の概要

在しない場合があるという問題がある．前島らはサンプリングの仕方自体に工夫を施すことで，効率的かつ高精度な着色予測を達成した．

しかし，この方法はメモリ効率に優れる一方で，局所的な情報のみに依存するため，画像の全体的なコンテキスト情報が必要な場合には正確な着色が困難となる．たとえば，線画のみでは類似した瞳のパッチであっても，異なるキャラクターに属しており，それぞれ異なる色の瞳（例えば黄色と緑色）を持つ場合がある．このような状況では，局所情報のみに基づく処理では誤った色が割り当てられてしまうことがある．このように，パッチベース学習は高解像度線画の着色における有効な戦略ではあるものの，局所的な情報に依存しすぎることで誤った着色結果を引き起こすという課題がある．

そこで本研究では，パッチベース学習の計算効率を保ちつつ，パッチ周辺のコンテキスト情報を取り込む新たな枠組みとして，階層的パッチ学習を提案する．本手法では，対象とするパッチの位置に対し，段階的にダウンサンプリングした線画から同一サイズのパッチを抽出し，それらを重ねて階層的パッチを構成することで，局所情報と周辺情報の統合を図る．

3.　提案手法

3.1　提案手法

本論文では，アニメーション制作工程における線画の着色作業の効率化を目的として，周辺情報を活用するための階層的パッチを提案する．階層的パッチとは，図 2 に示すように元の線画から複数回にわたって線画をダウンサンプリングし，それぞれから同一位置を中心としたパッチ画像を同一サイズで切り出したものである．すなわち，階層的パッチとはいずれも同じサイズの画像であり，かつ段階的に広域な範囲を取り出した画像のセットである．高解像度

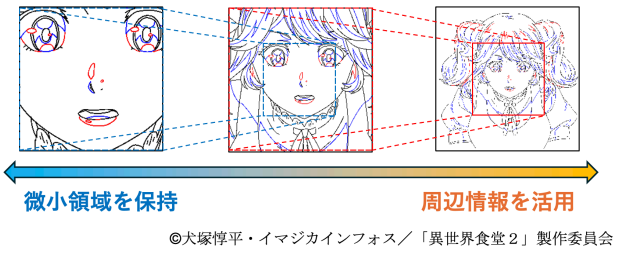

図 2　階層的パッチを構成する画像群

な画像が微小領域を保持する一方で，低解像度な画像がその周辺情報を保持する．これら複数スケールの画像をチャンネル方向に結合することで，階層的パッチを構成する．このパッチをネットワークに入力することで，ネットワークは複数の空間スケールを同時に参照できるようになり，従来の単一解像度のみのパッチでは得られなかった大域的なコンテキスト情報を補完する．

提案手法の全体の流れを図 1 に示す．まず，線画から階層的パッチを作成し，着色済み画像からはパッチ画像のみを切り出す．次に階層的パッチと着色済みのパッチ画像を組とし，データセットとして使用することでニューラルネットワークの学習を行う．学習を経て，学習済みのモデルを用いて推論を行う際には，学習時と同様に階層的パッチを入力し，パッチ画像の着色予測画像を得る．本実験ではオリジナルの解像度に加え，0.5 倍および 0.25 倍の解像度の画像を用いて，合計 9 チャンネルの入力として構成した．本研究のネットワーク構造には，FCN に一部スキップ接続を導入した FCN8s [8] を用いて検証実験を行った．

3.2　評価指標

出力される着色予測結果を定量的に評価するための指標について説明する．主な定量評価指標として，各ラベルごとの Intersection over Union (IoU) のスコアを領域ごとに

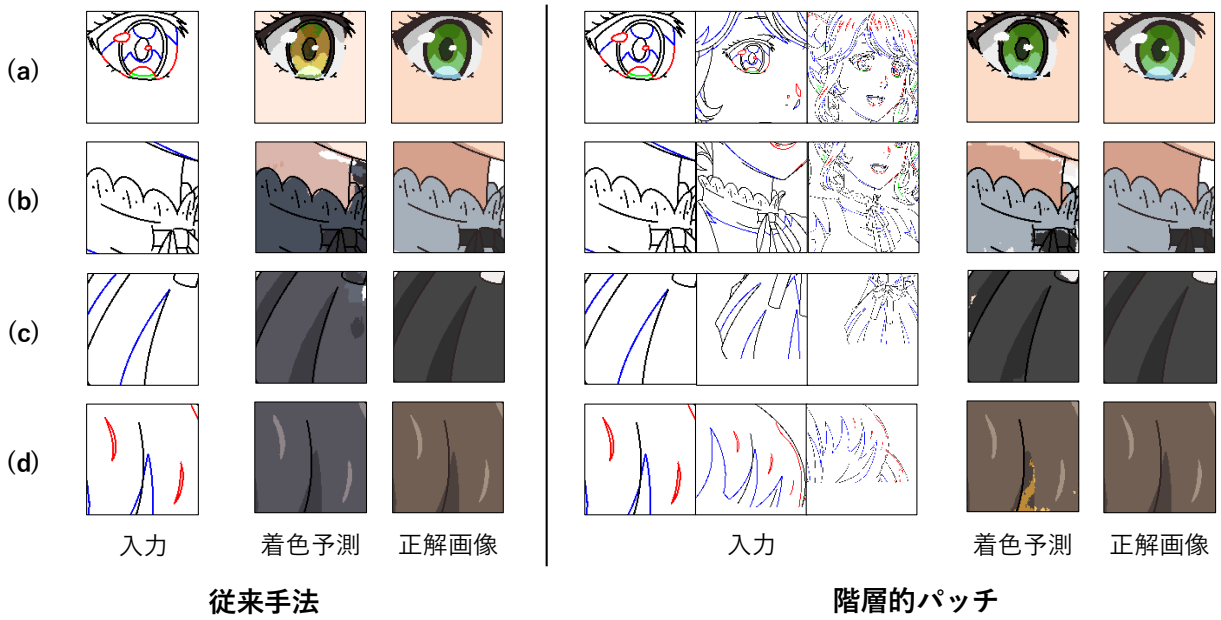

図 3 従来手法と提案手法の着色結果比較

平均した mIoU を用いる．IoU は，物体検出の分野において一般的に用いられる評価指標であり，着色予測画像と正解画像の色ラベルにおけるそれぞれの領域の，和集合と積集合を用いて計算される指標である．mIoU は 0 から 1 の範囲をとり，1 に近いほど精度良く予測されていると言える．正解としてクラス i がラベルづけされた画素を，クラス j として予測した数を N_{ij}，クラス数を K とすると，mIoU は以下の式で表される．

$$\mathrm{mIoU} = \frac{1}{K}\sum_{i=1}^{K}\frac{N_{ii}}{\sum_{j=1}^{K}(N_{ij}+N_{ji})-N_{ii}} \qquad (1)$$

さらに mIoU に加えて本研究では，CPA（Central Pixel Accuracy）を新たに導入した．CPA は，パッチの中心を基準とし，ユークリッド距離で半径 r 以内の領域からランダムに複数の画素をサンプリングし，その正解率を計測する指標である．この指標は，従来のパッチベース評価において，パッチの端部分における着色精度が大きく低下する傾向があることに着目し，パッチの中央部分におけるモデル本来の着色性能を純粋に評価することを目的として設計したものである．本論文では $r = 50$ と設定し，パッチの中央領域における精度傾向を分析した．

4. 実験

本研究では，階層的パッチとその正解となる着色済みパッチ画像のペアからなるデータセットを作成し，学習および評価を行った．なお，データの元となる素材は，黒髪かつ黄色い瞳を持つキャラクター 1 と，金髪かつ緑色の瞳を持つキャラクター 2 という 2 人の登場人物を含むテレビアニメシリーズの画像であり，許諾を得て使用している．

データセットの作成にあたっては，線画からランダムに 256×256 ピクセルのパッチを抽出し，各抽出位置に対して 3 節で定義した階層的パッチを生成した．本実験ではオリジナルの解像度に加え，0.5 倍および 0.25 倍の解像度の画像を用いて，計 3 枚のパッチ画像を使用し，これらを結合して 9 チャンネルの入力とした．最終的に，合計 25,116 組の画像ペアを作成し，それぞれ 18,943 組を学習用，3,560 組を検証用，2,613 組を評価用として分割した．

学習には，FCN8s [8] のネットワークを採用し，25 エポックの学習を行った．なお比較手法として，階層構造を持たない単一解像度のみのパッチ画像を入力とするベースラインモデルを同条件で学習させた．

5. 結果と考察

図 3 に，単一解像度のみのパッチ入力と階層的パッチ入力による着色結果の比較を示す．視覚的に，提案手法はキャラクター固有の色をより正確に保持しており，領域ごとの誤分類が抑えられていることが確認できる．

たとえば，図 3(a) に示す瞳の領域では，従来手法がキャラクター 1 の黄色い瞳をキャラクター 2 に誤って割り当てているが，階層的パッチ入力では本来の緑色を正しく再現している．このような誤りは，目のパッチ単体では十分なコンテキスト情報が得られず，キャラクター間で目の局

	従来手法	階層的パッチ (提案手法)
mIoU	43.67%	50.77%
CPA(r=50)	75.35%	80.34%

表 1 従来手法と提案手法における着色精度の定量評価

所的特徴が視覚的に類似しているために生じたと考えられる．一方，階層的パッチでは，ダウンサンプリングした複数の解像度の線画から抽出された周辺情報を参照できたため，キャラクター 2 の瞳であることを正しく識別し，適切な緑色で着色することができた．この結果は，階層的パッチによって，異なるキャラクター間で視覚的に類似した領域を識別する能力が向上することを示している．

図 3(c) においては，パッチ内に線画がほとんど含まれておらず，局所情報だけでは身体のどの部位かを識別するのが困難であったために，従来手法ではキャラクター 1 の髪の領域と誤認した着色をしてしまっている．一方，階層的パッチを用いた場合は，周辺情報を活用して該当部位がキャラクター 2 のエプロンであることを正しく推定し，本来の色を適用している．このように，パッチ内部の情報量が少ない場合でも，階層的パッチを用いることで周辺のコンテキスト情報を手掛かりとして補い，より正確な着色が可能となる．図 3(d) においても，類似した部位の誤着色が階層的パッチにより改善されることが確認された．

以上の定性的結果から，階層的パッチの導入は，特に局所情報が不十分な領域において，モデルによる空間的文脈の把握と部位の識別の精度を向上させる効果があると考えられる．

推論データ全体での mIoU，Pixel Accuracy による定量評価の結果を表 1 に示す．両評価指標において提案手法が従来手法を上回る性能を示した．mIoU においては，従来手法が平均約 44%であったのに対し，提案手法は約 51%を記録し，着色性能の向上が確認された．また CPA では，従来手法が平均約 75%であったのに対し，提案手法は約 80%を達成した．

これらの結果は，階層的パッチが，局所的な詳細を保持しつつ，キャラクター固有の色情報の整合性を保つために不可欠である大域的なコンテキスト情報を有効に提供していることを示している．

6. おわりに

本論文では，アニメーション制作工程における線画の着色作業の効率化を目的として，新たに階層的パッチを導入し，パッチの周辺情報を活用した新たなパッチベース学習を提案した．本手法は，mIoU および CPA といった定量指標において安定した精度向上を示した．また，パッチの周辺情報をネットワークに提供することで，従来手法では困難であった色の整合性の維持やキャラクター間の誤認識の抑制に寄与し，特に線画情報が乏しい領域において顕著な改善を示した．

今後の課題としては，各解像度の特徴量を効率的に統合するためのネットワーク構造の改良や，階層の深さおよびパッチサイズの最適化により，精度と計算コストのバランスをさらに追求する予定である．

謝辞

本論文は，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．

参考文献

[1] 品川政太朗, 久保尋之, 石井大地, 前島謙宣, 舩冨卓哉, 中村哲, 向川康博, “セマンティックセグメンテーションに基づくアニメ作品の自動彩色,” *Visual Computing*, 2020.

[2] A. Maejima, S. Shinagawa, H. Kubo, T. Funatomi, T. Yotsukura, S. Nakamura, and Y. Mukaigawa, “Continual few-shot patch-based learning for anime-style colorization,” *Computational Visual Media*, vol. 10, no. 4, pp. 705–723, 2024.

[3] M. Xie, C. Li, X. Liu, and T.-T. Wong, “Manga Filling Style Conversion with Screentone Variational Autoencoder,” *ACM Transactions on Graphics*, vol. 39, no. 6, pp. 1–15, 2020.

[4] H. Kim, H. Y. Jhoo, E. Park, and S. Yoo, “Tag2Pix: Line Art Colorization Using Text Tag with SECat and Changing Loss,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9056–9065, 2019.

[5] S. Iizuka, E. Simo-Serra, and H. Ishikawa, “Let There Be Color! Joint End-to-End Learning of Global and Local Image Priors for Automatic Image Colorization with Simultaneous Classification,” *ACM Transactions on Graphics*, vol. 35, no. 4, pp. 1–11, 2016.

[6] J.-W. Su, H.-K. Chu, and J.-B. Huang, “Instance-aware Image Colorization,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7968–7977, 2020.

[7] L. Zhang, C. Li, T.-T. Wong, Y. Ji, and C. Liu, “Two-Stage Sketch Colorization,” *ACM Transactions on Graphics*, vol. 37, no. 6, pp. 1–14, 2018.

[8] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3431–3440, 2015.