

# Iterative Inference Technique for Image-to-image Translation of Tile Art Image Generation

Naoki Matsumura, Yasuaki Ito and Koji Nakano  
Department of Information Engineering, Hiroshima University  
Kagamiyama 1-4-1, Higashi-Hiroshima, 739-8527 JAPAN

**Abstract**—Tile art made by assembling small pieces of tiles is one of the artistic techniques. Tile art image generation is generating a tile art image that resembles a given digital photos and illustrations on the computer. In our previous work, we proposed a tile art image generation method using conditional generative adversarial networks. This method can generate a tile art image only by inference computation of the trained network. However, in generated tile art images, some tiles have noises and lack of edges. The main contribution of this work is to show a quality improvement technique of the generated tile art image using iterative inference. In this technique, the generated tile art image is given to the trained generator again as an input image, and this iteration of inference is repeated several times. As a result, the generated tile art images obtained by iterative inference technique have less noise and clearer edges of tiles.

**Index Terms**—tile art, machine learning, conditional GAN

## I. はじめに

小さなタイルを組み合わせて作られるタイルアートは、芸術的な技法の一つであり、長い歴史がある [1]。一方、コンピュータグラフィックスの分野では、タイルアート画像はコンピュータ上で生成される。コンピュータによるタイルアート画像生成技術は、非写実的画像生成技術の一つとして知られている。現在までに、いくつかの非写実的画像生成技術に関する研究が行われている [2]。

近年、超解像 [3], [4], 画像の色付け [5]–[7], 画像補完 [8]–[10] など画像変換問題は、機械学習の技術を用いることで急速に改善されてきた。急速に改善された要因として、畳み込みニューラルネットワーク (CNN) [11] や敵対的生成ネットワーク (GAN) [12] の登場が挙げられる。畳み込みニューラルネットワークは、従来の全結合ネットワークでは難しかった画像の特徴抽出を可能にした。一方で、GAN はデータセットの拡張など、画像生成の分野で広く使用されている。GAN は生成ネットワークと識別ネットワークの二つのネットワークから構成されている。生成ネットワークは訓練データに似たデータを生成しようとするネットワークで、識別ネットワークは訓練データと生成ネットワークの出力データを正しく識別しようとするネットワークである。これら二つのネットワークを敵対的に訓練することにより、互いのネットワークの能力を向上させていく。また、画像変換問題には、畳み込みニューラルネットワークに基づいた GAN である DCGAN [13] がよく使用される。しかし、DCGAN は抽出した特徴のみから画像を変換するため、詳細を欠いた画像を出力しやすい。そのような状況下で、Isola らは *pix2pix* [14] と呼ばれる、様々な画像変換問題に対応できる汎用的なネットワークを提案した。彼らのアプローチは、GAN を拡張した、条件付き敵対的生成ネットワーク

(cGAN) [15] に基づいており、他にも様々な技術が使用されている。彼らは *pix2pix* を用いて、スケッチ画像に基づいた写真生成、画像の色付け、昼の風景から夜の風景への変換など、様々な画像変換問題に対する有用性を示した。

Tokura らは貪欲法を用いた点描画生成手法を提案した [16]。この手法は、与えられた入力画像を近似するよう、一枚ずつ正方形パターンをキャンバスに貼り付けることで、点描画を生成する。より高品質な点描画を生成するために、彼らは人間の視覚特性に基づいた考え方を導入した。この考え方では、出力画像の質の評価は、人間の目に投影された出力画像と与えられた入力画像の差を元に計算されている。この評価指針を用いて、出力画像の質が改善できなくなるまで、正方形パターンを一枚ずつ貼り付けていく。図 1(b) は画像サイズ 512 × 512 の Lena(図 1(a)) を入力画像として生成した正方形点描画である。生成された点描画は、入力画像をよく近似できていることがわかる。

本研究の主な貢献は、機械学習を用いたタイルアート画像生成手法を提案することである。まず、正方形点描画生成手法 [16] についてレビューを行い、タイルアート画像生成手法として拡張する。点描画生成では、回転角度を考慮した、均一な色をした正方形パターンを貼り付けていた。本研究では、それらのパターンの縁に幅 1 ピクセルの黒と白の枠を付加し、それをタイルとしてタイルアート画像を生成する。入力画像が与えられたとき、入力画像を再現するよう多量の枠付きタイルが貼り付けられる (図 2(a))。以降では、この点描画生成手法に基づいたタイルアート画像生成手法を局所探索法と呼ぶことにする。図 1(b) は局所探索法を用いて生成したタイルアート画像を示している。正方形パターンははっきりとした枠を持っているが、出力画像は入力画像をよく近似できていることが分かる。この局所探索法を元に、本研究では、機械学習を用いたタイルアート画像生成手法である機械学習法を提案する。より具体的に言い換えると、タイルアート画像を生成するディープニューラルネットワークを提案する。入力画像が与えられたとき、提案ネットワークの推論を行うことによってタイルアート画像を生成する (図 2(b))。このネットワークを構築するために、*pix2pix* [14] の考え方を採用する。また、訓練データセットには、画像データベースからランダムに選ばれた画像に局所探索法を適用し生成したタイルアート画像を使用する。提案ネットワークの目標は、可能な限り局所探索法によって生成されたタイルアート画像に近い画像を生成することである。訓練後の提案ネットワークは、入力画像の特徴を残した、また、タイルの構造を持ったタイルアート画像を生成できる。また、本研究では、機械学習法を用いて生成したタイルアート画像の高品質化手法についても提



Fig. 1. 局所探索法 [16] を用いた点描画と、 枠付きタイルを用いたタイルアートへの応用

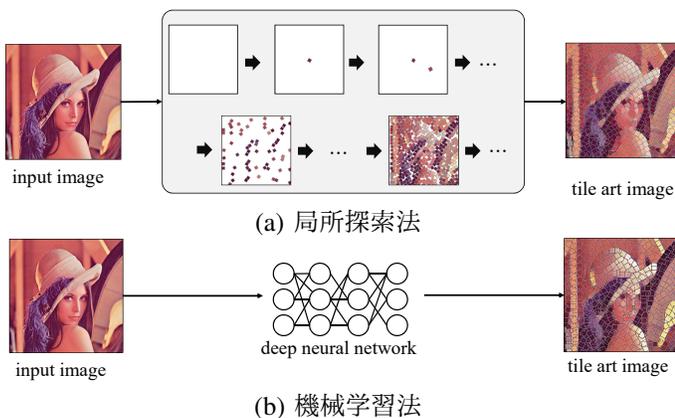


Fig. 2. 二つのタイルアート画像生成手法の概要

案する。この手法は、生成したタイルアート画像に反復推論テクニックを適用することで、より高品質なタイルアート画像を生成する。結果として、機械学習法のみで生成したタイルアート画像と比較して、よりノイズの少ない、タイルが際立った高品質なタイルアート画像を生成することができる。

本論文は以下のように構成されている。II 節では、貪欲法を用いた正方形点描画生成手法のレビューを行う。III 節では、機械学習を用いたタイルアート画像生成手法を提案する。IV 節では、生成されたタイルアート画像について評価する。V 節では、III 節で生成したタイルアート画像の高品質化手法を提案する。VI 節では、本研究を締めくくる。

## II. 貪欲法を用いた点描画生成

本節では、人間の視覚特性に基づく正方形点描画生成手法 [16] についてレビューする。この手法は、多量の同じ大きさである正方形パターンをキャンバス画像  $B$  に貼り付けることにより、与えられた入力画像  $A$  を再現する点描画を出力する。出力画像  $B$  の質は、人間の視覚特性に特性が類似しているガウシアンフィルタを用いて評価される。出力画像  $B$  の投影画像  $R$  は、出力画像  $B$  にガウシアンフィルタを適用することにより得られる。すなわち、二次元ガウ

シアン分布を持つ、要素の総和が 1 となるフィルタを  $G$  としたとき、投影画像  $R$  は出力画像  $B$  とガウシアンフィルタ  $R$  の畳み込みによって得られる。ガウシアンフィルタは人間の視覚特性に類似しているため、入力画像  $A$  と投影画像  $R$  の差が十分に小さい場合、出力画像  $B$  は入力画像  $A$  を再現していることになる。ここで、入力画像  $A$  と投影画像  $R$  の差には、各ピクセルの絶対値の差の平均値が用いられる。この評価指針と以下の貪欲法を用いて、正方形点描画を生成する。正方形パターンを貼り付けることで、質が最も改善される座標に正方形パターンを 1 枚貼り付ける。この操作は、出力画像  $B$  の質が改善されなくなるまで繰り返される。

本研究で使用するタイルは、[16] のような均一な色をした正方形パターンではなく、1 ピクセルの黒枠と白枠を持つ正方形パターンと丸形パターンである。これらのタイルパターンを用いて、上記手法からタイルアート画像を生成する。この枠付きタイルを使用した点描画生成手法を局所探索法と呼ぶことにする。次の節で使用する訓練データセットには、局所探索法を用いて生成したタイルアート画像を用いる。

## III. 機械学習を用いたタイルアート画像生成

本節では、機械学習を用いてタイルアート画像を生成する機械学習法について説明する。局所探索法では、多量のタイルパターンを繰り返しキャンバスに貼り付け、タイルアート画像を生成していた。一方で、機械学習法では、与えられた入力画像に対して、ディープニューラルネットワークを用いて直接タイルアート画像を生成する。つまり、訓練後のネットワークを用いて一回推論するだけで、タイルアート画像を生成する。提案するネットワークの構造は、pix2pix [14] を参考にしている。以下では、提案するネットワークのネットワークモデル、ネットワークアーキテクチャ、ネットワークの最適化手法について説明する。

### A. ネットワークモデル

機械学習法では、ネットワークモデルには敵対的生成ネットワーク (GAN) [12] を使用する。GAN はゲーム理論に基づいたディープニューラルネットワークモデルであり、生成ネットワーク ( $G$ ) と識別ネットワーク ( $D$ ) の二つのネッ

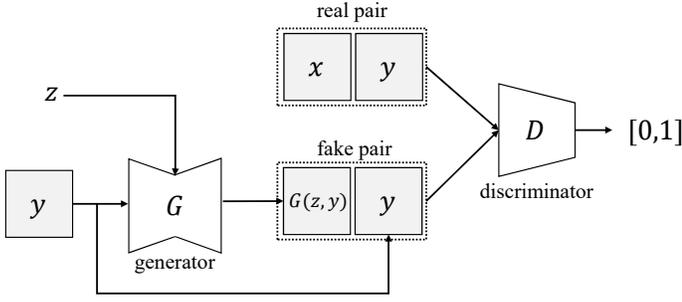


Fig. 3. 条件付き敵対的生成ネットワーク (cGAN)

トワークから構成されている。生成ネットワークは、ある対象領域の画像  $G(z)$  を出力するよう訓練する。  $z$  は潜在変数であり、通常ランダムノイズが使用される。ここで、訓練データセット中の画像を本物の画像、生成ネットワークが生成した画像を偽物の画像と呼ぶことにする。識別ネットワークは、本物の画像が偽物の画像かを正しく識別するよう訓練する。具体的に、識別ネットワークは 1 に近いほど本物の画像、0 に近いほど偽物の画像となる確率を出力しようとする。生成ネットワークと識別ネットワークは、互いに敵対的に、かつ、同時に訓練する。つまり、生成ネットワークは自身が生成した画像を偽物と特定されないよう、識別ネットワークは生成された画像を正しく偽物と特定できるよう訓練を行う。このように訓練を行うことで、生成ネットワークと識別ネットワークは互いに能力を向上させる。訓練後には、生成ネットワークは識別ネットワークを騙すことができる画像を生成できるようになる。この敵対的訓練を達成するため、以下の誤差関数  $\mathcal{L}_{GAN}(G, D)$  を定義する。

$$\mathcal{L}_{GAN}(G, D) = E_x[\log D(x)] + E_z[\log(1 - D(G(z)))] \quad (1)$$

ここで、  $E_x[\cdot]$  と  $E_z[\cdot]$  はそれぞれ訓練画像  $x$  と潜在変数  $z$  の確率分布の期待値である。

上記では、生成ネットワークは潜在変数  $z$  のみから画像を生成する。しかし、画像変換問題の領域では、生成された画像は変換前の画像に関連していなければならない。したがって、それらを結びつけるため、GAN を拡張した条件付き敵対的生成ネットワーク (cGAN) [15] を採用する。cGAN では、変換前の画像  $y$  は生成ネットワーク  $G$  と識別ネットワーク  $D$  の追加の入力として与えられる。図 3 は cGAN の概要を示している。生成ネットワーク  $G$  は二つの入力 (潜在変数  $z$  と変換前の画像  $y$ ) に対して、偽物の画像  $G(z, y)$  を出力する。一方で、識別ネットワークには real pair と fake pair のどちらかを入力する。ここで、real pair は訓練画像  $x$  と変換前の画像  $y$ 、fake pair は生成ネットワークが生成した画像  $G(z, y)$  と変換前の画像  $y$  から構成される。これら二つのネットワークを訓練するために、以下の式で表される cGAN の誤差関数  $\mathcal{L}_{cGAN}(G, D)$  を定義する。

$$\mathcal{L}_{cGAN}(G, D) = E_{x,y}[\log D(x, y)] + E_{z,y}[\log(1 - D(G(z, y), y))] \quad (2)$$

また、pix2pix では、生成された画像をより訓練画像に近づけるため、以下の誤差関数を式 2 に追加する。

$$\mathcal{L}_{L_1}(G) = E_{x,y,z}[\|x - G(z, y)\|_1] \quad (3)$$

ここで、  $\mathcal{L}_{L_1}(G)$  は訓練画像  $x$  と生成された画像  $G(z, y)$  の各ピクセルにおける差の絶対値の平均値である。

さらに、生成ネットワークの学習を安定させるため、以下の誤差関数で定義される *feature matching* [17] を導入する。

$$\mathcal{L}_{FM}(G, D) = E_{x,y,z} \sum_{t=1}^T \frac{1}{N_t} [\|D^{(t)}(x, y) - D^{(t)}(G(z, y), y)\|_1], \quad (4)$$

ここで、  $T$  は識別ネットワークの中間層の総数、  $D^{(t)}$  は識別ネットワークの  $t$  番目の中間層、  $N_t$  は  $D^{(t)}$  の要素数を示している。この誤差関数により、訓練画像と生成ネットワークが生成した画像を識別ネットワークに入力したときの、識別ネットワークの中間層の特徴量の期待値を一致させるよう、生成ネットワークは訓練される。結果として、生成ネットワークはより訓練画像の特徴量を再現するような、偽物の画像を生成することが出来るようになる。識別ネットワークは訓練により、訓練画像か生成された画像かを判別することのできる特徴量を見つけようとするため、この手法は GAN アルゴリズムにおいて本質的であるといえる。

最終的に、本研究では以下の目的関数  $\mathcal{L}(G, D)$  を使用する。

$$\mathcal{L}(G, D) = \mathcal{L}_{cGAN}(G, D) + \lambda_1 \mathcal{L}_{L_1}(G) + \lambda_2 \mathcal{L}_{FM}(G, D) \quad (5)$$

ここで、  $\lambda_1$  と  $\lambda_2$  はそれぞれ  $\mathcal{L}_{L_1}(G)$  と  $\mathcal{L}_{FM}(G, D)$  の目的関数に対する比重を制御するハイパーパラメータである。訓練中、生成ネットワークと識別ネットワークはそれぞれ  $\mathcal{L}(G, D)$  を最小化しようとする。つまり、本研究における訓練の目的は、以下の最適化問題から得られる生成ネットワーク  $G^*$  を見つけることである。

$$G^* = \arg \min_G \max_D \mathcal{L}(G, D)$$

## B. ネットワークアーキテクチャ

III-A 小節で説明したネットワークモデルで使用する、生成ネットワーク  $G$  と識別ネットワーク  $D$  の詳細について説明する。この詳細は主に pix2pix [14] に基づいている。

まず、生成ネットワークの詳細について説明する。生成ネットワークには、U-Net [18] を用いた。図 4 は提案する生成ネットワークの構造を示している。提案する生成ネットワークは、エンコーダネットワークとデコーダネットワークの二つのパートから構成されている。pix2pix [14] では、エンコーダネットワークとデコーダネットワークはそれぞれ 8 層であった。この層の数は、様々な画像変換問題に対応するために調整された層の数である。しかし、本研究では、タイルアート画像を生成することのできる構造であれば十分であるため、計 10 層の構造を採用した。層の数を計 16 層から計 10 層に減らすことで、訓練の簡易化や高速化が期待できる。ここから、エンコーダネットワークとデコーダネットワークの詳細について説明する。エンコーダネットワークは畳み込みニューラルネットワークから構成されている。エンコーダネットワークの各層は、ストライド 2 の  $4 \times 4$  畳み込みを実行しており、特徴マップをダウンサンプリングしている。各畳み込み層の後の活性化関数として、以下の式で表される leaky rectified linear unit (LeakyReLU) 関数 [19] を使用している。

$$f(x) = \begin{cases} x & (x \geq 0) \\ \gamma x & (x < 0) \end{cases} \quad (6)$$

ここで、  $\gamma$  は入力が負のときにどの程度値を小さくするか制御するハイパーパラメータ (値域 (0, 1)) である。一方で、デ

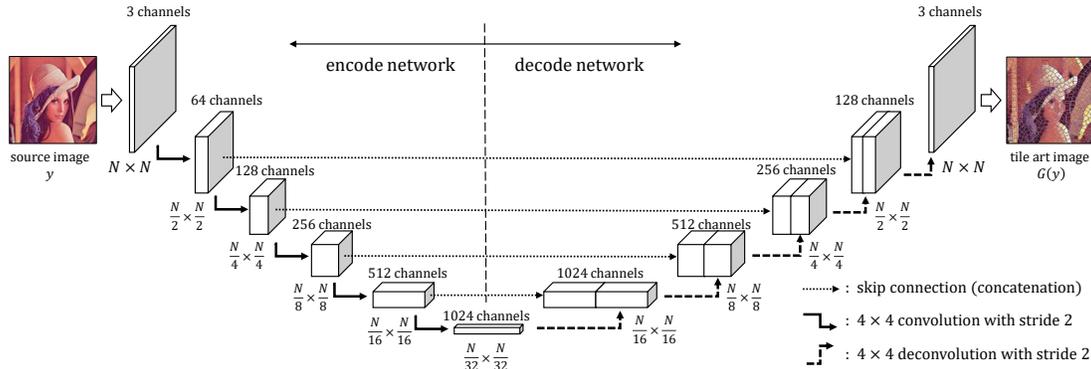


Fig. 4. 生成ネットワークの構造

コーダネットワークは逆畳み込みニューラルネットワークから構成されている。デコーダネットワークの各層は、ストライド2の $4 \times 4$ の逆畳み込みを実行しており、特徴マップをアップサンプリングしている。各逆畳み込み層の後には、以下の式で表される rectified linear unit(ReLU) 関数 [19] を使用している。

$$f(x) = \begin{cases} x & (x \geq 0) \\ 0 & (x < 0) \end{cases}$$

また、デコーダネットワークの各層には、skip connection [18] と呼ばれるエンコーダネットワークの特徴マップと入力とを結合する技術が使用されている。skip connection は、エンコーダネットワークでのダウンサンプリングによるピクセルレベルの情報の損失を防止し、画像変換問題の訓練を促進する技術である。生成ネットワークの出力の3チャンネルは、出力画像のRGBカラーモデルと一致している。U-Netでは、生成ネットワークの入力である潜在変数 $z$ の代わりにドロップアウトが適用されている。ドロップアウトは、訓練段階においてランダムに選ばれたノード(重み)をないものとして訓練を行い、ネットワークの過学習を抑制する訓練テクニックである。ランダムに選ばれるノードは、訓練ステップ毎に異なっている。ドロップアウトは入力である潜在変数 $z$ と同等の働きをすることを考えることが出来るため、本研究では潜在変数 $z$ を省略する。つまり、生成ネットワークの入力は変換前の画像 $y$ のみになる。

次に、識別ネットワークの詳細について説明する。図5は識別ネットワークの構造を示している。識別ネットワークの構造として、patchGAN [20] を使用する。上記で説明していた識別ネットワークの出力は一つの確率であった。一方で、patchGANでは、入力画像が複数の部分画像に分割され、各部分画像に対しての確率を出力する。このように複数の部分画像に分割することで、識別ネットワークのパラメータ数を削減でき、より効率的に訓練することが可能になる。識別ネットワークの入力には、二つの画像を結合した real pair または fake pair のいずれかが与えられる。real pair は変換前の画像 $y$ とその画像に対して局所探索法で生成した訓練用タイルアート画像 $x$ から構成される。一方で、fake pair は変換前の画像 $y$ と生成ネットワークがその画像に対して生成したタイルアート画像 $G(y)$ から構成される。識別ネットワークは、ストライド2の $4 \times 4$ の畳み込み層

3層とストライド1の $4 \times 4$ の畳み込み層2層であり、real pair または fake pair のダウンサンプリングを行う。各畳み込み層の後には、LeakyReLU 関数 [19] を適用している。最後に、 $N \times N$ の画像の pair に対して、識別ネットワークは $\frac{N}{8} \times \frac{N}{8}$ の識別結果を出力する。この出力の各要素は、分割した部分画像に対する識別結果の確率を表している。

### C. ネットワークの最適化

本小節では、提案ネットワークの訓練方法とその詳細について説明する。本研究では、ミニバッチ確率的勾配降下法 [11] を採用する。生成ネットワークと識別ネットワークの入力画像サイズは $256 \times 256$ 、つまり、 $N = 256$ に設定する。もし、入力画像が $256 \times 256$ より大きい場合でも、畳み込み演算または逆畳み込み演算を水平または垂直に繰り返し適用することで、タイルアート画像を出力することができる。入力画像が $256 \times 256$ より小さい場合については、入力画像の境界を $256 \times 256$ になるまで拡張し、生成ネットワークの入力とする。結果として得られた $256 \times 256$ の出力画像を、拡張する前の画像サイズで切り取ることで小さな画像に対してもタイルアート画像を生成することができる。したがって、どの大きさの画像に対しても、タイルアート画像を得ることが可能である。活性化関数について、式6の $\gamma$ は生成ネットワークと識別ネットワークの両方とも0.2に設定する。ドロップアウトは、50%の割合で生成ネットワークのデコーダネットワークの最初と2番目の層に適用する。また、識別ネットワークにおいても、同じ割合で最初の層と2番目の層に適用する。加えて、訓練を高速に行うために、Adam 最適化手法 [21] を使用する。Adam 最適化手法のハイパーパラメータとして、 $\alpha = 0.0001, \beta_1 = 0.5, \beta_2 = 0.99, \epsilon = 10^{-12}$ を使用する。これらのハイパーパラメータの詳細について興味がある場合は、参考文献 [21] を参照することを推奨する。生成ネットワークと識別ネットワークの全ての重みは、平均0かつ標準偏差0.02の正規分布を用いて初期化されている。生成ネットワークと識別ネットワークの学習率は、それぞれ0.0002と0.00001に設定した。式5のハイパーパラメータである $\lambda_1$ と $\lambda_2$ は1に設定した。

Algorithm 1は、提案ネットワークの訓練アルゴリズムを示している。ここで、ミニバッチサイズ $m$ は4に設定した。訓練アルゴリズムのループは、生成したタイルアート画像の質が変わらなくなるまで繰り返される。各ループは、1回のエポックに対応している。ここで、エポックとは、全体

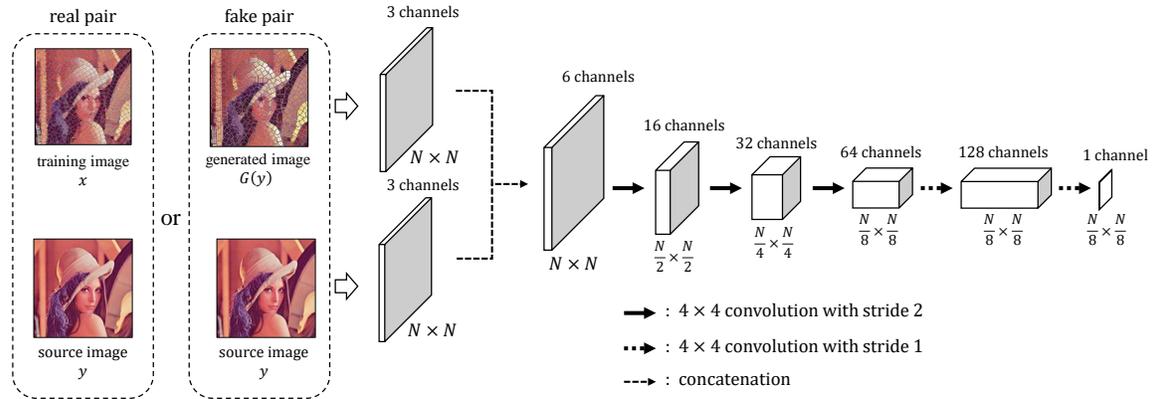


Fig. 5. 識別ネットワークの構造

の訓練データセットに対して順伝播と逆伝播を1回のみ実行することを指す。我々の経験では、通常200エポックだけループを繰り返すことで、生成されるタイルアート画像の質は変化しなくなる。このループは、生成ネットワークを訓練する段階、識別ネットワークを訓練する段階の2段階から構成されている。始めの段階で生成ネットワークを訓練し、その後、識別ネットワークを訓練する。識別ネットワークの訓練については、2回の訓練ステップにつき1回のみ訓練を実行することに注意する。この生成ネットワークと識別ネットワークの訓練回数についても、実験を繰り返して最適なものを選択した。

#### Algorithm 1 提案ネットワークの訓練アルゴリズム

- 1: 訓練データセット中の画像の全ピクセルを  $[-1, 1]$  に正規化
- 2: **loop**
- 3: ミニバッチ  $m$  分の変換前の画像  $y_1, \dots, y_m$  とそれに対応するタイルアート画像  $x_1, \dots, x_m$  を取得
- 4: 識別ネットワークのパラメータを固定し、以下の勾配から生成ネットワークのパラメータを更新

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^m (\log(1 - D(G(y_i), y_i)) + \lambda_1 (\|x_i - G(y_i)\|_1) + \lambda_2 (\sum_{t=1}^T \frac{1}{N_t} \|D^{(t)}(x_i, y_i) - D^{(t)}(G(y_i), y_i)\|_1)). \quad (7)$$

- 5: **if** 現在の訓練ステップが偶数 **then**
- 6: 生成ネットワークのパラメータを固定し、以下の勾配から識別ネットワークのパラメータを更新

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^m (\log D(x_i, y_i) + \log(1 - D(G(y_i), y_i))). \quad (8)$$

- 7: **end if**
- 8: **end loop**

#### IV. 機械学習法の評価実験

本節では、機械学習法により生成したタイルアート画像についての評価を行う。また、機械学習法の訓練過程についても説明する。評価用の画像として、大きさ  $512 \times 512$  の Lena 画像 [22] を使用する。

本研究では、正方形タイルアート画像生成ネットワークと丸形タイルアート画像生成ネットワークの二つのネットワークを訓練した。これらのネットワークを訓練するために、256種類のカテゴリに関する30607枚の画像を含む Caltech-256 データセット [23] を変換前の画像  $y$  として用いた。30607枚の内、ランダムに20000枚選択し、 $256 \times 256$  に拡大縮小する。これらの画像に対して、局所探索法を用いて大きさ  $256 \times 256$  の正方形タイルアート画像と丸形タイルアート画像をそれぞれ20000枚生成する。訓練データセット用のタイルアート画像生成には、タイルの大きさ  $23 \times 23$  を使用した。タイルを際立たせるため、タイルに黒と白の枠を追加した。また、正方形タイルアート画像について、正方形タイルの回転角は  $0, 30, 60$  度である。生成したタイルアート画像と変換前の画像をデータセットとして使用する。

III-C 小節で説明した最適化手法を用いて、提案ネットワークを訓練する。提案ネットワークは、生成したタイルアート画像の質が変化しなくなるまで訓練した。我々の経験から、出力画像の質は200エポック付近で変化しなくなるため、訓練回数として200エポックを採用する。図6は、訓練中のネットワークが Lena に対してタイルアート画像を生成した図である (Lena 画像は、訓練データセットには含まれていない)。訓練の初期段階では、タイルが構成されておらず、オリジナル画像が見えてしまっている。訓練が進むと、オリジナル画像の特徴を残した状態でタイルを構成している。150エポック後になると、タイルの特徴がより顕著になっている。

図7と図8は、大きさ  $512 \times 512$  の Lena 画像に対して、局所探索法と機械学習法を用いて生成した正方形タイルアート画像と丸形タイルアート画像を示している。機械学習法を用いて生成したタイルアート画像は、局所探索法のものと比較して、タイルを規則的に置く傾向がある。また、タイルの崩壊やタイル内の色が不均一な箇所がある。しかし、タイルを一枚ずつ置いていないにも関わらず、出力画像はタイルの構造をしており、タイルの重なりまでも表現している。正方形タイルアート画像と比較して、丸形タイルアート画像はより高品質であることがわかる。これは、丸形タイルが回転角を持たず、正方形タイルと比べてより簡単な形をしているためである。

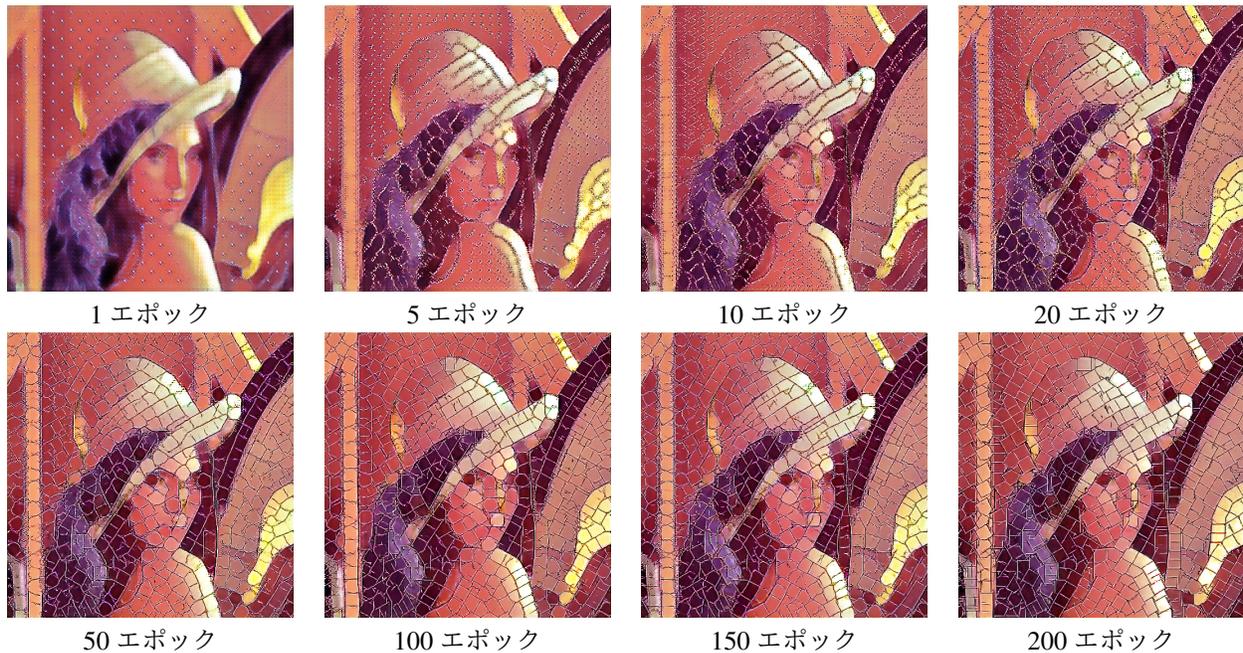


Fig. 6. 訓練中の生成ネットワークの出力画像

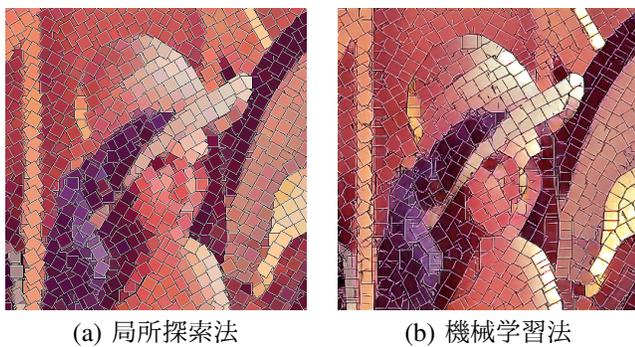


Fig. 7. 生成された正方形タイルアート画像

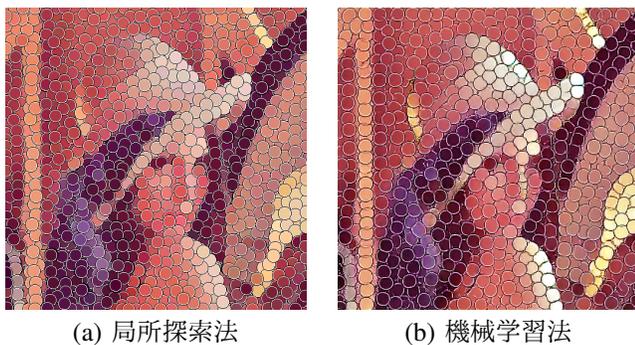


Fig. 8. 生成された丸形タイルアート画像

### V. 反復推論によるタイルアート画像の高品質化

III節では、cGANを用いたタイルアート画像生成手法である機械学習法を説明した。しかし、cGANの学習は困難であり、機械学習法で生成したタイルアート画像にはノイズやタイルが崩壊が見られた。そこで、本節では、機械学

習法で生成したタイルアート画像の高品質化テクニックについて説明する。このテクニックは非常にシンプルであり、訓練された生成ネットワークにより生成されたタイルアート画像を再び同じ生成ネットワークの入力として与え、反復して推論を行う(図11)。このテクニックの目的は、提案ネットワークが生成されたタイルアート画像のタイルの形をより際立たせることである。図12と図13は反復推論テクニックを正方形タイルアート画像と丸形タイルアート画像にそれぞれ適用したタイルアート画像を示している。これらから、図7(b)や図8(b)で示されている繰り返し推論を行わないタイルアート画像と比較して、タイルがより際立っていることがわかる。また、図7(b)や図8(b)にあったノイズも減少していることがわかる。しかし、推論を繰り返すと、入力画像になかったタイルの枠が影響しているため、画像全体が暗くなってしまっている。我々の経験上、反復推論を2回か3回行ったとき、タイルの特徴と全体の明るさのバランスが良いタイルアート画像が得られている。一方、タイルアート画像生成時間については、反復推論を行った回数だけ増加してしまう。しかし、機械学習法によるタイルアート画像生成時間はFullHD画像(1920×1080)に対して0.14秒と非常に短いため、実用的な場合に対しても有用性がある。

### VI. まとめ

本論文では、cGANを用いたタイルアート画像生成手法である機械学習法について提案した。提案ネットワークの訓練では、局所探索法で生成したタイルアート画像をデータセットとして用いた。訓練後の提案ネットワークは、1回の推論のみで元画像の特徴を捉えた、タイルの構造を持った画像を生成することができた。また、タイルアート画像の高品質化手法として、反復推論テクニックを提案した。このテクニックでは、タイルアート画像に対して推論計算を反

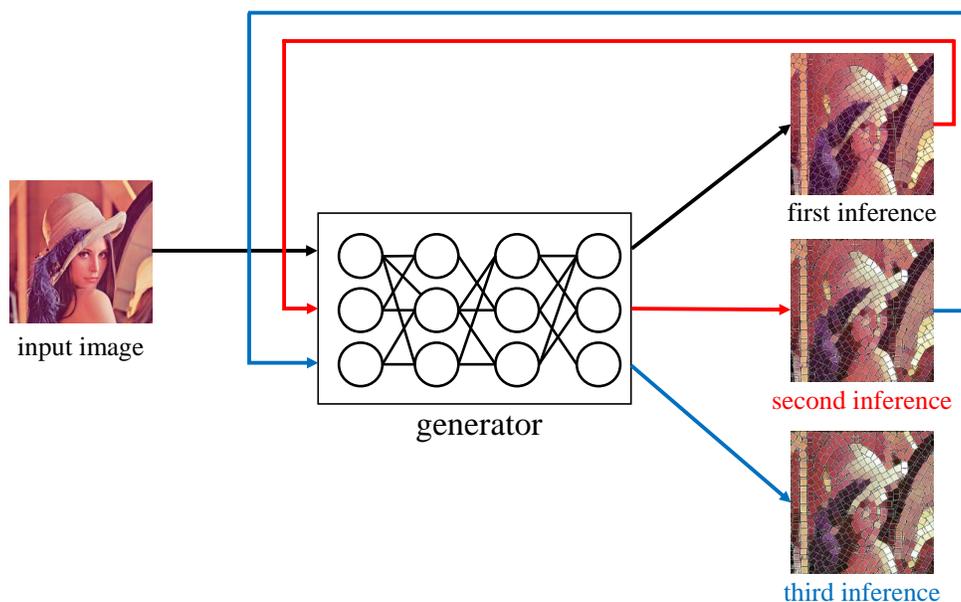


Fig. 9. 機械学習法に対する反復推論

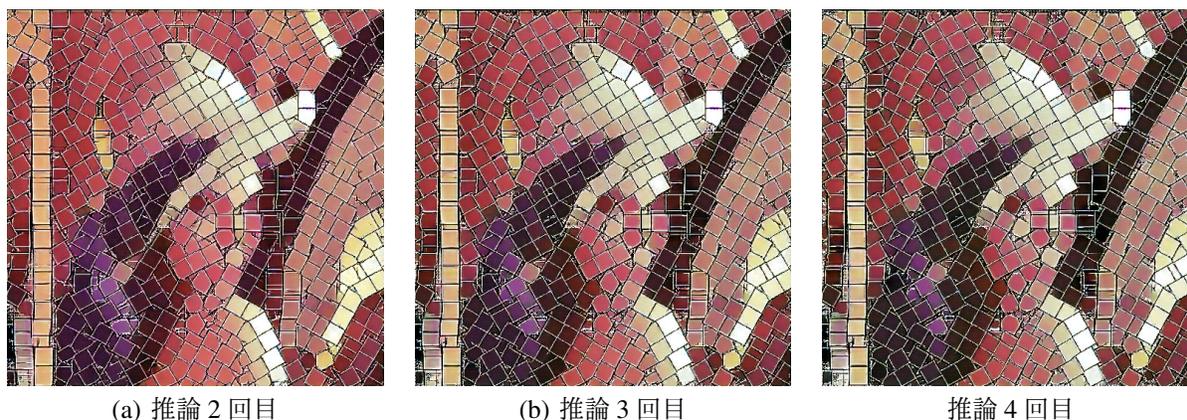


Fig. 10. 反復推論により生成された正方形タイルアート画像

復して行う。結果として、ノイズの少ない、タイルが際立った高品質なタイルアート画像を生成することに成功した。

## REFERENCES

- [1] Sarah Kelly. *The Complete Mosaic Handbook: Projects, Techniques, Designs*. Firefly Books, first edition, 2004.
- [2] S. Battiato, G. Di Blasi, G. M. Farinella, and G. Gallo. Digital mosaic frameworks - an overview. *Computer Graphics Forum*, 26(4):794–812, 2007.
- [3] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *Proc. of European Conference on Computer Vision*, pages 694–711, 2016.
- [4] Emily L Denton, Soumith Chintala, arthur szlam, and Rob Fergus. Deep generative image models using a Laplacian pyramid of adversarial networks. In *Advances in Neural Information Processing Systems 28*, pages 1486–1494. Curran Associates, Inc., 2015.
- [5] Richard Zhang, Phillip Isola, and Alexei A. Efros. Colorful image colorization. *CoRR*, abs/1603.08511, 2016.
- [6] Gustav Larsson, Michael Maire, and Gregory Shakhnarovich. Learning representations for automatic colorization. In *Proc. of European Conference on Computer Vision*, pages 577–593, 2016.
- [7] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi 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 (Proc. of SIGGRAPH 2016)*, 35(4):110, 2016.
- [8] Jia-Bin Huang, Sing Bing Kang, Narendra Ahuja, and Johannes Kopf. Image completion using planar structure guidance. *ACM Transactions on Graphics (Proc. of SIGGRAPH 2014)*, 33(4):129, 2014.
- [9] Deepak Pathak, Philipp Krähenbühl, Jeff Donahue, Trevor Darrell, and Alexei Efros. Context encoders: Feature learning by inpainting. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2536–2544, 2016.
- [10] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. Globally and locally consistent image completion. *ACM Transactions on Graphics (Proc. of SIGGRAPH 2017)*, 36(4):107, 2017.
- [11] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT press, 2016.
- [12] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems 27*, pages 2672–2680, 2014.
- [13] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *CoRR*, abs/1511.06434, 2016.

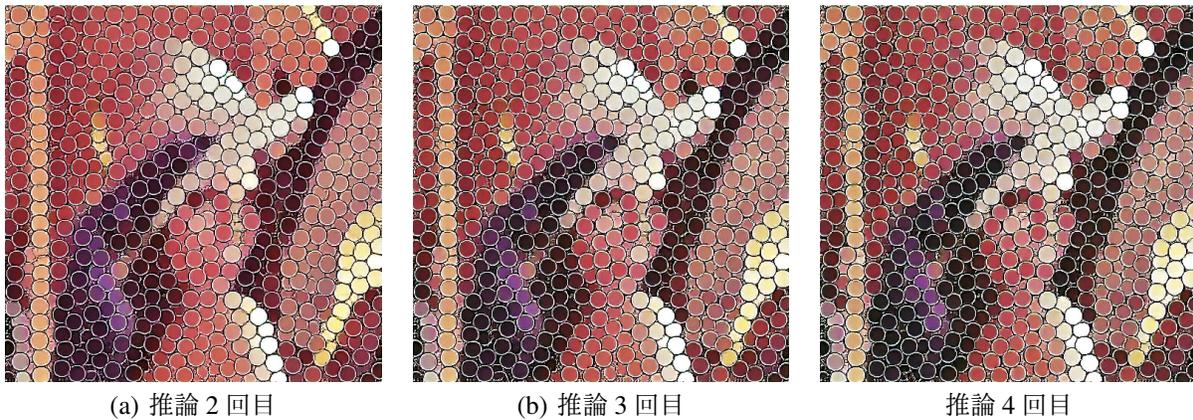


Fig. 11. 反復推論により生成された丸形タイルアート画像

- [14] Phillip Isola, Jun-Yan Zhu, and Tinghui Zhou. Image-to-image translation with conditional adversarial networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5967–5976, 2017.
- [15] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *CoRR*, abs/1411.1784, 2014.
- [16] Hiroki Tokura, Yuki Kuroda, Yasuaki Ito, and Koji Nakano. A square pointillism image generation, and its GPU acceleration. In *Proc. of International Symposium on Computing and Networking*, pages 38–47, 2017.
- [17] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training GANs. In *Advances in Neural Information Processing Systems*, pages 2234–2242, 2016.
- [18] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, 2015.
- [19] Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. Rectifier nonlinearities improve neural network acoustic models. In *Proc. of ICML Workshop on Deep Learning for Audio, Speech, and Language Processing (WDLASL 2013)*, 2013.
- [20] Chuan Li and Michael Wand. Precomputed real-time texture synthesis with Markovian generative adversarial networks. In *Computer Vision – ECCV 2016*, pages 702–716, 2016.
- [21] Diederik P. Kingma and Jimmy Lei Ba. ADAM: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2017.
- [22] The USC-SIPI image database. <http://sipi.usc.edu/database/>.
- [23] Gregory Griffin, Alex Holub, and Pietro Perona. Caltech-256 object category dataset. Technical report, California Institute of Technology, 2007.