

# DNN 中間層特徴マップを用いた合成テクスチャ画像の画質推定 Image Quality Estimation of Synthetic Texture using DNN Intermediate Layer Feature Maps

荒井 雅司<sup>†</sup>      八島 由幸<sup>†</sup>  
Masashi Arai    Yoshiyuki Yashima

## 1. はじめに

映像データの効率的な表現を行うため、従来の圧縮符号化技術とは異なるコンセプトでアプローチを行うテクスチャ合成符号化技術が研究されている[1]。従来のテクスチャ合成符号化は、パッチの抽出位置やパッチサイズといったパラメータによって違和感のある画像が生成されることがあり、これらのパラメータを効率的に決定するために、合成テクスチャ画像の画質を精度よく推定することが必要とされていた。本研究では、畳み込みニューラルネットワーク (CNN) の中間層出力である特徴マップを用いて、テクスチャ合成画像の画質を推定する手法を提案する。

## 2. テクスチャ合成画質推定

### 2.1 テクスチャ合成

テクスチャ合成の流れとしては、まず画像中の領域からその部分領域(パッチ)を切りし、次に原テクスチャ画像と同じサイズになるように切り出したパッチを貼り付ける処理を繰り返す[2]。テクスチャ合成画像はパッチの切り出し位置や切り出しサイズによって図 1 のように画質が大きく変わる。このように元の画像とは異なる質感の合成画像が生成されてしまうことは画像符号化の観点からは好ましくなく、違和感のない合成画像が得られるパッチの切り出し位置やパッチサイズを自動的に特定するためには、合成画質の定量的評価が必要であるが、PSNRやSSIM[3]といった旧来の手法は適切ではなく、その手法は確立されていない。

### 2.2 CNN 中間層を用いたテクスチャ類似度推定

本稿では、テクスチャ画像に対して主観画質推定を行うために、CNN の中間層を用いた類似度推定手法[4]を応用する。図 2 は CNN を利用した画質推定の流れ図である。まず、原画像と合成画像をそれぞれ CNN に入力し中間層の特徴マップを取得する。次に、取得した各特徴マップを、幅、又は高さが 7pixel 以下になるまで再帰的に 4 分割する。分割された特徴マップは  $F_{nkr}(i,j)$  と表せる。ここで、 $n$  は層番号、 $k$  は特徴マップ番号、 $k$  は分割領域番号となる。その後、 $n$  番目の層について、 $k$  番目の特徴マップの  $r$  番目の領域に対するヒストグラムが計算される。これらのヒストグラムを 1 つのヒストグラムに統合して、確率分布で表し、第  $n$  層の特徴ヒストグラム  $H_n$  を取得する。同様にして、テクスチャ合成画像の特徴ヒストグラム  $H_n'$  を取得する第  $n$  層の評価値  $Q(n)$  は、 $H_n \cap H_n'$  と定義される。ここで、 $0 \leq Q(n) \leq 1$  である。最後に、数式(1)に基づいて各層の評価値の幾何平均を計算することで、最終的な画質評価値を取得する。

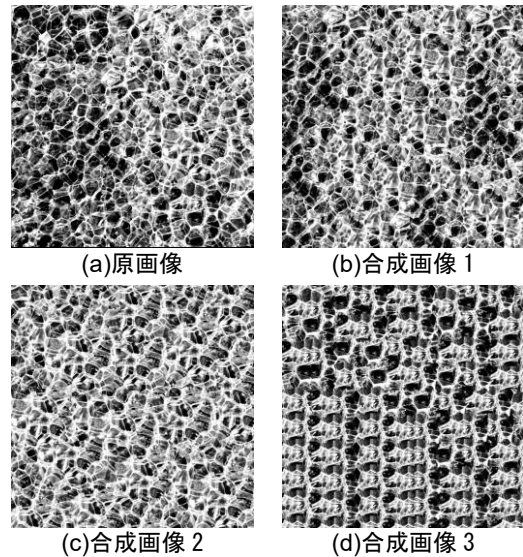


図 1 テクスチャ合成画像の例

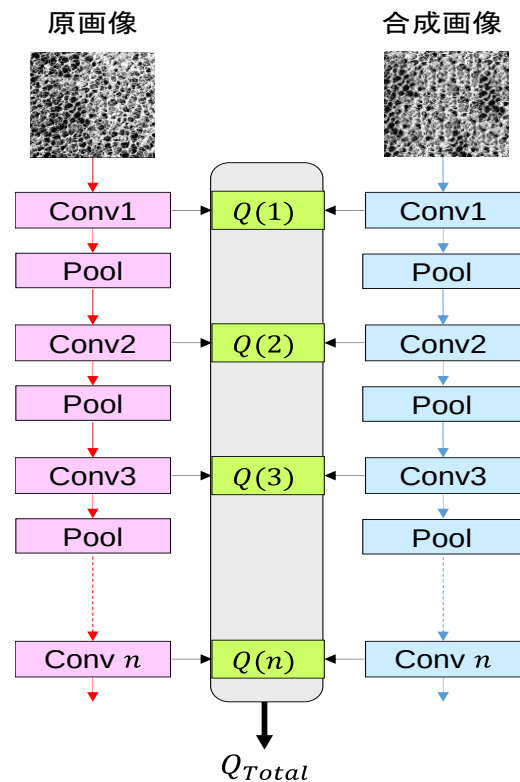


図 2 合成画像の画質推定の流れ

<sup>†</sup> 千葉工業大学大学院情報科学研究科, Graduate School of Information and Computer Science, Chiba Institute of Technology

$$Q_{Total} = (\prod_{n=1}^L Q(n))^{\frac{1}{L}} \quad (1)$$

ここで、 $L$ は使用する層の数である。

### 3. 実験結果

#### 3.1 実験条件

提案手法の有効性を示すために実験を行った。Brodatz データベースのテクスチャ画像 3 種の原画像(図 3)と、それをもとに様々な切り出し位置と切り出しサイズのパッチにより生成した合成画像を CNN への入力とし、出力値は(1)式によって求められる 2 つの画像の主観的テクスチャ類似度 (0~1) となる。実験で利用した CNN は 1000 クラス物体認識事前学習済みモデルで VGG-16[5]を利用した。また、15 名の評価者による 5 段階主観評価 (1~5) を実施し、推定評価値との相関を解析した。

#### 3.2 実験結果と考察

図 4 に画質の主観評価値と推定評価値の相関を示す。図中の  $\rho_p$  はピアソン係数、 $\rho_s$  はスピアマン順位相関係数である。また、ここで示している相関係数の値は、最も高い推定精度を示した中間層の組み合わせを使用した場合の結果を示している。VGG-16 は畳み込み層が 13 層あるため、層の組み合わせの総数は

$$\sum_{x=1}^{13} {}_{13}C_x = 8191 \quad (2)$$

となる。図 5 は、8191 パターンすべての推定評価値と主観評価値の相関係数を降順に示している。また、図 6 に、相関係数の上位 20 位と下位 20 位で使用された層の組み合わせパターンを示している。図 6 より、5~10 層目を使うと高い相関係数が得られることがわかる。一方、適切な層の組み合わせを選ばないと推定評価値と主観評価値の相関が低下し、特に最初の方の層を使用した場合、推定精度が大幅に低下することがわかる。

### 4. おわりに

CNN の中間層出力である特徴マップを用いてテクスチャ合成画像の主観画質を推定できる可能性を示した。今後は、データセットによる影響や、異なる CNN の利用を検証するとともに、実際の最適パッチ特定の検討を進める予定である。なお本研究は、JSPS 科学研究費 JP18K11360 の助成を受けて行った。

#### 参考文献

- [1] F. Racape, et al., "Adaptive pixel/patch-based synthesis for texture compression," IEEE International Conference on Image Processing, ICIP 2011.
- [2] Alexei A. Efros, William T. Freeman, "Image Quilting for Texture Synthesis and Transfer," Proceedings of SIGGRAPH 2001, pp.341-346, 2001
- [3] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, Eero P. Simoncelli, "The SSIM Index for Image Quality Assessment"
- [4] S. A. Amirshahi, M. Pedersen, and S. X. Yu, "Image quality assessment by comparing CNN features between images," Journal of Imaging Science and Technology, vol.60, no. 6, pp.060410-1-10, 2016.
- [5] Keras Applications <https://keras.io/ja/applications/>

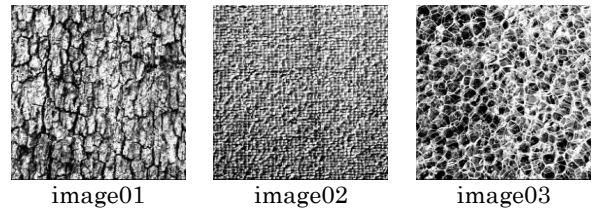


図 3 実験画像 (原画)

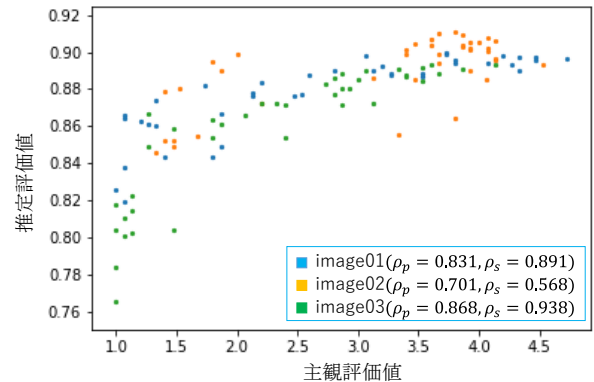


図 4 画質評価値の相関

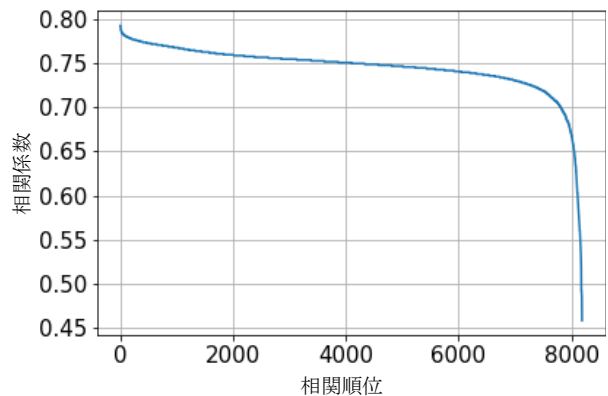


図 5 全組み合わせパターンの相関係数

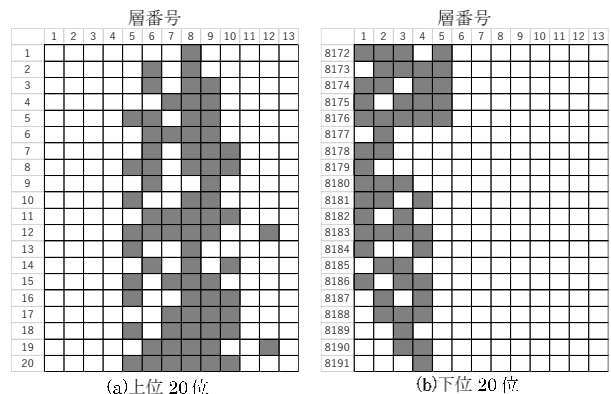


図 6 選択されたパターン