

CycleGAN を用いた宝飾品画像の補正における背景要素の影響

Effect of background on image enhancement of jewelry data using CycleGAN.

仲川 太陽[†] 鎮野 智宏[†] 石黒 大樹[†] 尾関 智子[†]
 Taiyo Nakagawa Tomohiro Shizuno Daiki Ishiguro Tomoko Ozeki

1. はじめに

近年ディープラーニングモデルを用いた画像補正についての研究が盛んである。先行研究[1]では、宝飾品画像の補正において CycleGAN[2]が特に有効であると示した。しかし、背景要素の有無による CycleGAN の学習及び補正能力の変化については検討されていなかった。本研究の目的は、影などの背景要素が宝飾品画像を補正する上で悪影響を及ぼすという仮定のもと、CycleGAN における宝飾品画像の補正能力の向上を実現することである。本研究では、Edge Flow[3]と呼ばれるアーキテクチャを導入したセグメンテーションツールである Eiseg を背景除去に用いる。背景を除去した宝飾品画像のデータセットを作成し、CycleGAN の学習に用いることで、背景除去の効果を検討する。

2. CycleGAN

CycleGAN[2]は、GAN (Generative Adversarial Network) の一種であり、ドメイン X からドメイン Y へのスタイル変換を実現する。GAN は、データの分布を捉える生成モデル G と、学習データと G から生成されたデータを判別する識別モデル D を Adversarial Loss (敵対的損失関数)

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (1)$$

により同時に学習させる。さらに CycleGAN では、Cycle Consistency Loss (サイクル一貫性損失関数)

$$L_{cyc}(G, F) = E_{x \sim p_{data}(x)} [\|F(G(x)) - x\|_1] + E_{y \sim p_{data}(y)} [\|G(F(y)) - y\|_1] \quad (2)$$

を導入することでドメイン X とドメイン Y の双方向の変換に矛盾が生じないような学習を実現する。本研究では、宝飾品の撮影画像をドメイン X 、熟練者が撮影画像を Photoshop を用いて加工した画像 (Photoshop 加工画像) をドメイン Y として CycleGAN を学習する。

3. Eiseg

Eiseg (Effective Interactive Segmentation) は、Edge Flow[3]を導入したクリックベースのインタラクティブなセグメンテーションツールである。Edge Flow は、edge-guided flow を用いることでユーザのクリック情報を活用するセグメンテーションアーキテクチャである (図 1)。

本研究では、生成された二値画像に 12×12 の平滑化フィルタをかける (図 2)。平滑化フィルタをかけた二値画像を用いて宝飾品画像の背景除去を行い、CycleGAN での学習に用いる。



図 1 Eiseg を用いた二値画像の生成

図 2 12×12 の平滑化フィルタをかけた二値画像

4. 評価指標

本研究では、CycleGAN によって生成された画像と Photoshop 加工画像との類似度を測る指標として以下の三つを用いる。第一に、画質の客観的な評価指標である PSNR (Peak Signal-to-Noise Ratio),

$$PSNR(x, y) = 10 \log_{10} \frac{MAX^2}{\frac{1}{N} \sum_i (x[i] - y[i])^2} \quad (3)$$

を用いる。PSNR は値が大きくなるほど評価が高い。PSNR は正の値を取り、上限値はない。

第二に、SSIM (Structure Similarity Index Measure),

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (4)$$

を用いる。SSIM は値が大きくなるほど評価が高い。SSIM は 0 から 1 の範囲の値を取る。

第三に、LAB 色空間におけるユークリッド距離 (LAB ユークリッド) を用いる。LAB ユークリッドは、2 枚の画像を RGB 色空間から LAB 色空間に変換し、ピクセルごとのユークリッド距離である色差を全ピクセルに対して平均することで算出できる。LAB ユークリッドは正の値を取り、値が小さいほど評価が高い。

5. 実験方法

本研究では、学習用データセットとして宝飾品会社より提供された画像を用いる。提供された画像は大きく二つに分けられる。第一は、宝飾品の撮影画像であり、サイズは 6000×4000 pixel である。第二は、熟練者による Photoshop 加工画像であり、サイズは 600×600 pixel である。撮影画像と Photoshop 加工画像はサイズや宝飾品のズーム度合いが異なることが問題となるが、先行研究[1]で用いられたイメージレジストレーションを行うことで解決できる。

[†] 東海大学 Tokai University

CycleGAN の学習用データセットを 2 種類用意する。第一に、撮影画像と Photoshop 加工画像共に背景除去を行った画像をデータセット A (背景なし) とする。第二に、撮影画像と Photoshop 加工画像共に背景除去を行わない画像をデータセット B (背景あり) とする。各データセットは撮影画像、Photoshop 加工画像ともに 2000 枚ずつ用意する。学習に用いる画像の枚数は 1000 枚, 1500 枚, 2000 枚の 3 パターン、画像サイズは 600pixel で用意する。最適化手法には Adam を用い、エポック数 200 で撮影画像を Photoshop 加工画像を模倣して変換するように CycleGAN[2]の学習を行う。

学習用データセットに含まれない 143 枚の画像をテスト用データセットとする。データセット A で学習を行ったモデルで補正する場合、事前にテスト用データセットの撮影画像に対しても背景除去を行う。各モデルで補正された画像に対して最後に背景除去を行い、評価指標を測ることで各モデルの性能評価を行う。

6. 実験結果

図 3 に撮影画像と各パターンで学習した CycleGAN による加工画像、Photoshop 加工画像を、表 1 に PSNR や SSIM, LAB ユークリッドを用いた定量的評価の平均値を示す。データセット B の 1500 枚を用いて学習を行ったモデルが指標において優れた結果を示し、主観的感覚とも一致する。データセット A で学習を行ったモデルは、図 4 に示すような背景要素が大きい画像に対しては、補正時に過剰露光やノイズが発生しないため、特に効果的である。また、データセット B を用いて学習を行ったモデルは、図 5 に示すように、それぞれ補正後に宝飾品が一部消失する場合がある。この現象は背景と宝飾品の境界が曖昧な画像群に対して発生する。

7. 考察と今後の展望

本研究では、CycleGAN で宝飾品画像を補正する際に背景要素が悪影響を及ぼすことを 2 種類の画像群において示した。第一に、背景要素が大きい画像群の補正に対しては、背景なしデータセットを用いることが有効である。要因として、背景要素の影響を受けずにモデルの学習が進んだことが考えられる。第二に、背景と宝飾品の境界が曖昧な画像群の補正に対しては、背景ありデータセットを用いることは有効でない。要因として、CycleGAN が背景と宝飾品の区別が曖昧なまま学習を行ったことが考えられる。背景ありデータセットを用いる場合、宝飾品を補正する方向への学習と同時に、背景を白くする方向への学習が進行する。背景と宝飾品の区別が曖昧であることで、宝飾品を白くする方向へ一部学習が進行した結果、図 5 のように宝飾品が一部消失すると考えられる。

一方、背景なしデータセットを用いた場合、背景ありデータセットを用いた場合に比べ性能が向上しなかった。背景なしデータセットでは、背景除去を行った撮影画像が Photoshop 加工画像にきわめて類似していた。その結果、識別モデル D が実画像と偽画像の違いを判別することが困難となり、生成モデル G の性能が十分に向上しなかったと考えられる。この問題の対策として、TTUR (Two Time-scale Update Rule) を導入することが挙げられる。TTUR を導入し、識別モデル D の学習率を大きく、生成モデル G の学習率を小さくすることで、この問題を解決する。



図 3 各パターンの CycleGAN による補正結果

表 1 600pixel の各評価指標の平均値

	画像数	PSNR	SSIM	LAB
データ セット A	1000	26.11	0.937	6.011
	1500	25.36	0.931	6.718
	2000	26.07	0.937	6.015
データ セット B	1000	25.98	0.936	5.983
	1500	26.27	0.938	5.825
	2000	26.25	0.939	5.852



図 4 背景要素が大きい画像に対する補正結果



図 5 補正後に宝飾品が一部消失する画像

謝辞

本研究は、株式会社セルビーの研究資金提供により研究が遂行されたものです。この場を借りて深く御礼申し上げます。同社の松谷稔哉氏、齊藤孝蔵氏、鈴木修康氏、山本真也氏には宝飾品画像と宝飾品に関する専門的知識、有益な助言をいただきました。心より感謝申し上げます。

参考文献

- [1] T. Shizuno, T. Nakagawa, D. Ishiguro, D. Hosokawa, Y. Oshigane, T. Ozeki, "CycleGAN を用いた宝飾品画像の補正", The 36th Annual Conference of the Japanese Society for Artificial Intelligence, (2022).
- [2] J-Y. Zhu, T. Park, P. Isola, A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks.", Proceedings of the IEEE international conference on computer vision, pp. 2223-2232, (2017).
- [3] Y. Hao, Y. Liu, Z. Wu, L. Han, Y. Chen, G. Chen, L. Chu, S. Tang, Z. Yu, Z. Chen, B. Lai, "Edgeflow: Achieving practical interactive segmentation with edge-guided flow.", Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1551-1560, (2021).