

# 人間の視覚特性に基づいた 深層学習による暗所画像鮮明化に関する研究

守屋広大<sup>1</sup> 楠房子<sup>2</sup> 稲垣成哲<sup>3</sup> 溝口博<sup>1</sup>

**概要**：博物館は様々な文化に触れる機会を提供することで、来場者の知的好奇心を育むことが可能である。このような博物館教育の効果を高めるためには、ユーザーエクスペリエンス(UX)を向上させることが必要である。UXを向上させる試みとして、物体検出を利用した情報提示ARアプリやロボットが注目されている。しかしながら、博物館内は展示物の保存のために照明を抑制しており、この暗所環境下では撮影画像が不鮮明となるため物体検出を行うことは困難である。そこで、本研究では人間の視覚特性に基づいた深層学習による暗所画像の鮮明化を提案する。

**キーワード**：深層学習、画像鮮明化

## 1. はじめに

博物館学習は、文化を体験して様々な興味を育むことが可能であることから、近年重要視されている[1]。この博物館学習の効果を高めるためには、ユーザーエクスペリエンス(UX)を向上させることが必要である[2]。UXを向上させる試みとして、従来は人間が博物館内を案内していたが、人的資源には限りがある。そこで、近年は深層学習を利用した情報提示ARアプリケーションやガイドロボット等が注目されている[3][4]。これらは、深層学習により展示物をリアルタイムに検出することで、展示物の情報を提供するものである。この情報により、利用者は展示物への理解を深めることを可能とし、UXが向上する。しかしながら、博物館内は展示物の保存のために照明を抑制しており、この暗所環境下では撮影画像が不鮮明となるため物体検出を行うことは困難である[5]。そのため、暗所での物体検出に向けて暗所画像の鮮明化が必要となる。

そこで、本研究では人間の視覚特性に基づいた深層学習による暗所画像の鮮明化を提案する。

## 2. 人間の視覚特性に基づいた暗所画像鮮明化

### 2.1 概要

本手法では、人間の視覚特性をモデル化した Retinex 理論[6]に基づいて暗所画像の鮮明化を行う。図1に示す通り、我々人間は、物体に当たり反射した光を目が捉えることで物を見ている。そのため、目が捉えた光は照明光成分と反射率成分に分離することが可能である。この Retinex 理論は画像に対しても適用することが可能であり、以下の式(1)で表される。

$$L = R \circ I \quad (1)$$

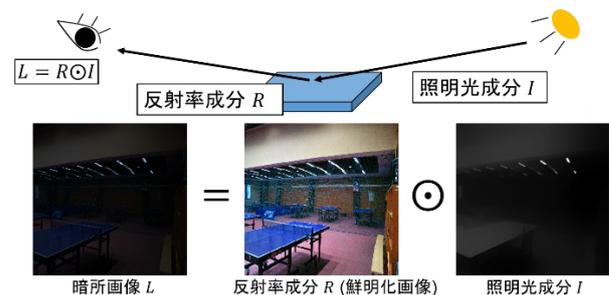


図1 Retinex 理論

ここで、 $L$  は暗所画像であり、 $R$  および  $I$  はそれぞれ反射率成分と照明光成分を表している。また、 $\circ$  はアダマール積である。この反射率成分は、照明光に依存しない物体が本来有している色を表している。そのため、本研究では反射率成分  $R$  の画像を鮮明化画像とみなし、式(2)に従い反射率成分を推定する。

$$R = L \oslash I \quad (2)$$

ここで、 $\oslash$  はアダマール除算である。反射率成分を推定するためには、暗所画像から正確な照明光成分を推定して分離する必要がある。本研究では、画像生成の分野において成果を挙げている GAN(Generative adversarial Network) [7] を用いて、暗所画像から照明光成分を生成した後に、式(2)に基づいて反射率成分を推定する。

### 2.2 ネットワーク構造

本ネットワークは、照明光成分を生成する Generator と入力画像が反射率成分の画像か通常の明所画像かを識別する Discriminator から構成されている。GAN では、これら二つのネットワークを競合させて学習させることにより、最終的に明所画像と見分けがつかないような鮮明化画像を生成することが可能となる。

<sup>1</sup> 東京理科大学  
Tokyo University of Science  
<sup>2</sup> 多摩美術大学  
Tama Art University

<sup>3</sup> 神戸大学  
Kobe University

Generator には、暗所画像  $L$  と Max-RGB 画像  $M$  がチャンネル方向に連結された画像  $L'$  を入力する。Max-RGB 画像  $M$  は、以下の式(3)に従い暗所画像  $L$  から生成される。

$$M = \max_{c \in \{R,G,B\}} L \quad (3)$$

この Max-RGB 画像は、正確な照明光成分に近いことが知られている[6]。そこで、本研究では Max-RGB 画像を暗所画像  $L$  に連結することで、4 チャンネル  $\{R,G,B,M\}$  の連結画像  $L'$  を作成し、鮮明化性能の向上を図る。Generator はこの連結画像を入力として、3 チャンネル  $\{R,G,B\}$  の照明光成分  $I^c$  を生成する。そして、Retinex 理論に基づき、暗所画像と照明光成分から鮮明化画像である反射率成分  $R^c$  を推定する。

Discriminator には、生成された反射率成分と明所画像がランダムに入力される。Discriminator は入力画像が明所画像である確率を出力し、その結果に基づいて Generator と Discriminator を学習させる。

## 2.3 損失関数

### 2.3.1 Adversarial Loss

GAN において、Generator は Discriminator を誤識別させるように学習する。一方、Discriminator は Generator が生成した画像と明所画像とを正しく識別するように学習する。これらの学習は、Adversarial Loss を最小化するように行われる。本研究では Adversarial Loss として、RaGAN [8] と LSGAN [9] を組み合わせたものを採用した。RaGAN の Discriminator は、入力された画像が生成された画像よりも本物に近い確率を出力する。この Discriminator の損失関数は以下の式(4)で表される。

$$\begin{aligned} D_{Ra}(x_r, x_f) &= \sigma(C(x_r) - \mathbb{E}_{x_f \sim P_{fake}}[C(x_f)]) \\ D_{Ra}(x_f, x_r) &= \sigma(C(x_f) - \mathbb{E}_{x_r \sim P_{real}}[C(x_r)]) \end{aligned} \quad (4)$$

ここで、 $x_r$  および  $x_f$  はそれぞれ明所画像と鮮明化画像を表している。 $P_{real}$  および  $P_{fake}$  はそれぞれ明所画像と鮮明化画像のデータ分布である。また、 $C$  と  $\sigma$  は Discriminator とシグモイド関数である。式(4)に対して LSGAN を適用させることで、最終的な Adversarial Loss は以下の式(5)で表される。

$$\begin{aligned} \mathcal{L}_G^{enc} &= \frac{1}{2} \mathbb{E}_{x_f \sim P_{fake}} \left[ (D_{Ra}^{enc}(x_f, x_r) - 1)^2 \right] \\ &\quad + \frac{1}{2} \mathbb{E}_{x_r \sim P_{real}} [D_{Ra}^{enc}(x_r, x_f)^2] \\ \mathcal{L}_G^{dec} &= \frac{1}{2} \mathbb{E}_{x_f \sim P_{fake}} \left[ \left( [D_{Ra}^{dec}(x_f, x_r)]_{i,j} - 1 \right)^2 \right] \\ &\quad + \frac{1}{2} \mathbb{E}_{x_r \sim P_{real}} \left[ [D_{Ra}^{dec}(x_r, x_f)]_{i,j}^2 \right] \end{aligned} \quad (5)$$

### 2.3.2 Self-Color Preserving Loss

生成した鮮明化画像の色を保持するために、Self-Color Preserving Loss を提案する。3ch の照明光成分による鮮明化は、それぞれの ch に対して別々の照明光成分を作用させるため、画像の色を保持することが困難である。そこで、本研究では Generator によって生成される照明光成分  $I^c$  とは別に 1ch の照明光成分による鮮明化画像と比較することで、色の保持を可能とする。ここで、1ch の照明光成分は以下の式(6)で表される。

$$R = L \oslash \left( 1 / \left( 1 + \left( \frac{\bar{L}^\alpha}{\exp(M)} \right)^5 \right) \right) \quad (6)$$

ここで、 $\bar{L}$  は暗所画像の全画素値の平均値である。また、 $\alpha$  は定数であり本研究では  $\alpha = -0.175$  とした。最終的に、Self-Color Preserving Loss は MSE Loss を用いて以下の式(7)で表される。

$$\mathcal{L}_{scp} = \text{MSE}(R^c, R) \quad (7)$$

ここで、 $R^c$  は 3ch の照明光成分によって得られた反射率成分である。

### 2.3.3 Color Constancy Loss

生成した鮮明化画像の色の恒常性を保つために、Stretch アルゴリズム[10]に基づいた Color Constancy Loss を提案する。この損失関数は以下の式(8)で表される。

$$\mathcal{L}_{col} = \sum_{c \in \{R, G, B\}} \left( R^c - \frac{1}{1 - \min L^c} \right)^2 \quad (8)$$

### 2.3.4 Self-Feature Preserving Loss

鮮明化画像の構造を保つために、Self-Feature Preserving Loss [11] を用いる。この損失関数は、画像の構造に関する特徴を有している VGG-16 の中間層を比較することで、画像の構造を保つ。Self-Feature Preserving Loss は以下の式(9)で表される。

$$\begin{aligned} \mathcal{L}_{sfp} &= \frac{1}{W_{k,l} H_{k,l}} \sum_{i=1}^{W_{k,l}} \sum_{j=1}^{H_{k,l}} \left( (\phi_{k,l}(L) - \phi_{k,l}(R^c))^2 \right. \\ &\quad \left. + (\phi_{k,l}(L) - \phi_{k,l}(I^c))^2 \right) \end{aligned} \quad (9)$$

ここで、 $\phi_{k,l}$  は学習済みの VGG-16 を表している。また、 $k$  および  $l$  はそれぞれ  $k$  番目のプーリング層と  $l$  番目の畳み込み層を表しており、 $W_{k,l}$  および  $H_{k,l}$  はそれぞれ特徴マップのサイズである。本研究では、鮮明化画像と照明光成分の両方に対して構造の制約を課している。

### 2.3.5 Exposure Control Loss

鮮明化画像の最適な露光補正のために, Exposure Control Loss [12]を用いる. この損失関数は, 画像内のある領域における平均輝度が定数 $E$ となるように, 制約を課す. 本研究では,  $E = 0.6$ とした. 損失関数は以下の式(10)で表される.

$$\mathcal{L}_{exp} = \frac{1}{K} \sum_{m=1}^K |Y_m - E| \quad (10)$$

ここで,  $K$ は局所領域の数であり,  $Y_m$ は領域内の平均輝度である.

### 2.3.6 Illumination Smoothness Loss

鮮明化画像における周囲の画素間との連続性を保つために, Illumination Smoothness Loss [12]を用いる. この損失関数は以下の式(11)で表される.

$$\mathcal{L}_{tv} = \sum_{c \in \xi} (|\nabla_x R^c| + |\nabla_y R^c|)^2, \xi = \{R, G, B\} \quad (11)$$

ここで,  $\nabla_x$ および $\nabla_y$ はそれぞれ水平および垂直方向の微分作用素である.

### 2.3.7 Spatial Consistency Loss

鮮明化画像の空間的な一貫性を保つために, Spatial Consistency Loss [12]を用いる. この損失関数は以下の式(12)で表される.

$$\mathcal{L}_{spa} = \frac{1}{J} \sum_{i=1}^J \sum_{j \in \Omega(i)} (|(R_i^c - R_j^c)| - |(L_i - L_j)|)^2 \quad (12)$$

ここで,  $J$ は局所領域の数である. また,  $\Omega(i)$ は位置 $i$ における近傍領域を表している.

### 2.3.8 Total Loss

ネットワークは, 以下の式(13)で表される Total Loss を最小化するように学習を行う.

$$\mathcal{L}_{total} = \mathcal{L}_G^{enc} + \mathcal{L}_G^{dec} + \mathcal{L}_{scp} + \mathcal{L}_{col} + \mathcal{L}_{sfp} + \mathcal{L}_{exp} + \mathcal{L}_{tv} + \mathcal{L}_{spa} \quad (13)$$

## 3. 評価実験

### 3.1 評価方法

提案手法の有効性を確認するために, 完全参照の画質評価指標である PSNR(Peak Signal-to-Noise Ratio)および SSIM(Structural similarity)を用いて従来手法との比較を行った. なお, ネットワークの学習には, ExDark データセット[13]から 1165 枚の暗所画像, Pascal VOC dataset [14]から 922 枚の明所画像を用いた. また, 評価には SICE データセ

表 1 従来手法との比較結果

手法	PSNR	SSIM
LIME [16]	15.76	0.83
RetinexNet [17]	16.73	0.81
EnlightenGAN [11]	16.30	0.84
Zero-DCE [12]	16.45	0.79
提案手法	17.05	0.84

ット [15]から 1778 枚のペア画像を用いた.

### 3.2 評価結果

PSNR と SSIM を用いた評価結果について, 表 1 に示す. 表 1 より, 提案手法は全ての従来手法よりも高い値を示した. この結果より, 本研究における鮮明化手法の有効性が確認できた.

## 4. おわりに

本論文では, 人間の視覚特性に基づいた深層学習による暗所画像鮮明化手法について述べた. 本研究では, Retinex 理論に基づいて暗所画像の照明光成分を GAN によって生成することで, 暗所画像の鮮明化を行った. 実験により, 提案手法は従来手法を上回る値が得られた. この結果から, 本研究における鮮明化手法の有効性が示唆された.

### 謝辞

本研究の一部は, JSPS 科研費 JP18H03660 の支援を受けた. 記して謝意を示す.

### 参考文献

- [1] Sung, Y. T., Chang, K. E., Lee, Y. H., and Yu, W. C., Effects of a mobile electronic guidebook on visitors' attention and visiting behaviors, *Educ. Technol. Soc.* 2008, vol. 11, no. 2, p. 67–80.
- [2] Damala, A., and Kockelkorn, H., A taxonomy for the evaluation of mobile museum guides, *ACM Int. Conf. Proceeding Ser.* 2006, vol. 159, p. 273–274.
- [3] Khan, M. A., Israr, S., Almgren, A. S., Din, I. U., Almgren, A., and Rodrigues, J. J. P. C., Using augmented reality and deep learning to enhance Taxila Museum experience, *J. Real-Time Image Process.* 2021, vol. 18, no. 2, p. 321–332.
- [4] Iio, T., Satake, S., Kanda, T., Hayashi, K., Ferreri, F., and Hagita, N., Human-Like Guide Robot that Proactively Explains Exhibits, *Int. J. Soc. Robot.* 2020, vol. 12, no. 2, p. 549–566.
- [5] Cuttle, C., Damage to museum objects due to light exposure, *Int. J. Light Res. Technol.* 1996, vol. 28, no. 1, p. 1–9.
- [6] Land, E. H., The Retinex Theory of Color Vision, *Sci. Am.* 1977, vol. 237, no.6, p.108–128.
- [7] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y., Generative Adversarial Nets, *Advances in Neural Information Processing Systems 27 (NIPS) 2014*, p.2672–2680.
- [8] Jolicœur-Martineau, A., The relativistic discriminator: a key element missing from standard GAN, *International Conference on Learning Representations*, 2019.
- [9] Mao, X., Li, Q., Xie, H., Lau, Y. K. R., Wang, Z., Smolley, P. S., Least Squares Generative Adversarial Networks, *IEEE*

- International Conference on Computer Vision (ICCV), 2017, p.2794-2802.
- [10] Nikitenko, D., Wirth, M., and Trudel, K., Applicability of white-balancing algorithms to restoring faded colour slides: An empirical evaluation, *J. Multimed.* 2018, vol. 3, no. 5, p. 9–18
  - [11] Jiang, Y. et al., EnlightenGAN: Deep Light Enhancement without Paired Supervision, *IEEE Trans. Image Process.* 2021, vol. 30, p. 2340–2349.
  - [12] Guo, C. et al., Zero-reference deep curve estimation for low-light image enhancement, *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 2020, p. 1777–1786.
  - [13] Loh, Y. P., and Chan, C. S., Getting to know low-light images with the Exclusively Dark dataset, *Comput. Vis. Image Underst.* 2019, vol. 178, p. 30–42
  - [14] Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., and Zisserman, A., The pascal visual object classes (VOC) challenge, *Int. J. Comput. Vis.* 2010, vol. 88, no. 2, p. 303–338.
  - [15] Cai, J., Gu, S., and Zhang, L., Learning a Deep Single Image Contrast Enhancer from Multi-Exposure Images, *IEEE Trans. Image Process.* 2018, vol. 27, no. 4, p. 2049–2062
  - [16] Guo, X., Li, Y., and Ling, H., LIME: Low-light image enhancement via illumination map estimation, *IEEE Trans. Image Process.* 2017, vol. 26, no. 2, p. 982–993
  - [17] Wei, C., Wang, W., Yang, W., and Liu, J., Deep retinex decomposition for low-light enhancement, *Br. Mach. Vis. Conf.* 2018, BMVC 2018