

## 機械学習によるグレースケール画像からの擬似分光画像の生成に関する研究

脇田 拓弥 指導教員：村上 和人

## 1 はじめに

カラーリゼーションにおける機械学習モデルの作成にはグレースケール画像とカラー画像を対としたデータセットが必要である。カラー画像は一般的に RGB カメラで撮影された画像であるため、RGB カメラによるデバイス固有の信号値に依存する。また、RGB カメラと人間の目による分光感度が異なるため、現実とカラー画像の色合いが異なるように見える場合もある。さらに RGB 表色系から XYZ 表色系や  $L^*a^*b^*$  表色系に変換してカラー画像を扱う場合、RGB カメラから出力された RGB 値が同じでも一般的にはスペクトルは異なるため、すべての色を正確に変換できないことが課題として挙げられる。そのため正確な色表現には、光の波長毎に強度を表したスペクトル画像（分光画像）を用いてカラー画像を生成する必要がある。

分光画像はマルチ・ハイパースペクトルカメラと呼ばれる多波長分光カメラを用いて取得可能である。しかし、導入コストが高いことや動画の撮影に不向きなことから普及には至っていない。擬似的な分光画像の生成は、分光画像の活用を容易にすると考えられる。機械学習を用いたカラー画像からの分光画像の再構成が提案されている [1]。しかしながら、グレースケール画像からの分光画像の再構成は提案されていない。

そこで本研究では、機械学習によるグレースケール画像から擬似分光画像を生成する手法について提案する。本手法により、正確な色表現とデバイス非依存のカラー画像生成が期待できる。

## 2 基本アイデア

グレースケール画像、カラー画像、分光画像 (Spectral Image: SI) の関係性を説明する。グレースケール画像は輝度のみの 1 チャンネル情報である。カラー画像は明度、彩度、色相情報の 3 チャンネル情報である。分光画像は最短波長  $\lambda_s$  [nm]、最長波長  $\lambda_t$  [nm] とした波長範囲  $[\lambda_s, \lambda_t]$  ( $\lambda_s < \lambda_t$ ) を波長間隔  $\Delta\lambda$  [nm] で割った  $C = \frac{(\lambda_t - \lambda_s)}{\Delta\lambda}$  の  $C$  チャンネル情報である。また、グレースケール画像やカラー画像の作成に必要な情報を含む。さらに反射した光束の分光密度と入射した光束の分光密度の比である分光反射率画像 (Spectral Reflectance Image: SRI) を用いると、人間の分光感度と類似したカラー画像を生成可能である。そこで多波長分光カメラ用いなくともグレースケール画像と分光画像の関係を機械学習モデルに学習させることで無数の組み合わせがあるスペクトルを導出可能だと考えた。本研究では、1 チャンネル情報から  $C$  チャンネル情報を求める不良設定問題を解くため、GAN の一種である pix2pix [2] を用い、グレースケール画像からの擬似的な分光反射率画像を生成する。

## 3 生成手法

GAN の派生モデル pix2pix [2] によるグレースケール画像からの擬似分光反射率画像の生成概要を図 1 に示す。真値の分光反射率画像 (実分光反射率画像) から三刺激値を成分とする XYZ 表色系に変換するブロックと、生成器 (Generator) および識別器 (Discriminator) 各々の学習ブロックから成る。

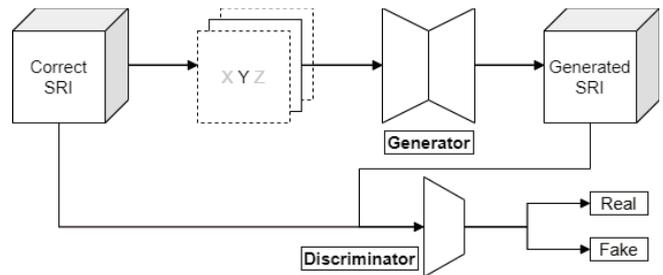


図 1: 擬似分光反射率画像の生成概要図

## 3.1 分光反射率画像から XYZ 表色系への変換

実分光反射率画像 (Correct SRI) から XYZ 表色系への変換は、

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \frac{1}{k} \int_{\lambda_s}^{\lambda_t} I(\lambda) R(\lambda) \begin{bmatrix} \bar{x}(\lambda) \\ \bar{y}(\lambda) \\ \bar{z}(\lambda) \end{bmatrix} d\lambda \quad (1)$$

$$k = \int_{\lambda_s}^{\lambda_t} I(\lambda) R(\lambda) \bar{y}(\lambda) d\lambda \quad (2)$$

のように計算する。ここで、 $I(\lambda)$  は光源の相対分光分布、 $R(\lambda)$  は分光反射率、 $\bar{x}(\lambda)$ 、 $\bar{y}(\lambda)$ 、 $\bar{z}(\lambda)$  は、X、Y、Z の三刺激値に対応した等色関数である。また、係数  $k$  は、輝度情報を持つ Y の範囲を  $0 < Y < 1$  に正規化するためのものである。さらに生成器への入力のために、三刺激値を正規化する必要がある。そのため式 (1) で  $R(\lambda) = 1$  とおいた式から光源の白色点  $X_w$ 、 $Y_w$ 、 $Z_w$  の三刺激値を導出する。三刺激値の正規化は

$$X' = \frac{X}{X_w}, Y' = \frac{Y}{Y_w}, Z' = \frac{Z}{Z_w} \quad (3)$$

のように計算する。ここで、 $X'$ 、 $Y'$ 、 $Z'$  は正規化された XYZ 値である。

## 3.2 生成・識別器学習

生成・識別器のネットワーク構造を図 2 に示す。図 2(a) は生成器、図 2(b) は識別器を示す。

生成器はグレースケール画像として輝度情報を表す三刺激値 Y を入力して、擬似分光反射率画像 (Generated SRI) を出力する。生成器のネットワークは畳み込みブロック (Conv) からなるエンコーダ (Encoder) と逆畳み込みブロック (DeConv) からなるデコーダ (Decoder) によって構成する。また、エンコーダとデコーダのレイヤー (ブロック) 数は  $N$  ( $N \geq 1$ ) である。畳み込みブロックは畳み込み層、活性化関数、正規化層の順で構成する。また、逆畳み込みブロックは逆畳み込み層、活性化関数、正規化層の順で構成する。エンコーダからの追加情報と勾配消失防止にエンコーダとデコーダ間を接続するスキップ接続 (Skip Connection) を用いる。

識別器は入力された実・擬似分光反射率画像に対して、出力画像の画素毎に本物 (Real) を 1、偽物 (Fake) を 0 とした確率  $[0, 1]$  を付ける。識別器のネットワークは生成器の畳み込みブロックを用いて構成する。また、レイヤー (ブロック) 数は  $M$  ( $M \geq 1$ ) である。

生成器は識別器に対して、偽物を本物と騙すように学習する。一方、識別器は本物と偽物を正確に見抜くように学習する。

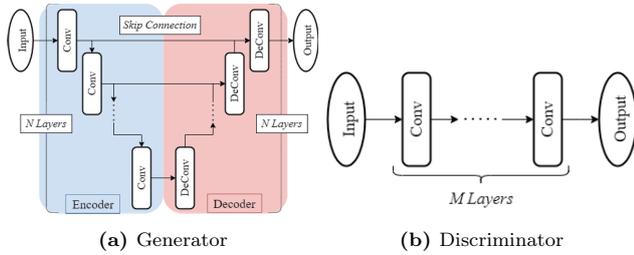


図 2: 生成器と識別器のネットワーク構造図

## 4 評価実験

本実験ではグレースケール画像およびカラー画像からの擬似分光反射率画像の生成を行う。また、客観的評価指標を用いた実・擬似分光反射率画像の評価と分光反射率画像から作成する RGB 画像の評価を有効性確認のため行う。

### 4.1 実験方法

本実験で用いた学習・テストデータセットは、409 枚の分光反射率画像を含む KAUST\_HS[3] と 31 枚の分光反射率画像を含む CAVE[4] とした。全データセットの分光反射率画像は最短波長  $\lambda_s = 400\text{nm}$ 、最長波長  $\lambda_t = 700\text{nm}$ 、波長間隔  $\Delta\lambda = 10\text{nm}$ 、チャンネル数  $C = 31$  とした。また、学習データセットの 81 枚 (20%) は、検証データセットとした。ハイパーパラメータは生成器のレイヤ数  $N = 3$ 、識別器のレイヤ数  $M = 1$ 、200 エポック、バッチサイズ 8 とした。モデルの最適化には生成器の学習率を 0.0007、識別器の学習率を 0.0002 とした Adam を用いた。

### 4.2 実験結果

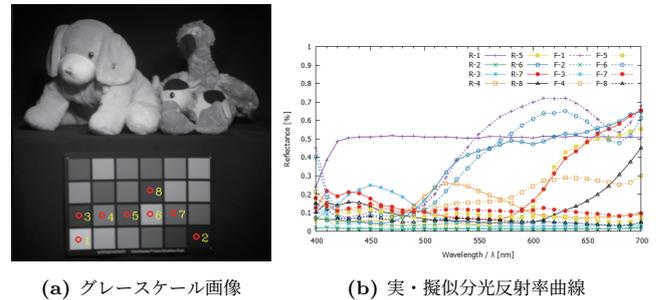
表 1 に、グレースケール画像 (Gray) またはカラー画像 (Color) を生成器に入力して得られた擬似分光反射率画像と、実分光反射率画像において客観的評価指標による評価結果の一例を示す。表 1 の矢印は上向きであれば高い数値ほど良く、下向きであれば低い数値ほど良い結果である。図 3 に、グレースケール画像の特定点に対応した実・擬似分光反射率曲線の一例を示す。図 3(a) のグレースケール画像の赤い丸で表記した特定点に振られた番号は図 3(b) の凡例における番号に対応する。また、実線は実分光反射率曲線、破線は擬似分光反射率曲線を表す。図 4(a) に擬似分光反射率画像、図 4(b) に実分光反射率画像から生成した RGB 画像の一例を示す。

## 5 おわりに

本研究では機械学習を用いたグレースケール画像からの擬似分光画像の生成手法を提案した。提案手法の有効性を確認するための評価実験を行った。実験結果の表 1 から、RMSE の数値差が 0.005 であるため、画素単位における反射率の誤差は小さかった。また、SSIM はほぼ一致していたことから、分光反射率画像の構造は類似していると考えられる。一方、PSNR の数値差が大きいため、画素単位における反射率の誤差は小さいが全体的な誤差は大きかったと考えられる。また、図 3(b) から実分光反射率と擬似分光反射率の双方における反射率が 0.15 未満の場合は誤差が小さかった。しかし、反射率が 0.15 以上の場合は実分光反射率と擬似分光反射率は大きく乖離して、誤差が大きくなった。特に図 3(b) では、波長 600nm 以上の反射率の誤差が大きかったことから、反射率の大きな変化に本研

表 1: 実・擬似分光反射率画像の評価結果の一例

Input	RMSE ↓	SSIM ↑	PSNR ↑
Color	0.040	0.824	28.054
Grayscale	<b>0.089</b>	<b>0.823</b>	<b>21.063</b>



(a) グレースケール画像

(b) 実・擬似分光反射率曲線

図 3: グレースケール画像に対応する実・擬似分光反射率曲線の一例



(a) 擬似分光反射率画像より生成

(b) 実分光反射率画像より生成

図 4: 生成した RGB 画像の比較

究の学習モデルは適応できていないと考えられる。さらに図 4 から、RGB 画像の構造はほぼ同一であったが、色表現は大きく異なった。理由として、先に述べた反射率の大きい変化に適応できなかったためと考えられる。以上の結果から、異なる機械学習モデルの活用や全体的な反射率の誤差を考慮した学習が今後の課題である。

## 参考文献

- [1] B. Arad *et al.*, “NTIRE 2022 Spectral Recovery Challenge and Data Set”, in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp.863–881, 2022.
- [2] P. Isola, J. Y. Zhu, T. Zhou, A. A. Efros, “Image-to-Image Translation with Conditional Adversarial Networks”, in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1125–1134, 2017.
- [3] Y. Li, Q. Fu, W. Heidrich, “Multispectral illumination estimation using deep unrolling network”, in *Proc. IEEE International Conference on Computer Vision*, pp. 2672–2681, 2021.
- [4] F. Yasuma, T. Mitsunaga, D. Iso, S. K. Nayar, “Generalized Assorted Pixel Camera: Postcapture Control of Resolution, Dynamic Range, and Spectrum”, *IEEE Transactions on Image Processing*, vol. 19, no. 9, pp. 2241–2253, 2010.