

深層学習を用いた基底関数表現によるスペクトル超解像

三家本 雄貴^{1,a)} 坂本 真啓² Raytchev Bisser² 金田 和文²

概要：スペクトル画像は多数のバンドを持つため、様々な分野で用いられている。しかし、スペクトル画像の取得にはコストがかかるといった問題がある。そのため、RGB 画像からスペクトル画像を生成するスペクトル超解像が注目されている。近年、深層学習を用いてスペクトル超解像を行う手法が開発されているが、それらの多くは離散表現で表されたスペクトルを用いている。自然光や白熱電球などの光源スペクトルや、通常の物体からの反射スペクトルは連続的なスペクトルとなる。本研究では基底関数の和を用いてスペクトルを連続的に表現する方法を用い、深層学習によりスペクトル超解像を行う手法を開発する。そして、提案手法と従来の離散表現との精度の違い等について検討する。

1. はじめに

1.1 スペクトル画像とスペクトル超解像

私たちが普段利用している画像は R, G, B の 3 チャンネルで構成されている。一方で、スペクトル画像は数十バンド以上の波長情報を含んだ画像である。RGB 画像と比べて画像に含まれている情報量が多いので、品質管理、農業等の様々な分野で活用されている。しかし、スペクトル画像を取得するためには、マルチスペクトルカメラなどの特殊なカメラが必要であり、スペクトル画像の取得にはコストがかかってしまう。

そのため、RGB 画像からスペクトル画像を生成するスペクトル超解像が注目されている。スペクトル超解像は入力画像の空間分解能を向上させるのではなく、スペクトル分解能を向上させる技術である。スペクトル超解像を用いて一般的なカメラで撮影された RGB 画像からスペクトル画像に変換することにより、低コストでスペクトル画像を取得することができる。グレースケール画像を RGB 画像へと変換する研究もスペクトル超解像を扱った手法と考えることができる [1]。

1.2 スペクトルの表現方法

スペクトルの表現方法は 2 つに大別することができる。

1.2.1 離散表現

スペクトルを離散サンプルすることにより表現する方法である。通常は一定間隔の波長ごとにサンプリングが行われる。スペクトルを忠実に表現するには、サンプリング間

隔を小さくしなければならないためデータ量が増大する。

また、深層学習によりスペクトル超解像を行う場合、入力として与えられる画像がスペクトルの離散表現を用いているため、出力のスペクトル超解像画像もスペクトルを離散表現する方法をとることが多い。すなわち、既存手法の多くがエンドツーエンドでスペクトル画像を学習する方法を採用している。

1.2.2 連続表現

連続的にスペクトルを表現する方法として、基底関数の和による表現がある。適切な基底関数を選択することにより少数の基底関数でスペクトルを高精度に表現することが出来る。また、基底関数は少数の係数で表現できるため、スペクトルを表現するデータ量は通常、離散表現に比べて少ない。

シングモイド関数を基底関数としてスペクトルを表現し、最適化問題を解くことにより、RGB 値からスペクトルに変換する手法が提案されている [2]。

本論文では深層学習を用いたスペクトル超解像において、従来の離散表現によるスペクトル超解像ではなく、基底関数の和を用いた連続表現によるスペクトル超解像を提案する。提案手法である基底関数の和を用いたスペクトル表現方法は従来の離散表現と比べて精度を向上させることができる。

2. 関連研究

2.1 辞書を用いたスペクトル超解像

RGB 画像からスペクトル画像に変換するための辞書を構築することによりスペクトル超解像を行う手法が開発

¹ 広島大学 工学部第二類情報工学課程 ビジュアル情報学研究室

² 広島大学 大学院先進理工系科学研究科

^{a)} yuki-mikamoto@hiroshima-u.ac.jp

されている。Arad ら [3] は画像構成のための辞書構築に K-SVD アルゴリズム [4] を利用した。後述する深層学習を用いたスペクトル超解像と比べて高速に処理できるという利点があるが、再構成精度は劣る。

2.2 深層学習を用いたスペクトル超解像

深層学習を用いた手法が近年主流となっており、RGB 画像からスペクトル画像に変換するマッピング関数をネットワークで学習することによりスペクトル画像を生成する。スペクトル超解像に用いるネットワークはセマンティックセグメンテーション用のモデルを改良したものが用いられることが多い。これはネットワークの出力画像サイズが入力画像サイズと同じためエンドツーエンドで学習できることと、対象ピクセルの周りの情報を考慮したスペクトルの生成ができるからである。

Galliani ら [5] は Fully Convolutional DenseNets[6], Yan ら [7] は U-Net[8] を改良したネットワークをそれぞれ利用している。これらの手法はスペクトル画像を直接出力するため、スペクトルは離散表現となる。

一方、坂本 ら [9] はスペクトルを輝度とスペクトル類似度に分けてスペクトル類似度のみを学習し、輝度を後からスケールアップすることによって精度向上を図った。その際に従来の離散表現だけではなく、基底関数の和によるスペクトルの連続表現を試みたが、基底関数の種類や数について十分に検討が行われていなかった。

3. 提案手法

坂本 ら [9] が提案した基底関数の和によるスペクトルの連続表現に注目し、基底関数の種類、数を変えることによってスペクトルの離散表現よりも精度の向上を図ることができる基底関数を見つけ、深層学習を用いた基底関数表現によるスペクトル超解像手法を確立する。

3.1 スペクトル画像の生成方法

坂本らの手法に基づいた方法でスペクトル画像を生成する。全体の流れは図 1 の通りである。

RGB 画像を入力するとニューラルネットワークはピクセル毎にスペクトル表現に用いる基底関数の各係数を出力する。このため、ネットワークの出力チャンネル数は係数の数と等しくなる。次にネットワークから出力された係数から基底関数を用いてスペクトル画像を生成するが、このスペクトル画像の輝度は正規化されている。そのため、入力された RGB 画像を利用して輝度変換を行うことにより最終的なスペクトル画像を得る。

輝度変換には XYZ 表色系の Y 成分を利用する。入力画像の Y 成分は (1) 式により算出される。また、正規化スペクトル画像のスペクトル $\hat{S}(\lambda)$ の Y 成分 ($\hat{Y}$) は等色関数 $\hat{y}$ を用いて (2) 式により算出される。

$$Y = 0.2126 \times R + 0.7152 \times G + 0.0722 \times B \quad (1)$$

$$\hat{Y} = \frac{1}{k} \int_{\Lambda} \hat{S}(\lambda) \bar{y}(\lambda) d\lambda \quad (k = \int_{\Lambda} \bar{y}(\lambda) d\lambda) \quad (2)$$

輝度変換後のスペクトル $S(\lambda)$ は (1) 式と (2) 式から得られた各 Y 成分を用いて次式により算出される。

$$S(\lambda) = \frac{Y}{\hat{Y}} \hat{S}(\lambda) \quad (3)$$

3.2 実験に用いるネットワーク

実験に用いるネットワークは、Galliani らの手法 [5] や坂本らの手法 [9] と同様に Fully Convolutional DenseNets[6] を用いる。このネットワークは、Jégou ら [6] が Densely Connected Convolutional Networks (DenseNet) [10] をセマンティックセグメンテーション用に完全畳み込みネットワークに拡張したものである。

損失関数には次式に示す CosineEmbeddingLoss (CosineE) を用い、ネットワークはスペクトル類似度を損失関数として学習する。

$$\text{CosineE}(a, b) = 1 - \frac{a \cdot b}{\|a\|_2 \|b\|_2} \quad (4)$$

3.3 スペクトル生成に用いる基底関数

スペクトルを生成するための基底関数は坂本 ら [9] が利用したシグモイド関数 (図 2(a))、正規分布の確率密度関数 (図 2(b)) と、さらにメキシカンハットウェーブレット (図 2(c)) を追加した合計 3 種類を用いた。

- シグモイド関数

Jakob ら [2] は二次多項式とシグモイド関数を組み合わせることによってスペクトルを表現した。

$$f(\lambda) = S(c_0 \lambda^2 + c_1 \lambda + c_2) \quad (5)$$

$$S(x) = \frac{1}{2} + \frac{x}{2\sqrt{1+x^2}} \quad (6)$$

この表現方法は二次多項式をシグモイド関数の引数とすることで値域が $[0, 1]$ となるため、正規化されたスペクトルを表現できるという利点がある。ネットワークは 1 つの関数につき、係数 c_0, c_1, c_2 と関数の重み ((9) 式の w_i) を含め 4 つの係数を学習する。

- 正規分布の確率密度関数

自然界に存在するスペクトルは一般的に連続で緩やかな変化をするという特性を持つ [11] ため、その特性に近い正規分布の確率密度関数を利用する。

$$f(\lambda) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(-\frac{(\lambda - c)^2}{2\sigma^2}\right) \quad (7)$$

ネットワークは 1 つの関数につき、係数 σ, c と関数の重み ((9) 式の w_i) を含め 3 つの係数を学習する。

- メキシカンハットウェーブレット

ガウス関数を 2 階微分したもので、メキシカンハット

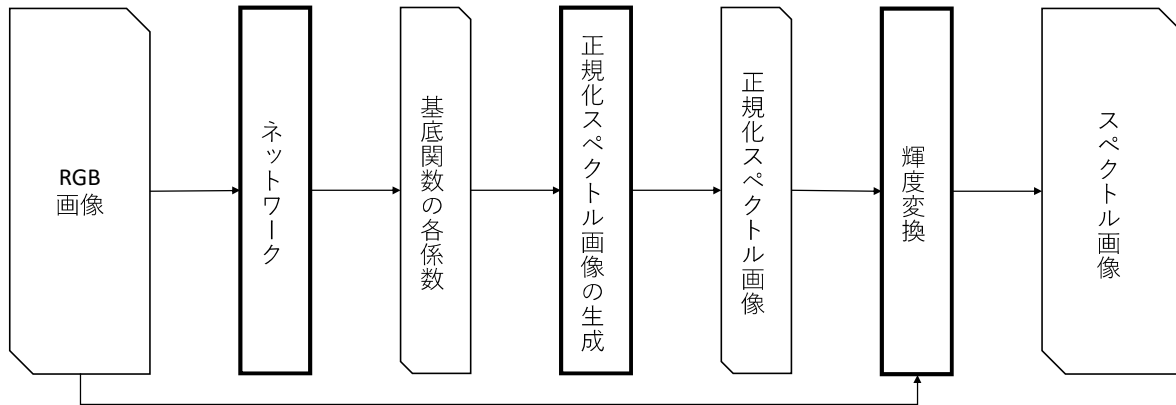
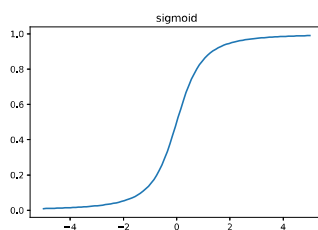
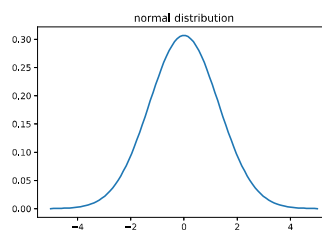


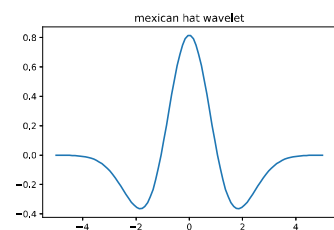
図 1: RGB 画像からのスペクトル画像生成の流れ



(a) シグモイド関数



(b) 正規分布の確率密度関数



(c) メキシカンハットウェーブレット

図 2: スペクトル表現に用いる基底関数

形状をしている関数である。正規分布の確率密度関数と似ているが、関数が負の値も取るので激しい変化をするスペクトルにも対応できると考えられる。

$$f(\lambda) = \frac{1}{\sqrt{a}} \left(1 - 2 \left(\frac{\lambda - c}{a} \right)^2 \right) \exp \left(- \left(\frac{\lambda - c}{a} \right)^2 \right) \quad (8)$$

ネットワークは 1 つの関数につき、係数 a, c と関数の重み ((9) 式の w_i) を含め 3 つの係数を学習する。

スペクトル $S(\lambda)$ はこれらの基底関数を用いて、次式により表される。

$$S(\lambda) = w_1 f_1 + w_2 f_2 + \dots + w_n f_n \quad (9)$$

ここで、 λ は波長、 f_i は基底関数、 w_i は各基底関数にかかる重み、 n は基底関数の数である。

4. 実験と結果

4.1 データセット

実験には CAVE データセット [12] と ICVL データセット [3] を用いる。

CAVE データセット

CAVE データセットはスペクトル超解像で多く用いられているデータセットである。32 枚の画像で構成されており、画像サイズは 512×512、波長 400nm から 700nm を 10nm 幅でサンプリングされた 31 バンドか

らなる。データの数 が 32 枚と少ないため、2 分割交差検証で評価をする。また、学習時間を削減するために画像サイズを 256×256 にダウンサンプリングして実験を行う。

ICVL データセット

ICVL データセットは 201 枚の画像で構成されており、画像サイズは一部を除いて 1300×1392、波長 400nm から 1000nm をおおよそ 1.25nm 幅でサンプリングされた 519 バンドからなる。また、波長 400nm から 700nm を 10nm 幅でサンプリングしてバンド数を 31 に圧縮されたデータも提供されているため本実験では後者のデータを用いる。

実験では画像サイズが 1300×1392 である 185 枚を利用する。検証用、テスト用に 1 割ずつ (19 枚, 18 枚) を無作為に選び、残りの 148 枚を訓練用に使用する。訓練用データはデータ増強を行っておりパッチサイズ 325×348、横方向のストライド 195、縦方向のストライド 261 で分割することにより 20 個のデータが得られる。これに画像サイズを縦、横それぞれ 1/4 にした全体画像を加えることで 1 枚の画像から 21 枚の訓練データが得られ、総数は 3108 枚となる。

テスト時は画像サイズを縦、横それぞれ 1/4 にした全体画像を利用する。

基底関数 (数)	RMSE ↓	SAM ↓
離散	4.817	8.526
シグモイド (5)	4.939	8.383
シグモイド (6)	4.638	8.134
正規分布 (31)	5.037	10.241
メキシカン (25)	4.768	8.620

表 1: CAVE データセットでの各評価値

基底関数 (数)	RMSE ↓	SAM ↓
離散	3.539	2.146
シグモイド (7)	3.617	2.255
シグモイド (31)	3.450	2.087
正規分布 (31)	3.783	2.283
メキシカン (25)	3.731	2.268

表 2: ICVL データセットでの各評価値

4.2 評価指標

評価指標には、ピクセル毎のスペクトル強度値の誤差を評価する二乗平均平方根誤差 (RMSE) とスペクトルの形状分布の誤差を評価する Spectral Angle Mapper (SAM) を用いる。

RMSE は次式で表される。

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (e_i - t_i)^2} \quad (10)$$

ここで、 n は画像のピクセル総数、 t_i は真値スペクトル、 e_i は予測スペクトルである。また、RMSE 算出前にスペクトルを 255 倍して 8bit 幅で表すため、RMSE の値域は 0 から 255 となる。

SAM は次式で表される。

$$SAM = \frac{1}{n} \sum_{i=1}^n \cos^{-1} \left(\frac{e_i \cdot t_i}{\|t_i\|_2 \|e_i\|_2} \right) \quad (11)$$

ここで、 n は画像のピクセル総数、 t_i は真値スペクトル、 e_i は予測スペクトルである。(11) 式で算出される SAM の値はラジアンとなるが、これを角度に変換して SAM の評価指標とする。

4.3 離散表現での学習

提案手法との比較を行うために従来手法の離散表現での実験も行った。離散表現の場合、ネットワークが出力するチャンネルはデータセットのスペクトル画像のサンプリング数に依存するため、ここでは 400nm から 700nm を 10nm 間隔でサンプリングされた 31 チャンネルで表現した。従って、提案手法と同じネットワークを用いるが、出力されるのは基底関数の係数ではなくサンプリングされた波長ごとのスペクトル値となる。

4.4 実験結果

定量的評価

CAVE データセットで得られた結果を図 3 と図 4 に示す。RMSE, SAM とともにスペクトルの表現に用いる基底の数を増やすと精度が向上していくことが分かる。また、シグモイド関数を用いた場合は基底の数を 6 以上とすれば評価指標 RMSE, SAM とともに従来の離散表現より精度が良くなっている。

さらに詳細に検討するために CAVE データセットで

の結果の一部を表 1 に示す。シグモイド関数を用いた場合、基底の数が 5 個のときに評価指標 SAM で従来の離散表現よりも良い結果が得られ、6 個にすることにより評価指標 RMSE でも上回ることが分かる。また、メキシカンハットウェーブレットを用いた場合も基底の数が 25 個のときに離散表現とほぼ同等の精度が得られている。しかし、正規分布の確率密度関数を基底関数として用いた場合は、基底の数を 31 個まで増やしても離散表現を上回ることが出来なかった。

表 2 は ICVL データセットでの結果である。

ICVL データセットの場合、CAVE データセットの時と同じ基底の数では従来の離散表現を上回ることが出来なかった。しかし、シグモイド関数を用いた場合、基底の数を十分に増やす (31 個) と評価指標 RMSE, SAM とともに従来の離散表現よりも良い結果が得られることが分かる。

定性的評価

図 5 の右上図は SAM の評価値が最も良い基底関数の種類をピクセル毎にカラーマッピングした図である。基底関数の数は一番良い SAM の評価値を残したものをを用いた。シグモイド関数が画像領域の多くを占めているが、局所的に他の関数が優れている部分もある。入力画像の紫色の領域のスペクトル (図 5 の右下図)を確認すると、メキシカンハットウェーブレットを用いたスペクトルが元のスペクトル (original) を一番良く再現できていることが分かる。

次に、正規分布の確率密度関数を基底関数として用いた場合の問題点について考察する。図 6 は図 5 と同様に最良の基底関数をカラーマッピング表示した図である。入力画像の白色の領域のスペクトル (図 6 の右下図)に着目すると、どの基底関数でも同じ傾向ではあるが、特に正規分布の確率密度関数は波打つ分布形状のスペクトルを表現できていないことが分かる。

図 5, 図 6 の右上図において黒色の領域が少ないことから基底の数を十分に増やした場合、従来の離散表現よりも精度よくスペクトルを再現できていることも分かる。

最後に基底の数によるスペクトルの違いについて考察する。図 7 の右上図はシグモイド関数のうち基底の数

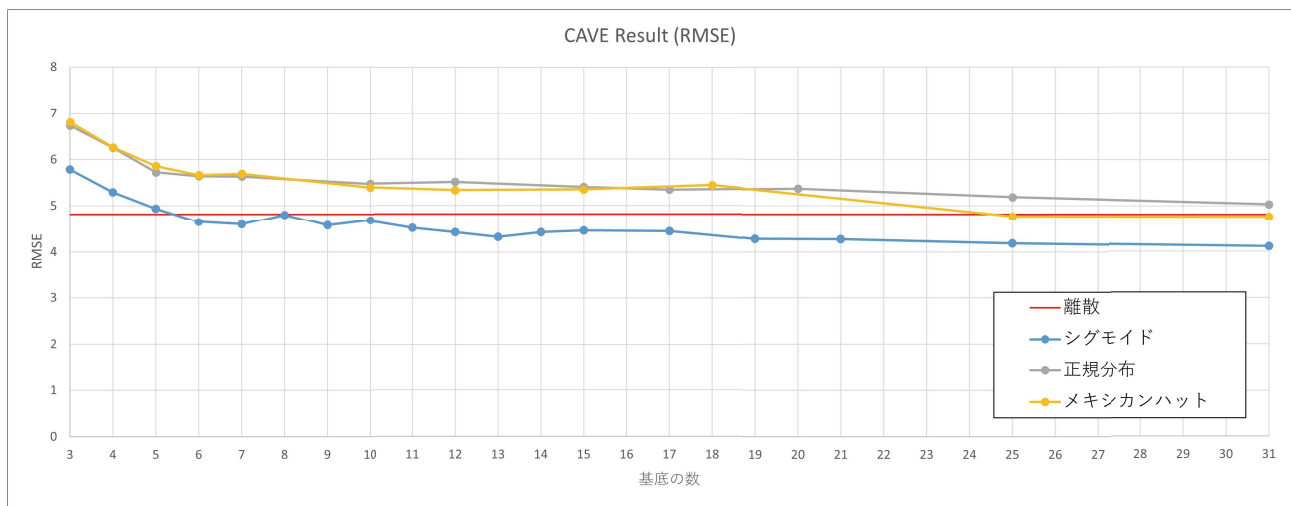


図 3: CAFE データセットの結果 (RMSE)

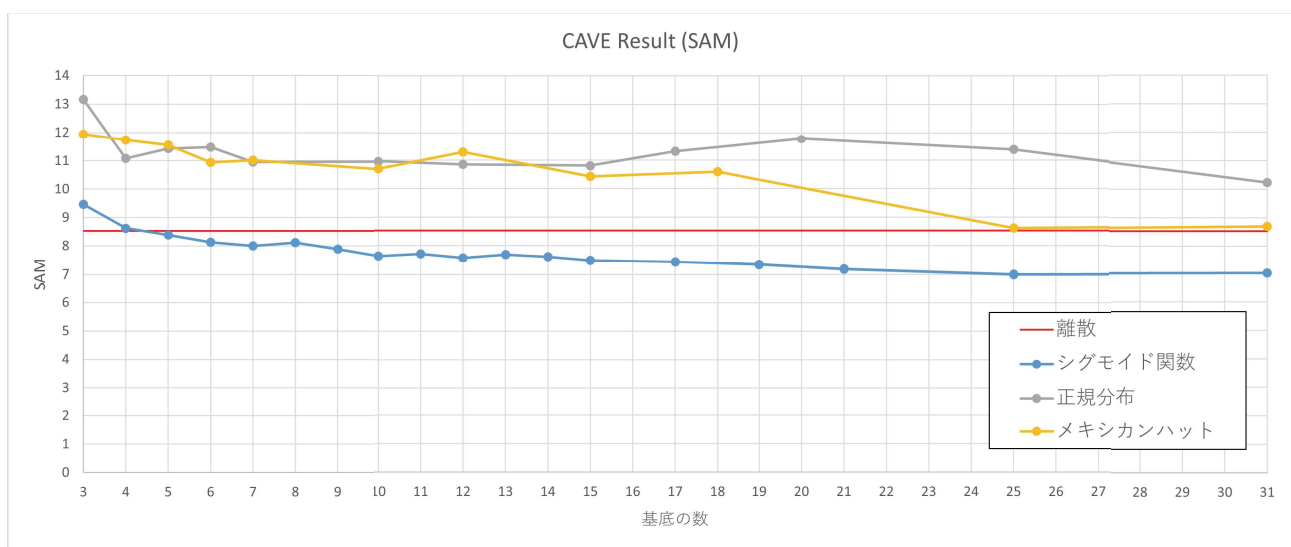


図 4: CAFE データセットの結果 (SAM)

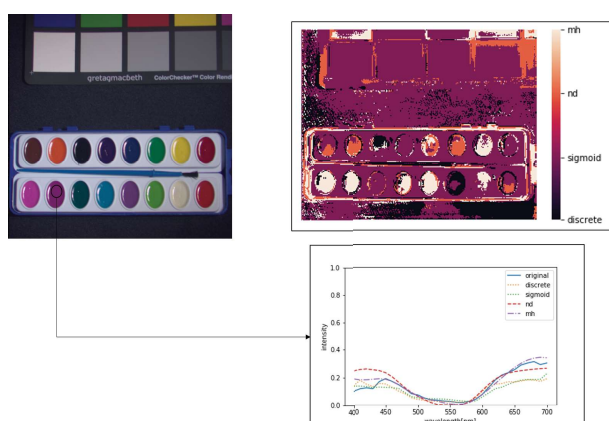


図 5: 基底関数の種類による SAM と波長の比較
discrete は離散, sigmoid はシグモイド関数, nd は正規分布の確率密度関数, mh はメキシカンハットウェーブレットを表している

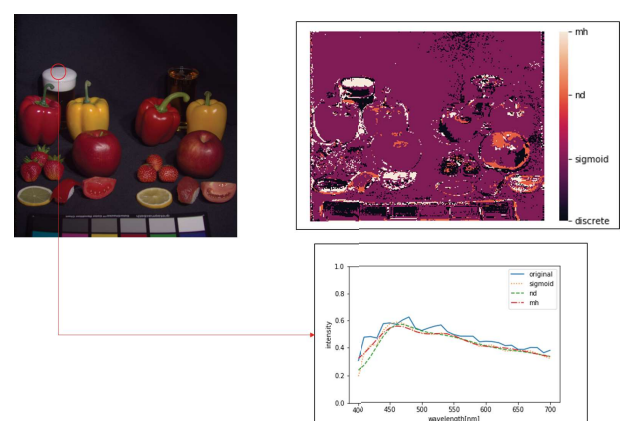


図 6: 正規分布の確率密度関数の問題点

が 3, 7, 31 のうち SAM の評価値が最も良い数をカラーマッピング表示した図である。画像中の大部分の領域は基底の数が最も多い 31 個の場合が優れているが, 3 個の場合が最も優れている領域も存在している。

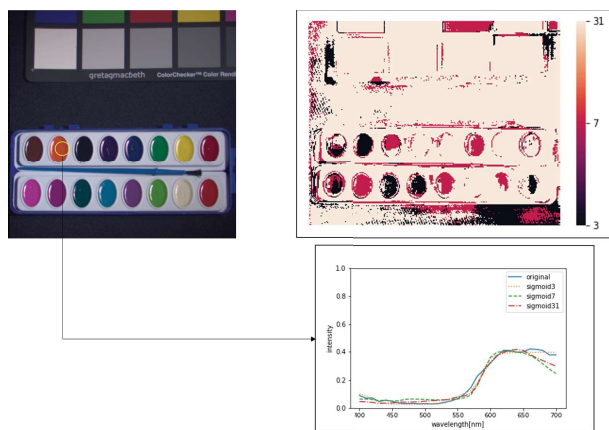


図 7: 基底関数の数による SAM と波長の比較

このような領域である入力画像のオレンジ色の領域のスペクトルを図 7 の右下図に示す。基底の数が 3 個の場合はスペクトル分布が単調となり、その結果元のスペクトルを最もよく再現できていることが分かる。

5. おわりに

本研究では、シグモイド関数あるいはメキシカンハットウェーブレット関数の和を用いたスペクトル表現を行い、深層学習によりスペクトル超解像を行う方法が従来の離散表現と同等、あるいはそれを超える優れた結果を与えることを明らかにした。また、平均的に良い結果をだす基底関数と特定の領域で良い結果をだす基底関数があるということ、領域ごとに最適な基底関数の数が異なるということも分かった。

今後の課題としては、他のスペクトル超解像用のネットワークモデルでも同様の結果が得られるかどうかを検証すること、放電ランプ等の輝線スペクトルをもつ光源環境で撮影された画像などが含まれるデータセットを用いて実験を行うことが挙げられる。さらに、複数種類の基底関数を組み合わせたスペクトル表現方法への拡張やピクセル毎に最適な基底関数を用いてスペクトル超解像を行うネットワークモデルの作成もチャレンジングな研究課題である。

参考文献

- [1] Iizuka, S., Simo-Serra, E. and Ishikawa, H.: 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*, Vol. 35, No. 4, pp. 110:1–110:11 (2016).
- [2] Jakob, W. and Hanika, J.: A Low-Dimensional Function Space for Efficient Spectral Upsampling, *Computer Graphics Forum*, Vol. 38, No. 2 (2019).
- [3] Arad, B. and Ben-Shahar, O.: Sparse Recovery of Hyperspectral Signal from Natural RGB Images, *European Conference on Computer Vision*, pp. 19–34 (2016).
- [4] Aharon, M., Elad, M. and Bruckstein, A.: K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation, *IEEE Transactions on Signal*

- Processing*, Vol. 54, No. 11, pp. 4311–4322 (2006).
- [5] Galliani, S., Lanaras, C., Marmanis, D., Baltasavias, E. and Schindler, K.: Learned spectral super-resolution, *arXiv preprint arXiv:1703.09470* (2017).
- [6] Jégou, S., Drozdal, M., Vazquez, D., Romero, A. and Bengio, Y.: The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation, *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1175–1183 (2017).
- [7] Yan, Y., Zhang, L., Li, J., Wei, W. and Zhang, Y.: Accurate Spectral Super-Resolution from Single RGB Image Using Multi-scale CNN, *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*, pp. 206–217 (2018).
- [8] Ronneberger, O., Fischer, P. and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 234–241 (2015).
- [9] 坂本真啓, Bisser, R., 玉木 徹, 金田和文: スペクトル類似度を考慮した深層学習による RGB 画像からスペクトル画像への変換手法, 研究報告, コンピュータグラフィックスとビジュアル情報学, Vol. 2020-CG-178, No. 5, pp. 1–7 (2020).
- [10] Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K. Q.: Densely Connected Convolutional Networks, *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2261–2269 (2017).
- [11] Maloney, L. T.: Evaluation of linear models of surface spectral reflectance with small numbers of parameters, *J. Opt. Soc. Am. A*, Vol. 3, No. 10, pp. 1673–1683 (1986).
- [12] Yasuma, F., Mitsunaga, T., Iso, D. and Nayar, S.: Generalized Assorted Pixel Camera: Post-Capture Control of Resolution, Dynamic Range and Spectrum, *Technical Report, Department of Computer Science, Columbia University CUCS-061-08* (2008).