Hyperspectral Reconstruction of Skin Through Fusion of Scattering Transform Features

NWC RIT on Applied Harmonic Analysisf

Brandon Kolstoe

Department of Mathematics
University of Maryland - College Park (USA)

May 6, 2024

Collaborators and Acknowledgements

- **Wojciech Czaja**: Department of Mathematics, University of Maryland – College Park (USA)
- **Jeremiah Emidih**: Department of Mathematics, University of Maryland – College Park (USA)
- **Richard G. Spencer**: National Institutes of Health, National Institute on Aging (USA)

- Work funded in part by the Intramural Research Program of the National Institute on Aging of the NIH.

Outline of Presentation

Table of Contents

Hyperspectral Images

- Standard images are stored in computers either as one matrix (grayscale) or 3 matrices (RGB).
  - Entry in each matrix corresponds to a pixel in the image.
    - The value of each entry corresponds to the intensity of the color.
  - In RGB, each matrix refers to a specific color (red, green, or blue).
- A **hyperspectral image** (**HSI**) consists of a large collection of matrices, each corresponding to a different wavelength of light.
  - Different materials react more to different wavelengths, and hence show up more in certain images.
    - Allows for improved identification of materials.
  - Call each image in this collection a **channel**.
- HSI are difficult to obtain (cameras are expensive/ sensitive).

Hyperspectral Images

Figure 1: An example of a hyperspectral image, courtesy of NASA.

Hyper-Skin Dataset

- The **Hyper-Skin dataset** [1] consists of 306 hypserspectral images of human faces.
  - Each image has spatial dimension 1024×1024 and 448 spectral channels.
  - The spectral channels are divided into the visual spectrum (**VIS**) and the near-infrared spectrum (**NIR**).
    - VIS is 400-700 nm, and NIR is 700-1000 nm.
- The images were taken with a hyperspectral camera of 51 participants.
  - 6 images were taken of each participant, depending on facial position (front, left, right) and facial expression (neutral, smile).

[1] Ng et al. "Hyper-Skin: A Hyperspectral Dataset for Reconstructing Facial Skin-Spectra from RGB Images". *Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track*. 2023.

Hyper-Skin Dataset

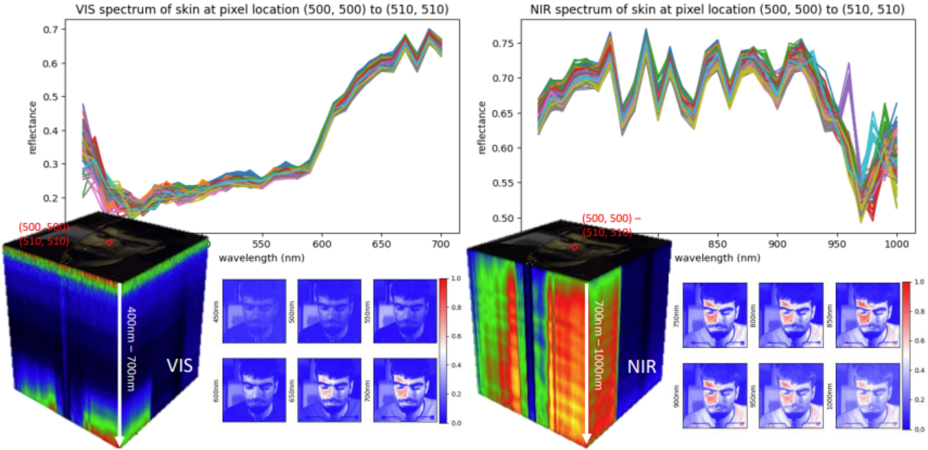


Figure 2: Figure 1 from (Ng et al.), of one of the images in the dataset. The left images are in the VIS spectrum, while the right images are in the NIR spectrum.

Table of Contents

Hyper-Skin Grand Challenge

- Every year, the International Conference on Acoustics, Speech, and Signal Processing (**ICASSP**) holds a number of Signal Processing Grand Challenges (SP GC).
- The goal of the **ICASSP 2024 SP Grand Challenge on Hyperspectral Skin Vision** was to reconstruct the Hyper-Skin dataset from multi-spectral images (**MSI**).
    - Each MSI, corresponding to one of the Hyper-Skin HSI, consists of the corresponding RGB image and the 960 nm near-infrared channel.
    - Want to find a model that will take in the MSI and return the coinciding HSI.

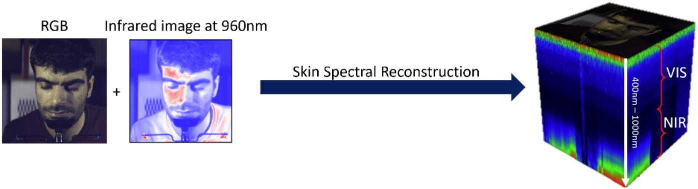


Figure 3: Diagram of the challenge goal, from `https://uoft-hyperskin.github.io/`.

Hyper-Skin Grand Challenge

- Each HSI in the dataset was downsampled in the channel dimension to reduce the number of channels from 448 to 61.
  - So, each HSI is of size $(\mathbf{1024}, \mathbf{1024}, \mathbf{61})$, where the 61 channels correspond to 400-1000 nm wavelengths in 10 nm increments.
  - The corresponding MSI are of size $(\mathbf{1024}, \mathbf{1024}, \mathbf{4})$, consisting of 3 channels from the RGB image and 1 channel of near-infrared.
- This has the appearance of an ill-posed problem, as we're essentially asked to find a map from 4 to 61 dimensions.
  - However, since real-world data tends to lie on a lower-dimensional manifold [2], there's hope that a deep learning approach could solve this problem.

[2] Lei et al. "A Geometric Understanding of Deep Learning". *Engineering*. 2020.

Hyper-Skin Grand Challenge

- 7 participants were removed from the Hyper-Skin dataset to make the **testing dataset**.
  - 12 more MSI images, taken with cameras different from the HSI camera of 2 participants in the training set, were also added to the testing set.
    - Wanted to see how well models generalize over different camera types.
- The **Spectral Angle Mapper** (**SAM**) score was used to test the results of the reconstruction.
  - Given two **spectra** (the vector of pixels from each HSI channel with the same coordinates), their spectral angle is

$$\mathrm{SA}(h_{i,j}, \tilde{h}_{i,j}) := \arccos\left(\frac{\langle h_{i,j}\tilde{h}_{i,j}\rangle}{||h_{i,j}||\ ||\tilde{h}_{i,j}||}\right)$$

  - The SAM score of two images is:

$$\mathrm{SAM}(h, \tilde{h}) = \frac{1}{NM}\sum_{i=0}^{N-1}\sum_{j=0}^{M-1}\mathrm{SA}(h_{i,j}, \tilde{h}_{i,j}))$$

- Only **skin spectra** was compared using the SAM score.

- All submitted models were compared against a baseline method: the **Multi-stage Spectral-wise Transformer** (**MST++**)[3].
    - MST++ placed first in a similar competition at the Computer Vision and Pattern Recognition Conference (CVPR) 2022.

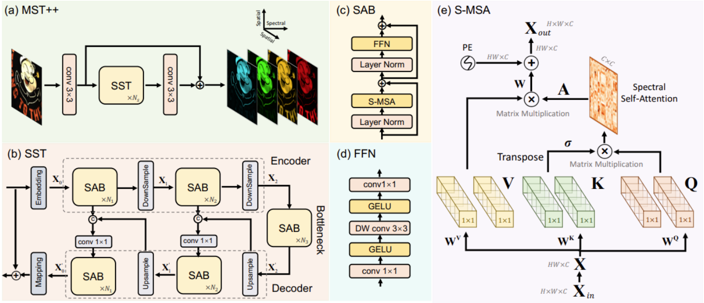


Figure 4: Figure 2 from (Cai et al.); the diagram of the MST++ model, whose main component consists of a spectral-wise transformer block.

[3] Cai et al. "MST++: Multi-stage spectral-wise transformer for efficient spectral reconstruction". *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.

Table of Contents

Our General Framework

- **Data fusion/ modality adaptation** perspective:
  - Have two correlated modalities and want a map between them.
- **General framework**: embed both modalities into a feature space and find a transformation that matches corresponding features.

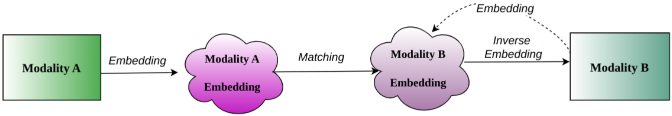


Figure 5: Diagram of general framework.

- Examples:
  - Diffusion map embeddings with rotation-based matching [4].
  - Laplacian Eigenmap embeddings with rotation-based matching [5].
  - Laplacian Eigenmap embeddings with graph matching [6].

[4] Coifman and Hirn. "Diffusion maps for changing data". *Applied and Computational Harmonic Analysis*. 2014.

[5] Cloninger, Czaja, and Doster. "The Pre-Image Problem for Laplacian Eigenmaps Utilizing L_1 Regularization with Applications to Data Fusion". *Inverse Problems*. 2017.

[6] Czaja and Emidih. "Heterogeneous Cancer Cell Line Data Fusion for Identifying Novel Response Determinants in Precision Medicine". Springer International Publishing. *Bioinformatics Research and Applications*. 2017.

Our General Framework

- For the hyperspectral skin reconstruction problem, we use the **scattering transform** [7] to embed the MSI and HSI modalities.
  - The scattering transform is a **feature extractor** with similar structure to a convolutional neural network (CNN) but using predefined directional wavelet filters.
    - Mathematical model of a CNN [8] with provable stability results.
    - Groups **size**, **location**, and **direction** features into scattering coefficients.
- Train a simple CNN to map from MSI to HSI scattering coefficients.
- Train a different CNN to invert the HSI scattering embedding.

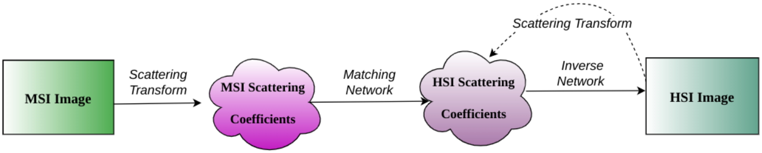


Figure 6: Diagram of specific framework.

[7] Mallat. "Group Invariant Scattering". *Communications in Pure and Applied Mathematics*. 2012.

[8] Mallat. "Understanding Deep Convolutional Networks". *Philosophical Transactions of the Royal Society A*. 2016.

The Scattering Transform

- Let ψ be the mother wavelet and fix $J, L \in \mathbb{N}$.
  - For $j, q \in \mathbb{Z}$, denote $\psi_{j,q}(u) = 2^{-2j}\psi(2^{-j} r_\theta u)$.
    - $\theta = \frac{q\pi}{L}$ and r_θ is corresponding rotation.
  - For ϕ a low-pass filter, let $\phi_J(u) = 2^{-2J}\phi(2^{-J}u)$.
- The (2 layer) **scattering transform** of a 2d image x is given by:

$$Sx = \{x * \phi_J, |x * \psi_{j,q}| * \phi_J, ||x * \psi_{j,q}| * \psi_{j',q'}| * \phi_J\}_{\substack{1 \leq j < j' \leq J \\ 1 \leq q, q' \leq L}}$$

  - $*$ denotes (periodic) convolution.
  - Due to redundant nature of the representation, the output of each scattering channel is downsampled by factor of 2^J.
- The components of the scattering transform are analogous to a CNN.
  - The filters $\psi_{j,q}$ are the (predefined) convolutional filters.
  - $|\cdot|$ is the nonlinearity.
  - ϕ_J is somewhat analogous to a pooling operation.

Properties of the Scattering Transform

- **Lipschitzity**: $|||Sx - Sy||| \leq ||x - y||_2$
  - Here, $|||Sx|||^2 = \sum_{g \in Sx} ||g||_2^2$.
- **Stability**[9]: Let $x_\tau(u) = x(u - \tau(u))$ for $\tau : \mathbb{R}^2 \to \mathbb{R}^2$.
  - If τ is sufficiently regular with $||\nabla\tau||_\infty < \frac{1}{4}$ and x is compactly supported, then

$$|||Sx_\tau - Sx|| \leq C||x||_2(2^{-J}||\tau||_\infty + ||\nabla\tau||_\infty)$$

- **Energy Decay** It is conjectured that the energy (i.e. norm) of each layer of the scattering transform decreases exponentially in the layer number.

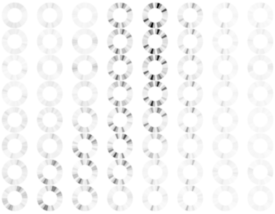
[9] Bruna and Mallat. "Invariant Scattering Convolution Networks". *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2013.

Implementation Details of the Scattering Transform

- The scattering transform is applied to each channel of an image.
  - If image is size (C, M, N) (where C is number of channels and M and N are the width and height), then scattering coefficients are size $(C, S, M/2^J, N/2^J)$, where $S = 1 + JL + \frac{J(J-1)L^2}{2}$.
  - We set **J=2** and **L=4** so that **S=25**, and we have **M=N=1024**.
- Implemented in Python (through package **Kymatio**) and MATLAB (as function **scatteringTransform**).
- In Kymatio, $\psi(u) = C_1 \frac{2}{\pi\sigma^2} e^{-\frac{2(u \cdot Du)}{\sigma^2}} (e^{i2(\xi \cdot u)} - C_2)$ is a Morlet Wavelet.
  - In this case, $\sigma = 0.8$, $C_1 = \frac{4}{L} = 1$, $D = \begin{pmatrix} 1 & 0 \\ 0 & 16/L^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $\xi = \begin{pmatrix} \frac{3}{4}\pi & 0 \end{pmatrix}^T$, and C_2 is chosen so that $\int\limits_{\mathbb{R}^2} \psi(u) du = 0$.
  - Also, $\phi(u) = \frac{2}{\pi\sigma^2} e^{\frac{-2|u|^2}{\sigma^2}}$ is a Gaussian.

Features of the Scattering Transform

(a) First layer scattering coefficients (*one filter applied*).



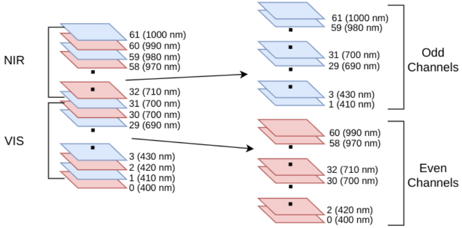
(b) Second layer scattering coefficients (*two filters applied*).

Figure 7: Channel of a testing image (left) and the **scattering coefficients** (center and right) of green highlighted region. In (a), each sector corresponds to a **rotation** and **dilation** ($j = 1$ for the outer layer, $j = 2$ for the inner layer). In (b), each rotation in (a) is split into 4 subsectors, each corresponding to a **second rotation**. The colors correspond to (relative) size of the scattering coefficients in that layer, from smallest (**white**) to largest (**black**).

Separating Channels

- HSI data is divided into **visual** spectrum (VIS) and **near-infrared** spectrum (NIR).
    - Label the channels in VIS from 0 to 30 and NIR from 31 to 61 (where 30 and 31 are the same wavelength: 700 nm).
    - **Even channels** refers to even indices, and similarly for **odd channels**
- Due to computational constraints, **2 separate pairs** of matching/ inverse networks are trained.
    - One pair of networks matches to, and inverts scattering coefficients of, even channels, while the other matches to and inverts odd channels.

Figure 8: Diagram of how HSI channels are split into even and odd channels.



Separating Channels

- Splitting into VIS and NIR channels leads to a **loss of correlations** between channels in the upper range of VIS and lower range of NIR (i.e. around the 650-750 nm range).
  - Reconstructing after splitting this way hence results in a misallignment in the predicted VIS and NIR images.
- Splitting into Even and Odd channels allows correlations throughout the full spectrum to be used in the Reconstruction.

Matching Network

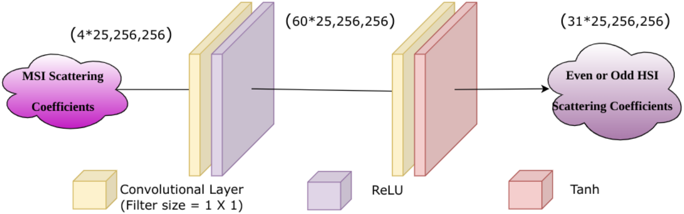


Figure 9: Diagram of **matching network**. The numbers correspond to the size of the data before and after each block of operations is applied (here, $M/2^J = N/2^J = 256$ is the size of the downsampled images, and $4 * 25$ and $31 * 25$ come from combining the scattering channels with the MSI and HSI channels, respectively).

Inverse Network

- **Inverse network** is similar to one proposed in [10].
  - Analogous to generator of **generative adversarial network** (GAN), with the scattering transform as the discriminator.

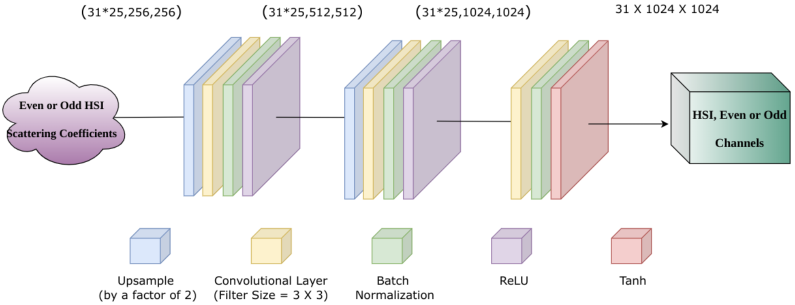


Figure 10: Diagram of **inverse network**, which is similar to a GAN. Numbers correspond to size of data before/ after each block of operations is applied.

[10] Angles and Mallat. “Generative networks as inverse problems with scattering transforms”. *Proc. ICLR*. 2018.

Multi-image Superresolution (MISR) Network

- Due to separation, some correlations are lost between adjoining HSI channels.
    - Apply a simple **multi-image superresolution** (MISR) network to improve alignment of predicted odd and even channels.
- First, (most) non-skin features in predicted HSI images are masked.
- Then, (linear) MISR network is applied to each **skin spectrum**.

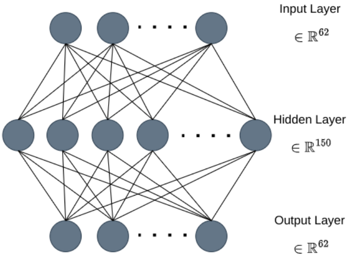


Figure 11: Diagram of MISR network, a **feedfoward neural network** (ReLU is used for first layer, tanh for second). The size of the data at each layer is indicated on the right.

Model Overview

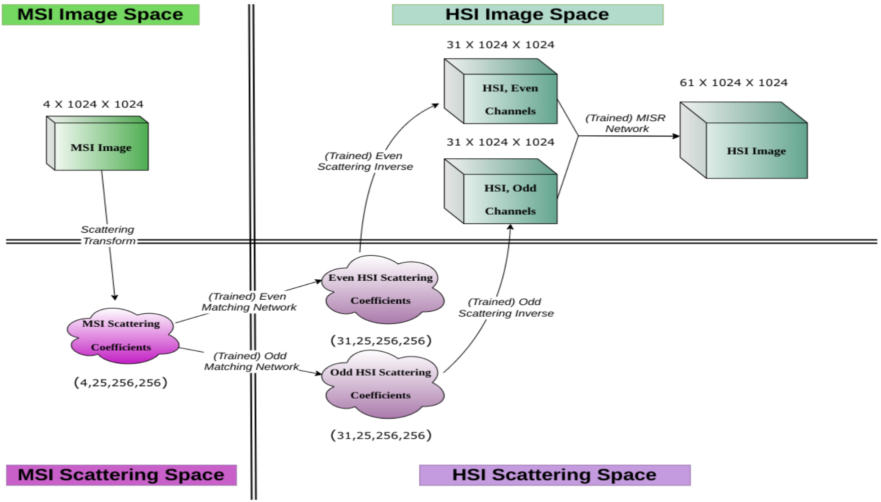


Figure 12: Diagram of full trained model, including **scattering transform** and **matching**, **inverse**, and **MISR** networks. Numbers represent sizes of data before/ after each part of model.

Table of Contents

Implementation Details

- Scattering transform is implemented by package **Kymatio** [11].

Table 1: Implementation Details of Neural Networks in Models

Network	Training Loss	Number of Epochs
Matching	MSE	100
Inverse	L^1	150
MISR	MSE	30 or 60

- All network architectures implemented using **PyTorch**.
  - Trained using **Adam optimizer** with learning rate 0.001.
- Outputs (even and odd channels) of models have two channels in common (700 nm).
  - For final output, **average** these two channels.

[11] Andreux et al. "Kymatio: Scattering Transforms in Python". *Journal of Machine Learning Research*. 2020.

Models Tested and Results

Table 2: Average SAM scores of skin reconstructions of the MST++ baseline and our proposed models.

Model	Average SAM score
MST++ (baseline)[3]	0.1182 ± 0.0200
Model 1 (no MISR)	0.1201 ± 0.0116
Model 2 (MISR, 30 Epochs)	0.1183 ± 0.0124
Model 3 (MISR, 60 Epochs)	$\mathbf{0.1179} \pm \mathbf{0.0129}$

- Models compared using SAM scores of reconstructed skin values.
  - Lower SAM score corresponds to better performance.
- **Source codes** for models available at: `https://github.com/BrandonKolstoe/Hyperskin_Scattering`

[3] Cai et al. "MST++: Multi-stage spectral-wise transformer for efficient spectral reconstruction". *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.

Model in Action: MSI

Figure 13: An example of a MSI image from the dataset. This image was not in the training set.

Model in Action: Visual Spectrum

Figure 14: Some reconstructed images from the visual spectrum (400-700 nm).

Model in Action: Near-Infrared Spectrum

Figure 15: Some reconstructed images from the near-infrared spectrum (700-1000 nm).

Current Work

Several improvements in our models:

1. **Dynamically scale** number of filters per layer in inverse network.
   - In a convolutional layer, train half as many filters as in previous layer.
   - Can train a single inverse/matching network pair which reconstructs all 61 channels.
   - Improves average SAM score on training set by $\approx \mathbf{0.024}$ (or $\approx \mathbf{13.1\%}$) with half the standard deviation compared to previous models.
2. **Denoising**: replace all sufficiently small scattering coefficients with 0.
   - Replace matched HSI scattering coefficients of magnitude at most 0.005 in 1st and 2nd layers.
   - Combined with (1), improves average SAM score on training set by $\approx \mathbf{0.032}$ (or $\approx \mathbf{17.3\%}$) with half the standard deviation compared to previous models.

Thank You!

- Name: **Brandon Kolstoe**
- Email: **bkolstoe@umd.edu**
- Paper: `https://arxiv.org/abs/2404.10030`
- GitHub: `https://github.com/BrandonKolstoe/Hyperskin_Scattering`