

HYPERSPECTRAL RECONSTRUCTION OF SKIN THROUGH FUSION OF SCATTERING TRANSFORM FEATURES

Wojciech Czaja, Jeremiah Emidih, Brandon Kolstoe

Richard G. Spencer

University of Maryland
Department of Mathematics
College Park, Maryland, United States

NIH, National Institute on Aging
Laboratory of Clinical Investigation
Baltimore, Maryland, United States

ABSTRACT

Hyperspectral imagery (HSI) is an established technique with an array of applications, but its use is limited due to both practical and technical issues associated with spectral devices. The goal of the ICASSP 2024 ‘Hyper-Skin’ Challenge is to extract skin HSI from matching RGB images and an infrared band. To address this problem we propose a model using features of the scattering transform - a type of convolutional neural network with predefined filters. Our model matches and inverts those features, rather than the pixel values, reducing the complexity of matching while grouping similar features together, resulting in an improved learning process.

Index Terms— HSI, scattering transform, data fusion, skin, machine learning

I. INTRODUCTION

Hyperspectral imagery (HSI), consisting of a set of 2D images representing different wavelengths, finds numerous applications in a wide range of fields [1]. In spite of its value, in many applications there are practical limitations for implementing this technology. The goal of the ICASSP 2024 Grand Challenge on Hyperspectral Skin Vision [2] is therefore to reconstruct HSI data from multispectral data (henceforth referred to as MSI) consisting of standard RGB images and one near-infrared band, with applications to skin analysis [3]. We propose a model based on matching and inverting scattering transform [4] features to address this challenge.

II. METHODS

2.1. Overview.

Our model takes advantage of representing and matching data in the feature space [5], instead of its original representation. Features are extracted by the scattering transform [4], which groups information about size, location, and direction

This work was funded in part by the Intramural Research Program of the National Institute on Aging of the NIH.

into scattering coefficients. Then, a transformation is learned by matching MSI scattering coefficients to those of the training HSI data. A preimage is then applied to obtain an HSI representation. Finally, a multi-image superresolution (MISR) network more closely aligns the skin values of the preimage.

2.2. Scattering Transform.

The scattering transform uses a convolutional neural network (CNN) structure but with predefined wavelet filters [4]. Let ψ be the mother wavelet, fix $L \in \mathbb{N}$, and for $j, q \in \mathbb{Z}$ denote $\psi_{j,q}(u) = 2^{-2j}\psi(2^{-l}r_\theta u)$, where $\theta = q\pi/L$ and r_θ is the corresponding rotation. Also, let $\phi_J(u) = 2^{-2J}\phi(2^{-J}u)$ where ϕ is a low-pass filter. Then, the 2-layer scattering transform of an image x is given by:

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

where $*$ denotes (periodic) convolution.

2.3. Matching Networks.

Due to computational constraints, two different matching networks are used. One network matches the scattering coefficients of MSI images to the scattering coefficients of even channels of HSI images, while the other network matches them to the coefficients of the odd channels. This was done to achieve compatible interpolations of both of these sets of HSI channels, rather than reconstructing results over essentially non-overlapping spectral bands. Each matching network consists of 2 linear layers. ReLU is used for the activation function of the first layer and tanh for the second.

2.4. Inverse Networks.

As above, one inverse network is trained for scattering coefficients of even channels and another for odd channels. We use an inverse similar to the one proposed in [6]. Two convolutional blocks, each consisting of upsampling by a factor of 2, a convolutional layer, batch normalization, and ReLU activation, are first applied. This is followed by one more convolutional layer, batch normalization, and tanh activation.

2.5. Multi-image Superresolution Network.

Some correlations are lost between adjoining HSI channels due to their separation above. This final network is applied to improve alignment of the predicted even and odd channel images. First, a mask is applied to the predicted HSI images which removes (most) non-skin features. Then, the network is applied to the skin spectra (i.e. the vectors of pixels from each HSI channel which have the same coordinates). The general network architecture is identical to the matching network.

III. IMPLEMENTATION AND RESULTS

3.1. Dataset.

We use the data in [3], as detailed in [2]. Note that the HSI data is divided into the visual range (VIS), from 400 to 700 nm in 10 nm increments, and the near-infrared range (NIR), from 700 to 1000 nm in 10 nm increments, with 700 nm in common. We label the channels in VIS from 0 to 31 and NIR from 32 to 61, so 31 and 32 are the same channel. Even channels refer to the even labels, and similarly for odd channels.

3.2. Implementation Details.

To compute the scattering transform, we use the package Kymatio [7]. Here, ψ is a Morlet wavelet and ϕ is a Gaussian. Additionally, the output is downsampled by a factor of 2^J . For our purposes, we choose $J = 2$ and $L = 4$, and we take the scattering transform of each channel of our images.

All network architectures were implemented using Pytorch, using the Adam optimizer with learning rate 0.001. The matching networks were trained for 100 epochs each using L^2 loss. The inverse networks were trained for 150 epochs each using L^1 loss. The MISR network was trained for either 30 or 60 epochs with L^2 loss.

Three models were tested. Model 1 doesn't use the MISR network, whereas Models 2 and 3 use the network trained for 30 and 60 epochs, respectively (in order to compare between different training times). Since the outputs of the models have one channel in common (700 nm), the final output for that channel is obtained by averaging these¹.

3.3. Results.

Table 1 summarizes the results of our three models and the baseline Mask-guided Spectral-Wise Transformer (MST++) method [8]. Models were compared using the average of the spectral angle mapper (SAM) scores [3] of the reconstructed skin values of the testing set; a lower SAM score corresponds to better performance. Our models perform comparably to the baseline model, with the third model performing best.

IV. CONCLUSION

Separate inverse networks (and MISR) were used due to limited computational resources. Our model's performance

¹Source codes for our models are available at https://github.com/BrandonKolstoe/Hyperskin_Scattering.

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

Model	Average SAM score
MST++ (baseline)	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)	0.1179 ± 0.0129

could likely be improved by applying PCA with dimensionality reduction to the HSI scattering coefficients [6], allowing us to train a single inverse network. This could also allow for deeper scattering transforms (i.e. with more details). We will also explore further improvement through use of scattering transforms better suited to HSI data [9].

V. REFERENCES

- [1] M. J. Khan et al., "Modern trends in hyperspectral image analysis: A review," *IEEE Access*, vol. 6, pp. 14118 – 14129, 2018.
- [2] P. C. Ng et al., "Hyperspectral skin vision challenge: Can your camera see beyond your skin?," in *ICASSP 2024 SPGC IEEE International Conference on Acoustics, Speech and Signal Processing Grand Challenge*, 2024.
- [3] P. C. Ng et al., "Hyper-skin: A hyperspectral dataset for reconstructing facial skin-spectra from RGB images," in *Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2023.
- [4] S. Mallat, "Group invariant scattering," *Communications in Pure and Applied Mathematics*, vol. 65, no. 10, pp. 1331–1398, 2012.
- [5] A. Cloninger, W. Czaja, and T. Doster, "The pre-image problem for Laplacian eigenmaps utilizing l_1 regularization with applications to data fusion," *Inverse Problems*, vol. 33, no. 7, pp. 074006, 2017.
- [6] T. Angles and S. Mallat, "Generative networks as inverse problems with scattering transforms," in *Proc. ICLR*, 2018.
- [7] M. Andreux et al., "Kymatio: Scattering transforms in Python," *Journal of Machine Learning Research*, vol. 21, no. 60, pp. 1–6, 2020.
- [8] Y. Cai et al., "MST++: Multi-stage spectral-wise transformer for efficient spectral reconstruction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 745–755.
- [9] I. Kavalero, W. Li, W. Czaja, and R. Chellappa, "3-d Fourier scattering transform and classification of hyperspectral images," *Transactions on Geoscience and Remote Sensing*, vol. 59, no. 12, pp. 10312–10327, 2021.