

# MatSpectNet: Material Segmentation Network with Domain-Aware and Physically-Constrained Hyperspectral Reconstruction (Supplementary Material)

Anonymous submission

## Project Code

Please refer to the README file in folder 'matspectnet' for more detail. The code will be published on GitHub once the paper is accepted.

## A. Additional Background: Laboratory Material Property Measurement

Apart from mobile devices, laboratory devices can measure spectral information under a constrained environment. For example, the spectrophotometer (Albert, Todt, and Davis 2012) quantitatively measures the absorptance and reflectance distribution against visible and infrared radiation wavelengths based on the amount of light absorbed by the material (Lv et al. 2022; van Nijnatten 2014). The streak camera (Bagayev et al. 2020; Horn 2009) measures time-dependent temporal point spread functions (TPSF) (Kirkby and Delpy 1996), which describe how light is reflected, refracted, scattered, or absorbed by the material. Additionally, femtosecond lasers (Lureau et al. 2020) can measure the thermal conductivity and mechanical properties of the material by emitting pulses of light and performing time-resolved measurements of temperature and laser-induced deformations (Guo et al. 2019). A spectrophotometer and a spectral camera are two devices that measure spectral information, but differ in their measurement capabilities. A spectrophotometer measures the amount of light absorbed or transmitted by a sample at a single point or small area, making it a precise tool for material characterisation and analysis in laboratory settings. In contrast, a spectral camera captures spectral information for an entire scene, making it useful for remote sensing and imaging applications. In the main paper, the samples in the spectraldb (Jakubiec 2022) are measured by a spectrophotometer, and samples in the ARAD\_1K (Arad et al. 2022) are measured by a spectral camera. We leverage both datasets by matching the measurements described in the Section Multi-Modal Fusion in the main paper.

## B. Network Training

The training of MatSpectNet consists of three phases: 1. the pre-training of the spectral recovery network  $S(x)$ . 2. the training of the physically-constrained recovery network

$S(x)$ ,  $R(h)$ . 3. the training of the material segmentation decoder. We will illustrate all three components in this section.

### B.1 Domain Discrimination Loss.

In order to align the features learned from two different datasets, the PatchGAN domain discriminator  $f_d$  (Chen et al. 2022) is used. The discriminators  $f_{d-spectral}$ ,  $f_{d-rgb}$  ensure that the hyperspectral images ( $h_s$ ,  $\hat{h}_m$ ) and the RGB images ( $R_s(h)$ ,  $x_m$ ) are domain-indistinguishable. When training the networks  $S(x)$ ,  $R(h)$ , the discriminators should not be updated and the loss term  $L_{domain}$  should reflect how successfully the network aligns the predictions:

$$L_{domain} = L_{MSE}(f_{d-spectral}(\hat{h}_m), 1) + L_{MSE}(f_{d-rgb}(R_s(h)), 1) \quad (1)$$

When training the discriminators, the loss  $L_{domain}$  should be configured to identify the sample domain:

$$L_{domain} = L_{MSE}(f_{d-spectral}(h_s), 1) + L_{MSE}(f_{d-spectral}(\hat{h}_m), 0) + L_{MSE}(f_{d-rgb}(x_m), 1) + L_{MSE}(f_{d-rgb}(R_s(h)), 0) \quad (2)$$

We use the label '1' to indicate the measured sample and '0' to indicate the reconstructed sample. MSE is used as suggested by GAN models such as CycleGAN (Zhu et al. 2017) and LSGANs (Mao et al. 2017), since MSE is not prone to vanishing gradients compared with cross-entropy.

### B.2. Data Preparation

We follow the dataset split method mentioned in (Heng et al. 2022b) to prepare both the LMD (Schwartz and Nishino 2020) and the OpenSurfaces (Bell et al. 2013). The models are monitored to choose the best parameters with the validation set and the reported performances are evaluated on the test set.

### B.3. Pre-training of the Spectral Recovery Network

We selected the MST++ (Cai et al. 2022) as the spectral recovery network  $S(x)$  to process the ARAD\_1K RGB images, which are normalised to a range of [0,1] using min-max normalisation. Specifically, for each channel in the

RGB image, we subtracted the minimum value and divided the result by the range. The corresponding hyperspectral image remained unchanged. Next, we randomly cropped the samples into  $128 \times 128$  patches and augmented the dataset with random vertical and horizontal flips. To optimise the network, we choose the mean relative absolute error ( $L_{MRAE}$ ) (Arad et al. 2022) and use the Adam optimiser with parameters  $\alpha = 4e^{-4}$ ,  $\beta_1 = 0.9$ , and  $\beta_2 = 0.999$  where  $\alpha$  is the initial learning rate. We employ the cosine annealing learning rate scheduler with a minimum learning rate of  $1e^{-6}$  and trained the network for 400 epochs with a batch size of 4 per GPU. The best model was selected based on its ability to produce the minimum  $L_{MRAE}$  calculated with the validation set.

#### B.4. Training of the Physically-Constrained Spectral Recovery Network

This section outlines the methodology for training the proposed physically-constrained spectral recovery network. The pre-trained  $S(x)$  is loaded and tuned in this section. We adopt the AdamW optimiser with an initial learning rate of  $6e^{-5}$ , and set the values of  $\beta_1$  and  $\beta_2$  to 0.5 and 0.999, respectively. Additionally, we employ the linear learning rate scheduler to adjust the learning rate during training. This step is trained for 400 epochs and the parameters that produce the minimum  $L_{trans}$  elaborated in Equation 1 from the main paper are kept for material segmentation.

#### B.5. Training of the Material Segmentation Decoder

In the main paper, Figure 1 depicts the recovered hyperspectral image  $\hat{h}$ , which is processed with a multi-layer perceptron (MLP) for material segmentation. In this section, we provide a detailed figure of the overall architecture by illustrating a material segmentation head, as illustrated in Figure 1. The MLP is configured to generate 128-channel features with hidden units of 64 channels. Recent studies have proposed that for material segmentation tasks, utilizing a combination of both material and contextual features, such as features related to objects or scenes, can reduce segmentation uncertainty and lead to improved performance (Heng et al. 2022b,a). As recovered hyperspectral images can provide reliable material features, we have incorporated an additional encoder-decoder segmentation network, namely the DBAT (Heng et al. 2022a), to extract contextual features from RGB images. The features extracted from both the MLP and the DBAT contain 128 channels each. These features are combined through concatenation and passed into a convolutional layer equipped with  $3 \times 3$  kernels, resulting in the final material label predictions.

During the training of the segmentation head, the parameters of the spectral recovery network  $S(x)$  are fixed. To augment the dataset, RGB images are randomly cropped into patches with dimensions of  $512 \times 512$  and randomly flipped. The AdamW optimizer is utilised with an initial learning rate of  $8e^{-5}$ ,  $\beta_1$  set to 0.9, and  $\beta_2$  set to 0.999. The cyclical learning rate scheduler (Smith 2017) is applied to gradually decrease the learning rate to  $7e^{-7}$  over 400 epochs. The

parameters that achieve the highest Pixel Acc are chosen to report the performance for each run.

### C. Physically-Constrained Spectral Recovery Experiments

This section analyses the configuration of the proposed RGB transformation network,  $R(h)$ , by validating the components of the simplified camera model and visualising the decay curve of  $L_{trans}$  and the recovered sRGB images to justify the model design and prepare for the material segmentation experiments.

#### C.1. [0,1] Normalisation and Brightness Factor

This section presents the decay curve of  $L_{trans}$ , which is introduced in Equation 1 in the main paper. Two training configurations are compared, by normalising the RGB images generated by  $R(h)$  into the range [0,1] or using a trainable brightness factor. The objective of  $L_{trans}$  is to assess the quality of the recovered RGB images  $\hat{x}_m$  obtained through the material data flow depicted in Figure 5 in the main paper. The normalisation operation acts as an additional constraint in regularising the network  $R(h)$ . It emphasises that the precise values of individual pixels are not crucial. Instead, the focus is on maintaining the relative values and the range difference of pixel intensities. Since the uniform brightness assumption works by scaling the pixels, it can be redundant when the [0,1] normalisation is used, which also scales the pixels by the range of the pixel values.

The decay curve of  $L_{trans}$  provides valuable insights into the impact of the normalisation operation on the training of  $R(h)$ . As depicted in Figure 2, the inclusion of the normalisation step in the training process has a significant impact on the convergence behaviour of  $R(h)$ . When the RGB images are normalised, the training of  $R(h)$  reaches a convergence level of 0.02301 at epoch 71. In contrast, using brightness factor, it takes approximately 216 epochs to achieve the same level of convergence. By utilising the normalisation constraint, the training of  $R(h)$  becomes more efficient. This indicates that scaling absolute pixel values is more difficult than learning the range difference and relative values.

#### C.2. Noise Level Tuning

This section tunes the noise level described in Section Camera System Noise and Brightness in the main paper. According to (Can Karaimer et al. 2019), the noise reduction process happens before the in-camera processing, by means of a high-pass filter. In the simplified camera model shown in Figure 2 in the main paper, the noise level is tuned to omit the noise reduction process, by minimising the  $L_{trans}$  in Equation 1 in the main paper without the network-modelled in-camera processing component. The evaluation with different noise levels are shown in Table 1. In the first column where no thermal noise is modelled, it is evident that the presence of Poisson shot noise significantly degrades the quality of the recovered RGB images, even when only a small amount of Poisson noise is applied (corresponding to high noise level  $\nu$ ). Therefore, for the subsequent experiments, the shot noise is excluded, and the thermal noise

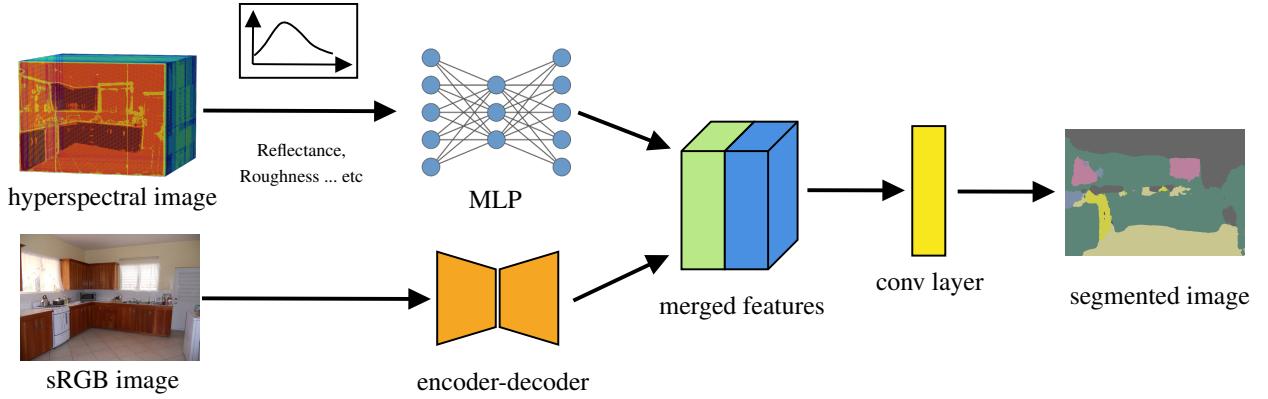


Figure 1: Feature merging for material segmentation.

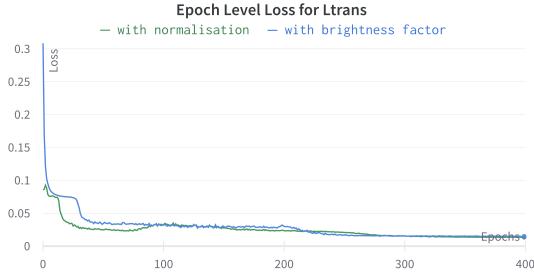


Figure 2: The loss decay curve of  $L_{trans}$  with or without normalisation.

noise level	no $\sigma$ noise	$\sigma \in [0, 0.001]$	$\sigma \in [0, 0.002]$	$\sigma \in [0, 0.005]$
no $\nu$ noise	0.006174	0.005990	0.005947	0.005989
$\nu \in [700, 900]$	0.008225	0.008305	0.008407	0.008775
$\nu \in [3000, 3500]$	0.007437	0.007499	0.007609	0.008028
$\nu \in [5000, 5500]$	0.007338	0.007398	0.007508	0.007928

Table 1: The converged  $L_{trans}$  with different noise level configurations.

level, denoted as  $\sigma$ , is constrained within the range of  $[0, 0.002]$ . This range represents the only configuration that surpasses the noise-free scenario, ensuring the integration of noise effects while omitting the noise reduction process.

### C.3. In-camera Processing Network Justification

The loss term of  $L_{trans}$  is shown in Figure 3. It converges to 0.005550 at epoch 388, which is 6.68% lower than the one without the in-camera processing network. The visual comparison in Figure 4 illustrates the difference between the recovered sRGB images, denoted as  $\hat{x}_m$ , and their corresponding ground truth images, denoted as  $x_m$ . Though the loss  $L_{trans}$  converges well, and the images look almost the same, the heatmap suggests that further improvements can be made especially for glass regions, probably caused by the uniform brightness assumption. There are two aspects that can be invested to improve the image quality.

Firstly, the field of spectral recovery remains an active area of research. Even with the use of a supervised dataset,

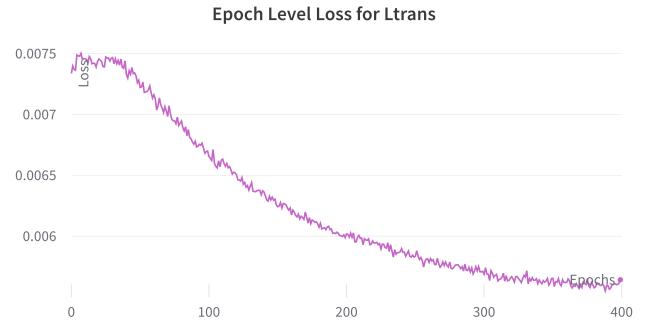


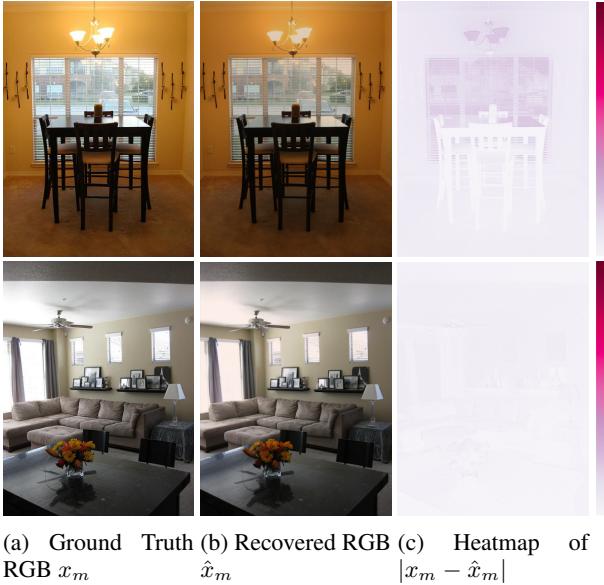
Figure 3: The loss decay curve of  $L_{trans}$  with tuned noise levels and in-camera processing.

the current performance of the spectral recovery network in generating hyperspectral images is not optimal. Through the implementation of an improved spectral recovery network, it is possible to achieve superior results by recovering more accurate and detailed information in the hyperspectral domain.

Secondly, the camera model incorporated in the network  $R(h)$  may suffer from the uniform brightness assumption, which does not hold in situations where multiple light sources exist. To address this issue, it is recommended to conduct experiments using more complex brightness estimation methods, such as (Afifi and Brown 2019). Moreover, it is possible to build a dataset-independent camera model with raw-RGB images instead of sRGB images. This adjustment allows for a simple but realistic camera model that better captures the transformation between hyperspectral and RGB images. Moreover, by eliminating the need for an additional network component to simulate the camera model, it is possible to streamline the image recovery process and potentially achieve improved results.

### C.4. Spectral Recovery Performance

Since there are no captured hyperspectral images in material datasets, the spectral recovery network is quantitatively evaluated on ARAD\_1K. As shown in table 2, with the ma-



(a) Ground Truth (b) Recovered RGB (c) Heatmap of  
RGB  $x_m$   $\hat{x}_m$   $|x_m - \hat{x}_m|$

Figure 4: The visualisation of two pairs of sRGB and recovered sRGB. The heatmap is measured by averaging the difference between normalised (range [0,1])  $x_m$  and  $\hat{x}_m$  across R, G, B channels. Pink means the difference is 0, and red means the difference is 1.

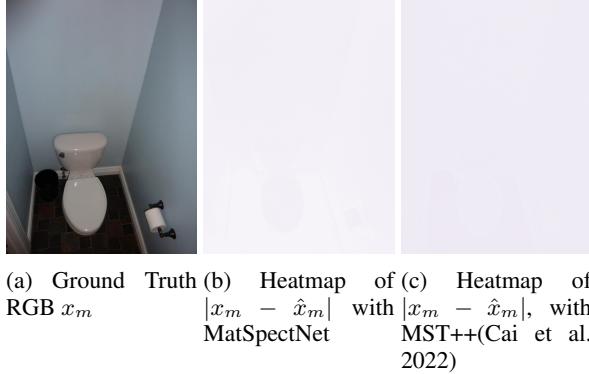
model	MRAE	RMSE
MST++	0.1645	0.0248
MatSpectNet	0.1774	0.0316

Table 2: The performance of the spectral recovery network  $S(x)$ , evaluated with metrics in (Arad et al. 2022).

terial dataset, the performance of the spectral recovery network  $S(x)$  decays. This is expected since the training objective is changed from recovering hyperspectral images in ARAD\_1K to material datasets. Luckily, the decay curve of  $L_{trans}$  in Figure 3 indicates that the quality of recovered hyperspectral images in material datasets is improved. Moreover, in Figure 5, with MatSpectNet, the differences between  $x_m$  and  $\hat{x}_m$  are smaller, supported by the whiter heatmap colour, compared with directly using the pre-trained MST++ (Cai et al. 2022).

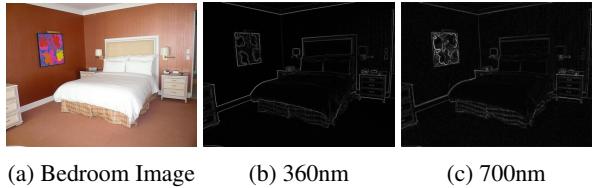
### C.5. Gradient Magnitude of Recovered Hyperspectral Images

The incorporation of hyperspectral images offers a unique advantage in material segmentation tasks as they capture the portion of light reflected by materials in the scene. As reflectance is a fundamental characteristic of materials, hyperspectral images provide vital discriminative features in addition to conventional RGB images. To validate the efficacy of hyperspectral images in material segmentation, one approach is to examine whether pixels representing the same material exhibit similar spectral profiles within a scene. The gradient magnitude plots of the hyperspectral images



(a) Ground Truth (b) Heatmap of (c) Heatmap of  
RGB  $x_m$   $|x_m - \hat{x}_m|$  with  $|x_m - \hat{x}_m|$ , with  
MatSpectNet MST++(Cai et al.  
2022)

Figure 5: Heatmaps of recovered RGB images with MatSpectNet or pre-trained MST++.



(a) Bedroom Image (b) 360nm (c) 700nm

Figure 6: The gradient magnitude of the recovered hyperspectral images.

at wavelengths 360nm and 700nm, as depicted in Figure 6, provide valuable insights in this regard.

For instance, in the bedroom image, at wavelength 360nm, the white quilt, pillow, and yellow headboard are all composed of fabric materials. The gradient magnitude plot shows no distinct boundaries, illustrating that hyperspectral images accurately capture reliable material properties. However, when analysing the painting on the wall at wavelength 700nm, clear boundaries between the colour spots emerge, potentially attributed to the diverse ingredients present in the painting. Similarly, the bed sheet reveals distinct regions, where the material of the strips may differ from the surrounding yellow fabric area. These observations underscore the challenges in annotating material segments and the complexities involved in defining appropriate material categories.

The analysis of hyperspectral image gradients provides compelling evidence for the potential of leveraging hyperspectral images in material segmentation tasks. The capacity to capture distinct spectral profiles of materials aids in accurately delineating material boundaries, thus offering a valuable contribution to the advancement of material segmentation methodologies. Furthermore, it highlights the need for robust annotation techniques and the exploration of innovative strategies for defining meaningful material categories in this challenging domain.

Model	DBAT	ResNet-152	ResNeSt-101	EfficientNet-B5	Swin-t	CAM-SegNet-DBA
Asphalt	88.66 ± 0.72	88.66 ± 0.17	94.35 ± 0.27	82.17 ± 2.80	91.83 ± 1.09	89.87 ± 1.94
Ceramic	68.31 ± 1.31	65.29 ± 3.19	62.86 ± 0.67	73.34 ± 0.42	75.35 ± 0.42	75.01 ± 0.64
Concrete	66.90 ± 1.07	50.89 ± 1.67	60.53 ± 2.00	59.36 ± 2.98	57.42 ± 4.88	69.20 ± 2.81
Fabric	93.14 ± 0.16	85.53 ± 0.22	86.420 ± 0.92	85.33 ± 0.20	88.71 ± 0.50	90.79 ± 0.43
Foliage	95.35 ± 0.12	93.55 ± 0.33	91.25 ± 1.16	88.21 ± 0.32	95.57 ± 0.45	94.04 ± 0.79
Food	93.27 ± 0.22	90.27 ± 0.22	94.96 ± 0.34	95.84 ± 0.14	92.51 ± 0.83	95.19 ± 0.24
Glass	73.27 ± 0.67	72.58 ± 2.50	68.33 ± 0.34	77.83 ± 0.94	77.95 ± 0.99	84.88 ± 1.11
Metal	79.99 ± 0.51	75.35 ± 0.94	80.66 ± 0.34	76.67 ± 0.28	81.54 ± 1.36	81.83 ± 0.48
Paper	73.83 ± 0.67	64.52 ± 2.87	71.14 ± 1.99	77.21 ± 0.13	63.05 ± 1.90	66.48 ± 1.43
Plaster	71.43 ± 0.71	68.01 ± 0.53	78.76 ± 0.62	73.11 ± 0.64	78.12 ± 1.90	72.37 ± 1.03
Plastic	50.62 ± 1.45	34.87 ± 1.21	36.07 ± 3.42	39.59 ± 0.64	51.64 ± 1.31	52.07 ± 2.28
Rubber	82.61 ± 1.01	77.08 ± 3.61	79.57 ± 1.62	69.73 ± 0.29	83.48 ± 0.67	81.63 ± 1.79
Soil	84.25 ± 0.50	73.27 ± 1.63	73.15 ± 2.67	79.73 ± 0.55	76.89 ± 1.11	80.39 ± 1.73
Stone	86.94 ± 0.95	69.66 ± 1.42	52.12 ± 0.93	70.07 ± 0.76	73.05 ± 1.92	60.73 ± 2.76
Water	97.12 ± 0.10	95.49 ± 0.33	97.54 ± 0.28	95.30 ± 0.32	95.78 ± 0.70	94.95 ± 0.69
Wood	90.53 ± 0.37	76.05 ± 1.08	76.71 ± 1.23	86.69 ± 0.24	82.03 ± 1.11	87.63 ± 0.98
PixelAcc	86.85 ± 0.08	80.68 ± 0.11	82.45 ± 0.20	83.17 ± 0.06	84.71 ± 0.26	86.12 ± 0.15
MeanAcc	81.05 ± 0.28	73.87 ± 0.25	75.31 ± 0.29	76.91 ± 0.06	79.06 ± 0.46	79.85 ± 0.28

Table 3: Per-category performance analysis. The networks are trained five times to report the uncertainty.

## D. Training Analysis

### D.1. Per-category Analysis

Table 3 displays the per-category performance of our proposed MatSpectNet and other networks. Notably, our model outperforms other models in eight categories, particularly those covering a small portion of the annotated samples, such as paper and plastic, as indicated by the numbers next to the category names. This suggests that the recovered hyperspectral images provide informative material features that enhance performance, even when the number of annotations is limited.

### D.2. Filter Analysis

The main paper introduces filters that aggregate spectral information using a weighted sum. These filters are designed to behave similarly to RGB filters, where the filter values represent the importance of each wavelength.

The spectral attention module computes spectral filters using an attention mechanism applied to the channels of the recovered hyperspectral images. As depicted in Figure 6 in the main paper, the hyperspectral image  $h \in \mathbb{R}^{n\_bands \times H \times W}$  is reshaped and permuted into  $h' \in \mathbb{R}^{HW \times n\_bands}$  to enable the application of self-attention to the spectral channels. Specifically, the spectral attention module employs linear projection to generate the query, key, and value matrices  $Q, K, V \in \mathbb{R}^{HW \times n\_bands}$  from the  $n\_bands$  hyperspectral image  $h'$  (Cai et al. 2022). Then the self-attention output  $A \in \mathbb{R}^{HW \times n\_bands}$  is acquired with Equation 3:

$$A = V \text{softmax}(K^T Q) \quad (3)$$

where the  $\text{softmax}$  function is applied to the spectral channel. The output  $A$  is then scaled to the range [0,1] with the min-max normalisation to align its physical meaning with the RGB response curves, which represent the sensitivity to each wavelength. In practice, we use  $n$  spectral attention modules in parallel to obtain  $n$  filters and construct an  $n$ -channel material image that contains dominant information for the segmentation task. The filters are analysed in the supplementary material.

In order to gain a better understanding of how these filters work, Figure 7 plots two of the filter weights for each wavelength alongside the RGB response functions. The filters show wavelength ranges where the weights are larger

than others, indicating the wavelengths that the network depends on, such as the range between 420 and 560 nm for filter 2. Moreover, unlike the RGB response functions, the learned filters can have more than one important region, as demonstrated by filter 1. This finding suggests that aggregating both short and long-wavelength information in the same filter is beneficial for material segmentation. Further investigation into the optimal weighting of different wavelength ranges can be a potential avenue for future research. Lastly, it is worth noting that the learned filters in the network exhibit a tendency to assign weights close to 0 for frequencies that are near those with large weights, such as 490nm and 550nm. This behaviour suggests that the training process eliminates frequencies that do not contribute significantly to material segmentation. One possible explanation for this phenomenon is that the spectral profiles often exhibit flat regions, as depicted in Figure 8. The filters effectively encode the spectral curve by sampling a few points from these flat regions and primarily focusing on the regions where the intensity varies.

### D.3. Quantitative Analysis

This section presents another segmentation visualisation in Figure 9. The table shown in the image lacks distinguishable texture, making it difficult even for humans to determine its material. As a result, in all other five models, the segmented images appear messy near the table region. The DBAT model struggles to distinguish between plaster and wood, whereas models such as Swin-t (Liu et al. 2021) and EfficientNet (Tan and Le 2019) are affected by reflected light and incorrectly recognize parts of the table as ceramic or glass. This indicates that RGB image does not provide reliable visual features for material segmentation task. In contrast, our MatSpectNet model avoids such noisy recognition and confidently identifies the table as stone. This indicates that the recovered hyperspectral image is a robust description of materials, irrespective of the illumination.

## Additional Figures

This section presents additional figures mentioned in the main paper. Figure 10 is the visualised response function of the R channel, for both the standard response and the shifted response.

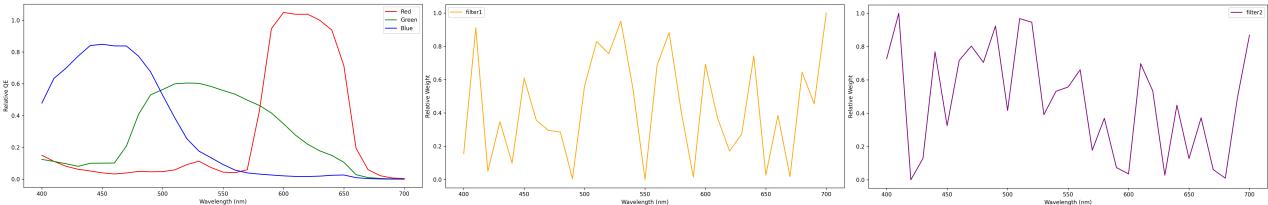


Figure 7: The filter weight of two filters for each wavelength.

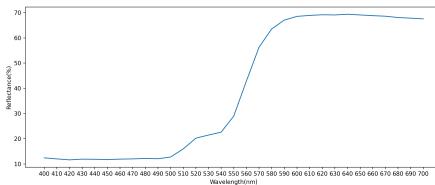


Figure 8: A sample of the recovered spectral profile for a pixel classified as plastic.

## References

- Afifi, M.; and Brown, M. S. 2019. Sensor-independent illumination estimation for DNN models. *arXiv preprint arXiv:1912.06888*.
- Albert, D. R.; Todt, M. A.; and Davis, H. F. 2012. A low-cost quantitative absorption spectrophotometer. *Journal of Chemical Education*, 89(11): 1432–1435.
- Arad, B.; Timofte, R.; Yahel, R.; Morag, N.; Bernat, A.; Cai, Y.; Lin, J.; Lin, Z.; Wang, H.; Zhang, Y.; et al. 2022. NTIRE 2022 Spectral Recovery Challenge and Data Set. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 863–881.
- Bagayev, S.; Averchenko, V.; Chekhonin, I.; Chekhonin, M.; Balmaev, I.; and Mekhov, I. 2020. Experimental new ultra-high-speed all-optical coherent streak-camera. In *Journal of Physics: Conference Series*, volume 1695, 012129. IOP Publishing.
- Bell, S.; Upchurch, P.; Snavely, N.; and Bala, K. 2013. OpenSurfaces: A Richly Annotated Catalog of Surface Appearance. *ACM Trans. on Graphics (SIGGRAPH)*, 32(4).
- Cai, Y.; Lin, J.; Lin, Z.; Wang, H.; Zhang, Y.; Pfister, H.; Timofte, R.; and Van Gool, L. 2022. Mst++: Multi-stage spectral-wise transformer for efficient spectral reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 745–755.
- Can Karaimer, H.; Khodadad, I.; Kazemzadeh, F.; and Brown, M. S. 2019. A Customized Camera Imaging Pipeline for Dermatological Imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 0–0.
- Chen, G.; Zhang, G.; Yang, Z.; and Liu, W. 2022. Multi-scale patch-GAN with edge detection for image inpainting. *Applied Intelligence*, 1–16.
- Guo, B.; Sun, J.; Lu, Y.; and Jiang, L. 2019. Ultrafast dynamics observation during femtosecond laser-material interaction. *International Journal of Extreme Manufacturing*, 1(3): 032004.
- Heng, Y.; Wu, Y.; Dasmahapatra, S.; and Kim, H. 2022a. Enhancing Material Features Using Dynamic Backward Attention on Cross-Resolution Patches. In *33rd British Machine Vision Conference 2022, BMVC 2022, London, UK, November 21-24, 2022*. BMVA Press.
- Heng, Y.; Wu, Y.; Kim, H.; and Dasmahapatra, S. 2022b. CAM-SegNet: A Context-Aware Dense Material Segmentation Network for Sparsely Labelled Datasets. In *17th International Conference on Computer Vision Theory and Applications (VISAPP)*, volume 5, 190–201.
- Horn, A. 2009. *Ultra-fast material metrology*. John Wiley & Sons.
- Jakubiec, J. A. 2022. Data-Driven Selection of Typical Opaque Material Reflectances for Lighting Simulation. *LEUKOS*, 1–14.
- Kirkby, D. R.; and Delpy, D. T. 1996. Measurement of tissue temporal point spread function (TPSF) by use of a gain-modulated avalanche photodiode detector. *Physics in Medicine & Biology*, 41(5): 939.
- Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; and Guo, B. 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 10012–10022.
- Lureau, F.; Matras, G.; Chalus, O.; Derycke, C.; Morbieu, T.; Radier, C.; Casagrande, O.; Laux, S.; Ricaud, S.; Rey, G.; et al. 2020. High-energy hybrid femtosecond laser system demonstrating  $2 \times 10$  PW capability. *High Power Laser Science and Engineering*, 8: e43.
- Lv, H.; Yang, Z.; Pan, H.; and Wu, R. 2022. Electromagnetic absorption materials: Current progress and new frontiers. *Progress in Materials Science*, 100946.
- Mao, X.; Li, Q.; Xie, H.; Lau, R. Y.; Wang, Z.; and Paul Smolley, S. 2017. Least squares generative adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, 2794–2802.
- Schwartz, G.; and Nishino, K. 2020. Recognizing Material Properties from Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(8): 1981–1995.
- Smith, L. N. 2017. Cyclical learning rates for training neural networks. In *2017 IEEE winter conference on applications of computer vision (WACV)*, 464–472. IEEE.

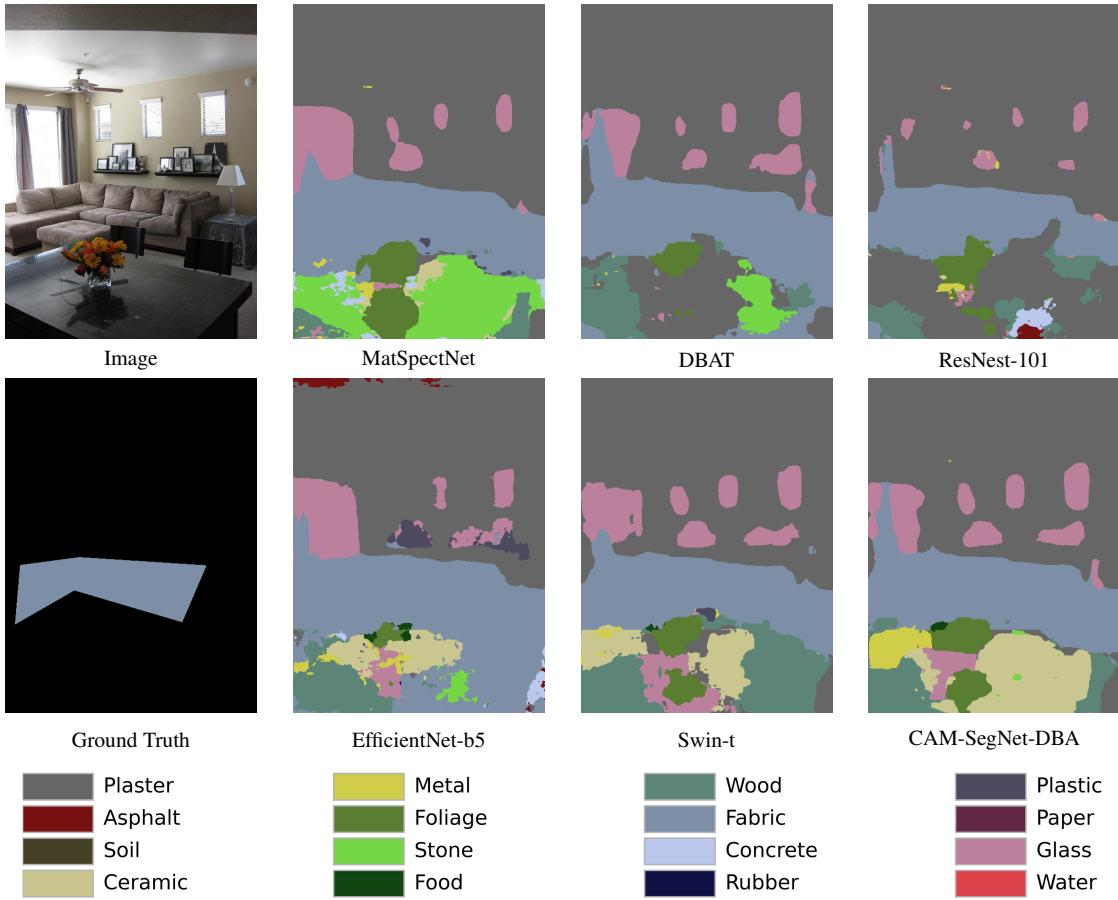


Figure 9: Predicted segmentation of one living room image.

Tan, M.; and Le, Q. 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, 6105–6114. PMLR.

van Nijnatten, P. A. 2014. Regular reflectance and transmittance. In *Experimental Methods in the Physical Sciences*, volume 46, 143–178. Elsevier.

Zhu, J.-Y.; Park, T.; Isola, P.; and Efros, A. A. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, 2223–2232.

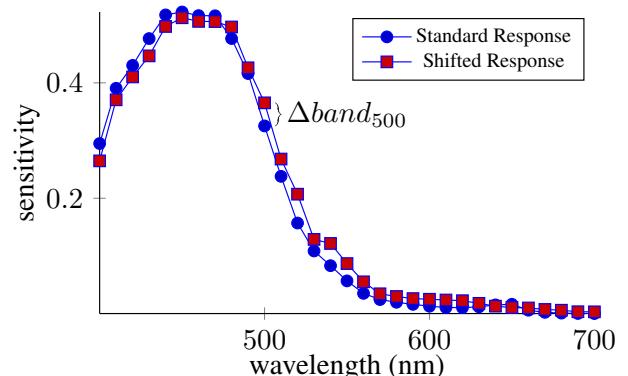


Figure 10: The illustration of the predicted band shift for the blue channel response curve.