Multispectral Imaging for Fine-Grained Recognition of Powders on Complex Backgrounds Supplementary Material

Tiancheng Zhi, Bernardo R. Pires, Martial Hebert and Srinivasa G. Narasimhan
Carnegie Mellon University
{tzhi,bpires,hebert,srinivas}@cs.cmu.edu

Figure 1. **Thick and Thin Powder Samples**.

1. Powder List

We collected 100 powders from 6 categories: Food (44 powders), Colorant (20), Skincare (10), Dust (10), Cleansing (9), and Other (7). See Table 1 for the list of powder names, categories, and legends for segmentation labels. See Figure 1 for snapshots of thick and thin powder samples.

2. Pseudocode for NNCV Band Selection

See Algorithm 1.

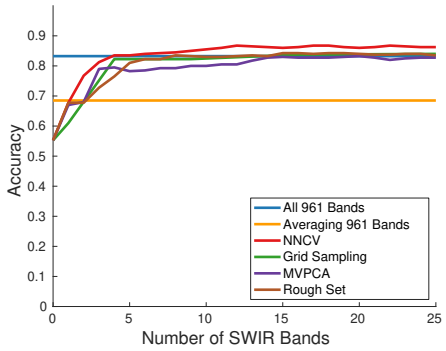
	Name	Category	Legend
1	Ajinomoto	Food	
2	Almond Flour	Food	
3	Aqua Glow	Colorant	
4	Aspirin	Other	
5	Baby Powder	Skincare	
6	Baking Soda	Food	
7	Barley Water	Food	
8	BB Powder	Skincare	
9	Beach Sand	Dust	
10	Blackboard Chalk	Dust	
11	Black Frit	Dust	
12	Black Iron Oxide	Colorant	
13	Black Pepper	Food	
14	Black Toner	Colorant	
15	Blue Pigment	Colorant	
16	Borax Detergent Booster	Cleansing	
17	Boric Acid	Cleansing	
18	Bronze Metallic	Colorant	
19	Brown Dye	Colorant	
20	Brown Sugar	Food	
21	Calcium Carbonate	Dust	
22	Cane Sugar	Food	
23	Caralluma	Food	
24	CC Powder	Skincare	
25	Celtic Sea Salt	Food	
26	Charcoal	Colorant	
27	Chaste Tree Berry	Food	
28	Chicken Bath	Dust	
29	Citric Acid	Food	
30	Cobalt Frit	Dust	
31	Cocoa	Food	
32	Coconut Flour	Food	
33	Coconut Oil	Food	
34	Coffe Mate	Food	
35	Corn Starch	Food	
36	Cream of Rice	Food	
37	Cream of Tartar	Food	
38	Cream of Wheat	Food	
39	Cyan Toner	Colorant	
40	Detox Powder	Dust	
41	Dragon Blood	Other	
42	Dry Milk	Food	
43	Espresso	Food	
44	Eye Shadow	Skincare	
45	Fake Moss	Other	
46	Flower Fuel	Other	
47	Fuchsia Dye	Colorant	
48	Fungicide	Cleansing	
49	Garlic	Food	
50	Ginger	Food	
51	Green Bean Water	Food	
52	Green Glow	Colorant	
53	Green Pigment	Colorant	
54	Guar Gum	Food	
55	Gym Chalk	Dust	
56	Hibiscus	Food	
57	Iodized Salt	Food	
58	Loose Powder	Skincare	
59	Lotus	Food	
60	Magenta Toner	Colorant	
61	Matcha	Food	
62	MCT Oil	Food	
63	Meringue	Food	
64	Milk Replacer	Food	
65	Moringa	Food	
66	Nail Dipping	Colorant	
67	Onion	Food	
68	Orange Glow	Colorant	
69	Orange Peel	Food	
70	Pearl Powder	Skincare	
71	Pet Moist	Cleansing	
72	Potassium Iodide	Other	
73	Potato Starch	Food	
74	Quick Blue Bleach	Cleansing	
75	Red Bean Water	Food	
76	Root Destroyer	Cleansing	
77	Sandalwood	Other	
78	Schorl Tourmaline	Dust	
79	Shaving Powder	Skincare	
80	Silver Metallic	Colorant	
81	Smelly Foot Powder	Cleansing	
82	Sodium Alginate	Food	
83	Spanish Paprika	Food	
84	Stain Remover	Cleansing	
85	Stevia	Food	
86	Stone Cement	Dust	
87	Sun Powder	Skincare	
88	Talcum	Skincare	
89	Teal Azul Dye	Colorant	
90	Tide Detergent	Cleansing	
91	Urea	Other	
92	Vanilla	Food	
93	Vitamin C	Food	
94	Wheat Grass	Food	
95	White Pepper	Food	
96	Yellow Dye	Colorant	
97	Yellow Glow	Colorant	
98	Yellow Pigment	Colorant	
99	Yellow Toner	Colorant	
100	Zinc Oxide	Skincare	

Table 1. **Powder List**. Powder names, categories, and legends for segmentation labels are listed.

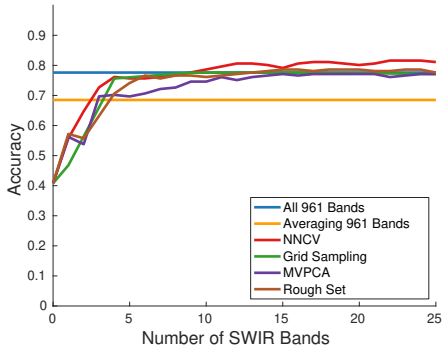
Algorithm 1 NNCV Band Selection

Input: Number of SWIR bands to be selected N_s
Output: Selectd SWIR bands B_s

$B_s \leftarrow \emptyset$
$B_n \leftarrow$ all SWIR bands
for $i = 1 : N_s$ **do**
  for each $b \in B_n$ **do**
    $score_b \leftarrow$ mean class accuracy of nearest neighbor cross validation using RGBN and $B_s \cup \{b\}$ bands
  end for
  $b \leftarrow \underset{b \in B_n}{\operatorname{argmax}}\, score_b$
  $B_s \leftarrow B_s \cup \{b\}$
  $B_n \leftarrow B_n - \{b\}$
end for



(a) Mean Sample Accuracy



(b) Mean Class Accuracy

Figure 2. **Band Selection Comparison on an Additional Patch Dataset**. Mean sample accuracy calculates the mean accuracy of all samples (200 powder samples and 200 common material samples), while mean class accuracy calculates the mean accuracy of 101 classes (100 powder classes and 1 background class). NNCV performs better than the others in most settings.

3. Band Selection Comparison on Additional Patch Dataset

We captured an additional patch dataset (200 powder patches and 200 common material patches) under light sources Set B. We classify a new patch by finding nearest neighbor in the patch dataset captured under Set A. The Split Cosine Distance between mean patch intensities is used. Figure 2 shows the accuracy *vs.* #SWIR bands curve of different selection methods. NNCV performs better than the others in most settings.

4. From Kubelka-Munk Model to Beer-Lambert Blending Model

The Beer-Lambert Blending model can be deduced from the Kubelka-Munk model [4] via approximation. The channel subscript c is ignored below.

Let R, R_∞, and R_g ($0 < R, R_\infty, R_g < 1$) be the absolute reflectance of thin powder, infinitely thick powder, and background, and S be the scattering coefficient. The Kubelka-Munk model is:

$$R = \frac{R_\infty^{-1}(R_g - R_\infty) - R_\infty(R_g - R_\infty^{-1})e^{Sx(R_\infty^{-1} - R_\infty)}}{(R_g - R_\infty) - (R_g - R_\infty^{-1})e^{Sx(R_\infty^{-1} - R_\infty)}} \tag{1}$$

Let $\kappa = S(R_\infty^{-1} - R_\infty)$, Equation 1 can be re-written as:

$$R = \frac{(1 - R_\infty^2)(R_g - R_\infty)}{(R_\infty R_g - R_\infty^2) - (R_\infty R_g - 1)e^{\kappa x}} + R_\infty \tag{2}$$

Since $0 < R_\infty, R_g < 1$, we assume R_∞^2 and $R_\infty R_g$ are small enough to be ignored. The approximate model is:

$$R = \frac{R_g - R_\infty}{e^{\kappa x}} + R_\infty = (1 - e^{-\kappa x})R_\infty + e^{-\kappa x}R_g \tag{3}$$

Under constant shading L, $I = LR$, $A = LR_\infty$, $B = LR_g$. Then we obtain the Beer-Lambert Blending Model:

$$I_c = (1 - e^{-\kappa_c x})A_c + e^{-\kappa_c x}B_c \tag{4}$$

Since $\kappa = S(R_\infty^{-1} - R_\infty)$, it means that if S does not change much across channels, the κ signature and SWIR signature (κ and R_∞) should show negative correlation.

5. Calibrating κ with Different Backgrounds

The Beer-Lambert Blending model assumes that the attenuation coefficient κ is independent of the background. This section checks if the calibrated κ is invariant to the background used for calibration.

We choose three different backgrounds (Black Aluminum Foil, Brown Leather Hide, Sand Paper) and image a thick sample, three thin samples, and three bare backgrounds in the same filed of view for each powder. We calibrate κ values using different backgrounds. We calculate the coefficient of variation c_v (std/mean) for each powder and each channel. Usually the data is considered low variance if $c_v < 1$. Figure 3 shows the histogram of mean c_v values for RGBN and SWIR channels separately. About

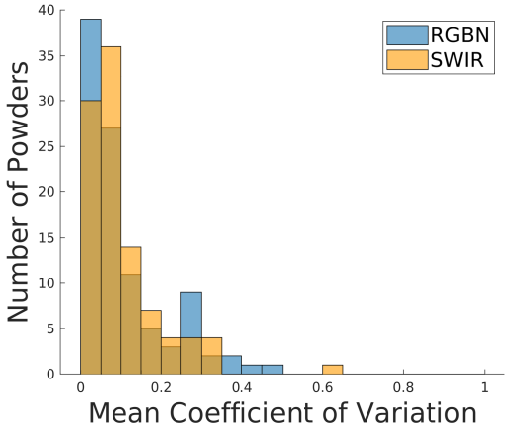


Figure 3. **Histogram of Coefficient of Variation c_v of κ Calibrated using Different Backgrounds**. Usually the data is considered low variance if $c_v < 1$. About 95% of the powders show $c_v < 0.3$, which means that κ calibrated with different backgrounds has a very low variance.

Blending	RMSE (mean±std)	
	RGBN	SWIR
Alpha	0.023±0.021	0.022±0.024
Beer-Lambert	**0.014±0.016**	**0.012±0.018**

Table 2. **Fitting error on Multi-background Dataset**. RMSE is calculated based on pixel values divided by white patch. Beer-Lambert Blending shows a smaller error than Alpha Blending.

95% of the powders show $c_v < 0.3$, which means that κ calibrated with different backgrounds has a very low variance.

We also calculate the fitting error on this multi-background dataset. As shown in Table 2, Beer-Lambert Blending has a smaller error than Alpha Blending.

6. Hyperparameters for Deep Learning

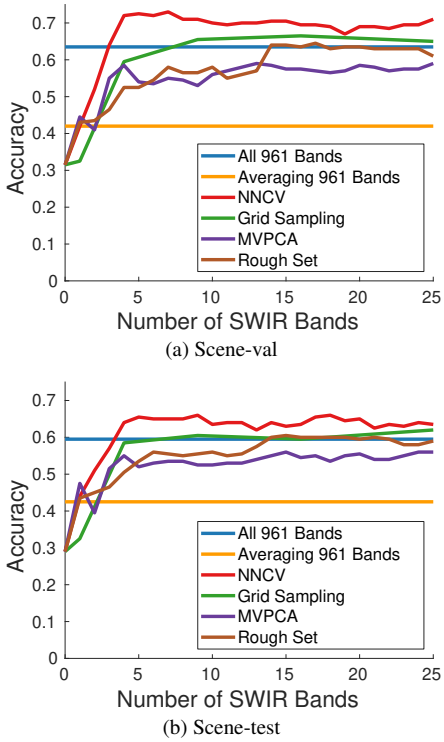
We use group normalization [6] instead of batch normalization [3] in DeepLab v3+ network [2]. CRF [1] postprocessing is used.

We use the AdamWR [5] optimizer and cosine annealing with warm restart scheduler. We set initial restarting period= 8, and $T_{mult} = 2$. Thus, the scheduler restarts at 8, 24, 56, 120, and 248 epochs. We use batch size = 8 and weight decay = 1e-4 for all experiments.

We first train the model from scratch on synthetic powders against synthetic backgrounds with initial learning rate = 1e-3 for 248 epochs. In each iteration, we find a random synthetic background for the current powder mask, and blend them to render a scene. We call it an epoch when it goes through all powder masks once.

Then we fine-tune it on synthetic powders against real backgrounds from *Scene-bg* and *Scene-sl-train* with initial learning rate = 1e-4 for 56 epochs.

We finally fine-tune it on real powders against real backgrounds from *Scene-sl-train* with initial learning rate = 5e-5 for at most 56 epochs. Model selection is done according to performance on validation set.



(a) Scene-val

(b) Scene-test

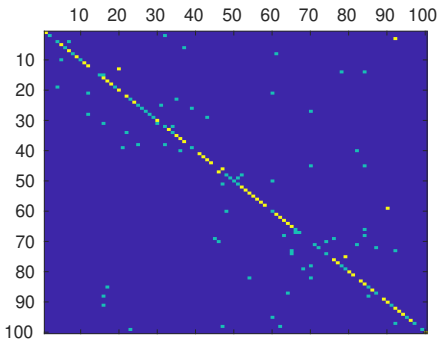
Figure 4. **Band Selection Comparison**. Grid Sampling is tested on square numbers only. The accuracy of NNCV saturates after four bands, outperforming the other methods.

7. More about Recognition with Known Powder Location

Band Selection: See Figure 4 for separate results on *Scene-val* and *Scene-test*.

Confusion Matrix: As shown in Figure 5, we visualized the confusion matrix of using inpainting background, Beer-Lambert Blending, and RGBN and 4 SWIR bands selected by NNCV. We show the confusion matrix for individual powders, and the one aggregated into 6 categories.

Top-N Accuracy: Top-N accuracy is meaningful for real applications. Table 3 show that top-3 predictions achieve over 80% accuracy and top-7 predictions achieve about 90% accuracy.



(a) 100 Powders

(b) Aggregated into 6 Categories

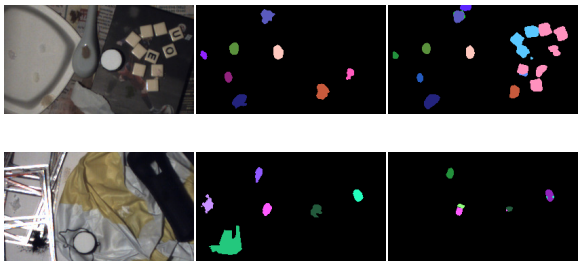
Figure 5. **Confusion Matrix**. The confusion matrix for individual powders, and the one aggregated into 6 categories are shown. The proposed method shows strong performance.

Top-N	Scene-val	Scene-test	Scene-sl-train	Scene-sl-test
1	72.0	64.0	62.50	62.50
3	81.0	86.0	82.00	83.50
5	87.0	88.5	87.25	87.75
7	90.0	92.5	89.50	90.75

Table 3. **Top-N Accuracy**. Inpainting background, Beer-Lambert Blending, RGBN channels and 4 SWIR bands selected by NNCV are used. Top-3 predictions achieve over 80% accuracy and top-7 predictions achieve about 90% accuracy.

8. More about Recognition with Unknown Powder Mask

PR curve: People usually care about whether there exists a specific powder in the image rather than the mask of the powder. We slightly modify the algorithm to answer this "Yes-or-No" question. We set a threshold of confidence, and check the confidence of being the specific powder for each pixel. We say "Yes" if there exists a pixel with confidence higher than the threshold. By adjusting the threshold , we plot the PR curve in Figure 8. The curve shows that our method significantly outperforms the baseline method.
More Qualitative Results: See Figure 7.



RGB Ground Truth Prediction

Figure 6. **Failure cases**. Row 1 misdetects small square objects; Row 2 misses powders on cluttered background.

Failure Cases: Figure 6 shows failure cases where the algorithm misdetects some small objects as powder, and misses some powders on cluttered background.

9. Spectral Specifications

White Patch: The white patch has a 98% reflectance with a approximately flat spectral curve in 400-1700nm. We divide the scene by the white patch intensity.
Light Source: The light sources are halogen and incandescent lamps made by 6 different manufacturers. Experiments show that our method works across different sources
Mirror/Beamsplitter: See Figure 9 for spectral specifications. The mirror has $> 70\%$ reflectance in 400-1700nm. Beamsplitter 1 transmits SWIR while reflects VIS and NIR. Beamsplitter 2 has transmittance:reflectance=50:50 in VIS-NIR range.
Theoretical SWIR Spectral Transmittance: Figure 10 shows the theoretical spectral transmittance of different voltage settings (step size=0.6V).

References

[1] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2018. 4

[2] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *European Conference on Computer Vision*. Springer, 2018. 4

[3] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning*, pages 448–456, 2015. 4

[4] Paul Kubelka and Franz Munk. An article on optics of paint layers. *Z. Tech. Phys*, 12(593-601), 1931. 3

[5] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. 4

[6] Yuxin Wu and Kaiming He. Group normalization. In *European Conference on Computer Vision*. Springer, 2018. 4

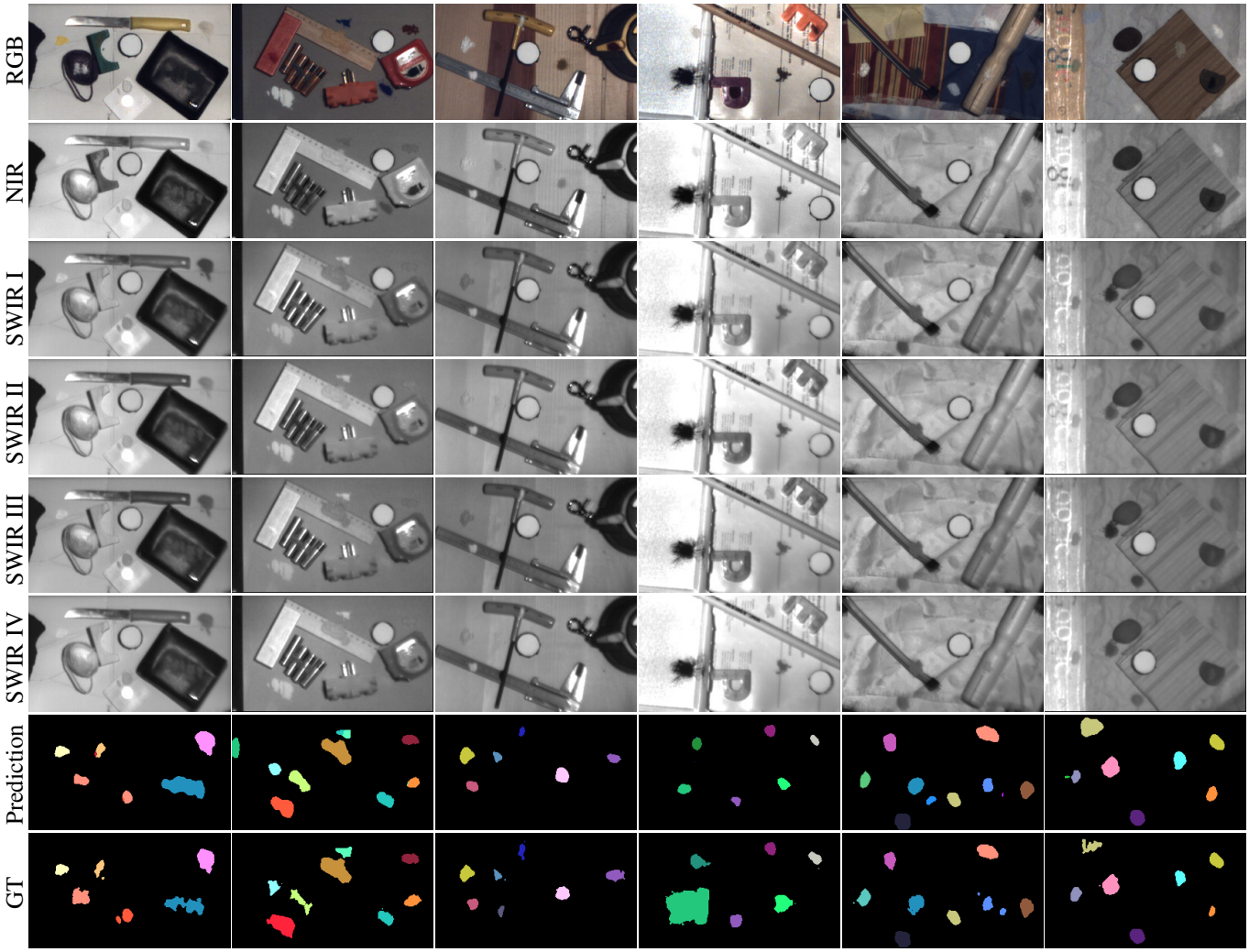


Figure 7. **More Qualitative Results**. Some images are from *Scene-sl-test*, which only has selected bands. But the two baseline methods (Per-pixel Nearest Neighbor and Standard Semantic Segmentation) require all bands. So the baseline results are not shown.

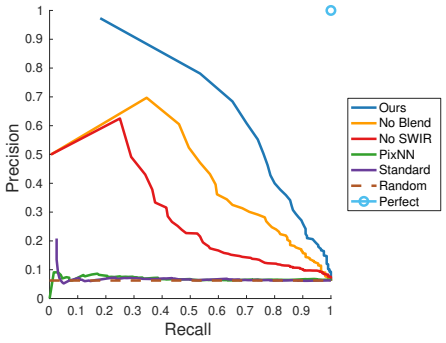
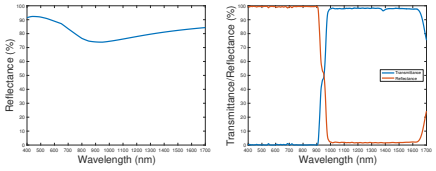
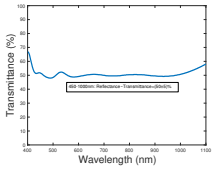


Figure 8. **PR curve on *Scene-test***. Incorporating band selection and data synthesis, our method outperforms Per-pixel Nearest Neighbor, and achieves huge improvement over simply training on limited real data with all bands.



(a) Mirror (b) Beamsplitter 1



(c) Beamsplitter 2

Figure 9. **Spectral Specifications** of Mirror/Beamsplitters.

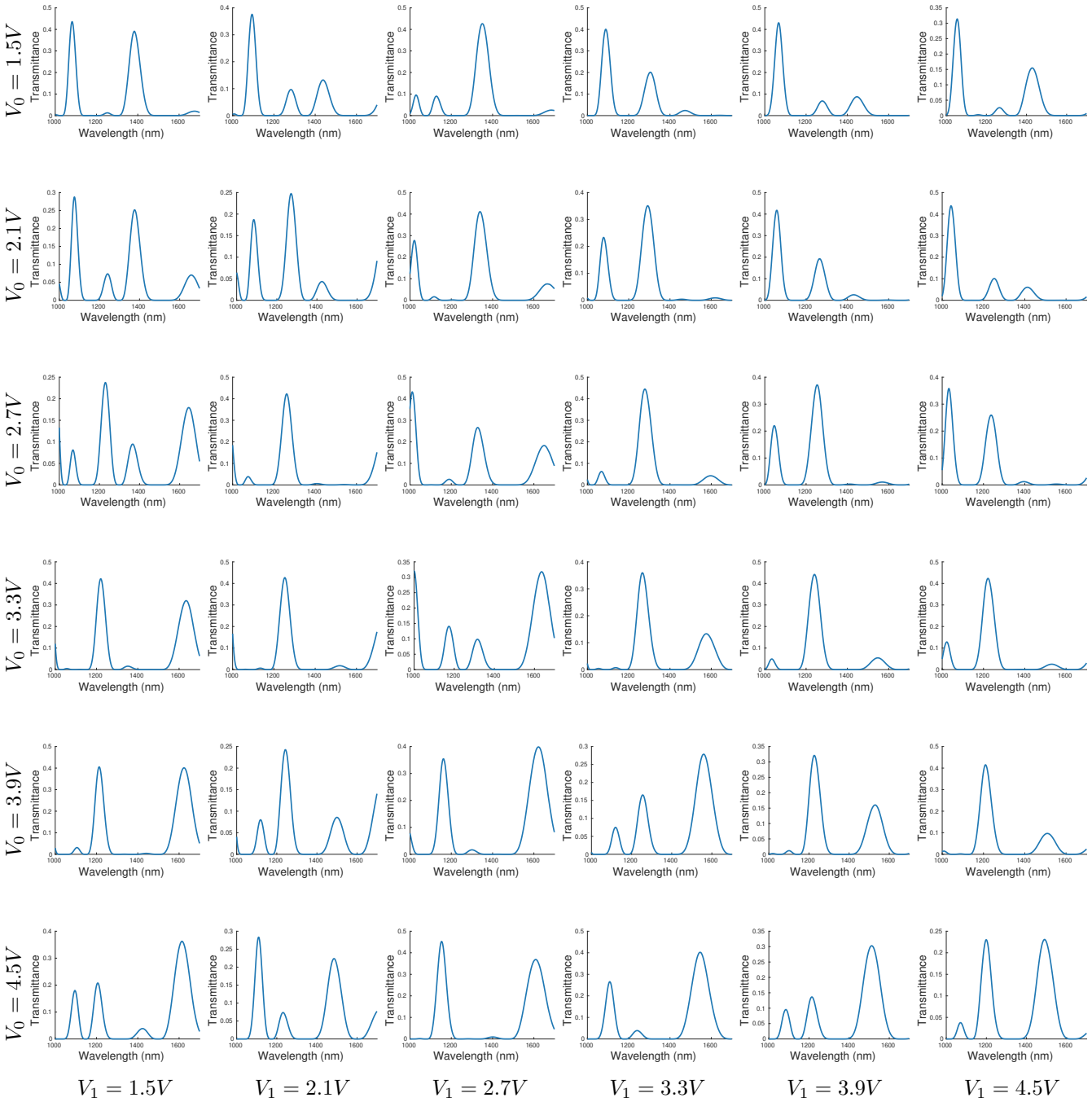


Figure 10. **Theoretical SWIR spectral transmittance** of different voltage settings.