Beyond RGB: A Real World Dataset for Multispectral Imaging in Mobile Devices Supplementary Material

Ortal Glatt[1] Yotam Ater[1] Woo-Shik Kim[2] Shira Werman[1] Oded Berby[1]
Yael Zini[1] Shay Zelinger[1] Sangyoon Lee[2] Heejin Choi[2] Evgeny Soloveichik[1]

Samsung Israel R&D Center[1], Israel {ortal.glatt,yotam.ater}@samsung.com
Samsung Advanced Institute of Technology[2], Korea wooshik.kim@samsung.com

S1. Method Implementation Details

The network was trained with the Adam optimizer [3] ($\beta_1 = 0.9$, $\beta_2 = 0.999$) with an initial learning rate of $3 \cdot 10^{-3}$. Training was performed for 30 epochs on random crops of size 256×256. We use 128×128 bins for the CCC histogram and set the loss weights for the final loss as $1, 30, 0.0001$ for w_1, w_2 and w_3 respectively. The channels triplets were chosen to be the following 21 triplets: (0, 1, 2), (1, 2, 3), (2, 3, 4), (3, 4, 5), (4, 5, 6), (5, 6, 7), (6, 7, 8), (7, 8, 9), (8, 9, 10), (9, 10, 11), (10, 11, 12), (11, 12, 13), (12, 13, 14), (13, 14, 15), (0, 5, 9), (1, 6, 10), (2, 7, 11), (3, 8, 12), (4, 9, 13), (5, 10, 14) and (3, 6, 15). The first 14 triplets are combinations of channels with large overlapping areas and the last 7 triplets are combinations of channels with sparse overlapping areas.

The network architecture is detailed in Table S1. The architecture of a single CCC pyramid is described in Table S2.

Layer	Configuration	Layer Output
Input	Create log-chroma histograms for M triplets	$M \times 128 \times 128$
M CCC pyramids	See Table S2	$M \times 128 \times 128$
Conv1	$c : 2M,\ k : 5 \times 5$	$2M \times 128 \times 128$
Batch normalization	$c : 2M$	$2M \times 128 \times 128$
Max pooling	$k : 4 \times 4$	$2M \times 32 \times 32$
Activation	ReLU	$2M \times 32 \times 32$
Conv2	$c : 4M,\ k : 5 \times 5$	$4M \times 32 \times 32$
Batch normalization	$c : 4M$	$4M \times 32 \times 32$
Max pooling	$k : 4 \times 4$	$4M \times 4 \times 4$
Activation	ReLU	$4M \times 4 \times 4$
Conv3	$c : 4M,\ k : 5 \times 5$	$4M \times 4 \times 4$
Batch normalization	$c : 4M$	$4M \times 4 \times 4$
Activation	ReLU	$4M \times 4 \times 4$
Flatten		$4M \cdot 4 \cdot 4$
Linear1	$o : 100$	100
Activation	ReLU	100
Linear2	$o : 50$	50
Activation	ReLU	50
Linear3	$o : 36$	36
Activation	Exponent	36

Table S1. Network architecture for the ISE problem. Here, c, k and o, stand for number of channels, kernel size and output size respectively.

Layer	Configuration	Layer Output
Input	Log-Chroma UV Histogram	128×128
Conv1	$c : 1$, $k : 5 \times 5$	128×128
Downsample	Bilinear interpolation, $sf : 0.5$	64×64
Conv2	$c : 1$, $k : 5 \times 5$	64×64
Downsample	Bilinear interpolation, $sf : 0.5$	32×32
Conv3	$c : 1$, $k : 5 \times 5$	32×32
Downsample	Bilinear interpolation, $sf : 0.5$	16×16
Conv4	$c : 1$, $k : 5 \times 5$	16×16
Downsample	Bilinear interpolation, $sf : 0.5$	8×8
Conv5	$c : 1$, $k : 5 \times 5$	8×8
Downsample	Bilinear interpolation, $sf : 0.5$	4×4
Conv6	$c : 1$, $k : 5 \times 5$	4×4
Downsample	Bilinear interpolation, $sf : 0.5$	2×2
Conv7	$c : 1$, $k : 5 \times 5$	2×2
Blur and upsample	Bilinear interpolation, $sf : 2$	4×4
Sum	Previous layer output and output of conv6	4×4
Blur and upsample	Bilinear interpolation, $sf : 2$	8×8
Sum	Previous layer output and output of conv5	8×8
Blur and upsample	Bilinear interpolation, $sf : 2$	16×16
Sum	Previous layer output and output of conv4	16×16
Blur and upsample	Bilinear interpolation, $sf : 2$	32×32
Sum	Previous layer output and output of conv3	32×32
Blur and upsample	Bilinear interpolation, $sf : 2$	64×64
Sum	Previous layer output and output of conv2	64×64
Blur and upsample	Bilinear interpolation, $sf : 2$	128×128
Sum	Previous layer output and output of conv1	128×128

Table S2. Architecture of a single CCC pyramid. Here, c, k and sf, stand for number of channels, kernel size and scale factor respectively.

The kernel for the blur operation in the pyramid was set to:

$$w = \begin{bmatrix} 0.0625 & 0.1250 & 0.0625 \\ 0.1250 & 0.2500 & 0.1250 \\ 0.0625 & 0.1250 & 0.0625 \end{bmatrix}. \tag{S1}$$

S2. Further Experiments & Results

S2.1. Expanded Statistics on Beyond RGB

We present expanded statistics of the results reported in the main manuscript. As is common in the color constancy literature, we report the 25th, 50th and 75th percentiles of each result in addition to standard deviation and mean. We report these statistics in Table S3 and Table S4 for $\Delta\mathrm{A}_{\mathrm{HS}}$, $\Delta\mathrm{A}_{\mathrm{MS}}$ and $\Delta\mathrm{A}_{\mathrm{XYZ}}$ on the Beyond RGB dataset for all the methods tested. Additionaly, we report results of PWIR [4] which we adapt to the MS+RGB modality by concatenating the MS and RGB data at the input of the network.

Dataset	Method	$\Delta\mathrm{A}_{\mathrm{HS}}\downarrow$					$\Delta\mathrm{A}_{\mathrm{MS}}\downarrow$				
		mean	25%	median	75%	std	mean	25%	median	75%	std
Lab	GrayWorld						6.17	3.30	4.42	8.07	4.09
	GrayEdge						8.97	5.05	7.49	12.56	5.02
	LRMF						21.59	19.94	21.21	22.83	2.28
	PWIR	27.10	18.55	23.91	35.68	10.86	11.03	5.37	8.80	16.90	6.81
	Ours	5.92	4.04	5.39	8.01	2.92	2.05	1.26	1.75	2.41	1.39
Field	GrayWorld						6.85	5.73	6.60	8.05	2.40
	GrayEdge						11.39	8.49	11.31	14.08	3.93
	LRMF						21.90	20.72	21.74	23.11	1.20
	PWIR	16.31	11.61	15.17	21.82	6.07	6.07	5.01	3.18	8.82	3.35
	Ours	7.22	3.31	6.14	9.89	5.54	2.73	1.06	1.91	3.47	2.46

Table S3. Expanded statistics of $\Delta\mathrm{A}_{\mathrm{HS}}$ and $\Delta\mathrm{A}_{\mathrm{MS}}$ comparison between the proposed method and other ISE methods on the Beyond RGB dataset. Green and yellow highlights respectively indicate best and second best results.

Dataset	Algorithm	Input Modality	$\Delta\mathrm{A}_{\mathrm{HS}}\downarrow$					$\Delta\mathrm{A}_{\mathrm{XYZ}}\downarrow$				
			mean	25%	median	75%	std	mean	25%	median	75%	std
Lab	Ours	RGB-O						3.82	1.23	2.62	4.12	3.96
		RGB-S						3.66	1.26	2.40	4.63	3.56
		MS	5.92	4.04	5.39	8.01	2.92	2.59	1.21	2.21	3.29	1.82
		Fusion-O	4.70	2.86	4.17	6.18	2.45	1.88	0.79	1.55	2.52	1.46
		Fusion-S	4.99	2.62	4.39	6.66	3.05	2.04	0.96	1.52	2.64	1.70
	PWIR	MS	27.10	18.55	23.91	35.68	10.86	13.35	3.05	12.91	24.25	10.60
		Fusion-O	26.46	17.45	24.34	34.90	10.49	12.82	2.73	11.05	23.16	10.15
		Fusion-S	29.76	22.12	26.63	37.46	9.20	14.04	3.76	12.57	25.25	10.86
Field	Ours	RGB-O						4.58	1.75	3.54	5.80	3.74
		RGB-S						5.41	2.10	4.26	6.35	5.12
		MS	7.22	3.31	6.14	9.89	5.54	3.52	1.33	2.27	4.72	3.74
		Fusion-O	7.64	4.58	6.40	9.06	5.42	4.13	1.39	3.14	5.38	3.88
		Fusion-S	6.90	4.26	5.31	7.53	5.85	3.35	1.18	2.44	3.47	4.22
	PWIR	MS	16.31	11.61	15.17	21.82	6.07	8.44	3.38	6.65	13.60	6.25
		Fusion-O	17.13	13.21	18.19	20.93	5.18	8.32	3.02	7.42	14.35	6.03
		Fusion-S	16.20	9.23	15.13	23.39	7.62	9.28	3.94	6.96	15.07	6.36

Table S4. Expanded statistics of $\Delta\mathrm{A}_{\mathrm{HS}}$ and $\Delta\mathrm{A}_{\mathrm{XYZ}}$ comparison of different versions of the proposed method on the Beyond RGB dataset. RGB-O and RGB-S respectively indicate RGB version using Oppo and Samsung devices data. Fusion-O and Fusion-S respectively indicate fusion version of MS data together with Oppo and Samsung devices RGB data. Green and yellow highlights respectively indicate best and second best results.

S2.2. Cross Dataset Validation

In this section, we showcase the outcomes of training our suggested approach using the KAUST dataset and subsequently testing it on Beyond RGB, and vice versa. In both scenarios, we utilized the full dataset for training. Specifically, when leveraging the KAUST dataset for training, we executed relighting based on the illuminants found in the Beyond RGB dataset and adapted the spectral responses to emulate the filters of our MS camera. As depicted in Table S5, there is a significant disparity in performance when trying to extrapolate from one dataset to the other. We note that such a performance degradation was not observed when training our method on KAUST and testing on the CAVE SR dataset as shown in Table 2 of the main manuscript. This leads us to hypothesize that there remains a domain gap between relit SR datasets and our own directly acquired MS dataset, highlighting the need for benchmarking algorithms on both modalities. We additionally emphasize that this evaluation protocol allowed the algorithm to test on illuminants which were present in the test split which does not happen in our normal evaluation protocol. This strengthens the assumption of the dataset domain gap, as the network had an opportunity to overfit on the illuminants but still fails and does worse than the inter-domain case where there exists a split between test and train illuminants.

Trained On	Tested On	$\Delta A_{HS} \downarrow$				
		mean	25%	median	75%	std
KAUST with Beyond RGB Illuminants	Beyond RGB Lab	12.54	8.47	11.56	16.00	5.79
	Beyond RGB Field	10.55	4.47	8.89	14.96	7.60
KAUST with Beyond RGB Illuminants	KAUST with Lab Illuminants	7.18	4.15	7.18	8.07	4.98
	KAUST with Field Illuminants	6.60	2.75	6.34	8.69	4.14
Beyond RGB	KAUST with Lab Illuminants	12.82	6.76	11.17	16.79	8.10
	KAUST with Field Illuminants	10.43	5.02	8.27	13.74	7.67
Beyond RGB	Beyond RGB Lab	5.92	4.04	5.39	8.01	2.92
	Beyond RGB Field	7.22	3.31	6.14	9.89	5.54

Table S5. Evaluation of the generalization of our method across Beyond RGB and KAUST datasets. ”Trained On” indicates the modality on which the model was trained and ”Tested On” indicates the modality on which inference was performed. ”Beyond RGB Illuminants” indicates illuminations measured across all of the Beyond RGB dataset, ”Lab” and ”Field” indicates only Beyond RGB lab and field data respectively.

S2.3. Cross Camera Evaluation

A common concern in developing color-constancy or spectral estimation algorithms is the sensitivity to the particular model of camera on which the algorithm is trained. We utilized the availability of multiple cameras in Beyon dRGB to assess the effect of cross-camera application of our method. For this assessment, we trained each model on 80% of the Beyond RGB dataset utilizing inputs from one camera while performing testing on 10% of the data of the other camera. In the fusion modality, the MS sensor used for training and testing is common and the RGB data comes from different cameras.

We observe a degradation in the quality of the results compared to the results of Table S4, indicating that our method is sensitive to variation of the camera.

Dataset	Trained on	Tested on	$\Delta A_{HS} \downarrow$					$\Delta A_{XYZ} \downarrow$				
			mean	25%	median	75%	std	mean	25%	median	75%	std
Lab	RGB-O	RGB-S						4.95	1.88	3.59	6.77	4.23
	RGB-S	RGB-0						4.58	1.51	3.14	6.56	4.13
	Fusion-O	Fusion-S	5.50	3.35	4.94	6.96	2.67	2.45	1.14	2.08	3.48	1.77
	Fusion-S	Fusion-O	5.92	4.04	5.39	8.01	2.92	2.59	1.21	2.21	3.29	1.82
Field	RGB-O	RGB-S						7.36	3.34	5.91	8.49	5.76
	RGB-S	RGB-0						6.09	1.89	5.24	8.32	5.08
	Fusion-O	Fusion-S	7.94	5.10	7.11	9.18	5.75	4.22	1.52	3.00	5.84	4.11
	Fusion-S	Fusion-O	7.78	5.06	6.81	9.09	4.60	3.70	1.39	2.73	4.93	3.39

Table S6. Results of cross-camera inference of models trained on Oppo RGB (RGB-O), Samsung RGB (RGB-S), MS+Oppo RGB (Fusion-O) and MS+Samsung (Fusion-S) inputs. ”Trained On” indicates the modality on which the model was trained and ”Tested On” indicates the modality on which inference was performed. Green and yellow highlights respectively indicate best and second best results.

S2.4. CNN Loss Ablation

In Table S7 we show the effect of utilizing the CNN and CCC losses. The results conclusively demonstrate that $\mathcal{L}_{\text{CCC}}$ is a crucial component of the network, and moreover that propagating $\mathcal{L}_{\text{CNN}}$ adversely affects performance, underscoring the importance of the CCC feature extraction step of the network.

Dataset	Method		$\Delta A_{\text{HS}} \downarrow$				
	$\mathcal{L}_{\text{CNN}}$ propagation	$\mathcal{L}_{\text{CCC}}$	mean	25%	median	75%	std
Lab	✓	✓	27.21	22.13	25.72	32.91	8.20
	✓	✗	26.84	21.12	27.01	30.56	7.29
	✗	✓	5.92	4.04	5.39	8.01	2.92
Field	✓	✓	20.61	14.25	21.54	24.99	5.71
	✓	✗	20.10	14.90	20.68	25.71	6.64
	✗	✓	7.22	3.31	6.14	9.89	5.54

Table S7. Ablation studies. We compare the hyper spectral angular error of our method, our method with propagation of $\mathcal{L}_{\text{CNN}}$ to the CCC blocks and with propagation of $\mathcal{L}_{\text{CNN}}$ to the CCC block, without $\mathcal{L}_{\text{CCC}}$. Green and yellow highlights respectively indicate best and second best results.

S3. Additional Dataset Information

S3.1. Multispectral Filter Responses

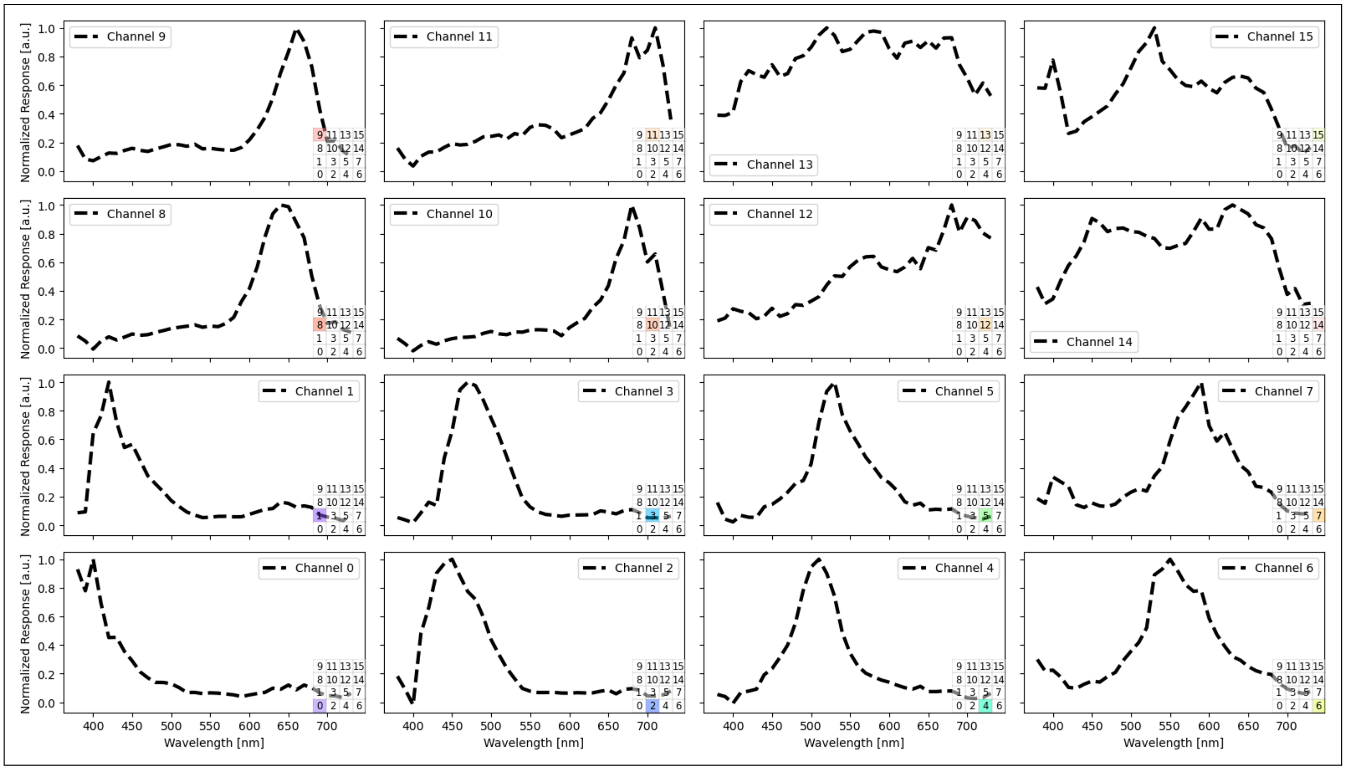


Figure S1. Normalized spectral responses of our MS sensor. Insets: the spatial arrangement of the 4x4 CFA pattern, with the relevant filter highlighted for each graph. The highlight color is the RGB equivalent of the channel response. Channel numbering is according to increasing peak wavelength of each channel.

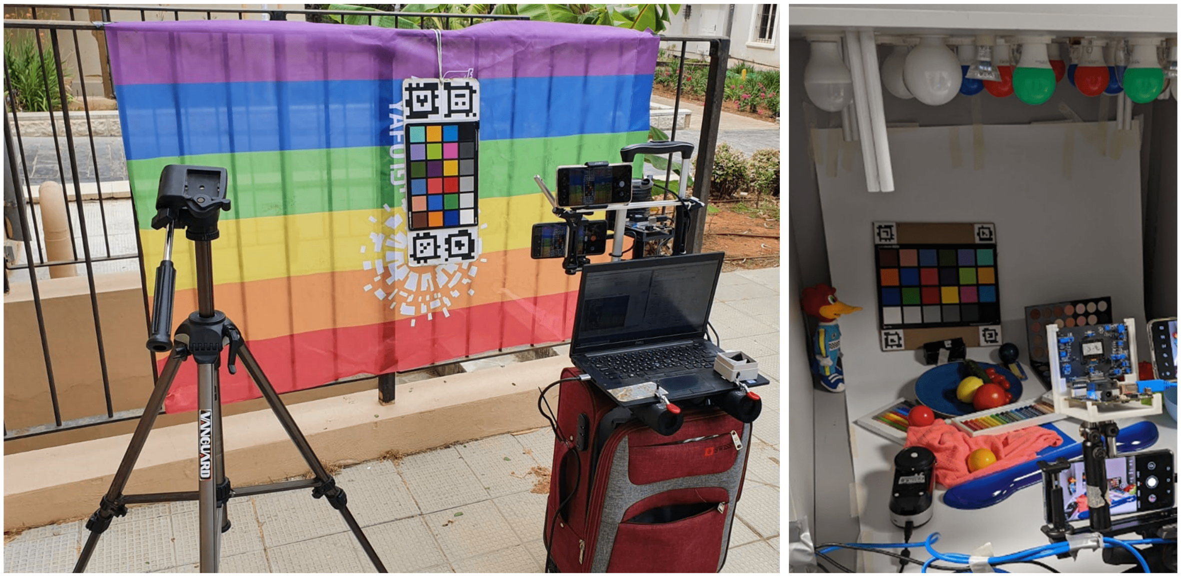
S3.2. Intrinsic and Extrinsic Calibration

Intrinsic Calibration: We employ OpenCV [1] to conduct camera intrinsic calibration using the conventional checkerboard technique. This procedure yields the camera's projection matrix, in addition to the radial and tangential distortion coefficients.

Extrinsic Calibration: Extrinsic calibration can be used to rectify image pairs and as a pre-processing step in many registration algorithms. The relative pose between the cameras is not consistent throughout the scenes in the Beyond RGB dataset. However, the ColorChecker calibration target outfitted with AprilTags offers a planar reference with well-established dimensions in all scenes where the color checker is presented. This allows us to find the pose of each camera relative to the calibration board and then solve for the relative poses [2]. Scenes in which the calibration target is not present will often have the same extrinsics as the counterpart scene which does include the calibration board, but this is not guaranteed.

S4. Beyond RGB Dataset Samples

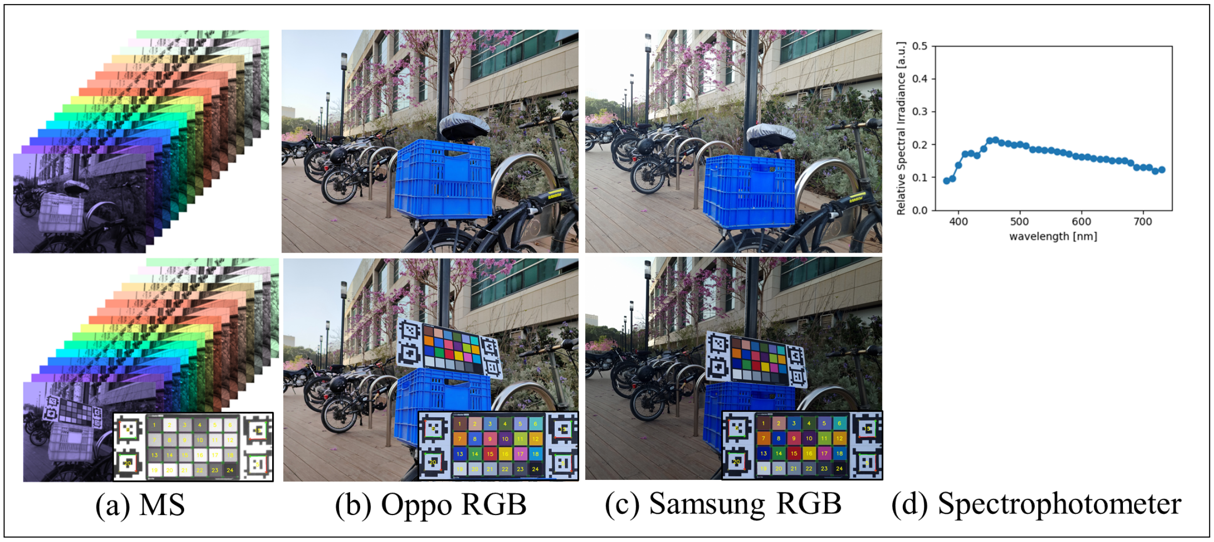
In the following section, we present a selection of samples from the Beyond RGB dataset. These examples have been chosen to represent the diversity of the dataset in terms of varying real world and laboratory conditions. Figure S2 shows examples of data collection in the field and in the lab. Figure S3 showcases the full data contents of a single scenario, including an MS image, two images captured by Android devices and a spectral irradiance measurement. Figure S4 highlights a sample of 100 images taken from the field portion of Beyond RGB. Figure S5 shows the 13 lab scenarios which were captured under varying illuminants.



(a) Field settings

(b) Lab settings

Figure S2. Beyond RGB dataset collection: (a) Capturing outdoor scene, with the color-chart presented using MS sensor and 2 Android devices. (b) Automated scene capturing in lab viewing booth, with the color-chart presented, using MS sensor and 2 Android devices, and an automated spectrophotometer measurement. Notice the various light sources present in the custom viewing booth.



(a) MS (b) Oppo RGB (c) Samsung RGB (d) Spectrophotometer

Figure S3. Beyond RGB dataset sample content: (a) MS (b) Oppo Find X5 Pro CPH230 and (c) Samsung Galaxy S21 Plus SM-G996B data with and without color-chart. Inset: detection and extraction of color chart coordinates. (d) Spectrophotometer illuminant spectra measurement.

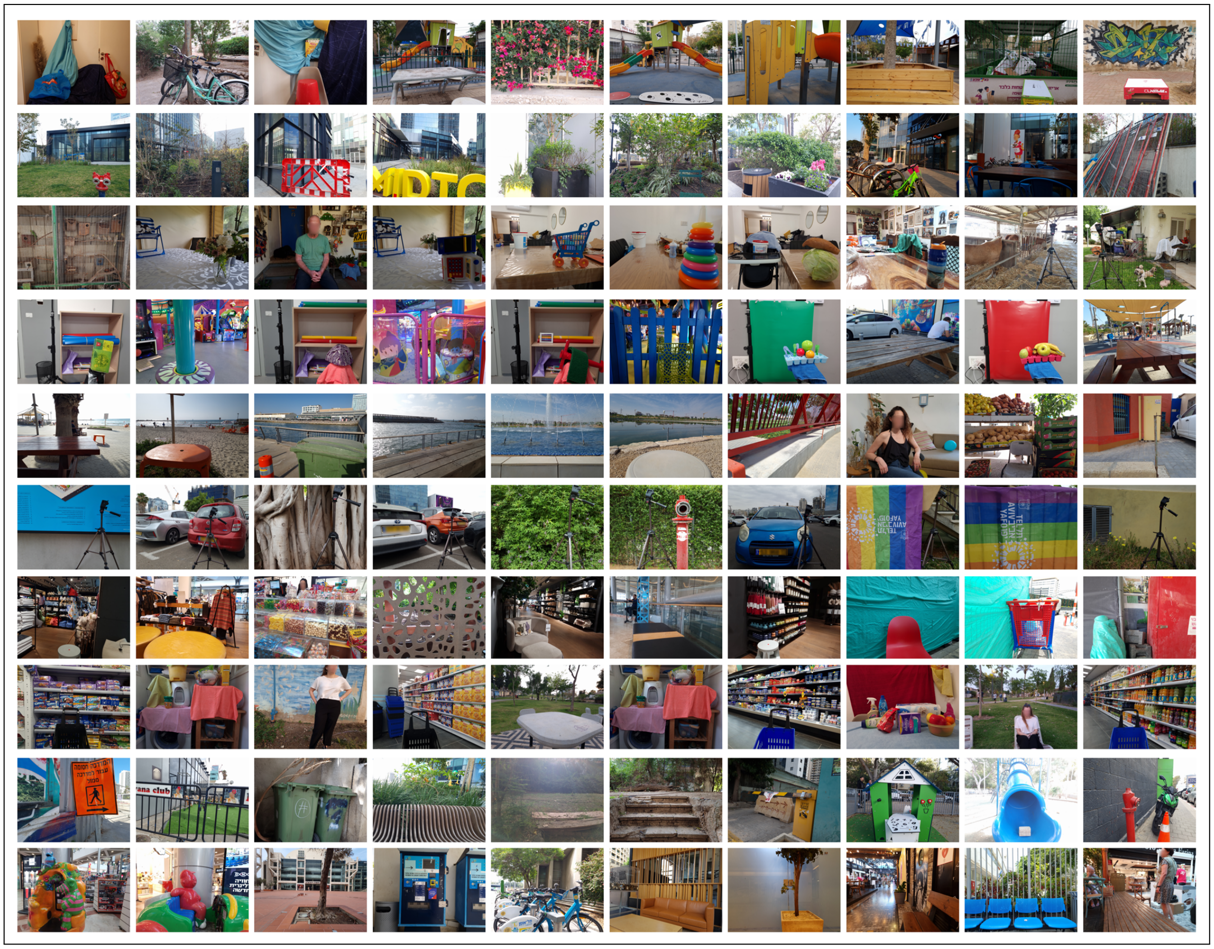


Figure S4. Beyond RGB Field scenarios examples without color-chart presented.

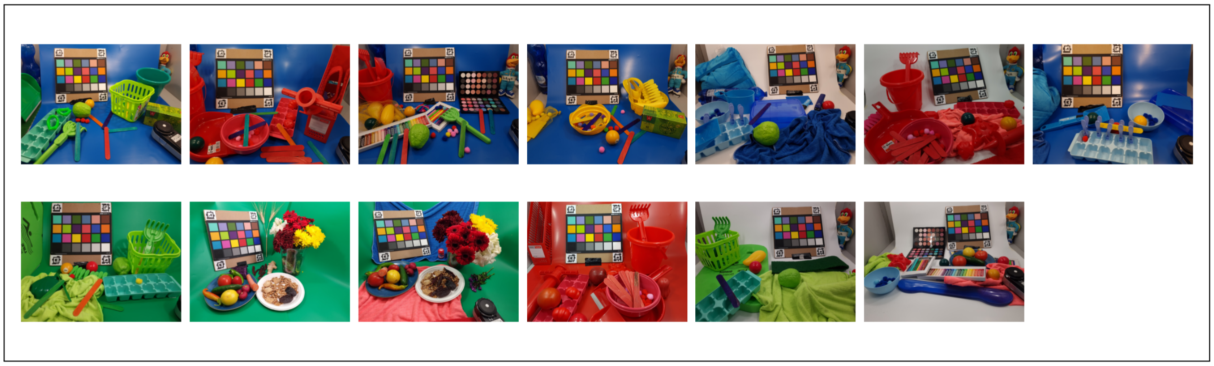


Figure S5. Beyond RGB lab scenarios with color-chart presented.

References

[1] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

[2] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[3] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

[4] Antonio Robles-Kelly and Ran Wei. A convolutional neural network for pixelwise illuminant recovery in colour and spectral images. In *2018 24th International Conference on Pattern Recognition (ICPR)*, pages 109–114. IEEE, 2018.