Ph.D. Dissertation

MODELING THE LIGHTING AS STYLE FACTOR VIA NEURAL NETWORKS FOR WHITE BALANCE CORRECTION

Furkan Kınlı

Advisor: Asst. Prof. M. Furkan Kıraç

Department of Computer Science
Özyeğin University

May 12, 2025

Acknowledgement

- Advisor, Asst. Prof. Furkan Kıraç
- Thesis committee
  - Prof. Lale Akarun
  - Prof. H. Fatih Uğurdağ
  - Prof. Erhan Öztop
  - Assoc. Prof. Aykut Erdem
- Vision and Graphics Lab @ Özyeğin
  - Dr. Barış Özcan
  - Sami Menteş, Doğa Yılmaz, Mert Erkol, Üveys Akbaş, Emir Ulurak
- Family

Outline

- Research Idea
- Style Factor & Feature Statistics
- Foundational Study (*IFRNet*)
- White Balance (WB) Correction
- Motivation for WB Correction
- First Attack: *Style WB*
- From Alignment To Exact Matching: *FDM WB*
- Feature Distribution Statistics As Loss Objective: *FDM Loss*
- Applications & Extensions
- Conclusion

Research Idea

Any disruptive factor in an image can be modeled as style factor.

Style Factors

- What is a style factor?
  - Extrinsic attribute changing the perceptual appearance, but not its content.
- In perceptual systems: Style ≠ Content †
  - accents, handwriting styles, brushstroke styles, etc.

† Tenenbaum, J. B. and Freeman, W. T. (2000). Separating style and content with bilinear models. In Neural Computation, vol 12:6.

Style Factors

- In existing literature, style factors are widely used in style transfer.
  - texture, color tone, brush stroke patterns, etc.

- Achieved by manipulating image/feature statistics.
  - Gram matrices[†‡], mean/variance[•], etc.

- **Novel approach:**
  - Instead of ~~transferring the style~~ one to another, use it to *remove* undesired style distortions. (*e.g.*, social media filters or color casts by complex illumination)

† Gatys, L. A., Ecker, A. S. and Bethge, M. (2015). A neural algorithm of artistic style. arXiv preprint arXiv:1508.06576.
‡ Li, Y., Wang, N., Liu, J. and Hou, X. (2017). Demystifying Neural Style Transfer. In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17) (pp. 2230-2236).
• Huang, X. and Belongie, S. (2017). Arbitrary style transfer in real-time with adaptive instance normalization. In Proceedings of the IEEE International Conference on Computer Vision (pp. 1501-1510).

(Deep) Feature Statistics

- Quantitative measure of feature representations.
  - describing the distribution and variability of features in **high-dimensional latent space**.
- For a feature map $\mathcal{F} \in \mathbb{R}^{C \times H \times W}$,

Mean: $$\mu_c = \frac{1}{HW} \sum_{h=1}^{H} \sum_{w=1}^{W} F_{c,h,w}$$

Skewness: $$\gamma_1 = \frac{1}{HW} \sum_{h=1}^{H} \sum_{w=1}^{W} \left(\frac{F_{c,h,w} - \mu_c}{\sigma_c} \right)^3$$

Variance: $$\sigma_c = \frac{1}{HW} \sum_{h=1}^{H} \sum_{w=1}^{W} (F_{c,h,w} - \mu_c)^2$$

Kurtosis: $$\gamma_2 = \frac{1}{HW} \sum_{h=1}^{H} \sum_{w=1}^{W} \left(\frac{F_{c,h,w} - \mu_c}{\sigma_c} \right)^4 - 3$$

- These statistics serve as the foundation for *manipulating* the feature distribution to achieve transferring or removing the style.

(Deep) Feature Statistics

- How can we manipulate the image/feature statistics?
  - Normalization techniques (*i.e.*, batch, instance, layer)
    - mostly want to *shift to arbitrary mean and variance for single instance regarding to a reference*, so definitely not a good idea.
  - Covariance matrices vs. Gram matrices
    - could be option in image space, *but feature representations have generally more information in more compact way*, so may not be a good idea.
  - Adaptive Instance Normalization (AdaIN)[†]

$$\mathrm{AdaIN}(x, y) = \sigma(y)\left(\frac{x - \mu(x)}{\sigma(x)}\right) + \mu(y)$$

    - however, it only **aligns** the channel-wise mean and variance of the feature maps to those of the style feature maps, so good option, but still not enough.

† Huang, X. and Belongie, S. (2017). Arbitrary style transfer in real-time with adaptive instance normalization. In Proceedings of the IEEE International Conference on Computer Vision (pp. 1501-1510).

(Deep) Feature Statistics

- Exact Feature Distribution Matching† (EFDM)
  - practical algorithm of matching empirical Cumulative Distribution Functions (eCDF) of feature maps.
  - not aligning, but exact matching.
  - grounded in the Glivenko-Cantelli theorem‡

$$\sup_x |\hat{F}_n(x) - F(x)| \xrightarrow{\text{a.s.}} 0 \text{ as } n \to \infty$$

- Practical implementation:

Algorithm 1 PyTorch-like pseudo-code for EFDM.

X: input vector, **Y**: target vector
_, IndexX $=$ torch.sort(**X**)
SortedY, _ $=$ torch.sort(**Y**)
InverseIndex $=$ IndexX.argsort(-1)
return X$+$ SortedY.gather(-1, InverseIndex) $-$**X**.detach()

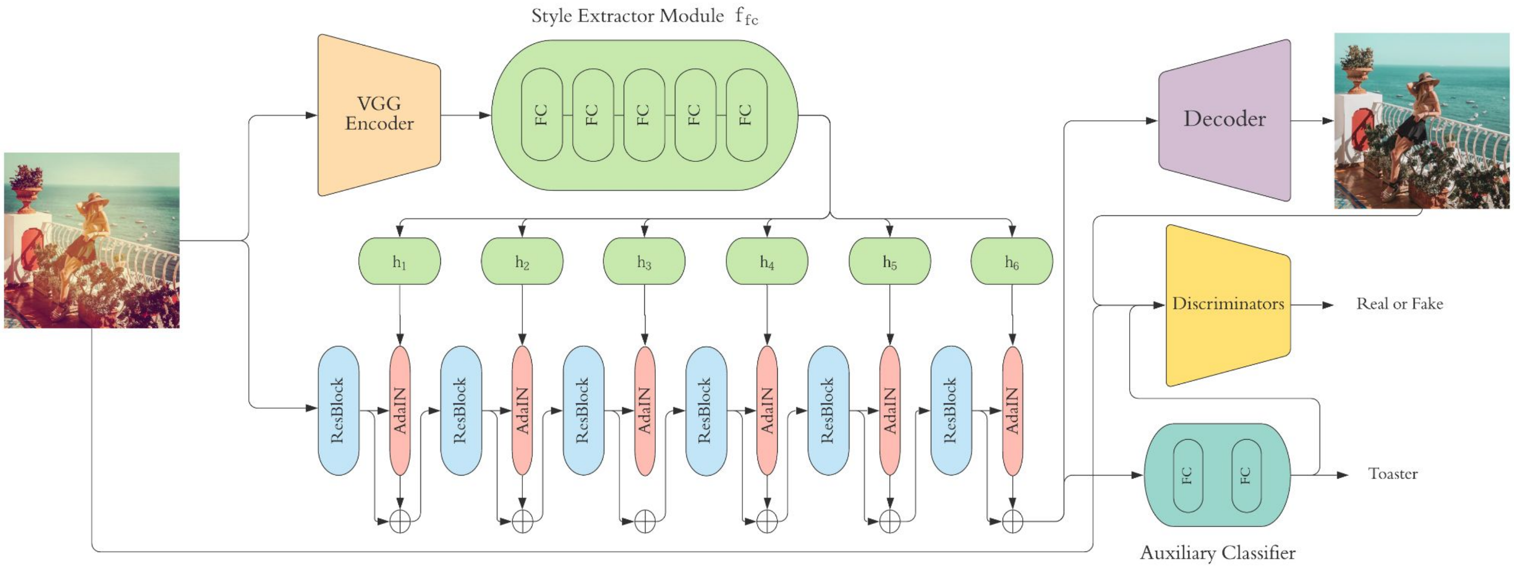
† Zhang, Y., *et al.* (2022). Exact Feature Distribution Matching for Arbitrary Style Transfer and Domain Generalization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 8035-8045).
‡ Van der Vaart, A. W. (2000). Asymptotic statistics. vol 3. Cambridge university press.
• Erkol, M., Kınlı, F., Özcan B. and Kıraç, F. (2023). [Re] Exact Feature Distribution Matching for Arbitrary Style Transfer and Domain Generalization. In ReScience Journal - ML Reproducibility Challenge 2022 vol 9:2.

Research Idea

Any disruptive factor in an image can be modeled as style factor.

For any non-linear filter applied to image 🔎

Foundational Study (IFRNet†)



† Kınlı, F., Özcan, B., and Kıraç, F. (2021). Instagram Filter Removal on Fashionable Images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (pp. 736-745).

Research Idea

Any disruptive factor in an image can be modeled as style factor.

For any non-linear filter applied to image ✅

Lighting? Illumination? 🔎

White Balance Correction

- What is White Balance?
  - White objects appear neutral white.
  - … regardless of the lighting conditions.
- Why does it matter?
  - Different light sources → different color temperatures
  - Shade (~7500**K**) or Tungsten (~2500**K**)
  - No correction → Unnatural color casts

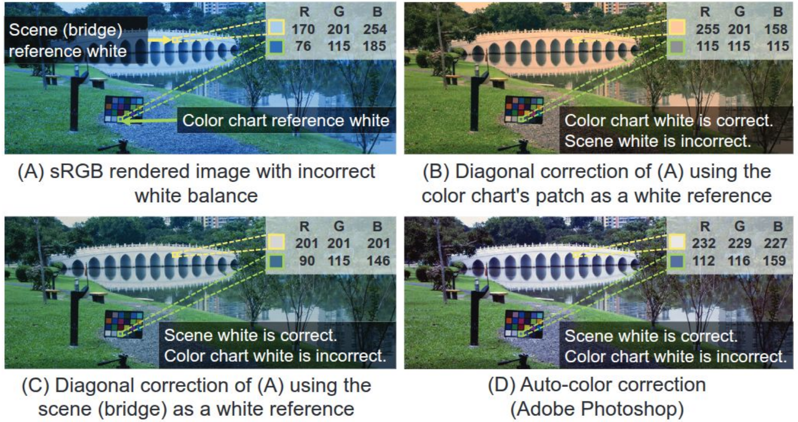
Tungsten

Shade

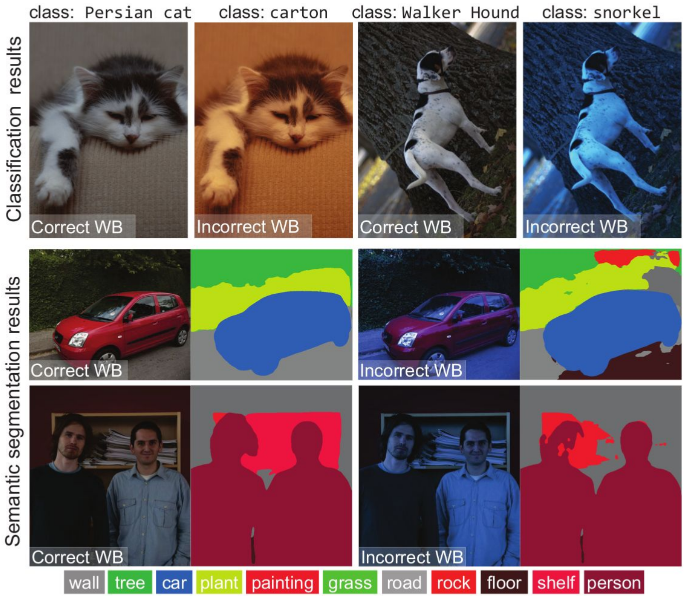
AWB Correction

Motivation for WB Correction

Auto-White Balance (AWB) is often inaccurate.†



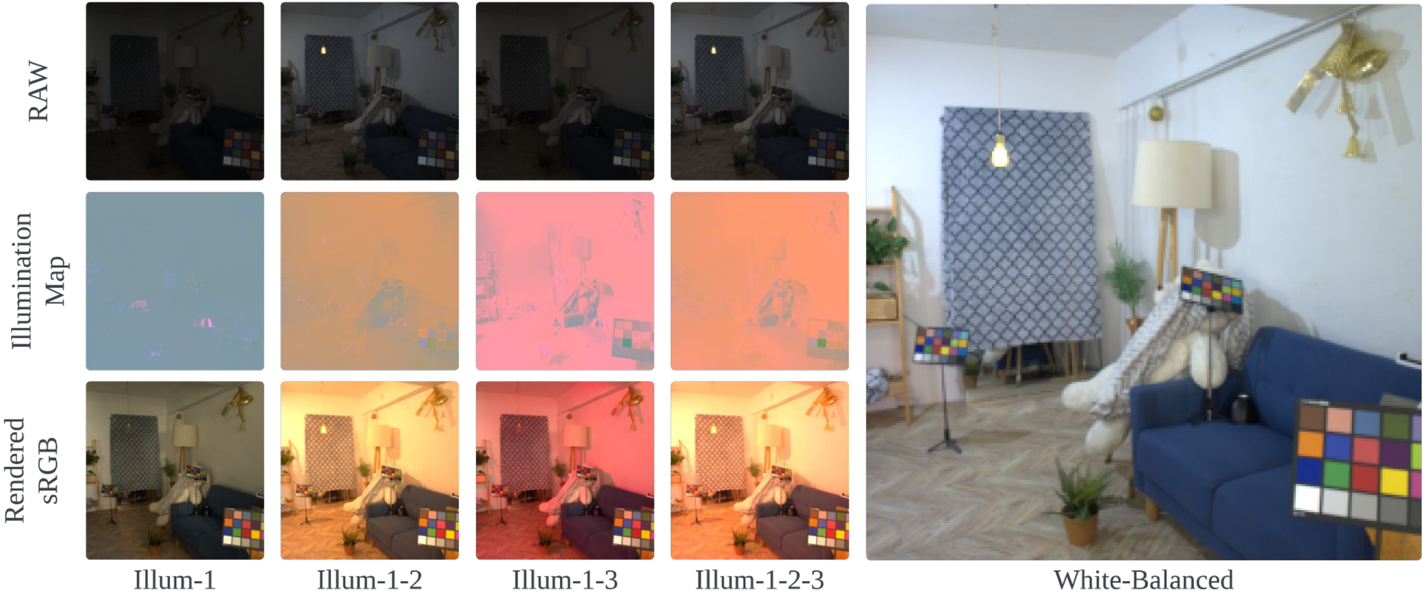
Color Constancy critical for downstream vision tasks.‡



† Afifi, M., Price, B., Cohen, S., and Brown, M. S. (2019). When Color Constancy Goes Wrong: Correcting Improperly White-Balanced Images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 1535-1544).
‡ Afifi, M. and Brown, M. S. (2019). What Else Can Fool Deep Learning? Addressing Color Constancy Errors on Deep Neural Network Performance. In Proceedings of the IEEE International Conference on Computer Vision (pp. 243-252).

Motivation for WB Correction

- But, it is challenging
  - Lighting is rarely uniform.



White Balance Correction

- Foundational approach
  - The Gray-World assumption[†] + Diagonal Correction
  - relying on Retinex Theory[‡]
  - lots of variants

$$R' = \frac{R}{\frac{1}{N}\sum_{i=1}^{N} R_i}, \quad G' = \frac{G}{\frac{1}{N}\sum_{i=1}^{N} G_i}, \quad B' = \frac{B}{\frac{1}{N}\sum_{i=1}^{N} B_i}$$

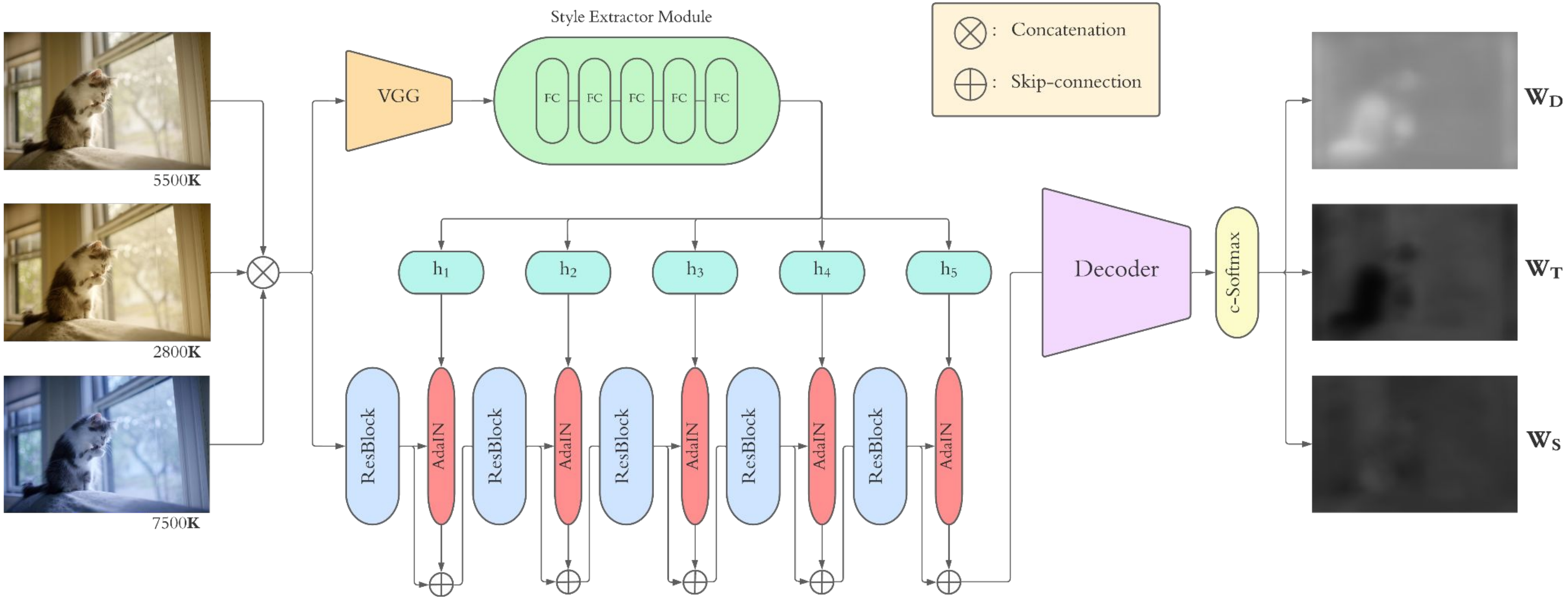
- Gamut-mapping
  - any deviation from the canonical gamut → a shift in the light source[•]
- Low-level statistical methods
  - Bayesian methods[‣] and its variants
- Scene semantics
  - based on prior knowledge of the world[ⱡ]
  - hard to have prior?
- (Deep) Neural Network methods
  - Powerful, but need data?

† Buchsbaum, G. (1980). A spatial processor model for object colour perception. In Journal of the Franklin institute, vol 310:1 (pp. 1-26).
‡ Land, E. H. (1977). The retinex theory of color vision. In Scientific american, vol 23:6 (pp. 108-129).
• Forsyth, D. A. (1990). A novel algorithm for color constancy. In International Journal of Computer Vision, vol 5:1 (pp. 5-35).
‣ Finlayson, G. D., Hordley, S. D. and Hubel, P. M. (2001). Color by correlation: A simple, unifying framework for color constancy. In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 23:11 (pp. 1209-1221)
ⱡ Gijsenij, A. and Gevers, T. (2010). Color constancy using natural image statistics and scene semantics. In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 33:4 (pp. 687-698)

Mid-break

- ~~Research Idea~~
- ~~Style Factor & Feature Statistics~~
- ~~Foundational Study (*IFRNet*)~~
- ~~White Balance (WB) Correction~~
- ~~Motivation for WB Correction~~
- First Attack: *Style WB*
- From Alignment To Exact Matching: *FDM WB*
- Feature Distribution Statistics As Loss Objective: *FDM Loss*
- Applications & Extensions
- Conclusion

First Attack: *Style WB*



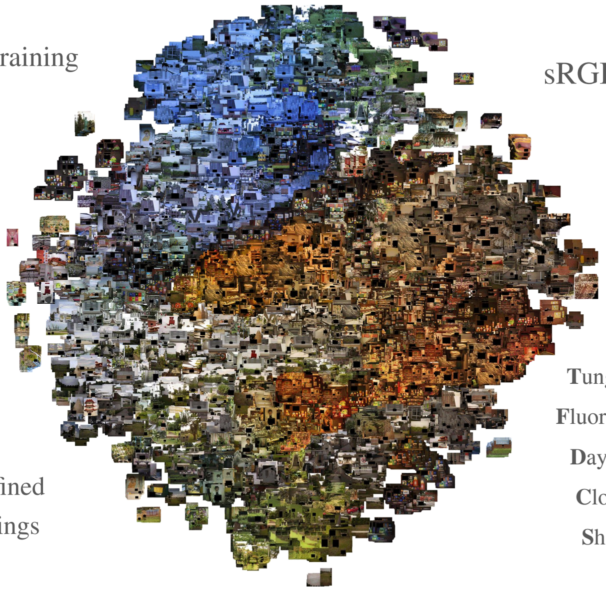
† Kınlı, F., Yılmaz, D., Özcan, B., and Kıraç, F. (2023). Modeling the Lighting in Scenes as Style for Auto White-Balance Correction. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 4903-4913).

First Attack: *Style WB*

RenderedWB Dataset†

Used for training

65,000
sRGB images



Tungsten 2800**K**

Fluorescent 3800**K**

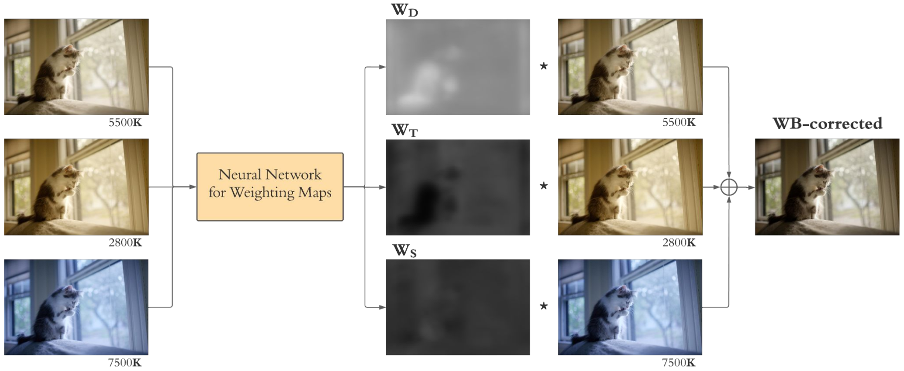
Daylight 5500**K**

Cloudy 6500**K**

Shade 7500**K**

5 pre-defined
WB settings

Learning weighting maps of WB Settings‡



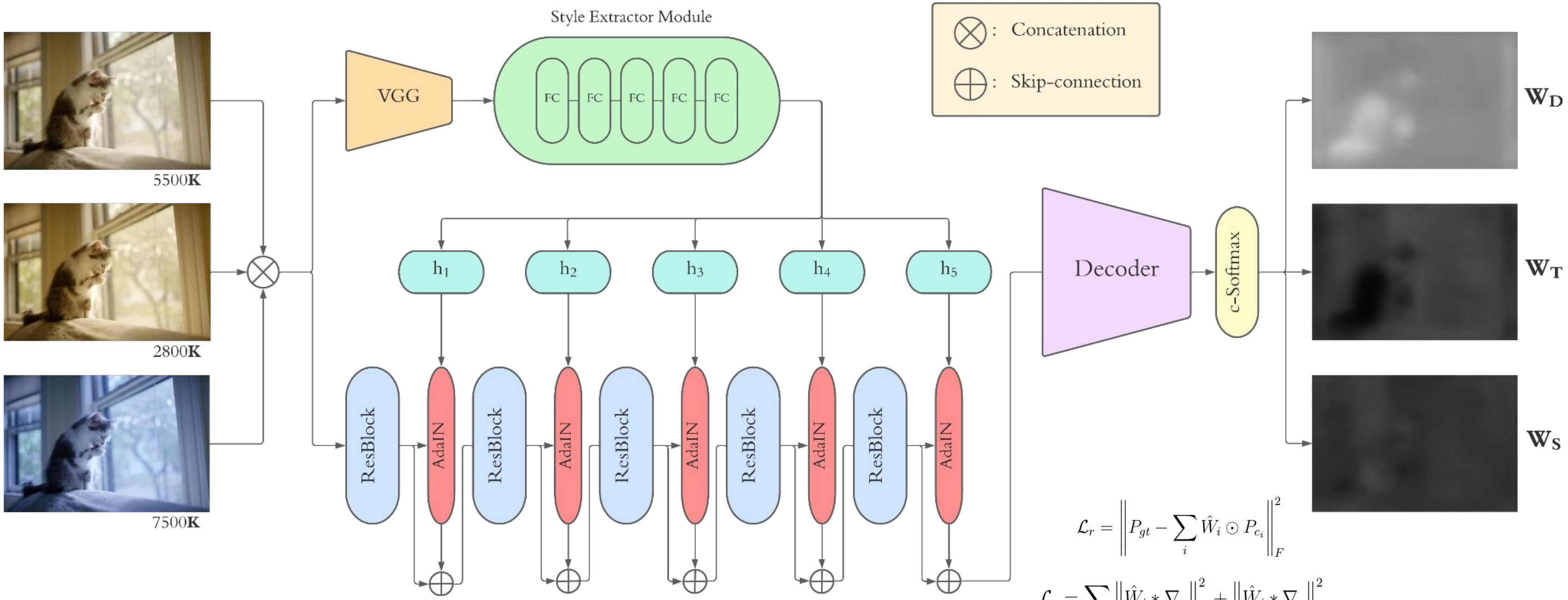
$$\tilde{I}_{corr} = \sum_i W_i \odot \tilde{I}_{c_i}$$

*Relying on Retinex Theory!

- Two configurations for WB settings: {t,f,d,c,s} | {t,d,s}

† Afifi, M., Price, B., Cohen, S., and Brown, M. S. (2019). When Color Constancy Goes Wrong: Correcting Improperly White-Balanced Images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 1535-1544).
‡ Afifi, M., Brubaker, M. A. and Brown, M. S. (2022). Auto white-balance correction for mixed-illuminant scenes. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 1210-1219).

First Attack: *Style WB*



$$\mathcal{L}_r = \left\| P_{gt} - \sum_i \hat{W}_i \odot P_{c_i} \right\|_F^2$$

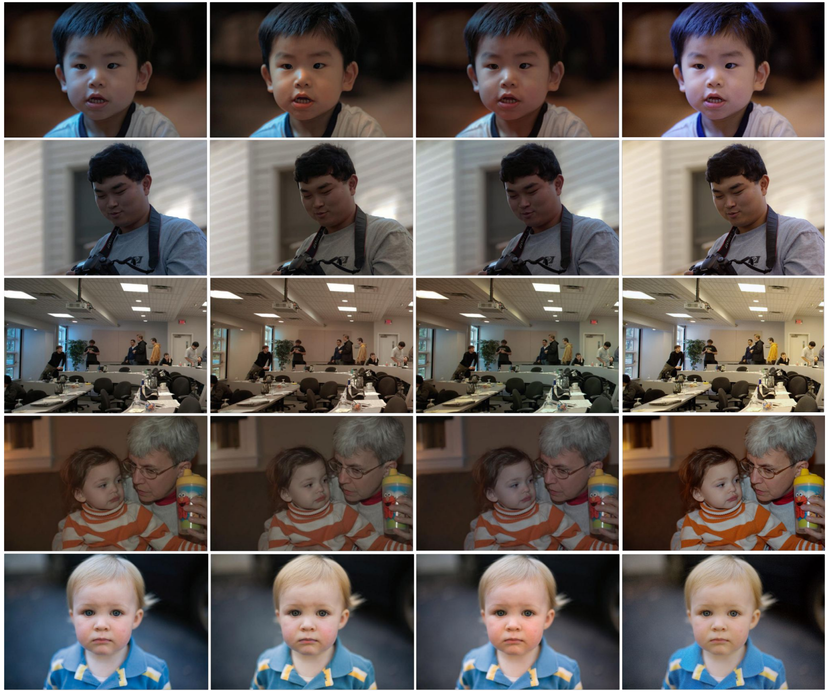
$$\mathcal{L}_s = \sum_i \left\| \hat{W}_i * \nabla_x \right\|_F^2 + \left\| \hat{W}_i * \nabla_y \right\|_F^2$$

† Kınlı, F., Yılmaz, D., Özcan, B., and Kıraç, F. (2023). Modeling the Lighting in Scenes as Style for Auto White-Balance Correction. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 4903-4913).

First Attack: *Style WB*

Table 5.2: *Benchmark on single-illuminant Cube+ dataset [1]. The top results are indicated with colored cells as, the best: **green**, the second: yellow, the third: red.*

Method	MSE				MAE				ΔE 2000				Size
	Mean	Q1	Q2	Q3	Mean	Q1	Q2	Q3	Mean	Q1	Q2	Q3	
FC4 [17]	371.90	79.15	213.41	467.33	6.49°	3.34°	5.59°	8.59°	10.38	6.60	9.76	13.26	5.89 MB
Quasi-U CC [19]	292.18	15.57	55.41	261.58	6.12°	1.95°	3.88°	8.83°	7.25	2.89	5.21	10.37	622 MB
KNN WB [6]	194.98	27.43	57.08	118.21	4.12°	1.96°	3.17°	5.04°	5.68	3.22	4.61	6.70	21.8 MB
Interactive WB [123]	159.88	21.94	54.76	125.02	4.64°	2.12°	3.64°	5.98°	6.20	3.28	5.17	7.45	**38 KB**
Deep WB [20]	**80.46**	15.43	33.88	74.42	3.45°	1.87°	2.82°	4.26°	4.59	2.68	3.81	5.53	16.7 MB
Mixed WB [2]													
$p = 64$, WB={t, d, s}	168.38	8.97	19.87	105.22	4.20°	1.39°	2.18°	5.54°	5.03	2.07	3.12	7.19	5.09 MB
$p = 64$, WB={t, f, d, c, s}	161.80	9.01	19.33	90.81	4.05°	1.40°	2.12°	4.88°	4.89	2.16	3.10	6.78	5.10 MB
$p = 128$, WB={t, f, d, c, s}	176.38	16.96	35.91	115.50	4.71°	2.10°	3.09°	5.92°	5.77	3.01	4.27	7.71	5.10 MB
Style WB (ours)													
$p = 64$, WB={t, d, s}	92.65	**6.52**	**14.23**	**35.01**	**2.47°**	**0.82°**	**1.44°**	**2.49°**	**2.99**	**1.36**	**2.04**	**3.32**	61.0 MB
$p = 64$, WB={t, f, d, c, s}	151.38	29.49	56.35	125.33	4.18°	2.13°	3.03°	4.81°	5.42	3.11	4.42	6.76	61.1 MB
$p = 128$, WB={t, d, s}	88.03	7.92	17.73	45.01	2.61°	0.93°	1.58°	2.85°	3.24	1.50	2.30	3.95	61.2 MB
$p = 128$, WB={t, f, d, c, s}	100.24	10.77	37.74	70.18	3.09°	1.15°	2.61°	3.87°	3.96	1.59	3.55	5.51	61.3 MB



Default AWB | Deep WB | Mixed WB | Style WB

Cube+: Banić, N., Koščević, K. and Lončarić, S. (2017). Unsupervised learning for color constancy. arXiv preprint arXiv:1712.00436.
MIT Adobe FiveK: Bychkovsky, V., Paris, S., Chan, E., and Durand, F. (2011). Learning photographic global tonal adjustment with a database of input/output image pairs. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 97-104).

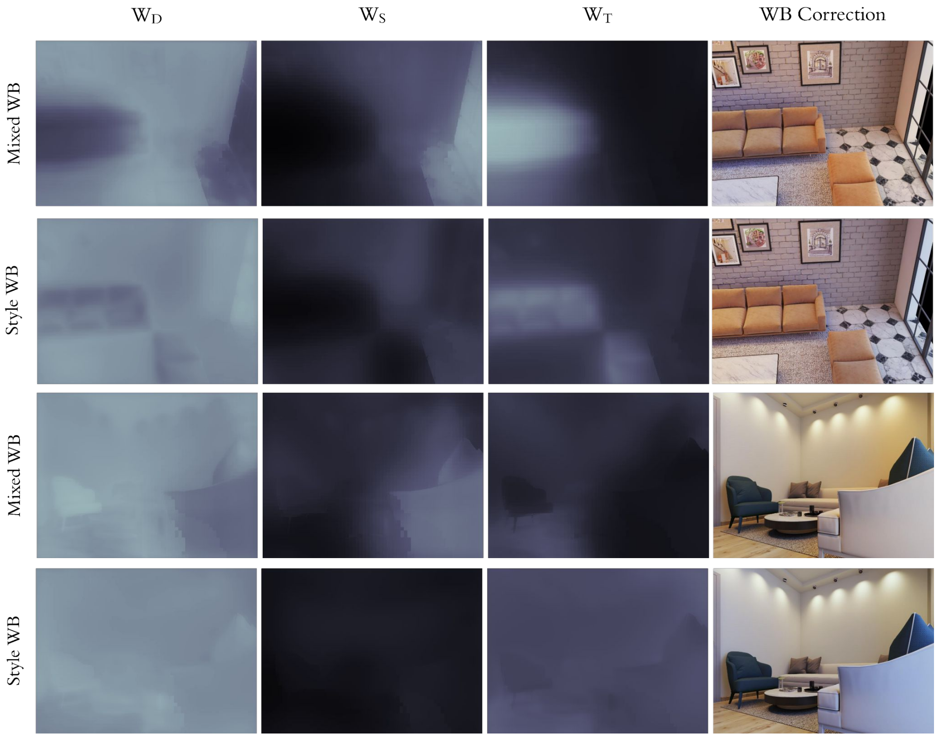
First Attack: *Style WB*

Table 5.3: *Benchmark on mixed-illuminant evaluation set [2]. The top results are indicated with colored cells as, the best: **green**, the second: yellow, the third: red.*

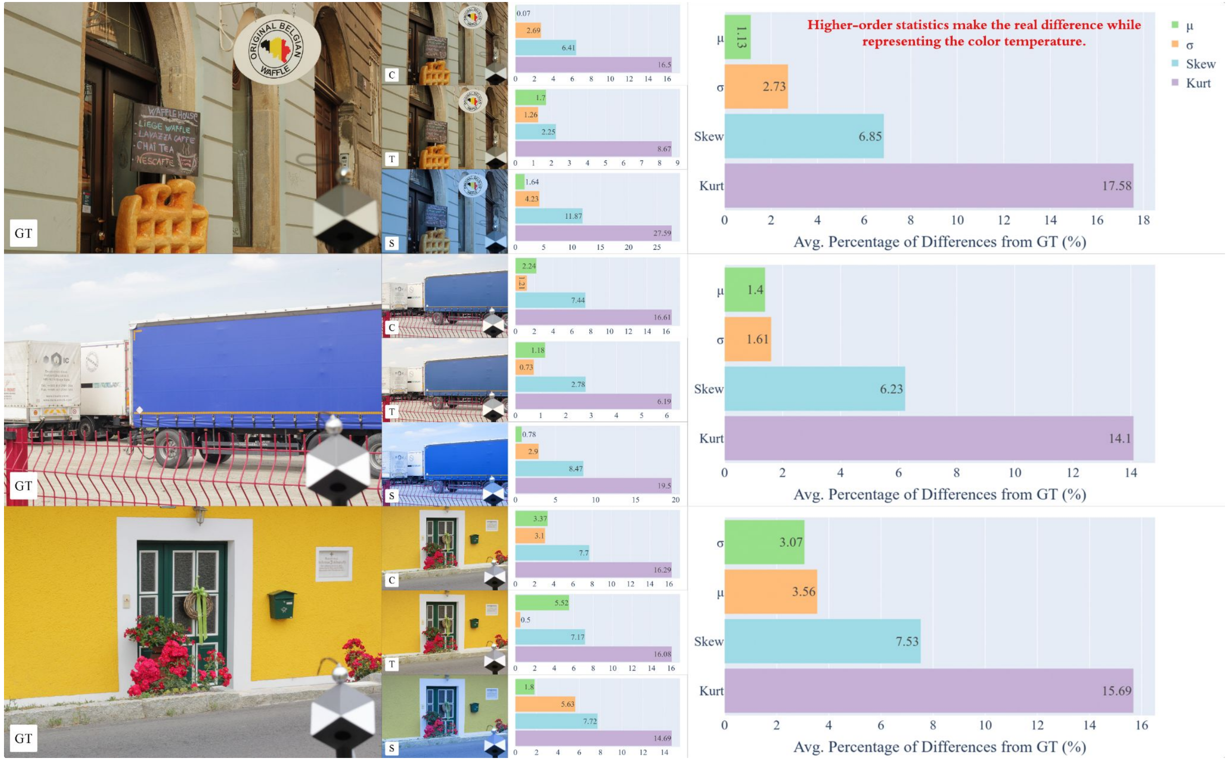
Method	MSE				MAE				Δ E 2000			
	Mean	**Q1**	**Q2**	**Q3**	**Mean**	**Q1**	**Q2**	**Q3**	**Mean**	**Q1**	**Q2**	**Q3**
Gray Pixel [101]	4959.20	3252.14	4209.12	5858.69	19.67°	11.92°	17.21°	27.05°	25.13	19.07	22.62	27.46
Grayness index [102]	1345.47	727.90	1055.83	1494.81	6.39°	4.72°	5.65°	7.06°	12.84	9.57	12.49	14.60
KNN WB [6]	1226.57	680.65	1062.64	1573.89	5.81°	4.29°	5.76°	6.85°	12.00	9.37	11.56	13.61
Interactive WB [123]	1059.88	616.24	896.90	1265.62	5.86°	4.56°	5.62°	6.62°	11.41	8.92	10.99	12.84
Deep WB [20]	1130.60	621.00	886.32	1274.72	**4.53°**	**3.55°**	4.19°	**5.21°**	10.93	**8.59**	**9.82**	11.96
Mixed WB [2]												
$p = 64$, WB={`t,d,s`}	**819.47**	655.88	845.79	1000.82	5.43°	4.27°	4.89°	6.23°	**10.61**	9.42	10.72	**11.81**
$p = 64$, WB={`t,f,d,c,s`}	938.02	757.49	961.55	1161.52	4.67°	3.71°	**4.14°**	5.35°	12.26	10.80	11.58	12.76
$p = 128$, WB={`t,d,s`}	830.20	584.77	853.01	**992.56**	5.03°	3.93°	4.78°	5.90°	11.41	9.76	11.39	12.53
$p = 128$, WB={`t,f,d,c,s`}	1089.69	846.21	1125.59	1279.39	5.64°	4.15°	5.09°	6.50°	13.75	11.45	12.58	15.59
Style WB (ours)												
$p = 64$, WB={`t,d,s`}	868.01	649.36	889.00	1026.98	5.73°	4.48°	5.42°	6.34°	12.11	10.42	12.12	13.36
$p = 64$, WB={`t,f,d,c,s`}	1051.07	760.86	1024.00	1332.50	6.30°	4.43°	6.01°	7.69°	14.43	11.90	13.11	16.15
$p = 128$, WB={`t,d,s`}	822.77	**576.52**	**840.67**	1025.26	5.11°	3.93°	4.85°	5.51°	11.65	10.63	11.86	13.02
$p = 128$, WB={`t,f,d,c,s`}	834.28	629.95	842.71	1005.59	5.71°	4.57°	5.54°	6.19°	11.79	9.84	12.19	13.00

Table 5.4: *The ablation study on using multi-scale (ms) weighting maps and applying edge-aware smoothing (eas) to weighting maps.*

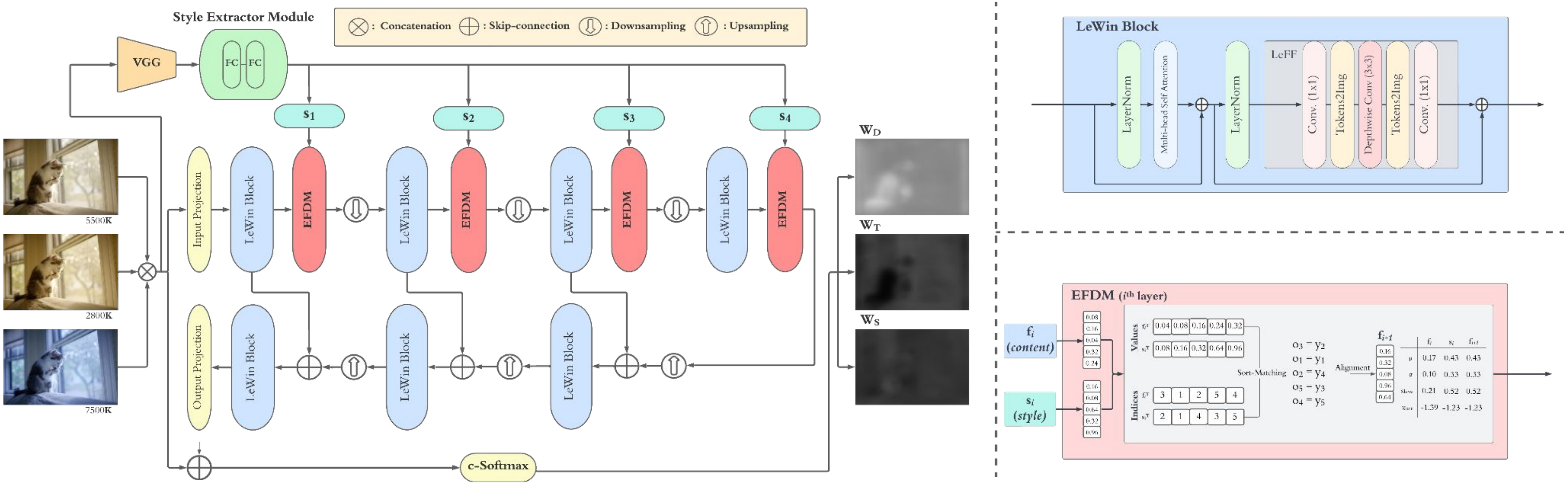
Models	**MSE**	**MAE**	**ΔE 2000**
Single-illuminant dataset, WB = {`t,d,s`}, $p = 64$			
$ms = 0, eas = 0$	98.55	2.71°	3.32
$ms = 1, eas = 0$	93.78	2.59°	3.15
$ms = 0, eas = 1$	97.20	2.66°	3.28
$ms = 1, eas = 1$	**92.65**	**2.47°**	**2.99**
Mixed-illuminant dataset, WB = {`t,d,s`}, $p = 128$			
$ms = 0, eas = 0$	878.58	5.05°	12.12
$ms = 1, eas = 0$	843.50	**5.04°**	11.70
$ms = 0, eas = 1$	843.64	**5.04°**	11.98
$ms = 1, eas = 1$	**822.77**	5.11°	**11.65**



From Alignment To Exact Matching: *FDM WB*



From Alignment To Exact Matching: *FDM WB*



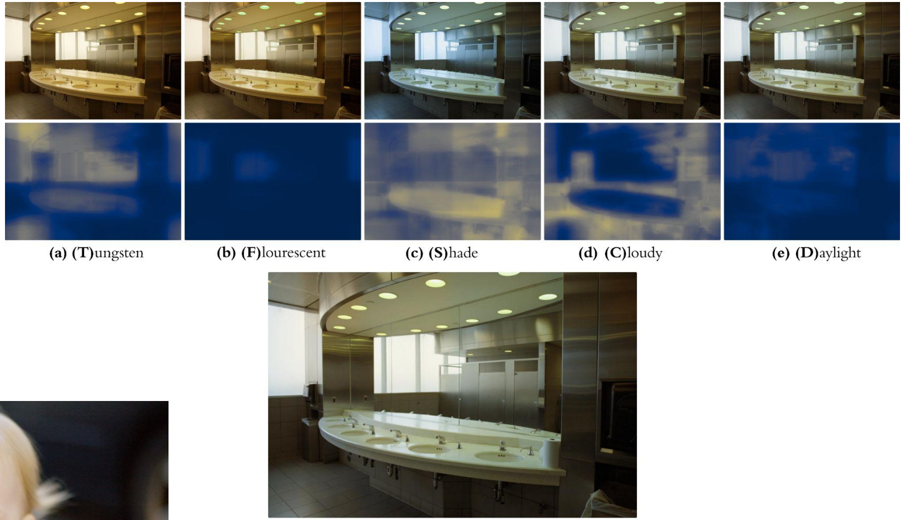
$$\mathcal{L}_r = \left\| P_{gt} - \sum_i \hat{W}_i \odot P_{c_i} \right\|_F^2 \qquad \mathcal{L}_s = \sum_i \left\| \hat{W}_i * \nabla_x \right\|_F^2 + \left\| \hat{W}_i * \nabla_y \right\|_F^2$$

† Kınlı, F., Özcan, B., and Kıraç, F. (2025). Advancing white balance correction through deep feature statistics and feature distribution matching. In Journal of Visual Communication and Image Representation, vol 108:4, 104412.

From Alignment To Exact Matching: *FDM WB*

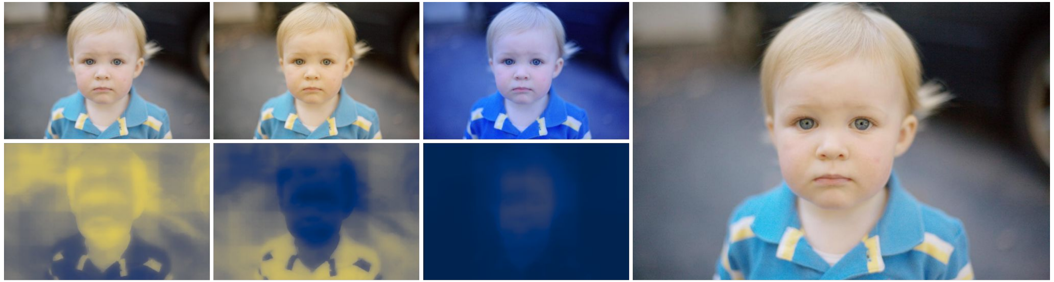
Table 5.5: *Benchmark on single-illuminant Cube+ dataset [1]. ↓ denotes that lower is better.*

	MSE ↓				MAE ↓				ΔE 2000 ↓			
Methods	Mean	Q1	Q2	Q3	Mean	Q1	Q2	Q3	Mean	Q1	Q2	Q3
FC4 [17]	371.90	79.15	213.41	467.33	6.49°	3.34°	5.59°	8.59°	10.38	6.60	9.76	13.26
Quasi-U CC [19]	292.18	15.57	55.41	261.58	6.12°	1.95°	3.88°	8.83°	7.25	2.89	5.21	10.37
KNN WB [6]	194.98	27.43	57.08	118.21	4.12°	1.96°	3.17°	5.04°	5.68	3.22	4.61	6.70
Interactive WB [123]	159.88	21.94	54.76	125.02	4.64°	2.12°	3.64°	5.98°	6.20	3.28	5.17	7.45
Deep WB [20]	80.46	15.43	33.88	74.42	3.45°	1.87°	2.82°	4.26°	4.59	2.68	3.81	5.53
MIMT [135]	-	-	-	-	2.52°	0.98°	1.38°	2.96°	2.88	1.94	2.42	2.87
Mixed WB [2]	161.80	9.01	19.33	90.81	4.05°	1.40°	2.12°	4.88°	4.89	2.16	3.10	6.78
Style WB [24]	88.03	7.92	17.73	45.01	2.61°	0.93°	1.58°	2.85°	3.24	1.50	2.30	3.95
DeNIM + Mixed WB [145]	99.70	13.89	24.71	43.88	2.49°	1.07°	1.62°	2.41°	3.44	1.95	2.74	3.78
DeNIM + Style WB [145]	83.41	13.23	21.46	37.44	1.93°	0.77°	1.09°	1.70°	2.73	1.62	2.03	2.71
FDM WB (**ours**)	**79.35**	**6.46**	**16.84**	**35.76**	**1.35°**	**0.56°**	**1.01°**	**1.66°**	**1.40**	**0.98**	**1.41**	**2.55**



(a) (T)ungsten **(b) (F)**lourescent **(c) (S)**hade **(d) (C)**loudy **(e) (D)**aylight

(f) AWB Correction



(a) (D)aylight **(b) (T)**ungsten **(c) (S)**hade **(d)** AWB Correction

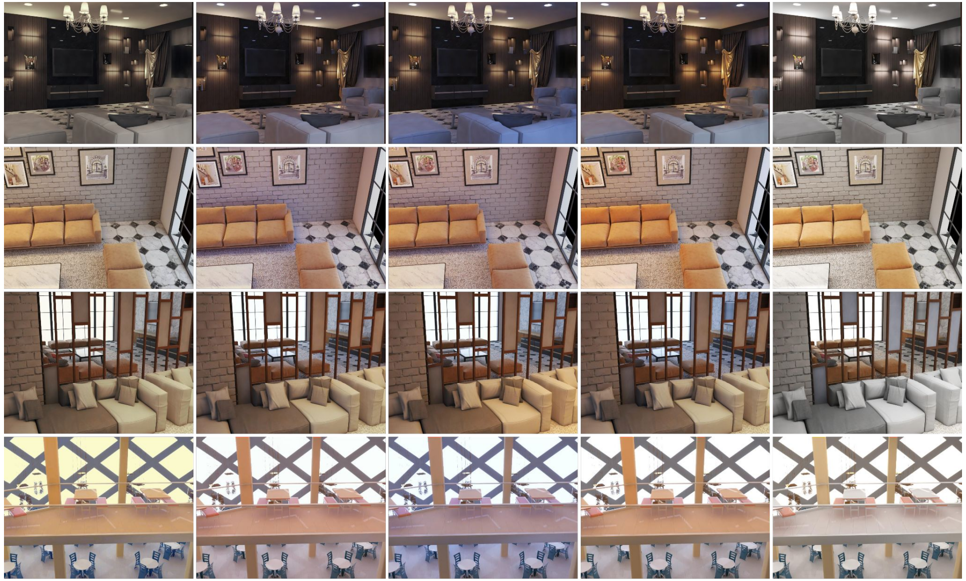
From Alignment To Exact Matching: *FDM WB*

Table 5.6: *Benchmark on mixed-illuminant evaluation set [2]. ↓ denotes that lower is better.*

Methods	MSE ↓				MAE ↓				ΔE 2000 ↓			
	Mean	Q1	Q2	Q3	Mean	Q1	Q2	Q3	Mean	Q1	Q2	Q3
Gray Pixel [101]	4959.2	3252.1	4209.1	5858.7	19.67°	11.92°	17.21°	27.05°	25.13	19.07	22.62	27.46
Grayness In. [102]	1345.5	727.9	1055.8	1494.8	6.39°	4.72°	5.65°	7.06°	12.84	9.57	12.49	14.60
KNN WB [6]	1226.6	680.7	1062.6	2573.9	5.81°	4.29°	5.76°	6.85°	12.00	9.37	11.56	13.61
Interact. WB [123]	1059.9	616.2	896.9	1265.6	5.86°	4.56°	5.62°	6.62°	11.41	8.92	10.99	12.84
Deep WB [20]	1130.6	621.0	886.3	1274.7	**4.53°**	**3.55°**	**4.19°**	**5.21°**	10.93	**8.59**	9.82	11.96
Mixed WB [2]	819.5	655.9	845.8	1000.8	5.43°	4.27°	4.89°	6.23°	10.61	9.42	10.72	11.81
Style WB [24]	822.8	576.5	840.7	1025.3	5.11°	3.93°	4.85°	5.51°	11.65	10.63	11.86	13.02
FDM WB (**ours**)	**761.9**	**513.9**	**818.4**	**969.3**	5.95°	4.64°	5.88°	6.90°	**10.16**	8.75	**9.81**	**11.69**

Table 5.7: *Ablation study on the impact of employing the Style Extractor module and EFDM on Cube+ dataset [1] and mixed-illuminant evaluation set [2].*

Method	MSE ↓	MAE ↓	ΔE 2000 ↓
Cube+ dataset			
$p = 64$, Uformer [5]	107.38	2.80°	3.46
$p = 64$, FDM WB	91.34	2.38°	2.88
$p = 128$, Uformer [5]	105.68	2.77°	3.39
$p = 128$, FDM WB	**79.35**	**1.35°**	**1.40**
Mixed-illuminant evaluation set			
$p = 64$, Uformer [5]	939.52	4.98°	12.97
$p = 64$, FDM WB	780.74	4.85°	10.84
$p = 128$, Uformer [5]	1067.37	5.99°	14.43
$p = 128$, FDM WB	**761.95**	**5.95°**	**10.16**

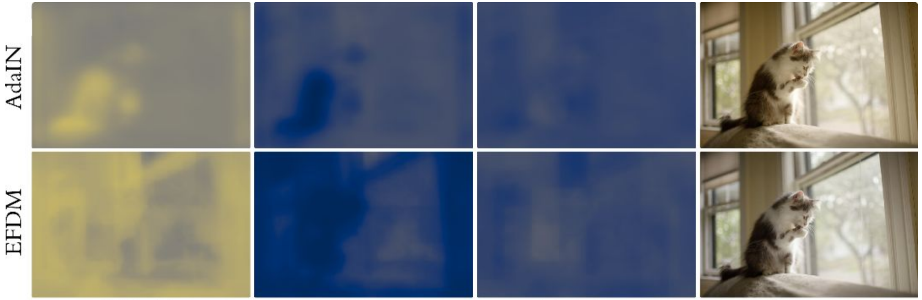


(a) Deep WB **(b)** Mixed WB **(c)** Style WB **(d)** FDM WB **(ours)** **(e)** GT

From Alignment To Exact Matching: *FDM WB*

Table 5.8: *Ablation study on style factor learning strategy on Cube+ dataset [1] and mixed-illuminant evaluation set [2].*

Method	MSE ↓	MAE ↓	ΔE 2000 ↓
Cube+ dataset			
AdaIN [42]	92.47	1.78°	1.94
EFDM	**79.35**	**1.35°**	**1.40**
Mixed-illuminant evaluation set			
AdaIN [42]	818.99	**5.41°**	11.01
EFDM	**761.95**	5.95°	**10.16**



(a) Daylight (b) Tungsten (c) Shade (d) WB Correction

Table 5.9: *Ablation study on changing patch size and using different WB settings on Cube+ dataset [1] and mixed-illuminant evaluation set [2].*

Method	MSE ↓	MAE ↓	ΔE 2000 ↓
Cube+ dataset			
$p = 64$, {t,d,s}	91.34	2.38°	2.88
$p = 64$, {t,f,d,c,s}	118.51	3.65°	4.56
$p = 128$, {t,d,s}	79.35	**1.35°**	**1.40**
$p = 128$,{t,f,d,c,s}	**78.76**	1.54°	1.69
Mixed-illuminant evaluation set			
$p = 64$, {t,d,s}	780.74	4.85°	10.84
$p = 64$, {t,f,d,c,s}	815.24	4.82°	11.36
$p = 128$, {t,d,s}	**761.95**	5.95°	**10.16**
$p = 128$,{t,f,d,c,s}	822.12	**4.73°**	11.08

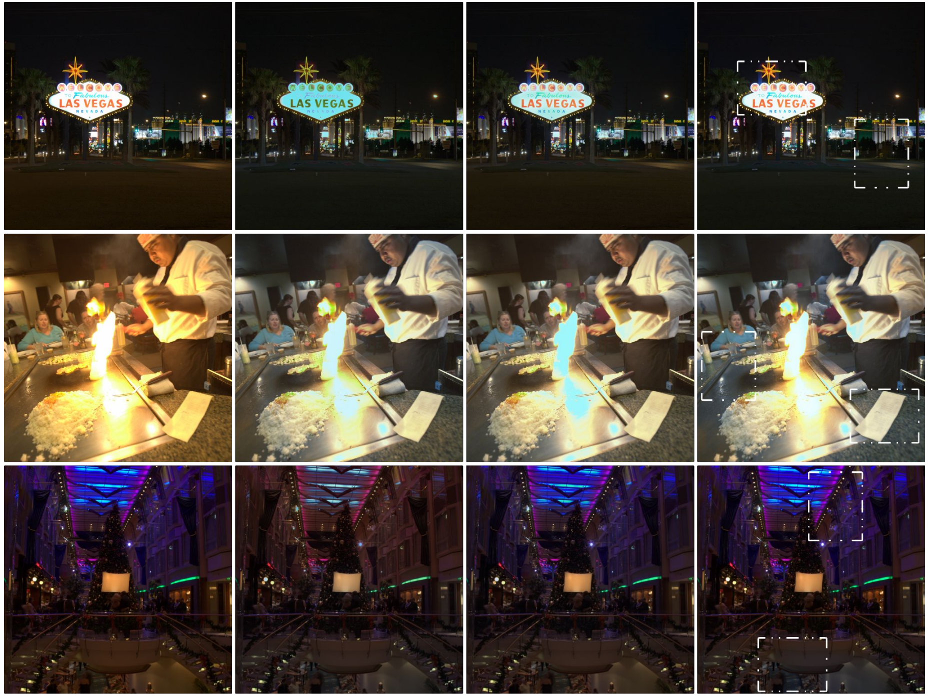
Table 5.10: *Ablation study on the effect of post-processing operation on the performance of our proposed model on Cube+ dataset [1].*

Method	MSE ↓	MAE ↓	ΔE 2000 ↓	Time (s)
ms ✗, eas ✗	85.20	1.33°	1.35	0.292
ms ✗, eas ✓	80.11	**1.29°**	**1.32**	11.051
ms ✓, eas ✗	80.72	1.37°	1.41	0.337
ms ✓, eas ✓	**79.35**	1.35°	1.40	11.228

From Alignment To Exact Matching: *FDM WB*

Table 5.11: *Comparison of the complexity of FDM WB and the prior methods with their post-processing tricks.*

Method	Time (s)	# of Params (M)	FLOPs (G)
Mixed WB [2] + ms + eas	10.390	**1.32**	**9.78**
Mixed WB [2] + eas	10.279		
Mixed WB [2] + ms	0.228		
Mixed WB [2]	**0.212**		
Style WB [24] + ms + eas	10.342	15.31	126.60
Style WB [24] + eas	10.307		
Style WB [24] + ms	0.232		
Style WB [24]	0.217		
FDM WB (ours) + ms + eas	11.228	20.53	61.92
FDM WB (ours) + eas	11.041		
FDM WB (ours) + ms	0.337		
FDM WB (ours)	0.292		



(a) Deep WB **(b)** Mixed WB **(c)** Style WB **(d)** FDM WB **(ours)**

From Alignment To Exact Matching: *FDM WB*

- Still lots of room for improvement…
  - Model complexity, too much?
  - EFDM, as a layer in architecture?
  - Only working on single illuminant data rendered with pre-defined WB settings?
  - Mean-squared error, as the objective function?
  - Better/efficient style representation, possible?

Feature Distribution Statistics As Loss Objective: *FDM Loss*

- **Research Question:** What happens if we remove *StyleExtractor* module and find better and efficient representation of style factor?

We can dramatically reduce the complexity if we remove StyleExtractor.

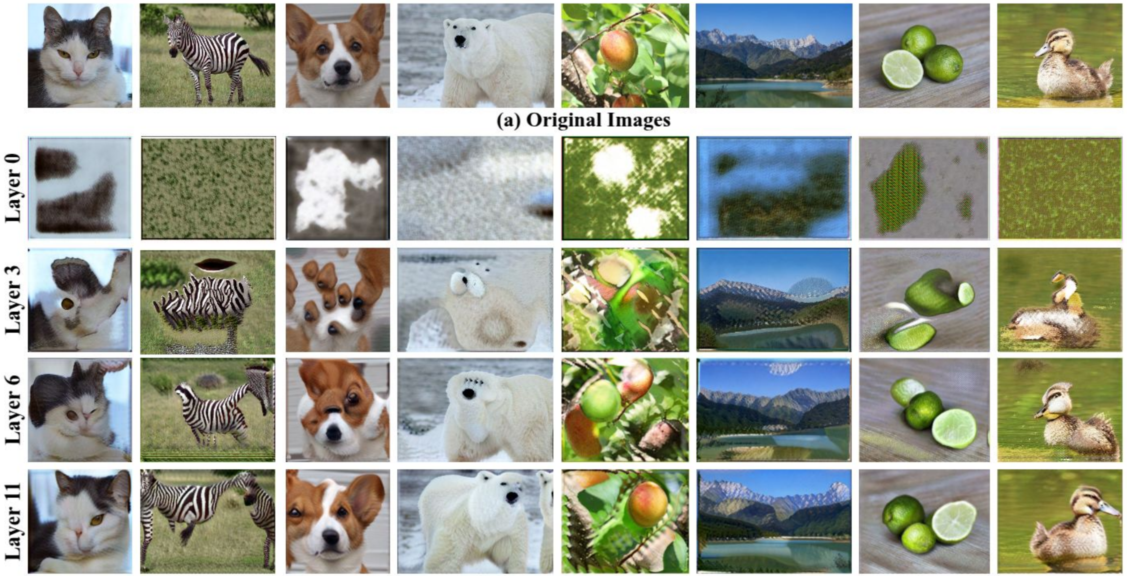
Visual Transformers (ViT) and its [CLS] *token has the potential to better and efficient represent the style factor*†.

- Still lots of room for improvement for *FDM WB*
  - ~~Model complexity, too much?~~
  - EFDM, as a layer in architecture?
  - Only working on single illuminant data rendered with pre-defined WB settings?
  - Mean-squared error, as the objective function?
  - ~~Better/efficient style representation, possible?~~

† Tumanyan, N., Bar-Tal, O., Bagon, S and Dekel, T. (2022). Splicing vit features for semantic appearance transfer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10748-10757).

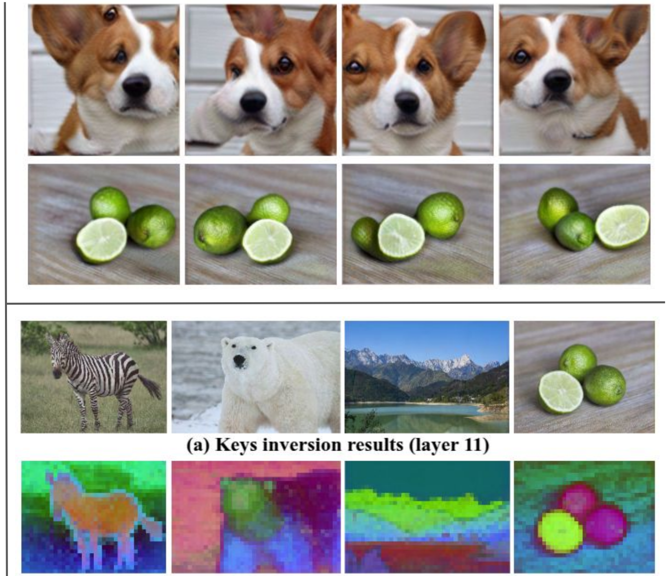
Feature Distribution Statistics As Loss Objective: *FDM Loss*

- Splicing ViT features for distilling appearance and semantics



(a) Original Images

(b) CLS token inversion results across layers



(a) Keys inversion results (layer 11)

(b) PCA of keys' self-similarity (3 leading components, layer 11)

† Tumanyan, N., Bar-Tal, O., Bagon, S and Dekel, T. (2022). Splicing vit features for semantic appearance transfer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10748-10757).

Feature Distribution Statistics As Loss Objective: *FDM Loss*

- **Research Question:** Do we have any dataset containing multi-illuminated scenes with corresponding ground truth illumination maps?

We have. Large scale Multi-illuminant Dataset (LSMI)†.

It comprises 7,486 meticulously annotated images captured in more than 2,700 diverse indoor and outdoor scenes, utilizing three different camera models: Samsung Galaxy Note 20 Ultra, Sony α9, and Nikon D810.

- Still lots of room for improvement for *FDM WB*
    - ~~Model complexity, too much?~~
    - EFDM, as a layer in architecture?
    - ~~Only working on single illuminant data rendered with pre-defined WB settings?~~
    - Mean-squared error, as the objective function?
    - ~~Better/efficient style representation, possible?~~

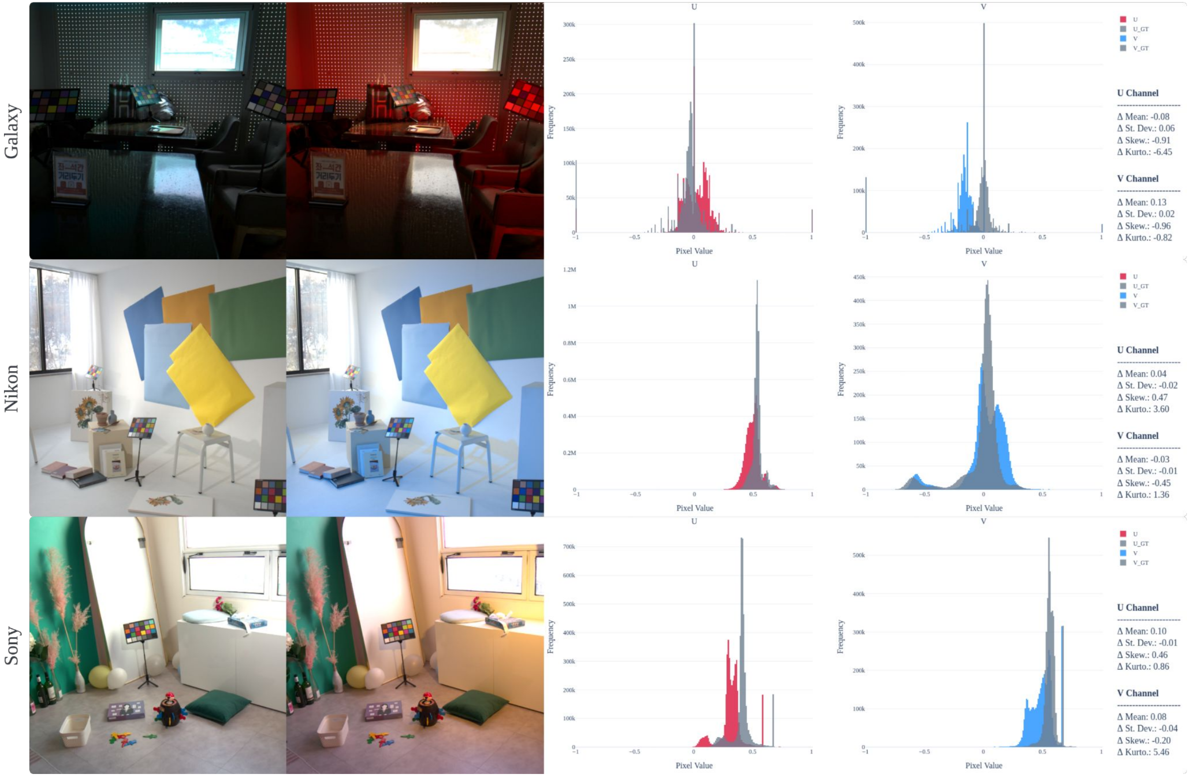
† Kim, D., *et al.* (2021). Large scale multi-illuminant (LSMI) dataset for developing white balance algorithm under mixed illumination. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10748-10757).

Feature Distribution Statistics As Loss Objective: *FDM Loss*

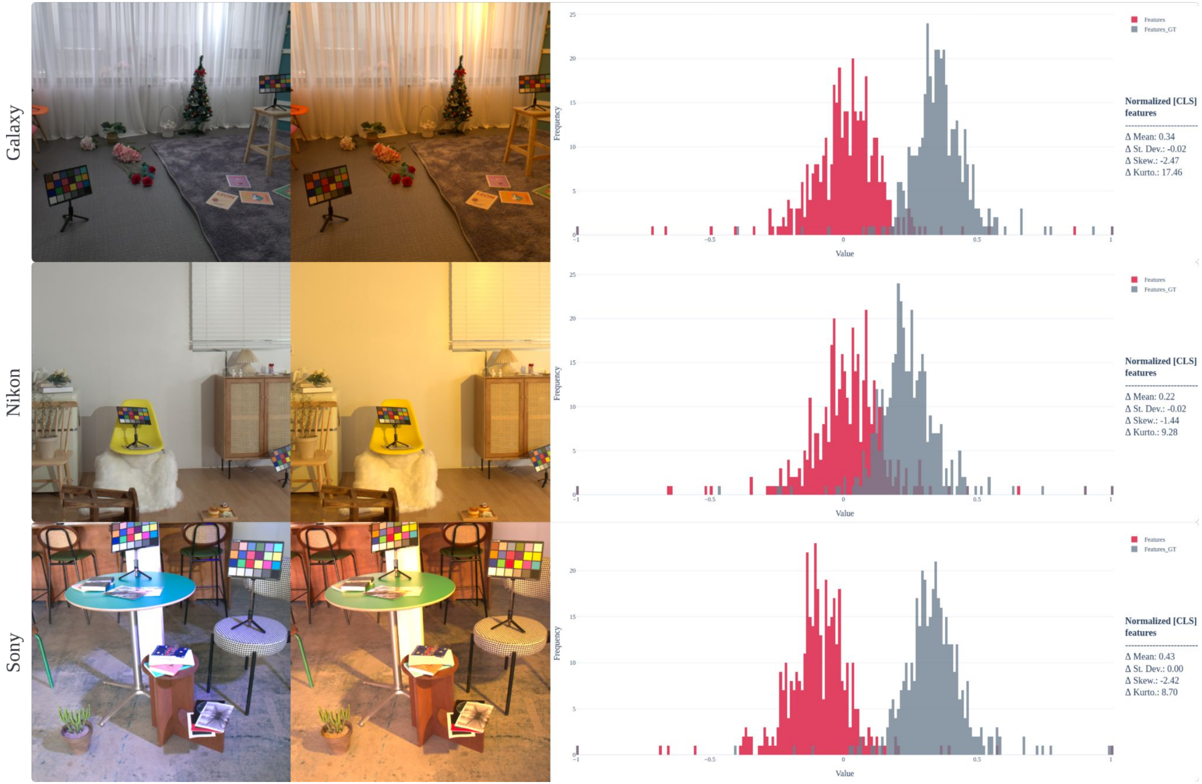
Single Illuminant Two Illuminants Three Illuminants

† Kim, D., *et al.* (2021). Large scale multi-illuminant (LSMI) dataset for developing white balance algorithm under mixed illumination. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10748-10757).

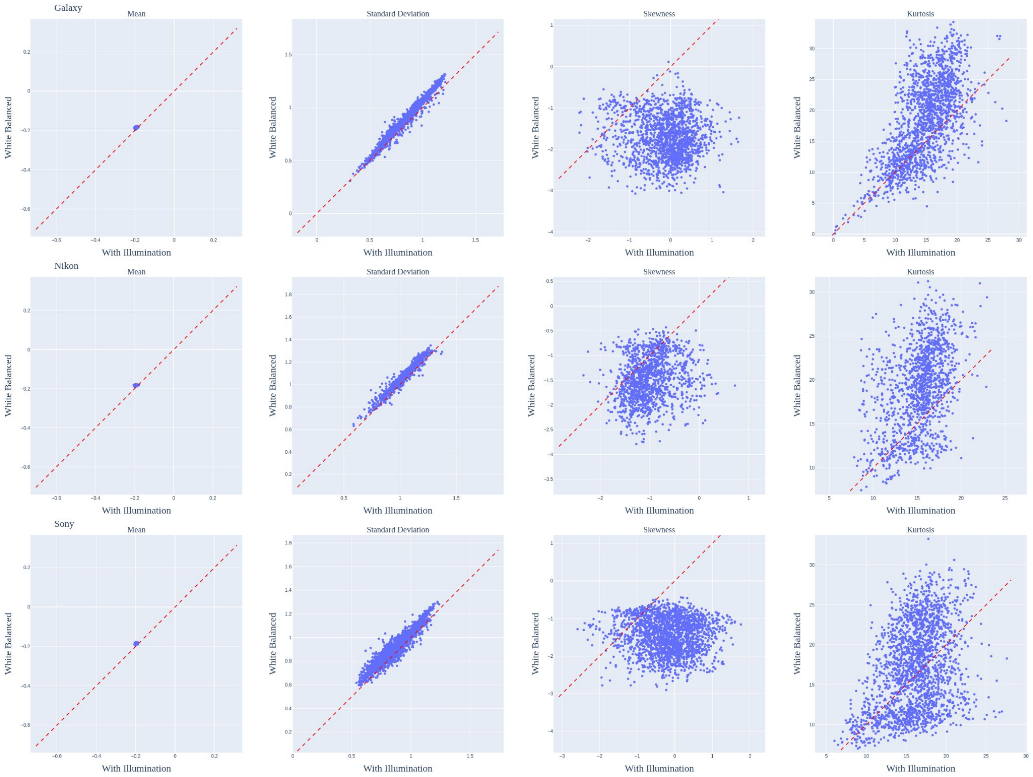
Feature Distribution Statistics As Loss Objective: *FDM Loss*



Feature Distribution Statistics As Loss Objective: *FDM Loss*



Feature Distribution Statistics As Loss Objective: *FDM Loss*



Feature Distribution Statistics As Loss Objective: *FDM Loss*

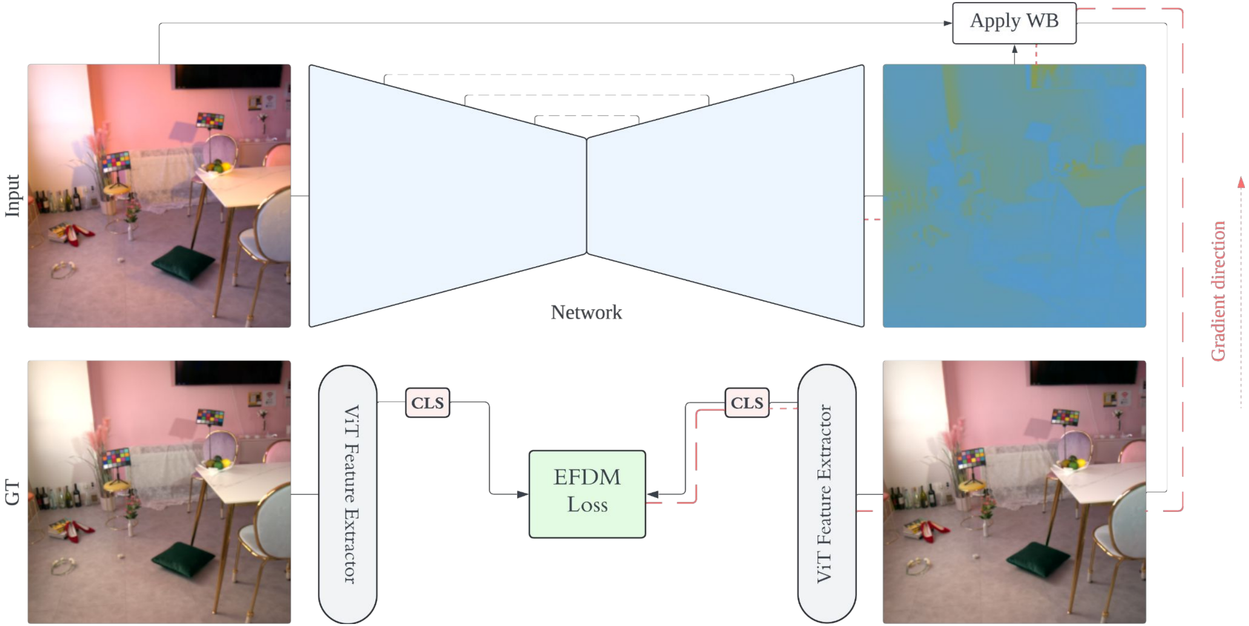
- **Research Question:** What happens if we use EFDM as the objective function, instead as a layer in architecture?

We can, it is just a function† that measures the distributional discrepancies and can be minimized for the optimization process.

- Still lots of room for improvement for *FDM WB*
    - ~~Model complexity, too much?~~
    - ~~EFDM, as a layer in architecture?~~
    - ~~Only working on single illuminant data rendered with pre-defined WB settings?~~
    - ~~Mean-squared error, as the objective function?~~
    - ~~Better/efficient style representation, possible?~~

† Zhang, Y., et al. (2022). Exact Feature Distribution Matching for Arbitrary Style Transfer and Domain Generalization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 8035-8045).

Feature Distribution Statistics As Loss Objective: *FDM Loss*†



$$L_{\text{EFDM}}(f_{\text{pred}}, f_{\text{gt}}) = \frac{1}{n}\sum_{i=1}^{n}[f_{\text{pred}}[i] - f_{\text{gt}}[\text{rank}(f_{\text{pred}}[i])]]^2$$

† Kınlı, F. and Kıraç, F. (2025). Feature distribution statistics as a loss objective for robust white balance correction. In Machine Vision and Applications, vol 36:58, 10.1007/s00138-025-01680-1.

Feature Distribution Statistics As Loss Objective: *FDM Loss*

- A Novel Metric for Generalization: *The Multi-to-Single Ratio (MSR)*[†]

$$\text{MSR} = \frac{\text{average MAE over multi-illuminant samples}}{\text{average MAE over single-illuminant samples}}$$

Evaluating the ability to generalize effectively to multi-illuminant conditions while avoiding overfitting to single-illuminant scenarios.

An MSR value closer to 1 indicates superior adaptability, while a higher/lower MSR *<generally higher>* reflects more performance degradation between single- and multi-illuminant conditions.

† Kınlı, F. and Kıraç, F. (2025). Feature distribution statistics as a loss objective for robust white balance correction. In Machine Vision and Applications, vol 36:58, 10.1007/s00138-025-01680-1.

Feature Distribution Statistics As Loss Objective: *FDM Loss*

Table 5.12: *Benchmark results on the LSMI dataset for the Galaxy camera. The Multi-to-Single Ratio reflects the robustness of the models in multi-illuminant scenarios.*

Model	Single		Multi		Mixed		MSR
	Mean	Median	Mean	Median	Mean	Median	
Pix2Pix [140]	6.53	2.17	4.28	2.63	5.66	2.44	**0.66**
Gijsenij *et al.* [79]	7.49	6.04	12.38	9.57	10.09	7.43	1.65
Bianco *et al.* [115]	4.15	3.30	5.56	4.33	4.89	3.83	1.34
HDRNet [149] r. [8]	2.85	2.20	3.13	2.70	3.06	2.54	1.10
HDRNet [149] r. [9]	-	-	-	-	3.06	2.54	-
UNet [57] r. [8]	2.95	1.86	2.35	2.00	2.63	1.91	0.80
UNet [57] r. [9]	2.85	-	2.55	-	2.68	2.17	0.90
One-Net [136]	**1.19**	0.75	2.16	1.53	**1.57**	0.93	*1.82*
AID [9]	**1.19**	-	2.03	-	1.66	1.41	*1.71*
Uformer + FDM (**ours**)	1.78	1.48	**1.87**	1.69	1.83	1.62	***1.05***

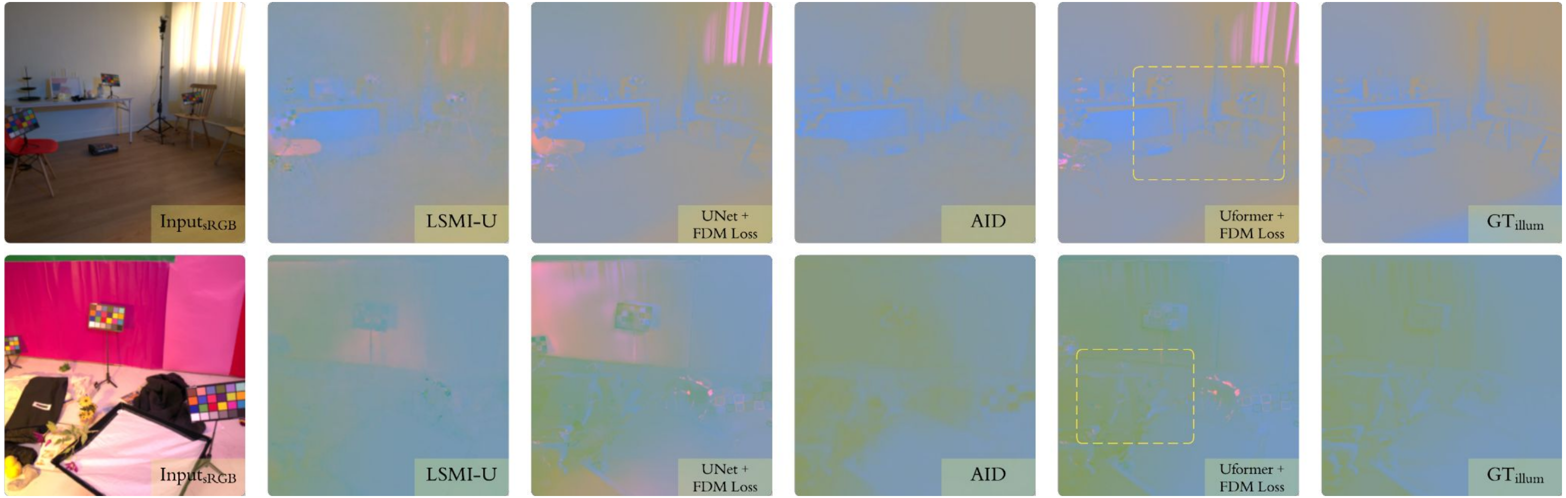
Table 5.13: *Benchmark results on the LSMI dataset for the Nikon camera. The Multi-to-Single Ratio reflects the robustness of the models in multi-illuminant scenarios.*

Model	Single		Multi		Mixed		MSR
	Mean	Median	Mean	Median	Mean	Median	
Pix2Pix [140]	6.1	2.27	4.18	2.76	5.41	2.49	**0.77**
Bianco *et al.* [115]	3.18	2.61	4.65	4.19	3.93	3.48	1.18
HDRNet [149] r. [8]	2.76	2.43	3.2	3.01	2.99	2.61	1.07
HDRNet [149] r. [9]	-	-	-	-	2.99	2.61	-
UNet [57] r. [8]	1.51	1.14	2.36	1.84	1.95	1.45	1.21
UNet [57] r. [9]	1.49	-	2.30	-	1.92	1.54	1.20
One-Net [136]	1.27	0.67	1.99	1.43	1.53	0.85	*1.30*
AID [9]	**1.11**	-	2.26	-	1.71	1.34	*1.32*
Uformer + FDM (**ours**)	1.26	0.97	**1.54**	1.13	**1.48**	1.05	***1.22***

Table 5.14: *Benchmark results on the LSMI dataset for the Sony camera. The Multi-to-Single Ratio reflects the robustness of the models in multi-illuminant scenarios.*

Model	Single		Multi		Mixed		MSR
	Mean	Median	Mean	Median	Mean	Median	
Pix2Pix [140]	4.08	1.72	4.37	3.26	4.20	2.20	**1.07**
Bianco *et al.* [115]	3.25	2.62	4.38	3.93	3.86	3.19	1.35
HDRNet [149] r. [8]	-	-	-	-	3.21	2.89	-
HDRNet [149] r. [9]	2.76	2.43	3.2	3.01	2.99	2.61	**1.07**
UNet [57] r. [8]	2.83	2.44	3.04	2.78	2.94	2.66	**1.07**
UNet [57] r. [9]	1.92	-	2.34	-	2.15	1.74	1.22
One-Net [136]	1.45	0.60	2.23	1.65	1.76	0.93	*1.54*
AID [9]	**1.01**	-	2.16	-	1.66	1.35	*2.14*
Uformer + FDM (**ours**)	1.52	1.39	**1.67**	1.57	**1.61**	1.47	***1.10***

Feature Distribution Statistics As Loss Objective: *FDM Loss*



- Superior adaptability to multi-illuminant conditions, particularly in fine scene details.

Feature Distribution Statistics As Loss Objective: *FDM Loss*

Table 5.15: *Ablation study on the proposed loss function using the Uformer architecture.*

Camera	Loss Function	Single Mean	Single Median	Multi Mean	Multi Median	Mixed Mean	Mixed Median	MSR
Galaxy	MSE	2.20	1.65	2.03	1.73	2.05	1.64	0.88
	FDM	**1.78**	1.48	**1.87**	1.69	**1.83**	1.62	1.05
Nikon	MSE	1.39	1.01	1.72	1.15	1.56	1.10	1.10
	FDM	**1.31**	0.98	**1.54**	1.12	**1.43**	1.05	1.08
Sony	MSE	2.15	1.54	2.03	1.73	2.08	1.68	0.94
	FDM	**1.52**	1.39	**1.67**	1.57	**1.61**	1.47	1.10

Table 5.16: *Ablation study on the proposed loss function using the UNet architecture.*

Camera	Loss Function	Single Mean	Single Median	Multi Mean	Multi Median	Mixed Mean	Mixed Median	MSR
Galaxy	MSE	2.95	1.86	2.35	2.00	2.63	1.91	0.80
	FDM	**2.42**	1.81	**2.14**	1.74	**2.27**	1.79	0.88
Nikon	MSE	1.51	1.14	2.36	1.84	1.95	1.45	1.21
	FDM	**1.40**	1.17	**1.89**	1.33	**1.66**	1.25	1.14
Sony	MSE	2.83	2.44	3.04	2.78	2.94	2.66	1.07
	FDM	**1.96**	1.63	**2.10**	1.74	**2.04**	1.67	1.07

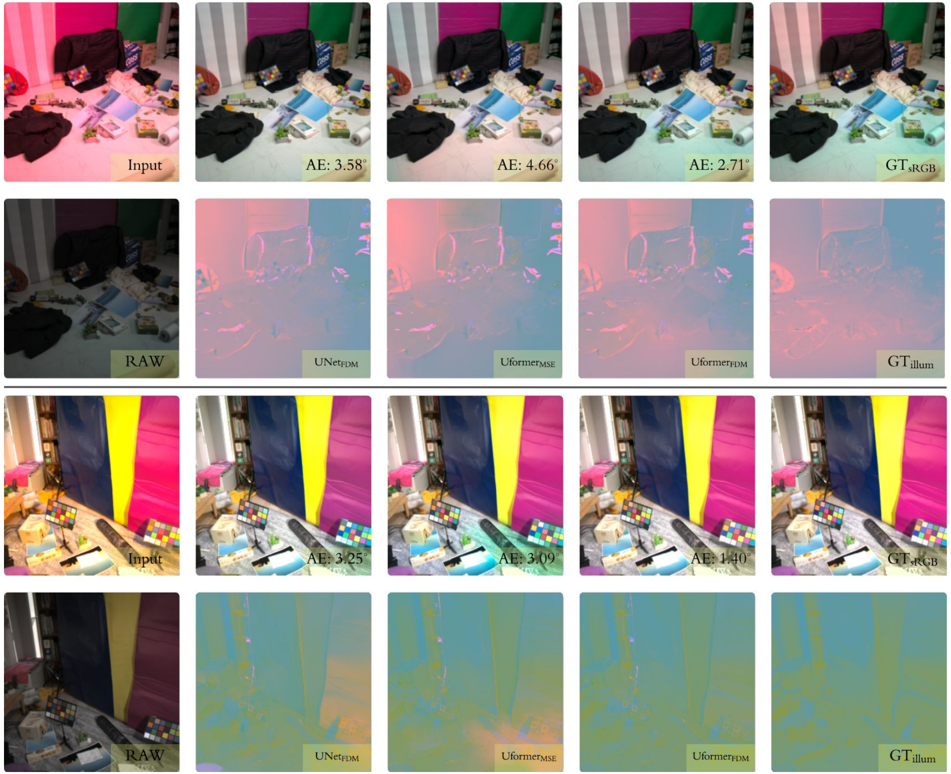
Galaxy

Feature Distribution Statistics As Loss Objective: *FDM Loss*

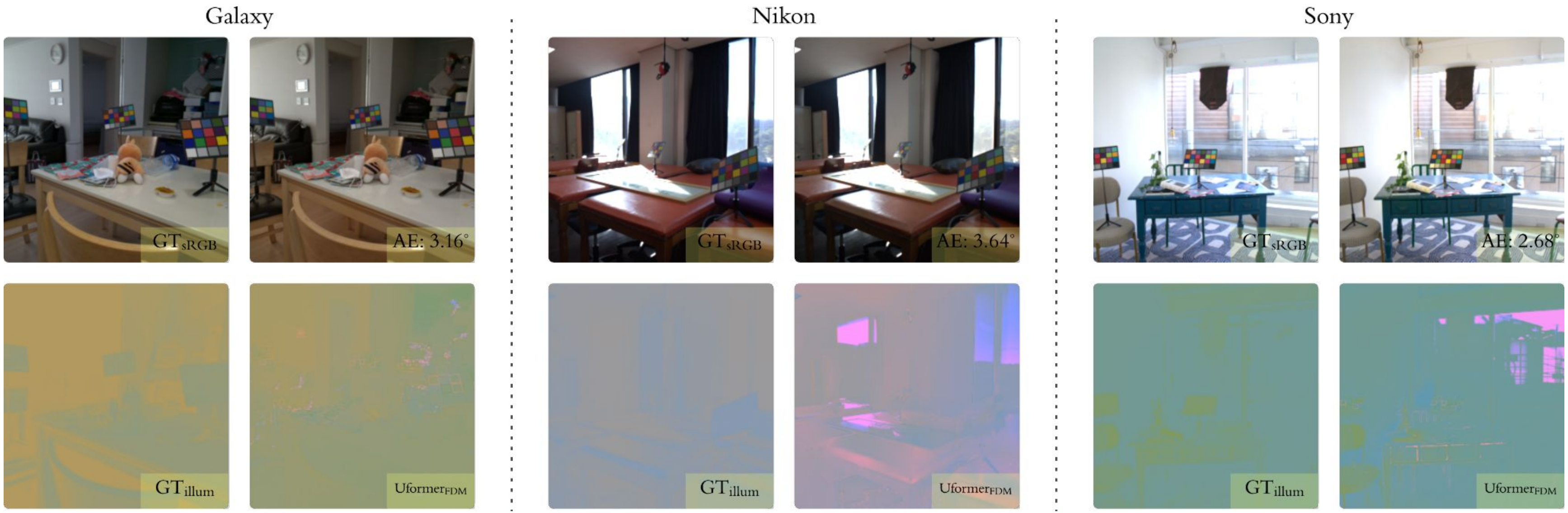
Nikon

Feature Distribution Statistics As Loss Objective: *FDM Loss*

Sony



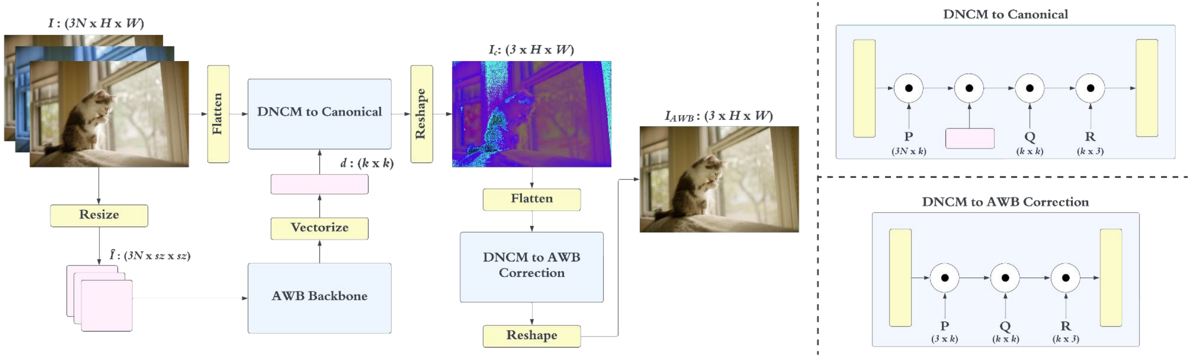
Feature Distribution Statistics As Loss Objective: *FDM Loss*



- Failure cases: strong directional illumination, highly reflective surface, strongly saturated regions

Applications & Extensions

- **De**terministic **N**eural **I**llumination **M**apping†



Method	MSE				MAE				ΔE 2000				Size
	Mean	Q1	Q2	Q3	Mean	Q1	Q2	Q3	Mean	Q1	Q2	Q3	
Mixed WB [2]													
$p = 64$, WB={t,d,s}	168.38	8.97	19.87	105.22	4.20°	1.39°	2.18°	5.54°	5.03	2.07	3.12	7.19	5.09 MB
$p = 64$, WB={t,f,d,c,s}	161.80	9.01	19.33	90.81	4.05°	1.40°	2.12°	4.88°	4.89	2.16	3.10	6.78	5.10 MB
$p = 128$, WB={t,f,d,c,s}	176.38	16.96	35.91	115.50	4.71°	2.10°	3.09°	5.92°	5.77	3.01	4.27	7.71	5.10 MB
Style WB [24]													
$p = 64$, WB={t,d,s}	92.65	**6.52**	**14.23**	35.01	2.47°	0.82°	1.44°	2.49°	2.99	**1.36**	2.04	3.32	61.0 MB
$p = 64$, WB={t,f,d,c,s}	151.38	29.49	56.35	125.33	4.18°	2.13°	3.03°	4.81°	5.42	3.11	4.42	6.76	61.1 MB
$p = 128$, WB={t,d,s}	88.03	7.92	17.73	45.01	2.61°	0.93°	1.58°	2.85°	3.24	1.50	2.30	3.95	61.2 MB
$p = 128$, WB={t,f,d,c,s}	100.24	10.77	37.74	70.18	3.09°	1.15°	2.61°	3.87°	3.96	1.59	3.55	5.51	61.3 MB
DeNIM + Mixed WB [2]													
$p = 64$, WB={t,d,s}	120.14	36.39	77.40	152.96	2.57°	1.53°	2.17°	3.19°	5.26	3.38	4.71	6.64	28.7 MB
$p = 64$, WB={t,f,d,c,s}	129.01	14.39	27.69	57.90	2.67°	0.99°	1.45°	2.29°	3.96	2.10	2.85	4.24	28.7 MB
$p = 128$, WB={t,d,s}	158.58	60.14	115.66	198.59	4.20°	2.38°	3.77°	5.63°	5.69	3.91	5.41	7.10	28.8 MB
$p = 128$, WB={t,f,d,c,s}	99.70	13.89	24.71	43.88	2.49°	1.07°	1.62°	2.41°	3.44	1.95	2.74	3.78	28.8 MB
DeNIM + Style WB [24]													
$p = 64$, WB={t,d,s}	**65.80**	10.06	16.98	**28.82**	2.03°	0.88°	1.23°	1.93°	2.95	1.79	2.33	3.18	196.3 MB
$p = 64$, WB={t,f,d,c,s}	83.41	13.23	21.46	37.44	**1.93°**	**0.77°**	**1.09°**	**1.70°**	**2.73**	1.62	**2.03**	**2.71**	196.3 MB
$p = 128$, WB={t,d,s}	80.53	17.59	27.80	44.35	2.16°	0.88°	1.34°	2.16°	3.08	1.86	2.37	3.30	196.4 MB
$p = 128$, WB={t,f,d,c,s}	89.10	11.27	19.34	43.01	2.49°	1.24°	1.64°	2.92°	3.16	1.87	2.53	3.35	196.4 MB

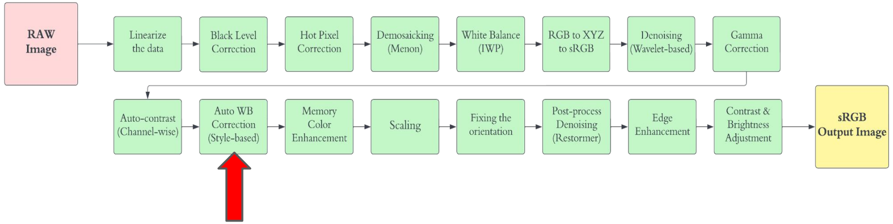
Table 6.2: *Comparison of the complexity of DeNIM and the prior methods with their post-processing tricks. ms: multi-scale weighting maps, eas: edge-aware smoothing.*

Model Architecture	Time (s)	Param (M)	FLOPS (G)
Mixed WB [2] + *ms* + *eas*	10.390	**1.32**	82.68
Mixed WB [2] + *ms*	0.228		
Mixed WB [2] + *eas*	10.279		
Mixed WB [2]	0.212		
Style WB [24] + *ms* + *eas*	10.342	15.31	76.80
Style WB [24] + *ms*	0.232		
Style WB [24] + *eas*	10.307		
Style WB [24]	0.217		
DeNIM + Mixed WB [2]	**0.006**	1.67	**2.14**
DeNIM + Style WB [24]	**0.010**	16.19	26.89

† Kınlı, F., Yılmaz, D., Özcan, B., and Kıraç, F. (2023). Deterministic Neural Illumination Mapping for Efficient Auto-White Balance Correction. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 1139-1147).
‡ Ke, Z. *et al.* (2023). Neural Preset for Color Style Transfer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 14173-14182).

Applications & Extensions

- Night Photography Challenge 23’†



Rank	Team	Mean Score
1	IVLTeam	0.670
2	DH_ImageAlgo	0.645
3	MiAlgo	0.626
4	BSSC	0.606
5	DH-AISP	0.583
6	Manual image enhancement	0.491
7	OzUVGL (**ours**)	0.453
8	The Majestic Mavericks	0.444
9	JMUCVLAB	0.439
10	NTU607	0.376
11	Baseline ISP	0.345

Mixed WB Style WB DeNIM + Style WB

† Shutova, A., *et al.* (2023). NTIRE 2023 challenge on night photography rendering. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (pp. 1982-1993).

Conclusion

- A style-based perspective for modeling illumination
- Proposed **three** complementary methods:
    - *Style WB*: Spatially-aware correction using style modulation
    - *FDM WB*: Better correction via EFDM
    - *FDM Loss*: A novel, efficient objective function for style-based WB correction
- Robust generalization performance to complex, multi-illuminant conditions
- New paradigm for image restoration <*removing injected style factor to restore*>
- Future?
    - exploring DeNIM for Transformer-based architecture
    - leveraging spatial priors
    - diffusion-based modeling approaches
    - flow matching & neural ODEs
    - architectural improvements for embedded & mobile deployments

Thank you!