

Supplementary Materials: Thinking Temporal Automatic White Balance: Datasets, Models and Benchmarks

Anonymous Authors

1 SUPPLEMENTARY EXPERIMENTS

1.1 Evaluation Metric Details

We assessed the automatic white balance (AWB) and temporal automatic white balance (TAWB) methods' performances in two dimensions: illumination estimation accuracy and stability.

For illumination estimation accuracy, we employed the established metrics - recovery angular error (AE) [3] - for evaluation, which computes angular error between predict illumination E and ground truth G :

$$AE(E, G) = \arccos\left(\frac{E \cdot G}{\|E\| \|G\|}\right), \quad (1)$$

where \cdot is the dot product, and $\|\cdot\|$ is the Euclidean norm. A lower value of AE means better illumination estimation accuracy.

To gauge the temporal stability of estimated illuminations, following [5], we utilized the maximum illumination change (MIC) and the illumination distribution scatter (STD).

MIC identifies abrupt illumination shifts in a sequence by quantifying the largest change in the estimated illuminations between consecutive frames:

$$MIC(S) = \max(AE(E_{S_i}, E_{S_{i+1}})), I = 1 \dots N_S - 1, \quad (2)$$

where E_{S_i} and $E_{S_{i+1}}$ represent the estimated illuminations of i^{th} and $(i + 1)^{th}$ frames within the sequence S , and N_S denotes the frame numbers of S .

To identify sequences spanning a large range of illuminations, we employed STD for scatteredness. In detail, the estimated illuminations of the sequence S are first converted into an Angle-Retaining Chromaticity diagram [4], a bi-dimensional representation where Euclidean distances correspond to angular distances in Eq.1. Then, STD is computed:

$$STD(S) = \sqrt{\sum_{i=1}^{N_S} \frac{(x_{S_i} - \bar{x}_S)^2}{N_S} + \sum_{i=1}^{N_S} \frac{(y_{S_i} - \bar{y}_S)^2}{N_S}}, \quad (3)$$

where (x_{S_i}, y_{S_i}) are the ARC coordinates of the i^{th} estimated illuminations in sequence S , and (\bar{x}_S, \bar{y}_S) represents the average of each coordinate for the sequence. The lower values of MIC and STD mean better illumination estimation stability.

1.2 Additional Training Details

We trained and tested all compared baseline models on 3090ti. For all compared methods, we followed the optimal settings in their papers [12, 8, 2, 11, 9, 10]. Throughout the training and testing phases of all methods, the color checker is masked out in each frame to avoid bias.

Unlike AWB methods, RCCNet, TCCNet and our CTANet estimated a single illumination color vector from consecutive frames. This approach can introduce inconsistency in experiments when the number of input frames varies. To maintain experimental consistency, we chose two neighboring frames, along with the target

frame, as inputs for RCCNet, TCCNet and our CTANet. *The trained models are in the Models fold of the supplementary material.*

1.3 Validation of Spatial-temporal Attention

Spatial-temporal (ST) attention is the key element of our CTANet, which aims to extract the temporal features that can represent the shared information of input frames. Directly, this will facilitate stable estimation of illumination in our CTANet. Ablation experiments of spatial-temporal (ST) attention are in Table.3 and Table.4 of the main text. Here, we additionally provided the visualizations of the temporal features highlighted by ST attention to validate its effect.

As in Fig.1(a), when the input frames contain similar content (two art paintings) but the illumination changes, to achieve a stable effect, the model should extract the features representing the shared content to estimate the illumination color of the target frame. In Fig.1(b), our ST attention achieves the above goal by identifying similar spatial features from adjacent frames. Specifically, when the spatial features of the target frame represent art printings located in the center and left, our ST attention recognizes similar information in adjacent frames (first and second lines). Even when the features of the target frame do not represent the information very clearly, our ST attention highlights the shared information in different frames through inter-frame similarity (third row), thus ensuring the effectiveness and stability of target illumination estimation. By the above mechanism, in Fig.1(c), the target illumination color estimated by our CTANet is extremely similar to the real illumination color (Ground Truth, GT), and the AE error between the corrected target frame and the real target frame (GT) is only 0.437.

1.4 Comparison of Visualization Effect Under Different Illuminations

Fig.2 presents qualitative comparisons between RCCNet [10], TCCNet [9], and our CTANet on the sequences with multiple and dark illuminations (on page 3). Under artificial dark illuminations such as Seq.1 and Seq.2, sequences corrected by RCCNet and TCCNet tend to have a cooler tone, whereas CTANet produces outputs that are more aesthetically pleasing. For artificial multiple illuminations (e.g., Seq.3 and Seq.4), RCCNet and TCCNet show a noticeable color bias towards red or blue tones, whereas CTANet introduces more natural tones. CTANet also performs well in handling natural dark and multiple illuminations, demonstrating improved single-frame color correction and temporal color consistency as seen in Seq.5 and Seq.6. Additionally, in scenes with dynamic illumination changes between frames (e.g., Seq.7 and Seq.8), frames corrected by RCCNet and TCCNet exhibit less temporal consistency compared to CTANet's color-corrected frames, which show smoother continuity. *For more details, please refer to the Visualization Comparison folder of the supplementary material.*

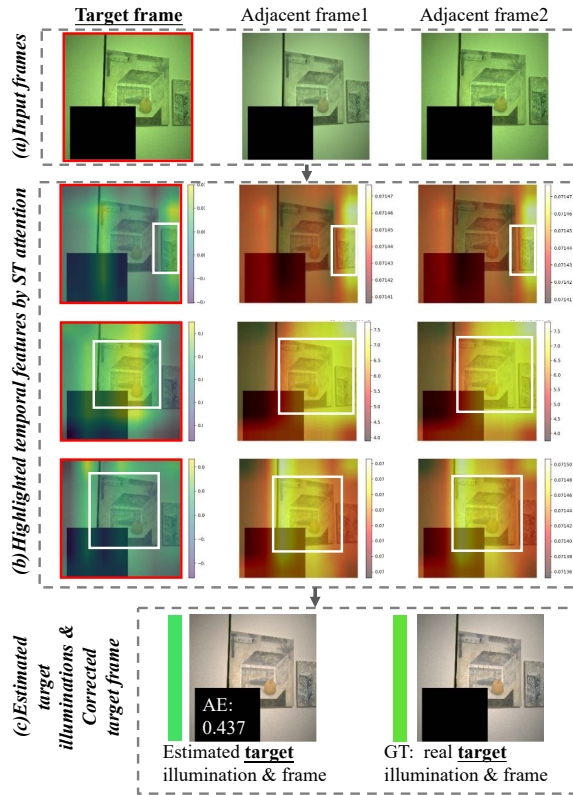


Figure 1: Visualizations of the temporal features highlighted by spatial-temporal (ST) attention in our CTANet (GT: Ground Truth).

2 SUPPLEMENTARY INFORMATION OF CTA DATASET

2.1 Reference White of Mobile Cameras

According to the Imaging Model[7], the illumination color labels (ground truths) are specific to cameras. Thus, directly comparing illumination color distributions is not practical. To address this, we employed the reference whites to transform diverse camera-specific illuminations into the camera-independent color space, i.e. Angle-Retaining Chromaticity diagram[4], facilitating a meaningful comparison of illumination distributions (Fig.2(c) of the main text).

Referring to [6], we opted for direct sunlight as our reference white. This choice is based on its simplicity in identifying images taken under these conditions. Adhering to this definition, we carefully selected the most representative image for each mobile camera and computed the reference white for our CTA dataset. The details are in Table. 1.

2.2 Dataset Processing

We positioned the ColorChecker in all frames to record the spatial and temporal illumination colors. To obtain the illumination color of a single frame, we manually computed the denoised RGB response from grey patches #20, #21, #22, #23 in the ColorChecker. We excluded the brightest grey patch #19, the darkest grey patch

Mobile Camera	R/G	B/G
Huawei P30 Pro	0.3545	0.4507
Huawei Mate30	0.3323	0.4292
iPhone 14 Pro Max	0.9785	1.0024
Xiaomi 11 Pro	1.2308	0.7691
Xiaomi 13	0.4962	0.5589
Vivo iqoo neo5	0.5599	0.5219

Table 1: Reference white (D55-like) for the different mobile cameras of our CTA dataset.

#24, and any additional saturated patches [1]. After capturing, we processed all sequences to obtain pure raw versions with Dcraw, including demosaic, black-level subtraction, and saturation point stretching. *The corresponding codes are in the Data Processing folder of the supplementary material.*

2.3 Dataset Statistics

Sec.3.1 of the main text detailed the methodology for capturing various natural and artificial illuminations. Natural illumination varies with weather conditions and time of day, while artificial illumination depends on the scene type. These factors are illustrated in Fig.5 of the main text. Comprehensive statistics regarding these conditions are provided in Table. 2 and Table. 3 on page 3, documenting the diversity of illumination environments captured. *Dataset is in the Dataset fold of the supplementary material.*

REFERENCES

- [1] Caglar Aytekin, Jarno Nikkanen, and Moncef Gabbouj. "INTEL-TUT dataset for camera invariant color constancy research". In: *arXiv preprint arXiv:1703.09778* (2017).
- [2] Simone Bianco and Claudio Cusano. "Quasi-supervised color constancy". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 12212–12221.
- [3] Marco Buzzelli, Simone Bianco, and Raimondo Schettini. "ARC: Angle-Retaining Chromaticity diagram for color constancy error analysis". In: *J. Opt. Soc. Am. A* 37.11 (Nov. 2020), pp. 1721–1730. doi: 10.1364/JOSAA.398692. URL: <https://opg.optica.org/josaa/abstract.cfm?URI=josaa-37-11-1721>.
- [4] Marco Buzzelli, Simone Bianco, and Raimondo Schettini. "ARC: Angle-Retaining Chromaticity diagram for color constancy error analysis". In: *J. Opt. Soc. Am. A* 37.11 (Nov. 2020), pp. 1721–1730. doi: 10.1364/JOSAA.398692. URL: <https://opg.optica.org/josaa/abstract.cfm?URI=josaa-37-11-1721>.
- [5] Marco Buzzelli and Ilaria Erba. "On the evaluation of temporal and spatial stability of color constancy algorithms". In: *J. Opt. Soc. Am. A* 38.9 (Sept. 2021), pp. 1349–1356. doi: 10.1364/JOSAA.434860. URL: <https://opg.optica.org/josaa/abstract.cfm?URI=josaa-38-9-1349>.
- [6] Marco Buzzelli et al. "Analysis of biases in automatic white balance datasets and methods". In: *Color Research & Application* 48.1 (2023), pp. 40–62. doi: <https://doi.org/10.1002/col.22822>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/col.22822>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/col.22822>.
- [7] Arjan Gijsenij and Theo Gevers. "Color constancy using natural image statistics and scene semantics". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33.4 (2010), pp. 687–698.
- [8] Yuanming Hu, Baoyuan Wang, and Stephen Lin. "FC4: Fully Convolutional Color Constancy with Confidence-Weighted Pooling". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017, pp. 330–339. doi: 10.1109/CVPR.2017.43.
- [9] Yanlin Qian et al. *A Benchmark for Temporal Color Constancy*. 2020. arXiv: 2003.03763 [cs.CV].
- [10] Yanlin Qian et al. "Recurrent color constancy". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 5458–5466.
- [11] Yuxiang Tang et al. "Transfer learning for color constancy via statistic perspective". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 36. 2. 2022, pp. 2361–2369.
- [12] Joost Van De Weijer, Theo Gevers, and Arjan Gijsenij. "Edge-based color constancy". In: *IEEE Transactions on image processing* 16.9 (2007), pp. 2207–2214.

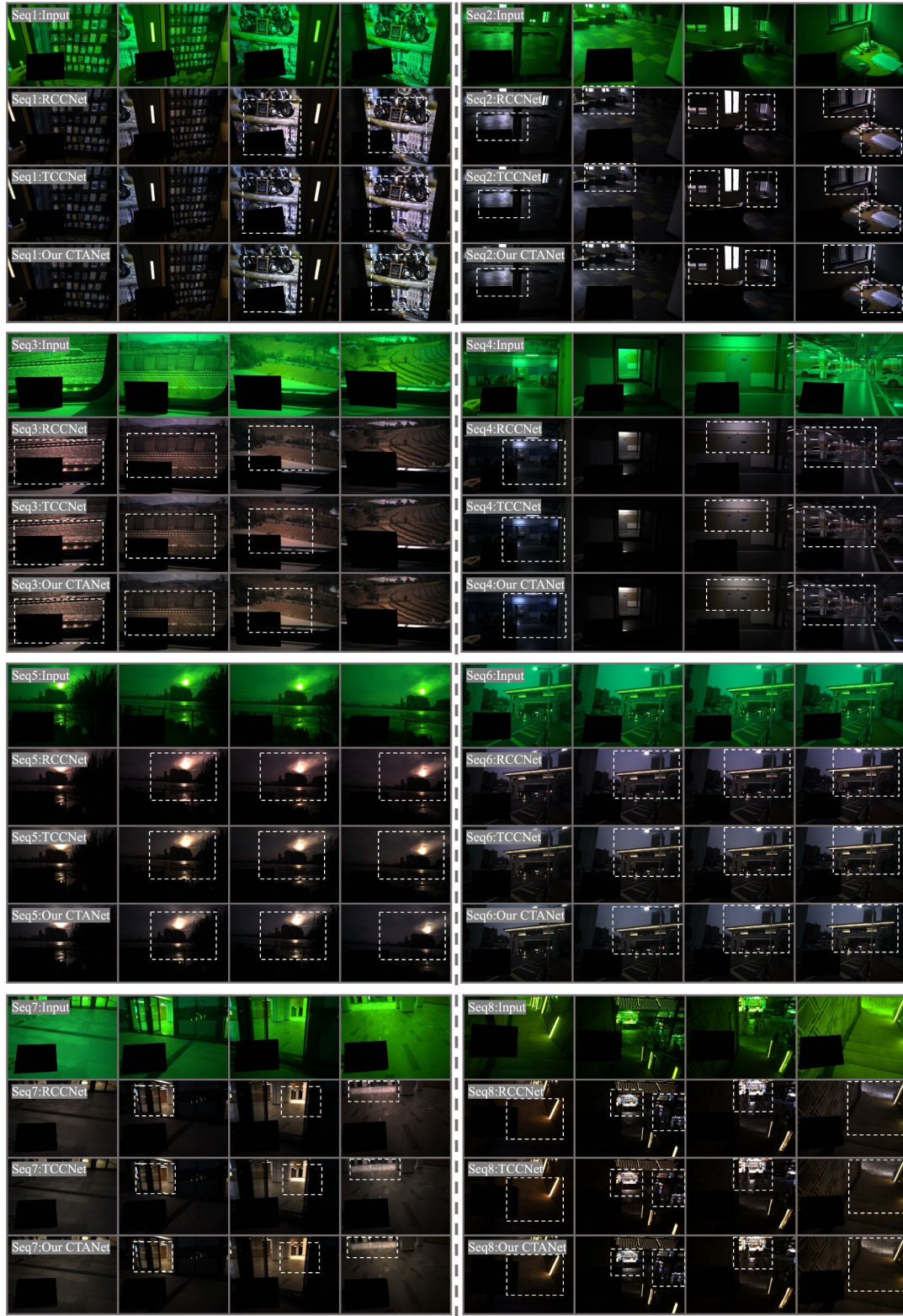


Figure 2: Visual comparisons of Our CTANet and state-of-the-art TAWB methods for the sequences with complex illuminations.

Scene Type	Location	Object	Count
Indoor	Room	browndoor	6
		orangegreysofa	6
		whitefurniture	8
		whitewall	6
		woodenfloor	11
	Staircase	stairs	10
		whitewall	6
	Undergroundparking	car	5
		greenwall	4
	Corridor	stair	7
		elevator	4
		greyfloor	5
		whitewall	6
	Lobby	greyfloor	4
		blackboard	6
	Classroom	greyfloor	5
		woodendesk	11
		woodenseat	8
	Train	greytrunk	1
		book	13
	Library	greydoor	6
		woodendesk	4
	Supermarket	painting	14
	Artgallery	painting	14
	Market	vegetables	1
	Parkinglot	greyfloor	4
		whitewall	4
	Mall	colorfulclothes	2
		floortile	3
Outdoor	City	building	114
		car	37
		greenery	108
		greyfloor	15
		lawn	17
		people	18
		street	116
		trail	32
		tree	138
	Wild	greenery	23
		greyfloor	11
		lake	6
		mountain	7
		redterrace	8
		stonebridge	5
		trail	28
		tree	43
In&Outdoor	-	brickwall	5
		building	9
		elevator	5
		greenery	9
		greyfloor	21
		stair	7
		street	9
		whitewall	5

Table 3: Statistics of scenes with locations and objects (Count: the number of captured frames).

Illumination Type	Weather	Time of Day	CT	Count
Natural illumination (Outdoor)	Cloudy	p.m.	cool	17
			neutral	30
			warm	31
			dynamic	26
		a.m.	neutral	1
	Rainy	p.m.	warm	4
			dynamic	5
			neutral	7
			neutral	5
	Sunny	p.m.	cool	4
			neutral	48
			warm	31
			dynamic	61
		a.m.	cool	5
			neutral	18
			warm	2
			dynamic	10
Dynamic illumination (In&outdoor)	Cloudy	p.m.	cool	2
			neutral	8
			dynamic	7
		a.m.	neutral	1
		a.m.	dynamic	3
	Rainy	p.m.	neutral	1
			warm	1
			dynamic	1
			neutral	5
	Sunny	p.m.	warm	3
			dynamic	15
			neutral	2
		a.m.	dynamic	1
Artificial illumination (Indoor)	Room	-	cool	22
			neutral	4
			warm	17
			dynamic	16
	Artgallery	-	cool	10
			neutral	4
			dynamic	1
	Classroom	-	neutral	11
			warm	1
			dynamic	7
	Corridor	-	cool	4
			neutral	7
			warm	4
	Library	-	cool	4
			neutral	7
			warm	4
	Lobby	-	dynamic	4
			cool	1
			neutral	3
	Mall	-	warm	2
			dynamic	3
			cool	7
	Market	-	cool	3
			neutral	2
			dynamic	2
	Parkinglot	-	neutral	1
			cool	1
			dynamic	1
	Staircase	-	neutral	2
			cool	1
			dynamic	1
	Supermarket	-	cool	2
			neutral	2
			dynamic	1
	Train	-	cool	1
			neutral	4
			warm	2
	Undergroundparking	-	dynamic	3

Table 2: Statistics of illumination with scene, weather, time and color temperature (CT) (Count: the number of captured frames).