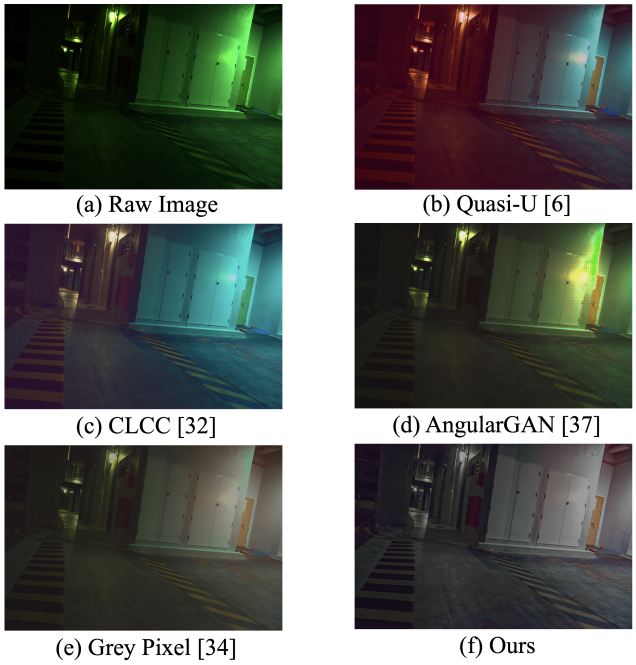
NightCC: Nighttime Color Constancy via Adaptive Channel Masking

Shuwei Li
National University of Singapore
shuwei@u.nus.edu

Robby T. Tan
ASUS Intelligent Cloud Services
National University of Singapore
robby.tan@nus.edu.sg

Abstract

Nighttime conditions pose a significant challenge to color constancy due to the diversity of lighting conditions and the presence of substantial low-light noise. Existing color constancy methods struggle with nighttime scenes, frequently leading to imprecise light color estimations. To tackle nighttime color constancy, we propose a novel unsupervised domain adaptation approach that utilizes labeled daytime data to facilitate learning on unlabeled nighttime images. To specifically address the unique lighting conditions of nighttime and ensure the robustness of pseudo labels, we propose adaptive channel masking and light uncertainty. By selectively masking channels that are less sensitive to lighting conditions, adaptive channel masking directs the model to progressively focus on features less affected by variations in light colors and noise. Additionally, our model leverages light uncertainty to provide a pixel-wise uncertainty estimation regarding light color prediction, which helps avoid learning from incorrect labels. Our model demonstrates a significant improvement in accuracy, achieving 21.5% lower Mean Angular Error (MAE) compared to the state-of-the-art method on our nighttime dataset.



(a) Raw Image (b) Quasi-U [6] (c) CLCC [32] (d) AngularGAN [37] (e) Grey Pixel [34] (f) Ours

Figure 1. The comparison between our method and state-of-the-art methods on a nighttime outdoor image with complex lighting conditions. Unlike both supervised and unsupervised methods, our model is more effective in removing the majority of color casts.

1. Introduction

Most existing color constancy methods are designed under the assumption of well-lit conditions, a scenario that contrasts with typical nighttime settings. The differences arise from the more complex lighting conditions and significant low-light noise present in nighttime images. Consequently, when applied to nighttime scenes, existing methods yield suboptimal results, as shown in Fig. 1.

Traditional approaches, ranging from low-level statistical analysis (e.g., [10, 27, 40]) to physics-based methods (e.g., [14, 15, 39, 43]), are mostly tailored for single light color settings. In common nighttime scenarios with multiple light colors, they tend to fail to estimate all the light colors. While some methods can handle multiple light colors [8, 9, 17, 34], they are constrained by the necessity to assume the number of light colors, compromising their performances.

Deep-learning-based methods (e.g., [1, 6, 19, 32, 37]) achieve superior performance on well-lit images by relying on training with labeled data. However, existing datasets [3, 11, 23, 26] are primarily collected under single-light daytime or multi-light indoor conditions. In contrast, nighttime scenes present a more uneven distribution of light colors and significant low-light noise. This discrepancy creates a significant domain gap, leading to inaccurate light color estimations in nighttime conditions. A straightforward approach to address this issue would be to create a

labeled nighttime dataset. Yet, this task is particularly challenging, requiring pixel-level ground-truths.

In this paper, we introduce a novel unsupervised domain adaptation method that leverages labeled daytime to learn on unlabeled nighttime data. We employ the mean teacher model [38] to generate pseudo labels to guide the student to learn about nighttime data. To overcome the nighttime unique lighting conditions and improve the robustness of the pseudo labels, we propose adaptive channel masking and light uncertainty.

Adaptive channel masking is designed to reduce the sensitivity of the model's feature channels to nighttime main attributes: light colors and noise. The insensitive channels are identified through our night augmentation and masked out in the student model during training on night data. This masking compels the model to progressively focus on learning features that are less sensitive to light colors and noise.

The light uncertainty mechanism aims to ensure that our model does not learn from incorrect pseudo-labels. It involves feeding the predicted white-balanced image to the model and measuring the deviation between the estimated light color of this image and white light. This approach enables our model to compute pixel-level uncertainty of light color estimation for our pseudo-labels. To train and evaluate our model, we have collected a new nighttime dataset containing 720 unlabeled images and 64 sparsely annotated images. In summary, our contributions are:

- To the best of our knowledge, our work is the first effort to tackle nighttime color constancy. We introduce a novel method based on domain adaptation mean teacher that effectively learns from unlabeled nighttime images.
- A new adaptive channel masking scheme is introduced to enhance the model's generalization to nighttime lighting conditions. By selectively masking feature channels, our model is encouraged to be less sensitive to the presence of various light colors and noise in estimating surface colors.
- To avoid noisy labels to be learned, we design a novel light uncertainty mechanism. Such uncertainty allows for a pixel-level estimation of the confidence in pseudo labels to avoid incorrect labels being learned.

Our model demonstrates a significant improvement in accuracy, achieving 21.5% lower Mean Angular Error (MAE) compared to the state-of-the-art method on our nighttime dataset. We further demonstrate the effectiveness of our approach when applied to indoor multi-illuminant datasets.

2. Related Work

Color constancy, commonly known as white balancing, is essential in digital imaging for eliminating color casts caused by scene illumination. Its primary challenge lies in estimating the illumination of an image, broadly categorized into two types: uniform color constancy, assuming a single illumination color, and multi-illuminant color constancy, considering the presence of multiple illumination colors.

Historically, the focus has predominantly been on single light color constancy. The classic Gray World algorithm [9] operates on the assumption that the average of each color channel independently reflects the illuminant color signal. Expanding upon this, the Grey Edge [40] framework was developed to unify a range of unsupervised methods, incorporating higher-order statistics and derivatives for enhanced performance. These foundational assumptions have paved the way for more advanced statistics-based methods like White Patch [27].

Recently, data-driven approaches, especially neural networks, have superseded statistical methods. In uniform color constancy, CNN-based solutions (e.g., [19, 44, 45]) have emerged, with CLCC [32] currently leading through the introduction of global contrastive learning. These methods are trained on the collected daytime/indoor datasets [3, 11, 26]. Studies like MSCC [20] and Quasi-U [20] have explored learning from unlabeled data in single-illuminant scenes using visual cues and self-supervised training, respectively. Nonetheless, these methods fall short in complex nighttime conditions. Due to the assumption of uniform light color, these single light color methods cannot fully remove the color cast when multiple illuminant colors exist, which is common in nighttime scenarios.

Early attempts to address both uniform and multi-illuminant problems, such as [5, 8, 9, 17], relied on statistical methods based on empirical priors. However, these methods often faced challenges in accurately setting parameters for general scenes. Hussain et al. [21] introduced an approach that segments the input image into multiple parts using the K-means algorithm, aiming to enhance color adjustment accuracy. Another method, Gray Pixels [34], was developed for handling multiple illuminants, but it tends to falter in scenarios where it's difficult to determine whether a scene is single-illuminant or not. Moreover, Gray Index [35] employs the Dichromatic Reflection model and ranks pixels based on their grayness level. The N-white balancing [2], introduced by Akazawa et al., suggests adjusting source white points to align with corresponding ground truth values. In multi-illuminant color constancy, the scarcity of extensive datasets has resulted in fewer learning-based methods. Recently, translation-base methods [29, 37] are proposed to directly predict the surface color. These methods can be trained on a recent large-scale indoor dataset LSMI [23].

Nighttime scenarios have received special attention in low-level vision research [30, 31, 41] due to their distinct illumination conditions. When facing nighttime scenarios, existing single-illuminant methods often falter due to their assumption of uniform light color, which is typically not the case at nighttime. While learning-based multi-

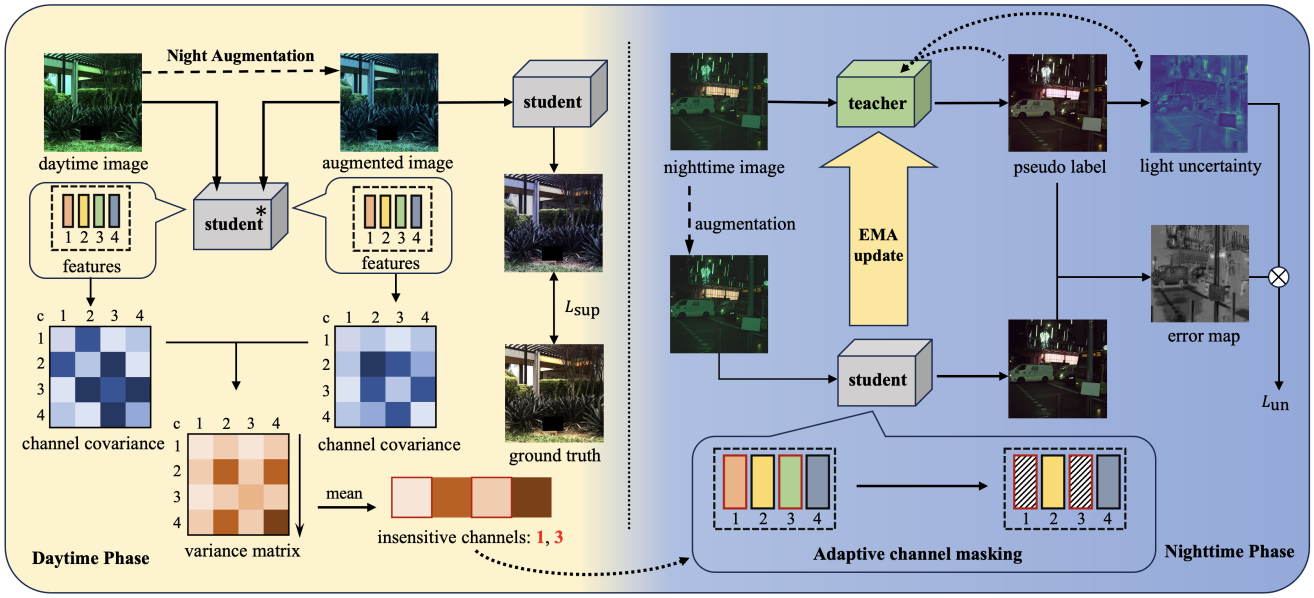


Figure 2. Overview of our method. In the daytime phase, the student model learns from labeled images. By comparing the feature covariance matrices between the original image and Night Augmented image, a variance matrix can be calculated, identifying the channels insensitive to the light color and noise. During the nighttime phase, the teacher model generates pseudo labels on the nighttime images. The student model applies Adaptive Channel Masking based on the insensitive channels determined on daytime. The error map is weighted by the Light Uncertainty map. The teacher model's parameters are updated by Exponential Moving Average (EMA).

illuminant methods could be applied due to their generally superior performance, they tend to yield suboptimal results when faced with the substantial domain gap between daytime/indoor datasets and the distinct conditions of nighttime scenarios.

3. Proposed Method

Our goal is to estimate the surface color y_{N} from a raw image x_{N}, under complex nighttime lighting conditions. To tackle this, we propose a semi-supervised learning approach based on the mean teacher [38]. Our method involves learning from (1) daytime images X_{D} and corresponding surface-color ground truth Y_{D}; (2) unlabeled nighttime images X_{N}.

3.1. Unsupervised Domain Adaptation Training

Our unsupervised domain adaptation framework is built on mean teacher [38], as shown in Fig. 2. To be specific, the framework includes two neural networks, named the teacher and the student, which are structurally identical U-Net [36] but differ in two key areas: how their weights are updated and how the input data is augmented.

Weights Update The teacher's weights θ_t are updated by the exponential moving average (EMA) of the student's weights θ_s:

$$\theta_t = \epsilon\theta_t + (1-\epsilon)\theta_s, \tag{1}$$

where $\epsilon \in (0, 1)$ is the momentum. This updating strategy allows the teacher model to integrate the student's latest learnings after each training step. Research has shown that this approach of averaging weights over time can make the training process more stable and effective compared to traditional gradient descent [33]. The weights of student model θ_s are optimized by gradient descent. This typically involves minimizing a loss function:

$$L_{\text{total}} = L_{\text{sup}} + L_{\text{unsup}}, \tag{2}$$

where L_{sup} denotes the supervised loss and L_{unsup} denotes the unsupervised teacher-student consistency loss.

Augmentation Nighttime images suffer from complications that are not present in daytime scenes. To be specific, nighttime scenarios can have more than one light source with different colors, drastically different from the more uniform and natural lighting of daytime. These light sources are also artificial, introducing diverse illuminant color distribution compared to daytime images. Moreover, nighttime images often suffer from higher levels of noise, especially in low-light areas.

Keeping in mind the characteristics of nighttime images, we propose Night Augmentation, aiming to replicate the nighttime lighting challenges on the daytime image. This involves creating multi-light conditions to simulate various artificial lighting scenarios and adding sensor read noise to mimic the noise in a raw image. For the light sources, nighttime light sources can be categorized into: (1) Planckian sources (e.g., candles) that follow the Planckian locus on the chromaticity diagram; (2) non-Planckian sources (e.g., LEDs) that can have a large range of colors. We simulate these lighting conditions using Gaussian distributions of random Planckian and non-Planckian colors, applied to daytime ground truth images via an inverse von Kries transform [25]. Additionally, we model sensor read noise following [42] – comprising thermal noise, source follower noise, and banding pattern noise – using a Tukey lambda distribution [22]:

$$N_{\text{read}} \sim \text{TL}\left(\lambda; 0, \sigma_{\text{TL}}\right), \tag{3}$$

where λ and σ represent the shape and scale parameters.

In the daytime phase, the full Night Augmentation is applied to daytime images. With our Night Augmentation, we can synthesize the challenges of nighttime on a daytime image and estimate the sensitivity of different channels in our adaptive channel masking scheme. For the input of the student model on the nighttime phase, only the sensor read noise is applied.

3.2. Adaptive Channel Masking Scheme

The Adaptive Channel Masking Scheme plays a crucial role in enhancing our model's robustness under challenging nighttime lighting conditions. This approach is built on the principle that the output of our model is only surface color related, so the extracted features should not be sensitive to the nighttime lighting challenges: complex lighting conditions and noise. Traditionally, models tend to not manage the sensitivity to the light color and noise, leading to suboptimal performance when facing nighttime scenarios. Our scheme identifies and prioritizes channels that have relatively low response to lighting conditions and noise. By selectively masking out channels less sensitive to the night augmentation during training, the model is guided to focus on identifying features that are less relevant to the light color and noise, such as surface color information.

To be specific, during daytime training, for every batch, the input images x_i and their Night Augmented versions $\tilde{x_i}$ are fed into the model. The resulting feature maps can be denoted as W_i and $\tilde{W}_i \in \mathbb{R}^{C\times HW}$, allowing us to calculate the feature covariance maps [12] Σ_i and $\tilde{\Sigma}_i$ as:

$$\boldsymbol{\Sigma_i} = \frac{1}{HW} W_i {W_i}^\top \in \mathbb{R}^{C\times C}. \tag{4}$$

Afterward, we derive a variance matrix from the differences between the covariance matrices Σ_i and $\tilde{\Sigma}_i$. The variance matrix can be defined as:

$$\mathbf{V} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\sigma}_i^2 \in \mathbb{R}^{C\times C}, \tag{5}$$

from mean and variance for each element from two different covariance matrices of the i-th image in the batch, i.e.,

$$\boldsymbol{\mu}_{\boldsymbol{\Sigma}_i} = \frac{1}{2}\left(\boldsymbol{\Sigma}_i + \tilde{\boldsymbol{\Sigma}}_i\right), \tag{6}$$

$$\boldsymbol{\sigma}_i^2 = \frac{1}{2}\left(\left(\boldsymbol{\Sigma}_i - \boldsymbol{\mu}_{\boldsymbol{\Sigma}_i}\right)^2 + \left(\tilde{\boldsymbol{\Sigma}}_i - \boldsymbol{\mu}_{\boldsymbol{\Sigma}_i}\right)^2\right), \tag{7}$$

where N denotes the batch size. As a result, the mean value of the variance value of i-th channel can be calculated:

$$\mathbf{S}_i = \frac{1}{C}\sum_{j=1}^{C} \mathbf{V}_{ij}, \tag{8}$$

$\mathbf{S} \in \mathbb{R}^{C\times 1}$ implies the sensitivity of different channels to the augmentation [12], which involves light color and sensor read noise in our Night Augmentation. Therefore, the channels with high variance values contain features that are more responsive to light color and sensor read noise, causing instability when processing nighttime images.

To push the student model to learn more insensitive features, we randomly mask out the lowest sensitivity channels in the student model. The teacher-student consistency loss L_{unsup} ensures the student model remains aligned with the teacher model's prediction, even with the absence of masked-out information. This process will guide the student model to acquire the information that is more invariant to the augmentation, e.g., surface color. Moreover, it will eventually benefit the teacher model through the EMA updating mechanism.

3.3. Light Uncertainty

In unsupervised domain adaptation, accurately estimating uncertainty is fundamental to the successful training of the model. As mentioned above, our approach uses the teacher model to generate pseudo ground truths for unlabeled nighttime images. However, this introduces a risk: reliance on potentially incorrect pseudo labels generated by the model. Without proper management, this could create a detrimental feedback loop, reinforcing the model's inaccuracies [46]. To mitigate this risk, an effective uncertainty mechanism is vital. Specifically, we require a mechanism that functions well with a color constancy model and provides pixel-wise uncertainty estimation. Consequently, the student model can avoid learning from the possible incorrect area in the pseudo ground truths.

Previous works adopt Monte-Carlo uncertainty and ensemble uncertainty, which can provide pixel-wise uncertainty estimation. However, these approaches have significant drawbacks. The Monte-Carlo method usually requires

more than 20 times of forward computation [16], resulting in substantial computational complexity. An ensemble of neural networks can provide pixel-wise uncertainty estimation [28], but this demands the training and storage of multiple networks, which can be resource-intensive.

Algorithm 1 Light Uncertainty Calculation

Require: img: Input image in camera RGB space
Ensure: uncertaintyMap: Pixel-wise uncertainty map
function LIGHTUNCERTAINTY(img)
input1 ← teacherAug(img) ▷ Apply augmentation
output1 ← teacher(input1) ▷ First pass
input2 ← RGB2CAM(output1)
output2 ← teacher(input2) ▷ Second pass
lightMap ← calculateLight(input2, output2)
uncertaintyMap ← CosineSimilarity(lightMap, RGB2CAM(whiteLight))
return uncertaintyMap
end function

To address these limitations, we have developed a novel uncertainty mechanism, namely light uncertainty, focused on providing pixel-level uncertainty estimation for the color constancy model. This mechanism hinges on the teacher network's function, which is to predict the surface color of objects in a scene, invariant of the illuminant. If the teacher's initial prediction on the surface color is accurate, reprocessing this color-corrected image should yield no change, as the true surface colors are already revealed. Deviations in the light results provide a direct indication of uncertainty.

The implementation involves a two-step prediction cycle using the teacher network, as detailed in Algorithm 1. After the network produces the initial surface color estimation, we transform this output back to camera RGB space and reintroduce it to the network. The angle between the estimated light color from this second output and a standard achromatic light indicates the level of uncertainty. Such a localized measurement provides a nuanced understanding of the model's predictive confidence.

3.4. Overall Optimization Objectives

The overall optimization objective of our method consists of supervised loss and unsupervised loss, corresponding to daytime and nighttime training data. For both supervised and unsupervised loss, the L_1 loss and the mean angular error (MAE) loss [37] are used. To calculate MAE, in brief, the estimated illuminant map is first obtained by dividing the input image by the output image on the pixel level. Similarly, the ground truth illuminant map is obtained by dividing the input image by the ground truth image. MAE is the mean of the pixel-wise angular difference E_{MAE} between the two light color maps.

Our supervised loss is represented as:

$$L_{\text{sup}} = \lambda_1 L_1 + \lambda_2 L_{\text{MAE}}, \tag{9}$$

where λ's are the weights of different loss components. L_1 is the mean average error loss between the ground truths and predicted images.

In our unsupervised learning loss, pseudo ground truths are taken as the training target. The loss is weighted by the uncertainty map as $1-U$, where higher uncertainty leads to a lower weight for the loss at each pixel. This is expressed as:

$$L_{\text{unsup}} = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}(1-U_{ij})\left(\lambda_1|Y_{ij}-\hat{Y}_{ij}| + \lambda_2 E_{\text{MAE},ij}\right), \tag{10}$$

where Y and $\hat{Y}$ are the ground truth image and predicted images respectively, $E_{\text{MAE},ij}$ represents the pixel-wise angular difference for pixel (i,j), and H and W are the height and width of the images. U_{ij} is the light uncertainty value for pixel (i,j) in the uncertainty map U.

In conclusion, our overall optimization objective can be expressed as:

$$L_{\text{total}} = L_{\text{sup}} + L_{\text{unsup}}. \tag{11}$$

4. Experiment

4.1. Implementation Details

Network Details In our framework, both the teacher and student models utilize identical architectures, which are based on a U-Net structure with ResNet blocks as described in [18]. The feature dimensions at each level of the U-Net [36] are configured to 64, 128, 256, and 512, respectively. Each level of the network comprises two ResNet blocks. The total number of parameters in our model amounts to 27.4 million. For processing each image, the model demonstrates an inference time of 16.9 ms.

Training Details Our method is implemented based on the PyTorch library and trained on NVIDIA RTX 3090 GPUs. The training images are resized to the size of 256×256, while the same resolution is set for the output images. The training of our model uses the Adam [24] optimizer with a step decay learning rate scheduler. The initial learning rate is 0.001 and it decays to 0 after 140,000 iterations. During the pretraining, the model is trained for 140,000 iterations. The batch size is set as 16. During the unsupervised domain adaptation training, the student model is trained for 140,000 iterations. Following [37], the weights of λ_1 is set to 100, while λ_2 is set to 1.

(a) Raw Test Image (b) Visualized Test Image (c) Visualized Ground Truth (d) Evaluation Masks

Figure 3. Illustration of the image pairs and evaluation masks used in our dataset. (a) The original raw image, which will be the testing image. (b) The testing image processed with post-processing for visualization. (c) The ground truth image with post-processing for visualization, which includes the achromatic paper for illumination estimation. (d) Masks highlighting the areas covered by the achromatic paper, used in calculating the error in evaluation.

4.2. Datasets

Nighttime Dataset In our study, we address the limitations of previous color constancy datasets [3, 11, 23, 26], which predominantly comprise daytime or indoor images. Although some datasets like Cube+ [3] and Cube++ [13] have explored nighttime challenges, their reliance on calibration objects placed in image corners for ground truth estimation does not provide direct pixel-level annotation.

To overcome this, we have collected a new nighttime dataset with sparse, pixel-wise annotations by a Sony A7R3 camera. As shown in Fig. 3, our approach involves capturing two images of the same scene in quick succession to avoid changes in lighting conditions. In the second image, parts of the surface are covered with achromatic paper, which exhibits the local light color.

During testing, the image without the paper is taken as input. The model's performance is then evaluated based on the areas annotated with achromatic paper. For this purpose, we have compiled a dataset comprising 64 labeled images for evaluation and an additional 720 unlabeled images, which are utilized for unsupervised domain adaptation.

NUS-8 Dataset The NUS-8 dataset, a benchmark in single-illuminant color constancy research as detailed in [11], comprises 1,736 linear raw RGB images. This dataset is predominantly composed of images captured in daytime outdoor settings. Additionally, it includes a selection of indoor images, but these are characterized by the presence of a single light color. Some of the images are indoor but with a single light color. For our nighttime evaluation, the Sony dataset serves as the labeled daytime dataset in both pretraining and unsupervised domain adaptation training.

LSMI Dataset The Large Scale Multi-Illuminant (LSMI) dataset [23] is an extensive multi-illuminant indoor/daytime dataset comprising 7,486 images. These images are captured over 2,700 distinct scenes. Each scene features 1 to 3 light sources, encompassing a mix of natural and artificial lighting. For our evaluation, we divided the LSMI dataset into training, validation, and test sets in a 0.7, 0.2, and 0.1 ratio, respectively, through a random selection process. In the context of our multi-illuminant dataset evaluation, the training set was further subdivided into labeled and unlabeled data to facilitate semi-supervised learning.

4.3. Nighttime Evaluation

We compare our method with state-of-the-art methods, including supervised single-illuminant, unsupervised domain adaptation, unsupervised single-illuminant, supervised multi-illuminant, and unsupervised multi-illuminant methods. For the supervised methods, we trained on the daytime dataset. For the semi-supervised methods, we trained it using both daytime and unlabeled nighttime datasets.

In the evaluation, we first calculate the mean angular error (MAE) of the estimated light color on each evaluation mask. Then, for each image, we calculate the mean of the MAE among the masks.

Quantitative Evaluation Our method's performance was assessed using the Mean Angular Error (MAE) metric. As

(a) Test Image (b) Quasi-U [6] (c) CLCC [32] (d) AngularGAN [37] (e) Ours

Figure 4. The comparison between our method and state-of-the-art methods on the nighttime dataset. For visualization purposes, the images have been auto-brightened. In these challenging environments, our method yielded the most satisfactory results, effectively handling diverse lighting scenarios. In contrast, significant color casts remained evident in the results produced by other methods.

indicated in Table 1, our approach surpasses current state-of-the-art methods in both the mean and median of MAE. In the realm of deep learning-based methods, our method achieves a 23.7% reduction in mean MAE compared to the leading single-illuminant method, CLCC [32]. Additionally, when measured against the top-performing multiple-illuminant method, our approach shows a 21.5% improvement in mean MAE.

Qualitative Evaluation We conducted visual qualitative comparisons of our method against Quasi-U [6], CLCC [32], and AngularGAN [37] on our nighttime dataset. Figure 4 showcases these comparisons, where our method excels in generating color-corrected images of higher quality under challenging nighttime conditions, a task at which the state-of-the-art methods falter, especially with complex illuminants. The single illuminant methods Quasi-U [6] and CLCC [32] cannot fully remove the light colors. Although AngularGAN is tailored for multi-illuminant color scenarios, it tends to perform inconsistently under the complex lighting conditions of nighttime. In contrast, our method consistently produces results free of light color cast, even in these challenging nighttime environments. This underscores the robustness and effectiveness of our approach in navigating and correcting the intricate lighting nuances characteristic of nighttime scenes.

4.4. Multi-Illuminant Indoor Dataset Evaluation

In addition to addressing nighttime color constancy, one may wonder if our method can also work for daytime/indoor datasets. To this end, we investigate its applicability to daytime and indoor datasets, aiming to reduce the reliance on labeled data. This is evaluated through experiments on the LSMI dataset [23]. To train our model in a semi-supervised way, we use 10% of the training data with ground truths, complemented by 0% and 90% without ground truths, respectively. The labeled training data are used for the initial pretraining of our model.

We compare our method with both supervised learning and unsupervised learning methods. It is important to note that for the supervised learning methods, they have full access to all the ground truths in the training data.

As shown in Table. 2, our approach demonstrates competitive performance compared to other baselines, with only 10% labeled data. Furthermore, our method's performance improved progressively with the increase in the proportion of labeled data used, underscoring its adaptability across different data scenarios. Notably, in comparison to not utilizing unlabeled data, our method achieves a 33.8% improvement in the mean value of Mean Angular Error (MAE), underscoring its effectiveness in leveraging unlabeled data for enhanced learning outcomes.

Table 1. Comparison of the overall mean angular error on the nighttime dataset. We compare them using mean angular error (MAE). Our method achieves the best result among the baseline methods, including supervised, semi-supervised, and unsupervised methods for single and multi-illuminant color constancy.

Method	Supervised	Single light	Mean	Median
Gray-Edge (1st) [40]	✗	✓	12.18	12.20
Quasi-U [6]	✗	✓	10.18	8.60
FC4 (SqueezeNet) [19]	✓	✓	9.47	8.83
MSCC [20]	✗	✓	7.64	6.91
White-Patch [27]	✗	✓	7.45	5.52
CLCC [32]	✓	✓	6.61	6.39
FFCC[4]	✓	✓	6.32	6.14
Hussain WP [21]	✗	✗	12.36	11.40
Gijsenij *et al.* [17]	✗	✗	11.88	11.38
Patch CNN [7]	✓	✗	8.46	8.90
Gray-Pixel (M=2) [34]	✗	✗	7.78	7.27
Gray-Pixel (M=3) [34]	✗	✗	7.65	7.21
AngularGAN [37]	✓	✗	7.18	6.74
Gray-Pixel (M=1) [34]	✗	✗	7.03	7.33
Gray-Index [35]	✗	✗	6.19	6.04
MIMT [29]	✓	✗	<u>6.14</u>	<u>5.84</u>
Ours	✗	✗	**4.82**	**4.69**

4.5. Ablation Studies

In the ablation study, we analyze the performance of our model with and without the main components on the nighttime dataset. Table 3 reveals that using only the daytime dataset results in the least effective performance. Incorporating unlabeled nighttime data into our unsupervised domain adaptation framework results in enhanced performance, albeit with a modest margin. The introduction of our light uncertainty mechanism and adaptive channel masking leads to a significant reduction in the Mean Angular Error (MAE). To further demonstrate the effectiveness of our light uncertainty, we also provide analysis in our supplementary materials.

5. Conclusion

In this paper, we presented a novel framework for nighttime color constancy, by effectively utilizing unlabeled nighttime data. To enhance the model's adaptability to diverse lighting conditions, an adaptive channel masking scheme is designed to mask out the channels insensitive to light color and noise, guiding the model to progressively focus on learning such features. To avoid noisy pseudo label to be learned in the unlabeled data, we designed a light uncertainty mechanism that provides local estimation by reintroducing the white-balanced image into the model. Our experimental results, on both nighttime and indoor datasets, demonstrate that our method shows competitive results compared to existing baselines, underscoring its efficacy in learning from unlabeled images. Our method has also shown promising potential in semi-supervised learning applications. This adaptability suggests broader utility across various tasks in color constancy with limited labeled data.

Table 2. Comparison of the overall mean angular error on the LSMI dataset [23]. When training our method by semi-supervised learning, we use 10% of the training data with ground truths, 0% and 90% without ground truths, respectively.

Method	Labeled	Unlabeled	Mean	Median
White-Patch [27]	0%	100%	12.8	14.3
Gray-Edge (1st) [40]	0%	100%	12.1	10.8
Gray-World [9]	0%	100%	11.3	8.8
Gijsenij *et al.* [17]	0%	100%	18.0	17.0
Hussain WP [17]	0%	100%	17.7	16.9
Gray-Pixel(M=2) [34]	0%	100%	17.1	17.4
Gray-Index [35]	0%	100%	15.1	16.0
N-WB (GW) [2]	0%	100%	13.9	13.1
N-WB (GP) [2]	0%	100%	12.4	11.1
N-WB (GE) [2]	0%	100%	12.1	10.8
N-WB (PCA) [2]	0%	100%	8.3	7.4
Patch CNN [7]	100%	0%	4.82	4.24
AngularGAN [37]	100%	0%	4.69	3.88
MIMT [29]	100%	0%	**2.48**	**2.00**
Ours	10%	0%	5.71	4.86
Ours	10%	90%	<u>3.04</u>	<u>2.62</u>

Table 3. Ablation studies of the nighttime data and key components. We compare the mean angular error with and without: unlabeled nighttime data, light uncertainty (LU), and adaptive channel masking (ACM). The best performance is achieved when all the components are implemented.

Day	Night	LU	ACM	Mean	Median
✓	✗	✗	✗	7.36	6.65
✓	✓	✗	✗	6.95	6.21
✓	✓	✓	✗	5.40	4.97
✓	✓	✓	✓	4.82	4.69

References

[1] Mahmoud Afifi, Marcus A Brubaker, and Michael S Brown. Auto white-balance correction for mixed-illuminant scenes. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1210–1219, 2022. 1

[2] Teruaki Akazawa, Yuma Kinoshita, Sayaka Shiota, and Hitoshi Kiya. N-white balancing: White balancing for multiple illuminants including non-uniform illumination. *IEEE Access*, 10:89051–89062, 2022. 2, 8

[3] Nikola Banić, Karlo Koščević, and Sven Lončarić. Unsupervised learning for color constancy. *arXiv preprint arXiv:1712.00436*, 2017. 1, 2, 6

[4] Jonathan T Barron and Yun-Ta Tsai. Fast fourier color constancy. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 886–894, 2017. 8

[5] Shida Beigpour, Christian Riess, Joost Van De Weijer, and Elli Angelopoulou. Multi-illuminant estimation with conditional random fields. *IEEE Transactions on Image Processing*, 23(1):83–96, 2013. 2

[6] Simone Bianco and Claudio Cusano. Quasi-unsupervised color constancy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12212–12221, 2019. 1, 7, 8

[7] Simone Bianco, Claudio Cusano, and Raimondo Schettini. Single and multiple illuminant estimation using convolutional neural networks. *IEEE Transactions on Image Processing*, 26(9):4347–4362, 2017. 8

[8] David H Brainard and Brian A Wandell. Analysis of the retinex theory of color vision. *JOSA A*, 3(10):1651–1661, 1986. 1, 2

[9] Gershon Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin institute*, 310(1): 1–26, 1980. 1, 2, 8

[10] Ayan Chakrabarti, Keigo Hirakawa, and Todd Zickler. Color constancy with spatio-spectral statistics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(8):1509–1519, 2011. 1

[11] Dongliang Cheng, Dilip K Prasad, and Michael S Brown. Illuminant estimation for color constancy: why spatial-domain methods work and the role of the color distribution. *JOSA A*, 31(5):1049–1058, 2014. 1, 2, 6

[12] Sungha Choi, Sanghun Jung, Huiwon Yun, Joanne T Kim, Seungryong Kim, and Jaegul Choo. Robustnet: Improving domain generalization in urban-scene segmentation via instance selective whitening. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11580–11590, 2021. 4

[13] Egor Ershov, Alexey Savchik, Illya Semenkov, Nikola Banić, Alexander Belokopytov, Daria Senshina, Karlo Koščević, Marko Subašić, and Sven Lončarić. The cube++ illumination estimation dataset. *IEEE Access*, 8:227511–227527, 2020. 6

[14] Graham D Finlayson and Gerald Schaefer. Convex and non-convex illuminant constraints for dichromatic colour constancy. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, pages I–I. IEEE, 2001. 1

[15] Graham D Finlayson and Gerald Schaefer. Solving for colour constancy using a constrained dichromatic reflection model. *International Journal of Computer Vision*, 42:127–144, 2001. 1

[16] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059. PMLR, 2016. 5

[17] Arjan Gijsenij, Rui Lu, and Theo Gevers. Color constancy for multiple light sources. *IEEE Transactions on image processing*, 21(2):697–707, 2011. 1, 2, 8

[18] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 5

[19] Yuanming Hu, Baoyuan Wang, and Stephen Lin. Fc4: Fully convolutional color constancy with confidence-weighted pooling. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4085–4094, 2017. 1, 2, 8

[20] Xinwei Huang, Bing Li, Shuai Li, Wenjuan Li, Weihua Xiong, Xuanwu Yin, Weiming Hu, and Hong Qin. Multi-cue semi-supervised color constancy with limited training samples. *IEEE Transactions on Image Processing*, 29:7875–7888, 2020. 2, 8

[21] Md Akmol Hussain and Akbar Sheikh Akbari. Color constancy algorithm for mixed-illuminant scene images. *IEEE Access*, 6:8964–8976, 2018. 2, 8

[22] Brian L Joiner and Joan R Rosenblatt. Some properties of the range in samples from tukey's symmetric lambda distributions. *Journal of the American Statistical Association*, 66 (334):394–399, 1971. 4

[23] Dongyoung Kim, Jinwoo Kim, Seonghyeon Nam, Dongwoo Lee, Yeonkyung Lee, Nahyup Kang, Hyong-Euk Lee, ByungIn Yoo, Jae-Joon Han, and Seon Joo Kim. Large scale multi-illuminant (lsmi) dataset for developing white balance algorithm under mixed illumination. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2410–2419, 2021. 1, 2, 6, 7, 8

[24] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 5

[25] VON KRIES. Influence of adaptation on the effects produced by luminous stimuli. *handbuch der Physiologie des Menschen.*, 3:109–282, 1905. 4

[26] Firas Laakom, Jenni Raitoharju, Jarno Nikkanen, Alexandros Iosifidis, and Moncef Gabbouj. Intel-tau: A color constancy dataset. *IEEE access*, 9:39560–39567, 2021. 1, 2, 6

[27] Edwin H Land and John J McCann. Lightness and retinex theory. *Josa*, 61(1):1–11, 1971. 1, 2, 8

[28] Keuntaek Lee, Ziyi Wang, Bogdan Vlahov, Harleen Brar, and Evangelos A Theodorou. Ensemble bayesian decision making with redundant deep perceptual control policies. In *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, pages 831–837. IEEE, 2019. 5

[29] Shuwei Li, Jikai Wang, Michael S Brown, and Robby T Tan. Transcc: Transformer-based multiple illuminant

color constancy using multitask learning. *arXiv preprint arXiv:2211.08772*, 2022. 2, 8

[30] Yun Liu, Zhongsheng Yan, Sixiang Chen, Tian Ye, Wenqi Ren, and Erkang Chen. Nighthazeformer: Single nighttime haze removal using prior query transformer. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 4119–4128, 2023. 2

[31] Yun Liu, Zhongsheng Yan, Jinge Tan, and Yuche Li. Multi-purpose oriented single nighttime image haze removal based on unified variational retinex model. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(4):1643–1657, 2023. 2

[32] Yi-Chen Lo, Chia-Che Chang, Hsuan-Chao Chiu, Yu-Hao Huang, Chia-Ping Chen, Yu-Lin Chang, and Kevin Jou. Clcc: Contrastive learning for color constancy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8053–8063, 2021. 1, 2, 7, 8

[33] Boris T Polyak and Anatoli B Juditsky. Acceleration of stochastic approximation by averaging. *SIAM journal on control and optimization*, 30(4):838–855, 1992. 3

[34] Yanlin Qian, Said Pertuz, Jarno Nikkanen, Joni-Kristian Kämäräinen, and Jiri Matas. Revisiting gray pixel for statistical illumination estimation. *arXiv preprint arXiv:1803.08326*, 2018. 1, 2, 8

[35] Yanlin Qian, Joni-Kristian Kämäräinen, Jarno Nikkanen, and Jiri Matas. On finding gray pixels. In *IEEE International Conference of Computer Vision and Pattern Recognition*, 2019. 2, 8

[36] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 3, 5

[37] Oleksii Sidorov. Conditional gans for multi-illuminant color constancy: Revolution or yet another approach? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019. 1, 2, 5, 7, 8

[38] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in neural information processing systems*, 30, 2017. 2, 3

[39] Shoji Tominaga. Multichannel vision system for estimating surface and illumination functions. *JOSA A*, 13(11):2163–2173, 1996. 1

[40] Joost Van De Weijer, Theo Gevers, and Arjan Gijsenij. Edge-based color constancy. *IEEE Transactions on image processing*, 16(9):2207–2214, 2007. 1, 2, 8

[41] Wenhui Wang, Anna Wang, and Chen Liu. Variational single nighttime image haze removal with a gray haze-line prior. *IEEE Transactions on Image Processing*, 31:1349–1363, 2022. 2

[42] Kaixuan Wei, Ying Fu, Jiaolong Yang, and Hua Huang. A physics-based noise formation model for extreme low-light raw denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2758–2767, 2020. 4

[43] Sung-Min Woo, Sang-Ho Lee, Jun-Sang Yoo, and Jong-Ok Kim. Improving color constancy in an ambient light environment using the phong reflection model. *IEEE Transactions on Image Processing*, 27(4):1862–1877, 2017. 1

[44] Bolei Xu, Jingxin Liu, Xianxu Hou, Bozhi Liu, and Guoping Qiu. End-to-end illuminant estimation based on deep metric learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3616–3625, 2020. 2

[45] Huanglin Yu, Ke Chen, Kaiqi Wang, Yanlin Qian, Zhaoxiang Zhang, and Kui Jia. Cascading convolutional color constancy. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 12725–12732, 2020. 2

[46] Lequan Yu, Shujun Wang, Xiaomeng Li, Chi-Wing Fu, and Pheng-Ann Heng. Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22*, pages 605–613. Springer, 2019. 4