Domain Adversarial Learning for Color Constancy

Zhifeng Zhang, **Xuejing Kang**, **Anlong Ming***
School of Computer Science (National Pilot Software Engineering School), Beijing University of Posts and Telecommunications
{zhangzhifeng, kangxuejing, mal}@bupt.edu.cn

Abstract

Color Constancy aims to eliminate the color cast of RAW images caused by non-neutral illuminants. Though contemporary approaches based on convolutional neural networks significantly improve illuminant estimation, they suffer from the seriously insufficient data problem. To solve this problem by effectively utilizing multi-domain data, we propose the Domain Adversarial Learning Color Constancy (DALCC) which consists of the Domain Adversarial Learning Branch (DALB) and the Feature Reweighting Module (FRM). In DALB, the Camera Domain Classifier and the feature extractor compete against each other in an adversarial way to encourage the emergence of domain-invariant features. At the same time, the Illuminant Transformation Module performs color space conversion to solve the inconsistent color space problem caused by those domain-invariant features. They collaboratively avoid model degradation of multi-device training caused by the domain discrepancy of feature distribution, which enables our DALCC to benefit from multi-domain data. Besides, to better utilize multi-domain data, we propose the FRM that reweights the feature map to suppress Non-Primary Illuminant regions, which reduces the influence of misleading illuminant information. Experiments show that the proposed DALCC can more effectively take advantage of multi-domain data and thus achieve state-of-the-art performance on commonly used benchmark datasets.

1 Introduction

Color Constancy (CC) aims to eliminate the influence of non-neutral illuminants on the canonical color of objects, which is an essential part of the Image Signal Processor and can improve many downstream high-level computer vision tasks, such as visual recognition[Chen *et al.*, 2015], image segmentation[Afifi and Brown, 2019b].

Recent learning methods[Hernandez *et al.*, 2020; Lo *et al.*, 2021] based on Convolutional Neural Networks (CNN) achieve better illuminant estimation than traditional statistics-based approaches[Land, 1977; Buchsbaum, 1980]. Because those CNNs can learn accurate illuminant regression from labeled data. However, due to the expensive data acquisition in CC[Lo *et al.*, 2021] and the sensor domain gap[Hernandez *et al.*, 2020], those CNNs suffer from insufficient data, resulting in learning spurious correlations[Lo *et al.*, 2021] and limiting model capacity[Xiao *et al.*, 2020].

To solve the insufficient data problem, some methods have been proposed. Commonly used technologies, such as pre-training[Lou *et al.*, 2015] and data augmentation[Yu *et al.*, 2020], can not enrich RAW images and realistic illuminants, achieving marginal improvement. Multi-device training[Hernandez *et al.*, 2020] with multi-domain data is ideal for enriching RAW images and realistic illuminants. However, commonly used backbones of the existing CNN-based methods suffer from the domain discrepancy of feature distribution in multi-device training, resulting in model degradation or limited improvement. Besides, almost all previous methods utilize the illuminant information of all images regions indiscriminately, which endures the misleading illuminant information from Non-Primary Illuminant (NPI) regions, leading to the underutilization of multi-domain data.

In this paper, we propose the Domain Adversarial Learning Color Constancy (DALCC) to solve the domain discrepancy of feature distribution problem and reduce the influence of misleading illuminant information. First, to solve the domain discrepancy problem, we propose the Domain Adversarial Learning Branch (DALB) that consists of the Camera Domain Classifier (CDC) and the Illuminant Transformation Module (ITM). The CDC competes against the feature extractor to get domain-invariant features by minimizing $\mathcal{H}$-divergence between feature distributions. The ITM performs color space conversion to solve the inconsistent color space problem caused by those domain-invariant features. They collaboratively avoid model degradation in multi-device training, enabling our DALCC to utilize multi-domain data to solve the insufficient data problem. Besides, to reduce the influence of misleading illuminant information, we propose the Feature Reweighting Module (FRM) to reweight feature map by assigning lower confidence to misleading features from NPI regions, which is beneficial for utilizing multi-domain data better and achieving more accurate and stable illuminant regression. Our contributions can be summarized as follows:

*Contact Author

1. This paper introduces domain adversarial learning to CC for the first time, which provides a new way to utilize multi-domain data to alleviate the insufficient data problem.

2. The proposed DALB solves the domain discrepancy of feature distribution problem by minimizing $\mathcal{H}$-divergence and the consequent inconsistent color space problem by color space conversion, avoiding model degradation in multi-device training.

3. The proposed FRM reduces the influence of misleading illuminant information on CC by reweighting the feature map, which is beneficial for utilizing multi-domain data better and achieving more accurate and stable illuminant regression.

4. The proposed DALCC method achieves the state-of-the-art CC performance on commonly used benchmark datasets.

2 Related Work

2.1 Overview for Color Constancy

Methods for CC fall into two categories: the statistics-based and the learning-based. The formers [Buchsbaum, 1980; Van De Weijer *et al.*, 2007; Land, 1977; Gijsenij and Gevers, 2010] assume some statistical properties of the nature scene's surface reflectance are achromatic and estimate the illuminant color by the deviations between those statistical properties and grayness. Despite being fast and straightforward, their assumptions are often violated in natural scenes and easily affected by NPI regions, limiting their illuminant estimation abilities. The learning-based CNN methods[Hu *et al.*, 2017; Yu *et al.*, 2020] significantly improve illuminant estimation performance by learning from labeled data. However, due to the expensive data acquisition and the sensor domain gap, those CNN-based methods face seriously insufficient data, resulting in learning spurious correlations[Lo *et al.*, 2021] and limiting model capacity[Xiao *et al.*, 2020].

2.2 Color Constancy with Insufficient Data

Significant attempts have been made to tackle this challenging problem. [Lou *et al.*, 2015] pretrains their model with ImageNet, but images from ImageNet have no accurate illuminant labels, resulting in marginal improvement. [Bianco and Cusano, 2019] performs gray pixel detection on RAW data by learning semantic features from JPEG data, but this detection is another challenging task. [Afifi and Brown, 2019a; Hernandez *et al.*, 2020; Afifi *et al.*, 2021] design the cross-camera CC models to alleviate insufficient data and thus only achieve comparable camera-dependent performances. [Hu *et al.*, 2017; Yu *et al.*, 2020] perform data augmentation, such as random cropping, illuminant relighting to generate more training data. However, they can not enrich image scenes and realistic illuminants and suffer from misleading illuminant information. Multi-device training with multi-domain data is ideal for enriching images and realistic illuminants. However, commonly used backbones of the existing CNN-based methods suffer from the domain discrepancy of feature distribution in multi-device training, resulting in model degradation or limited improvement[Xiao *et al.*, 2020].

Recently, some researchers have started to develop methods to alleviate this domain discrepancy of feature distribution problem. [Xiao *et al.*, 2020] employs a domain-specific channel reweighting module to adjust the feature map, which alleviates this problem by assigning different weights to features from different domains. [Tang *et al.*, 2022] proposes the CGA-Branch that extracts image-specific color features to regularize the feature map, reducing the domain discrepancy through normalization. However, they are restricted in alleviating domain discrepancy of feature distribution through feature post-processing, resulting in limited ability in utilizing multi-domain data. Besides, they indiscriminately utilize the illuminant information of all image regions and thus suffer from misleading illuminant information from NPI regions.

In this paper, instead of reweighting or normalizing features, we initially introduce domain adversarial learning to modify the feature itself to be domain-invariant by minimizing $\mathcal{H}$-divergence, which can avoid the emergence of feature with domain discrepancy and enable our DALCC to benefit from multi-domain data. Besides, we reduce the influence of misleading illuminant information by assigning lower confidence to misleading features from NPI regions, which is beneficial for utilizing multi-domain data better and achieving more accurate illuminant regression.

3 Problem Description

3.1 Preliminary for Color Constancy

Under the prevalent single illuminant assumption and Lambert reflection model[Geusebroek *et al.*, 2003], the RAW image can be modeled as[Hernandez *et al.*, 2020]:

$$I(c, X) = \int_{\Omega} F(\lambda)R(\lambda, X)C(c, \lambda)d\lambda, c \in \{r, g, b\} \quad (1)$$

where $I(c, X)$ is the pixel intensity of color channel c at location X, Ω is visible spectrum and $\lambda \in \Omega$, $F(\lambda)$ is spectral power spectrum, $R(\lambda, X)$ is surface reflectance, $C(c, \lambda)$ is Camera Spectral Sensitivity (CSS). The goal of CC becomes the estimation of illuminant $L(c)$ where:

$$L(c) = \int_{\Omega} F(\lambda)C(c, \lambda)d\lambda, c \in \{r, g, b\} \quad (2)$$

3.2 Domain Discrepancy of Feature Distribution

Existing CNN-based methods face the insufficient training data problem, and multi-device training is ideal for enriching training data. However, due to the sensor domain gap, commonly used backbones of the existing CNN-based methods can only perform well in single-device training. When using them in multi-device training, those backbones suffer from the domain discrepancy of feature distribution, resulting in model degradation or limited improvement.

Next, we experimentally analyze this domain discrepancy problem. In multi-device training, the loss function is framed as[Xiao *et al.*, 2020]:

$$\mathcal{L}_{angular} = \sum\nolimits_{k=1}^{K} \mathcal{L}(E_{\phi}(G_{\theta}(I_k)), L_k) \quad (3)$$

where $G_{\theta}(.)$ and $E_{\phi}(.)$ are the feature extractor and the illuminant estimator parameterized by θ and ϕ. $\mathcal{L}(.)$ is the angular loss function; I_k, L_k represents RAW images and associated illuminant labels from the k-th camera domain. K is the total number of camera domains.

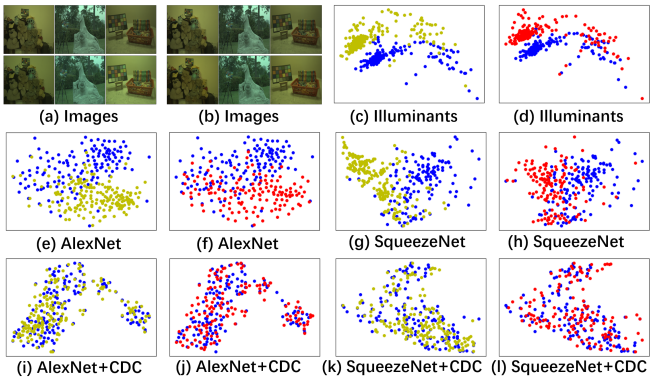

Figure 1: Domain discrepancy of feature distribution. (a) Images from Canon600D and SonyA57. (b) Images from Canon600D and NikonD5200. (c) Illuminants from Canon600D and SonyA57. (d) Illuminants from Canon600D and NikonD5200. Yellow, blue, and red represent camera domains: SonyA57, Canon600D, and NikonD5200. (e)-(l): Feature visualization of different models on different camera domains.

As shown in Figure 1(a)-1(b), RAW images I_k are camera-dependent because CSS is RAW images' component (Equation 1), which leads the illuminant labels L_k (Figure 1(c)-1(d)) extracted from I_k to be camera-dependent. Existing commonly used backbones (AlexNet and SqueezeNet) directly estimate the camera-dependent illuminant from the RAW image, which also leads the extracted feature $G_\theta(I_k)$ to be camera-dependent without other supervisory signals. We use t-SNE[Van der Maaten and Hinton, 2008] projection to visualize the feature $G_\theta(I_k)$ with color-coding the domains in Figure 1. There are clear non-overlaps between the feature distributions of different domains (Figure 1(e)-1(h)), which is due to the camera-dependent characteristic of feature. Those non-overlaps indicate that those features can be classified under the premise of a minor loss. It is the manifestation of a considerable $\mathcal{H}$-divergence between feature distributions[Ben-David *et al.*, 2010], and therefore the existing backbones suffer from the domain discrepancy of feature distribution problem. This problem leads the existing backbones can not effectively learn from multi-domain data and thus endure model degradation or limited improvement in multi-device training.

To solve the domain discrepancy of feature distribution problem, we start a pioneer work by introducing domain adversarial learning to get the domain-invariant feature.

4 Domain Adversarial Learning for Color Constancy

This section presents the DALCC consisting of DALB and FRM, which can promote effective learning from multi-domain data to alleviate the insufficient data problem.

4.1 Domain Adversarial Learning Branch

Camera Domain Classifier To solve the domain discrepancy of feature distribution problem, we propose the CDC that introduces domain adversarial learning to minimize the $\mathcal{H}$-divergence between feature distributions. As shown in Figure 2, the CDC adopts the traditional fully connected architecture and gradient reversal layer. We adopt n CDCs ($n = K - 1$), the j-th CDC distinguishes whether the feature from target domain T or source domain S_j ($T \neq S_j$). To minimize domain discrepancy, we design the adversarial loss $\mathcal{L}_A^j$ of the j-th CDC as the approximation of $\mathcal{H}$-divergence.

$$\mathcal{L}_A^j = d_p log(D_\varphi^j(F_p)) + (1 - d_p) log(1 - D_\varphi^j(F_p)) \quad (4)$$

where D_φ^j is the j-th CDC, $p = \{T, S_j\}$, d_p represents classification label with $d_T = 0$ and $d_{S_j} = 1$, $F_p = G_\theta(I_p)$ is the feature of image I_p. With CDC, we rewrite the loss function in multi-device training as:

$$\mathcal{L}_{loss} = \mathcal{L}_{angular} + \lambda * \sum\nolimits_{j=1}^{n} \mathcal{L}_A^j \quad (5)$$

where λ is a trade-off coefficient.

From Equation 5, the CDC is dedicated to distinguishing which camera domain the feature comes from while the feature extractor tries to fool the CDC. They compete against each other in an adversarial way to minimize the angular loss and adversarial loss $\mathcal{L}_A$ simultaneously. Since minimizing $\mathcal{L}_A$ directly minimizes the $\mathcal{H}$-divergence between feature distributions of different domains[Ben-David *et al.*, 2010], it leads the features of different domains to align gradually. As shown in Figure 1(i)-1(l), by adding our CDC, the features from different domains overlap with each other, which confirms that our CDC is beneficial for obtaining domain-invariant features from multi-domain data.

Illuminant Transformation Module Benefiting from our CDC, we get the domain-invariant feature through domain adversarial training. However, it brings another problem, that is, the inconsistent color space, which prevents us from learning accurate illuminant regression on multi-domain data. As shown in Figure 3(a), the illuminant estimation of different domains overlap with each other because the domain-invariant feature leads them to lie in a common space. However, the CC task requires achieving illuminant regression on multi-domain data, that is, the illuminant estimation should lie in different camera spaces, just like the original illuminant labels (Figure 3(c)).

To solve this inconsistent problem, we design the ITM that performs color space conversion to build a bridge between the illuminant estimation and illuminant label. As shown in Figure 2, the ITM estimates the Color Space Transfer Matrix (CSTM) by neural network and applies it to the illuminant estimation. To achieve accurate conversion from common space to camera space, the CSTM is expected to be camera-dependent. We realize this goal by performing convolution operations on the camera labels, which helps to obtain camera-dependent parameters. Then we multiply those parameters with the global feature from global average pooling to get the estimated CSTM. Therefore, the final angular loss function can be expressed as:

$$\mathcal{L}_{angular}^* = \sum\nolimits_{k=1}^{K} \mathcal{L}(M_k E_\phi(G_\theta(I_k)), L_k) \quad (6)$$

where M_k is the estimated CSTM for camera k.

In order to achieve the minimum angular error, Equation 6 will lead M_k to learn the CSTM from the common space to

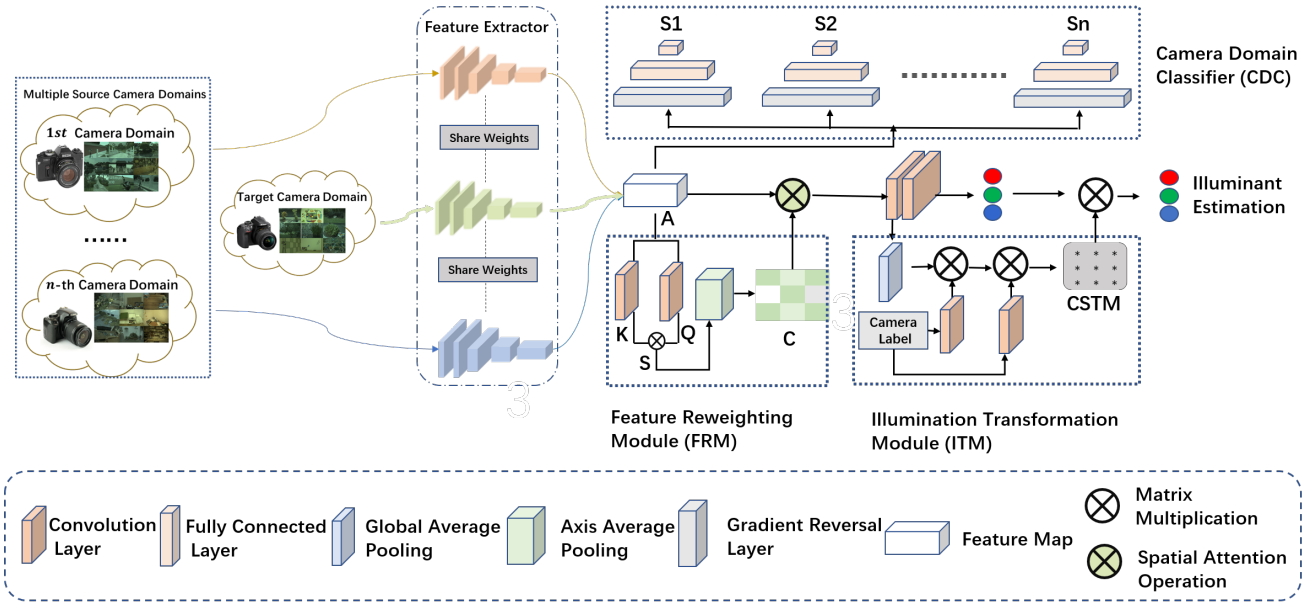

Figure 2: The network of domain adversarial learning color constancy.

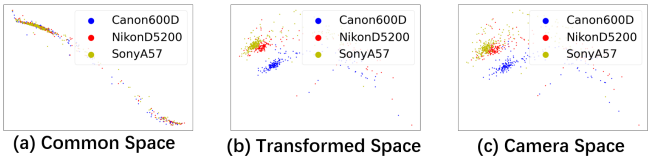

Figure 3: Illuminant distributions in different spaces.

the k-th camera space, which enables us to convert illuminant estimation to the transformed space. As shown in Figure 3(b), the illuminant distribution in the transformed space is aligned with the original illuminant label (Figure 3(c)), which confirms that our ITM solves the problem of inconsistent color space, enabling our DALCC to learn accurate illuminant regression on multi-domain data.

So far, by solving the domain discrepancy of feature distribution problem and the consequent inconsistent color space problem, our DALCC avoids model degradation in multi-device training and thus can learn better illuminant regression from multi-domain data.

4.2 Feature Reweighting Module

By now, few researchers are aware that images' NPI regions decrease illuminant estimation accuracy. Figure 4(a) shows that the NPI regions differ from most standard regions in illuminant information. If indiscriminately utilizing all illuminant information, the illuminant estimation abilities of CNNs will be damaged by the misleading illuminant information from NPI regions, leading to the underutilization of multi-domain data. Based on the fact that the total feature similarity of NPI regions is less than standard regions (Figure 4(b)), we design the FRM to reduce the influence of misleading illuminant information by reweighting the feature map, which can

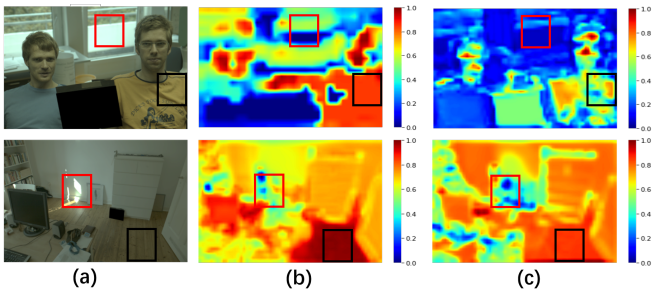

Figure 4: The FRM can suppress the NPI regions. (a): Natural images with NPI regions. Red boxes denote NPI regions, and black denotes standard regions. (b): The heatmap of total feature similarity. (c): The heatmap of confidence map generated by our FRM.

be expressed as:

$$A' = A \circledast C = A \circledast Sigmoid(AAP((QK^T) * m^{-1})) \quad (7)$$

where $\circledast$ represents the spatial attention operation.

As shown in Figure 2, the input feature A first generates two intermediate features Q and K by convolution operations. To get the feature similarity S of each local feature with others, we compute the dot product similarity between Q and K^T and regularize it with regularization term m. Then, we apply Axis Average Pooling (AAP) and sigmoid activation function on dot product similarity S to get the total feature similarity, i.e., confidence map C. As shown in Figure 4(c), our confidence map C assigns lower confidence to NPI regions and higher confidence to meaningful regions. Therefore, reweighting feature A with C to get the refined feature A' can suppress the NPI regions, which reduces the influence of misleading illuminant information. It promotes better utilization of multi-domain data by masking out NPI regions and thus leads to more stable and accurate illuminant regression.

Method	NUS-8[Cheng *et al.*, 2014]					Cube+[Banić *et al.*, 2017]				
	Mean	Med	Tri.	Best 25%	Worst 25%	Mean	Med	Tri.	Best 25%	Worst 25%
White Patch[Land, 1977]	9.91	7.44	8.78	1.44	21.27	6.80	3.85	5.21	0.68	16.93
Gray World[Buchsbaum, 1980]	4.59	3.46	3.81	1.16	9.85	3.52	2.55	2.82	0.60	7.98
Shades-of-Gray[Finlayson and Trezzi, 2004]	3.67	2.94	3.03	0.98	7.75	3.22	2.12	2.44	0.43	7.77
1st-order Gray-Edge[Van De Weijer *et al.*, 2007]	3.35	2.58	2.76	0.79	7.18	3.06	2.05	2.32	0.55	7.22
2nd-order Gray-Edge[Van De Weijer *et al.*, 2007]	3.36	2.70	2.80	0.89	7.14	3.28	2.34	2.58	0.66	7.44
Cheng[Cheng *et al.*, 2014]	2.18	1.48	1.64	0.46	5.03	-	-	-	-	-
Color Dog[Banic and Loncaric, 2015]	2.83	1.77	2.03	0.48	7.04	3.32	1.19	1.60	0.22	10.22
FC4-Squeezenet[Hu *et al.*, 2017]	2.23	1.57	1.72	0.47	5.15	1.35	0.93	1.01	0.30	3.24
FFCC[Barron and Tsai, 2017]	1.99	1.31	1.43	**0.35**	4.75	1.38	0.74	0.89	**0.19**	3.67
APAP using GW[Afifi *et al.*, 2019]	2.40	1.76	-	0.55	5.42	2.01	1.36	-	0.38	4.71
SIIE[Afifi and Brown, 2019a]	2.05	1.50	-	0.52	4.48	2.14	1.44	-	0.44	5.06
Quisa[Bianco and Cusano, 2019]	1.97	1.41	-	-	-	2.69	1.76	2.00	0.49	6.45
Daniel[Hernandez *et al.*, 2020]	2.35	1.55	1.73	0.46	5.62	-	-	-	-	-
MDLCC[Xiao *et al.*, 2020]	1.78	1.29	1.40	0.42	3.97	1.24	0.83	0.92	0.26	**2.91**
IGTN[Xu *et al.*, 2020]	1.85	1.24	-	0.36	4.58	-	-	-	-	-
C4[Yu *et al.*, 2020]	1.96	1.42	1.53	0.48	4.40	-	-	-	-	-
CLCC[Lo *et al.*, 2021]	1.84	1.31	1.42	0.41	4.20	-	-	-	-	-
C5[Afifi *et al.*, 2021]	1.77	1.37	1.46	0.48	3.75	1.39	0.79	0.93	0.24	3.55
TLCC[Tang *et al.*, 2022]	1.60	1.27	1.33	0.44	3.35	1.31	0.97	1.03	0.29	2.99
DALCC	**1.42**	**1.06**	**1.13**	0.38	**3.09**	**1.20**	**0.71**	**0.82**	**0.19**	3.06

Table 1: Color constancy results by different methods on NUS-8, Cube+. The best metric is shown in black.

5 Experiment

5.1 Dataset

We verify the effectiveness of our proposed DALCC on two public datasets, the NUS-8 dataset[Cheng *et al.*, 2014] and the Cube+ dataset[Banić *et al.*, 2017]. The NUS-8 dataset consists of 1736 RAW images captured by eight different cameras. The Cube+ dataset is composed of 1707 RAW images captured by Canon550D. All images are processed by demosaicing, black-level subtraction, and saturated pixel removal to get the linear RGB images. Following[Tang *et al.*, 2022; Xiao *et al.*, 2020], we adopt the three-fold cross-validation in all experiments.

5.2 Implementation Details

Angular Loss and Evaluation Metrics Following[Yu *et al.*, 2020; Xiao *et al.*, 2020], we measure our method by the angular loss between illuminant label L and illuminant estimation $\tilde{L}$ (Equation.8) and report the standard statistics (Mean, Median, Tri-mean, Best 25%, Worst 25%.) of angular loss to summarize the results over the investigated datasets.

$$Angularloss = \frac{\pi}{180} \arccos \frac{\tilde{L} \cdot L}{||\tilde{L}||||L||} \tag{8}$$

Network Implementation and Data Augmentation We adopt the backbone of FC4-Alexnet[Hu *et al.*, 2017] and implement our network on Pytorch with CUDA support. We train our model about 3000 epochs by setting the learning rate to 1×10^{-4}. The batch size is 16. We use Adam to optimize the network. In the training phase, we first mask the calibration object in the image. Then, random image cropping, rotating, and flipping are used[Hu *et al.*, 2017]. Finally, we resize the image to 256×256. In the test phase, we mask the calibration object and directly resize the image to 256×256 to speed up the test process.

Experiments	Test	Mean	Med.	Tri.	Best 25%	Worst 25%
A:Backbone	Cube+	1.35	0.83	0.96	0.21	3.40
A:Backbone	NUS	2.08	1.68	1.75	0.59	4.39
B:Backbone (M)	Cube+	1.42	0.90	1.02	0.23	3.49
B:Backbone (M)	NUS	1.78	1.30	1.38	0.44	3.95
C:Backbone w DALB (M)	Cube+	1.25	0.72	0.86	0.21	3.30
C:Backbone w DALB (M)	NUS	1.49	1.13	1.22	0.42	3.19

Table 2: The effectiveness of domain adversarial learning to color constancy. M denotes multi-device training.

5.3 Comparison with State-of-the-art Methods

As shown in Table 1, our DALCC method outperforms all start-of-the-art methods in NUS-8 and Cube+ datasets. For the NUS-8 dataset, compared to previous methods, our DALCC achieves the improvements of 11.2%, 14.5%, 15.0%, 7.7% in mean, median, tri-mean and worst 25 % metrics respectively. For the Cube+ dataset, our DALCC also achieves the improvements of 3.2%, 4.0%, 7.8% in mean, median, tri-mean metrics respectively. It confirms that by eliminating the domain discrepancy of feature distribution and reducing the influence of misleading illuminant information, our DALCC can learn from multi-domain data better and thus achieve more stable and accurate illuminant regression.

5.4 Effectiveness of Domain Adversarial Learning

We conduct three groups of experiments to verify the effectiveness of the domain adversarial learning to color constancy. As shown in Table 2, A and B are the experimental groups that our backbone is trained separately on single-domain and multi-domain data. C is the experimental group that trains our DALCC on multi-domain data. For comparability, FRM is removed from DALCC. By comparing A and B, we can confirm that direct multi-device training results in model degradation or limited improvement. We introduce domain adversarial learning to CC and solve the do-

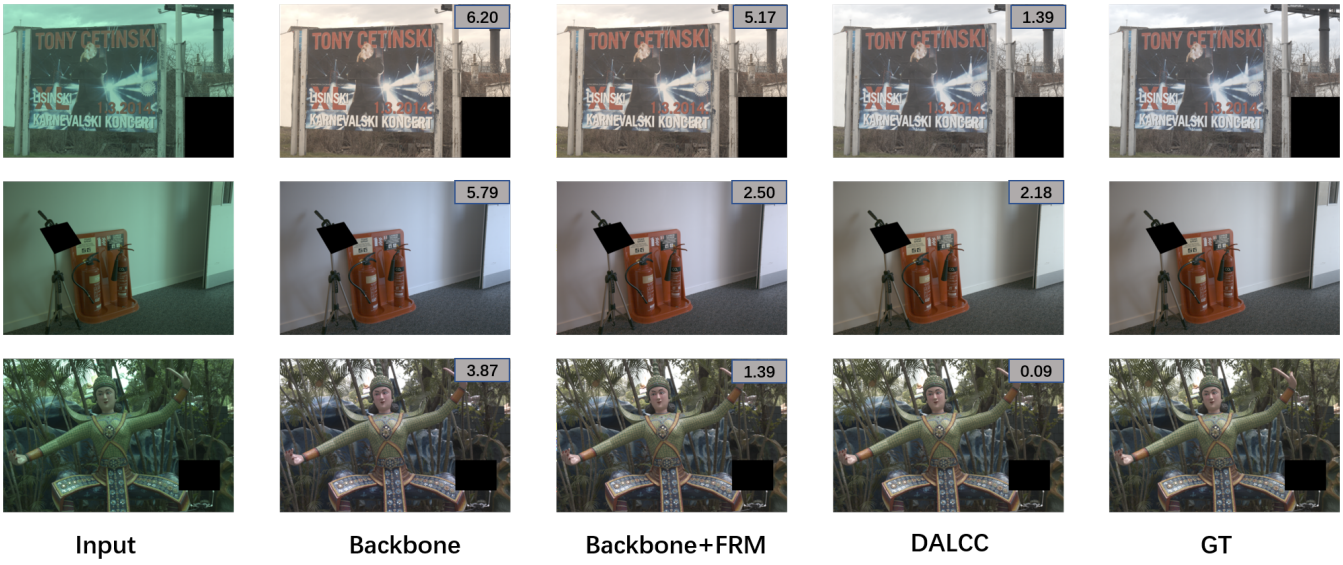

Figure 5: Visualization of color constancy results by different network architectures. DALCC denotes Backbone with FRM and DALB. Images are performed gamma correction for visualization. The upper right corner of image is angular loss.

Architectures	Mean	Med.	Tri.	Best 25%	Worst 25%
(1): Backbone	1.42	0.90	1.02	0.23	3.49
(2): Backbone w CDC	2.20	1.54	1.68	0.82	4.74
(3): Backbone w DALB	1.25	0.72	0.86	0.21	3.30
(4): Backbone w FRM	1.35	0.93	1.01	0.30	3.24
(5): DALCC	**1.20**	**0.71**	**0.82**	**0.19**	**3.06**

Table 3: The ablation study of our DALCC on Cube+ dataset.

Methods	Mean Angular Error	Parameters
FC4-Alexnet[Hu *et al.*, 2017]	2.12	2.93M
C4[Yu *et al.*, 2020]	1.96	5.19M
IGTN[Xu *et al.*, 2020]	1.85	500M
CLCC[Lo *et al.*, 2021]	1.84	**1.73M**
TLCC[Tang *et al.*, 2022]	1.60	30.52M
DALCC	**1.42**	3.00M

Table 4: Model complexity versus mean angular error on NUS-8 dataset.

main discrepancy of feature distribution problem. It forces our DALCC to extract the domain-invariant feature, which avoids model degradation and benefits from multi-domain data. Therefore, experimental group C achieves significant improvement over experimental group A.

5.5 Ablation Study and Analysis

As shown in Table 3, we carry out ablation study on the Cube+ dataset to evaluate the effectiveness of our DALCC model. Though CDC can solve the domain discrepancy of feature distribution, when directly transferring it to the CC task, it suffers from the inconsistent color space problem, resulting in a significant decrease in experiment (2) relative to experiment (1). Our ITM solves this problem by performing color space conversion, which supports our model to learn accurate illuminant regression with the domain-invariant feature, avoiding model degradation in multi-device training. Thus, experiment (3) achieves significant improvement to experiments (2) and (1). To learn from multi-domain data more effectively, our FRM reduces the influence of misleading illuminant information by assigning lower confidence to NPI regions. Thereby, experiment (5) achieves a more accurate and stable illuminant estimation than experiment (3). Adding the FRM to the backbone alone can also promote more effective learning from multi-domain data, which can be verified by comparing the experiment (1) and experiment (4). Some visualization results are illustrated in Figure 5.

5.6 Model Complexity Versus Angular Loss

We compare the model complexity and the mean angular loss in the NUS-8 dataset of our DALCC with the existing advanced methods, such as TLCC, CLCC, ICTN, C4, and FC4. As shown in Table 4, our DALCC achieves a 33.0% improvement in mean angular loss compared to our backbone FC4-Alexnet with only a 2.3% increment of model complexity. The CLCC has a more lightweight model but achieves minor improvement than us. Some competitive methods, such as C4 and IGTN, use more model parameters but gently improve. Though TLCC gives the most similar improvement, their model complexity is $10\times$ than ours.

6 Conclusion

This paper proposes the DALCC method to alleviate the insufficient data problem of the CC task by effectively utilizing multi-domain data. We achieve this by minimizing the domain discrepancy of feature distribution and reducing the influence of misleading illuminant information. Experiment results show that the proposed DALCC can avoid model degradation in multi-device training, suppress NPI regions and thus achieve state-of-the-art performance on commonly used benchmark datasets. In future work, we plan to extend our DALCC method as a CC framework.

Acknowledgements

This work was supported by the national key R & D program intergovernmental international science and technology innovation cooperation project (2021YFE0101600).

References

[Afifi and Brown, 2019a] M. Afifi and M. S. Brown. Sensor-independent illumination estimation for dnn models. In *BMVC 2019*, 2019.

[Afifi and Brown, 2019b] Mahmoud Afifi and Michael S Brown. What else can fool deep learning? addressing color constancy errors on deep neural network performance. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 243–252, 2019.

[Afifi *et al.*, 2019] Mahmoud Afifi, Abhijith Punnappurath, Graham Finlayson, and Michael S Brown. As-projective-as-possible bias correction for illumination estimation algorithms. *JOSA A*, 36(1):71–78, 2019.

[Afifi *et al.*, 2021] Mahmoud Afifi, Jonathan T Barron, Chloe LeGendre, Yun-Ta Tsai, and Francois Bleibel. Cross-camera convolutional color constancy. In *Proc. IEEE Int. Conf. Comput. Vis.*, pages 1981–1990, 2021.

[Banic and Loncaric, 2015] Nikola Banic and Sven Loncaric. Color dog-guiding the global illumination estimation to better accuracy. In *VISAPP*, pages 129–135, 2015.

[Banić *et al.*, 2017] Nikola Banić, Karlo Koščević, and Sven Lončarić. Unsupervised learning for color constancy. *arXiv preprint arXiv:1712.00436*, 2017.

[Barron and Tsai, 2017] 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.

[Ben-David *et al.*, 2010] Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Machine learning*, 79(1):151–175, 2010.

[Bianco and Cusano, 2019] 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.

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

[Chen *et al.*, 2015] Yu-Hsiu Chen, Ting-Hsuan Chao, Sheng-Yi Bai, Yen-Liang Lin, Wen-Chin Chen, and Winston H Hsu. Filter-invariant image classification on social media photos. In *Proc. ACM Conf. Multimed.*, pages 855–858, 2015.

[Cheng *et al.*, 2014] 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.

[Finlayson and Trezzi, 2004] Graham D Finlayson and Elisabetta Trezzi. Shades of gray and colour constancy. In *Color and Imaging Conference*, volume 2004, pages 37–41. Society for Imaging Science and Technology, 2004.

[Geusebroek *et al.*, 2003] Jan-Mark Geusebroek, Rein van den Boomgaard, Arnold WM Smeulders, and Theo Gevers. Color constancy from physical principles. *Pattern Recognition Letters*, 24(11):1653–1662, 2003.

[Gijsenij and Gevers, 2010] Arjan Gijsenij and Theo Gevers. Color constancy using natural image statistics and scene semantics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(4):687–698, 2010.

[Hernandez *et al.*, 2020] Daniel Hernandez, Sarah Parisot, Benjamin Busam, Ales Leonardis, Gregory Slabaugh, and Steven McDonagh. A multi-hypothesis approach to color constancy. In *Proceedings of the Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 2270–2280, 2020.

[Hu *et al.*, 2017] 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.

[Land, 1977] Edwin H Land. The retinex theory of color vision. *Scientific american*, 237(6):108–129, 1977.

[Lo *et al.*, 2021] 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.

[Lou *et al.*, 2015] Zhongyu Lou, Theo Gevers, Ninghang Hu, Marcel P Lucassen, et al. Color constancy by deep learning. In *BMVC*, pages 76–1, 2015.

[Tang *et al.*, 2022] Yuxiang Tang, Xuejing Kang, Chunxiao Li, Zhaowen Lin, and Anlong Ming. Transfer learning for color constancy via statistic perspective. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022.

[Van De Weijer *et al.*, 2007] Joost Van De Weijer, Theo Gevers, and Arjan Gijsenij. Edge-based color constancy. *IEEE Trans. Image Process.*, 16(9):2207–2214, 2007.

[Van der Maaten and Hinton, 2008] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

[Xiao *et al.*, 2020] Jin Xiao, Shuhang Gu, and Lei Zhang. Multi-domain learning for accurate and few-shot color constancy. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 3258–3267, 2020.

[Xu *et al.*, 2020] Bolei Xu, Jingxin Liu, Xianxu Hou, Bozhi Liu, and Guoping Qiu. End-to-end illuminant estimation based on deep metric learning. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 3616–3625, 2020.

[Yu *et al.*, 2020] 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*, volume 34, pages 12725–12732, 2020.