Object-based attention mechanism for color calibration of UAV remote sensing images in precision agriculture.

HUANG, H., TANG, Y., TAN, Z., ZHUANG, J., HOU, C., CHEN, W. and REN, J.

2022

Object-based attention mechanism for color calibration of UAV remote sensing images in precision agriculture

Huasheng Huang, Yu Tang, *Member, IEEE*, Zhiping Tan, Jiajun Zhuang, Chaojun Hou, Weizhao Chen, and Jinchang Ren, *Senior Member, IEEE*

***Abstract*—Color calibration is a critical step for Unmanned Aerial Vehicles (UAV) remote sensing, especially in precision agriculture, which relies mainly on correlating color changes to specific quality attributes, e.g., plant health, disease, and pest stresses. In UAV remote sensing, the exemplar-based color transfer is popularly used for color calibration, where the automatic search for the semantic correspondences is the key to ensuring the color transfer accuracy. However, the existing attention mechanisms encounter difficulties in building the precise semantic correspondences between the reference image and the target one, in which the normalized cross correlation is often computed for feature reassembling. As a result, the color transfer accuracy is inevitably decreased by the disturbance from the semantically unrelated pixels, leading to semantic mismatch due to the absence of semantic correspondences. In this paper, we proposed an unsupervised object-based attention mechanism (OBAM) to suppress the disturbance of the semantically unrelated pixels, along with a further introduced weight-adjusted AdaIN (WAA) method to tackle the challenges caused by the absence of semantic correspondences. By embedding the proposed modules into a photorealistic style transfer method with progressive stylization, the color transfer accuracy can be improved while better preserving the structural details. We evaluated our approach on the UAV data of different crop types including rice, beans, and cotton. Extensive experiments demonstrate that our proposed method outperforms several state-of-the-art methods. As our approach requires no annotated labels, it can be easily embedded into the off-the-shelf color transfer approaches. Relevant codes and configurations will be available at http://github.com/huanghsheng/object-based-attention-mechanism.**

***Index Terms*—Unmanned aerial vehicles (UAV); semantic correspondences; attention mechanism; color transfer.**

I. Introduction

Color is essential in many precision agriculture applications for assessing plant health, disease, and pest stresses. Unfortunately, color cast is inevitable due to the rapid ambient temperature changes and light irradiance under natural field conditions [1]. Due to the limited coverage area of a single UAV imagery, an image sequence with constant overlap is often used to produce an orthomosaics to cover the whole field. For the same flight, the random color variation causes not only the color error in the UAV imagery but also color inconsistency in the final orthomosaics. According to Afifi et al. [2], images' color cast negatively impacts image classification and segmentation. Therefore, color cast may cause inaccuracies in assessing crop stress/yield and nutrient deficiency, even using the newly emerged deep learning models [3].

In general, computational color constancy and exemplar-based color transfer methods are usually used to calibrate the color variations of the UAV imagery [4]. The former automatically adjusts the color value according to the illumination changes [5, 6], whilst the latter considers one image with the correct color as the reference image before transferring the color pattern of the reference image to other images captured in the same UAV flight [7, 8]. In practice, the sensor is hard to detect illumination changes during the high-speed flight process, thus, the commercial sensors established with color constancy still suffer from the color cast. Therefore,

Manuscript received September 12, 2022. This work was supported by the National key research and development project of China (Grant No. 2021YFD2000104-02), the National Natural Science Foundation of China (Grant No. 32071895), the Planned Science and Technology Project of Guangdong Province, China (Grant Nos. 2019B020216001, 2019A050510045, and 2021A0505030075), the Natural Science Foundation of Guangdong Province, China (Grant Nos. 2020B1515120070, and 2021A1515010824), the Key Project of Universities in Guangdong Province, China (Grant No. 2020ZDZX1061), the Innovation Team Project of Universities in Guangdong Province, China (Grant No. 2021KCXTD010), the Planned Science and Technology Project of Guangzhou, China (Grant Nos. 202002020063, 202007040007, and 202103000028), the Basic and Applied Basic Research Fund in Guangdong Province, China (Grant No. 2021A1515110756), the Basic and Applied Basic Research Project of Guangzhou City, China (Grant No. 202201011391), the Project of Educational Commission of Guangdong Province, China (Grant No. 2021KQNCX044), and the Rural Revitalization Strategy Project of Guangdong Province, China (Grant No. 2019KJ138). (Corresponding author: Yu Tang.)

Huasheng Huang and Yu Tang are with the College of Computer Sciences, Guangdong Polytechnic Normal University, Guangzhou, China, and also with the Academy of Interdisciplinary Studies, Guangdong Polytechnic Normal University, Guangzhou, China. (e-mail: huanghsheng@gpnu.edu.cn; yutang@gpnu.edu.cn).

Zhiping Tan and Weizhao Chen are with the Academy of Interdisciplinary Studies, Guangdong Polytechnic Normal University, Guangzhou, China. (e-mail: tanzp@gpnu.edu.cn; weizhao.chen@foxmail.com).

Jiajun Zhuang and Chaojun Hou are with the Academy of Contemporary Agriculture Engineering Innovations, Zhongkai University of Agriculture and Engineering, Guangzhou, China. (e-mail: zhuangjiajun@zhku.edu.cn; houchaojun@zhku.edu.cn).

Jinchang Ren is with the College of Computer Sciences, Guangdong Polytechnic Normal University, Guangzhou, China, and also with the Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow G1 1XW, U.K. (e-mail: jinchang.ren@ieee.org).

the exemplar-based color transfer becomes the most popular way to calibrate the color variation in the UAV imagery [4].

There are two main steps in the exemplar-based color calibration. The first is to detect the semantic correspondences between the reference image and the target image, and the second is to conduct the color transfer between the homogeneous regions of these two images. For the first step, the semantic segmentation or attention mechanism is often applied to generate the semantic correspondences. The semantic color transfer employs the semantic segmentation models to generate the semantic maps of the reference image and the target one. The color transfer is only conducted between the regions with the same semantic category [9]. The attentional color transfer method usually computes the normalized cross correlation between the representation of the image pair, and reassembles the deep features according to the cross correlation for the image synthesis [10, 11]. In conventional color transfer methods, the mean, standard deviation, or other statistics measures are often employed in the second step to design a linear or nonlinear transform function [12]. However, these low-level statistics features often fail to capture the semantic layout, leading to poor photorealism within the stylization results. Recently, with the rapid development of deep learning, the style transfer has successfully transferred the color from the reference to the target images while preserving the spatial details of the latter [13-15]. Despite the great success of arbitrary style transfer, the automatic search for semantic correspondences is still challenging. The semantic style transfer approach performs the transfer process between the regions with the same semantic category. However, this strategy requires a large number of manual annotations to train the semantic segmentation model, which is very labor intensive and time consuming thus hard to be applied to unknown scenarios. The attentional style transfer method reassembles the feature via the cross correspondences, where the color transfer accuracy is inevitably affected by the semantically unrelated pixels. Also, both approaches ignore the absence of semantic correspondences, which can easily cause significant semantic mismatch and noticeable structural artifacts.

To address the aforementioned challenges, a novel unsupervised color calibration method is proposed in this paper. As our method requires no annotated labels, it can be easily embedded into the off-the-shelf color transfer models without extra training. The major contributions of this paper can be summarized as follows.

1) We propose an object-based attention mechanism (OBAM) to search the semantic correspondences between the target and the reference images, which can effectively suppress the disturbance from the semantically unrelated elements.

2) We introduce a weight-adjusted AdaIN (WAA) method to address the absence of semantic reference, which has improved the perceptual quality of the whole imagery.

3) We conduct comprehensive experiments on the UAV imageries of different plant species and achieve state-of-the-art performance. The code and data will be made publicly available to further benefit the community.

II. Related Work

Existing exemplar-based style transfer methods include global and local fashion. The global style transfer method matches the global statistics from the reference to the target image, and often fail the mission when the image pair have a different semantic distribution. Local transfer algorithms only perform the transfer task between the semantically related regions, increasing the color transfer accuracy and better preserving the structural details. The main component of local transfer is to obtain the dense semantic correspondences, where the mainstream literatures can be divided into two categories, i.e. semantic style transfer and attentional style transfer. The semantic style transfer method utilizes the semantic maps from manual labeling or semantic segmentation models to guide the accurate style transfer for each semantic class. In contrast, the attentional style transfer method automatically generates the semantic correspondences by the attention mechanism.

A. Semantic style transfer

Semantic style transfer algorithms utilize semantic information to guide the transfer between the regions with the same semantic class, where the semantic information is either from manual labeling or the semantic segmentation models [16]. Luan et al. [17] proposed the locally affine transformation in RGB color space and expressed this module as an energy term. However, solving the optimization problem for the energy term requires heavy computational costs, which limits their practical usage. Li et al. [18] and Yoo et al. [19] proposed to transform the representation into a one-dimensional vector, and perform the style transfer with whitening and coloring transform, making it easy to be incorporated with the semantic maps. Anokhin et al. [20] proposed to append an extra semantic segmentation branch for the decoder, and this architecture proved to help to preserve the semantic context in the style transfer. Zhu et al. [21] proposed semantic region-adaptive normalization (SEAN), which is designed for accurate style transfer using the given semantic masks. Ma et al. [22] proposed to use semantic information to guide the image reconstruction in order to better preserve the content details. Though semantic style transfer obtains higher accuracy in color transfer, they need annotated labels to train the semantic segmentation model or guide the local transfer, which requires tedious manual tags and prevents its applications in generic scenarios.

B. Attentional style transfer

Recently, the attention mechanism has been raised as the fundamental tool to automatically search for the dense semantic correspondences in style transfer. Liao et al. [23] applied the image analogy to the deep features, and built the dense semantic correspondences via the nearest-neighbor field (NNF) method. He et al. [24] further extended this work by jointly optimizing the dense semantic correspondences and the linear transformation models, preventing the content mismatching that occurred in the previous work. However, the NNF searching process of Liao et al. [23] and He et al. [24] consumes too much computation and cannot be implemented with GPU acceleration due to its repeated logical judgement, preventing its further development.

Instead, the normalized cross-correlation is widely employed for correspondence estimation and image reconstruction. Chen et al. [25] proposed patch-wise similarity matching between the content and style activation patches using normalized cross-correlation. Avatar-Net [26] further extended the AdaIN [27] with a multi-scale fusion strategy, and integrated the feature matching with projection. Huang et al. [28] utilized patch attention to address the problem of pixel isolation, and adopted multi-level fusion for better stylization. Park et al. [29] introduced a style-attentional network (SANet) to exploit the semantic correlation in a self-attention mode. On this basis, Chen et al. [30] proposed a novel loss design to address the artifacts in style transfer. The loss function involves internal statistics, external information, and two contrastive losses. Zhang et al. [31] proposed a transfer network that automatically search for the semantic correlations for semantic-level transfer.

Generally speaking, the attentional style transfer methods with normalized cross-correlation are the state of the art technique in terms of color transfer accuracy and detail preservation. This technique mainly computes the normalized cross-correlation to estimate the dense semantic correspondences and reconstructs the image by reassembling the features according to the correspondence scores. In the scenario of color calibration for UAV remote sensing in precision agriculture, three are two main constraints. First, the existing approaches ignore the absence of semantic correspondences in many occasions, which may easily cause semantic mismatch and structural artifacts. Second, image reconstruction with weighted reassembling is inevitably influenced by the semantically unrelated elements, decreasing the color transfer accuracy. To address these challenging issues, we propose in this paper an OBAM method, and extensive experiments have demonstrated the efficacy of our method as detailed in the following sections.

III. Method

In this section, we describe our proposed object-based attention mechanism (OBAM) method for color calibration of UAV remote sensing in precision agriculture, as shown in Fig. 1. Our research is motivated by the style transfer framework, therefore the regular terms of style transfer researches were employed in this paper. Conceptually, the style image in this paper refers to the reference image, and the content image denotes the target image. First, we propose an object-based attention mechanism (OBAM) to search the semantic correspondences between the image pair under the unsupervised mode, which is described in section III-A. In this stage, the content map, style map, and the confidence score for each pixel were generated, and the whole content image was divided into the areas with strong semantic correspondences and weak semantic correspondences. Next, we apply the wavelet corrected whitening and coloring transforms (wavelet corrected WCT) to transfer the color for the areas with strong semantic correspondences, as given in section III-B. Finally, we introduce the weight-adjusted AdaIN (WAA) method to address the absence of semantic correspondences for the areas with weak semantic correspondences, as detailed in section III-C.

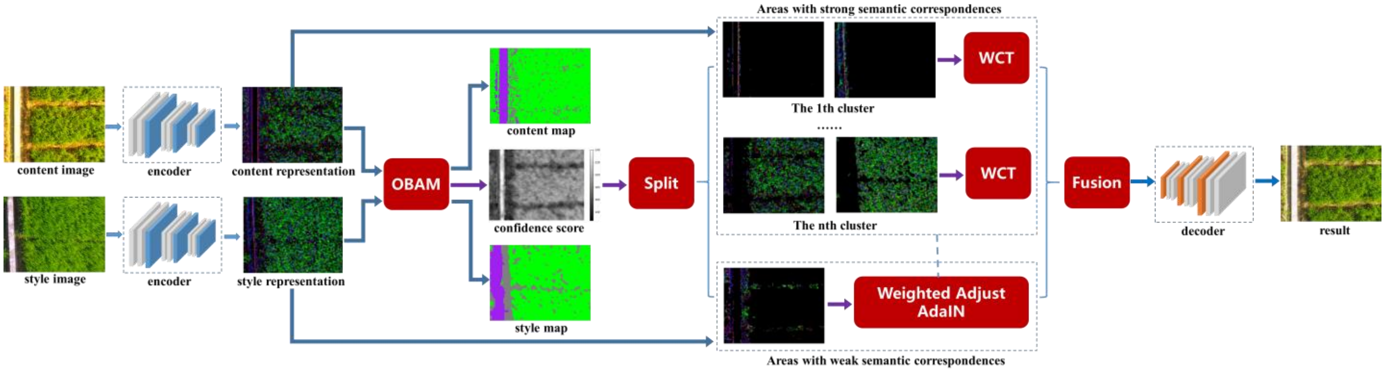

Fig. 1. The overall workflow of our proposed OBAM model

A. Object-based attention mechanism

To address the unrelated disturbance brought by the general attention mechanism, we proposed an object-based attention mechanism (OBAM) to search the pair-wise semantic correspondences in an unsupervised manner, as shown in Fig. 2. Different from the reassembled approaches, the OBAM considers the regional correlations and only responds to the position with the maximum correspondence, which can thus help to avoid the pixel isolation and the disturbance from unrelated pixels.

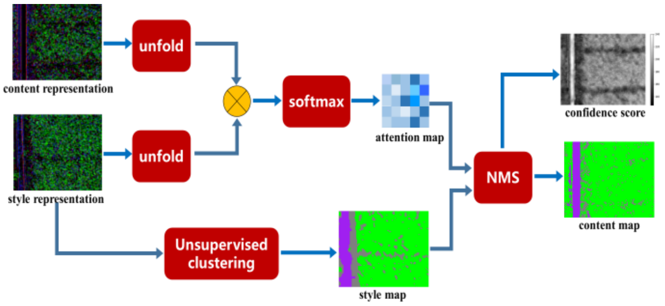

Fig. 2. The architecture of our proposed OBAM

First, we apply unsupervised clustering to the style representation to obtain the style map M_s. For the extracted representation, the neighboring information was considered in the representational space by unfolding patches at each position. Specifically, the unfolding operation gathers the spatial neighbors to each point in the representational space, and the matrix multiplication of the unfolding results builds the quantitative measurement of mutual correlation, as shown in formula (1). Let the size of the original representation be $\mathrm{C} \times \mathrm{H} \times \mathrm{W}$, where C, H, and W represent the channel size, height and width of the representation, respectively. Afterwards, we can obtain the unfolding features with a size of $\mathrm{C} \times \mathrm{P} \times \mathrm{P} \times \mathrm{H} \times \mathrm{W}$, where P denotes the patch size. Later, the unfolding results is reshaped to $(\mathrm{H} \times \mathrm{W}) \times (\mathrm{C} \times \mathrm{P} \times \mathrm{P})$ for the normalized cross-correlation, which can be expressed as:

$$M_A^{ij} := \mathrm{Unfold}(z_c^i) \cdot \mathrm{Unfold}(z_s^j) \quad (1)$$

$$\overline{M_A^{ij}} := \frac{\exp(M_A^{ij})}{\sum_{k=1}^{N} \exp(M_A^{ik})}, where\ N := |z_s| \quad (2)$$

where z_c and z_s denote the representation of the content and style images, and M_A and $\overline{M_A}$ are the attention maps before and after normalization. The symbol of := indicates assignment operation.

After this stage, the attention map $\overline{M_A}$ is obtained with a size of $(\mathrm{H} \times \mathrm{W}) \times (\mathrm{H} \times \mathrm{W})$, where $\overline{M_A^{ij}}$ represents the correlation between the i^{th} element in z_c and the j^{th} element in z_s. Different from the mainstream research on spatial attention, we do not use the attention map $\overline{M_A}$ for feature reassembling. Instead, we propose to use the nonmaximum suppression (NMS) method to address the disturbance of the unrelated representations as follows:

$$t_i := argmax\left(\overline{M_A^{ik}}\right), where\ k := 1,2, \dots, N, and\ N := |z_s| \quad (3)$$

$$M_c^i := M_s^{t_i}, where\ i := 1,2, \dots M, and\ M := |z_c| \quad (4)$$

where t_i refers to the position in style representation z_s with maximum correspondence to the ith element in content representation z_c, and M_c represents the content map. Also, we can have the confidence score S, whose element S_i denotes the maximum correspondence from the style representation to the i^{th} element in content representation z_c, as given by:

$$S_i := max\left(\overline{M_A^{ik}}\right), where\ k := 1,2, \dots, N, and\ N := |z_s| (5)$$

As seen in Fig. 2, the OBAM method drives all pixels of the representations into several clusters, and the output map is similar to the results of the semantic segmentation. Though the output map of OBAM carries no information on specific semantic categories, the content and style map reflect the semantic correspondence between the image-pair, which builds the foundation for accurate color transfer between the homogeneous regions. Also, the confidence score reflects the possibility of semantic alignment for each position. In our research, one threshold value is introduced to split the content representation into homogeneous and heterogeneous regions. The homogeneous regions' confidence score is higher than a given threshold and defined as the regions with strong correspondences. On contrast, the heterogeneous regions' confidence score is lower than the threshold and denoted as the regions with weak correspondences.

B. Wavelet corrected whitening and coloring transforms

Wavelet corrected whitening and coloring transform (wavelet corrected WCT) is applied to perform the color transfer for each cluster in the areas with strong correspondences, as shown in Fig. 3. As seen from Fig. 3, the wavelet corrected WCT generally follows the encoding and decoding architecture of image transform. To better preserve the details in the content images, the downsampling in the encoder is replaced by the wavelet pooling, and the upsampling in the decoder is replaced by the wavelet unpooling.

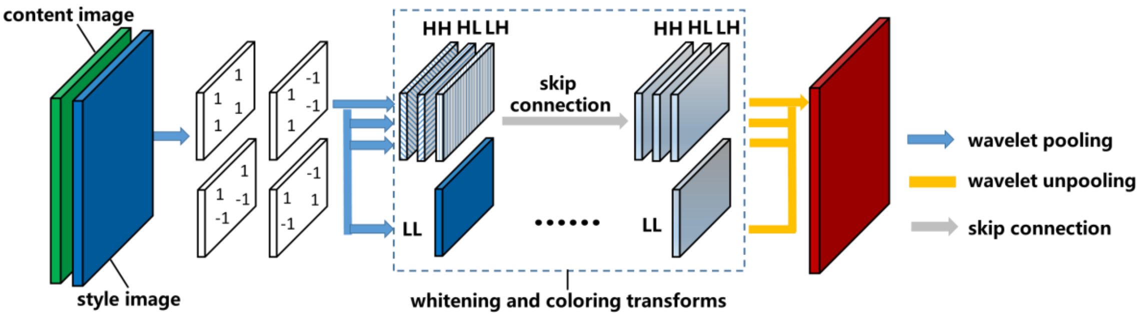

Fig. 3. The structure of the wavelet corrected WCT

Following the definition of the Haar wavelet, the low filter (L) and high filter (H) are defined as:

$$L := \frac{1}{\sqrt{2}}[1\ 1],\ \ H := \frac{1}{\sqrt{2}}[-1\ 1] \quad (6)$$

These filters consist of four kernels for the wavelet pooling: LL^T, LH^T, HL^T, HH^T, which represent respectively the low frequency, vertical, horizontal, and the diagonal edge information. During the encoding process, only the low-frequency information is passed to the next layers, and the high-frequency signals are employed in the decoding process via skip connections. We progressively transform features in a single forward path, and the WCT is applied for color transform between the homogeneous regions at each scale. Let the z_{c-i} and z_{s-i} denote the content and style representation at the i^{th} scale, the WCT can be expressed as:

$$z_{cs-i} := P_{s-i} P_{c-i} z_{c-i} \quad (7)$$

$$P_{c-i} := E_{c-i} \Lambda_{c-i}^{-\frac{1}{2}} E_{c-i}^{\mathrm{T}} \quad (8)$$

$$P_{s-i} := E_{s-i}\Lambda_{s-i}^{-\frac{1}{2}}E_{s-i}^{\mathrm{T}} \quad (9)$$

where Λ_{c-i} and Λ_{s-i} denote the diagonal matrices with the eigenvalues of the covariance matrix $z_{c-i}z_{c-i}^{\mathrm{T}}$ and $z_{s-i}z_{s-i}^{\mathrm{T}}$, and the E_{c-i} and E_{s-i} are the corresponding orthonormal matrices. The z_{cs-i} represents the stylization result at the i^{th} scale, where the decoder will process the last stylization result for image reconstruction.

C. Weight-adjusted AdaIN

It is hard to adjust the color cast for the regions with weak correspondences since there are no appropriate style patterns as reference. To address this problem, we propose the weight-adjusted AdaIN (WAA) for bias estimation in the representation space, as shown in Fig. 4. The feature statistics were computed across each cluster in the areas with strong semantic correspondences. The weighted sum of the feature statistics was used as the bias estimation of the color cast in the areas with weak semantic correspondences, and the weight for each cluster was measured by its valid pixels, as detailed below.

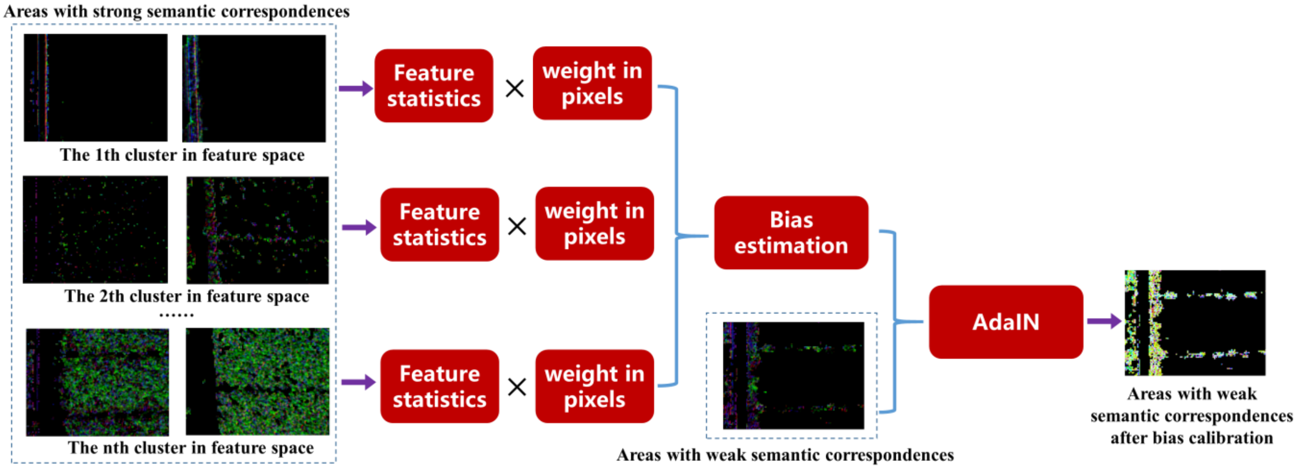
Areas with strong semantic correspondences
The 1th cluster in feature space
The 2th cluster in feature space
......
The nth cluster in feature space
Feature statistics
×
weight in pixels
Feature statistics
×
weight in pixels
Feature statistics
×
weight in pixels
Bias estimation
Areas with weak semantic correspondences
AdaIN
Areas with weak semantic correspondences after bias calibration

Fig. 4. The structure of the weighted adjusted AdaIN

We compute the channel-wise mean and variance in the areas with strong semantic correspondences for each cluster by:

$$\mu_c^i := \frac{1}{\sum_{a=1}^{H}\sum_{b=1}^{W}(M_c^{ab}=i)}\sum_{a=1}^{H}\sum_{b=1}^{W} z_c^{ab}\,(M_c^{ab}=i) \quad (10)$$

$$\sigma_c^i := \sqrt{\frac{1}{\sum_{a=1}^{H}\sum_{b=1}^{W}(M_c^{ab}=i)}\sum_{a=1}^{H}\sum_{b=1}^{W}(z_c^{ab}-\mu_c^i)^2+\varepsilon} \quad (11)$$

where μ_c^i and σ_c^i represent the mean and variance of the i^{th} cluster in the feature space, and H and W denote the height and width of the feature maps, respectively. The symbol of $=$ denotes the numerical equality, and the Boolean operator $(M_c^{ab}=i)$ indicates that the statistics are computed for each cluster. The feature statistics was combined by the weighted average of all clusters, and the weight for each cluster was measured by the ratio of the cluster pixels. The larger cluster in the representation space plays more important role in the bias estimation, which inspire our proposed WAA method as follows:

$$\mu_c := \sum_{i=1}^{n}\frac{\sum_{a=1}^{h}\sum_{b=1}^{w}(M_c^{ab}=i)}{\sum_{i=1}^{n}\sum_{a=1}^{h}\sum_{b=1}^{w}(M_c^{ab}=i)}\mu_c^i \quad (12)$$

$$\sigma_c := \sum_{i=1}^{n}\frac{\sum_{a=1}^{h}\sum_{b=1}^{w}(M_c^{ab}=i)}{\sum_{i=1}^{n}\sum_{a=1}^{h}\sum_{b=1}^{w}(M_c^{ab}=i)}\sigma_c^i \quad (13)$$

where μ_c and σ_c represent the channel-wise mean and variance of the content image, which will be later used in the AdaIN method for color calibration of the areas with weak semantic correspondences.

IV. Experiments

In this section, we will evaluate the performance of our method on the publicly available dataset of UAV remote sensing for precision agriculture: CropUAV [32]. To the best of our knowledge, CropUAV is the only public dataset concerning on the color cast problem of UAV remote sensing in precision agriculture, which drives us selecting this dataset for evaluation. First, we give a detailed description of the evaluated dataset in section IV-A. Next, we present the implementation details on the model architecture, training, and testing in section IV-B. We further demonstrate the effectiveness of the proposed modules through a careful ablation study, given in section IV-C. Finally, we compared the performance of our method with the state-of-the-arts semantic style transfer and attentional style transfer algorithms, as detailed in section IV-D. Moreover, we will release our implementation in http://github.com/huanghsheng/object-based-attention-mechanism.

A. Data Set Description

To evaluate the effectiveness of our proposed method, we conducted comprehensive experiments on the public dataset: CropUAV [32]. CropUAV is a dataset designed for crop monitoring using UAV remote sensing in precision agriculture, and the involved plant species include rice, beans, and cotton. It contains 8850 training images and 6825 validation images, where the image size is 600 × 800, as shown in Table I. The

UAV imagery were captured at a low altitude under the natural field conditions, where the image sequences in one flight present significant color cast, as shown in Fig. 5. It can be seen from Fig. 5 that the color of the crops varies significantly, which may lead to misjudgment of the crop conditions and the following field management.

Table I. Details of the CropUAV dataset

Plant species	dataset partition	dataset size
rice	training set	2050
	validation set	1500
beans	training set	4200
	validation set	2075
cotton	training set	2600
	validation set	3250

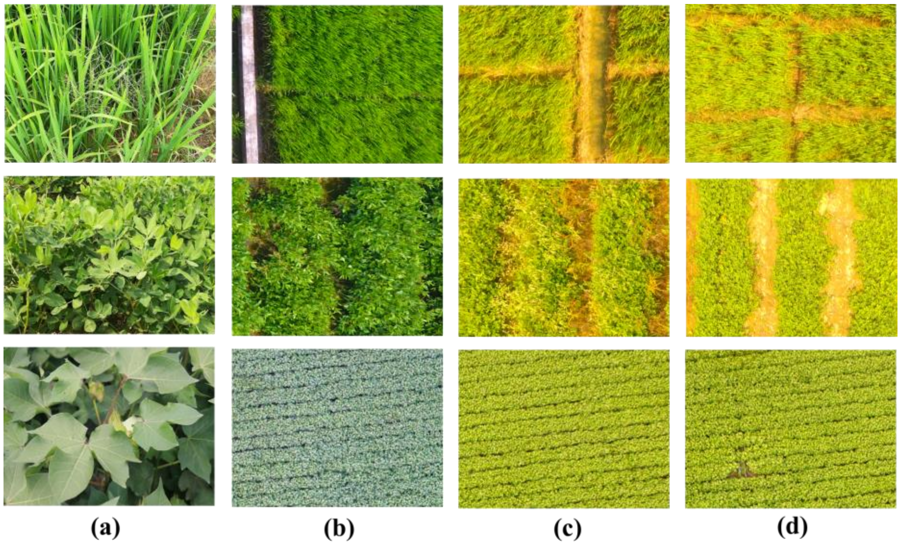

Fig. 5. Demonstration for the color cast of the UAV imagery in the dataset. (a) in field photograph of the crops; (b) UAV imagery with correct color; (c-d) UAV imagery with color cast.

B. Implementation Details

Our color calibration network generally follows the encoder-transfer-decoder architecture. The encoder module is fixed to the first few layers of the pretrained VGG-19 network, where the pooling layers are replaced with wavelet pooling layers. Different from the default configuration from others [18, 19], our encoder only applies the *conv_1* to *conv_3* in the VGG-19 network to better preserve the details in the stylization results. The transfer module contains our proposed OBAM workflow, as discussed in section III. The decoder is the inverse architecture of the encoder; only the wavelet pooling is replaced by wavelet unpooling. The decoder is pretrained on the Microsoft COCO dataset like most researches on style transfer [28-30], and the training loss includes the content loss, style loss, and the L_2 reconstruction loss. The pretraining applied the Adam optimizer with a fixed learning rate of 10^{-3}. To perform the quantitative evaluation of our proposed method, we used the Kullback-Leibler divergence (KL) and Hellinger distance (Hel) to denote the precision of color calibration since these are the general metrics to measure the color precision [33]. Also, we used the gradient difference (M_{grad}) to represent the detail preservation, and employed the HIGRADE-1 [34] to measure the stylization results' image quality. To avoid the disturbance from the difference of semantic distribution, the KL and Hel are only computed across the crop areas. However, the M_{grad} and HIGRADE-1 are measured within the whole image since the detail preserving and image quality are free from the semantic distribution. All experiments were conducted on a computer with a i7 CPU and a NVIDIA RTX 2080 TI GPU.

C. Ablation Study and Analysis

The scientific contributions of this research mainly include the OBAM and WAA methods for accurate color transfer between the target and reference images. We decompose our methodology step by step to reveal the effectiveness of these proposed modules. Table II gives the quantitative results on different plant species. Obviously, the introduction of the OBAM method significantly increases the accuracy of color transfer. We argue that the OBAM method builds the semantic correspondences between the homogeneous regions, which avoid the disturbance of unrelated pixels thus improving the transfer precision. However, the utilization of the OBAM method suffers from semantic mismatch caused by the absence of semantic correspondences, decreasing the color transfer accuracy and perceptual quality. To solve this problem, we propose the WAA method for bias estimation of the areas with weak semantic correspondences. The WAA method estimated the color cast for the areas with weak correspondences by weighting the statistics of the areas with strong correspondences, which eliminate the structural artifacts and calibrate the color cast, as shown within the yellow frames of the last three samples in Fig. 5. From the aspect of efficiency, the OBAM module consumes more 0.17 s for one image, and the utilization of the WAA method contrarily decrease the inference time. We argue that the WAA reduces the number of clusters in color transfer, thus saving more computational time. It can also be seen from Table II that the KL and Hel present slight inconsistency on the evaluation of color precision. We speculate that the KL and Hel belong to asymmetrical measure and symmetrical measure, thus the evaluation on the difference between two probability distribution present slight variation. However, these metrics generally reflect the precision of color calibration, and the evaluation on these metrics shares approximate results.

Table Ⅱ**.** The ablation study for the proposed modules

Plant species	**OBAM**	**WAA**	**KL** ↓	**Hel** ↓	M_{grad} ↓	**HIGRADE-1** ↑	**Time /s** ↓
rice			0.2382	0.2129	**0.0561**	**-0.0322**	**0.3564**
	√		**0.1120**	**0.1487**	0.0785	-0.1251	0.5228
	√	√	0.1710	0.1638	0.0605	-0.0830	0.4668
beans			0.4164	**0.1989**	0.0793	**-0.0852**	**0.3525**
	√		**0.2333**	0.2182	0.0421	-0.1400	0.5166
	√	√	0.2998	0.2394	**0.0286**	-0.1514	0.4712
cotton			0.9562	0.2392	0.0522	-0.4891	**0.3413**
	√		0.0704	**0.1214**	0.0381	-0.5107	0.5163
	√	√	**0.0692**	0.1237	**0.0367**	**-0.4714**	0.4698

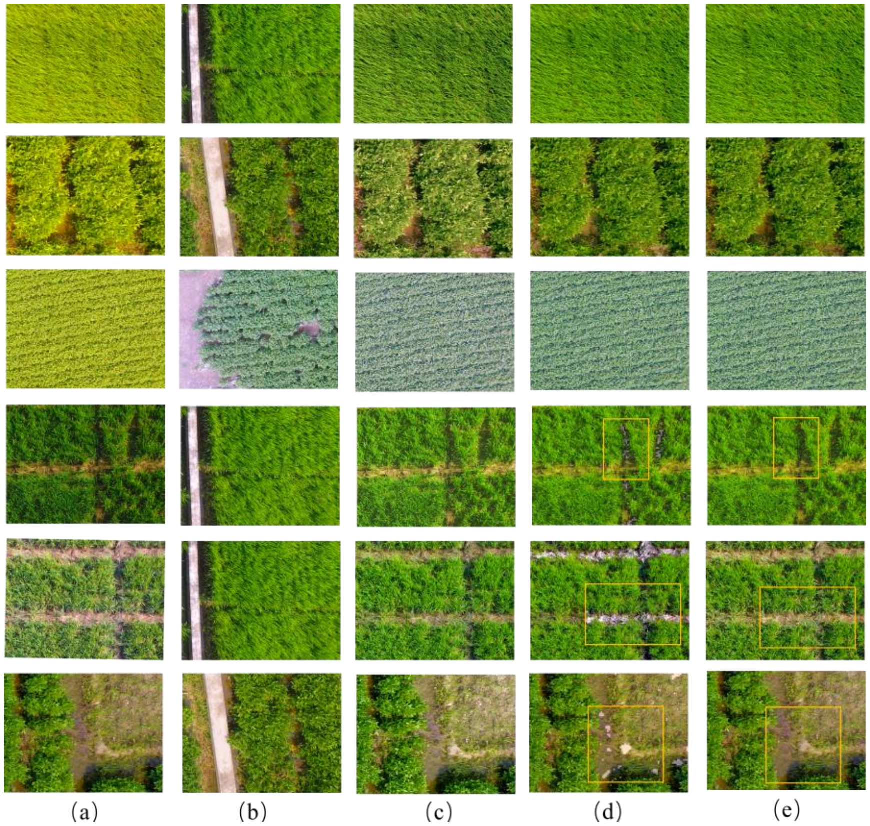

Fig. 6. Visual comparison on ablation study. (a) target images; (b) reference; (c) outputs without OBAM and WAA; (d) outputs with only OBAM; (e) outputs with OBAM and WAA.

Fig. 7 gives a visual explanation on WAA's effectiveness to remove structural artifacts. As seen from Fig. 7, most of the artifacts appear in the regions with low confidence scores, as shown in Fig. 7 (c) and Fig. 7 (d). Though the OBAM searches for the largest correspondence in the style representation, the corresponding features may not be the homogeneous points, thus leading to semantic mismatch. The WAA method gives proximate bias estimation by combining the corresponding clusters with strong correspondences, effectively avoiding the semantic mismatch and removing the unexpected artifacts, as shown in Fig. 7(e).

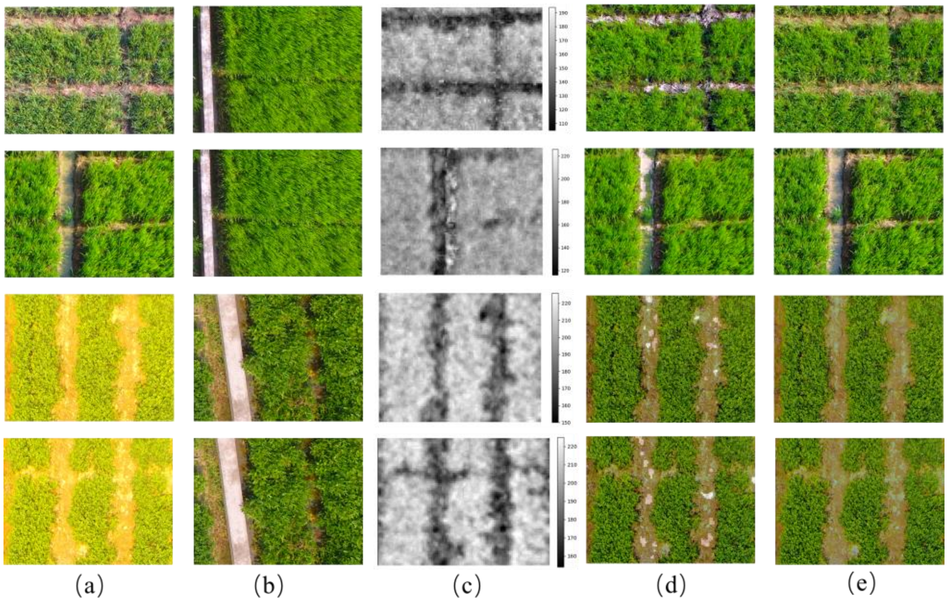

Fig. 7. Visual explanation on the structural artifacts. (a) target images; (b) reference; (c) confidence map of the OBAM; (d) outputs with only OBAM; (e) outputs with OBAM and WAA.

D. Comparison with the State of the Arts

This section compares our method with the semantic style transfer and attentional style transfer algorithms. Qualitative and quantitative comparisons were conducted to demonstrate the superiority of the proposed method.

1) Comparison with the semantic style transfer

We adopted three representative semantic style transfer algorithms for comparison, i.e. Class Based Styling (CBS) [35], Photo whitening and coloring (PhotoWCT) [18], and whitening and coloring v2 (WCT2) [19]. For all the compared methods, a fully convolutional network was employed to perform dense classification on the validation set. Each pixel of the UAV imagery was classified into three classes: crop, others, and background, where the background category indicates the areas outside the investigated fields. The *others* class include many semantic categories such as cement, soil, and mulch, et al. Thus the color transfer accuracy was only evaluated in the crop areas. However, the detail preservation and image quality assessment was reported on the whole image. The published official implementations of these methods were employed for a fair comparison in our experiments.

Table III gives the quantitative comparisons of our method with the semantic style transfer algorithms, where our method achieves competitive results for all the involved plant species. Specifically, the evaluations are conducted for each crop separately, since the crop monitoring tasks generally targets on a specific crop type for one farmer. CBS [35] applied the feed-forward network proposed by Johnson et al. [36] for stylization, resulting in low transfer accuracy from global color transfer workflow. Specifically, CBS [35] ignores the integration of semantic related regions thus achieves best performance in efficiency. However, this method relies on the trained reference image. When the reference image is changed, the whole network has to be trained again. Compared with CBS [35] and PhotoWCT [18], WCT2 [19] demonstrates its superiority in accurate color transfer. We argue that the stylization between the semantically related regions improves the transfer accuracy and photorealism, and the integration of wavelet pooling and unpooling prevents the detail loss. According to the results in Table III, WCT2 [19] achieves best color precision in some cases. However, WCT2 [19] still relies on annotated labels for semantic segmentation, which is hard to be applied to unknown scenarios. In comparison, our method automatically searches the semantic correspondences, and obtains better results without the guidance of semantic maps. Also, the proposed method requires no annotations, which significantly releases the massive manual labor that is required in conventional approaches.

Table III. Quantitative results of our method and the semantic style transfer algorithms

Plant species	Algorithm	KL ↓	Hel ↓	M_{grad} ↓	HIGRADE-1 ↑	Time /s ↓
rice	CBS [35]	0.2728	0.2180	0.1005	**0.1118**	**0.2290**
	PhotoWCT [18]	0.5119	0.3621	0.1309	-0.2295	0.5416
	WCT2 [19]	0.1854	0.1804	0.0980	-0.0142	0.8283
	ours	**0.1710**	**0.1638**	**0.0605**	-0.0830	0.4668
beans	CBS [35]	0.4443	0.3089	0.0431	**0.1038**	**0.2290**
	PhotoWCT [18]	0.3689	0.2514	0.0886	-0.3285	0.5416
	WCT2 [19]	0.3937	**0.1724**	0.0695	-0.1081	0.8283
	ours	**0.2998**	0.2394	**0.0286**	-0.1514	0.4712
cotton	CBS [35]	0.1001	0.1604	0.0616	-0.5357	**0.2290**
	PhotoWCT [18]	0.1072	0.1864	0.0780	-0.7362	0.5416
	WCT2 [19]	0.0914	**0.1173**	0.0617	**-0.3868**	0.8283
	ours	**0.0692**	0.1237	**0.0367**	-0.4714	0.4698

Fig. 8 gives the qualitative results of our method and the semantic style transfer algorithms. For the outputs by CBS [35], the crop areas and the soils are presented in different color spaces, significantly decreasing the perceptual quality of the UAV imagery, as shown in the first two samples in Fig. 8 (c). When the image patterns of the content and style share dissimilarity, color transfer accuracy may decrease as shown in the last three samples in Fig. 8 (c). With the semantic information applied for local color transfer, PhotoWCT [18] and WCT2 [19] increase the color transfer accuracy. However, PhotoWCT [18] suffers from blurring artifacts due to the max pooling operation in feature encoding, as shown in the second sample in Fig. 8(d). WCT2 [19] integrates the wavelet pooling and wavelet unpooling in image encoding and decoding, improving rice's texture details. However, the *others* category in agricultural monitoring often comprises many semantic classes, and the style transfer between this category may produce unsatisfied results due to the semantic difference. It can be seen that PhotoWCT [18] and WCT2 [19] transfer the color of the cement to the soils, decreasing the photorealism of the stylization results, as shown in the first three samples in Fig. 8 (d) and Fig. 8 (e). However, our proposed method can well address this problem, as shown in Fig. 8 (f), thanks to the OBAM module for detecting the semantic difference, and the following WAA method for robust color calibration in the representation space.

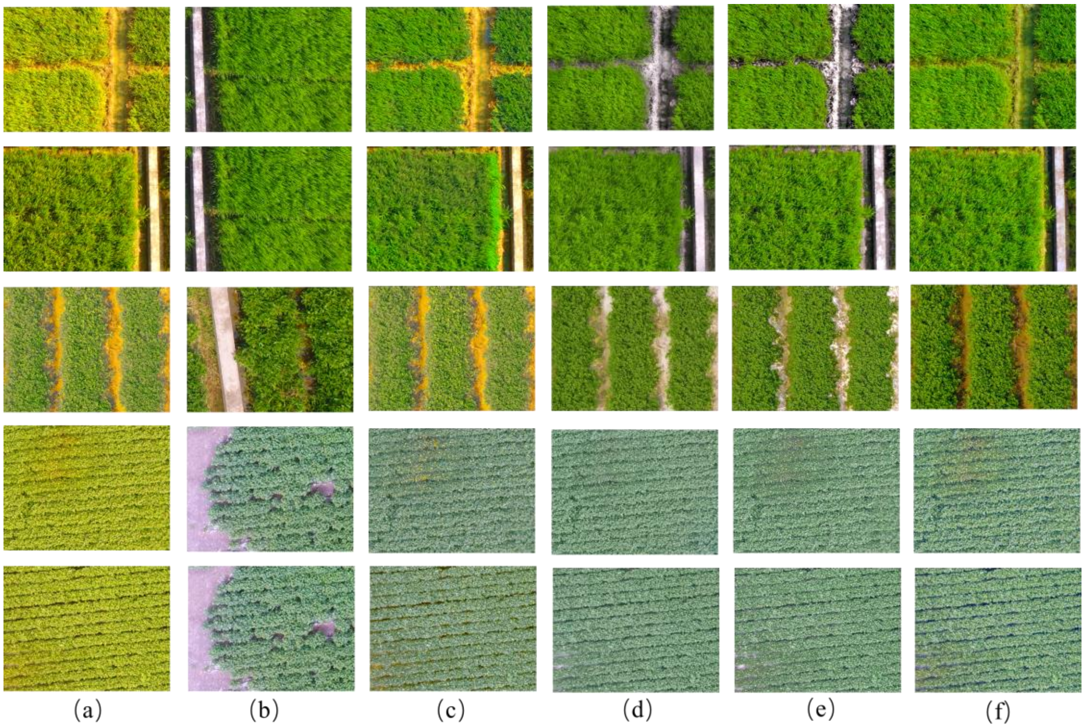

Fig. 8. Qualitative results of our method and the semantic style transfer algorithms. (a) target images; (b) reference; (c) outputs by CBS [35]; (d) outputs by PhotoWCT [18]; (e) outputs by WCT2 [19]; (f) outputs by our method.

It is worth noting that the calibrated results should retain the original imagery's anomaly information during color transfer, as seen in the last two samples of Fig. 8, which are the UAV imagery collected in the cotton fields infected with spider mites, ad detailed in [37]. However, the stylization results by the compared counterparts weaken the symptoms of mite infection as shown in Fig. 8 (c-e). We argue that the compared methods focus only on the color transfer accuracy, which inevitably ignores the anomaly areas presenting color differences with the reference image. In contrast, the stylization of our method retains the anomaly appealing caused by the spider mites, thanks to OBAM’ s capability to process the color calibration for the regions with low semantic correspondences.

2) Comparison with the attentional style transfer

Compared with semantic style transfer, the attentional style transfer requires no labels for training, raising researchers' attention in recent years. We compare our proposed method with the mainstream attentional style transfer approaches, including Avatar-net [26], StyleMixer [28], and IEContraAST [30]. For fair comparison, the published official implementations of the compared methods are utilized to conduct the experiments. Our experiments found that the outputs by Avatar-net [26] and IEContraAST [30] suffer from the blurring artifacts. To address this problem, we only retain the *conv1* to *conv3* in the encoder derived from the VGG-19 network, and the symmetric architectures are preserved in the decoder. All the layers in StyleMixer [28] were retained due to its multi-level feature fusion strategy. According to our experiments, the three compared methods demonstrated poor performance with their default parameters. Therefore, all the compared counterparts were carefully tuned on our training set before evaluation.

Table IV gives the quantitative results of our method and the three compared attentional style transfer algorithms. The accuracy of Avatar-net [26] and StyleMixer [28] is relatively low for the involved plant species. Especially for the cotton, Avatar-net [26] and StyleMixer [28] only obtain 0.5376 and 0.7165 in KL. One possible reason is that the StyleMixer [28] computes the cross correlation and uses correspondence score for feature reassembling. However, the disturbance from the semantically unrelated pixels cannot be avoided in the feature reassembling process. Avatar-net [26] proposed an improved patch matching strategy so that every element in the content representation can find a semantically nearest element in the style representation. However, this workflow ignores the absence of semantic correspondences for some regions of the content image. Under this circumstance, Avatar-net [26] suffered from semantic mismatches and structural artifacts. Also, Avatar-net [26] is too slow to be applied in applications, where the processing for one 600×800 image consumes 3.7654 seconds. We argue that the proposed style decorator performs the patch matching with convolution in high dimensional representation space, which consumes too much GPU memory and execution time. IEContraAST [30] obtains better performance than Avatar-net [26] and StyleMixer [28] in all metrics. We debate that the proposed external loss and contrastive loss help better learn the content and style representation by considering the stylization-to-stylization relations. From the evaluated metrics, IEContraAST [30] obtains competitive performance with our proposed method on some crop types such as cotton. However, IEContraAST [30] is a training-based approach and requires that the reference images should be approximate with the trained templates. In contrast, our proposed OBAM requires no extra training and can be directly incorporated to unknown scenarios. Generally, our method achieves state-of-the-art or close to state of the art performance, especially in the color transfer accuracy and detail preserving. We argue that the proposed OBAM addresses the pixel isolation problem by feature clustering, building a better foundation for searching the semantic correspondences.

Table IV. Quantitative results of our method and the attentional style transfer algorithms

Plant species	**Algorithm**	**KL ↓**	**Hel ↓**	M_{grad} ↓	**HIGRADE-1 ↑**	**Time /s ↓**
rice	Avatar-net [26]	0.1840	0.1954	0.1918	-0.3320	3.7654
	StyleMixer [28]	0.3750	0.2672	0.2293	**0.1794**	0.5432
	IEContraAST [30]	0.2972	0.2661	0.1331	-0.1843	**0.3371**
	ours	**0.1710**	**0.1638**	**0.0605**	-0.0830	0.4668
beans	Avatar-net [26]	0.3173	0.2310	0.1491	-0.5279	3.7654
	StyleMixer [28]	**0.2496**	**0.1611**	0.1973	-0.0519	0.5432
	IEContraAST [30]	0.2662	0.2263	0.0989	**-0.1374**	**0.3371**
	ours	0.2998	0.2394	**0.0286**	-0.1514	0.4712
cotton	Avatar-net [26]	0.5376	0.1877	0.1856	-1.0694	3.7654
	StyleMixer [28]	0.7165	0.2121	0.2274	-0.4932	0.5432
	IEContraAST [30]	0.0783	**0.1131**	0.1215	**-0.1242**	**0.3371**
	ours	**0.0692**	0.1237	**0.0367**	-0.4714	0.4698

Fig. 9 provides the visual comparison of our method with the attentional style transfer algorithms. As seen, Avatar-net [26] performs poorly in photorealism. We argue that the style decorator module matches the soils of the target images to the cement regions in the reference image, leading to significant structural artifacts in the reconstruction process, as shown in Fig. 9 (c). The same problem occurs in the outputs produced by StyleMixer [28]. We debate that StyleMixer [28] applies the weighted reassembling to perform the image reconstruction, thus still suffering from the problem of semantic mismatch. Generally, IEContraAST [30] and our method produce pleasing results for all the plant species, in terms of the color transfer accuracy, detail preserving, and photorealism. However, IEContraAST [30] can only be incorporated into the learning-based methods. When the reference images are too different from the training styles, IEContraAST [30] may generate unsatisfactory results. This limitation is also observed in our research. When the pretrained weights were directly applied for evaluation, IEContraAST [30] demonstrated poor performance in all metrics, especially in detail preserving, as shown in Fig. 10. Semantic mismatches were observed from the first two samples of Fig. 10 (c), and obvious structural artifacts were presented in the last sample. This problem was later addressed by careful fine-tuning on our training set, as shown in Fig. 10 (d). On the contrary, our proposed OBAM and WAA module require no extra training. Specifically, we embed the proposed modules into the WCT2 pretrained on the COCO dataset and use the incorporated model for evaluation on our validation set without extra training. Therefore, our method is superior in zero-shot arbitrary style transfer, enhancing its extension to more generic applications.

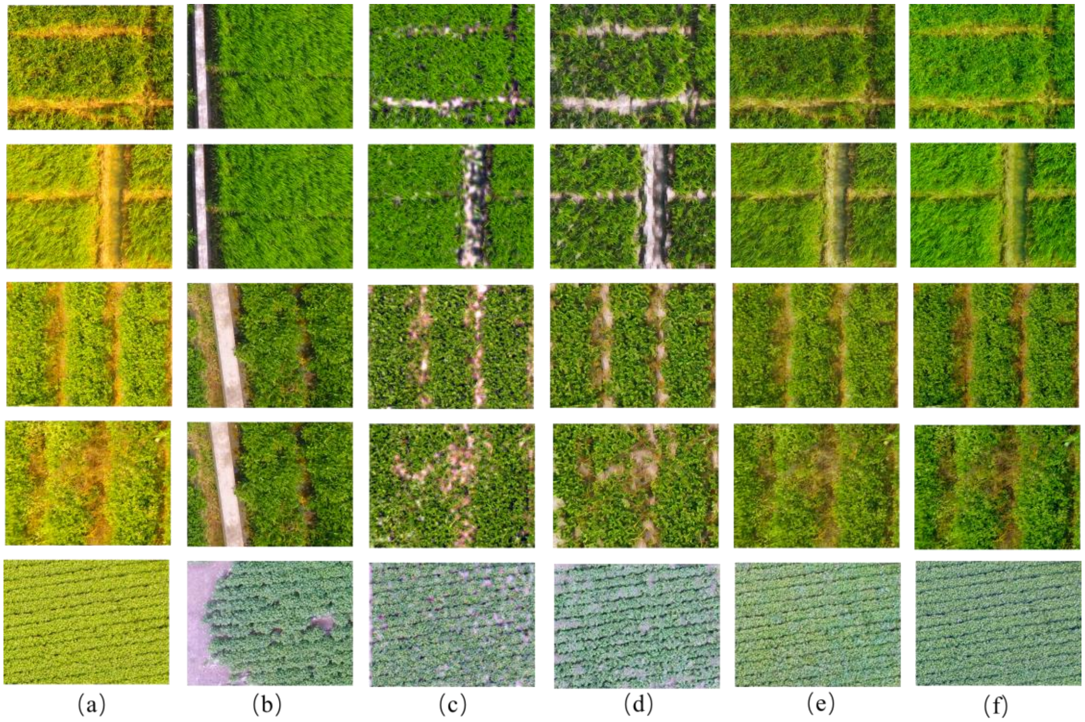

Fig. 9. Qualitative results of our method and the three attentional style transfer algorithms. (a) target images; (b) reference; (c) outputs by Avatar-net [26]; (d) outputs by StyleMixer [28]; (e) outputs by IEContraAST [30]; (f) outputs by our method.

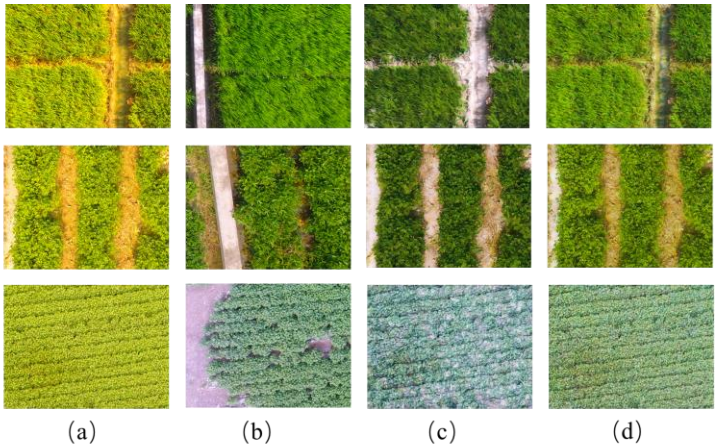

Fig. 10. Comparison on IEContraAST [30] before and after finetuning. (a) target images; (b) reference; (c) outputs by the default parameters; (d) outputs by the finetuned version.

E. Limitations

The main limitation of this work is that the proposed OBAM may be stuck with the semantic mismatch in some cases. Fig.11 presents some failure cases of our approach. When one homogenous object is wrongly split into homogenous and heterogeneous regions, these areas will be processed with different calibration modes thus result in structural artifacts, as shown in the red brackets of the first two samples in Fig. 11. Also, the proposed WAA module applied the statistics results of the areas with strong semantic correspondences as the bias estimation for the areas without semantic reference. When the areas with strong semantic correspondences are too small, the estimation error will be increased, as shown in the third sample in Fig. 11 where the purple bracket indicates the small areas with semantic correspondences. The research to increase the accuracy for the bias estimation will be the direction of our future work.

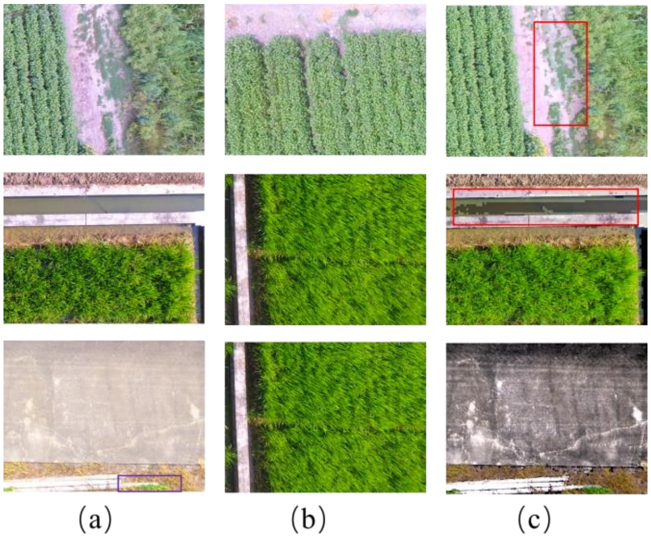

Fig. 11. Failure cases of our approach on color calibration. (a) target images; (b) reference; (c) outputs by our method.

V. Conclusion

In this paper, we proposed a novel OBAM method for color calibration of UAV imagery in precision agriculture. The proposed OBAM method searches the semantic correspondences of the target and reference images via unsupervised clustering and non-maximum suppression, which has successfully suppressed the disturbance from the semantically unrelated elements. To address the absence of semantic correspondences in certain regions, we further propose a WAA module for bias estimation and color calibration. The proposed modules were carefully evaluated through an ablation study, and visual explanations were conducted to exploit the reason for performance boosts. Later, we compared our method with the state-of-the-art semantic style transfer and attentional style transfer algorithms. Qualitative and quantitative results have demonstrated that our method consistently outperformed others on all plant species. Also, our proposed module requires no annotated labels, which can be easily embedded into other color transfer models. The contribution of this work is expected to enrich the research of attention mechanism in style transfer and build a general color calibration framework for UAV remote sensing in precision agriculture and beyond.

References

[1]. S. Sunoj, C. Igathinathane, N. Saliendra, J. Hendrickson, and D. Archer. Color calibration of digital images for agriculture and other applications[J]. ISPRS journal of photogrammetry and remote sensing, 2018, 146: 221-234.

[2]. Mahmoud Afifi and Michael S. Brown. What else can fool deep learning? Addressing color constancy errors on deep neural network performance[C]. in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019: 243-252.

[3]. Jiale Jiang, Qiaofeng Zhang, Wenhui Wang, Yapeng Wu, Hengbiao Zheng, Xia Yao, Yan Zhu, Weixing Cao, and Tao Cheng. MACA: A Relative Radiometric Correction Method for Multiflight Unmanned Aerial Vehicle Images Based on Concurrent Satellite Imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-14.

[4]. Alwaseela Abdalla, Haiyan Cen, Elfatih Abdel-Rahman, Liang Wan, and Yong He. Color calibration of proximal sensing RGB images of oilseed rape canopy via deep learning combined with K-means algorithm[J]. Remote Sensing, 2019, 11(24): 3001.

[5]. Simone Bianco and Claudio Cusano. Quasi-unsupervised color constancy[C]. in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 12212-12221.

[6]. Daniel Hernandez-Juarez, Sarah Parisot, Benjamin Busam, Ales Leonardis, Gregory Slabaugh, and Steven McDonagh. A multi-hypothesis approach to color constancy[C]. in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 2270-2280.

[7]. Karen Panetta, Long Bao and Sos Agaian. Fast Hue-Division-Based Selective Color Transfer[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(9): 2853-2866.

[8]. Yifei Huang, Sheng Qiu, Changbo Wang, and Chenhui Li. Learning Representations for High-Dynamic-Range Image Color Transfer in a Self-Supervised Way[J]. IEEE Transactions on Multimedia, 2021, 23: 176-188.

[9]. Ming Lu, Hao Zhao, Anbang Yao, Feng Xu, Yurong Chen, and Li Zhang. Decoder network over lightweight reconstructed feature for fast semantic style transfer[C]. in Proceedings of the IEEE international conference on computer vision (ICCV), 2017: 2469-2477.

[10]. Zhibo Rao, Mingyi He, Zhidong Zhu, Yuchao Dai, and Renjie He. Bidirectional Guided Attention Network for 3-D Semantic Detection of Remote Sensing Images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(7): 6138-6153.

[11]. Yingxiao Xu, Hao Chen, Chun Du, and Jun Li. MSACon: Mining Spatial Attention-Based Contextual Information for Road Extraction[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-17.

[12]. Graham D. Finlayson, Michal Mackiewicz and Anya Hurlbert. Color correction using root-polynomial regression[J]. IEEE Transactions on Image Processing, 2015, 24(5): 1460-1470.

[13]. Yingying Deng, Fan Tang, Weiming Dong, Haibin Huang, Chongyang Ma, and Changsheng Xu. Arbitrary video style transfer via multi-channel correlation[C]. in Proceedings of the AAAI Conference on Artificial Intelligence, 2021: 1210-1217.

[14]. Songhua Liu, Tianwei Lin, Dongliang He, Fu Li, Meiling Wang, Xin Li, Zhengxing Sun, Qian Li, and Errui Ding. Adaattn: Revisit attention mechanism in arbitrary neural style transfer[C]. in Proceedings of the IEEE/CVF international conference on computer vision (ICCV), 2021: 6649-6658.

[15]. Pei Wang, Yijun Li and Nuno Vasconcelos. Rethinking and improving the robustness of image style transfer[C]. in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 124-133.

[16]. Alex J. Champandard. Semantic style transfer and turning two-bit doodles into fine artworks[J]. arXiv preprint arXiv:1603.01768, 2016.

[17]. Fujun Luan, Sylvain Paris, Eli Shechtman, and Kavita Bala. Deep photo style transfer[C]. in Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), 2017: 4990-4998.

[18]. Yijun Li, Ming-Yu Liu, Xueting Li, Ming-Hsuan Yang, and Jan Kautz. A closed-form solution to photorealistic image stylization[C]. in Proceedings of the European Conference on Computer Vision (ECCV), 2018: 453-468.

[19]. Jaejun Yoo, Youngjung Uh, Sanghyuk Chun, Byeongkyu Kang, and Jung-Woo Ha. Photorealistic style transfer via wavelet transforms[C]. in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019: 9036-9045.

[20]. Ivan Anokhin, Pavel Solovev, Denis Korzhenkov, Alexey Kharlamov, Taras Khakhulin, Aleksei Silvestrov, Sergey Nikolenko, Victor Lempitsky, and Gleb Sterkin. High-resolution daytime translation without domain labels[C]. in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 7488-7497.

[21]. Peihao Zhu, Rameen Abdal, Yipeng Qin, and Peter Wonka. Sean: Image synthesis with semantic region-adaptive normalization[C]. in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 5104-5113.

[22]. Zhuoqi Ma, Jie Li, Nannan Wang, and Xinbo Gao. Image style transfer with collection representation space and semantic-guided reconstruction[J]. Neural Networks, 2020, 129: 123-137.

[23]. Jing Liao, Yuan Yao, Lu Yuan, Gang Hua, and Sing Bing Kang. Visual attribute transfer through deep image analogy[J]. ACM Transactions on Graphics (TOG), 2017, 36(4): 1-15.

[24]. Mingming He, Jing Liao, Dongdong Chen, Lu Yuan, and Pedro V. Sander. Progressive Color Transfer With Dense Semantic Correspondences[J]. ACM Transactions on Graphics, 2019, 38(2): 1-18.

[25]. Tian Qi Chen and Mark Schmidt. Fast patch-based style transfer of arbitrary style[J]. arXiv preprint arXiv:1612.04337, 2016.

[26]. Lu Sheng, Ziyi Lin, Jing Shao, and Xiaogang Wang. Avatar-net: Multi-scale zero-shot style transfer by feature decoration[C]. in Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), 2018: 8242-8250.

[27]. Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization[C]. in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017: 1501-1510.

[28]. Zixuan Huang, Jinghuai Zhang and Jing Liao. Style Mixer: Semantic‐aware Multi‐Style Transfer Network[C]. in Computer Graphics Forum, 2019: 469-480.

[29]. Dae Young Park and Kwang Hee Lee. Arbitrary style transfer with style-attentional networks[C]. in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 5880-5888.

[30]. Haibo Chen, Zhizhong Wang, Huiming Zhang, Zhiwen Zuo, Ailin Li, Wei Xing, and Dongming Lu. Artistic style transfer with internal-external learning and contrastive learning[J]. Advances in Neural Information Processing Systems, 2021, 34: 26561-26573.

[31]. Mohan Zhang, Jing Liao and Jinhui Yu. Deep Exemplar-based Color Transfer for 3D Model[J]. IEEE Transactions on Visualization and Computer Graphics, 2022, 28(8): 2926-2937.

[32]. Huasheng Huang, Aqing Yang, Yu Tang, Jiajun Zhuang, Chaojun Hou, Zhiping Tan, Sathian Dananjayan, Yong He, Qiwei Guo, and Shaoming Luo. Deep color calibration for UAV imagery in crop monitoring using semantic style transfer with local to global attention[J]. International Journal of Applied Earth Observation and Geoinformation, 2021, 104: 102590.

[33]. Mahmoud Afifi, Marcus A. Brubaker and Michael S. Brown, HistoGAN: Controlling Colors of GAN-Generated and Real Images via Color Histograms [C]. in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022: 7937-7946.

[34]. Yifei Huang, Sheng Qiu, Changbo Wang, and Chenhui Li. Learning Representations for High-Dynamic-Range Image Color Transfer in a Self-Supervised Way[J]. IEEE Transactions on Multimedia, 2021, 23: 176-188.

[35]. Lironne Kurzman, David Vazquez and Issam Laradji. Class-based styling: Real-time localized style transfer with semantic segmentation[C]. in Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCV), 2019: 3189-3192.

[36]. Justin Johnson, Alexandre Alahi and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution[C]. in European conference on computer vision (ECCV), 2016: 694-711.

[37]. Huasheng Huang, Jizhong Deng, Yubin Lan, Aqing Yang, Xiaoling Deng, Lei Zhang, Sheng Wen, Yan Jiang, Gaoyu Suo, and Pengchao Chen. A two-stage classification approach for the detection of spider mite-infested cotton using UAV multispectral imagery[J]. Remote Sensing Letters, 2018, 9(10): 933-941.

Huasheng Huang received the M.Sc. degrees in pattern recognition from South China Agricultural University, Guangzhou, China, in 2013, and the Ph.D. degree in agricultural electrification from South China Agricultural University, Guangzhou, China, in 2019.

He is currently an associate professor of the college of computer sciences, Guangdong Polytechnic Normal University, Guangzhou, China. His research interests include image processing, UAV remote sensing and intelligent agriculture.

Yu Tang (Member, IEEE) received the B.Sc. degree in electrical engineering from the Civil Aviation University of China, Tianjin, China, in 2005, and received the M.S. degree in optical engineering and the Ph.D. degree in microelectronics and solid state electronics from the South China University of Technology, Guangzhou, China, in 2008 and 2013, respectively. From 2013 to 2015, he worked as a postdoctoral researcher at the South China university of technology, Guangzhou, China.

He is currently a Professor and the Dean of the Academy of Interdisciplinary Studies, Guangdong Polytechnic Normal University, Guangzhou, China. His current research interests include artificial intelligence, image/video processing, and IntelliSense and autonomous control.

Zhiping Tan received the B.Sc. degree in electronic engineering from Xiangtan University and the Ph.D. degree from the College of Mathematics and Informatics, South China Agricultural University. He has published over ten research articles. His research interests include evolutionary optimization, Deep learning and machine vision.

Jiajun Zhuang received the B.Sc. and M.Sc. degrees in measurement and control technology from Guangdong University of Technology in 2007 and 2010, respectively, and the Ph.D. degree in computer science from South China University of Technology in 2013.

He is currently an associate professor with the College of Mathematics and Data Science, Zhongkai University of Agriculture and Engineering, Guangzhou, China. His current research interest includes computer vision, machine learning and agricultural engineering.

Chaojun Hou received his B. Sc. and M. Sc. degrees from South China University of Technology, Guangzhou, China, in 2001 and 2004, respectively. He received his Ph.D. degree from the Sun Yat-Sen University, Guangzhou, China, in 2009.

He is currently an associated professor in the College of Mathematics and Data Science, Zhongkai University of Agriculture and Engineering, Guangzhou, China. His current research interests include agriculture engineering and artificial intelligence in agriculture.

Weizhao Chen received his M. Sc. degree from school of information engineering at Guangdong University of Technology, Guangzhou, China, in 2020.

His research interests are Machine Learning, Deep learning and Hyperspectral image processing. He currently focuses on the application of machine learning in feature extraction of hyperspectral image.

Jinchang Ren (Senior Member, IEEE) received the B.Eng. degree in computer software, the M.Eng. degree in image processing, and the D.Eng. degree in computer vision from Northwestern Polytechnical University, Xi’an, China, in 1992, 1997, and 2000, respectively, and the Ph.D. degree in electronic imaging and media communication from the University of Bradford, Bradford, U.K., in 2009.

He is a Reader with the Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, U.K. With over 300 scientific articles, his research interests include image processing, machine learning, hyperspectral imaging, remote sensing, and big data analytics.