UCloudNet: A Residual U-Net with Deep Supervision for Sky/Cloud Image Segmentation

Yijie Li, Hewei Wang, *Student Member, IEEE,* Shaofan Wang, *Member, IEEE,* Yee Hui Lee, *Senior Member, IEEE,* Muhammad Salman Pathan, and Soumyabrata Dev, *Member, IEEE*

Abstract—**Recently, there is a growing interest in the use of ground-based sky cameras for meteorology applications. The segmentation of sky/cloud images obtained from these sky cameras offers us great insights into the cloud coverage computation and understanding various atmospheric events. Most of the existing research in sky/cloud image segmentation are based on traditional computer vision methods, including the use of color features and gradient variation in the images. With the development of deep learning architectures, convolutional neural networks (CNNs) have been widely used for sky/cloud image segmentation. However, CNNs require larger training time and higher number of epochs to converge. This limits its widespread use on onboard sky camera's computing systems. In this paper, we introduce a residual U-Net with deep supervision for cloud segmentation which provides better performance than other CNN-based approaches, and with less training consumption. Our proposed model achieves the highest F-score value as compared to the state-of-the-art cloud segmentation techniques. In the spirit of reproducible research, the model code, dataset, and results of the experiments in this paper are available at: https://github.com/Att100/UCloudNet.**

Index Terms—**deep learning, cloud segmentation, deep supervision, U-Net, residual network.**

I. Introduction

CLOUD information analysis is necessary and important for meteorology research. The distribution or form of the cloud can reflect specific information which can be used to learn the weather and generate advanced prediction. Generally, cloud images are taken by the meteorological satellite at the near-earth orbit, but in recent years, ground-based sky cameras [1], [2] have been widely used because of its better temporal and spatial resolutions. Several datasets of optical RGB images captured by these sky cameras are released to the community, including SWIMSEG [3], SWINSEG [4], and SWINySEG [5]. To better extract the cloud information from those images, the deep learning method with fully convolution network (FCN) is widely used for cloud segmentation which consists a series of encoder and decoder. However, this early design without extra modified structure is hard to aggregate feature from the first few layers. Additionally, the lack of short-cut skip connection in encoder will still result in the difficulty of training such deep convolution neural network. In this paper, we introduce the residual U-Net model with deep supervision for cloud segmentation task, called UCloudNet. UCloudNet consists a series of convolution block with residual connection which enhance feature aggregation ability of original U-Net by including more feature map fusion operations. The experiments prove that our proposed UCloudNet can achieve better performance with less training time and iterations than previous approaches.

The main contributions of our work are,

- We introduced a novel CNN-based approach, the UCloudNet, which have better performance than previous approaches.
- We adopt deep supervision in our proposed UCloudNet which significantly reduce the training time consumption.
- We evaluate our UCloudNet on three different benchmark datasets and prove its effectiveness.

II. Related Works

To solve the sky/cloud image segmentation task, several techniques were developed, which can be broadly divided into traditional computer vision methods [6]–[8] and deep learning methods [5], [9], [10]. Some of those traditional methods use color features, pre-defined fixed convolution filters, and gradient of pixels, for example, Dev *et al.* (2014) [7] use principal component analysis (PCA) and fuzzy cluster to evaluate the color model aims to capture the greatest color variance between cloud ad sky. Those approaches can successfully capture the overall distribution of sky part, but many details are lost during the extraction procedure, which yield a poor segmentation accuracy and the generated binary masks of sky/cloud images are not well defined and do not conform the image boundaries. With the advent of deep learning methods to sky/cloud image segmentation task, the generation of binary cloud masks have been improved. Dev *et al.* (2019) [5] introduce a novel approach, CloudSegNet, which use a standard FCN structure which first down-sample the original images to compress information into high-dimensional feature maps and then perform a series of up-sample operation to recover the segmentation results which significantly improve the overall quality, and the details of boundary are more accurate than previous methods. In the same year, Dev *et al.* [10] introduce another approach for multi-label sky/cloud

Manuscript received Month Day, 2022; revised Month Day, 2022; accepted Month Day 2022.

Y. Li and H. Wang are with the Beijing-Dublin International College, Beijing University of Technology, Beijing 100124, China (e-mail: yijie.li@ucdconnect.ie, hewei.wang@ucdconnect.ie).

S. Wang is with Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China (email: wangshaofan@bjut.edu.cn).

Y. H. Lee is with Nanyang Technological University Singapore, Singapore (e-mail: EYHLee@ntu.edu.sg).

M. S. Pathan is with Maynooth University, Ireland (e-mail: muhammad.pathan@mu.ie).

S. Dev is with the School of Computer Science, University College Dublin (e-mail: soumyabrata.dev@ucd.ie).

segmentation which consider the label of each sky/cloud image as three classes, including thin cloud, thick cloud and sky. They trained a multi-classes U-Net to generate prediction which allows researcher to perform more accurate analysis on three-classes segmentation map.

III. Architecture

Our UCloudNet is based on the U-Net [11] structure which contains a series of decoders and encoders with channels concatenation in each stage. To compare with the original U-Net structure, we use a hyper-parameter k to control the parameters amount and inspired by He *et al.* [12], we add residual connection in each convolution block in encoder which is helpful for training the deeper layers. As for the training strategy, we use deep supervision [13] to support the training process. The architecture of our proposed model is shown in Fig. 1. Our model contains a series of 'Double Convolution' blocks, 'Down Sample' blocks, and 'Up Sample' blocks. We explain these blocks in the following sections.

A. Double Convolution Block (DCB)

Figure 2 describes the structure of the double convolution block. The 'Double Convolution Block' contains two group of layers which include an original 3x3 convolution layer, a ReLU6 activation layer, and a batch-normalization layer. This group of layers is called 'BasicConv2d' in our implementation. The structure of DCB module is different in encoder and decoder, we only apply residual connection in the DCB in encoder while the structure of DCB in decoder is just a simple stack of 'Conv-Bn-ReLu6' groups. The different design of DCB can result from the short-cut concatenation in U-Net structure. The U-Net structure already has short connection between encoder and decoder, and therefore there is no need to apply short-cut in decoder again.

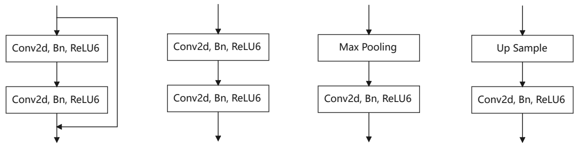

Fig. 2: The structure of 'Double Convolution Block' in encoder, decoder, 'Down Sample Block', and 'Up Sample Block' (from left to right).

B. Down Sample Block (DSB) and Up Sample Block (UPB)

The 'Down Sample Block' include a Max-pooling layer and a 'BasicConv2d' block. The 'Up Sample Block' is a combination of an Up-sample layer and a 'BasicConv2d' block. The UPB module receive the output of a previous DCB and the output of this block will be concatenated with output of another DCB in encoder with same feature map size.

C. Deep supervision

The deep supervision training strategy can effectively support the training of deep neural network and improve the regularization ability. In our implementation, we use two additional auxiliary loss branch to enable the deep supervision. These two loss branches are located at stages with $1/2$ and $1/4$ output size, as shown at the top left of Fig. 1. In order to perform the binary cross entropy loss, we interpolate the target map into $1/2$ and $1/4$ of original size. Additionally, a group of weight values is considered in our total loss so that the significance of high-resolution prediction map is larger than low-resolution prediction map.

D. Parameters control and loss function

The number of parameters of our model can be controlled by a hyper-parameter k. The number of the filters of DCB in encoder can be specified by $k * 2^s$ while the configuration of DCB in decoder is $k * 2^{3-s}$. The filters of the convolution layer in "Down Sample" block can be calculated by $k * 2^{s+1}$ and the filters number of "Up Sample" block is $k * 2^{4-s}$, while $s = 0, 1, 2, 3$.

We use binary cross entropy as loss function and total loss function can be represented as follow:

$$L(p, y) = -\frac{1}{N} * \sum_{i=0}^{N} y_i * \log p_i + (1 - y_i) * \log (1 - p_i) \quad (1)$$

$$L_{total}(p, y) = L(p, y) + 0.4 * L(p_2, y_2) + 0.2 * L(p_4, y_4) \quad (2)$$

IV. Datasets & Configurations

A. Dataset

We obtain the cloud segmentation data set from Singapore Whole sky Nychthemeron Image SEGmentation Database (SWINySEG) which contains 6078 day-time cloud images and 690 night-time cloud images. These images are captured in Singapore using a calibrated camera. In our experiments, we train our model with three different splits which are day-time images (augmented SWIMSEG), night-time images (augmented SWINSEG), and full SWINySEG dataset.

B. Training Configurations

We use PaddlePaddle to implement our model and perform the training on a single NVIDIA Tesla V100-SXM2 16GB GPU. We split the data set with a ratio of 80%: 20% for training and testing. We set training batch-size to 16 and training for 100 epochs in all experiments. As for the optimizer, we use Adam with initialized learning rate of 1e-3, beta1 to 0.9, beta2 to 0.999, and epsilon with 1e-8. We also use exponential learning-rate decay with gamma equal to 0.95 after each training epoch. We evaluate our model on the test set after every 5 epochs.

V. Experiments & Results

We conduct experiments under the training setup that introduced in previous section. On augmented SWIMSEG (day-time) and augmented SWINSEG (night-time) data set, we train our model with $k = 2$ and $k = 4$ with auxiliary loss and learning rate decay. On full SWINySEG data set, we train our model with four different configurations, $k = 2$ and $k = 4$ with auxiliary loss and learning rate decay, $k = 4$ with learning rate decay, and $k = 4$ without additional training strategy.

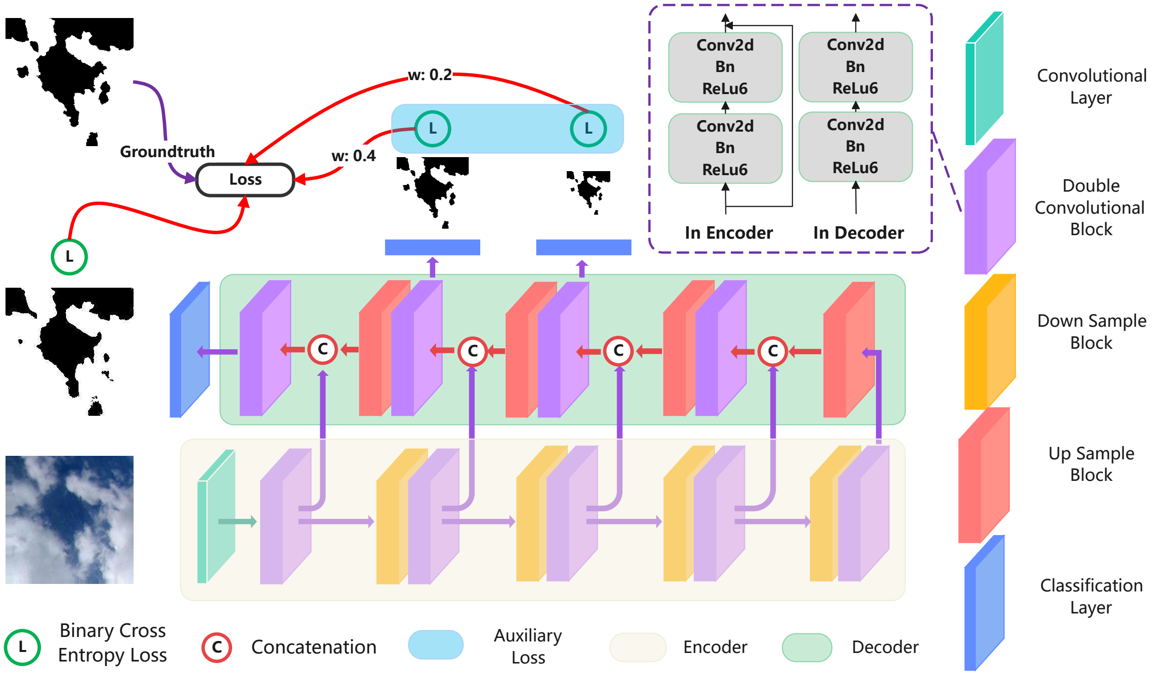

Fig. 1: The architecture of the UCloudNet model. The procedure between the output of model and the segmentation mask has been omitted in this figure.

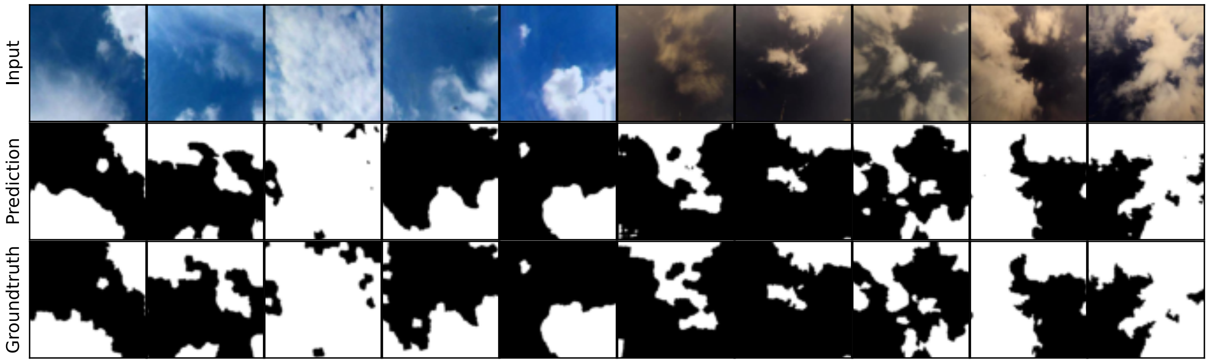

Fig. 3: Results of cloud segmentation for day-time (1-6 columns) and night-time (7-12 columns).

TABLE I: Comparison of UCloudNet with other cloud segmentation methods

Dataset	Method	Precision	Recall	F-measure	Error-rate
SWINySEG (day) (augmented SWIMSEG)	Dev et al. 2014 [7]	0.89	0.92	0.89	0.09
	Long et al. [6]	0.89	0.82	0.81	0.14
	Li et al. [14]	0.81	**0.97**	0.86	0.12
	CloudSegNet [5]	0.92	0.88	0.89	0.07
	Souza et al. [15]	**0.99**	0.53	0.63	0.18
	UCloudNet(k=2)+aux+lr-decay	0.90	0.93	0.92	0.07
	UCloudNet(k=4)+aux+lr-decay	0.92	0.94	**0.93**	**0.06**
SWINySEG (night) (augmented SWINSEG)	Dev et al. 2017 [4]	0.94	0.74	0.82	0.13
	Yang et al. 2009 [16]	**0.98**	0.65	0.76	0.16
	Yang et al. 2010 [8]	0.73	0.33	0.41	0.37
	Gacal et al. [17]	0.48	**0.99**	0.62	0.50
	CloudSegNet [5]	0.88	0.91	0.89	0.08
	UCloudNet(k=2)+aux+lr-decay	0.92	0.94	0.93	0.06
	UCloudNet(k=4)+aux+lr-decay	0.95	0.95	**0.95**	**0.04**
SWINySEG (day+night)	CloudSegNet [5]	0.92	0.87	0.89	0.08
	UCloudNet(k=2)+aux+lr-decay	0.90	0.92	0.91	0.08
	UCloudNet(k=4)	0.92	0.90	0.91	0.08
	UCloudNet(k=4)+lr-decay	0.91	0.94	0.92	0.07
	UCloudNet(k=4)+aux+lr-decay	**0.92**	**0.94**	**0.93**	**0.06**

A. Metrics

In our experiments, we evaluate our model with four widely-used metrics: precision, recall, F-measure, and error-rate. F-measure is usually used to describe the overall performance of a model which is equal to the harmonic mean of precision and recall, $\frac{2 \times Precision \times Recall}{Precision + Recall}$. Precision, can be expressed as $\frac{TP}{TP+FP}$, recall, is equal to $\frac{TP}{TP+FN}$, and error-rate can be expressed as $\frac{FP+FN}{P+N}$. Additionally, we also use PR curve with 256 thresholds to observe the performance of our proposed model.

B. Quantitative Analysis

Quantitative evaluation results of our methods are shown in Table 1, which show the precision, recall, F-measure, and error-rate of our proposed UCloudNet with other methods on different data sets. On full SWINySEG data set, UCloudNet (k=4) with deep supervision and learning rate decay have the best performance on all the four metrics while the UCloudNet (k=4) with only learning rate decay have the second best performance which can prove that deep supervision with auxiliary loss can improve our model performance. On day-time images, the overall performance is better than others. As for the night-time images, our model have the lowest error-rate amongst all methods.

C. Qualitative Analysis

Qualitatively, we use several day-time and a night-time images as our evaluation samples. We use these images as input and perform a threshold with p=0.5 on the sigmoid output, shown in Fig. 3. We observed our binary prediction maps and compare them with input images and ground-truths, it is proved that our proposed model can accurately extract the overall feature and generate binary prediction correctly. However, there still exists some negative phenomenons, for example, the lack of accuracy on segmentation on edges. Besides, we evaluate our proposed model with PR (Precision-Recall) curve, shown in Fig. 4. In order to receive a clearer and more precise results, we set 256 thresholds in total to generate the curve, the area under the curve is bigger, the overall performance of the model is better.

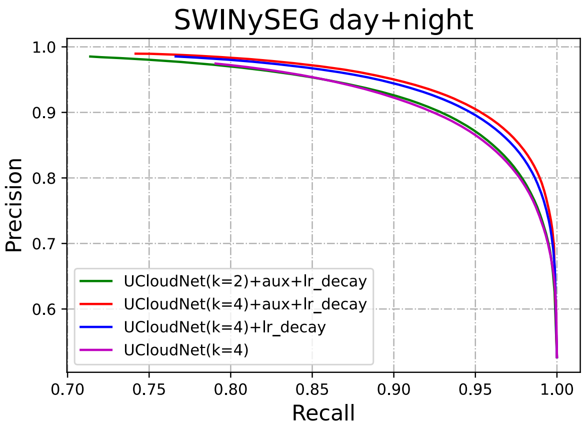

Fig. 4: PR curve of our proposed model with different training configuration on full SWINySEG ground-based cloud segmentation data set.

Additionally, we evaluate the training status of our proposed model qualitatively by observe curves of final loss together with auxiliary loss branches, shown in Fig. 5. The loss curves show that the loss of the final output converge much faster than the loss of x2-down-sample loss and x4-down-sample loss in the early 2500-iterations, but the tendency tend to be the same after it, which can prove that the auxiliary loss is helpful to the training of deep convolution neural net work at the beginning. From another point of view, the loss can achieve a low and stable level in less than 10,000 iterations, and then keep dropping slowly until it finally converge which prove that the speed and ability of fitting of our proposed model is considerable.

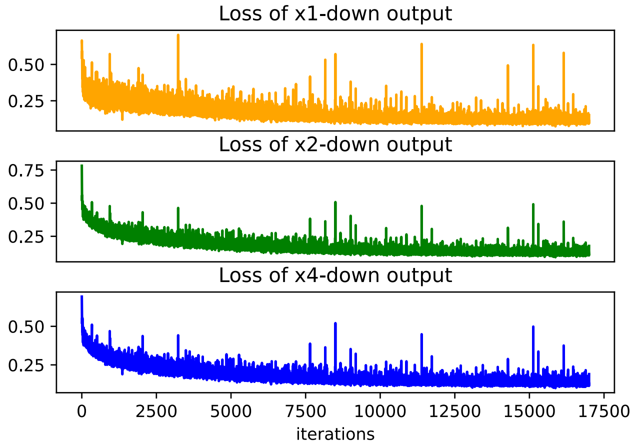

Fig. 5: Loss curve of the final output and auxiliary outputs.

VI. Conclusion and Future works

In this paper, we introduce a residual U-Net with deep supervision for cloud-sky segmentation. We train our model with different configurations on various splits of SWINySEG dataset. Our proposed method achieves better performance as compared to the other methods and our experiments prove that deep supervision with auxiliary loss can gain better performance. Additionally, our model only need less than 17500 iterations (100 epochs with batch-size 16) to converge which can significantly save the training consumption, as compared to other deep learning methods. In the future, we intend to focus on the new structure design with fewer parameters and lower training time.

Acknowledgment

This research was conducted with the financial support of Science Foundation Ireland under Grant Agreement No. 13/RC/2106_P2 at the ADAPT SFI Research Centre at University College Dublin. ADAPT, the SFI Research Centre for AI-Driven Digital Content Technology, is funded by Science Foundation Ireland through the SFI Research Centres Programme.

References

[1] M. Jain, I. Gollini, M. Bertolotto, G. McArdle, and S. Dev, “An extremely-low cost ground-based whole sky imager,” in *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2021, pp. 8209–8212.

[2] S. Dev, F. M. Savoy, Y. H. Lee, and S. Winkler, “Design of low-cost, compact and weather-proof whole sky imagers for high-dynamic-range captures,” in *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2015, pp. 5359–5362.

[3] S. Dev, Y. H. Lee, and S. Winkler, “Color-based segmentation of sky/cloud images from ground-based cameras,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 1, pp. 231–242, 2016.

[4] S. Dev, F. M. Savoy, Y. H. Lee, and S. Winkler, “Nighttime sky/cloud image segmentation,” in *Proc. IEEE International Conference on Image Processing (ICIP)*. IEEE, 2017, pp. 345–349.

[5] S. Dev, A. Nautiyal, Y. H. Lee, and S. Winkler, “Cloudsegnet: A deep network for nychthemeron cloud image segmentation,” *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 12, pp. 1814–1818, 2019.

[6] C. N. Long, J. M. Sabburg, J. Calbó, and D. Pagès, “Retrieving cloud characteristics from ground-based daytime color all-sky images,” *Journal of Atmospheric and Oceanic Technology*, vol. 23, no. 5, pp. 633–652, 2006.

[7] S. Dev, Y. H. Lee, and S. Winkler, “Systematic study of color spaces and components for the segmentation of sky/cloud images,” in *Proc. IEEE International Conference on Image Processing (ICIP)*. IEEE, 2014, pp. 5102–5106.

[8] J. Yang, W. Lv, Y. Ma, W. Yao, and Q. Li, “An automatic groundbased cloud detection method based on local threshold interpolation,” *Acta Meteorologica Sinica*, vol. 68, no. 6, pp. 1007–1017, 2010.

[9] M. Jain, C. Meegan, and S. Dev, “Using GANs to augment data for cloud image segmentation task,” in *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2021, pp. 3452–3455.

[10] S. Dev, S. Manandhar, Y. H. Lee, and S. Winkler, “Multi-label cloud segmentation using a deep network,” in *2019 USNC-URSI Radio Science Meeting (Joint with AP-S Symposium)*. IEEE, 2019, pp. 113–114.

[11] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Proc. International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[12] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[13] L. Wang, C.-Y. Lee, Z. Tu, and S. Lazebnik, “Training deeper convolutional networks with deep supervision,” *arXiv preprint arXiv:1505.02496*, 2015.

[14] Q. Li, W. Lu, and J. Yang, “A hybrid thresholding algorithm for cloud detection on ground-based color images,” *Journal of atmospheric and oceanic technology*, vol. 28, no. 10, pp. 1286–1296, 2011.

[15] M. P. Souza-Echer, E. B. Pereira, L. Bins, and M. Andrade, “A simple method for the assessment of the cloud cover state in high-latitude regions by a ground-based digital camera,” *Journal of Atmospheric and Oceanic Technology*, vol. 23, no. 3, pp. 437–447, 2006.

[16] Q. Yang, L. Tang, W. Dong, and Y. Sun, “Image edge detecting based on gap statistic model and relative entropy,” in *Proc. Sixth International Conference on Fuzzy Systems and Knowledge Discovery*, vol. 5. IEEE, 2009, pp. 384–387.

[17] G. F. B. Gacal, C. Antioquia, and N. Lagrosas, “Ground-based detection of nighttime clouds above manila observatory (14.64° n, 121.07° e) using a digital camera,” *Applied Optics*, vol. 55, no. 22, pp. 6040–6045, 2016.