2024 IEEE International Conference on Robotics and Automation (ICRA 2024)
May 13-17, 2024. Yokohama, Japan

CrackNex: a Few-shot Low-light Crack Segmentation Model Based on Retinex Theory for UAV Inspections

Zhen Yao[1], Jiawei Xu[1], Shuhang Hou[1] and Mooi Choo Chuah[1]

***Abstract*— Routine visual inspections of concrete structures are imperative for upholding the safety and integrity of critical infrastructure. Such visual inspections sometimes happen under low-light conditions, e.g., checking for bridge health. Crack segmentation under such conditions is challenging due to the poor contrast between cracks and their surroundings. However, most deep learning methods are designed for well-illuminated crack images and hence their performance drops dramatically in low-light scenes. In addition, conventional approaches require many annotated low-light crack images which is time-consuming. In this paper, we address these challenges by proposing CrackNex, a framework that utilizes reflectance information based on Retinex Theory to learn a unified illumination-invariant representation. Furthermore, we utilize few-shot segmentation to solve the inefficient training data problem. In CrackNex, both a support prototype and a reflectance prototype are extracted from the support set. Then, a prototype fusion module is designed to integrate the features from both prototypes. CrackNex outperforms the SOTA methods on multiple datasets. Additionally, we present the first benchmark dataset, LCSD, for low-light crack segmentation. LCSD consists of 102 well-illuminated crack images and 41 low-light crack images. The dataset and code are available at https://github.com/zy1296/CrackNex.**

I. Introduction

Cracks are common defects on pavement and in concrete structures. Overloading, structural changes, and environmental hazards may accelerate these deteriorations, causing a significant safety risk [1]. Therefore, regular inspection of roads and bridges to identify damage and repair defects is essential to maintain building and traffic safety. In recent years, a variety of deep-learning algorithms [2]–[10] have been proposed.

However, these algorithms exhibited limited effectiveness in real-world scenarios due to variations in lighting conditions, the presence of shadows, and other factors [11]–[13]. In real-life scenarios, there exist numerous instances of low-light conditions where distinguishing cracks becomes challenging. For example, cracks on the underside of bridge piers, in tunnel walls, and in historical buildings are in remote or hard-to-reach areas that have limited natural light. Engineers often rely on artificial lighting, which may not provide optimal visibility and may cause information loss. Thus, computer vision-based low-light crack segmentation is necessary for safety inspections.

*This work was supported by National Science Foundation Grant CPS 1931867.

[1]All authors are with the Department of Computer Science and Engineering, P.C. Rossin College of Engineering and Applied Science, Lehigh University, Bethlehem, PA 18015, USA. {zhy321, jix519, shh420, mcc7}@lehigh.edu

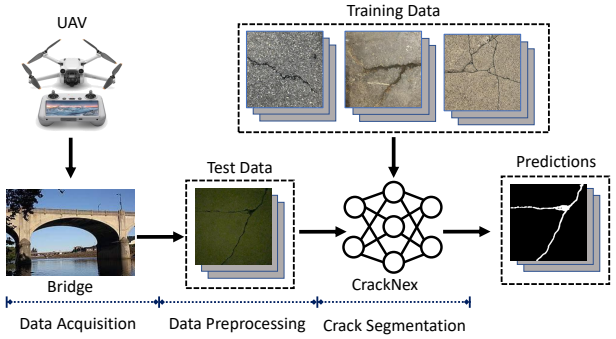

Fig. 1: System overview for UAV inspections

Low-light crack segmentation is challenging because most deep-learning methods were trained on normal-light crack images captured under well-lit conditions [14]–[16]. In unfavorable low-light conditions, the segmentation performance drops significantly, suffering from both low image contrast and ambiguity of object boundaries. Therefore, we introduce the Retinex Theory [17] to address this challenge. It, rooted in human color perception, posits that the observed color image can be decomposed into reflectance and illumination. Reflectance represents the intrinsic attributes of captured objects, which remain consistent under different light conditions. Illumination describes the luminance values present on objects. By using a pre-trained Decompose Network [18], we can estimate the reflectance features and help the model learn a unified and illumination-invariant representation.

Additionally, deep learning is essentially data-driven while collecting such low-light crack images and making high-quality annotations at pixel level are not only time-consuming but also prone to human errors. To tackle this challenge (scarcity of annotated low-light data), we utilize the few-shot segmentation method. The increasingly popular few-shot learning is a promising direction to address the limitation of insufficient data by training models to generalize to new classes with a small number of examples (or shots) [19]. This is particularly valuable in situations where acquiring abundant data is difficult. Unlike common deep-learning approaches, few-shot segmentation has better generalization ability which makes it more adaptable to new or unseen classes [20]–[25]. In few-shot crack segmentation, our goal is to train a model using a sufficient number of normal-light crack images and during inferring, this trained model can be used to segment low-light cracks given a few labeled low-light crack images, as illustrated in Figure 1.

979-8-3503-8457-4/24/$31.00 ©2024 IEEE

In this work, we propose a reflectance-guided few-shot low-light crack segmentation network, CrackNex. It leverages reflectance to enhance the learning of contrast and recover lost details in low-light images. In addition to extracting the support prototype from the support set, we generate a complementary reflectance prototype from the reflectance features. The prototype fusion module (PFM) is proposed to integrate the support prototype with the reflectance prototype and it is capable of uncovering the inner connection between two prototypes with attention weights learned through the co-attention mechanism. Moreover, an Atrous Spatial Pyramid Pooling (ASPP) module [26] is applied to improve the features extracted by the CNN backbone in a multi-scale manner.

In summary, our contributions to this paper include:

- To the best of our knowledge, we propose the first few-shot method for low-light crack segmentation.
- We introduce reflectance information from Retinex Theory and propose a novel reflectance-guided network, CrackNex. Our work highlights a new direction for low-light segmentation. Reflectance indicates the intrinsic properties of different objects and can help distinguish cracks from other non-crack regions.
- We conduct experiments to evaluate our CrackNex model on two crack segmentation datasets and demonstrate its effectiveness using standard segmentation metrics. Compared to several SOTA models, our CrackNex achieves the SOTA performance on these 2 datasets.
- We present a new crack segmentation dataset, LCSD, with both well-illuminated and low-light crack images for the benefit of the research community.

II. Related works

A. Few-Shot Learning

Few-shot learning, the task of training models to recognize and generalize from a limited number of examples, has garnered significant attention in recent years due to its applicability in various domains. Existing methods are mainly meta-learning and metric learning.

Several meta-learning approaches [27]–[31] have been proposed to learn transferable knowledge from diverse learning tasks, leading to substantial advancements in the field. Elsken [32] combined meta-learning with gradient-based Neural Architecture Search (NAS) [33] methods and designed a meta-learning framework capable of customizing the meta-architecture to task-specific architectures. Baik [34] proposed a novel framework to learn a task-adaptive loss function through two meta-learners, employing two distinct meta-learners: one responsible for learning the loss function and another for learning parameters for the loss function.

Metric learning [35] leverages distance measurement to optimize the distance or similarity between the images and regions. Fan [36] proposed a novel self-support network by leveraging self-support matching to solve the appearance discrepancy problem. Okazawa [37] proposed a novel few-shot segmentation approach that effectively enhanced the distinction between the target class and closely resembling classes, yielding improved separation performance. Our work is inspired by the metric-based approach to generate better prototypes utilizing reflectance information.

B. Crack Segmentation

Semantic segmentation, a pixel-level image classification task, is essential for understanding and interpreting visual data. The deep-learning method [2], [3], [11], [12], [38]–[40] has shown promising results in the crack segmentation task.

Early works rely on conventional semantic segmentation models. Liu [41] applied U-Net for pavement crack segmentation and proposed an unmanned aerial system for UAV inspections [42]–[45]. Liu [5] extended the U-Net architecture by incorporating additional convolutional layers to develop a pavement crack segmentation method. Sarmiento [46] used another successful segmentation model, DeepLabv3, to segment pavement cracks.

Recently, Transformers and attention mechanisms have also been widely used for crack segmentation. Wang [11] designed a segmentation model that utilizes a hierarchical Transformer as the encoder and integrates a top-down structure. Xiang [47] introduced a dual encoder–decoder model by using both transformers and CNNs to achieve precise segmentation of crack images.

III. Our proposed low-light crack segmentation model

A. Problem Formulation

Assessing structural integrity and identifying potential damages of infrastructure such as bridges or old buildings is a critically important task for repairs that can be carried out in-time to ensure the safety of human lives. However, sometimes, limited budgets or lack of workers may prevent such task from being carried out as frequently as one wishes. With the recent emergence of affordable drones, such inspections can be conducted effectively without much cost. In Fig.1, we show an example of how Unmanned Aerial Vehicles (UAV) can be used to inspect the lateral side and underside of bridges.

B. Background

Few-shot segmentation aims to generate pixel-level segmentation predictions of novel classes when only a limited number of annotated labels are available. The setup is defined in an N-ways-K-shot format, where N represents the number of classes, and K indicates the number of support images needed for a query image. Our crack segmentation task involves learning to predict 1-way-1-shot and 1-way-5-shot segmentations. In the 1-shot setting, the model uses one single support image as a reference, and in the 5-shot setting, it utilizes 5 support images to generate predictions. The foreground refers to cracks, while the background is non-crack regions.

We use SSP [36] as our baseline few-shot segmentation model. Fan et al. proposed this self-support matching

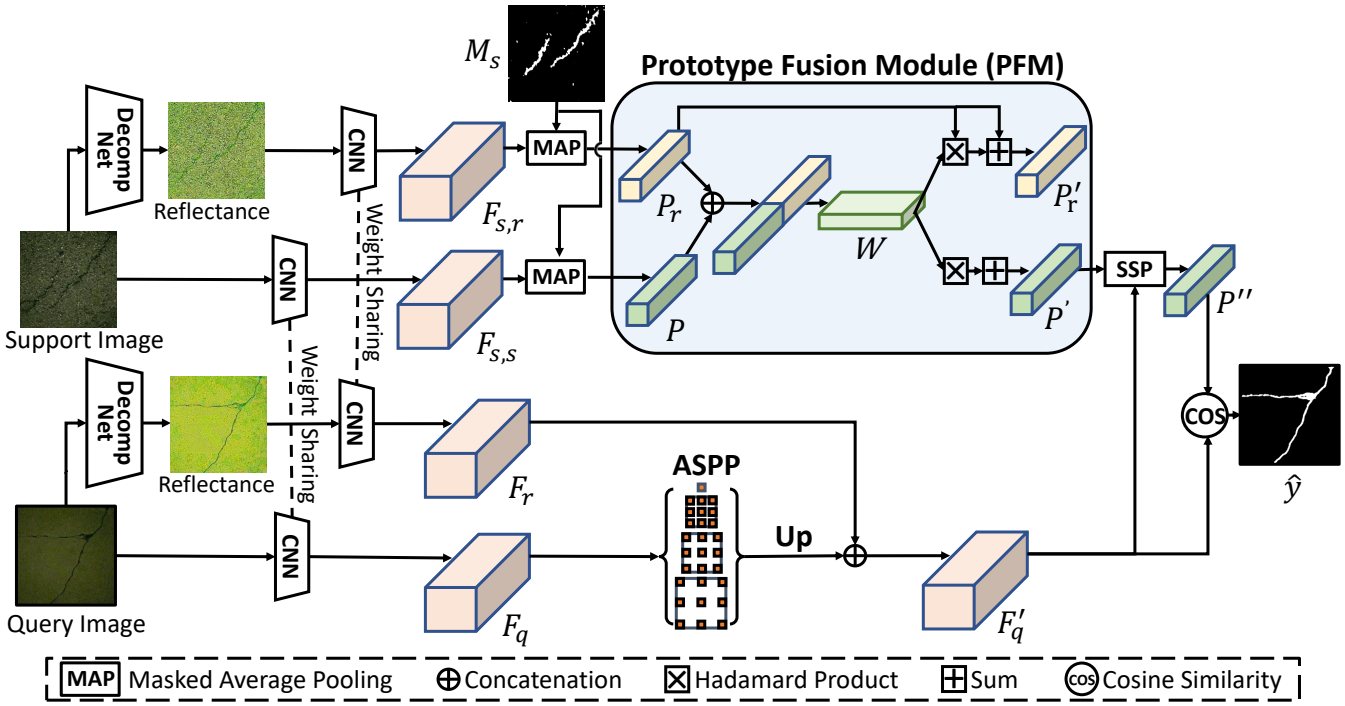

Fig. 2: Illustration of CrackNex: We generate reflectance features on query image and support images respectively by using Decompose Net. Afterwards, we generate the support and reflectance prototypes and update both prototypes by using the Prototype Fusion Module. Meanwhile, reflectance query features are utilized as low-level features in the Atrous Spatial Pyramid Pooling (ASPP) module to preserve details. Finally, we further update the support prototype in the Self-Support Prototype (SSP) module to perform matching with query features and calculate the loss.

framework, which utilized query features to generate self-support prototypes. It addresses the intra-class appearance discrepancy problem in few-shot segmentation by effectively reducing the gap between support prototypes and query features. Such capability is useful for our work.

The training set contains well-illuminated crack images and we represent the normal-light crack as base class $\mathbf{C_{train}}$. Then, we evaluate the trained model using low-light crack images (novel class $\mathbf{C_{test}}$). Following a previous work [48], an episode-based sampling strategy is employed during both training and evaluating. Specifically, every sampled episode $e_i = \{\mathbf{S_i}, \mathbf{Q_i}\}$ of each class $\mathbf{C}$ consists of the support set and the query set. The support set $\mathbf{S_i} = \{(\mathbf{I_s^k}, \mathbf{M_s^k}), k \in \{1, \ldots, K\}\}_i$ consists of a collection of support images, where $\mathbf{I_s^k}$,$\mathbf{M_s^k}$ are the k^{th} support image and its masks, and the query set $\mathbf{Q_i} = \{(\mathbf{I_q}, \mathbf{M_q})\}_i$ refers to a collection of images for which the model needs to perform segmentation, where $\mathbf{I_q}$ is the query image and its ground truth mask $\mathbf{M_q}$.

C. Architecture Overview of CrackNex

Our model consists of a Decompose Net, a Prototype Fusion module (PFM), an ASPP module [26], and an SSP module as illustrated in Fig. 2. Our unique contribution includes utilizing reflectance information to improve segmentation performance and designing the PFM module to fuse extracted support features and reflectance features. The Prototype Fusion Module is described in Section III-D and the ASPP module is described in Section III-E.

We first use a Decompose Net from RetinexNet [18] pre-trained on the LOL dataset [18] to generate reflectance images on both support and query images. The Decompose Net is responsible for separating an input image into a reflectance image and an illumination image. The reflectance features can be further used to highlight object boundaries and capture details, e.g., color, texture, and surface characteristics, leading to more accurate edge and robust segmentation predictions.

We then apply two CNN backbones pre-trained with ImageNet [49] to extract feature maps $\mathbf{F_q} \in \mathbb{R}^{H/8 \times W/8 \times C}$ and $\mathbf{F_r} \in \mathbb{R}^{H/8 \times W/8 \times C}$ for the query image and query reflectance image, respectively. The image features are integrated with the reflectance features through the ASPP module at multiple scales. The output of the ASPP module is updated query features $\mathbf{F'_q} \in \mathbb{R}^{H/4 \times W/4 \times C}$.

Another two backbones are also applied to generate features on support images and support reflectance images. Note that the two backbones shared weights with the former backbones in pairs. Afterwards, extracted support features $\mathbf{F_{s,s}} \in \mathbb{R}^{H/8 \times W/8 \times C}$ and support reflectance features $\mathbf{F_{s,r}} \in \mathbb{R}^{H/8 \times W/8 \times C}$, along with ground truth mask $\mathbf{M_s}$, are fed into the masked average pooling layer to generate the support prototype $\mathbf{P} \in \mathbb{R}^{1 \times 1 \times C}$ and reflectance prototype $\mathbf{P_r} \in \mathbb{R}^{1 \times 1 \times C}$, respectively. By using the proposed Prototype Fusion Module, support prototype $\mathbf{P}$ is integrated with the

reflectance prototype $\mathbf{P_r}$ through the co-attention mechanism.

Updated support prototype $\mathbf{P'} \in \mathbb{R}^{1\times1\times C}$ is then fed into the SSP module with updated query features $\mathbf{F'_q}$. The output is augmented prototype $\mathbf{P''} \in \mathbb{R}^{1\times1\times C}$.

Finally, following the SSP setting, we compute the cosine distance and estimate a similarity map between the augmented prototype $\mathbf{P''}$ and query features $\mathbf{F'_q}$ to generate the final predictions $\hat{\mathbf{y}} \in \mathbb{R}^{H\times W\times1}$:

$$\hat{\mathbf{y}} = \text{softmax}(\text{cosine}(\mathbf{P''}, \mathbf{F'_q})) \tag{1}$$

D. Prototype Fusion Module

A Prototype Fusion Module (PFM) is introduced to interactively fuse the support prototype $\mathbf{P}$ and reflectance prototype $\mathbf{P_r}$. This fusion is achieved through the co-attention mechanism. Specifically, it learns attention weights from both prototypes and integrates the weights into the prototype pair. By using our proposed Prototype Fusion Module, we effectively fuse the features extracted from different representations that carry complementary information. Such fusion results in better enhanced images in terms of improved image quality and visual perception.

Given the prototype $\mathbf{P}$ and the reflectance prototype $\mathbf{P_r}$, the concatenated features of $\mathbf{P}$ and $\mathbf{P_r}$ are fed to a convolutional layer. Next, we apply normalization on the features and use two fully connected layers to learn the attention weights as follows:

$$\mathbf{W} = \text{sigmoid}(f_2(\sigma(f_1(\mathbf{X})))) \tag{2}$$

where $\mathbf{X} \in \mathbb{R}^{1\times1\times2C}$ represents the concatenated prototype features after normalization, f represents fully connected layers and σ represents the activation function.

Then, we update the support and reflectance prototypes using the following equations:

$$\mathbf{P'} = (1 + \alpha\mathbf{W} \otimes \mathbf{P}) \tag{3}$$

$$\mathbf{P'_r} = (1 + \alpha\mathbf{W} \otimes \mathbf{P_r}) \tag{4}$$

where α is a learnable parameter and $\otimes$ represents the Hadamard product.

E. ASPP Module

We apply the Atrous Spatial Pyramid Pooling (ASPP) module based on the work of DeepLabV3 [26]. The ASPP module has atrous convolutions at multiple dilation rates and therefore, captures contextual information at various scales, addressing both local and global contexts. By incorporating multi-scale features, the ASPP module helps the network distinguish between objects of varying sizes and complex scenes with diverse textures.

In our work, the ASPP module is employed to capture multi-scale information from high-level query features $\mathbf{F_q}$. The resulting feature map is then upsampled and concatenated with the reflectance features $\mathbf{F_r}$ (referred as low-level features). The reflectance features help in preserving edge details, which is essential for accurately separating cracks and backgrounds. Finally, the concatenated features $\mathbf{F'_q}$ are fed to the SSP module as the second input to update the support prototype $\mathbf{F'_r}$. It is also used in generating the final matching predictions.

F. Loss

We train our model for the final prediction under supervision:

$$L_{seg} = \text{BCE}(\hat{\mathbf{y}}, \mathbf{M_q}) \tag{5}$$

where BCE is the binary cross-entropy loss, $\hat{y}$ is the final segmentation prediction and $\mathbf{M_q}$ is the ground truth label of the query image. L_{seg} ensures that the predictions are consistent with the ground truth label.

To further facilitate the SSP matching procedure, we apply the self-support loss mentioned in SSP [36] to measure the support and reflectance prototypes:

$$L_s = \text{BCE}(\text{cosine}(\mathbf{P'}, \mathbf{F_{s,s}}), \mathbf{M_s}) + \text{BCE}(\text{cosine}(\mathbf{P'_r}, \mathbf{F_{s,r}}), \mathbf{M_s}) \tag{6}$$

We apply the same procedure to the query features to introduce the query self-support loss L_q:

$$L_q = \text{BCE}(\text{cosine}(\text{MAP}(\mathbf{F'_q}), \mathbf{F'_q}), \mathbf{M_q}) \tag{7}$$

where MAP is the Masked Average Pooling layer used to generate a prototype on query features.

Finally, we train our model by optimizing all aforementioned losses jointly:

$$L = L_{seg} + \lambda_1 L_s + \lambda_2 L_q \tag{8}$$

where λ_1 = 1.0, λ_2 = 0,2 are the loss weights.

IV. Experiments

In this section, we describe the experiments we conduct to compare CrackNex and SOTA few-shot segmentation models using two datasets, namely (a) ll_CrackSeg9k and (b) LCSD datasets. Our results show that our model design achieves better performance than state-of-the-art models. We also provide a detailed analysis of our design features via several ablation studies.

A. Datasets

1) **ll_CrackSeg9k***:* The CrackSeg9k dataset [50] is a popular crack-related dataset that researchers used. We select 9000 crack images from CrackSeg9k as our training set and another 1500 crack images as our test set. Since 1500 images in the test set are considered normal light images, we use Restormer [51] pre-trained on LDIS [52] dataset to convert them into synthetic low-light images

2) **LCSD***:* To evaluate our method under real-world lowlight conditions, we additionally collect our own crack dataset, LCSD, with 102 well-illuminated crack images as the training set and 41 low-light crack images as the test set within the Lehigh University campus. All images are taken by iPad Pro 1st generation and are resized to 400×400 for efficiency. We further annotate each crack image pixel-wise and generate a binary label.

TABLE I: Baseline comparisons on the ll_CrackSeg9k and LCSD dataset in terms of mIOU↑

Method	Backbone	CrackSeg9k		LCSD	
		1-shot	5-shot	1-shot	5-shot
VAT [53]	ResNet50	54.32	57.45	54.28	56.53
MLC [54]		56.54	58.72	55.48	57.41
SSP [36]		60.42	64.25	56.41	63.30
CrackNex (Ours)		**63.00**	**69.66**	**63.85**	**65.17**
VAT [53]	ResNet101	59.83	61.27	55.25	59.24
MLC [54]		56.73	62.99	57.18	58.11
SSP [36]		56.45	65.29	56.61	63.16
CrackNex (Ours)		**65.90**	**70.59**	**66.10**	**68.82**

B. Implementation Details

For the backbone CNN, We adopt the ResNet50 and ResNet101 [55] pre-trained on ImageNet-1K dataset [49]. We train the entire framework using SGD optimizer [56] with the 0.9 momentum. The initial learning rate is 10e-3 and decayed by 10 times every 2,000 iterations. Our network is trained on one single NVIDIA TITAN RTX GPU for 6000 iterations with a batch size of 4. Both images and masks are augmented with random horizontal flipping while the evaluation is performed on the original image.

For comparison with our proposed scheme, we additionally evaluate the performance of several SOTA methods on the LCSD dataset.

C. Quantitative Results

In terms of the evaluation metrics, we use the popular mean Intersection-over-Union (mIOU↑) to evaluate our model under 1-shot and 5-shot settings. We evaluate performance on both ll_CrackSeg9k and LCSD benchmarks. The main results are presented in the Table. I where we compare the performance of CrackNex with other state-of-the-art methods using the ll_CrackSeg9k and LCSD datasets. We test SOTA methods (VAT [53], MLC [54] and SSP [36]) using their default settings. From the table, we see that CrackNex achieves an mIOU of 63.00 and 69.66 respectively under 1-shot and 5-shot settings using the ResNet50 backbone and 65.90 and 70.59 using the ResNet101 backbone [55]. It outperforms SOTA methods on the ll_CrackSeg9k dataset.

We additionally compare the performance of CrackNex with other state-of-the-art methods on the LCSD dataset. From the table, we found that CrackNex achieves an mIOU of 63.85 and 65.17 respectively under 1-shot and 5-shot settings using the ResNet50 backbone and 66.10 and 68.82 using the ResNet101 backbone [55]. It outperforms other SOTA models by a large margin.

D. Qualitative Results

To better analyze our proposed model, we visualize several segmentation results from the LCSD dataset, as shown in Fig. 3. We compare the qualitative results of our method with two SOTA models, SSP [36] and MLC [54] using their default settings. We can see that CrackNex generates more accurate boundaries and more discriminative cracks compared with existing SOTA approaches, which demonstrates the effectiveness of CrackNex.

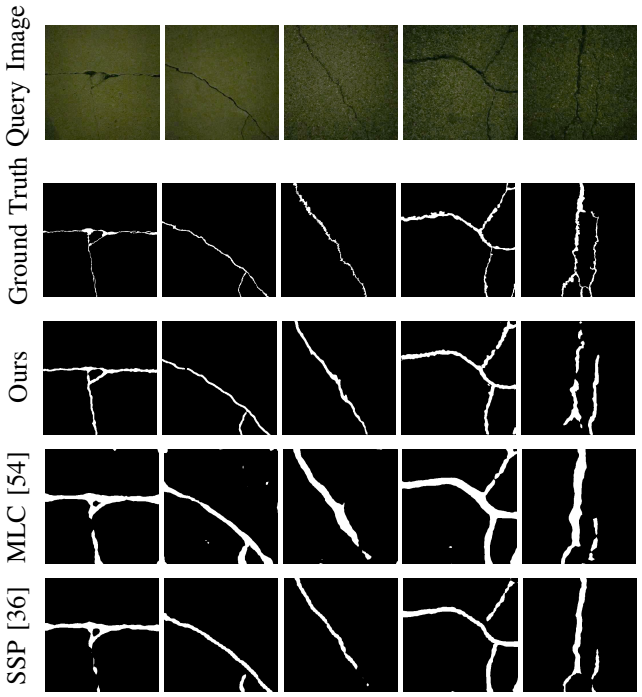

Fig. 3: Qualitative results on LCSD dataset. Zoom in for details. Images have been brightened to improve their visibility.

TABLE II: CrackNex compared with data-driven crack segmentation models on the ll_CrackSeg9k dataset

Method	mIOU↑
DDRNet [57]	69.23
DeepLabV3 [26]	71.16
STDCSeg [58]	70.98
HrSegNet-B16 [59]	71.32
HrSegNet-B32 [59]	72.45
CrackNex (Ours)	70.59

To further evaluate the UAV inspection system in real-world scenarios, we use a drone to conduct the inspections of a nearby bridge built in 1924. As shown in Fig. 4, the segmentation results are accurate and demonstrate the effectiveness of our method.

E. Compared with data-driven crack segmentation models

We additionally compare our model with several SOTA data-driven crack semantic segmentation models on the ll_CrackSeg9k dataset. The main results are presented in the Table. II. We test SOTA methods (DDRNet [57], DeepLabV3 [26], STDCSeg [58] and HrSegNet [59]) using their default settings. From the table, we see that CrackNex achieves mIOU of 70.59 under 5-shot settings and achieves comparable results compared with these data-driven crack segmentation models. Note that these data-driven methods require a large amount of labeled low-light crack images to achieve

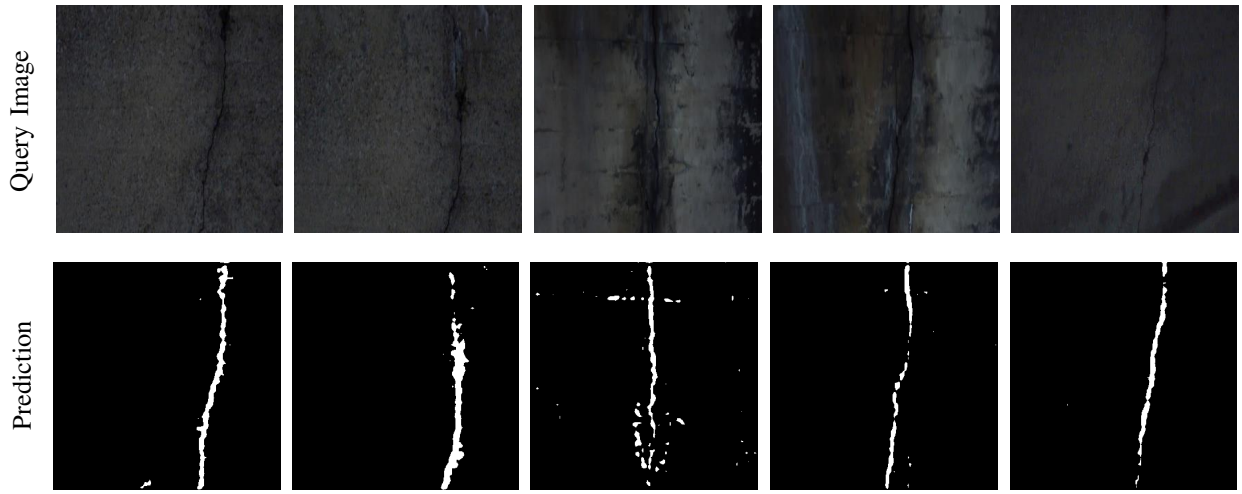

Fig. 4: Qualitative results on UAV-based real-site low-light bridge cracks. Images have been brightened to improve their visibility.

TABLE III: Ablation study of adding different components on LCSD dataset

Reflectance features	PFM Module	ASPP module	mIOU↑ 1-shot	mIOU↑ 5-shot
			56.61	63.16
✓			63.33	65.79
✓	✓		65.06	67.37
✓	✓	✓	**66.10**	**68.82**

optimal performance, whereas our model doesn't need to be trained on low-light crack images.

F. Ablation Study

We further perform ablation studies on the LCSD dataset using the ResNet101 backbone [55] to investigate the contribution of key technical components in our method.

Specifically, we evaluate four variants of CrackNex: (i) baseline architecture, (ii) adding reflectance features, (iii) utilization of Prototype Fusion Module (PFM), and (iv) employment of all the designed components. The results of the ablation study are summarized in Table III.

The second row of Table III investigates the effectiveness of reflectance features in CrackNex. We directly concatenate support features and support reflectance features and feed the concatenated features to generate one single support prototype. Our results demonstrate that incorporating the reflectance features can significantly enhance performance.

We further test a variant "w/ PFM" where we utilize the PFM module to generate dual prototypes. Instead of early concatenating support features and support reflectance features, we use a co-attention mechanism to interactively update the support prototype. The results are reported in row 3. Experimental results show that PFM helps in generating better prototypes.

As for the ASPP module, our experimental results (compared row 3 with 4) demonstrate that adding the ASPP module provides further feature extraction capabilities and yields better performance.

V. Conclusion

In this paper, we propose a novel reflectance-guided few-shot low-light crack segmentation framework, CrackNex. We utilize few-shot segmentation to solve the problem of having to annotate many training images. In addition, we introduce reflectance information to improve segmentation predictions during low-light environments. We validate our framework on two low-light crack datasets, ll_CrackSeg9k and LCSD, and demonstrate significant improvements in the mIOU metric. Our results highlight the importance of incorporating reflectance features to capture details and enhance object boundaries. Additionally, we release a new crack dataset with both well-illuminated and low-light crack images for the benefit of the research community.

Acknowledgments

This work was supported by National Science Foundation Grant CPS 1931867. We would also like to express our gratitude to the anonymous reviewers for their insightful comments and suggestions.

References

[1] Y. Yan, S. Zhu, S. Ma, Y. Guo, and Z. Yu, "Cycleadc-net: A crack segmentation method based on multi-scale feature fusion," *Measurement*, vol. 204, p. 112107, 2022.

[2] W. Choi and Y.-J. Cha, "Sddnet: Real-time crack segmentation," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 9, pp. 8016–8025, 2019.

[3] Y. Liu, J. Yao, X. Lu, R. Xie, and L. Li, "Deepcrack: A deep hierarchical feature learning architecture for crack segmentation," *Neurocomputing*, vol. 338, pp. 139–153, 2019.

[4] D. Kang, S. S. Benipal, D. L. Gopal, and Y.-J. Cha, "Hybrid pixel-level concrete crack segmentation and quantification across complex backgrounds using deep learning," *Automation in Construction*, vol. 118, p. 103291, 2020.

[5] J. Liu, X. Yang, S. Lau, X. Wang, S. Luo, V. C.-S. Lee, and L. Ding, "Automated pavement crack detection and segmentation based on two-step convolutional neural network," *Computer-Aided Civil and Infrastructure Engineering*, vol. 35, no. 11, pp. 1291–1305, 2020.

[6] A. Rezaie, R. Achanta, M. Godio, and K. Beyer, "Comparison of crack segmentation using digital image correlation measurements and deep learning," *Construction and Building Materials*, vol. 261, p. 120474, 2020.

[7] Z. Zheng, X. Ying, Z. Yao, and M. C. Chuah, "Robustness of trajectory prediction models under map-based attacks," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 4541–4550.

[8] Y.-T. Hsieh, K. Anjum, S. Huang, I. Kulkarni, and D. Pompili, "Neural network design via voltage-based resistive processing unit and diode activation function-a new architecture," in *2021 IEEE International Midwest Symposium on Circuits and Systems (MWSCAS)*. IEEE, 2021, pp. 59–62.

[9] Z. Wang, L. Zhang, L. Wang, and M. Zhu, "Landa: Language-guided multi-source domain adaptation," *arXiv preprint arXiv:2401.14148*, 2024.

[10] Y. Zhu, Y. Qiu, Q. Wu, F. L. Wang, and Y. Rao, "Topic driven adaptive network for cross-domain sentiment classification," *Information Processing & Management*, vol. 60, no. 2, p. 103230, 2023.

[11] W. Wang and C. Su, "Automatic concrete crack segmentation model based on transformer," *Automation in Construction*, vol. 139, p. 104275, 2022.

[12] L. Fan, S. Li, Y. Li, B. Li, D. Cao, and F.-Y. Wang, "Pavement cracks coupled with shadows: A new shadow-crack dataset and a shadow-removal-oriented crack detection approach," *IEEE/CAA Journal of Automatica Sinica*, vol. 10, no. 7, pp. 1593–1607, 2023.

[13] W. Li, Z. Shen, and P. Li, "Crack detection of track plate based on yolo," in *2019 12th international symposium on computational intelligence and design (ISCID)*, vol. 2. IEEE, 2019, pp. 15–18.

[14] G. Li, Q. Liu, W. Ren, W. Qiao, B. Ma, and J. Wan, "Automatic recognition and analysis system of asphalt pavement cracks using interleaved low-rank group convolution hybrid deep network and segnet fusing dense condition random field," *Measurement*, vol. 170, p. 108693, 2021.

[15] Y. Wang, K. Song, J. Liu, H. Dong, Y. Yan, and P. Jiang, "Renet: Rectangular convolution pyramid and edge enhancement network for salient object detection of pavement cracks," *Measurement*, vol. 170, p. 108698, 2021.

[16] J. Dong, N. Wang, H. Fang, Q. Hu, C. Zhang, B. Ma, D. Ma, and H. Hu, "Innovative method for pavement multiple damages segmentation and measurement by the road-seg-capsnet of feature fusion," *Construction and Building Materials*, vol. 324, p. 126719, 2022.

[17] E. H. Land, "The retinex theory of color vision," *Scientific american*, vol. 237, no. 6, pp. 108–129, 1977.

[18] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," *arXiv preprint arXiv:1808.04560*, 2018.

[19] N. Zhao, T.-S. Chua, and G. H. Lee, "Few-shot 3d point cloud semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 8873–8882.

[20] K. Wang, J. H. Liew, Y. Zou, D. Zhou, and J. Feng, "Panet: Few-shot image semantic segmentation with prototype alignment," in *proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 9197–9206.

[21] N. Dong and E. P. Xing, "Few-shot semantic segmentation with prototype learning." in *BMVC*, vol. 3, no. 4, 2018.

[22] J. Min, D. Kang, and M. Cho, "Hypercorrelation squeeze for few-shot segmentation," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 6941–6952.

[23] B. Mao, X. Zhang, L. Wang, Q. Zhang, S. Xiang, and C. Pan, "Learning from the target: Dual prototype network for few shot semantic segmentation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 2, 2022, pp. 1953–1961.

[24] L. Sun, C. Li, X. Ding, Y. Huang, Z. Chen, G. Wang, Y. Yu, and J. Paisley, "Few-shot medical image segmentation using a global correlation network with discriminative embedding," *Computers in biology and medicine*, vol. 140, p. 105067, 2022.

[25] P. Pan, Z. Fan, B. Y. Feng, P. Wang, C. Li, and Z. Wang, "Learning to estimate 6dof pose from limited data: A few-shot, generalizable approach using rgb images," *arXiv preprint arXiv:2306.07598*, 2023.

[26] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.

[27] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *International conference on machine learning*. PMLR, 2017, pp. 1126–1135.

[28] V. Garcia and J. Bruna, "Few-shot learning with graph neural networks," *arXiv preprint arXiv:1711.04043*, 2017.

[29] T. Munkhdalai and H. Yu, "Meta networks," in *International conference on machine learning*. PMLR, 2017, pp. 2554–2563.

[30] Z. Wang, M. Ye, X. Zhu, L. Peng, L. Tian, and Y. Zhu, "Metateacher: Coordinating multi-model domain adaptation for medical image classification," *Advances in Neural Information Processing Systems*, vol. 35, pp. 20 823–20 837, 2022.

[31] C. Li, X. Lin, Y. Mao, W. Lin, Q. Qi, X. Ding, Y. Huang, D. Liang, and Y. Yu, "Domain generalization on medical imaging classification using episodic training with task augmentation," *Computers in biology and medicine*, vol. 141, p. 105144, 2022.

[32] T. Elsken, B. Staffler, J. H. Metzen, and F. Hutter, "Meta-learning of neural architectures for few-shot learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 12 365–12 375.

[33] B. Zoph and Q. V. Le, "Neural architecture search with reinforcement learning," *arXiv preprint arXiv:1611.01578*, 2016.

[34] S. Baik, J. Choi, H. Kim, D. Cho, J. Min, and K. M. Lee, "Meta-learning with task-adaptive loss function for few-shot learning," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 9465–9474.

[35] X. Ying, X. Li, and M. C. Chuah, "Weakly-supervised object representation learning for few-shot semantic segmentation," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 1497–1506.

[36] Q. Fan, W. Pei, Y.-W. Tai, and C.-K. Tang, "Self-support few-shot semantic segmentation," in *European Conference on Computer Vision*. Springer, 2022, pp. 701–719.

[37] A. Okazawa, "Interclass prototype relation for few-shot segmentation," in *European Conference on Computer Vision*. Springer, 2022, pp. 362–378.

[38] S. L. Lau, E. K. Chong, X. Yang, and X. Wang, "Automated pavement crack segmentation using u-net-based convolutional neural network," *Ieee Access*, vol. 8, pp. 114 892–114 899, 2020.

[39] J. König, M. D. Jenkins, M. Mannion, P. Barrie, and G. Morison, "Optimized deep encoder-decoder methods for crack segmentation," *Digital Signal Processing*, vol. 108, p. 102907, 2021.

[40] Y. Qiu, Y. Shen, Z. Sun, Y. Zheng, X. Chang, W. Zheng, and R. Wang, "Sats: Self-attention transfer for continual semantic segmentation," *Pattern Recognition*, vol. 138, p. 109383, 2023.

[41] K. Liu, X. Han, and B. M. Chen, "Deep learning based automatic crack detection and segmentation for unmanned aerial vehicle inspections," in *2019 IEEE international conference on robotics and biomimetics (ROBIO)*. IEEE, 2019, pp. 381–387.

[42] C. Sun, S. Huang, and D. Pompili, "Hmaac: Hierarchical multi-agent actor-critic for aerial search with explicit coordination modeling," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 7728–7734.

[43] Y. Wei, Z. Wei, Y. Rao, J. Li, J. Zhou, and J. Lu, "Lidar distillation: Bridging the beam-induced domain gap for 3d object detection," in *European Conference on Computer Vision*. Springer, 2022, pp. 179–195.

[44] B. Dang, D. Ma, S. Li, X. Dong, H. Zang, and R. Ding, "Enhancing kitchen independence: Deep learning-based object detection for visually impaired assistance," *Academic Journal of Science and Technology*, vol. 9, no. 2, pp. 180–184, 2024.

[45] D. Ma, B. Dang, S. Li, H. Zang, and X. Dong, "Implementation of computer vision technology based on artificial intelligence for

medical image analysis," *International Journal of Computer Science and Information Technology*, vol. 1, no. 1, pp. 69–76, 2023.
[46] J.-A. Sarmiento, "Pavement distress detection and segmentation using yolov4 and deeplabv3 on pavements in the philippines," *arXiv preprint arXiv:2103.06467*, 2021.
[47] C. Xiang, J. Guo, R. Cao, and L. Deng, "A crack-segmentation algorithm fusing transformers and convolutional neural networks for complex detection scenarios," *Automation in Construction*, vol. 152, p. 104894, 2023.
[48] X. Ying, X. Li, and M. C. Chuah, "Weakly-supervised object representation learning for few-shot semantic segmentation," in *Proceedings of the IEEE WACV*, 2020.
[49] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.
[50] S. Kulkarni, S. Singh, D. Balakrishnan, S. Sharma, S. Devunuri, and S. C. R. Korlapati, "Crackseg9k: a collection and benchmark for crack segmentation datasets and frameworks," in *European Conference on Computer Vision*. Springer, 2022, pp. 179–195.
[51] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang, "Restormer: Efficient transformer for high-resolution image restoration," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 5728–5739.
[52] X. Ying, B. Lang, Z. Zheng, and M. C. Chuah, "Delving into light-dark semantic segmentation for indoor scenes understanding," in *Proceedings of the 1st Workshop on Photorealistic Image and Environment Synthesis for Multimedia Experiments*, 2022, pp. 3–9.
[53] S. Hong, S. Cho, J. Nam, S. Lin, and S. Kim, "Cost aggregation with 4d convolutional swin transformer for few-shot segmentation," in *European Conference on Computer Vision*. Springer, 2022, pp. 108–126.
[54] L. Yang, W. Zhuo, L. Qi, Y. Shi, and Y. Gao, "Mining latent classes for few-shot segmentation," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 8721–8730.
[55] 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.
[56] J. Kiefer and J. Wolfowitz, "Stochastic estimation of the maximum of a regression function," *The Annals of Mathematical Statistics*, pp. 462–466, 1952.
[57] J. Peng, Y. Liu, S. Tang, Y. Hao, L. Chu, G. Chen, Z. Wu, Z. Chen, Z. Yu, Y. Du *et al.*, "Pp-liteseg: A superior real-time semantic segmentation model," *arXiv preprint arXiv:2204.02681*, 2022.
[58] M. Fan, S. Lai, J. Huang, X. Wei, Z. Chai, J. Luo, and X. Wei, "Rethinking bisenet for real-time semantic segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 9716–9725.
[59] Y. Li, R. Ma, H. Liu, and G. Cheng, "Real-time high-resolution neural network with semantic guidance for crack segmentation," *Automation in Construction*, vol. 156, p. 105112, 2023.