Artificial Intelligence Research and Development
I. Sanz et al. (Eds.)

doi:10.3233/FAIA230657

Stacking Up for Success: A Cascade Network Model for Efficient Road Crack Segmentation

Ammar M. OKRAN [a,1], Adel SALEH [b], Domenec PUIG [a] and Hatem A. RASHWAN [a]
[a] *DEIM, Rovira i Virgili University, 43007 Tarragona, Spain*
[b] *Gaist Solutions Ltd, U.K*
ORCiD ID: Ammar M. Okran https://orcid.org/0000-0001-9264-0301, Adel Saleh https://orcid.org/0000-0001-5502-100X, Domenec Puig https://orcid.org/0000-0002-0562-4205, Hatem A. Rashwan https://orcid.org/0000-0001-5421-1637

Abstract. This paper proposes an integrated framework for automatically segmenting road surface cracks that utilize a Multi-Attention-Network and a modified U-Net, combined through neural network stacking, to segment the crack regions accurately. To evaluate the effectiveness of the proposed framework, we introduce a road crack dataset containing complex environmental noise. We explore several stacking scenarios and perform thorough evaluations to assess the performance of the proposed model. Our results show that the proposed method improves the IOU score of 1.5% compared to the original network, indicating its effectiveness in segmenting road cracks. The proposed framework can be a valuable tool for road maintenance and inspection, enabling timely detection and repair of cracks and improving road safety and longevity. Our findings demonstrate the importance of exploring various stacking scenarios and performing comprehensive evaluations to establish the efficacy of the proposed framework.

Keywords. Road Cracks, Deep learning, Semantic segmentation, Cascade network

1. Introduction

Road networks are vital for society, but maintaining them is challenging. Cracks on road surfaces are a common problem, leading to accidents and economic losses. Spain had over 1,000 fatal accidents in 2022, many caused by road cracks. Factors like weather and traffic contribute to crack formation. Prompt repair using methods like asphalt or concrete filling is crucial for road safety[2].

These accidents caused human losses and inflicted considerable material damages and direct property destruction. Consequently, the Spanish government has allocated a

[1] Corresponding Author: Ammar M. Okran, ammar.okran@urv.cat.

[2] Directorate-General for Traffic (DGT): https://www.dgt.es/comunicacion/notas-de-prensa/1.145-personas-fallecieron-en-siniestros-de-trafico-durante-2022

significant budget of 2.75 billion euros for 2023 to address road construction, maintenance, and repairs, as reported by the Spanish government [3].

Accurate and efficient identification of road cracks is crucial for effective maintenance. Traditional visual inspections are subjective and time-consuming. Deep learning (DL) technologies, like CNNs, have emerged as essential tools for crack detection, such as Yolov7 [1] and semantic segmentation, UNet [2]. They can accurately identify and segment cracks by extracting high-level features from images. DL-based algorithms have improved crack detection performance, with solutions presented in the CRDDC'2022 [3] using one-stage detectors and ensembling approaches [4,5]. Ensembles of one-stage and two-stage detectors [6], have also been proposed. These advancements provide more efficient and accurate solutions for automating crack detection and characterization.

However, road crack detection may not be the most efficient method when it comes to cost calculation. Accurate cost estimation for road maintenance and repair requires precise segmentation of road cracks to determine their size and area, which provides information on the extent of damage, needed repairs, and associated material and labor costs [7,8]. Consequently, several studies have proposed various methods for crack segmentation. For instance, DeepCrack [9] employs an auto-encoder comprising encoder and decoder networks similar to SegNet [10] that the encoder and decoder networks generate convolutional features at the same scale, which are pairwise fused to obtain the final feature representations. In turn, the authors of [11] proposed using gated skip connections that enable the decoder layers to selectively incorporate crack-aware feature representations from the encoder layers. The gating mechanism assigns higher weights to crack-relevant features from the encoder layers and lower weights to irrelevant features.

The existing methods for crack segmentation can encounter challenges due to factors like illumination changes, texture issues, noise, occlusions, and complex object structures. Cascading and stacking DL networks offer a suitable solution by extracting hierarchical features [12], integrating contextual information [13] and improving robustness [14]. This approach ensures accurate and reliable segmentation results [15,16]. In the cascaded or stacked networks architecture, lower-level features capture fine-grained crack details and local information, while higher-level features capture abstract and global contextual information [17]. This hierarchical feature extraction aids in precise object boundary delineation and semantic understanding while handling noise, occlusions, and complex structures. Cascading and stacking DL networks enhance the robustness of crack segmentation systems [17].

The performance of a cascaded DL model heavily relies on the design choices, such as the number of stages, the architecture of each step, and the information flow between stages [18]. Finding an optimal design configuration can be challenging and require extensive experimentation, especially with road crack segmentation and its challenges. Thus, careful experimentation and validation involve finding the right balance regarding model capacity, receptive field size, and complexity for each stage [18]. Iterative refinement and tuning of the cascade architecture are often required to achieve the best tradeoff between localization accuracy and segmentation refinement [18]. Consequently, this paper proposes a road crack segmentation model with two main contributions.

[3] Spanish Government: https://www.sepg.pap.hacienda.gob.es/sitios/sepg/es-ES/Presupuestos/PGE/ProyectoPGE2023/Documents/LIBROAMARILLO2023.pdf

- Introducing a new road crack segmentation dataset derived from the RDD2022 dataset, widely used for road damage detection assessment. This dataset provides a comprehensive and diverse set of crack images for training and evaluating crack segmentation models. The new dataset will provide a more diverse and challenging set of crack images for training and evaluating crack segmentation models.
- Proposing a novel cascade/stack network architecture is proposed for accurate road crack segmentation. It utilizes skip and dense connections to fuse features across layers, improving efficiency. Ensemble learning combines different models for enhanced performance and robustness in crack detection.

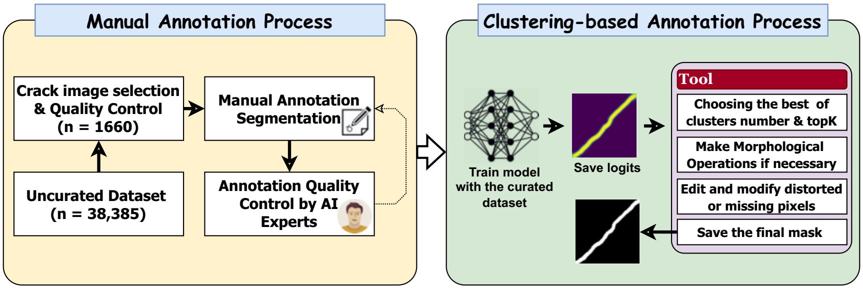

Manual Annotation Process
Crack image selection & Quality Control (n = 1660)
Manual Annotation Segmentation
Uncurated Dataset (n = 38,385)
Annotation Quality Control by AI Experts
Clustering-based Annotation Process
Train model with the curated dataset
Save logits
Tool
Choosing the best of clusters number & topK
Make Morphological Operations if necessary
Edit and modify distorted or missing pixels
Save the final mask

Figure 1. The two stages of the dataset construction process.

The structure of this work is as follows: Section 2 provides a detailed explanation of the dataset construction and the proposed road crack segmentation system. Section 3 presents the experimental results obtained from our proposed method. Finally, Section 4 concludes the paper by summarizing the findings and suggesting potential future research directions.

2. Methodology

The study methodology includes two subsections: dataset construction and road crack segmentation method. The dataset creation process and labeling procedure are detailed in the first subsection. The second subsection presents the innovative DL-based road crack segmentation method using a cascade network architecture. It covers components like the loss function and training procedure. The aim is to clearly understand the methodology, dataset, and proposed approach for road crack segmentation.

2.1. *Road Crack Segmentation Dataset*

The available road crack datasets often rely on manual annotations, which can introduce annotation variations and subjectivity. Different annotators may interpret crack boundaries differently, resulting in inconsistencies in the ground truth annotations. This variability can affect the training and evaluation of models and impact their performance. Thus, in this work, we implement standardized annotation protocols merging two stages of manual and automatic annotations to reduce annotation variations and improve the consistency of ground truth labels. Regular quality control checks and inter-annotator agreement assessments can ensure higher annotation accuracy and reliability. The process of constructing the road crack segmentation dataset consists of two stages, an extension of our previous work [19], illustrated in Figure 1.

In the first stage, 1,660 sub-images were selected from the RDD2022 dataset [20], which includes 38,385 training images from diverse acquisition systems in six countries, i.e., Japan, the Czech Republic, India, the United States of America, Norway, and China. The selection process involved careful review to ensure acceptable image quality. Rejected images, such as those that were blurry or dark, were excluded. An experienced engineer provided manual annotations for each crack in the dataset to ensure accuracy and consistency. AI experts conducted quality control to identify errors or inconsistencies in the annotations. The engineer re-evaluated and manually corrected any rejected annotations to ensure a high-quality final dataset.

We trained a DL model in the second stage using the UNet [2] and DeepLabv3+ [14] networks. The best checkpoint was selected and saved for further use. Subsequently, the model was used to generate logit matrices for each new image after applying the sigmoid function, which produced values ranging from $[0, 1]$. The critical stage of the process involved filtering the results and refining them to create masks that closely resembled those generated manually. Thus, we cluster the segmented regions related to the crack for each image based on the K-means algorithm. We then select the optimal number of clusters by choosing the top-K clusters produced masks that closely resembled the manual annotations. As a postprocessing stage, we applied morphological operations as necessary and manually edited and refined the masks to ensure the highest level of accuracy. Finally, the masks were saved, completing the dataset creation process.

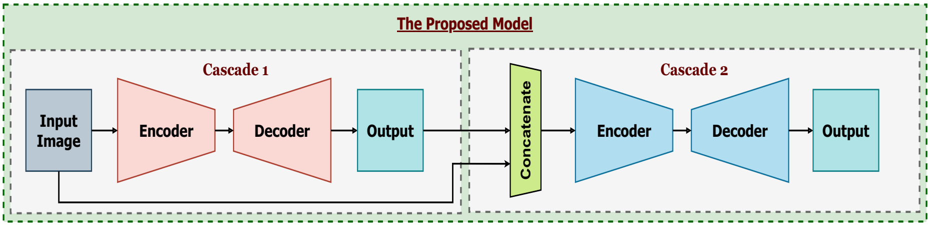
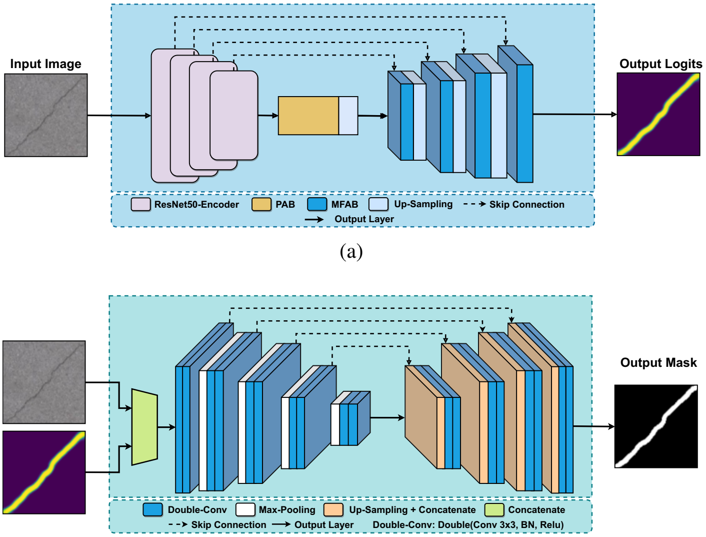


Figure 2. The proposed cascaded model.

In total, we obtained $6,246$ images. The training set comprised $5,041$ images taken from the training set of the RDD2022 dataset, while the testing set contained $1,205$ images taken from the test set of RDD2022. Each image is resized to 128×128.

2.2. *Road Crack Segmentation Method Using a Cascade Network Architecture*

This subsection outlines the methodology employed in our proposed model, which consists of two cascaded networks. The overall model architecture is illustrated in Figure 2. The first cascade network utilizes a Multi-Scale Attention Network (MA-Net) [21], a modular architecture designed to leverage the attention mechanism for capturing rich contextual dependencies. This network is instrumental in capturing comprehensive spatial information and contextual relationships. In turn, the second cascade network in our model adopts a Deeper-UNet architecture. It is fed by concatenating the output from the previous cascade network and the original image as input. The Deeper-UNet architecture is well-suited for capturing fine-grained details and refining the segmentation results.



(a)

(b)

Figure 3. The architecture of the first (a) and the second (b) stages of the cascaded model.

2.2.1. *Cascade 1:* The Multi-Scale Attention Network (MA-Net)

We tested different state-of-the-art segmentation models to select the best model for Cascade 1. Among them, the MANet yields the best results for road crack segmentation. The MANet (cascade 1) [21], incorporates a ResNet-50 [22] encoder as its backbone, as illustrated in Fig 3-(a). The choice of ResNet-50 as the backbone is motivated by its reputation as a state-of-the-art architecture for different image segmentation tasks. The ResNet-50 network is particularly advantageous due to its implementation of residual connections, which address the issue of gradient vanishing. These connections allow the output from a preceding layer to be directly added to a subsequent layer, facilitating the flow of gradients during training. By mitigating the gradient vanishing problem, ResNet-50 enables the efficient propagation of information through the network, extracting meaningful features from the road crack images. The ResNet-50 in our model can encode the input road crack image and extract significant and abstract features at various scales or levels of detail. Cascade 1 also depends on two attention networks for accurate road crack segmentation.

Position-Wise Attention Block (PAB) To capture spatial dependencies between any two positions of feature maps, we employ a PAB [23] at the neck of our network. This approach enhances the model's capability to incorporate broader spatial contextual information across local feature maps. By leveraging the PAB, our network becomes more adept at capturing relevant features for road crack segmentation. This means the PAB is vital in enabling the network to effectively understand the spatial relationships and dependencies within the road crack images, facilitating accurate and precise segmentation.

Multi-Scale Fusion Attention Block (MFAB) Our model incorporates the Multi-MFAB from [21] to capture interdependencies among feature channels. Inspired by the human visual system, MFAB selects salient information for accurate crack segmentation without introducing spatial dimensions. This mechanism enhances useful feature maps while suppressing less relevant ones, improving segmentation performance. By employ-ing MFAB, our model effectively captures contextual dependencies, facilitating precise crack segmentation. The Multi-MFAB is vital in identifying and prioritizing relevant information across scales, leading to reliable and accurate crack segmentation results.

Cascade 1 comprises four encoder blocks from ResNet-50 and four MFAB blocks in the decoder. The MFAB blocks receive feature maps from previous MFAB blocks, PAB blocks, or corresponding ResNet encoder layers. These feature maps include low-level features from skip connections and high-level features from the previous MFAB or PAB blocks. High-level feature maps undergo 3x3 and 1x1 convolutions, Batch Normalization (BatchNorm2D), and ReLU activation. The resulting feature maps are upsampled and passed through a SE-Block to capture high-level attention [24]. The SE-Block includes Adaptive Average Pooling, Convolutional layers, ReLU, Convolutional with a 1x1 kernel, and Sigmoid activation. This combination of layers and blocks in Cascade 1 effectively integrates low-level and high-level features, captures spatial dependencies, and utilizes attention mechanisms for refined and improved feature maps, leading to accurate and robust road crack segmentation.

Additionally, we incorporate another SE-Block to process the low-level feature maps and obtain the low-level attention feature maps. Following this, we perform an element-wise addition of the attention feature maps produced by the SE-Blocks. The resulting feature map is then multiplied by the high-level up-sampled feature maps [21] and concatenated with the low-level feature maps obtained from the Skip-Connection. We apply two successive 3x3 convolutional layers, each followed by Batch Normalization (BatchNorm2D) and Rectified Linear Unit (ReLU) activation, to refine further and enhance the feature maps. These layers can capture meaningful and semantically rich information, allowing for the extraction of more discriminative features.

By incorporating these operations within the MFAB block, our model effectively integrates attention-based mechanisms for low- and high-level feature maps. This enables the network to emphasize relevant information, refine feature representations, and generate more precise crack segmentation results.

*2.2.2. **Cascade 2:** The Deeper-UNet Network*

The proposed cascaded model leverages the Deeper-UNet architecture as Cascade 2, as shown in Fig 3-(b). Deeper-UNet is a popular DL model for image segmentation tasks. It is an extension of the original UNet model [2] consisting of a deep encoder network followed by a decoder network with skip connections between corresponding layers.

The Deeper U-Net architecture consists of an input layer followed by a series of convolutional layers, each responsible for extracting increasingly complex features from the input image. The output of each convolutional layer is then passed to a pooling layer, which reduces the spatial dimension of the feature maps. This is followed by convolutional layers that upsample the feature maps and restore the original spatial dimension. The output of these convolutional layers is then concatenated with the corresponding feature maps from the encoder network to form the skip connections.

In our model, the Deeper-UNet network accepts a 4-channel image as input. We create the 4-channel image by concatenating the logit output of Cascade 1 with the original image. This concatenated image serves as the input for the Deeper-UNet model. The input passes through convolutional layers, doubling the number of channels at each layer until reaching 512 channels. Skip connections are used by upsampling and concatenating feature maps from the encoder with corresponding decoder feature maps. These connections integrate local and global information for more accurate segmentation. The model's output is obtained through a 1x1 convolutional layer, reducing channels to 2 classes (background and defect) and generating the segmentation map for road cracks.

3. Experiments and Results

This section describes the experiments conducted to assess the efficacy of the proposed model. It includes details of the experimental setup, the evaluation metrics employed, and a comprehensive analysis of the resulting outcomes.

We used our constructed dataset for our experiments, which consisted of $6,246$ images. The training set comprised $5,041$ images, each with a size of 128 x 128, and the testing set contained $1,205$ images. In the training pipeline, we utilized augmentation techniques such as horizontal and vertical flipping, rotation, and random brightness contrast for both Cascade 1 and Cascade 2 models. These techniques increased data diversity and improved the models' generalization ability. The models were trained with the Adam optimizer, using a batch size 32 and learning rates of 0.0001 and 0.001 for Cascade 1 and 2, respectively. Cascade 1 employed the binary cross-entropy loss function, while the cross-entropy loss function was used for Cascade 2. These training measures aimed to ensure accurate segmentation while maintaining robust and practical training for the models. The proposed model was evaluated in this work using various evaluation metrics, which included Intersection over Union (IoU), F1-score, Accuracy, and Precision.

In this subsection, we present the experimental results of our road crack segmentation models. First, we compared our proposed models with the baseline models, including DeepLabv3 [13], DeepLabv3+ [14], UNet [2], UNet++ [25], FPN [26], LinkNet [27], PAN [28], PSPNet [29], MANet [21], and Deeper-UNet. All networks used ResNet-50 as a backbone, except Deeper-UNet.

Table 1. Performance of the baseline segmentation models on the curated dataset.

Method	Backbone	IoU	F1-score	Accuracy	Precision
DeepLabv3	ResNet50	0.77640	0.86835	0.91499	0.80596
DeepLabv3+	ResNet50	0.78903	0.87605	0.91896	0.82091
UNet	ResNet50	0.79403	0.87949	0.91996	0.83489
UNet++	ResNet50	0.78430	0.87335	0.91743	0.81934
FPN	ResNet50	0.78732	0.87462	0.91630	0.82627
LinkNet	ResNet50	0.7921	0.87836	0.91680	0.83809
PAN	ResNet50	0.78286	0.87174	0.91614	0.82293
PSPNet	ResNet50	0.77707	0.86867	0.91285	0.81
MANet	ResNet50	**0.80252**	**0.88537**	0.9222	**0.84377**
Deeper-UNet	CNN	0.79639	0.88107	**0.93073**	0.83237

Table 1 shows the results of the baseline models as Cascade 1, where we evaluated the performance of each model in terms of IOU score, F1-score, Accuracy, and Precision. As shown in Table 1, MANet achieved the highest IOU score of 0.802, the highest F1-score of 0.885, and the highest Precision of 0.844 among the ten baselines tested. While Deeper-UNet achieved the highest Accuracy of 0.931. Additionally, we observed that some models, such as LinkNet and PSPNet, performed relatively poorly compared to other models. The results showed the effectiveness of MANet in accurately capturing the extent of road cracks.

In the second experiment, we froze the weights of Cascade 1 and added Cascade 2 as the Deeper-UNet network to train the stacked network. The Deeper-UNet architecture incorporates an expanded network depth compared to the traditional UNet model with multiple stacked convolutional layers that capture more detailed and high-level representations, enhancing the overall segmentation performance. Table 2 shows the results of the proposed models combined with the nine networks of Cascade 1 mentioned above and the Deeper-UNet as a Cascade 2. As shown in Table 2, the proposed model with MANet, combined with Deeper-UNet, achieved the highest IOU score of 0.817, F1-score of 0.895, and Precision of 0.888. However, DeepLabv3 with Deeper-UNet achieved the highest Accuracy of 0.939. The cascaded models yielded more improvement in the four evaluation metrics than the one-stage segmentation models of the first experiment, with an improvement of 1.5% in terms of IOU.

Table 2. The results of the proposed models combined with the nine networks of Cascade 1.

Method1	Method2	IoU	F1-score	Accuracy	Precision
DeepLabv3	**Deeper-UNet**	0.80541	0.88784	**0.93935**	0.86173
DeepLabv3+	**Deeper-UNet**	0.81255	0.89165	0.93352	0.87491
UNet	**Deeper-UNet**	0.81543	0.89359	0.92403	0.88769
UNet++	**Deeper-UNet**	0.81358	0.89259	0.93171	0.87767
FPN	**Deeper-UNet**	0.80928	0.88926	0.93057	0.87473
LinkNet	**Deeper-UNet**	0.80779	0.88892	0.92753	0.87583
PAN	**Deeper-UNet**	0.80443	0.88615	0.93142	0.86879
PSPNet	**Deeper-UNet**	0.80180	0.88569	0.91994	0.87651
MANet	**Deeper-UNet**	**0.81709**	**0.89538**	0.92480	**0.88818**

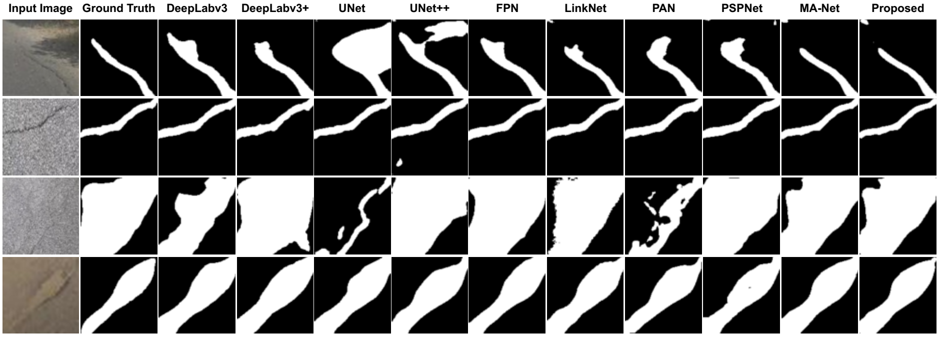


Figure 4. Segmentation results of different models. Each row presents a different crack—row1: Longitudinal Crack, row2: Transverse Crack, row3: Alligator Crack, and row4: Pothole Damage.

The experimental results depicted in Figure 4 demonstrate the effectiveness of our proposed cascaded models for road crack segmentation. The cascaded models consistently outperformed the one-stage models, showcasing their ability to accurately identify and segment road cracks. A quantitative evaluation of the results further supports this claim. Figure 4 visually compares the cascaded models with the ground truth, revealing that our approach generates segmentation results closely resembling the annotated ground truth. This indicates that our cascaded models capture fine details, handle complex crack structures, and produce accurate segmentation maps. Overall, these findings underscore the significance of cascading the MANet and Deeper-UNet networks, as it enables the integration of complementary features and contextual information from multiple stages, leading to improved segmentation performance. By leveraging the strengths of each model and incorporating them in a cascaded manner, our approach achieves superior results in road crack segmentation tasks.

4. Conclusion and Future work

In this work, we introduced a novel road crack segmentation dataset and developed a cascade network architecture designed explicitly for road crack segmentation. We conducted comprehensive experiments using our newly constructed dataset and compared our results with state-of-the-art methods in the field. Our proposed approach outperformed existing methods in terms of various evaluation metrics, including IoU, F1-score, Accuracy, and Precision, with an improvement of 1.5%, 1%, 0.8%, and 4.5%, respectively. These results provide that the cascade network architecture can achieve notable advancements in road crack segmentation performance. Future work will develop a comprehensive road crack detection merging crack multi-class identification and segmentation.

Acknowledgements

Diputació de Tarragona and Universitat Rovira i Virgili partly supported this study through Projecte d'Especialització i Competitivitat Territorial (PECT).

References

[1] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv preprint arXiv:2207.02696*, 2022.

[2] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.

[3] Deeksha Arya, Hiroya Maeda, Sanjay Kumar Ghosh, Durga Toshniwal, Hiroshi Omata, Takehiro Kashiyama, and Yoshihide Sekimoto. Crowdsensing-based road damage detection challenge (crddc-2022). *arXiv preprint arXiv:2211.11362*, 2022.

[4] Dongjun Jeong and Jua Kim. Road damage detection using yolo with image tiling about multi-source images. In *2022 IEEE International Conference on Big Data (Big Data)*, pages 6401–6406. IEEE, 2022.

[5] Ammar M Okran, Mohamed Abdel-Nasser, Hatem A Rashwan, and Domenec Puig. Effective deep learning-based ensemble model for road crack detection. In *2022 IEEE International Conference on Big Data (Big Data)*, pages 6407–6415. IEEE, 2022.

[6] Wenchao Ding, Xu Zhao, Bingke Zhu, Yinglong Du, Guibo Zhu, Tao Yu, Lei Li, and Jinqiao Wang. An ensemble of one-stage and two-stage detectors approach for road damage detection. In *2022 IEEE International Conference on Big Data (Big Data)*, pages 6395–6400. IEEE, 2022.

[7] Wadea Sindi and Bismark Agbelie. Assignments of pavement treatment options: Genetic algorithms versus mixed-integer programming. *Journal of Transportation Engineering, Part B: Pavements*, 146(2):04020008, 2020.

[8] S Jana, S Thangam, Anem Kishore, Venkata Sai Kumar, and Saddapalli Vandana. Transfer learning based deep convolutional neural network model for pavement crack detection from images. *International Journal of Nonlinear Analysis and Applications*, 13(1):1209–1223, 2022.

[9] Yahui Liu, Jian Yao, Xiaohu Lu, Renping Xie, and Li Li. Deepcrack: A deep hierarchical feature learning architecture for crack segmentation. *Neurocomputing*, 338:139–153, 2019.

[10] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017.

[11] M Jabreel and M Abdel-Nasser. Promising crack segmentation method based on gated skip connection. *Electronics Letters*, 56(10):493–495, 2020.

[12] Yi Li, Haozhi Qi, Jifeng Dai, Xiangyang Ji, and Yichen Wei. Fully convolutional instance-aware semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2359–2367, 2017.

[13] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017.

[14] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018.

[15] Stanislav Fort, Huiyi Hu, and Balaji Lakshminarayanan. Deep ensembles: A loss landscape perspective. *arXiv preprint arXiv:1912.02757*, 2019.

[16] Li Deng, Xiaodong He, and Ji Gao. Deep stacking networks for information retrieval. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3153–3157. IEEE, 2013.

[17] Yu-Cheng Liu, Mohammad Shahid, Wannaporn Sarapugdi, Yong-Xiang Lin, Jyh-Cheng Chen, and Kai-Lung Hua. Cascaded atrous dual attention u-net for tumor segmentation. *Multimedia tools and applications*, 80:30007–30031, 2021.

[18] Shervin Minaee, Yuri Boykov, Fatih Porikli, Antonio Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos. Image segmentation using deep learning: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 44(7):3523–3542, 2021.

[19] Ammar M Okran, Mohamed Abdel-Nasser, Hatem A Rashwan, and Domenec Puig. A curated dataset for crack image analysis: Experimental verification and future perspectives. 2022.

[20] D Arya, Hiroya Maeda, Yoshihide Sekimoto, Hiroshi Omata, Sanjay Kumar Ghosh, Durga Toshniwal, Madhavendra Sharma, Van Vung Pham, Jingtao Zhong, Muneer Al-Hammadi, Mamoona Birkhez Shami, Du Nguyen, Hanglin Cheng, Jing Zhang, Alex Klein-Paste, Helge Mork, Frank Lindseth, Toshikazu Seto, Alexander Mraz, and Takehiro Kashiyama. Rdd2022 - the multi-national road damage dataset released through crddc'2022, 2022.

[21] Tongle Fan, Guanglei Wang, Yan Li, and Hongrui Wang. Ma-net: A multi-scale attention network for liver and tumor segmentation. *IEEE Access*, 8:179656–179665, 2020.

[22] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE CVPR*, pages 770–778, 2016.

[23] Jun Fu, Jing Liu, Haijie Tian, Yong Li, Yongjun Bao, Zhiwei Fang, and Hanqing Lu. Dual attention network for scene segmentation. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3141–3149, 2019.

[24] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.

[25] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*, pages 3–11. Springer, 2018.

[26] Alexander Kirillov, Kaiming He, Ross Girshick, and Piotr Dollár. A unified architecture for instance and semantic segmentation, 2017.

[27] Abhishek Chaurasia and Eugenio Culurciello. Linknet: Exploiting encoder representations for efficient semantic segmentation. In *2017 IEEE visual communications and image processing (VCIP)*, pages 1–4. IEEE, 2017.

[28] Hanchao Li, Pengfei Xiong, Jie An, and Lingxue Wang. Pyramid attention network for semantic segmentation. *arXiv preprint arXiv:1805.10180*, 2018.

[29] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2881–2890, 2017.