

Fusion of Thermal and RGB Images for Automated Deep Learning Based Marble Crack Detection

Eleni Vrochidou
MLV Research Group

Department of Computer Science
International Hellenic University
Kavala, Greece
evrochid@cs.ihu.gr

Athanasios G. Ouzounis
MLV Research Group

Department of Computer Science
International Hellenic University
Kavala, Greece
athouzo@emt.ihu.gr

Andreas Stamkos
Intermek A.B.E.E.

Kavala, Greece
a.stamkos@intermek.gr

George K. Sidiropoulos
MLV Research Group

Department of Computer Science
International Hellenic University
Kavala, Greece
georsidi@teiemt.gr

Ilias T. Sarafis

Department of Chemistry
International Hellenic University
Kavala, Greece
isarafis@chem.ihu.gr

Ioannis Tsimperidis
MLV Research Group

Department of Computer Science
International Hellenic University
Kavala, Greece
tsimperidis@cs.ihu.gr

Vassilis Kalpakis
Intermek A.B.E.E.

Kavala, Greece
v.kalpakis@intermek.gr

George A. Papakostas
MLV Research Group

Department of Computer Science
International Hellenic University
Kavala, Greece
gpapak@cs.ihu.gr

Abstract—Research is constantly turning towards the development of image-based inspection tools that leverage deep learning models to automate surface crack detection in materials. Yet, most efforts involve color imaging which may lead to poor models' performance in different lighting variations. Thermal images are more constant in changes of lighting, however, lack crispness of details. In this work, thermal-*RGB* image fusion techniques are investigated for accurate marble crack detection. The proposed method uses a Feature Pyramid Network (FPN) combined with three different backbone architectures and an augmented dataset of thermal-*RGB* image pairs. The performances of different image modalities, i.e., *RGB*, thermal and fused images, are comparatively evaluated, to test the impact of image type, fusion techniques, deep networks, and feature extraction backbones, on the segmentation results, considering two different experimental evaluation approaches: a model-based and a backbone-based approach. Results verify the initial hypothesis that fused images are richer in spatial information. Model-based evaluation resulted in scores of up to 85.07% in terms of mean Intersection over Union (mIoU), which is 2.9% and 4.75% higher than by using separately the corresponding *RGB* (82.67% mIoU) and thermal (81.21% mIoU) images, respectively. Backbone-based evaluation highlighted *resnext50* as the optimal feature extraction network combination for FPN, reporting an mIoU of 87.04%.

Keywords—*image fusion, marble crack detection, deep learning, thermal images, computer vision, semantic segmentation*

I. INTRODUCTION

Marble surface crack detection is crucial to preserve the high quality and aesthetic value of structures, and thus to prevent the risks of their early deterioration and natural decay through maintenance actions [1]. To date, marble surface crack detection is performed manually by experienced human inspectors. However, cracks may be too thin and difficult to be detected by observation, especially in randomly textured marble surfaces. Automated methods to facilitate or entirely

replace conventional visual inspection of marble cracks, are therefore needed.

Recent advancements in deep learning (DL) algorithms have gained considerable attention from both industry and academia towards automatic fault diagnosis [2]. Semantic segmentation based on deep neural networks has been proposed in numerous crack detection applications, proving that automatic DL-based methods could be highly efficient [3]–[5]. Most of the current semantic segmentation methods, use single-modal sensory data, mostly referring to red-green-blue (*RGB*) images [6]–[8] or thermal images [9]–[11]. The sensing of *RGB* images is closer to human perception, while thermal cameras detect infrared radiation emitted from the surface of physical objects. Visible detection based on *RGB* images is more commonly used, due to the relatively affordable high resolution and contrast commercial cameras. However, in varying lighting conditions, *RGB* images may not be effective. Thermal detection is light invariant; yet, thermal cameras are comparatively pricey, while the images lack in crispness of detail, are of lower resolution and contrast. In order to take advantage of both modality-specific information, multimodal *RGB*-Thermal image fusion techniques were introduced [12]–[14]. Results indicated that multi-modal fusion can significantly improve the performance of DL segmentation models towards defects detection in materials [11].

It should be noted here that DL-based marble crack detection is scarce in the literature; two reported works on marble crack detection exploit visible detection [1], [15], while thermal crack detection was only recently investigated [16]. However, *RGB*-Thermal fusion has never been applied for marble crack segmentation.

To this end, this work investigates for the first time, the performance of *RGB*-Thermal image fusion techniques for marble crack detection. More specifically, 10 image fusion techniques (17 including their variations) are examined for pairs of *RGB* and thermal images. Fused images are used to

train a Feature Pyramid Network (FPN) with three different feature extraction backbone segmentation models. To gain new insights regarding modality-specific data, this work comparatively evaluates the performance of RGB, thermal, and fused images, following two different evaluation approaches: a model-based and a backbone-based evaluation.

The rest of the paper is structured as follows. Section II discusses related works. Section III presents materials and methods, including the proposed approach, the used DL models and the examined fusion techniques and dataset. Experimental results and discussion are summarized in Section IV. Finally, Section V concludes the paper.

II. RELATED WORKS

Although much work has been done in DL-based crack detection for mainly structural materials such as concrete, crack detection in marble is particularly limited. Only two reported works on DL-based marble crack detection exploiting visible detection [1], [15] can be found in the literature. In [15], five different convolutional neural networks (CNNs) were trained to detect cracks on RGB marble images: MobileNet - v2, Xception, ResNet - 18, Inception - ResNet - v2 and ResNet - 50, combined with three different optimization algorithms (stochastic gradient descent with momentum (SGDM), Adam, and root mean square propagation (RMSprop)). ResNet - 50 with the RMSprop optimizer resulted as the best architecture, providing a mean Intersection over Union (mIoU) of 67.2%. In [1], the authors conducted a performance evaluation of 112 DL segmentation model architectures, combining four models and 28 feature extraction backbones, using RGB marble slab images. Results indicated the FPN model as the most efficient architecture, providing mIoU of 71.35%, and SE-ResNet as the most effective backbone family for the problem under study. Thermal marble crack detection was only recently investigated [16]. The authors comparatively evaluated

several DL models, using both thermal and RGB images, towards evaluating the impact of different image modalities on the segmentation results. Experimental results showed FPN as best performing model architecture with 71.61% and 68.07% mIoU, for RGB and thermal images, respectively, while the best performing backbone, was the efficientnetb4 with 80.07% and 75.49% mIoU for RGB and thermal images, respectively. Yet, it should be noted that RGB-Thermal fusion has never been reported in the literature for marble crack segmentation.

III. MATERIALS AND METHODS

A. The Proposed Methodology

All image types, fused, RGB and thermal were subjected to basic pre-processing before being processed by the DL model. RGB-only images were subjected to contrast-limited adaptive histogram equalization (CLAHE) [17] towards reducing noise amplification. Image embossing to raise crack patterns against the background was applied to thermal-only and RGB-thermal fused images, along with Principal Components Analysis color augmentation (Fancy PCA) [18]. Ten different fusion techniques were then tested. The input images were used to train and test a DL segmentation model. One semantic segmentation model, FPN, was combined with three feature extraction networks. Results included the output segmentation image and the numerical results in terms of well-known segmentation metrics and were evaluated based on two different approaches, a model-based evaluation and a backbone-based evaluation.

The same process was repeated for fused, RGB and thermal images of the same cracks for comparative reasons. Therefore, a comparative study of thermal versus color versus RGB-thermal-fused imaging took place. Fig.1 illustrates the conceptual flow of the proposed methodology.

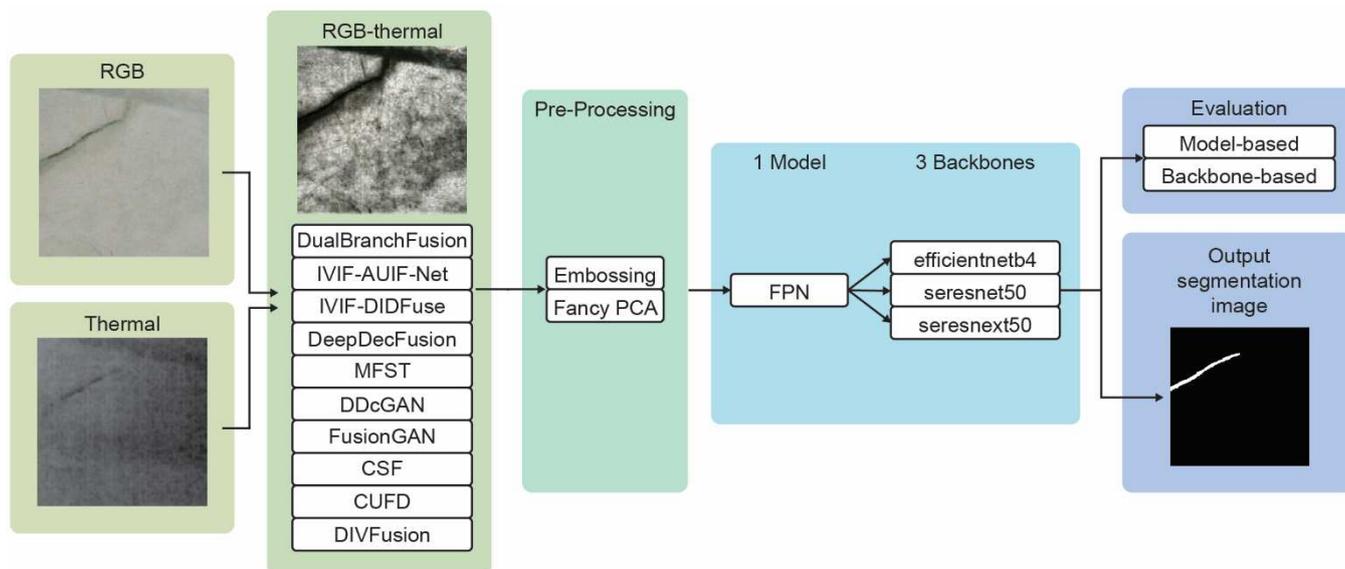


Fig. 1. The proposed methodology.

B. RGB-Thermal Fusion Techniques

Image fusion is a subject of interest in the research area of image processing. This is due to the fact that fused images can contain more comprehensive information, since they extract

and combine effective features from multi-modal images. Visible and thermal image fusion is expected to maintain both the prominent textural details from RGB images and highlight the thermal targets from thermal images.

In this work, ten of the most recent and advanced algorithms in the field of image fusion were used to generate RGB-Thermal fused images based on their popularity and reported performances, and are compared:

- Dual-branch Network (DualBranchFusion, 2021) with two feature strategies (addition strategy (Add) and channel strategy (Ch)) [19],
- Infrared and Visible Image Fusion via Algorithm Unrolling (IVIF-AUIF-Net, 2021) with three fusion strategies (addition (Add), Average (Av), l_1 -attention addition (Norm)) [20],
- Deep Image Decomposition (IVIF-DIDFuse, 2020) with three fusion strategies (Add, Av, Norm) [21],
- Deep Learning-based encoder/decoder for fusing infrared and visible images (DeepDecFusion, 2021) trained with three different image datasets (medical images (Med dataset [22], TNO dataset [23], Road dataset [24]) [25],
- Multi-Modal Feature Self-Adaptive Transformer (MFST, 2022) [26],
- Dual-Discriminator Conditional Generative Adversarial Network (DDcGAN, 2020) [27],
- Generative Adversarial Network for Infrared and Visible Image Fusion (FusionGAN, 2019) [28],
- Classification Saliency-Based Rule for Visible and Infrared Image Fusion (CSF, 2021) [29],
- Encoder-Decoder Network for Visible and Infrared Image Fusion based on Common and Unique Feature Decomposition (CUFD, 2022) [30],
- Darkness-Free Infrared and Visible Image Fusion (DIVFusion, 2023) [31].

C. The Dataset

The dataset [16] used in this work is derived from 38 marble tiles with cracks of up to 2 mm wide from the marble quarrying company Solakis S.A. [32] in Drama, Greece. All tiles were photographed from 90 cm. For the high-resolution RGB images, it was used an MV_CA050-10GM/GC digital camera with an MVLMF0824M-5MP lens. The thermal images were acquired after the tiles being heated with an infrared source by using a thermal heat-sensitive 206×156 Seek Compact XR camera [33].

RGB and thermal images were paired, while the visible cracks were manually annotated on the RGB images with the LabelMe annotation tool [34]. The original dataset included a total of 24 pairs of RGB-thermal images. To comparatively test the influence of different modality input data, the performance of the DL model was tested separately on the RGB, thermal and fused images. RGB, thermal and fused images were augmented with random rotation (between 0 and 90°), horizontal flip and vertical flip (both with 50% chance) concluding in 244 fused images. Five-fold cross-validation was applied to the final dataset to boost the confidence of the model's performance.

Fig. 2 indicatively illustrates a pair of images, RGB image and thermal image of the same marble crack, and the fused image resulting from the DIVFusion algorithm.

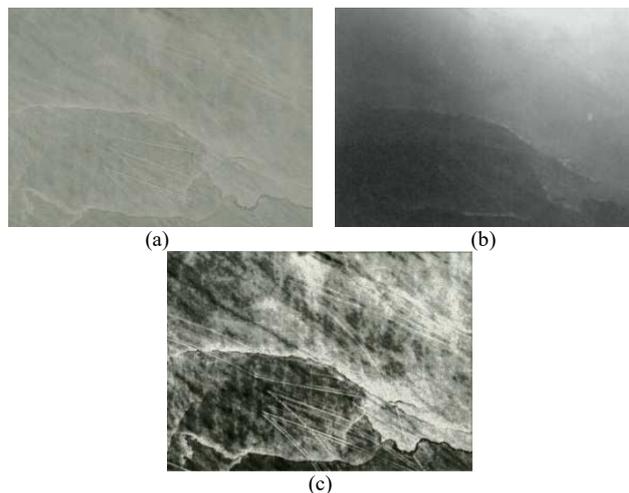


Fig. 2. Marble crack images: (a) RGB, (b) thermal, (c) RGB-thermal with DIVFusion fusion algorithm.

D. DL Segmentation Models

The DL segmentation models are obtained by the combinations of FPN [35] and three feature extraction backbone networks: efficientnetb4, seresnet50 and seresnext50.

The selection of model and backbones is based on a recent comparative performance evaluation between state-of-the-art DL models on the same thermal and RGB input images of marble cracks [16]. Based on [16], the best performing segmentation model resulting from the conducted model-based evaluation experiment and the three best performing feature extraction methods resulted from the backbone-based evaluation experiment, are selected.

The aim is to investigate the comparative performance of different modality-specific input data, fusion algorithms and network architectures for marble crack detection.

IV. RESULTS AND DISCUSSION

All experiments in this work were implemented in Python 3.9 engaging Tensor-Flow and Keras and run on an Nvidia RTX 3090 GPU. Thermal, RGB and fused images were resized into 256×256 pixels size to be input to the FPN model. Activation function was sigmoid, Adam optimizer was used and the loss function was the sum of focal and dice loss as suggested in [16]. The model was trained for 50 epochs with 20 steps per epoch and the learning rate was 0.0005. The 75% of the FPN layers were frozen and the rest 25% were trainable. For better convergence, the three backbone networks were pretrained on ImageNet [21]. The same configuration was applied to all model combinations.

All experimental results presented in this work are calculated after 5-fold cross-validation in terms of the most used semantic segmentation metrics: Inter-section over Union (IoU), Precision (P), Recall (R) and F1-score.

A. Experimental Results

a) *Model-Based Evaluation*: In this approach, the average performance of the FPN model with the three most efficient [16] backbones (Efficientnetb4, Seresnet50, Seresnext50) are studied. Table I includes the performance results (mean values) after 5-fold cross-validation on the testing set. The Table includes the segmentation results for all images: RGB, thermal and fused for all fusion algorithms.

TABLE I. MODEL-BASED SEGMENTATION RESULTS (% MEAN VALUES) FOR ALL IMAGE MODALITIES. BEST MEAN IOU IS MARKED IN BOLD

Image Type		Performance Metric			
	Fusion Algorithm	IoU	P	R	F1-score
	Fused	DualBranchFusion_Add	81.76	91.84	88.32
DualBranchFusion_Ch		80.67	93.15	85.99	84.16
IVIF-AUIF-Net_Add		81.97	92.51	87.94	85.51
IVIF-AUIF-Net_Av		82.25	92.93	87.93	85.75
IVIF-AUIF-Net_Norm		81.32	92.69	87.14	84.82
IVIF-DIDFuse_Add		80.04	90.89	87.31	83.57
IVIF-DIDFuse_Av		82.95	93.58	87.88	86.43
IVIF-DIDFuse_Norm		83.31	93.03	88.76	86.81
DeepDecFusion_Med		83.57	92.57	89.47	87.15
DeepDecFusion_TNO		82.45	93.70	87.25	86.05
DeepDecFusion_Road		84.09	93.32	89.30	87.66
MFST		79.91	91.14	87.06	83.63
DDcGAN		83.58	91.97	90.15	87.26
FusionGAN		83.71	92.04	90.12	87.56
CSF		82.06	92.25	88.14	85.51
CUFD	81.79	92.39	87.63	85.39	
DIVFusion	85.07	93.51	90.31	88.72	
RGB		82.67	94.07	87.37	86.14
Thermal		81.21	93.29	86.35	84.33

As it can be observed from Table I, RGB-Thermal fused images result in better performances compared to RGB and thermal, regardless of the fusion algorithm. The latter verifies the hypothesis that multi-modal fusion techniques are capable of improving the performance of DL defect detection models compared to single-modal images. More specifically, results highlight the DIVFusion as the most efficient fusion algorithm; marble crack detection accuracy is as high as 85.07% mIoU, which is 2.9% and 4.75% higher than by using separately the corresponding RGB (82.67 % mIoU) and thermal (81.21% mIoU) images, respectively. DeepDecFusion_Road also reveals a high segmentation performance of 84.09% mIoU, followed by FusionGAN with 83.71% mIoU and DDcGAN with 83.58% mIoU. Fig. 3 illustrates an indicative segmentation result for FPN with DIVFusion.

b) Backbone-Based Evaluation: FPN is evaluated for three different feature extraction backbones. Table II includes the performance results for each backbone and image modality: RGB, thermal and RGB-thermal (fused) for each fusion algorithm.

Considering the backbone-based experimental approach, results in Table II once again indicate that multimodal images can lead the DL model to better segmentation performances. In all examined cases of backbones, the model resulted in better results for RGB-thermal images. Best IoU of 87.04 % was reported with Seresnext50 and fused images from the DIVFusion algorithm.

Evaluating the overall performance of feature extraction backbones, efficientnetb4 reported a higher mIoU (83.19%) considering all fusion algorithms, compared to the other two backbones (83.97% for seresnext50 and 80.08% for seresnet50). Results verify the hypothesis that the selection of a specific feature extraction network can improve the segmentation performance of a model. Figure 4 indicatively illustrates segmentation results of the FPN model with fused images by DDcGAN of the testing set, with two different backbones.

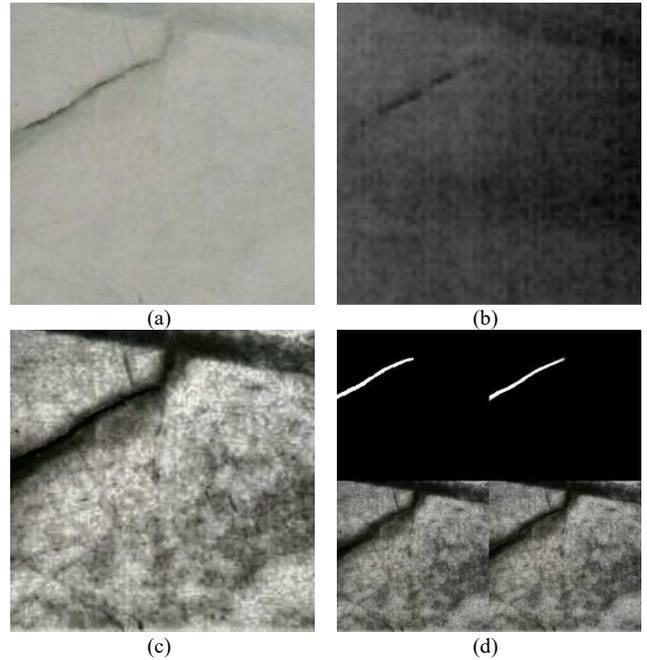


Fig. 3. Indicative results of FPN to fused testing images with DIVFusion algorithm: (a) RGB image, (b) thermal image, (c) fused image and (d) segmentation result corresponding to IoU=97.25% with efficientb4 (up-left=ground truth image, down-left=fused image, up-right=output segmentation, down-right=output segmentation mask applied to the input image)).

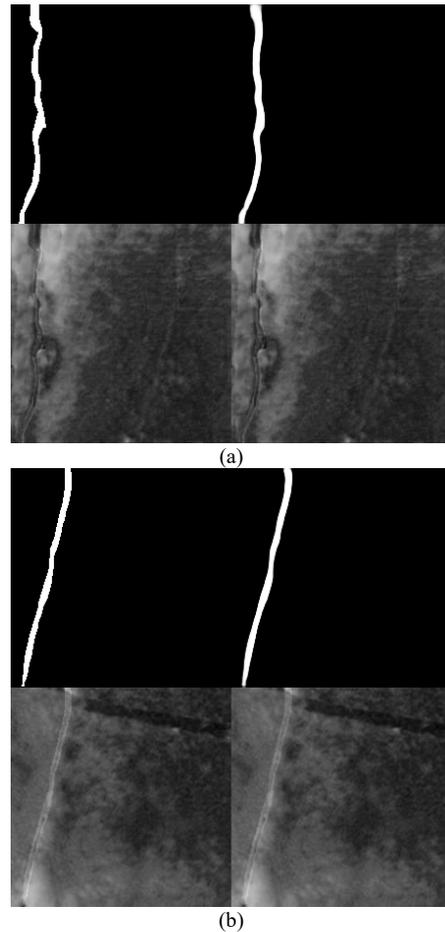


Fig. 4. Indicative results of FPN to fused testing images with DDcGAN fusion (up-left=ground truth image, down-left=fused image, up-right=output segmentation, down-right=output segmentation mask applied to the input image): (a) with efficientb4 backbone (IoU=96.20%), (b) with seresnext50 backbone (IoU=95.76%).

TABLE II. BACKBONE-BASED SEGMENTATION RESULTS (% MEAN VALUES) FOR ALL IMAGE MODALITIES. BEST MEAN IOU IS MARKED IN BOLD

Image Type		Backbone	Performance Metric			
Fusion Algorithm	IoU		P	R	F1-score	
		Fused				DualBranchFusion_Add
Seresnet50	80.32		92.23	86.53	83.88	
Seresnext50	82.36		92.17	88.82	85.87	
DualBranchFusion_Ch	Efficientnetb4		82.57	92.78	88.14	86.26
	Seresnet50		76.75	93.02	81.89	80.20
	Seresnext50		82.70	93.64	87.95	86.03
IVIF-AUIF-Net_Add	Efficientnetb4		81.94	93.57	86.44	85.58
	Seresnet50		81.42	91.44	88.56	84.87
	Seresnext50		82.54	92.51	88.82	86.07
IVIF-AUIF-Net_Av	Efficientnetb4		82.31	93.09	87.59	85.95
	Seresnet50		81.47	92.77	87.66	84.72
	Seresnext50		82.97	92.93	88.53	86.59
IVIF-AUIF-Net_Norm	Efficientnetb4		82.81	92.68	88.32	86.44
	Seresnet50		77.87	92.46	84.00	81.29
	Seresnext50		83.28	92.94	89.09	86.74
IVIF-DIDFuse_Add	Efficientnetb4		85.02	93.57	90.12	88.57
	Seresnet50		71.40	86.12	82.52	74.99
	Seresnext50		83.71	92.98	89.30	87.15
IVIF-DIDFuse_Av	Efficientnetb4		84.19	93.73	88.87	87.71
	Seresnet50		80.73	93.39	85.91	84.25
	Seresnext50		83.92	93.63	88.86	87.33
IVIF-DIDFuse_Norm	Efficientnetb4		83.89	92.82	89.42	87.48
	Seresnet50		80.62	92.09	86.63	84.19
	Seresnext50		85.41	94.18	90.22	88.76
DeepDecFusion_Med	Efficientnetb4		83.53	92.49	89.11	87.18
	Seresnet50		83.00	92.64	88.70	86.66
	Seresnext50		84.17	92.59	90.59	87.62
DeepDecFusion_TNO	Efficientnetb4		83.58	93.14	88.65	87.18
	Seresnet50		80.31	93.24	85.72	83.89
	Seresnext50		83.47	94.72	87.39	87.09
DeepDecFusion_Road	Efficientnetb4		84.07	93.30	89.11	87.67
	Seresnet50		83.10	93.84	87.78	86.74
	Seresnext50		85.10	92.82	91.00	88.58
MFST	Efficientnetb4		80.01	90.86	87.33	83.89
	Seresnet50		77.80	90.69	85.45	81.50
	Seresnext50		81.93	91.87	88.40	85.49
DDcGAN	Efficientnetb4		84.13	92.57	90.11	87.93
	Seresnet50		81.88	90.92	89.41	85.58
	Seresnext50		84.74	92.43	90.94	88.28
FusionGAN	Efficientnetb4		84.65	92.95	89.70	88.56
	Seresnet50		81.49	92.70	87.33	85.38
	Seresnext50		85.00	90.48	93.33	88.74
CSF	Efficientnetb4		82.00	92.55	87.36	85.51
	Seresnet50		80.85	92.05	87.31	84.34
	Seresnext50		83.34	92.16	89.74	86.69
CUFD	Efficientnetb4		83.02	92.67	88.42	86.75
	Seresnet50		78.24	92.95	83.50	81.77
	Seresnext50		84.11	91.55	90.98	87.64
DIVFusion	Efficientnetb4		84.01	94.78	87.95	87.88
	Seresnet50		84.16	91.93	90.90	87.87
	Seresnext50	87.04	93.83	92.09	90.41	
RGB	Efficientnetb4	84.99	93.63	90.01	88.87	
	Seresnet50	80.18	94.78	84.27	83.36	
	Seresnext50	85.90	94.33	90.50	89.46	
Thermal	Efficientnetb4	82.89	92.82	88.39	86.07	
	Seresnet50	79.54	93.68	84.08	82.68	
	Seresnext50	82.42	93.43	87.61	85.55	

B. Discussion

In this work, the influence of the feature extraction backbone network and different image types (RGB, thermal and RGB-thermal) on the segmentation performance of a DL model was investigated. Furthermore, different RGB-thermal fusion algorithms were considered. Results from both conducted experimental approaches included in Tables I and

II, reveal that for fused images the comparison metrics of all model configurations were of the same range for all fusion techniques, outperforming the corresponding results with RGB and thermal images.

Since the cracks only refer to a small part of the marble slab, certain metrics may be affected by the sample imbalance between the target and background pixels. This is the reason

why considering four well-known metrics, including F1-score, for the comparative results.

The experimental results show that a model improves its marble crack segmentation performance using RGB-thermal images as input, especially when certain backbone networks, such as the seresnext50, are embedded in the FPN model.

It should be noted here that only the RGB images in this work were taken under controlled conditions (diffusion box), while the thermal images were acquired separately in natural conditions. This is reflected in the results of RGB and thermal images, where the thermal image leads to lower segmentation performance compared to RGB. At this point, it should also be highlighted that the RGB camera was of high resolution, while the thermal camera was of low resolution, which can also be observed in Fig.3. In both cases however, only one marble crack was presented in each image. The latter refers to ideal image acquisition conditions, which poses a limitation in practice, such as the thermal camera being affected by the outside temperature or the RGB camera being affected by varying lighting conditions.

An additional limitation was the finite number of marble crack images, including only 24 image pairs. Future work includes the acquisition of a large dataset of RGB and thermal image pairs of marble cracks, under different lighting conditions, and by using high resolution sensors. Moreover, augmentation techniques based on adversarial Generative adversarial networks (GANs) for image augmentation will be explored.

Regardless of these limitations, the proposed approach is the first to present RGB-thermal image fusion comparative evaluation for marble crack detection, considering different image modalities, different feature extraction networks and two different experimental approaches. This work confirmed the hypothesis that RGB-thermal fused images can outperform the RGB-only and thermal-only DL models. The DL model is able to obtain additional insights provided by thermal images and become more robust, suitable for surface marble crack detection tasks in the production lines.

V. CONCLUSIONS

This work introduced a comprehensive comparison between ten different RGB-thermal fusion techniques and tested different image modalities as inputs to an FPN segmentation model with three different backbone combinations, towards marble crack detection. Multi-modal images were proven more efficient in all cases, considering two experimental approaches, a model-based and a backbone-based evaluation approach. Results verified the hypothesis that fused images were richer in spatial information, resulting in accuracies of up to 85.07% in terms of mIoU, compared to RGB-only (82.67% mIoU) and thermal-only (81.21% mIoU) images. As the best performing backbone model was indicated the seresnext50, reaching mIoU of up to 87.04% with fused images.

This study highlighted the efficiency of applying RGB-thermal image fusion techniques towards marble crack detection. However, due to the limited size of the training dataset, the limited class of cracks, and the low-resolution thermal camera used in this work for the thermal image acquisition, it can be concluded that there is still room for further improvements, which will be investigated in future studies.

ACKNOWLEDGMENT

This research has been co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH — CREATE — INNOVATE (project code: T2EAK-00238).

REFERENCES

- [1] E. Vrochidou *et al.*, “Towards Robotic Marble Resin Application: Crack Detection on Marble Using Deep Learning,” *Electronics*, vol. 11, no. 20, p. 3289, Oct. 2022.
- [2] Y. Hamishebahar, H. Guan, S. So, and J. Jo, “A Comprehensive Review of Deep Learning-Based Crack Detection Approaches,” *Appl. Sci.*, vol. 12, no. 3, p. 1374, Jan. 2022.
- [3] J. Huyan, W. Li, S. Tighe, Z. Xu, and J. Zhai, “CrackU-net: A novel deep convolutional neural network for pixelwise pavement crack detection,” *Struct. Control Heal. Monit.*, vol. 27, no. 8, Aug. 2020.
- [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,” *Autom. Constr.*, vol. 118, p. 103291, Oct. 2020.
- [5] Y. Z. Ayele, M. Aliyari, D. Griffiths, and E. L. Droguett, “Automatic Crack Segmentation for UAV-Assisted Bridge Inspection,” *Energies*, vol. 13, no. 23, p. 6250, Nov. 2020.
- [6] Z. Fan, H. Lin, C. Li, J. Su, S. Bruno, and G. Loprencipe, “Use of Parallel ResNet for High-Performance Pavement Crack Detection and Measurement,” *Sustainability*, vol. 14, no. 3, p. 1825, Feb. 2022.
- [7] S. Guo, “Research on Segmentation Method of Civil Construction Image based on Unet Algorithm Model,” in *2022 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)*, 2022, pp. 1503–1506.
- [8] D. Li *et al.*, “Automatic defect detection of metro tunnel surfaces using a vision-based inspection system,” *Adv. Eng. Informatics*, vol. 47, p. 101206, Jan. 2021.
- [9] K. Wang, J. Zhang, H. Ni, and F. Ren, “Thermal Defect Detection for Substation Equipment Based on Infrared Image Using Convolutional Neural Network,” *Electronics*, vol. 10, no. 16, p. 1986, Aug. 2021.
- [10] X. Peng, X. Zhong, A. Chen, C. Zhao, C. Liu, and Y. F. Chen, “Debonding defect quantification method of building decoration layers via UAV-thermography and deep learning,” *Smart Struct. Syst.*, vol. 28, pp. 55–67, 2021.
- [11] P. G. Minatel, B. C. Oliveira, and A. Albertazzi, “Comparison of Unet and Mask R-CNN for impact damage segmentation in lock-in thermography phase images,” in *Automated Visual Inspection and Machine Vision IV*, 2021, p. 26.
- [12] C. Chen, S. Chandra, and H. Seo, “Automatic Pavement Defect Detection and Classification Using RGB-Thermal Images Based on Hierarchical Residual Attention Network,” *Sensors*, vol. 22, no. 15, p. 5781, Aug. 2022.
- [13] C. Chen, S. Chandra, Y. Han, and H. Seo, “Deep Learning-Based Thermal Image Analysis for Pavement Defect Detection and Classification Considering Complex Pavement Conditions,” *Remote Sens.*, vol. 14, no. 1, p. 106, Dec. 2021.
- [14] Y. Sun, W. Zuo, P. Yun, H. Wang, and M. Liu, “FuseSeg: Semantic Segmentation of Urban Scenes Based on RGB and Thermal Data Fusion,” *IEEE Trans. Autom. Sci. Eng.*, vol. 18, no. 3, pp. 1000–1011, Jul. 2021.
- [15] S. A. Akosman, M. Oktem, O. T. Moral, and V. Kilic, “Deep Learning-based Semantic Segmentation for Crack Detection on Marbles,” in *2021 29th Signal Processing and Communications Applications Conference (SIU)*, 2021, pp. 1–4.
- [16] E. Vrochidou *et al.*, “RGB and Thermal Image Analysis for Marble Crack Detection with Deep Learning,” in *International Conference on Paradigms of Communication, Computing and Data Analytics (PCCDA 2023)*, 2023.
- [17] K. Zuiderveld, “Contrast Limited Adaptive Histogram Equalization,” in *Graphics Gems*, Elsevier, 1994, pp. 474–485.
- [18] N. An, J. Xie, X. Zheng, and X. Gao, “Application of PCA in Concrete Infrared Thermography Detection,” in *Proceedings of the 2015 2nd International Workshop on Materials Engineering and Computer Sciences*, 2015.

- [19] Y. Fu and X.-J. Wu, "A Dual-Branch Network for Infrared and Visible Image Fusion," in *2020 25th International Conference on Pattern Recognition (ICPR)*, 2021, pp. 10675–10680.
- [20] Z. Zhao, S. Xu, J. Zhang, C. Liang, C. Zhang, and J. Liu, "Efficient and Model-Based Infrared and Visible Image Fusion via Algorithm Unrolling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 3, pp. 1186–1196, Mar. 2022.
- [21] Z. Zhao, S. Xu, C. Zhang, J. Liu, J. Zhang, and P. Li, "DIDFuse: Deep Image Decomposition for Infrared and Visible Image Fusion," in *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, 2020, pp. 970–976.
- [22] Y. Liu, X. Chen, J. Cheng, and H. Peng, "A medical image fusion method based on convolutional neural networks," in *2017 20th International Conference on Information Fusion (Fusion)*, 2017, pp. 1–7.
- [23] A. Toet, "The TNO Multiband Image Data Collection," *Data Br.*, vol. 15, pp. 249–251, Dec. 2017.
- [24] H. Xu, J. Ma, Z. Le, J. Jiang, and X. Guo, "FusionDN: A Unified Densely Connected Network for Image Fusion," *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 07, pp. 12484–12491, Apr. 2020.
- [25] Y. Fu, X.-J. Wu, and J. Kittler, "Effective method for fusing infrared and visible images," *J. Electron. Imaging*, vol. 30, no. 03, May 2021.
- [26] X. Liu, H. Gao, Q. Miao, Y. Xi, Y. Ai, and D. Gao, "MFST: Multi-Modal Feature Self-Adaptive Transformer for Infrared and Visible Image Fusion," *Remote Sens.*, vol. 14, no. 13, p. 3233, Jul. 2022.
- [27] J. Ma, H. Xu, J. Jiang, X. Mei, and X.-P. Zhang, "DDcGAN: A Dual-Discriminator Conditional Generative Adversarial Network for Multi-Resolution Image Fusion," *IEEE Trans. Image Process.*, vol. 29, pp. 4980–4995, 2020.
- [28] J. Ma, W. Yu, P. Liang, C. Li, and J. Jiang, "FusionGAN: A generative adversarial network for infrared and visible image fusion," *Inf. Fusion*, vol. 48, pp. 11–26, Aug. 2019.
- [29] H. Xu, H. Zhang, and J. Ma, "Classification Saliency-Based Rule for Visible and Infrared Image Fusion," *IEEE Trans. Comput. Imaging*, vol. 7, pp. 824–836, 2021.
- [30] H. Xu, M. Gong, X. Tian, J. Huang, and J. Ma, "CUFD: An encoder–decoder network for visible and infrared image fusion based on common and unique feature decomposition," *Comput. Vis. Image Underst.*, vol. 218, p. 103407, Apr. 2022.
- [31] L. Tang, X. Xiang, H. Zhang, M. Gong, and J. Ma, "DIVFusion: Darkness-free infrared and visible image fusion," *Inf. Fusion*, vol. 91, pp. 477–493, Mar. 2023.
- [32] Solakis, "Solakis Marble Enterprises." [Online]. Available: <https://www.solakismarble.com/>. [Accessed: 14-Mar-2023].
- [33] Seek Thermal, "CompactXR." [Online]. Available: <https://www.thermal.com/compact-series.html>. [Accessed: 14-Mar-2023].
- [34] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, "LabelMe: A Database and Web-Based Tool for Image Annotation," *Int. J. Comput. Vis.*, vol. 77, no. 1–3, pp. 157–173, May 2008.
- [35] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2117–2125.