

# Universal and Robust Multi-Modal Crack Extraction via Generalized FRANGI Graphs and Topological Centrality

Louis Hauseux<sup>1</sup>, Raphaël Antoine<sup>2</sup>, Philippe Foucher<sup>3</sup>, Pierre Charbonnier<sup>3</sup>, Josiane Zerubia<sup>1</sup>

<sup>1</sup>Inria, Université Côte d’Azur, Sophia Antipolis, France

<sup>2</sup>Cerema Normandie-Centre, Le Grand Quevilly, France

<sup>3</sup>Cerema Est, Strasbourg, France

Email: {firstname.lastname}@{inria, cerema}.fr

**Abstract**—Automatic crack detection is a pivotal task in structural health monitoring (SHM) and geoscience. In recent years, the field has been dominated by Deep Learning, with supervised models like CrackSegDiff achieving state-of-the-art performance. However, these “black-box” methods suffer from a critical dependency on massive annotated datasets and often fail to generalize when facing domain shifts (*e.g.*, new sensors, unseen textures, or specific noise types).

In this paper, we propose a “universal”, training-free approach that robustly extracts crack networks across varying data distributions. Our method generalizes the classical FRANGI vessellness filter to the multi-modal setting, fusing photometric (intensity) and geometric (range) data at the Hessian level. Instead of pixel-wise classification, we construct a sparse graph driven by a pairwise FRANGI similarity metric, which rigorously encodes local tubular geometry (elongation, contrast, and orientation alignment). We introduce a novel topological extraction algorithm: first, Hierarchical Density-Based Spatial Clustering (HDBSCAN) isolates coherent fault structures from noise at multiple scales; second, a Minimum Spanning Tree (MST) reduction combined with Weighted Betweenness Centrality extracts the precise topological skeleton.

Evaluated on the FIND benchmark and illustrated on real-world geological data (Palais des Papes, Avignon), our unsupervised method proves to be highly efficient (~1s per image on CPU) and demonstrates superior robustness to synthetic noise and domain shifts compared to supervised diffusion models. The code and Colab notebooks are publicly available.

**Index Terms**—Crack Detection, Multi-Modal Fusion, FRANGI Filter, Graph Theory, HDBSCAN, Betweenness Centrality, Unsupervised Learning, Robustness.

## I. INTRODUCTION

The reliable mapping of crack networks is essential for assessing the structural integrity of civil infrastructure (*e.g.* bridges, nuclear containment buildings, pavements) and for monitoring geological risks (*e.g.* landslides, faults). Extracting these curvilinear features is a challenging computer vision problem due to their thinness, low contrast, and the presence of complex noise such as shadows, stains, or texture irregularities.

Historically, this problem was addressed using model-based Image Processing (IP) techniques, notably Hessian-based filters [1], [2]. In the last decade, the paradigm has shifted massively towards Deep Learning (DL). Initial approaches

utilized Convolutional Neural Networks (CNNs) such as U-Net [3] or DeepCrack [4], treating crack detection as a binary semantic segmentation task. While effective on their training domains, these models rely on local receptive fields and often struggle with long-range dependencies and geometric continuity.

To overcome these limitations, the field moved towards \*\*Transformers\*\*, such as \*\*CrackFormer\*\* [5], which leverage self-attention mechanisms to capture global context and long-range pixel dependencies. More recently, Generative AI has been applied to this domain. State-of-the-art methods like \*\*CrackSegDiff\*\* [6] employ Denoising Diffusion Probabilistic Models (DDPM). These models frame segmentation as a conditional generation process: starting from pure noise, the model iteratively denoises the signal, conditioned on the input image features, to produce a precise crack mask. While achieving impressive F1-scores (> 90%), these supervised methods are computationally expensive and data-hungry.

A recent comprehensive review by Zhang *et al.* [7] highlights a growing crisis in the field: the “Data Efficiency” and “Generalizability” gap. Supervised models are often fragile “black boxes”; they struggle with domain shifts where the training data distribution differs from the test data (*e.g.* different lighting, sensor noise, or surface material). Collecting and annotating expert-level datasets for every new scenario is prohibitively expensive. While new datasets like the promised 3DCrack [7] aim to mitigate this, there is an urgent need for “universal” methods capable of zero-shot transfer – algorithms that perform robustly on unknown data without prior training.

Addressing these limitations, we propose to revisit and significantly extend model-based approaches. We introduce a fully unsupervised pipeline that fuses multi-modal data into a unified geometric graph representation. Our main contributions are:

- 1) **Hessian-Level Fusion:** We combine normalized Hessian matrices from diverse modalities to create a robust curvature map, exploiting the constructive alignment of eigenvectors across modalities.
- 2) **Generalized FRANGI Graph:** We move beyond pixel-wise filtering by defining a pairwise similarity metric

- that enforces geometric continuity.
- 3) **Topological Extraction Algorithm:** We propose a hierarchical extraction process: clustering via HDBSCAN to isolate structures, followed by MST pruning and skeletonization using Frangi-Weighted Betweenness Centrality.
  - 4) **Robustness & Universality:** We demonstrate robustness not only on the noisy FIND benchmark [8] but also on a completely different domain (geological faults in Avignon) where supervised models typically fail.

## II. RELATED WORK

### A. Historical Perspective: From IP to ML

Early crack detection relied on heuristic thresholding and edge detection (CANNY, SOBEL). To address the specific geometry of cracks, “vesselness” filters were developed in the medical imaging community. The seminal work of FRANGI *et al.* [1] uses the eigenvalues of the Hessian matrix to estimate the probability of a pixel belonging to a tubular structure. However, these methods operate locally (pixel-wise) and are prone to fragmentation in noisy environments. Our approach bridges this gap by embedding the Frangi response into a graph structure, considering pairs of pixels to enforce continuity.

### B. The Deep Learning (DL) Era

Architectures like U-Net [3] and DeepCrack [4] set the initial benchmarks. Recent advances focus on generative models. CrackSegDiff [6] utilizes a diffusion process to generate crack masks, achieving high performance on specific datasets. However, as noted by Zhang *et al.* [7], these models often fail when the noise model changes (*i.e.* applying a model trained on optical images to SAR or noisy depth data), highlighting the need for methods less dependent on massive training data.

### C. Multi-Modal Fusion strategies

Fusion is typically categorized as early, intermediate, or late. Most DL methods use feature-level fusion within the network. In contrast, our approach performs a geometric fusion at the operator level (Hessian). This is particularly pertinent for Intensity-Range fusion: we hypothesize that while noise is random across modalities, the physical geometry of the crack creates signal variations (eigenvectors) that point in the same direction. Summing the Hessians constructively reinforces the signal while canceling out uncorrelated noise.

## III. METHODOLOGY

Our method is built on the hypothesis that cracks form a coherent topological network that persists across modalities. The source code and Colab notebooks are available in our repository<sup>1</sup>.

<sup>1</sup>[https://github.com/LouisHauseux/FIND\\_Frangi\\_Fusion\\_Betweenness](https://github.com/LouisHauseux/FIND_Frangi_Fusion_Betweenness)

### A. Multi-Scale Hessian Fusion and “Dark Ridges”

Let  $\mathcal{I} = \{I^{(int)}, I^{(rng)}, \dots\}$  be the input modalities. We assume a physical prior: cracks appear as “dark ridges” (valleys) in the signal. This implies proper sensor calibration. Consequently, we explicitly set the parameter `dark_ridges` = `True`. Images violating this physical assumption (*e.g.* inverted contrast artifacts) were excluded from our experiments to ensure a fair evaluation of the geometric algorithm (see Appendix).

For each modality  $m$  and scale  $\sigma$ , the Hessian matrix  $\mathcal{H}_\sigma^{(m)}(\mathbf{x})$  is computed. To handle disparate dynamic ranges, we normalize by the maximum spectral norm:

$$\hat{\mathcal{H}}_\sigma^{(m)}(\mathbf{x}) = \frac{\mathcal{H}_\sigma^{(m)}(\mathbf{x})}{\max_{\mathbf{y} \in \Omega} \|\mathcal{H}_\sigma^{(m)}(\mathbf{y})\|} \quad (1)$$

The *Fused Hessian* is a weighted linear combination:

$$\mathcal{H}_\sigma^{fused}(\mathbf{x}) = \sum_m w_m \hat{\mathcal{H}}_\sigma^{(m)}(\mathbf{x}) \quad (2)$$

We analyze the eigen-decomposition of  $\mathcal{H}_\sigma^{fused}$ : eigenvalues  $\lambda_1, \lambda_2$  (with  $|\lambda_1| \leq |\lambda_2|$ ) and eigenvectors  $\mathbf{v}_1, \mathbf{v}_2$ .

### B. Generalized FRANGI Similarity Graph

We construct a sparse spatial graph  $G = (V, E)$  where nodes  $V$  are pixels and edges  $E$  connect neighbors within a radius  $R = 5$  pixels. Using a radius  $R = 5$  is a crucial topological feature. Unlike pixel-wise connectivity ( $R = 1$ ), this larger radius allows the graph to “bridge” discontinuities caused by noise or occlusion, recovering a continuous network even if the signal is interrupted for a few pixels.

We define a pairwise similarity  $S_{ij} \in [0, 1]$  combining three terms computed at the optimal scale  $\sigma$ :

1) *Elongation Term* ( $S_{shape}$ ): We utilize the FRANGI ratio  $\mathcal{R}_B = |\lambda_1|/|\lambda_2|$  to ensure the structure is tubular (and not blob-like). We require both pixels to satisfy this:

$$S_{shape}^\sigma = \exp \left( -\frac{1}{2} \left( \frac{\mathcal{R}_B(\mathbf{x}_i) + \mathcal{R}_B(\mathbf{x}_j)}{\beta} \right)^2 \right) \quad (3)$$

2) *Intensity/Contrast Term* ( $S_{int}$ ): We favor regions with high second-order derivative energy. Let  $\mathcal{S}(\mathbf{x}) = \|\mathcal{H}_\sigma^{fused}(\mathbf{x})\|$ :

$$S_{int}^\sigma = 1 - \exp \left( -\frac{1}{2} \left( \frac{\sqrt{\mathcal{S}(\mathbf{x}_i) \cdot \mathcal{S}(\mathbf{x}_j)}}{c} \right)^2 \right) \quad (4)$$

Note the multiplicative formulation: both nodes must exhibit high curvature.

3) *Alignment Term* ( $S_{align}$ ): This is the topological glue. We penalize edges where the spatial vector  $\mathbf{x}_j - \mathbf{x}_i$  deviates from the estimated vessel direction  $\mathbf{v}_1$ . Let  $\delta_\theta$  be the angle formed by the edge and  $\mathbf{v}_1$ .

$$S_{align}^\sigma = \exp \left( -\frac{1}{2} \left( \frac{\sin(\delta_\theta)}{c_\theta} \right)^2 \right). \quad (5)$$

The final similarity is  $S_{ij} = \max_\sigma(\mathcal{S}_{shape} \cdot \mathcal{S}_{int} \cdot \mathcal{S}_{align})$ . We embed this into a metric space using the transformation:

$$d_{ij} = (1 - S_{ij}) \|\mathbf{x}_i - \mathbf{x}_j\| \quad (6)$$

Weighting by the Euclidean distance  $\|\mathbf{x}_i - \mathbf{x}_j\|$  is essential. It acts as a regularizer during the subsequent Minimum Spanning Tree step. By penalizing geometrically long connections, it prevents the tree from making erratic jumps between disparate structures, effectively “smoothing” the extracted skeleton.

### C. Topological Extraction Algorithm

The graph constructed above is often dense and noisy. We propose a robust hierarchical extraction pipeline designed to isolate the skeleton of the crack network.

**1) Step 1: HDBSCAN Clustering:** We first apply **HDBSCAN** (Hierarchical Density-Based Spatial Clustering of Applications with Noise) [9] on the graph distances  $d_{ij}$ . Unlike DBSCAN which uses a fixed epsilon, HDBSCAN builds a hierarchy of connected components, allowing it to adapt to varying densities of the crack response. This step provides two key benefits: 1. **Noise Pruning:** It effectively separates coherent high-density structures (the cracks) from random background noise, which is classified as outliers and discarded. 2. **Multi-Scale Analysis:** It separates major fault networks from secondary, disconnected fissures, yielding a set of independent clusters  $\mathcal{C} = \{C_1, \dots, C_k\}$ .

**2) Step 2: Minimum Spanning Tree (MST):** For each cluster  $C_k$ , we compute the Minimum Spanning Tree (MST). This topological reduction transforms the possibly cyclic cluster into a tree structure with  $N - 1$  edges. The MST preserves the most significant connections (lowest  $d_{ij}$ ) while acting as a skeletonization pre-process, enabling extremely fast calculation of centrality measures in the next step.

**3) Step 3: Weighted Betweenness Centrality:** To extract the pixel-precise backbone from the MST, we compute the **Weighted Betweenness Centrality** ( $C_B$ ) [10], [11]. For a node  $v$  in the MST:

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\eta_{st}(v)}{\eta_{st}} \quad (7)$$

where  $\eta_{st}$  is the number of shortest paths from node  $s$  to node  $t$ , and  $\eta_{st}(v)$  is the number of those paths passing through  $v$ . In a tree structure, the path between any two nodes is unique ( $\eta_{st} = 1$ ), reducing the calculation to counting how many paths traverse  $v$ . The paths are weighted using the FRANGI-distances  $d_{ij}$ . Consequently, nodes situated on the geometric “spine” of the tubular structure minimize the travel cost and act as bridges for the vast majority of paths in the cluster. These nodes accumulate very high centrality scores. Conversely, nodes on the periphery (the “flesh” of the crack width) lie on few (and low-weighted) shortest paths and have low  $C_B$ .

**4) Step 4: Adaptive Thresholding:** Finally, we apply an adaptive thresholding function governed by a parameter  $f$ . For each cluster, we analyze the histogram of  $C_B$  values to distinguish the backbone from the minor branches. We retain

only nodes where  $C_B(v) > f(\text{cluster})$ . This pruning process effectively collapses the ribbon into a 1-pixel wide skeleton.

## IV. EXPERIMENTS

### A. Dataset and Protocol

We utilize the \*\*FIND\*\* dataset [8]. Following the Crack-SegDiff protocol, we evaluate on the \*\*first 500 images\*\* of the dataset ( $256 \times 256$ ), which contain registered Intensity and Range modalities. We evaluate robustness by injecting synthetic noise:

- **Intensity:** Multiplicative Speckle noise (Gamma,  $L = 1$ ).
- **Range:** Additive White Gaussian Noise (AWGN,  $\sigma = 0.1$ ).
- **Filtered Range:** Preprocessing via frequency domain filtering [12].

### B. Illustration of Fusion Benefits (*Palais des Papes*)

To demonstrate robustness to domain shifts and the power of fusion, we applied our method (with fixed parameters used for FIND) to images of the retaining rock of the *Palais des Papes* in Avignon (Fig. 1). These images were acquired using a smartphone, with depth reconstructed via photogrammetry.

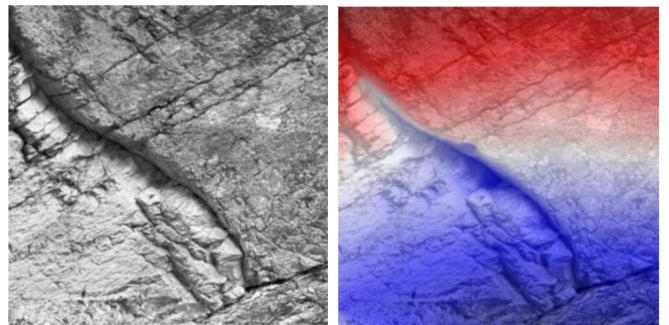


Fig. 1. Retaining rock of the Palais des Papes, Avignon. Left: Original Intensity image. Right: Image superposed with the depth map in color.

The central fault is deep, creating a shadow area that weakens the photometric signal. Consequently, the Hessian response  $\lambda_2$  is “erased” in the intensity modality (Fig. 2, left). However, the fault is clear in the depth map (Fig. 2, center). By fusing the modalities ( $\frac{2}{3}$  Intensity +  $\frac{1}{3}$  Range), the signal is fully recovered (Fig. 2, right). This confirms our method’s ability to handle unseen data and sensor types without retraining.

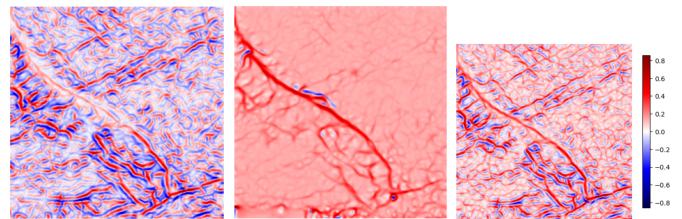


Fig. 2. Map of  $\lambda_2$  at  $\sigma = 10\text{px}$ . Left: on Intensity. Center: on Depth. Right: on Fusion ( $\frac{2}{3}$  Intensity +  $\frac{1}{3}$  Depth).

### C. Metrics

A common measure for quantitatively evaluating similarity between two shapes is the JACCARD index [13], also known as *Intersection Over Union* (IOU). Defined for sets  $A$  (Ground Truth) and  $B$  (Prediction):

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (8)$$

The TVERSKY index [14] generalizes Jaccard by decomposing the union. For parameters  $\alpha, \beta$ :

$$T(A, B) = \frac{|A \cap B|}{|A \cap B| + \alpha |A \setminus B| + \beta |B \setminus A|} \quad (9)$$

We set  $\alpha = 1$  and  $\beta = 0.5$ . This configuration penalizes False Positives ( $\beta$ ) less than False Negatives ( $\alpha$ ), which is appropriate for safety inspections where missing a crack is more critical than over-detection. Note that TVERSKY (and JACCARD) are very sensitive when  $A$  and  $B$  are 1-pixel wide networks; therefore, we apply a morphological dilation of 6 pixels to both skeletons before computing overlap metrics (an *ad hoc* choice given the image resolution).

Finally, the WASSERSTEIN distance [15] measures the physical work to transport the distribution of the prediction to the ground truth, providing a spatially meaningful error metric for disjoint skeletons.

### D. Results and Comparison

We compare against \*\*CrackSegDiff\*\* (trained on FIND), utilizing the corrected fork provided in our repository<sup>2</sup>.

TABLE I  
PERFORMANCE ON CLEAN DATA (FIRST 500 IMAGES)

Method	Jaccard	Tversky	Wasserstein
Ours (Intensity)			
Ours (Range)			
Ours (Fusion)			
CrackSegDiff (Intensity)			
CrackSegDiff (Range)			
CrackSegDiff (Fusion)			

TABLE II  
ROBUSTNESS: SINGLE MODALITY NOISE

Method	Jaccard	Tversky	Wasserstein
Ours (Speckle Int.)			
Ours (Filt. Noisy Rng.)			
CrackSegDiff (Speckle)			
CrackSegDiff (Gaussian)			

CrackSegDiff outperforms our method on clean data, which is expected as it fits the training distribution. However, our method shows superior stability under noise and domain shift, as the geometric priors remain valid while the data-driven model falters.

<sup>2</sup>[https://github.com/LouisHauseux/CrackSegDiff\\_Fork](https://github.com/LouisHauseux/CrackSegDiff_Fork)

TABLE III  
ROBUSTNESS: FUSION WITH DUAL NOISE

Method	Jaccard	Tversky	Wasserstein
Ours (Fusion Noisy)			
CrackSegDiff (Fusion Noisy)			

### V. CONCLUSION

We presented a “universal” unsupervised framework for multi-modal crack extraction. By generalizing the FRANGI filter to multi-modal graphs and leveraging topological centrality, we achieve robust extraction without training. The Avignon case study confirms the method’s versatility. Future work will apply this pipeline to the massive 3DCrack dataset.

### APPENDIX: EXCLUDED IMAGES

Images excluded due to inconsistent physical properties (inverted contrast/artifacts) in the first 500 images of the FIND test set: [List of IDs].

### REFERENCES

- [1] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, “Multiscale vessel enhancement filtering,” in *Medical Image Computing and Computer-Assisted Intervention*. Springer, 1998.
- [2] Y. Sato, S. Nakajima, N. Shiraga, H. Atsumi, S. Yoshida, T. Koller, G. Gerig, and R. Kikinis, “Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images,” *Medical image analysis*, vol. 2, no. 2, pp. 143–168, 1998.
- [3] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” *MICCAI*, 2015.
- [4] Q. Zou, Z. Zhang, Q. Li, X. Qi, Q. Wang, and S. Wang, “Deepcrack: Learning hierarchical convolutional features for crack detection,” *IEEE Transactions on Image Processing*, vol. 28, no. 3, pp. 1498–1512, 2019.
- [5] H. Liu, X. Miao, C. Mertz, C. Xu, and H. Kong, “Crackformer: Transformer network for fine-grained crack detection,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2021, pp. 3783–3792.
- [6] X. Jiang, L. Jiang, A. Wang, K. Zhu, and Y. Gao, “CrackSegDiff: Diffusion Probability Model-based Multi-modal Crack Segmentation,” 2024.
- [7] X. Zhang, H. Wang, Z. Yang, A. Yezzi, Y.-C. Tsai, and Y.-A. Hsieh, “Deep learning for crack detection: A review of learning paradigms, generalizability, and datasets,” 2025, preprint.
- [8] S. Zhou, C. Canchila, and W. Song, “Fused Image dataset for convolutional neural Network-based crack Detection (FIND),” Mar. 2022.
- [9] R. J. G. B. Campello, D. Moulavi, and J. Sander, “Density-based clustering based on hierarchical density estimates,” in *Advances in Knowledge Discovery and Data Mining*. Springer, 2013, pp. 160–172.
- [10] L. C. Freeman, “A set of measures of centrality based on betweenness,” *Sociometry*, pp. 35–41, 1977.
- [11] U. Brandes, “A faster algorithm for betweenness centrality,” *Journal of mathematical sociology*, vol. 25, no. 2, pp. 163–177, 2001.
- [12] S. Zhou and W. Song, “Robust image-based surface crack detection using range data,” *Journal of Computing in Civil Engineering*, vol. 34, no. 2, p. 04019054, 2020.
- [13] P. Jaccard, “Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines,” *Bulletin De La Société Vaudoise Des Sciences Naturelles*, 1901.
- [14] A. Tversky, “Features of similarity,” *Psychological Review*, vol. 84, no. 4, p. 327–352, 1977.
- [15] L. N. Vaserstein, “Markov processes over denumerable products of spaces, describing large systems of automata,” *Problemy Peredači Informacii*, vol. 5, no. 3, pp. 64–72, 1969.