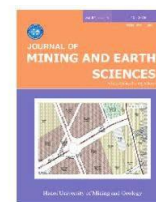




Journal of Mining and Earth Sciences

Website: <https://jmes.humg.edu.vn>



Geometric evaluation in crack segmentation: an analysis of the complementary relationship between region-based (IoU) and boundary-based (BF1) metrics



Phuong Ngoc Vu¹, Dung Trung Pham^{2,*}

¹ University of Transport and Communications, Hanoi, Vietnam

² Hanoi University of Mining and Geology, Hanoi, Vietnam

ARTICLE INFO

Article history:

Received 15th June 2026

Revised 02nd Sept. 2026

Accepted 12th Sept. 2026

Keywords:

Boundary F1-score (BF1),
Crack segmentation,
Geometric evaluation,
Intersection over Union (IoU),
U-Net.

ABSTRACT

Crack segmentation from images plays a crucial role in infrastructure inspection and maintenance, where the reliability of predicted masks directly affects the assessment of structural deterioration. Although the Intersection over Union (IoU) metric is widely adopted for segmentation evaluation, it primarily measures region overlap and may not adequately capture geometric boundary accuracy for thin, elongated structures such as cracks. This study investigates the complementary relationship between the region-based IoU metric and the boundary-based Boundary F1-score (BF1), with the aim of clarifying their respective strengths and limitations in crack segmentation assessment. Experiments were conducted using a U-Net model with a ResNet-34 encoder on two datasets with distinct labeling characteristics, including a multi-source ground-image dataset and a consistently annotated UAV dataset. In addition, a controlled boundary-deviation experiment was designed to directly evaluate the response of both metrics to predefined geometric distortions. The results indicate that IoU tends to increase with crack width and mainly reflects area overlap, whereas BF1 is more sensitive to boundary localization errors and annotation quality. For wide cracks, IoU achieved similar values across both datasets (0.706), while BF1 clearly discriminated data quality, increasing from 0.273 on the multi-source dataset to 0.875 on the UAV dataset. The controlled experiments further demonstrated that BF1 decreases once boundary deviations exceed the tolerance threshold θ , whereas IoU degrades more gradually with increasing geometric distortion. These findings demonstrate that IoU and BF1 characterize different geometric aspects of segmentation quality and therefore should not be used interchangeably. Based on this evidence, the study recommends a standardized evaluation framework that reports both region-based and boundary-based metrics to provide a more reliable assessment of thin linear structure segmentation.

Copyright © 2026 Hanoi University of Mining and Geology. All rights reserved.

*Corresponding author

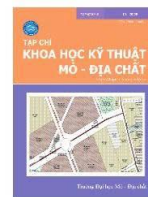
E - mail: phamtrungdung@humg.edu.vn

DOI: 10.46326/JMES.2026.67(5).09



Tạp chí Khoa học Kỹ thuật Mỏ - Địa chất

Trang điện tử: <https://tapchi.humg.edu.vn>



Đánh giá hình học trong phân đoạn vết nứt: phân tích tính tương hỗ giữa chỉ số vùng (IoU) và chỉ số biên (BF1)

Vũ Ngọc Phượng¹, Phạm Trung Dũng^{2,*}

¹ Trường Đại học Giao thông Vận tải, Hà Nội, Việt Nam

² Trường Đại học Mỏ - Địa chất, Hà Nội, Việt Nam

THÔNG TIN BÀI BÁO

Quá trình:

Nhận bài 15/6/2026

Sửa xong 02/9/2026

Chấp nhận đăng 12/9/2026

Từ khóa:

Boundary F1-score,
Đánh giá hình học,
Intersection over Union,
Phân đoạn vết nứt,
U-Net.

TÓM TẮT

Phân đoạn vết nứt từ ảnh là một nhiệm vụ quan trọng trong kiểm tra và bảo trì kết cấu hạ tầng, trong đó độ tin cậy của kết quả dự báo ảnh hưởng trực tiếp đến việc đánh giá mức độ hư hỏng. Mặc dù chỉ số Intersection over Union (IoU) được sử dụng rộng rãi để đánh giá kết quả phân đoạn, chỉ số này chủ yếu phản ánh mức độ chồng lấp vùng và chưa mô tả đầy đủ độ chính xác hình học của ranh giới đối với các đối tượng dạng tuyến mảnh như vết nứt. Nghiên cứu này phân tích mối quan hệ bổ sung giữa chỉ số vùng IoU và chỉ số biên Boundary F1-score (BF1), nhằm làm rõ vai trò và hạn chế của từng chỉ số trong đánh giá phân đoạn vết nứt. Thử nghiệm được thực hiện với mô hình U-Net sử dụng bộ mã hóa ResNet-34 trên hai bộ dữ liệu có đặc tính nhãn khác nhau, gồm tập ảnh đa nguồn mặt đất và tập ảnh UAV được gán nhãn nhất quán. Bên cạnh đó, một thí nghiệm sai lệch biên có kiểm soát được xây dựng để đánh giá trực tiếp phản ứng của các chỉ số trước các biến dạng hình học. Kết quả cho thấy IoU có xu hướng tăng theo bề rộng vết nứt và chủ yếu phản ánh mức độ chồng lấp diện tích, trong khi BF1 nhạy hơn với sai lệch vị trí ranh giới và chất lượng nhãn gán. Trên nhóm vết nứt rộng, IoU đạt giá trị tương đương giữa hai bộ dữ liệu (0,706), nhưng BF1 phân biệt rõ chất lượng dữ liệu, tăng từ 0,273 ở tập đa nguồn lên 0,875 ở tập UAV. Thí nghiệm sai lệch có kiểm soát cho thấy BF1 suy giảm mạnh khi sai lệch biên vượt quá ngưỡng dung sai, trong khi IoU biến đổi tương đối từ từ theo mức độ biến dạng. Những kết quả này chứng minh rằng IoU và BF1 phản ánh các khía cạnh hình học khác nhau và không thể thay thế cho nhau trong đánh giá phân đoạn vết nứt. Trên cơ sở đó, nghiên cứu đề xuất quy trình đánh giá chuẩn hóa theo hướng báo cáo đồng thời chỉ số vùng và chỉ số biên nhằm nâng cao độ tin cậy trong đánh giá các cấu trúc dạng tuyến mảnh.

© 2026 Trường Đại học Mỏ - Địa chất. Tất cả các quyền được bảo đảm.

*Tác giả liên hệ

E - mail: phamtrungdung@humg.edu.vn

DOI: 10.46326/JMES.2026.67(5).09

1. Mở đầu

Khảo sát và đánh giá vết nứt là nhiệm vụ quan trọng trong công tác quản lý và bảo trì hạ tầng. Sự phát triển của công nghệ xử lý ảnh và học sâu cho phép phát hiện và phân đoạn vết nứt trực tiếp từ ảnh ở cấp độ điểm ảnh (pixel), qua đó trích xuất các đặc trưng hình học như chiều dài, chiều rộng và hướng phát triển của vết nứt. Tuy nhiên, phân đoạn vết nứt vẫn là một bài toán thách thức do nền ảnh phức tạp, điều kiện chiếu sáng thay đổi, kết cấu bề mặt không đồng nhất và đặc trưng hình học mảnh và không đồng đều của vết nứt (Du và nnk., 2021; Kheradmandi & Mehranfar, 2022). Các tiêu chuẩn đánh giá tình trạng đường bộ cũng được quy định trong các tài liệu kỹ thuật hiện hành (Bộ Khoa học và Công nghệ, 2024; Miller & Bellinger, 2014).

Để giải quyết những khó khăn này, nhiều nghiên cứu đã tập trung vào việc cải tiến các mô hình phân đoạn dựa trên học sâu. Các phương pháp ban đầu chủ yếu sử dụng mạng tích chập và kiến trúc encoder-decoder như FCN hoặc U-Net, sau đó được mở rộng với mạng phần dư, cơ chế chú ý và Transformer nhằm cải thiện khả năng nhận dạng vết nứt mảnh trong nền phức tạp (Al-Huda và nnk., 2023; Dung & Anh, 2019; Fei và nnk., 2020; Kang & Cha, 2022; Yu và nnk., 2023). Gần đây, các hướng tiếp cận mới như mô hình lai (hybrid) hoặc mô hình khuếch tán cũng được đề xuất nhằm tăng cường năng lực biểu diễn và khả năng tổng quát hóa (Chen và nnk., 2024; Goo và nnk., 2025). Nhìn chung, phần lớn các nghiên cứu hiện nay tập trung vào việc nâng cao hiệu quả mô hình và chất lượng dữ liệu.

Khi hiệu suất mô hình ngày càng được cải thiện, cách đánh giá chất lượng mặt nạ phân đoạn trở thành vấn đề quan trọng cần được xem xét. Trong nhiều nghiên cứu, các chỉ số đánh giá theo điểm ảnh hoặc chồng lấp vùng như IoU, Dice vẫn được sử dụng phổ biến. Tuy nhiên, các chỉ số này chủ yếu phản ánh mức độ trùng khớp về diện tích mà không đánh giá trực tiếp độ chính xác hình học theo ranh giới. Trong khi đó, một số chỉ số dựa trên biên như Boundary F1-score (BF1), Boundary IoU hay cDice đã được đề xuất nhằm bổ sung hạn chế này trong các bài toán phân đoạn tổng quát (Cheng và nnk., 2021; Csurka và nnk., 2013; Korczynski, 2026; Shit và nnk., 2021). Tuy nhiên, vai trò của các chỉ số này đối với phân đoạn

vết nứt chưa được phân tích một cách có hệ thống. Các chỉ số này đã được khảo sát chủ yếu trên đối tượng có vùng lớn và biên khép kín trong phân đoạn ảnh tổng quát. Trong khi Boundary IoU được đánh giá trên COCO và Cityscapes (Cheng và nnk., 2021), cDice lại hướng tới bảo toàn tính liên thông của cấu trúc dạng ống (Shit và nnk., 2021). Hiện nay, các nghiên cứu về phân đoạn vết nứt hầu như vẫn chỉ sử dụng các chỉ số IoU, Dice hoặc F1 theo điểm ảnh. Các chỉ số đánh giá biên vẫn chưa được áp dụng để phân tích cho phân đoạn vết nứt. Khoảng trống nghiên cứu này là cơ sở để triển khai các phân tích trong bài báo.

Đối với vết nứt, sai lệch nhỏ ở mức vài điểm ảnh có thể dẫn đến cách diễn giải khác nhau về chất lượng phân đoạn. Một dự báo bám đúng tuyến phát triển có thể vẫn có IoU thấp do diện tích chồng lấp nhỏ, trong khi một mặt nạ có IoU cao vẫn có thể lệch biên, ảnh hưởng đến việc xác định vị trí hoặc chiều rộng vết nứt. Điều này cho thấy việc chỉ sử dụng chỉ số vùng chồng lấp có thể chưa đầy đủ để đánh giá đầy đủ chất lượng phân đoạn đối với đối tượng dạng tuyến như vết nứt.

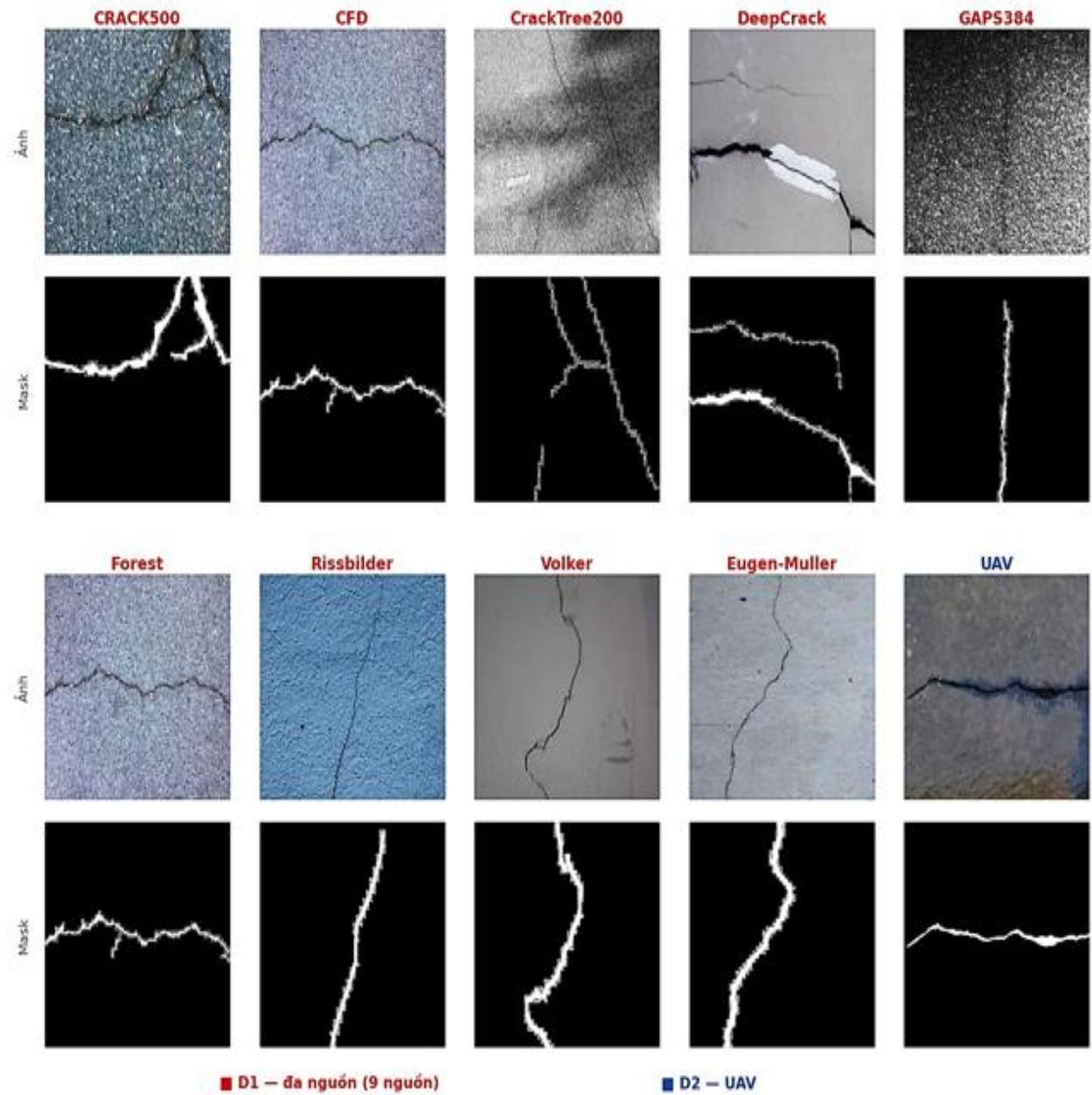
Nghiên cứu này phân tích đặc điểm và tính bổ sung giữa chỉ số vùng (IoU) và chỉ số biên (BF1) trong phân đoạn vết nứt. Đóng góp chính của bài báo là làm rõ vai trò của BF1, khi chỉ số này cung cấp thông tin ranh giới quan trọng mà IoU thiếu sót, đặc biệt đối với vết nứt mảnh, hỗ trợ đánh giá toàn diện hơn chất lượng dự báo của mô hình.

2. Dữ liệu và phương pháp nghiên cứu

2.1. Dữ liệu

Nghiên cứu sử dụng hai bộ dữ liệu vết nứt với đặc trưng khác biệt về nguồn thu nhận và chất lượng nhãn. Bộ dữ liệu thứ nhất (D1) gồm 1.695 ảnh chụp mặt đất tổng hợp từ 9 nguồn dữ liệu công khai khác nhau. Đặc điểm của các bộ dữ liệu này là nhãn được gán bởi nhiều nhóm và thiếu tính đồng nhất. Bộ dữ liệu thứ hai (D2) gồm 315 ảnh thu nhận từ thiết bị bay không người lái (UAV), với điều kiện thu nhận ổn định và nhãn được xây dựng theo một cách nhất quán (Hình 1).

Tất cả ảnh của cả hai bộ D1 và D2 đều được chuẩn hóa về kích thước 448×448 pixel và được chia thành các tập train/validation/test theo tỉ lệ tương ứng (Bảng 1). Ảnh không trùng lặp được phân chia ngẫu nhiên, độc lập thành ba tập huấn luyện, xác thực và kiểm tra theo tỷ lệ 70:20:10.



Hình 1. Ảnh và mặt nạ nhãn tương ứng của hai bộ dữ liệu D1 (đa nguồn) và D2 (UAV).

Bảng 1. Tóm tắt đặc điểm của các bộ dữ liệu vết nứt sử dụng trong nghiên cứu.

Bộ dữ liệu	Số lượng ảnh	Kích thước ảnh (pixel)	Phân chia dữ liệu train/val/test	Đặc điểm nhãn	Nguồn
D1 (ảnh mặt đất đa nguồn).	1.695	448×448	1.185/338/172	Nhiều nhóm gán nhãn, không đồng nhất, đặc biệt tại biên.	09 nguồn công khai mặt đất (Middha, 2020).
D2 (ảnh UAV).	315	448×448	220/63/32	Cùng một nhóm gán nhãn theo giao thức nhất quán.	Một nguồn UAV, góc chụp thẳng đứng (Ziya, 2025).

2.2. Mô hình và thiết lập thông số huấn luyện

Để tạo các kết quả phân đoạn phục vụ phân tích chỉ số đánh giá, nghiên cứu sử dụng mô hình U-Net (Ronneberger và nnk., 2015) với bộ mã hóa ResNet-34 được khởi tạo từ trọng số tiền huấn luyện trên tập ImageNet. Trong khi mạng U-Net đã được chứng minh hiệu quả cho phân đoạn vết nứt ở cấp độ điểm ảnh (Kang & Cha, 2022; Yu và nnk., 2023) thì ResNet-34 cho phép trích xuất đặc trưng sâu mà vẫn ổn định khi huấn luyện (He và nnk., 2016). Bên cạnh đó, độ sâu 34 cân bằng giữa năng lực biểu diễn và chi phí tính toán, phù hợp với tập dữ liệu vết nứt hạn chế trong nghiên cứu này.

Ảnh đầu vào là ảnh màu RGB kích thước 448×448 pixel, đầu ra là mặt nạ nhị phân vết nứt/nền với ngưỡng 0,5. Mô hình được huấn luyện bằng hàm mất mát Dice và bộ tối ưu hóa AdamW, kết hợp lịch trình điều chỉnh tốc độ học theo cơ chế khởi động ấm (warm-up) và suy giảm theo hàm cosine. Thực nghiệm được triển khai trên nền tảng PyTorch kết hợp thư viện segmentation_models_pytorch, thực thi trên bộ xử lý đồ họa NVIDIA RTX 4000 Ada (21,5 GB) ở chế độ độ chính xác hỗn hợp tự động (AMP/FP16). Các siêu tham số huấn luyện chính được tóm tắt trong Bảng 2.

Bảng 2. Thiết lập siêu tham số và cấu hình huấn luyện kiểm soát.

Tham số	Giá trị
Kiến trúc mạng	U-Net, encoder ResNet-34 (tiền huấn luyện ImageNet)
Kích thước ảnh đầu vào	448 × 448, RGB
Hàm mất mát	Dice Loss
Bộ tối ưu	AdamW
Learning rate (decoder/encoder)	$1,2 \times 10^{-4}$ / $1,2 \times 10^{-5}$
Weight decay	10^{-2}
Lịch learning rate	Warmup tuyến tính 5 epoch → cosine
Batch size	16
Epoch tối đa/ dừng sớm	100 / patience = 20
Ngưỡng nhị phân hóa	0,5
Số lần lặp (seed khác nhau)	3

Các giá trị khởi tạo được thiết lập theo cấu hình tiêu chuẩn cho việc tinh chỉnh bộ mã hóa tiền huấn luyện (He và nnk., 2016; Ronneberger và nnk., 2015); sau đó các siêu tham số như tốc độ học (learning rate), hệ số suy giảm trọng số (weight decay) và lịch trình tối ưu được hiệu chỉnh thực nghiệm dựa trên hiệu năng trên tập xác thực của bộ dữ liệu nghiên cứu.

2.3. Chỉ số đánh giá độ chính xác mô hình

Để đánh giá kết quả phân đoạn, nghiên cứu sử dụng hai chỉ số đại diện cho hai khía cạnh hình học khác nhau là IoU (Intersection over Union) và BF1 (Boundary F1-score). IoU đo mức độ chồng lấp vùng giữa mặt nạ dự báo và nhãn, trong khi BF1 phản ánh mức độ tương đồng hình học tại ranh giới thông qua ngưỡng dung sai khoảng cách θ . Việc sử dụng song song hai chỉ số này giúp cung cấp cái nhìn bổ sung về chất lượng phân đoạn, đặc biệt đối với các đối tượng dạng tuyến như vết nứt.

Trong phân đoạn nhị phân, các chỉ số được xây dựng từ ba đại lượng cơ bản ở cấp điểm ảnh gồm (dương tính thật - TP, dương tính giả - FP và âm tính giả - FN). Nhóm chỉ số chồng lấp vùng phổ biến bao gồm IoU, F1-score và Dice (Rahman & Wang, 2016; Sudre và nnk., 2017). Do các chỉ số này có quan hệ chặt chẽ, nghiên cứu lựa chọn IoU làm đại diện cho nhóm đánh giá sai lệch vùng để so sánh với chỉ số biên BF1. Chỉ số IoU được định nghĩa như sau:

$$IoU = \frac{|P \cap A|}{|P \cup A|} = \frac{TP}{TP + FP + FN} \in [0,1] \quad (1)$$

Trong đó: P và A lần lượt là mặt nạ dự báo và nhãn thực tế. IoU phản ánh mức độ chồng lấp diện tích trên toàn ảnh, do đó đại diện cho sai lệch vùng.

Khác với IoU, BF1 đánh giá mức độ tương đồng giữa biên dự báo và biên nhãn thông qua ngưỡng dung sai khoảng cách θ . Một điểm biên được xem là khớp nếu khoảng cách tới điểm biên gần nhất không vượt quá θ . Chỉ số BF1 được xác định là trung bình điều hòa giữa độ chính xác biên $P_B(\theta)$ và độ phủ biên $R_B(\theta)$ (Martin và nnk., 2004):

$$BF1(\theta) = \frac{2 \cdot P_B(\theta) \cdot R_B(\theta)}{P_B(\theta) + R_B(\theta)} \in [0,1] \quad (2)$$

Trong đó:

$$P_B(\theta) = \frac{1}{|B_p|} \sum_{q \in B_p} 1 \left(\min_{p \in B_g} \|q - p\| \leq \theta \right) \quad (3)$$

$$R_B(\theta) = \frac{1}{|B_g|} \sum_{p \in B_g} 1 \left(\min_{q \in B_p} \|p - q\| \leq \theta \right) \quad (4)$$

Với B_p và B_g lần lượt là tập điểm biên dự báo và biên chuẩn; $|B_p|$, $|B_g|$ là số lượng điểm trong mỗi tập; $1(\cdot)$ là hàm chỉ báo. Cách xác định điểm khớp dựa trên khoảng cách gần nhất giúp BF1 phản ánh trực tiếp sai lệch hình học của ranh giới thay vì chỉ đo mức chồng lấp vùng. Theo Martin và nnk. (2004), đánh giá biên cần cho phép sai khác nhỏ về vị trí do sự không chắc chắn trong gán nhãn ở mức điểm ảnh. Nghiên cứu của Csurka và nnk. (2013) đề xuất ngưỡng dung sai θ theo tỷ lệ đường chéo ảnh. Tuy nhiên, với các đối tượng mảnh như vết nứt, ngưỡng này có thể lớn hơn chiều rộng thực tế của đối tượng. Vì vậy, trong nghiên cứu này, θ được xác định trực tiếp từ đặc trưng hình học của vết nứt trên dữ liệu. Cụ thể, chiều rộng cục bộ của vết nứt được tính trên trục giữa (medial axis) của mặt nạ nhãn, thông qua khoảng cách tới nền, và giá trị trung vị được sử dụng làm đại diện nhằm giảm ảnh hưởng của nhiễu và các giá trị ngoại lai (Fan và nnk., 2020; Wang & Pang, 2025). Phân tích phân bố cho thấy phần lớn vết nứt có độ rộng không vượt quá 4 pixel. Do đó, nghiên cứu này chọn ngưỡng dung sai ở mức $\theta = 2$ pixel.

Ngưỡng dung sai θ trong chỉ số BF1 cần được xác định chặt chẽ thay vì nói rộng tùy ý, bởi việc tăng θ luôn làm điểm số tăng ảo do đặc tính của các chỉ số dựa trên dung sai. Ngưỡng $\theta = 2$ pixels được xác định là lựa chọn tối ưu duy nhất dựa trên hai giới hạn hình học:

(1) Cận trên ($\theta \leq 2$ pixels): Với bề rộng vết nứt trung vị bằng 2 pixels, nếu $\theta > 2$, các điểm ranh giới dự báo nằm hoàn toàn ngoài vết nứt vẫn được tính là chính xác, làm mất ý nghĩa hình học của phép đo.

(2) Cận dưới ($\theta \geq 2$ pixels): Nếu giảm θ xuống 1 pixel, dung sai không đủ để bù đắp sai số cỡ 1 điểm ảnh vốn có từ thao tác gán nhãn thủ công (Martin và nnk., 2004). Vì θ bắt buộc là số

nguyên, hai giới hạn trên loại trừ tất cả các giá trị ngoại trừ $\theta = 2$ pixels. Do đó, nghiên cứu không áp dụng phương pháp tính θ theo tỷ lệ đường chéo ảnh (Csurka và nnk., 2013), vốn tạo ra dung sai lớn hơn nhiều so với kích thước nét mảnh của vết nứt.

Về ý nghĩa vật lý, ngưỡng dung sai biên tương ứng với khoảng cách thực địa $d = \theta \times \text{GSD}$ (Ground Sample Distance). Mỗi quan hệ này cho phép tham chiếu dung sai đánh giá với các tiêu chuẩn phân cấp hư hỏng mặt đường khi dữ liệu có kèm thông số hiệu chuẩn (như hướng dẫn của FHWA (Miller & Bellinger, 2014)).

2.4. Thí nghiệm sai lệch biên có kiểm soát

Trong thực nghiệm, độ nhạy của các chỉ số đánh giá chịu tác động đồng thời từ nhiều yếu tố: sai lệch hình học, hiệu năng mô hình, đặc tính dữ liệu huấn luyện và chất lượng hình ảnh. Để cô lập riêng yếu tố hình học, nghiên cứu thiết kế một thí nghiệm biến dạng có kiểm soát trên mặt nạ nhãn chuẩn của tập kiểm tra D1. Tương tự phương pháp phân tích độ nhạy của Cheng và nnk. (2021), nhãn chuẩn được sử dụng làm tham chiếu để tổng hợp các mặt nạ dự báo mô phỏng với loại hình sai số và biên độ biến dạng được xác định trước. Qua đó, nghiên cứu định lượng chính xác độ nhạy của IoU và BF1 đối với từng dạng và cấp độ sai lệch hình học.

Ba dạng biến đổi hình học được lựa chọn gồm giãn biên (dilation), co biên (erosion) và dịch mặt nạ (shift), đại diện cho các dạng lỗi phổ biến trong phân đoạn ảnh. Trong đó, giãn và co biên được sử dụng để mô phỏng sai lệch tỉ lệ vùng phân đoạn, còn dịch mặt nạ mô phỏng sai lệch định vị đối tượng (Chen và nnk., 2024). Khi áp dụng cho vết nứt, giãn biên tương ứng với dự báo thừa làm vùng nứt dày hơn nhãn. Co biên phản ánh dự báo thiếu làm vết nứt bị mảnh đi hoặc mất đoạn, trong khi dịch mặt nạ giữ nguyên hình dạng tuyến nhưng làm lệch vị trí so với nhãn chuẩn.

Mức độ sai lệch được kiểm soát theo đơn vị pixel, cho phép phân tích định lượng sự thay đổi của IoU và BF1 theo từng mức biến dạng tăng dần. Nhờ đó, thí nghiệm không chỉ giúp quan sát trực tiếp phản ứng của từng chỉ số trước các dạng sai lệch hình học, mà còn cung cấp cơ sở để diễn giải các kết quả thực nghiệm thu được từ mô hình.

3. Kết quả thực nghiệm

3.1. Kết quả định tính

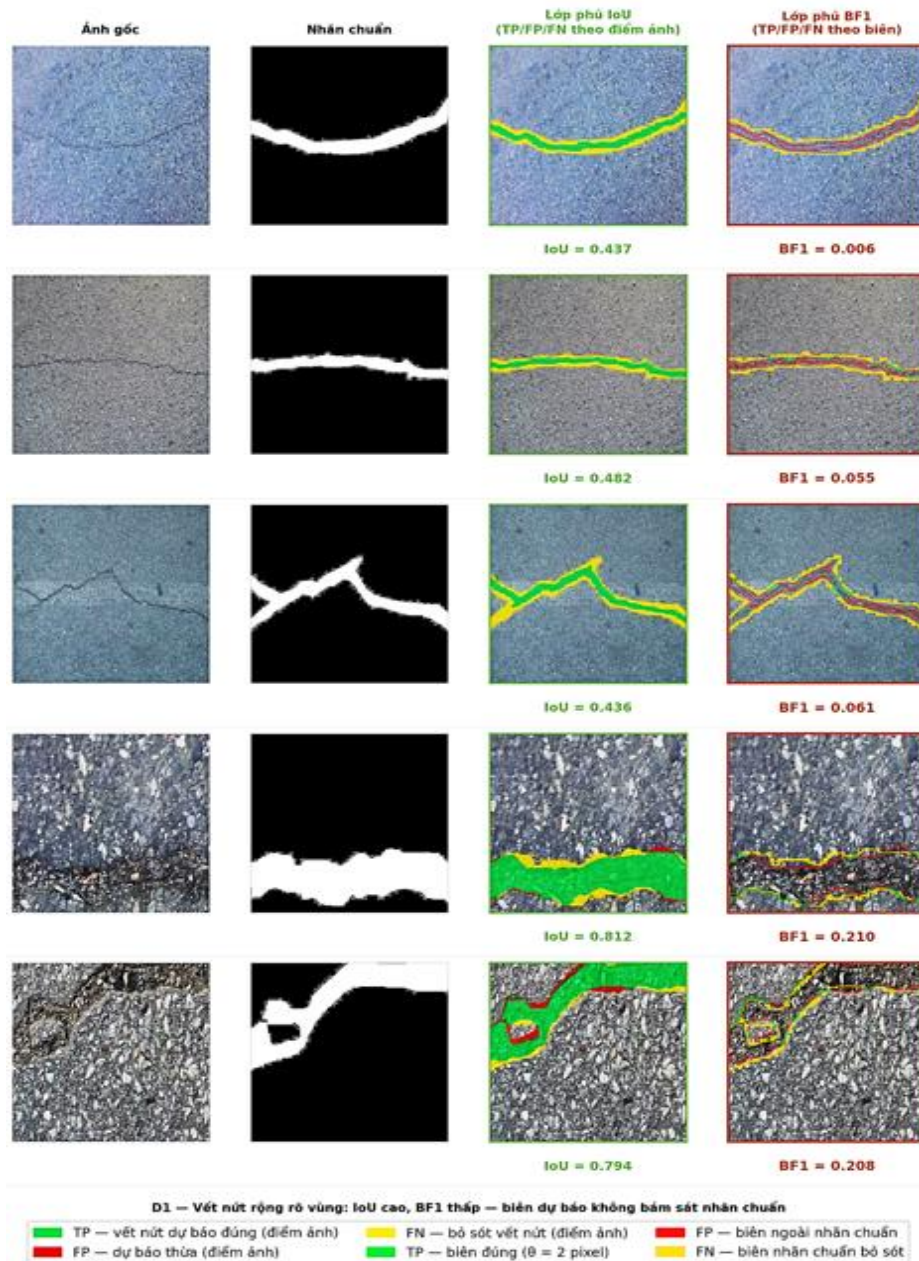
Kết quả dự đoán vết nứt của mô hình trên hai tập dữ liệu kiểm tra D1 (ảnh mặt đất) và D2 (ảnh UAV) được trực quan hóa qua các lớp phủ điểm ảnh (đại diện cho chỉ số IoU) và lớp phủ đường biên trong dung sai $\theta = 2$ pixel (đại diện cho chỉ số BF1).

Đối với nhóm vết nứt rộng, rõ vùng

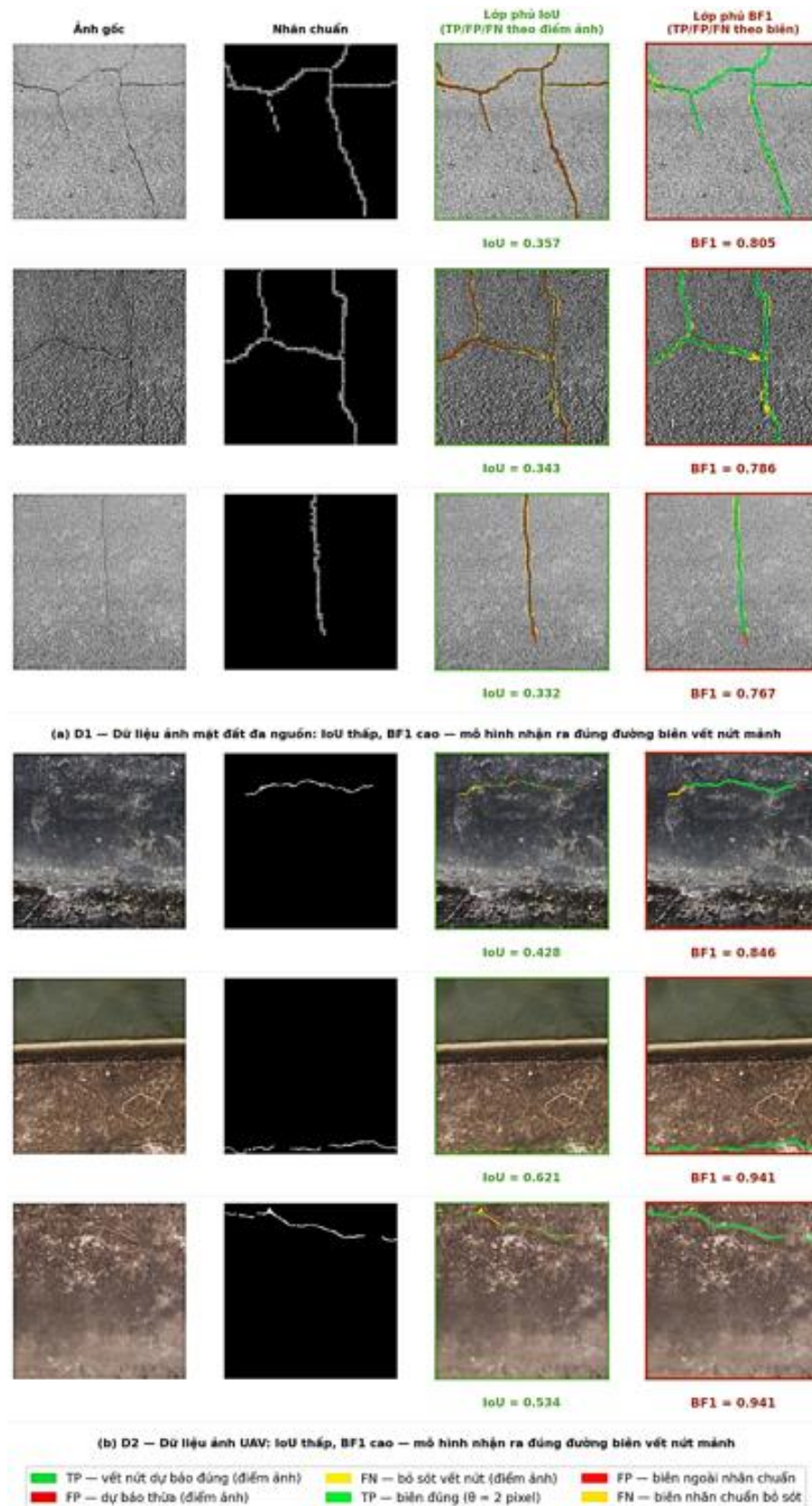
Hình 2 thể hiện vùng màu xanh lá (True Positive - TP) chiếm diện tích chủ đạo. Tuy nhiên,

tại các khu vực rìa và vùng tiếp giáp xuất hiện rải rác các điểm màu vàng (False Negative - FN) và màu đỏ (False Positive - FP). Trên cột lớp phủ BF1, các đường màu vàng (FN biên) xuất hiện dày đặc dọc theo cạnh vết nứt, trong khi các đoạn màu xanh lá (TP biên) chiếm tỷ lệ rất thấp. Riêng đối với mẫu có hình dạng kéo dài theo đường thẳng ngang đơn giản (mẫu 4 và mẫu 5), các đoạn biên dự báo chính xác có xu hướng tăng lên.

Đối với nhóm vết nứt mảnh (Hình 3)



Hình 2. Trường hợp IoU cao nhưng BF1 thấp trên vết nứt rộng. Các phép overlay thể hiện TP/FP/FN theo điểm ảnh (IoU) và theo biên (BF1).



Hình 3. Trường hợp BF1 cao hơn IoU trên vết nứt mảnh. IoU đánh giá theo chồng lớp vùng, trong khi BF1 phản ánh mức độ bám theo biên.

Trên cả hai tập dữ liệu D1 và D2, cột lớp phủ IoU cho thấy các điểm màu vàng (FN) chiếm ưu thế lớn, diện tích chồng lấp xanh lá (TP) xuất hiện rất ít. Ngược lại, trên cột lớp phủ BF1, đường màu xanh lá (TP biên) bám rõ theo tuyến vết nứt, tuy trên D1 vẫn xen kẽ các đoạn bỏ sót. Trên tập D2, đường biên đúng liên tục và bao phủ gần hết chiều dài vết nứt, tiêu biểu là mẫu 5 và mẫu 6.

3.2. Kết quả định lượng

Hiệu suất định lượng của mô hình trên hai tập dữ liệu kiểm tra D1 và D2 được tổng hợp và phân loại theo độ rộng vết nứt (ngưỡng phân tách $2\theta = 4$ pixel) trong Bảng 3.

Chỉ số IoU

Giá trị IoU trung bình của nhóm vết nứt rộng gần như bằng nhau trên cả hai tập dữ liệu (0,706). Khi chuyển sang nhóm vết nứt mảnh, chỉ số này sụt giảm xuống mức 0,542 (trên D1) và 0,546 (trên D2). Xét về độ ổn định, nhóm D2-nứt mảnh có độ lệch chuẩn thấp nhất ($\sigma = 0,091$), trong khi nhóm D1-nứt mảnh có độ lệch chuẩn cao nhất ($\sigma = 0,144$). Phân tích phân bố liên tục (Hình 4) ghi nhận hệ số tương quan Spearman giữa độ rộng vết nứt và IoU đạt giá trị dương trên cả hai tập: $\rho = +0,38$ (D1) và $\rho = +0,60$ (D2).

Chỉ số BF1 và các thành phần biên

Trên tập D1, chỉ số BF1 trung bình của nhóm vết nứt rộng ở mức thấp (0,273), nhưng tăng đáng kể lên 0,531 ở nhóm vết nứt mảnh. Ngược lại, trên tập D2, chỉ số BF1 duy trì ở mức cao và ổn định trên cả hai nhóm vết nứt rộng (0,875) và nứt mảnh (0,842). Hệ số tương quan Spearman giữa độ rộng vết nứt và BF1 (Hình 4) đạt giá trị âm rõ rệt trên D1 ($\rho = -0,55$) nhưng đạt giá trị dương nhẹ gần bằng không trên D2 ($\rho = +0,15$).

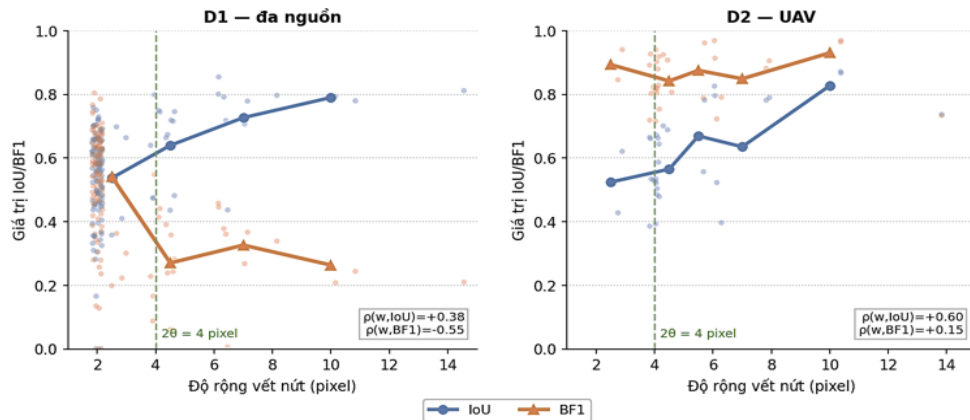
Đối với các chỉ số thành phần biên, hiệu số giữa độ chính xác biên ($P_B(\theta)$) và độ phủ biên ($R_B(\theta)$) đạt giá trị lớn nhất ở nhóm D1-nứt mảnh ($0,767 - 0,419 = +0,348$) và nhóm D1-nứt rộng ($0,410 - 0,207 = +0,203$). Trong khi đó, hiệu số này giảm xuống mức rất thấp trên tập D2 với $+0,039$ (nứt rộng) và $+0,035$ (nứt mảnh).

Phân tích độ nhạy $\theta \in [1,5]$ pixels

Kết quả Bảng 4 cho thấy thứ tự hiệu năng giữa các nhóm được giữ nguyên ở mọi ngưỡng. Chỉ số IoU giữ nguyên do bản chất đánh giá theo vùng, trong khi BF1 tăng dần theo θ trên cả hai tập dữ liệu (mức tăng ở D_1 là $+0,375-0,421$, cao hơn so với D_2 là $+0,179-0,217$). Sự gia tăng này khiến khoảng cách BF1 giữa hai tập dữ liệu giảm liên tục

Bảng 3. Kết quả định lượng của IoU, BF1 và các thành phần biên trên hai tập dữ liệu.

Bộ dữ liệu	Loại vết nứt	IoU	BF1	$P_B(\theta)$	$R_B(\theta)$
D1 (Đa nguồn mặt đất)	Nứt rộng (> 4 pixel)	$0,706 \pm 0,124$	$0,273 \pm 0,127$	0,410	0,207
	Nứt mảnh (≤ 4 pixel)	$0,542 \pm 0,144$	$0,531 \pm 0,166$	0,767	0,419
D2 (Ảnh UAV)	Nứt rộng (> 4 pixel)	$0,706 \pm 0,138$	$0,875 \pm 0,085$	0,897	0,858
	Nứt mảnh (≤ 4 pixel)	$0,546 \pm 0,091$	$0,842 \pm 0,069$	0,868	0,833



Hình 4. Quan hệ giữa IoU, BF1 và độ rộng vết nứt trên tập kiểm tra D1 và D2. IoU có xu hướng tăng theo độ rộng, trong khi BF1 thay đổi tùy theo nguồn dữ liệu. Các điểm mờ biểu diễn từng ảnh, đường nối là giá trị trung bình theo khoảng độ rộng.

Bảng 4. Giá trị trung bình của IoU và BF1 trên các ảnh của từng nhóm, tại các ngưỡng dung sai khác nhau. Cột $\theta = 2$ pixel tương ứng với giá trị trung bình đã trình bày trong Bảng 3.

Bộ dữ liệu	Loại vết nứt	IoU	BF1 ($\theta=1$)	BF1 ($\theta=2$)	BF1 ($\theta=3$)	BF1 ($\theta=4$)	BF1 ($\theta=5$)
D1 (Đa nguồn mặt đất)	Nứt rộng (> 4 pixel)	0,706	0,177	0,273	0,374	0,461	0,552
	Nứt mảnh (≤ 4 pixel)	0,542	0,377	0,531	0,655	0,736	0,798
D2 (Ảnh UAV)	Nứt rộng (> 4 pixel)	0,706	0,729	0,875	0,922	0,936	0,946
	Nứt mảnh (≤ 4 pixel)	0,546	0,720	0,842	0,879	0,891	0,899

từ 0,343 xuống 0,101 đối với nhóm nứt mảnh, và đạt đỉnh tại $\theta = 2$ pixels (0,602) trước khi thu hẹp về 0,394 đối với nhóm nứt rộng.

3.3. Kết quả thực nghiệm sai lệch biên có kiểm soát

Kết quả phản ứng độc lập của hai chỉ số IoU và BF1 đối với ba phép biến đổi hình học giả lập trực tiếp từ mặt nạ nhãn chuẩn của tập D1 (không qua mô hình) được mô tả chi tiết trong Hình 5 và Bảng 5.

4. Thảo luận

4.1. Sự khác biệt bản chất giữa IoU và BF1

Sự khác biệt bản chất giữa IoU và BF1 xuất phát từ cơ chế phản ánh sai lệch hình học: IoU đánh giá dựa trên diện tích chồng lấp nên nhạy hơn với sai lệch khe hở hơn đối với vết nứt mảnh so với vết nứt rộng (ví dụ, lệch 1 pixel gây tổn thất 50% bề rộng vết nứt 2 pixels nhưng chỉ chiếm 10% ở vết nứt 10 pixels), tương thích với đặc tính được khẳng định bởi Cheng và nnk., (2021). Ngược lại, BF1 đo lường độ bám biên theo khoảng cách thực tế trong ngưỡng dung sai θ độc lập với diện tích, song tác động thực sự của θ vẫn chịu sự chi phối từ bề rộng vết nứt. Khi bề rộng nhỏ hơn 2θ , hai dải dung sai chồng lấp bao phủ toàn bộ vết nứt. Còn khi bề rộng lớn hơn 2θ , hai dải tách biệt buộc mô hình phải định vị chính xác cả hai ranh giới. Vì vậy, với $\theta = 2$ pixels, mốc hình học $2\theta = 4$ pixels đóng vai trò ngưỡng cốt lõi để phân nhóm định lượng vết nứt mảnh. Chính sự khác biệt này là tiền đề cho sự tương hỗ giữa hai chỉ số, qua đó bù đắp những hạn chế của nhau trong đánh giá hình học vết nứt.

4.2. Vai trò của chất lượng dữ liệu và phân tích lỗi hệ thống

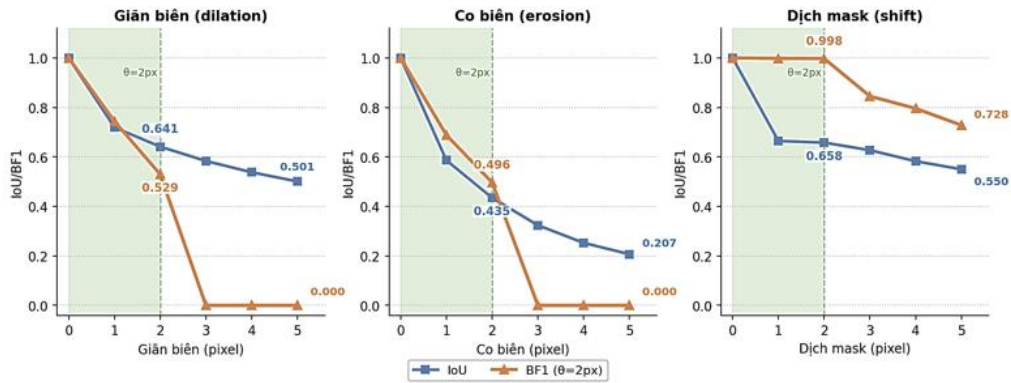
So sánh trên hai bộ dữ liệu khẳng định BF1 nhạy hơn IoU đối với chất lượng ranh giới và độ

đồng nhất của nhãn gán. Trong khi IoU duy trì mức ổn định giữa D_1 và D_2 , BF1 cho thấy khả năng phân biệt rõ ràng hơn giữa các mức chất lượng dữ liệu. Tốc độ tăng của BF1 theo ngưỡng dung sai θ phản ánh phân bố sai lệch biên thay vì chỉ đo mức chính xác trung bình. Cụ thể, BF1 trên D_2 đạt điểm cao ngay tại $\theta = 1$ pixel và sớm bão hòa, chứng tỏ sai số ranh giới hầu hết không vượt quá 1 pixel. Ngược lại, BF1 trên D_1 chỉ tăng mạnh khi θ đạt 4-5 pixels, phản ánh sự thiếu đồng nhất từ dữ liệu gán nhãn đa nguồn. Tuy nhiên, do nghiên cứu mới sử dụng hai bộ dữ liệu, nhận định này cần được kiểm chứng thêm trước khi khái quát hóa.

Phân tích các thành phần độ chính xác biên (P_B) và độ phủ biên (R_B) đã chỉ ra các dạng lỗi hệ thống trên bộ dữ liệu D_1 . Đối với vết nứt mảnh, P_B đạt mức cao nhưng R_B chỉ đạt 0,419, chứng tỏ ranh giới dự báo bị đứt quãng. Đối với vết nứt rộng, cả P_B và R_B đều sụt giảm, phản ánh sự kết hợp giữa sai lệch vị trí và thiếu bao phủ. Ngược lại, hai chỉ số này trên bộ dữ liệu D_2 duy trì trạng thái cân bằng ở cả hai nhóm hình thái.

4.3. Nguyên tắc lựa chọn ngưỡng dung sai

Thực nghiệm quét dung sai đã xác nhận $\theta = 2$ pixels là lựa chọn tối ưu theo cơ sở hình học (Mục 2.3). Việc vi phạm cận dưới ($\theta = 1$ pixel) khiến BF1 nhóm nứt rộng trên D_1 giảm mạnh xuống 0,177 do sai số gán nhãn thủ công cũng bị tính là sai. Trong khi đó, vi phạm cận trên ($\theta \geq 4$ pixels) làm suy giảm mạnh khả năng phân biệt dữ liệu khi khoảng cách BF1 ở nhóm nứt mảnh thu hẹp còn 0,155 ($\theta = 4$) và 0,101 ($\theta = 5$). Ngưỡng $\theta = 2$ pixels đạt sự cân bằng giữa hai giới hạn, tạo ra chênh lệch BF1 cực đại (0,602) giữa hai tập dữ liệu ở nhóm nứt rộng. Nguyên tắc tổng quát được thiết lập: ngưỡng θ bị chặn trên bởi bề rộng vết nứt và chặn dưới bởi độ bất định của thao tác gán nhãn, đòi hỏi phải hiệu chỉnh linh hoạt theo độ phân giải không gian mặt đất (GSD).



Hình 5. So sánh độ nhạy của IoU và BF1 dưới các sai lệch biến có kiểm soát trên tập D1. Ba dạng sai lệch gồm giãn biên, co biên và dịch mặt nạ được tạo từ nhân chuẩn. Vùng nền thể hiện dung sai $\theta = 2$ pixel; các giá trị được đánh dấu tại mức θ và tại điểm cuối của mỗi đường.

Bảng 5. Giá trị IoU và BF1 dưới các phép sai lệch hình học có kiểm soát.

Dạng sai lệch giả lập	Chỉ số	Tại mức lệch 2 pixel (θ)	Tại mức lệch 5 pixel (θ)	Xu hướng suy giảm tổng thể
Giãn biên (Dilation)	IoU	0,641	0,501	Giảm chậm, tuyến tính.
	BF1	0,529	0,000	Giảm nhanh, chạm đáy tại 5 pixel.
Co biên (Erosion)	IoU	0,435	0,207	Giảm nhanh hơn so với dạng giãn biên.
	BF1	0,496	0,000	Giảm nhanh về ranh giới 0 từ mức 3 pixel.
Dịch mặt nạ (Shift)	IoU	0,658	0,550	Suy giảm ngay từ các mức dịch nhỏ nhất.
	BF1	0,998	0,728	Gần như không đổi tại mức ≤ 2 pixel, giảm dần sau đó.

5. Kết luận

Nghiên cứu đã chỉ ra tính bổ sung hình học giữa IoU (đánh giá mức trùng lặp vùng) và BF1 (đánh giá độ chính xác ranh giới trong ngưỡng dung sai θ) cho bài toán phân đoạn vết nứt dạng tuyến mảnh. Thực nghiệm chỉ ra rằng IoU duy trì sự ổn định giữa các nguồn dữ liệu trên cùng nhóm độ rộng: 0,706 ở nhóm nứt rộng; 0,542-0,546 ở nhóm nứt mảnh. Trong khi BF1 cho mức độ nhạy cao hơn với độ bám biên và tính nhất quán của nhãn gán, thể hiện qua sự phân hóa rõ rệt giữa bộ dữ liệu đa nguồn D_1 (0,273 và 0,531) và bộ dữ liệu UAV D_2 (0,875 và 0,842). Thí nghiệm sai lệch có kiểm soát tiếp tục cung cấp bằng chứng trực tiếp: khi dịch chuyển mặt nạ từ 2 đến 5 pixels, BF1 giảm mạnh từ 0,998 xuống 0,728 và tiến về 0 khi bị giãn/co biên nghiêm trọng, trong khi IoU suy giảm tuyến tính chậm. Điều này chứng tỏ BF1 bộc lộ rõ các lỗi hình học như đứt đoạn hay lệch biên mà chỉ số vùng IoU bỏ qua.

Từ những phát hiện này, nghiên cứu đề xuất việc xây dựng một quy trình đánh giá chuẩn hóa cho phân đoạn đối tượng dạng tuyến mảnh gồm bốn phần: (i) báo cáo song song chỉ số vùng (IoU) và chỉ số biên (BF1); (ii) công bố rõ ngưỡng dung sai θ kèm cơ sở hình học xác lập; (iii) phân tầng kết quả theo độ rộng đối tượng và báo cáo các thành phần biên (P_B , R_B) để chẩn đoán dạng lỗi; (iv) kiểm tra độ nhạy của chỉ số bằng thí nghiệm sai lệch có kiểm soát trước khi diễn giải. Trong phạm vi bằng chứng hiện có, đây là khuyến nghị thực hành nhằm tránh diễn giải sai khi chỉ dùng chỉ số vùng, và cần được kiểm chứng thêm trên nhiều bộ dữ liệu và kiến trúc khác.

Hạn chế hiện tại của nghiên cứu nằm ở phạm vi dữ liệu, ngưỡng dung sai cố định và việc thiếu thông số GSD. Chính điều này khiến kích thước biểu diễn theo pixel chưa thể đổi sang milimét để đối chiếu với tiêu chuẩn bảo trì FHWA (Miller & Bellinger, 2014). Những hạn chế trên cũng mở ra hướng nghiên cứu tiếp theo. Việc khảo sát thêm các kiến trúc khác, trong đó có Transformer, sẽ

cho biết kết luận về tính tương hỗ có giữ nguyên hay không. Với dữ liệu có hiệu chuẩn GSD, ngưỡng dung sai có thể quy đổi sang đơn vị chiều dài và đối chiếu với tiêu chuẩn phân cấp hư hỏng. Các hàm mất mát nhận thức biên và bảo tồn cấu trúc là hướng có thể xem xét để cải thiện cả độ chính xác vùng lẫn biên.

Đóng góp của các tác giả

Vũ Ngọc Phượng - chịu trách nhiệm chính trong việc xây dựng ý tưởng nghiên cứu, thiết kế thí nghiệm, phân tích kết quả và soạn thảo bản thảo; Phạm Trung Dũng - tham gia xây dựng phương pháp, phân tích kết quả, góp ý về nội dung khoa học, chỉnh sửa bản thảo và hoàn thiện bài báo. Tất cả các tác giả đã đọc và đồng ý với phiên bản cuối cùng của bài báo.

Tuyên bố sử dụng công cụ hỗ trợ AI

Các tác giả đã sử dụng các công cụ AI như Microsoft Copilot, Gemini và Claude để hỗ trợ chỉnh sửa ngôn ngữ và cải thiện diễn đạt. Mọi nội dung khoa học và kết luận trong bài đều do các tác giả chịu trách nhiệm.

Tài liệu tham khảo

Al-Huda, Z., Peng, B., Algburi, R. N. A., Al-Antari, M., Al-Jarazi, R., & Zhai, D. (2023). A hybrid deep learning pavement crack semantic segmentation. *Eng. Appl. Artif. Intell.*, 122, 106142. <https://doi.org/10.1016/j.engappai.2023.106142>.

Bộ Khoa học và Công nghệ. (2024). *Bảo dưỡng thường xuyên đường bộ - Yêu cầu kỹ thuật* (TCVN 14182:2024). Bộ Khoa học và Công nghệ. <https://tieuchuan.vsqi.gov.vn/tieuchuan/view?sohieu=TCVN+14182%3A2024>.

Chen, Z., Shamsabadi, E. A., Jiang, S., Shen, L., & Dias-Da-Costa, D. (2024). Vision Mamba-based autonomous crack segmentation on concrete, asphalt, and masonry surfaces. *ArXiv, abs/2406.16518*. <https://doi.org/10.48550/arxiv.2406.16518>.

Cheng, B., Girshick, R., Dollár, P., Berg, A. C., & Kirillov, A. (2021). *Boundary IoU: Improving Object-Centric Image Segmentation Evaluation* (arXiv:2103.16562). arXiv. <https://doi.org/10.48550/arXiv.2103.16562>.

Csurka, G., Larlus, D., & Perronnin, F. (2013). What is a good evaluation measure for semantic segmentation? *Proceedings of the British Machine Vision Conference 2013*, 32.1-32.11. <https://doi.org/10.5244/C.27.32>.

Du, Z., Yuan, J., Xiao, F., & Hettiarachchi, C. (2021). Application of image technology on pavement distress detection: A review. *Measurement*, 184, 109900. <https://doi.org/10.1016/j.measurement.2021.109900>.

Dung, C. V., Anh, L. D. (2019). Autonomous concrete crack detection using deep fully convolutional neural network. *Automation in Construction*, 99, 52-58. <https://doi.org/10.1016/j.autcon.2018.11.028>.

Fan, Z., Li, C., Chen, Y., Mascio, P. D., Chen, X., Zhu, G., & Loprencipe, G. (2020). Ensemble of Deep Convolutional Neural Networks for Automatic Pavement Crack Detection and Measurement. *Coatings*, 10(2), 152. <https://doi.org/10.3390/coatings10020152>.

Fei, Y., Wang, K. C. P., Zhang, A., Chen, C., Li, J. Q., Liu, Y., Yang, G., & Li, B. (2020). Pixel-Level Cracking Detection on 3D Asphalt Pavement Images Through Deep-Learning- Based CrackNet-V. *IEEE Transactions on Intelligent Transportation Systems*, 21(1), 273-284. <https://doi.org/10.1109/TITS.2019.2891167>.

Goo, J. M., Milidonis, X., Artusi, A., Boehm, J., & Ciliberto, C. (2025). Hybrid-Segmentor: Hybrid approach for automated fine-grained crack segmentation in civil infrastructure. *Automation in Construction*, 170, 105960. <https://doi.org/10.1016/j.autcon.2024.105960>.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770-778. <https://doi.org/10.1109/CVPR.2016.90>.

Kang, D., & Cha, Y. (2022). Efficient attention-based deep encoder and decoder for automatic crack segmentation. *Structural Health Monitoring*, 21(5), 2190-2205. <https://doi.org/10.1177/14759217211053776>.

Kheradmandi, N., & Mehranfar, V. (2022). A critical review and comparative study on

- image segmentation-based techniques for pavement crack detection. *Construction and Building Materials*, 321, 126162. <https://doi.org/10.1016/j.conbuildmat.2021.126162>.
- Korcynski, N. (2026). *A Boundary-Metric Evaluation Protocol for Whiteboard Stroke Segmentation Under Extreme Imbalance* (arXiv:2603.00163). arXiv. <https://doi.org/10.48550/arXiv.2603.00163>.
- Martin, D. R., Fowlkes, C. C., & Malik, J. (2004). Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(5), 530-549. <https://doi.org/10.1109/TPAMI.2004.1273918>.
- Middha, L. (2020). *Crack Segmentation Dataset [Data set]*. <https://www.kaggle.com/datasets/lakshaymiddha/crack-segmentation-dataset>.
- Miller, J. S. & Bellinger, W. Y. (2014). *Distress Identification Manual for the Long-Term Pavement Performance Program*. (5th rev.ed.; Report No. FHWA-HRT-13-092.) Federal Highway Administration. <https://www.fhwa.dot.gov/publications/research/infrastructure/pavements/ltp/13092/13092.pdf>.
- Rahman, M. A., & Wang, Y. (2016). Optimizing Intersection-Over-Union in Deep Neural Networks for Image Segmentation. In G. Bebis, R. Boyle, B. Parvin, D. Koracin, F. Porikli, S. Skaff, A. Entezari, J. Min, D. Iwai, A. Sadagic, C. Scheidegger, & T. Isenberg (Eds.), *Advances in Visual Computing* (pp. 234-244). Springer International Publishing. https://doi.org/10.1007/978-3-319-50835-1_22.
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015* (pp. 234-241). Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4_28.
- Shit, S., Paetzold, J. C., Sekuboyina, A., Ezhov, I., Unger, A., Zhylka, A., Pluim, J. P. W., Bauer, U., & Menze, B. H. (2021). cDice-A Novel Topology-Preserving Loss Function for Tubular Structure Segmentation. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 16560-16569. <https://doi.org/10.1109/CVPR46437.2021.01629>.
- Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., & Cardoso, M. J. (2017). Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* (Vol. 10553, pp. 240-248). https://doi.org/10.1007/978-3-319-67558-9_28 (Original work published In: DLMIA Workshop, MICCAI.).
- Wang, Z., & Pang, J. (2025). *Fast Measuring Pavement Crack Width by Cascading Principal Component Analysis* (arXiv:2511.02144). arXiv. <https://doi.org/10.48550/arXiv.2511.02144>.
- Yu, G., Dong, J., Wang, Y., & Zhou, X. (2023). RUC-Net: A Residual-Unet-Based Convolutional Neural Network for Pixel-Level Pavement Crack Segmentation. *Sensors*, 23(1), 53. <https://doi.org/10.3390/s23010053>.
- Ziya. (2025). *UAV-Based Crack Detection Dataset [Data set]*. <https://www.kaggle.com/datasets/ziya07/uav-based-crack-detection-dataset>.