

APMFUSION: MÔ HÌNH TỔNG HỢP HÌNH ẢNH HỒNG NGOẠI VÀ KHẢ KIẾN DỰA TRÊN AGF VÀ PFCFUSE

Lê Thị Hồng Hà¹, Trịnh Văn Hậu², Nguyễn Văn Cường¹, Lê Đức Thọ¹,
Lê Đình Nghiệp³, Nguyễn Thế Cường¹

TÓM TẮT

Tổng hợp hình ảnh hồng ngoại và khả kiến đóng vai trò quan trọng trong việc nâng cao khả năng quan sát mục tiêu và nhận thức bối cảnh trong các điều kiện quan sát bất lợi. Tuy nhiên, nhiều phương pháp hiện nay vẫn chưa đạt được sự cân bằng hiệu quả giữa bảo toàn cường độ sáng, độ tương phản và khả năng duy trì thông tin cấu trúc. Để khắc phục hạn chế này, chúng tôi đề xuất APMFusion, một mô hình tổng hợp kết hợp bộ lọc AGF, mô hình học sâu PFCFuse và toán tử MRE. Kết quả thực nghiệm trên 80 cặp ảnh từ bộ dữ liệu công khai MSRS cho thấy APMFusion đạt hiệu năng tốt nhất hoặc tiệm cận tốt nhất trên sáu chỉ số đánh giá quan trọng gồm Q_{MLI} , Q_{CI} , Q_E , Q_{MI} , Q_{SSIM} và $Q^{AB/F}$, khi so sánh với năm phương pháp tiên tiến được công bố trong giai đoạn 2024-2025. Code nguồn, các kết quả thực nghiệm được công bố tại https://github.com/house3173/AGF_IVIF.git.

Từ khóa: AGF, PFCFuse, MRE, tổng hợp ảnh hồng ngoại và khả kiến.

DOI: <https://doi.org/10.70117/hdujs.84.2.2026.1149>

1. ĐẶT VẤN ĐỀ

Tổng hợp ảnh hồng ngoại (Infrared Image - IR) và ảnh khả kiến (Visible Image - VIS) là một bài toán quan trọng trong lĩnh vực xử lý ảnh, đặc biệt trong các ứng dụng như giám sát an ninh, phát hiện mục tiêu trong điều kiện thiếu. Ảnh VI cung cấp thông tin phong phú về màu sắc và kết cấu, trong khi ảnh IR phản ánh đặc trưng nhiệt của đối tượng. Việc tổng hợp hai mô thức ảnh này nhằm khai thác các ưu điểm bổ sung của từng loại, từ đó tạo ra ảnh hợp nhất giàu thông tin hơn về cả cấu trúc không gian và đặc trưng nhiệt.

Các phương pháp tổng hợp ảnh IR và VIS truyền thống thường được bao gồm ba bước: biến đổi ảnh, tổng hợp các thành phần và biến đổi ngược [1]. Trong giai đoạn biến đổi ảnh, các kỹ thuật dựa trên bộ lọc [2][3] được áp dụng rộng rãi nhằm tách ảnh nguồn thành các thành phần cơ sở và chi tiết. Ở giai đoạn tổng hợp, các thành phần cơ sở thường được kết hợp bằng các quy tắc đơn giản như trung bình cộng [4][5], trong khi các thành phần chi tiết thường được xử lý thông qua các chiến lược dựa trên năng lượng [3]. Cuối cùng, giai đoạn biến đổi ngược được thực hiện tương ứng với kỹ thuật biến đổi đã sử dụng ở bước đầu.

¹ Khoa Kỹ thuật, Công nghệ và Truyền thông, Trường Đại học Hồng Đức; Email: nguyenthecuong@hdu.edu.vn

² Công ty VNPT-Media Hà Nội

³ Trung tâm Công nghệ thông tin, Truyền thông và Thư viện, Trường Đại học Hồng Đức

Mặc dù có ưu điểm về hiệu quả tính toán, các phương pháp truyền thống vẫn tồn tại một số hạn chế. Cụ thể, khi tổng hợp các thành phần cơ sở, việc áp dụng quy tắc trung bình [3][4], thường dẫn đến hiện tượng suy giảm cường độ sáng và độ tương phản của ảnh tổng hợp. Đối với các thành phần chi tiết, các phương pháp dựa trên năng lượng vẫn chưa bảo toàn hiệu quả thông tin đường biên và cấu trúc. Vì vậy, việc nghiên cứu các phương pháp tổng hợp ảnh IR-VIS hiệu quả nhằm nâng cao chất lượng thị giác và khả năng bảo toàn thông tin cấu trúc là rất cần thiết. Gần đây, nhờ khả năng học và trích xuất đặc trưng mạnh mẽ, các mô hình học sâu đã được kết hợp với các kỹ thuật tổng hợp truyền thống để tạo thành phương pháp lai, tận dụng ưu điểm của cả hai hướng tiếp cận. Bằng cách sử dụng bộ lọc trong quá trình phân rã ảnh và mô hình học sâu để trích xuất, tổng hợp đặc trưng, các phương pháp này cho phép xây dựng chiến lược giúp cải thiện đáng kể chất lượng ảnh.

Trong bài báo này, chúng tôi đề xuất APMFusion, một mô hình tổng hợp ảnh khả kiến - hồng ngoại (IR-VIS) kết hợp Alternating Guided Filtering (AGF) [6], mô hình học sâu PoolFormer và CNN Fusion (PFCFuse) [7], và toán tử năng lượng vùng cực đại (Maximum Region Energy - MRE) với nhân nhị phân. Mô hình được thiết kế nhằm tăng cường cường độ sáng, độ tương phản của ảnh tổng hợp, đồng thời bảo toàn hiệu quả thông tin đường biên và các đặc trưng cấu trúc quan trọng. Đóng góp chính của nghiên cứu bao gồm: Đề xuất phương pháp tổng hợp ảnh IR - VIS mới, APMFusion, giúp cải thiện cường độ sáng, độ tương phản và bảo toàn đường biên và các đặc trưng cấu trúc quan trọng so với các phương pháp hiện có.

Cấu trúc của bài báo được tổ chức như sau: Phần 2 trình bày các kiến thức nền tảng liên quan đến AGF, PFCFuse và MRE; Phần 3 mô tả phương pháp tổng hợp ảnh được đề xuất và phân tích kết quả thực nghiệm; Phần 4 đưa ra kết luận và định hướng nghiên cứu.

2. KIẾN THỨC NỀN TẢNG

2.1. Bộ lọc AGF (Alternating Guided Filtering)

AGF [6] là một kỹ thuật làm trơn bảo toàn biên, được xây dựng dựa trên việc áp dụng bộ lọc kết hợp (joint filter). AGF bao gồm ba bước, được mô tả như sau:

Bước 1: áp dụng JBF (joint bilateral filter) lên ảnh đầu vào I , điểm ảnh (i, j) , N là hệ số chuẩn hóa, $f(\cdot)$ và $g(\cdot)$ là nhân lọc không gian và nhân lọc theo miền giá trị; Ω biểu thị miền hỗ trợ không gian của nhân lọc. Kết quả lọc tại vòng lặp thứ $k + 1$, ký hiệu F^{k+1} , được tính bằng cách sử dụng kết quả của vòng lặp trước F^k làm ảnh hướng dẫn, theo (2.1).

$$F_i^{k+1} = \frac{1}{N} \sum_{j \in \Omega} I_j \cdot f(i - j) \cdot g(|F_i^k - F_j^k|). \quad (2.1)$$

Bước 2: JBF tiếp tục được áp dụng lên kết quả của bước thứ nhất, trong đó ảnh nguồn I được sử dụng làm ảnh hướng dẫn. Quá trình này được mô tả như công thức (2.2).

$$F_i^{k+1} = \frac{1}{N} \sum_{j \in \Omega} F_j^k \cdot f(i - j) \cdot g(|I_i - I_j|). \quad (2.2)$$

Bước 3: bộ lọc trung vị được áp dụng lên kết quả của bước thứ hai nhằm loại bỏ các nhiễu dư, đặc biệt tại các vùng lân cận của đường biên lớn.

AGF phân rã ảnh đầu vào thành thành phần cơ sở thông qua lọc trực tiếp, trong khi thành phần chi tiết được xác định bằng hiệu giữa ảnh đầu vào và thành phần cơ sở.

2.2. Mô hình PFCFuse (Poolformer and CNN Fusion)

PFCFuse [7] được phát triển dựa trên ý tưởng của mô hình CDDFuse, sử dụng kiến trúc kết hợp PoolFormer-CNN cùng chiến lược huấn luyện hai giai đoạn. Ở giai đoạn thứ nhất, các cặp ảnh IR và VIS được đưa vào encoder để trích xuất đặc trưng chung ban đầu. Sau đó, các đặc trưng tần số thấp và tần số cao lần lượt được trích xuất thông qua Base PoolFormer Encoder (BPE) và Detail CNN Encoder (DCE), rồi được kết hợp và đưa vào decoder để tái tạo ảnh gốc. Ở giai đoạn thứ hai, encoder đã được huấn luyện được sử dụng để trích xuất đặc trưng cơ sở và đặc trưng chi tiết, từ đó tạo ra ảnh tổng hợp cuối cùng thông qua decoder. Trong cả hai giai đoạn, mô hình sử dụng encoder song nhánh, cùng với lớp tổng hợp đặc trưng cơ sở (Basic Feature Layer - BFL) và lớp tổng hợp đặc trưng chi tiết (Detail Fusion Layer - DFL).

Encoder của PFCFuse bao gồm shared encoder (SE) và independent encoder (IE). SE được xây dựng từ các khối Restormer, trong khi IE gồm BPE và DCE sử dụng các khối mạng nơ-ron khả nghịch (invertible neural network - INN).

Với cặp ảnh IR và VIS, các đặc trưng nông được trích xuất theo công thức (2.3).

$$\{\phi_{ir}^S, \phi_{vis}^S\} = \{SE(IR), SE(VIS)\}. \quad (2.3)$$

BPE trích xuất các đặc trưng nền tảng, trong khi DCE thu nhận các đặc trưng chi tiết. BPE bao gồm các phép Layer Normalization, pooling, PoolMLP và DropPath, được mô tả như các công thức (2.4) và (2.5).

$$\phi_{ir}^{S'} = LN(Pooling(LN(\phi_{ir}^S))), \quad (2.4)$$

$$\phi_{ir}^B = DropPath(PoolMLP(\phi_{ir}^{S'})). \quad (2.5)$$

trong đó, $PoolMLP(\cdot)$ là perceptron đa tầng sử dụng các lớp tích chập 1×1 .

Để trích xuất hiệu quả các đặc trưng chi tiết, PFCFuse tích hợp INN vào DCE. Quá trình này được mô tả như các công thức (2.6) và (2.7).

$$\phi_{ir}^{S,k+1} = INN(\phi_{ir}^{S,k}), \quad (2.6)$$

$$\phi_{ir}^D = \phi_{ir}^{S,N}. \quad (2.7)$$

Ở giai đoạn thứ hai, các đặc trưng cơ sở và chi tiết của ảnh hồng ngoại và ảnh khả kiến được tổng hợp như các công thức (2.8) và (2.9), trong đó, \oplus là phép cộng đặc trưng.

$$\phi_f^B = BFL(\phi_{ir}^B \oplus \phi_{vis}^B), \quad (2.8)$$

$$\phi_f^D = DFL(\phi_{ir}^D \oplus \phi_{vis}^D). \quad (2.9)$$

Đầu ra của decoder là ảnh gốc ở giai đoạn thứ nhất và ảnh tổng hợp ở giai đoạn thứ hai như công thức (2.10) và (2.11), trong đó, $CAT(\cdot)$ biểu thị phép nối theo chiều kênh và $D(\cdot)$ là decoder.

$$I' = D(CAT(\phi_I^B, \phi_I^D)), V' = D(CAT(\phi_V^B, \phi_V^D)), \quad (2.10)$$

$$F = D(CAT(\phi_I^B, \phi_I^D)). \quad (2.11)$$

2.3. Năng lượng vùng cực đại (Maximum Region Energy - MRE)

Năng lượng vùng cực đại là một chiến lược được sử dụng phổ biến trong bài toán tổng hợp ảnh nhằm kết hợp các thành phần chi tiết từ nhiều ảnh nguồn. Năng lượng vùng (Regional Energy - RE) tại điểm ảnh (i, j) được xác định như công thức (2.12).

$$RE(i, j) = \sum_{m=-x}^x \sum_{n=-x}^x \omega(m+x, n+x) \cdot (I(i+m, j+n))^2 \quad (2.12)$$

trong đó, ω là nhân trọng số áp dụng cho vùng lân cận cục bộ có kích thước $(2x+1) \times (2x+1)$, lấy tâm tại điểm ảnh (i, j) và $I(i+m, j+n)$ là giá trị cường độ tại vị trí tương ứng trong cửa sổ này.

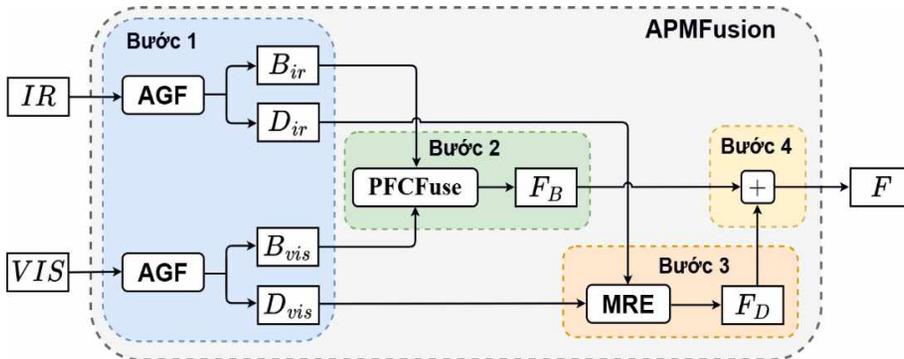
Dựa trên giá trị năng lượng vùng, chiến lược MRE thực hiện tổng hợp các thành phần ảnh bằng cách lựa chọn giá trị điểm ảnh tại vị trí tương ứng từ thành phần có năng lượng vùng lớn hơn. Nguyên lý tổng hợp này có thể được mô tả như sau:

$$I_F(i, j) = \begin{cases} I_A(i, j), & \text{nếu } RE_A(i, j) \geq RE_B(i, j), \\ I_B(i, j), & \text{nếu } RE_A(i, j) < RE_B(i, j), \end{cases} \quad (2.13)$$

trong đó I_A và I_B là hai thành phần ảnh cần tổng hợp, còn $RE_A(i, j)$ và $RE_B(i, j)$ lần lượt là năng lượng vùng tại điểm ảnh (i, j) của hai thành phần tương ứng.

3. KẾT QUẢ NGHIÊN CỨU VÀ THẢO LUẬN

3.1. Mô hình đề xuất



Hình 1. Mô hình tổng hợp hình ảnh APMFusion

Mô hình tổng hợp ảnh được đề xuất, APMFusion, được minh họa trong Hình 1 và được mô tả chi tiết như sau:

Bước 1 (Phân rã ảnh bằng AGF): Hai ảnh đầu vào hồng ngoại IR và khả kiến VIS được phân rã thành các thành phần cơ sở B_{ir}, B_{vis} và các thành phần chi tiết D_{ir}, D_{vis} bằng bộ lọc AGF theo các công thức (3.1) và (3.2).

$$[B_{ir}, D_{ir}] = \mathbf{AGF}(IR). \quad (3.1)$$

$$[B_{vis}, D_{vis}] = \mathbf{AGF}(VIS). \quad (3.2)$$

Bước 2 (Tổng hợp thành phần cơ sở bằng PFCFuse): Các thành phần cơ sở B_{ir} và B_{vis} được tổng hợp bằng mô hình học sâu PFCFuse, theo công thức (3.3).

$$F_B = \mathbf{PFCFuse}(B_{ir}, B_{vis}). \quad (3.3)$$

Bước 3 (Tổng hợp thành phần chi tiết bằng MRE với nhân nhị thức): Các thành phần chi tiết D_{ir} và D_{vis} được tổng hợp bằng chiến lược MRE kết hợp với nhân nhị thức, theo công thức (3.4).

$$F_D = \mathbf{MRE}(D_{ir}, D_{vis}), \quad (3.4)$$

Bước 4 (Tái tạo ảnh tổng hợp): ảnh tổng hợp F được tái tạo bằng cách cộng các thành phần đã tổng hợp ở bước 2 và bước 3 theo công thức (3.5).

$$F = F_B + F_D \quad (3.5)$$

3.2. Thực nghiệm

3.2.1. Dữ liệu thực nghiệm

Để đánh giá hiệu quả của phương pháp đề xuất, chúng tôi tiến hành thực nghiệm trên bộ dữ liệu ảnh hồng ngoại - khả kiến MSRS (Multi-Spectral Road Scenarios) [8]. Bộ dữ liệu này bao gồm các cặp ảnh VIS và IR được thu nhận trong các kịch bản giao thông dưới điều kiện ánh sáng yếu hoặc ban đêm và đã được đồng đăng ký ở mức điểm ảnh. 80 cặp ảnh IR - VIS từ bộ dữ liệu MSRS, ký hiệu là tập dữ liệu D_1 , được sử dụng để thực nghiệm.

3.2.2. Các giải thuật so sánh

Để đánh giá hiệu quả của phương pháp đề xuất, năm thuật toán tổng hợp ảnh đã được chọn để so sánh, bao gồm: BTSFusion (Balancing Texture and Saliency, 2024) [9], CrossFuse (2024) [10], PFCFuse (2024) [7], MaeFuse (Masked Autoencoder Fuse, 2025) [11] và MPCFusion (2024) [12]. Các thuật toán này lần lượt được ký hiệu từ T_1 đến T_5 . Phương pháp đề xuất được ký hiệu là T_6 .

3.2.3. Chỉ số đánh giá

Sáu chỉ số được sử dụng để đánh giá hiệu quả của các giải thuật bao gồm: chỉ số cường độ sáng Q_{MLI} , chỉ số độ tương phản Q_{CI} , chỉ số Entropy Q_E , chỉ số thông tin tương hỗ Q_{MI} [13], chỉ số bảo toàn cấu trúc Q_{SSIM} [14], chỉ số bảo toàn cạnh $Q^{AB/F}$ [15].

3.2.4. Môi trường thực nghiệm

Tất cả các thực nghiệm được thực hiện trên hệ điều hành Windows 11, với môi trường phần mềm được quản lý bằng Anaconda (v24.5.0) nhằm đảm bảo tính nhất quán của các thư viện phụ thuộc. Python (v3.10.14) được sử dụng làm ngôn ngữ lập trình chính, kết hợp với PyTorch (v2.3.1) và NumPy (v1.24.3) cho quá trình phát triển mô hình và tính toán số. Về phần cứng, hệ thống sử dụng bộ xử lý Intel Core i5-1335U (1.3 GHz). Các thí nghiệm liên quan đến mô hình học sâu được thực hiện bổ sung trên Google Colab với hỗ trợ GPU, qua đó đảm bảo tính tái lập và hiệu quả tính toán của toàn bộ quá trình thực nghiệm.

3.2.5. Mô tả thực nghiệm

Thực nghiệm 1: nhằm đánh giá hiệu quả của mô hình học sâu PFCFuse trong tổng hợp thành phần cơ sở của APMFusion (A_6). Cụ thể, PFCFuse lần lượt được thay thế bằng ba chiến lược tổng hợp, bao gồm: quy tắc trung bình [4], phương pháp dựa trên năng lượng vùng [3] và phương pháp dựa trên bản đồ độ nổi thị giác [16]. Các biến thể tương ứng của APMFusion được ký hiệu từ L_1 đến L_3 . Hiệu năng được đánh giá trên bộ dữ liệu D_1 thông qua ba chỉ số: Q_{MLI} , Q_{CI} và Q_E .

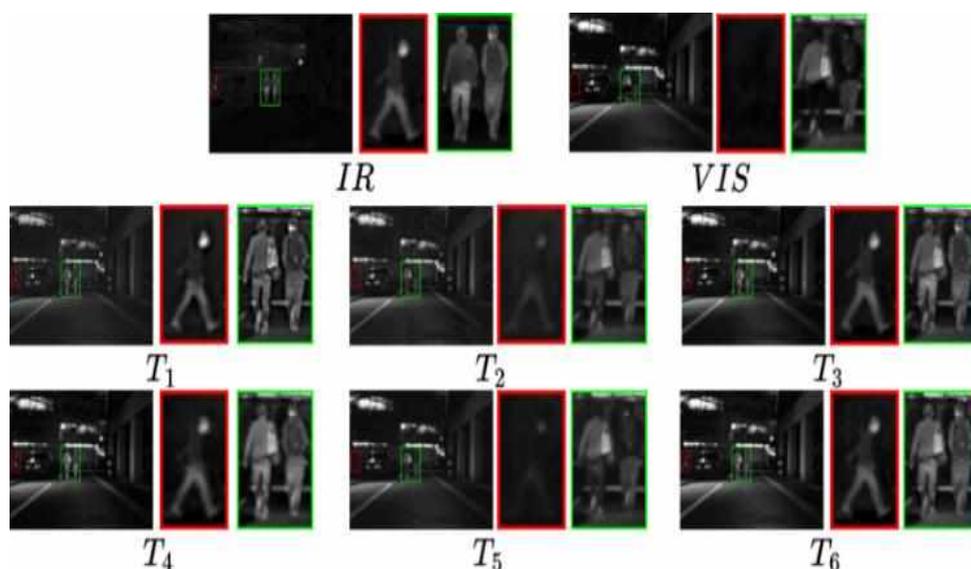
Thực nghiệm 2: nhằm đánh giá hiệu năng của thuật toán MRE trong tổng hợp các thành phần chi tiết của APMFusion. Ba chiến lược tổng hợp thay thế, bao gồm: quy tắc lựa chọn giá trị lớn nhất, mô hình PAPCNN (Parameterized Adaptive Pulse Coupled Neural

Network) [17], phương pháp dựa trên năng lượng cục bộ có trọng số [18]. Các biến thể APMFusion tương ứng được ký hiệu từ H_1 đến H_3 . Đánh giá định lượng được thực hiện thông qua ba chỉ số: Q_{MI} , $Q^{AB/F}$ và Q_{SSIM} .

Thực nghiệm 3: So sánh APMFusion với năm thuật toán tổng hợp ảnh gần đây (T_1 - T_5), được trình bày trong Mục 3.2.2. Thực nghiệm được thực hiện trên tập dữ liệu D_1 . Nhằm đảm bảo đánh giá có độ tin cậy cao, sáu chỉ số đánh giá được sử dụng, bao gồm: Q_{MLI} , Q_{CI} , Q_E , Q_{MI} , $Q^{AB/F}$ và Q_{SSIM} .

3.3. Đánh giá kết quả thực nghiệm

3.3.1. Đánh giá trực quan



Hình 2. So sánh trực quan và các lát cắt của ảnh tổng hợp thu được từ sáu mô hình.

Dễ dàng nhận thấy trong Hình 2, ảnh hồng ngoại *IR* cho phép quan sát rõ một người đi bộ trong vùng khoanh đỏ và hai người đi bộ trong vùng khoanh xanh, trong khi trên ảnh khả kiến *VIS* chỉ có thể nhận diện hai người đi bộ trong vùng khoanh xanh. Điều này cho thấy ảnh hồng ngoại thể hiện rõ các đối tượng mục tiêu nhưng thiếu thông tin về bối cảnh xung quanh, trong khi ảnh khả kiến bảo toàn tốt các kết cấu nền phong phú nhưng làm suy giảm độ nổi bật của mục tiêu. Các ảnh tổng hợp đã khắc phục phần lớn hạn chế của từng mô thức riêng lẻ. Tuy nhiên, kết quả của các phương pháp T_2 và T_5 vẫn bộc lộ những hạn chế đáng kể, khi người đi bộ trong vùng khoanh đỏ khó được nhận diện rõ ràng so với các phương pháp còn lại. Trong khi đó, các phương pháp T_1 , T_3 và T_4 cho ảnh tổng hợp có cường độ sáng và độ tương phản thấp hơn so với phương pháp đề xuất T_6 .

3.3.2. Đánh giá khách quan

Bảng 1 và Bảng 2 lần lượt trình bày kết quả đánh giá khách quan đối với các giải thuật tổng hợp thành phần cơ sở và thành phần chi tiết. Đối với giải thuật tổng hợp thành phần cơ sở (Bảng 1), T_6 đạt giá trị cao nhất trên cả ba chỉ số $Q_{MLI} = 0,2915$, $Q_{CI} = 0,1904$

và $Q_E = 6,8371$. So với phương pháp đối sánh tốt nhất trong nhóm còn lại (L_3), T_6 cải thiện lần lượt 47,8%, 23,6% và 9,2%; so với phương pháp có hiệu năng thấp nhất (L_2), mức cải thiện đạt tới 81,4%, 87,0% và 12,7%. Những kết quả này cho thấy việc sử dụng PFCFuse giúp bảo toàn hiệu quả cường độ sáng, tăng cường độ tương phản của ảnh tổng hợp.

Bảng 1. So sánh các phương pháp tổng hợp thành phần cơ sở trên D_1 .

STT	Giải thuật	Q_{MLI}	Q_{CI}	Q_E
1	L_1	0,1748	0,1018	6,0678
2	L_2	0,1607	0,1503	6,1487
3	L_3	0,1972	0,1541	6,2630
4	T_6	0,2915	0,1904	6,8371

Bảng 2. So sánh các phương pháp tổng hợp thành phần chi tiết trên D_1 .

STT	Giải thuật	Q_{MI}	Q_{SSIM}	$Q^{AB/F}$
1	H_1	4,1382	0,4746	0,6027
2	H_2	4,1806	0,4933	0,6579
3	H_3	4,0336	0,4635	0,5996
4	T_6	4,3298	0,4807	0,6769

Đối với giải thuật tổng hợp thành phần chi tiết (Thực nghiệm 2), kết quả trong Bảng 2 cho thấy T_6 đạt giá trị cao nhất đối với $Q_{MI} = 4,3298$ và $Q^{AB/F} = 6,6769$. So với phương pháp có kết quả tốt nhất, Q_{MI} được cải thiện khoảng 3,6%, trong khi $Q^{AB/F}$ tăng khoảng 2,9%. Chỉ số Q_{SSIM} đạt 0,4807, cao hơn H_1 và H_3 , đồng thời xấp xỉ với H_2 (0,4933), cho thấy phương pháp MRE duy trì tốt tính nhất quán cấu trúc tổng thể của ảnh tổng hợp.

Bảng 3. So sánh APMFusion với các phương pháp tổng hợp ảnh trên D_1 .

STT	Giải thuật	Q_{MLI}	Q_{CI}	Q_E	Q_{MI}	Q_{SSIM}	$Q^{AB/F}$	Thời gian (giây)
1	T_1	0,2429	0,1449	6,3889	2,3610	0,3900	0,4968	0,1184
2	T_2	0,2683	0,1592	6,6194	3,3290	0,4556	0,5969	0,7298
3	T_3	0,2748	0,1873	6,7884	5,4795	0,5034	0,7181	0,7444
4	T_4	0,2065	0,1637	6,6656	2,2489	0,4325	0,5112	0,8054
5	T_5	0,2532	0,1708	6,3099	3,8052	0,4088	0,5784	0,1700
6	T_6	0,2915	0,1904	6,8371	4,3298	0,4807	0,6769	0,8580

Bảng 3 trình bày kết quả so sánh giữa phương pháp APMFusion (T_6) và năm phương pháp đối sánh $T_1 - T_5$. Kết quả cho thấy T_6 đạt giá trị tốt nhất hoặc tiệm cận tốt nhất trên sáu chỉ số. Cụ thể, APMFusion (T_6) cao nhất đối với các chỉ số Q_{MLI} , Q_{CI} và Q_E , lần lượt cao hơn phương pháp đối sánh tốt nhất T_3 khoảng 6,1%, 1,7% và 0,7%. Mặc dù phương pháp T_3 đạt giá trị cao nhất đối với Q_{MI} , $Q^{AB/F}$ và Q_{SSIM} , phương pháp T_6 vẫn thể hiện hiệu năng ổn định và cân bằng hơn trên toàn bộ sáu chỉ số đánh giá, đặc biệt nổi trội về bảo toàn độ sáng và độ tương phản. Nhìn chung, các kết quả trên cho thấy T_6 không chỉ cải thiện từng thành phần riêng lẻ mà còn đạt hiệu năng tổng thể cân bằng và ổn định, vượt trội so với các phương pháp so sánh trên nhiều nhóm chỉ số đánh giá.

Tuy nhiên, Bảng 3 cũng cho thấy hạn chế của APMFusion (T_6) đó là thời gian thực thi cao so với các phương pháp đối sánh. Cụ thể, thời gian tổng hợp trung bình cho một cặp

ảnh của T_6 là 0,8580 giây, xếp thứ 6/6 trong các thuật toán được so sánh, cao hơn khoảng 6,5% so với thuật toán xếp thứ 5/6 (T_4 : 0,8054 giây). Nguyên nhân chủ yếu dẫn đến thời gian thực thi cao là do mô hình đề xuất tích hợp nhiều mô-đun xử lý, bao gồm AGF, PFCFuse và MRE. Thời gian thực thi là một hạn chế của mô hình đề xuất, gây khó khăn cho việc triển khai trong các ứng dụng thực tế, đặc biệt là các hệ thống yêu cầu xử lý thời gian thực.

3.3.3. Kiểm định Friedman

Để đánh giá độ tin cậy thống kê của các kết quả thực nghiệm, nghiên cứu sử dụng kiểm định Friedman. Kiểm định Friedman là một kiểm định phi tham số phù hợp cho việc so sánh nhiều phương pháp trên cùng một tập dữ liệu ghép cặp và không yêu cầu giả định dữ liệu tuân theo phân phối chuẩn. Kết quả trong Bảng 4 cho thấy tất cả các giá trị p-value đều rất nhỏ và thấp hơn đáng kể so với mức ý nghĩa 0,05, chứng tỏ tồn tại sự khác biệt có ý nghĩa thống kê giữa các phương pháp tổng hợp ảnh trên tất cả các chỉ số đánh giá. Do đó, giả thuyết không (các phương pháp có hiệu năng tương đương) có thể bị bác bỏ với độ tin cậy cao.

Bảng 4. Kết quả p-value của kiểm định Friedman

Chỉ số đánh giá	Giá trị Friedman p-value
Q_{MLI}	$1,3841 \times 10^{-50}$
Q_{CI}	$1,7262 \times 10^{-66}$
Q_E	$1,7961 \times 10^{-62}$
Q_{MI}	$5,1578 \times 10^{-79}$
Q_{SSIM}	$1,8434 \times 10^{-59}$
$Q^{AB/F}$	$1,9372 \times 10^{-66}$

4. KẾT LUẬN

Trong bài báo này, chúng tôi đã đề xuất APMFusion, một mô hình tổng hợp ảnh hồng ngoại IR và khả kiến VIS dựa trên sự kết hợp giữa bộ lọc AGF, mô hình học sâu PFCFuse và toán tử MRE với nhân nhị thức. Các kết quả thực nghiệm trên 80 cặp ảnh của bộ dữ liệu MSRS cho thấy APMFusion đạt hiệu năng tốt nhất trên ba chỉ số đánh giá Q_{MLI} , Q_{CI} , và Q_E , và tiệm cận tốt nhất trên ba chỉ số đánh giá Q_{MI} , Q_{SSIM} và $Q^{AB/F}$ khi so sánh với các phương pháp tổng hợp ảnh tiên tiến được công bố trong giai đoạn 2024 - 2025. Điều này khẳng định hiệu quả và tính ổn định của mô hình đề xuất trong bài toán tổng hợp ảnh IR và VIS. Mặc dù đạt được những kết quả khả quan, APMFusion vẫn tồn tại hạn chế về chi phí tính toán. Cụ thể, chi phí tính toán của mô hình PFCFuse có thể ảnh hưởng đến khả năng triển khai trong các hệ thống yêu cầu xử lý thời gian thực hoặc có tài nguyên tính toán hạn chế. Trong các nghiên cứu tiếp theo, chúng tôi sẽ tập trung tối ưu hóa độ phức tạp tính toán của mô hình và mở rộng đánh giá trên nhiều bộ dữ liệu ảnh đa phổ khác nhằm nâng cao khả năng ứng dụng thực tiễn. Cụ thể, chúng tôi dự kiến cải tiến mô hình PFCFuse nhằm nâng cao chất lượng ảnh tổng hợp đồng thời giảm thời gian thực thi của thuật toán.

TÀI LIỆU THAM KHẢO

- [1] Yang, K., Xiang, W., Chen, Z., Zhang, J., Liu, Y. (2024), *A review on infrared and visible image fusion algorithms based on neural networks*, Journal of Visual Communication and Image Representation, 101, 104179.

- [2] Zhu, H., Zhang, W. (2025), *Infrared and Visible Image Fusion Based on Image Enhancement and Target Extraction*, IEEE Access, 13, 61862-61875.
- [3] Luo, Y., He, K., Xu, D., Shi, H., Yin, W. (2025), *Infrared and visible image fusion based on hybrid multi-scale decomposition and adaptive contrast enhancement*, Signal Processing: Image Communication, 130, 117228.
- [4] Zhu, F., Liu, W. (2024), *Infrared-visible image fusion method based on multi-scale shearing Co-occurrence filter*, Infrared Physics and Technology, 136, 105009.
- [5] Li, Y., Liu, G., Bavirisetti, D. P., Gu, X., Zhou, X. (2023), *Infrared-visible image fusion method based on sparse and prior joint saliency detection and LatLRR-FPDE*, Digital Signal Processing, 134, 103910.
- [6] Toet, A. (2016), *Alternating guided image filtering*, PeerJ Computer Science, 2, e72. Portico.
- [7] Hu, X., Liu, Y., Yang, F. (2024), *PFCFuse: A Poolformer and CNN Fusion Network for Infrared-Visible Image Fusion*, IEEE Transactions on Instrumentation and Measurement, 73, 1-14.
- [8] Ha, Q., Watanabe, K., Karasawa, T., Ushiku, Y., Harada, T. (2017), *MFNet: Towards real-time semantic segmentation for autonomous vehicles with multi-spectral scenes*, 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 5108-5115.
- [9] Qian, Y., Liu, G., Tang, H., Xing, M., Chang, R. (2024), *BTSFusion: Fusion of infrared and visible image via a mechanism of balancing texture and salience*, Optics and Lasers in Engineering, 173, 107925.
- [10] Li, H., Wu, X.-J. (2024), *CrossFuse: A novel cross attention mechanism based infrared and visible image fusion approach*, Information Fusion, 103, 102147.
- [11] Li, J., Jiang, J., Liang, P., Ma, J., Nie, L. (2025), *MaeFuse: Transferring Omni Features With Pretrained Masked Autoencoders for Infrared and Visible Image Fusion via Guided Training*, IEEE Transactions on Image Processing, 34, 1340-1353.
- [12] Cheng, C., Xu, T., Wu, X.-J., Li, H., Li, X., Tang, Z., Kittler, J. (2025), *TextFusion: Unveiling the power of textual semantics for controllable image fusion*, Information Fusion, 117, 102790.
- [13] Chen, Y., Xue, Z., Blum, R. S. (2008), *Theoretical analysis of an information-based quality measure for image fusion*, Information Fusion, 9(2), 161-175.
- [14] Zhou Wang, Bovik, A. C., Sheikh, H. R., Simoncelli, E. P. (2004), *Image quality assessment: from error visibility to structural similarity*, IEEE Transactions on Image Processing, 13(4), 600-612.
- [15] Xydeas, C. S., Petrović, V. (2000), *Objective image fusion performance measure*, Electronics Letters, 36(4), 308-309.
- [16] Luo, Y., He, K., Xu, D., Yin, W., Liu, W. (2022), *Infrared and visible image fusion based on visibility enhancement and hybrid multiscale decomposition*, Optik, 258, 168914.
- [17] Zhang, C., Li, C., Yang, X. (2024), *Research on fractional wavelet transform combined with parameter adaptive PCNN for infrared and visible image fusion algorithm*, Optics Communications, 573, 131026.
- [18] Wei, L., Zhu, R., Li, X., Zhao, L., Hu, X., & Zhang, X. (2024). "Pixel-level structure awareness for enhancing multi-modal medical image fusion". *Biomedical Signal Processing and Control*, 97, 106694.

APMFUSION: AN INFRARED AND VISIBLE IMAGE FUSION MODEL BASED ON AGF AND PFCFUSE

Le Thi Hong Ha, Trinh Van Hau, Nguyen Van Cuong, Le Duc Tho,
Le Dinh Nghiep, Nguyen The Cuong

ABSTRACT

Infrared and visible image fusion plays an important role in enhancing target observation and scene understanding under adverse viewing conditions. However, many existing methods still fail to achieve an effective balance between preserving brightness, contrast and maintaining structural information. To address these limitations, we propose APMFusion, a fusion framework that integrates the AGF filter, the PFCFuse deep learning model, and the MRE operator. Experimental results conducted on 80 image pairs from the publicly available MSRS benchmark dataset demonstrate that APMFusion achieves the best or near-best performance across six widely used evaluation metrics, Q_{MLI} , Q_{CI} , Q_E , Q_{MI} , Q_{SSIM} , and $Q^{AB/F}$, when compared with five state-of-the-art fusion models published in 2024 and 2025. Source code and experimental results are publicly available at: https://github.com/house3173/AGF_IVIF.git.

Keywords: AGF, PFCFuse, MRE, infrared and visible image fusion.

* Ngày nộp bài: 05/02/2026; Ngày gửi phản biện: 11/02/2026; Ngày duyệt đăng: 28/02/2026

* Bài báo là kết quả nghiên cứu từ đề tài NCKH cấp cơ sở (mã số ĐT-2025-23) của Trường Đại học Hồng Đức.