1 DECEMBER 2025

DEA-NET

SINGLE IMAGE DEHAZING BASED ON DETAIL-ENHANCED CONVOLUTION AND CONTENT-GUIDED ATTENTION

IEEE TRANSACTIONS ON IMAGE PROCESSING (IEEE TIP)

Giới thiệu tổng quan

TIÊU ĐỀ

DEA-Net: Single image dehazing based on detail-enhanced convolution and content-guided attention

Khử mờ hình ảnh đơn dựa trên phép tích chập tăng cường chi tiết và cơ chế chú ý theo nội dung

TÁC GIẢ

Zixuan Chen

Zewei He

Zhe-Ming Lu

Cả 3 đều thuộc khoa hàng không và du hành vũ trụ, Đại học Chiết Giang, Hàng Châu, Trung Quốc

NĂM & NƠI CÔNG BỐ

Bản tiền ấn (preprint) trên arXiv đăng 12 Jan 2023; công trình sau đó được đăng/ghi nhận trên IEEE Transactions on Image Processing (TIP, 22/01/2024) — repo mã nguồn công khai trên GitHub.

Giới thiệu tổng quan

IEEE TRANSACTIONS ON IMAGE PROCESSING (IEEE TIP)

TỔNG QUAN NGHIÊN CỨU

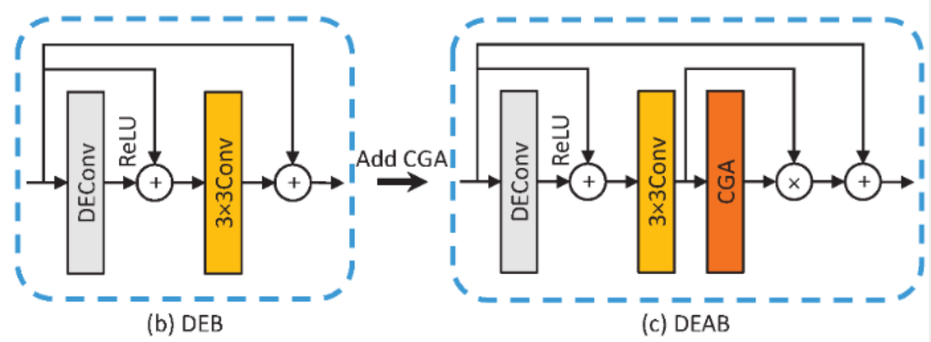
***Abstract*—Single image dehazing is a challenging ill-posed problem which estimates latent haze-free images from observed hazy images. Some existing deep learning based methods are devoted to improving the model performance via increasing the depth or width of convolution. The learning ability of convolutional neural network (CNN) structure is still under-explored. In this paper, a detail-enhanced attention block (DEAB) consisting of the detail-enhanced convolution (DEConv) and the content-guided attention (CGA) is proposed to boost the feature learning for improving the dehazing performance. Specifically, the DEConv integrates prior information into normal convolution layer to enhance the representation and generalization capacity. Then by using the re-parameterization technique, DEConv is equivalently converted into a vanilla convolution with NO extra parameters and computational cost. By assigning unique spatial importance map (SIM) to every channel, CGA can attend more useful information encoded in features. In addition, a CGA-based mixup fusion scheme is presented to effectively fuse the features and aid the gradient flow. By combining above mentioned components, we propose our detail-enhanced attention network (DEA-Net) for recovering high-quality haze-free images. Extensive experimental results demonstrate the effectiveness of our DEA-Net, outperforming the state-of-the-art (SOTA) methods by boosting the PSNR index over 41 dB with only 3.653 M parameters. The source code of our DEA-Net will be made available at https://github.com/cecret3350/DEA-Net.**

***Index Terms*—Image dehazing, Detail-enhanced convolution, Content-guided attention, Fusion scheme.**

Giới thiệu tổng quan

TỔNG QUAN NGHIÊN CỨU

Bài báo đề xuất một khối mới (Detail-Enhanced Attention Block — DEAB) và cả một kiến trúc tổng thể (DEA-Net) nhằm cải thiện chất lượng gỡ sương (dehazing) cho ảnh đơn (single-image).



Điểm nổi bật: tăng khả năng biểu diễn chi tiết trong CNN bằng cơ chế detail-enhanced convolution (DEConv) và một cơ chế chú ý theo nội dung (Content-Guided Attention, CGA).
Đồng thời họ dùng *kỹ thuật re-parameterization* để giữ chi phí tính toán/ tham số thấp ở thời gian chạy — dẫn tới mô hình nhẹ (~3.653 M tham số) nhưng đạt hiệu năng rất cao (PSNR $\gtrsim$ 41 dB trên bộ dữ liệu chuẩn).

$$\mathrm{MSE}=\frac{1}{m*n}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j)-K(i,j)]^2$$

$$\mathrm{PSNR}=10.\log_{10}\left(\frac{MAX_I^2}{MSE}\right)=20.\log_{10}\left(\frac{MAX_I^2}{\sqrt{MSE}}\right)$$

Những đóng góp này quan trọng vì dehazing là bài toán nghịch đảo, khó (ill-posed) và ứng dụng rộng.

Giới thiệu tổng quan

LÝ DO CHỌN PAPER

Thứ nhất: tuy mô hình nhỏ nhưng đạt kết quả cạnh tranh / vượt SOTA (*state-of-the-art: vượt trội hơn các phương pháp tiên tiến nhất hiện nay*) — phù hợp với các ứng dụng cần hiệu năng cao nhưng tài nguyên hạn chế.

Thứ hai: giới thiệu ý tưởng thiết kế module (*Sử dụng DEConv kết hợp CGA và mixup fusion cùng với re-parameterization*) mang tính sáng tạo và có thể tái sử dụng cho các tác vụ phục hồi ảnh khác.

Thứ ba: có mã nguồn chính thức trên GitHub — thuận tiện để reproduce / thử nghiệm trực tiếp.

Thứ tư: Tập trung vào các keyword sau đây để tìm hiểu về paper: DEConv, re-parameterization, CGA, mixup fusion

IEEE TIP

DEA-NET

Nội dung và phương pháp chính

IEEE TRANSACTIONS ON IMAGE PROCESSING (IEEE TIP)

MỤC TIÊU

Từ một ảnh bị sương mù (hazy) dự đoán ảnh sạch (haze-free) tương ứng — tức là ước lượng và phục hồi chi tiết, màu sắc và độ tương phản cho bức ảnh.

ĐỘNG CƠ

Nhiều mạng CNN hiện đại cải thiện bằng cách tăng chiều sâu/chiều rộng (deep/wide), nhưng vẫn còn hạn chế về khả năng nắm bắt chi tiết và chú ý theo nội dung (CGA) trong các đặc trưng (features). Đồng thời, mô hình lớn gây tốn tài nguyên. DEA-Net nhắm tới:

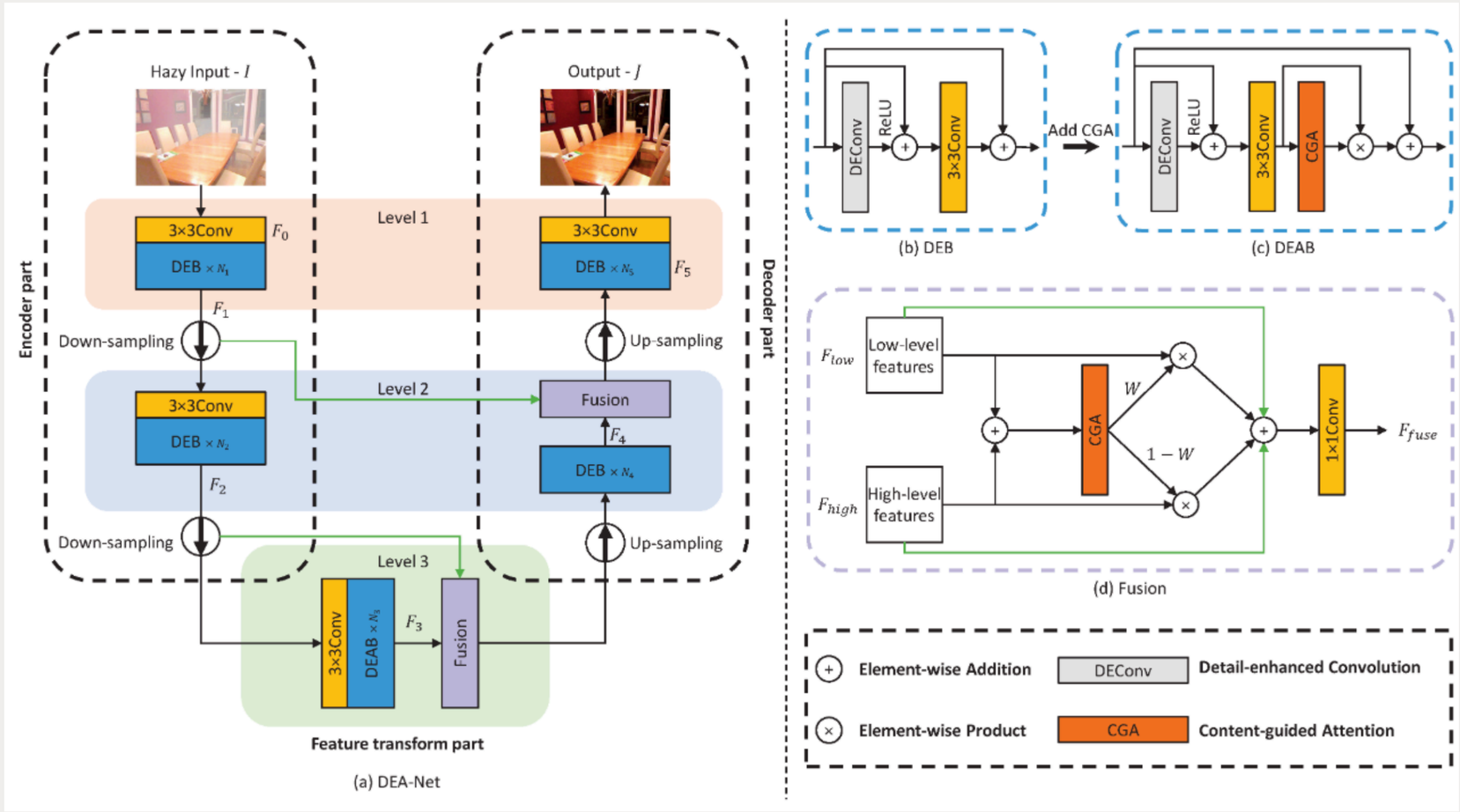
(1) Tăng khả năng học biểu diễn chi tiết

(2) Tập trung vào vùng hữu ích của ảnh theo kênh

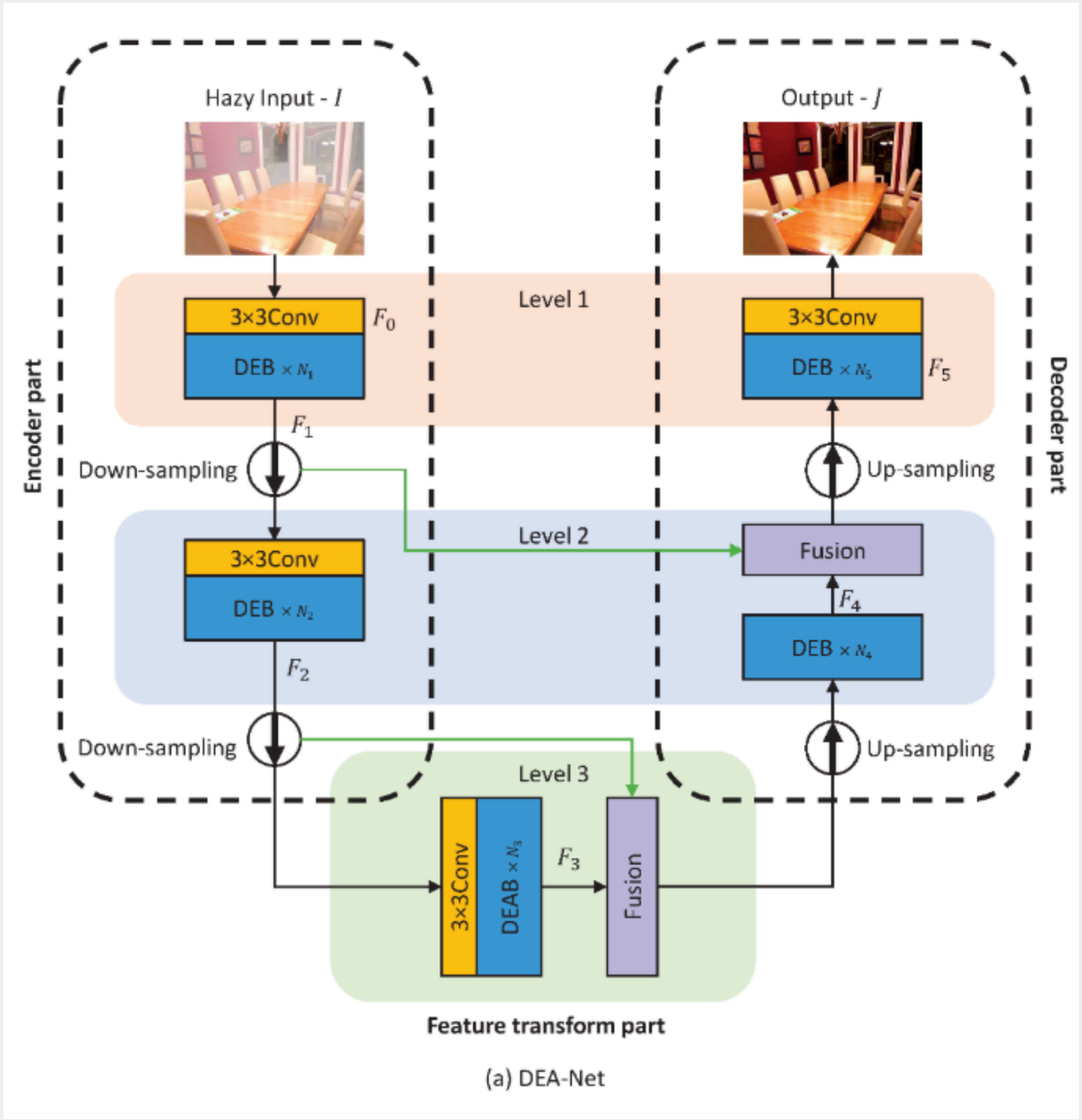
(3) Giữ mô hình gọn nhẹ khi suy luận (inference) bằng reparameterization.

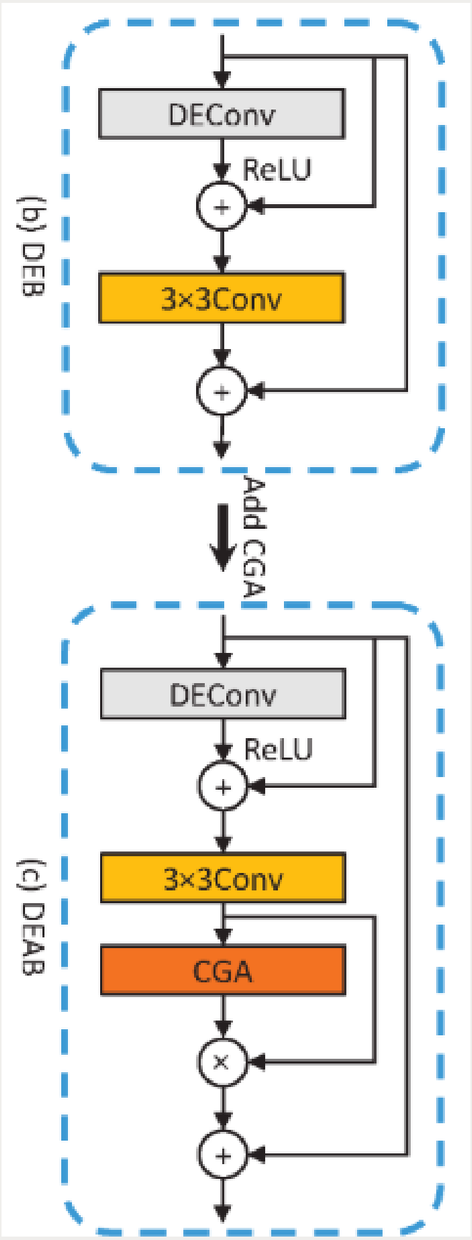
Nội dung và phương pháp chính

MÔ TẢ THUẬT TOÁN — CÁC THÀNH PHẦN CHÍNH VÀ PIPELINE



MÔ TẢ THUẬT TOÁN — CÁC THÀNH PHẦN CHÍNH VÀ PIPELINE





DEA-Net bao gồm ba phần:
Phần mã hóa, phần biến đổi đặc trưng và phần giải mã.

Là phần cốt lõi của DEA-Net, phần biến đổi đặc trưng áp dụng các khối chú ý tăng cường chi tiết (DEAB) xếp chồng lên nhau để học các đặc trưng không bị mờ.

Cấu trúc phân cấp có ba cấp, sử dụng các khối khác nhau ở các cấp độ khác nhau để trích xuất các đặc trưng tương ứng (cấp độ 1 và 2: DEB, cấp độ 3: DEAB)

MÔ TẢ THUẬT TOÁN — CÁC THÀNH PHẦN CHÍNH VÀ PIPELINE

Tích chập tăng cường chi tiết (DEConv)

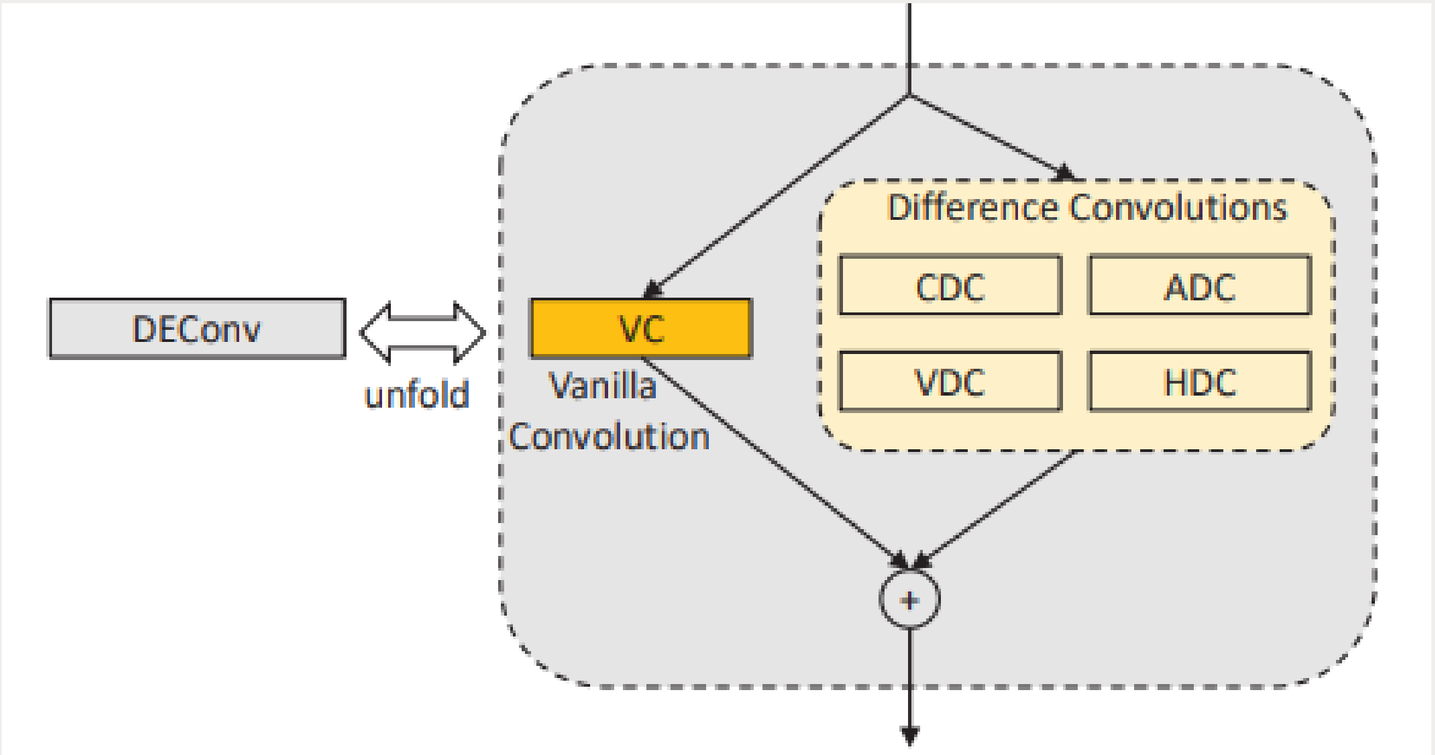
DEConv hoạt động bằng cách kết hợp:

1. Vanilla Convolution (VC)
2. Difference Convolutions (4 loại: CDC, ADC, VDC, HDC)

Tất cả sau đó được cộng (element-wise sum) để tạo ra feature "được tăng chi tiết"

DEConv được chèn bên trong DEB – Detail-Enhanced Block, là block chính dùng để:

+ Tăng cường chi tiết bị mất do sương mù
+ Giảm hiện tượng hình ảnh bị trơn, bị lem
+ Giúp mô hình tập trung vào các cấu trúc (edges, shapes) quan trọng



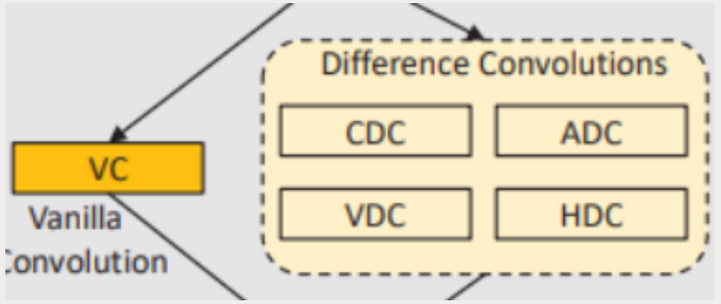
MÔ TẢ THUẬT TOÁN — CÁC THÀNH PHẦN CHÍNH VÀ PIPELINE

1. Unfold – Tách ảnh thành từng patch nhỏ

Ảnh được "giải nén" thành những cửa sổ nhỏ (patch 3×3 hoặc 5×5).

Mục tiêu:

- Giúp mô hình nhìn chi tiết cục bộ rõ hơn.
- Chuẩn bị dữ liệu cho difference convolution vốn cần ma trận nhiều chiều.
- Khi tổng hợp lại vừa rõ nét vừa giữ nguyên cấu trúc



2. Hai nhánh xử lý song song

Nhánh 1: Vanilla Convolution (VC)

- Nhánh này là convolution 3×3 thông thường.
- Nó giữ lại các nội dung "bình thường" của ảnh như: màu sắc tổng thể, texture cơ bản, cấu trúc không quá sắc nét
- Bản thân VC không phải để tăng chi tiết, mà để giữ lại thông tin gốc.

Nhánh 2: Difference Convolutions

- Difference Convolution = phép convolution mô phỏng gradient – directional derivative – variation theo các hướng khác nhau.
- Nó giúp mô hình nhận ra đường biên – cạnh sắc – chi tiết có thay đổi mạnh.

CDC – Central Difference Convolution

- Tính sự thay đổi quanh pixel trung tâm.
- Giúp lộ ra các biên rõ ràng và phần chuyển tiếp gắt.

VDC – Vertical Difference Convolution

- Nhạy với biến đổi theo hướng dọc.
- Bắt được chi tiết như: cạnh cửa, cạnh ghế, đường thẳng đứng trong phòng…

ADC – Angular Difference Convolution

- Bắt các chi tiết theo đường xiên, thường khó xử lý với conv chuẩn.
- Giúp phát hiện cạnh góc chéo, vật nghiêng, đường sát mặt bàn…

HDC – Horizontal Difference Convolution

- Nhạy với thay đổi hướng ngang.
- Bắt cạnh bàn, mép sàn, viền tường…

MÔ TẢ THUẬT TOÁN — CÁC THÀNH PHẦN CHÍNH VÀ PIPELINE

Tất cả 4 loại difference convolution tạo ra một không gian đặc trưng rất giàu chi tiết biên và texture, thứ mà ảnh bị haze thường làm mờ nhòe.

3. Cộng gộp (Element-wise Addition)

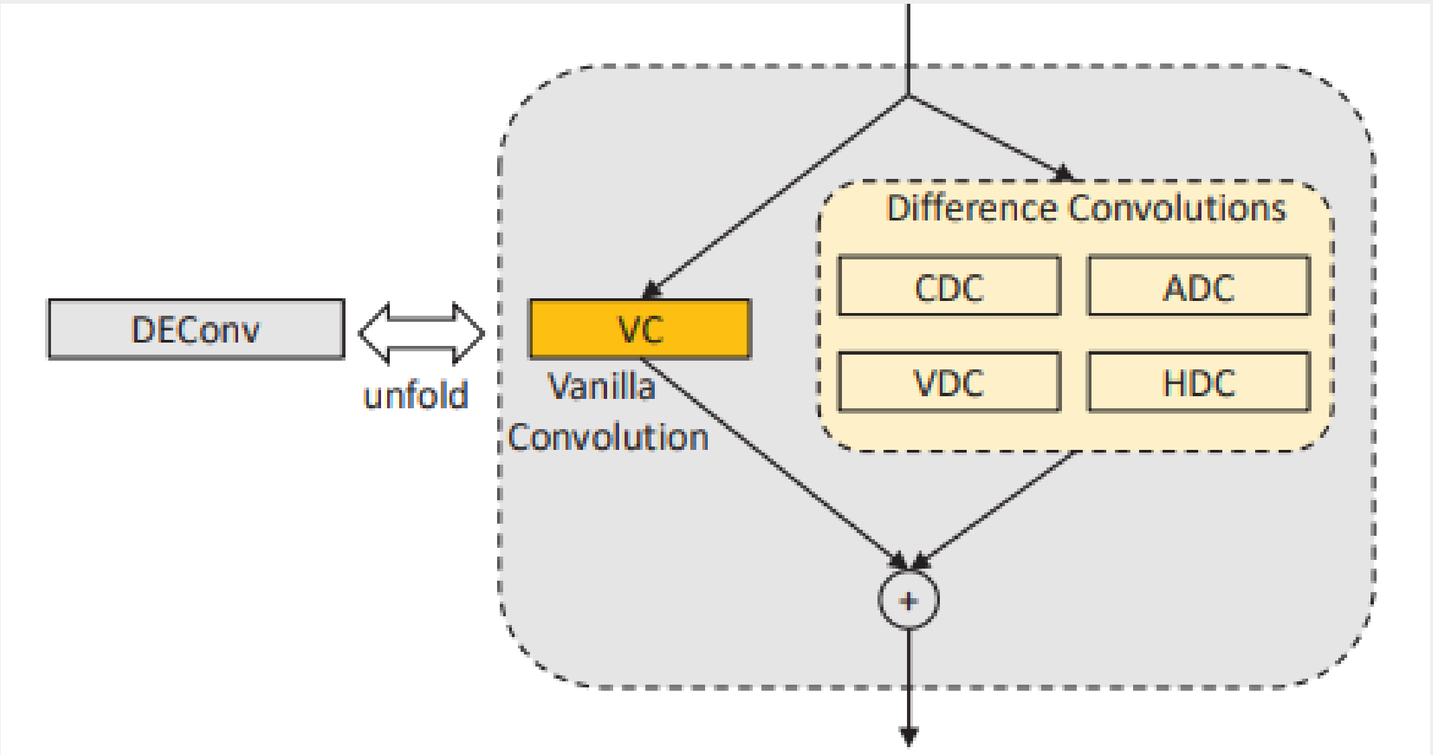
Hai nhánh:

- Kết quả VC
- Kết quả Difference Convolutions (tổng cộng từng loại)

→ được cộng trực tiếp với nhau.

Nhờ đó:

- VC giữ nội dung kiến trúc
- Difference Convolution bơm chi tiết mạnh
- Tổng hợp lại vừa rõ nét vừa giữ nguyên cấu trúc



MÔ TẢ THUẬT TOÁN — CÁC THÀNH PHẦN CHÍNH VÀ PIPELINE

Re-parameterization ở DEA-Net là kỹ thuật gộp nhiều nhánh tích chập (multiple convolution branches) vào một kernel duy nhất khi suy luận (inference), nhưng giữ nguyên cấu trúc nhiều nhánh khi training.

Trong quá trình Training (Backpropagation)
Mỗi nhánh có một kernel riêng: Các kernel này chạy song song, sinh ra 5 output feature maps → sau đó được cộng/kết hợp.
Điều này giúp mô hình học được các kiểu gradient khác nhau và khác biệt chi tiết (detail enhancement).

Trong quá trình suy luận (Forward Inference)
Thay vì chạy 5 convolution, ta gộp về 1 kernel duy nhất:

$$\text{Re-Parameterization: } w_i = \sum_{j=1}^{5} w_{i,j}$$

→ Chỉ cần 1 lần conv, biểu diễn và tốc độ như một mạng nhẹ.

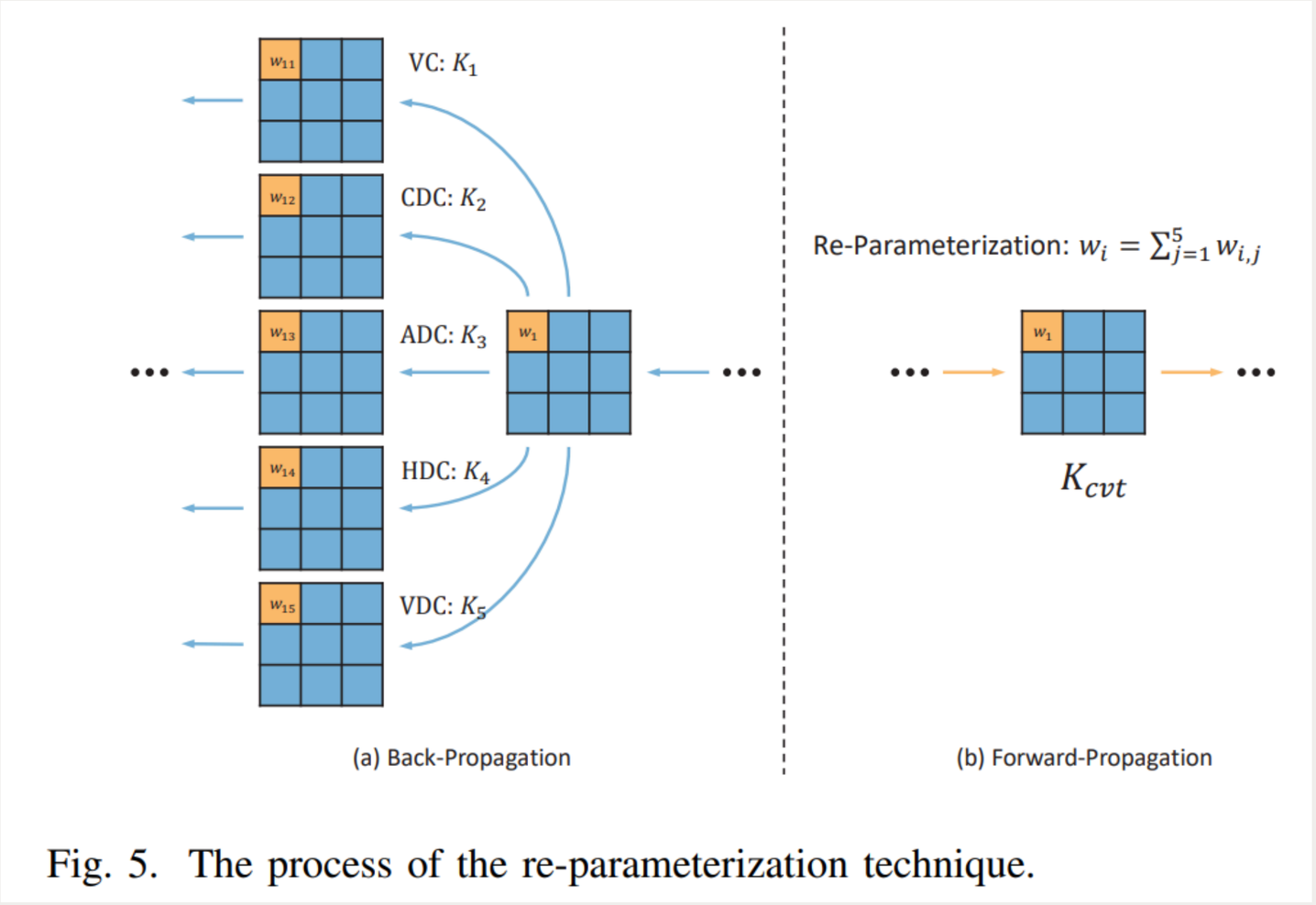


Fig. 5. The process of the re-parameterization technique.

Re-parameterization trong DEA-Net là kỹ thuật gộp các kernel của 5 nhánh tích chập lại thành 1 kernel duy nhất khi suy luận để giữ hiệu năng cao mà vẫn đảm bảo tốc độ thực thi nhanh.

MÔ TẢ THUẬT TOÁN — CÁC THÀNH PHẦN CHÍNH VÀ PIPELINE

1. Spatial Attention Branch (Nhánh màu xanh)

Mục tiêu: học vị trí quan trọng trên ảnh.

Gồm các bước:

(a) GMP – Global Max Pooling (theo kênh)

Tính giá trị max trên mỗi kênh → cho biết vùng nào nổi bật nhất.

(b) GAP – Global Average Pooling (theo kênh)

Tính giá trị trung bình → cho biết đặc trưng tổng quan.

(c) Channel-wise Concatenation

Ghép GMP và GAP theo chiều kênh → tạo đặc trưng chung:

(d) 7×7 Convolution

Ý nghĩa: Ws cho biết "pixel nào **quan trọng**".

2. Channel Attention Branch (Nhánh màu vàng)

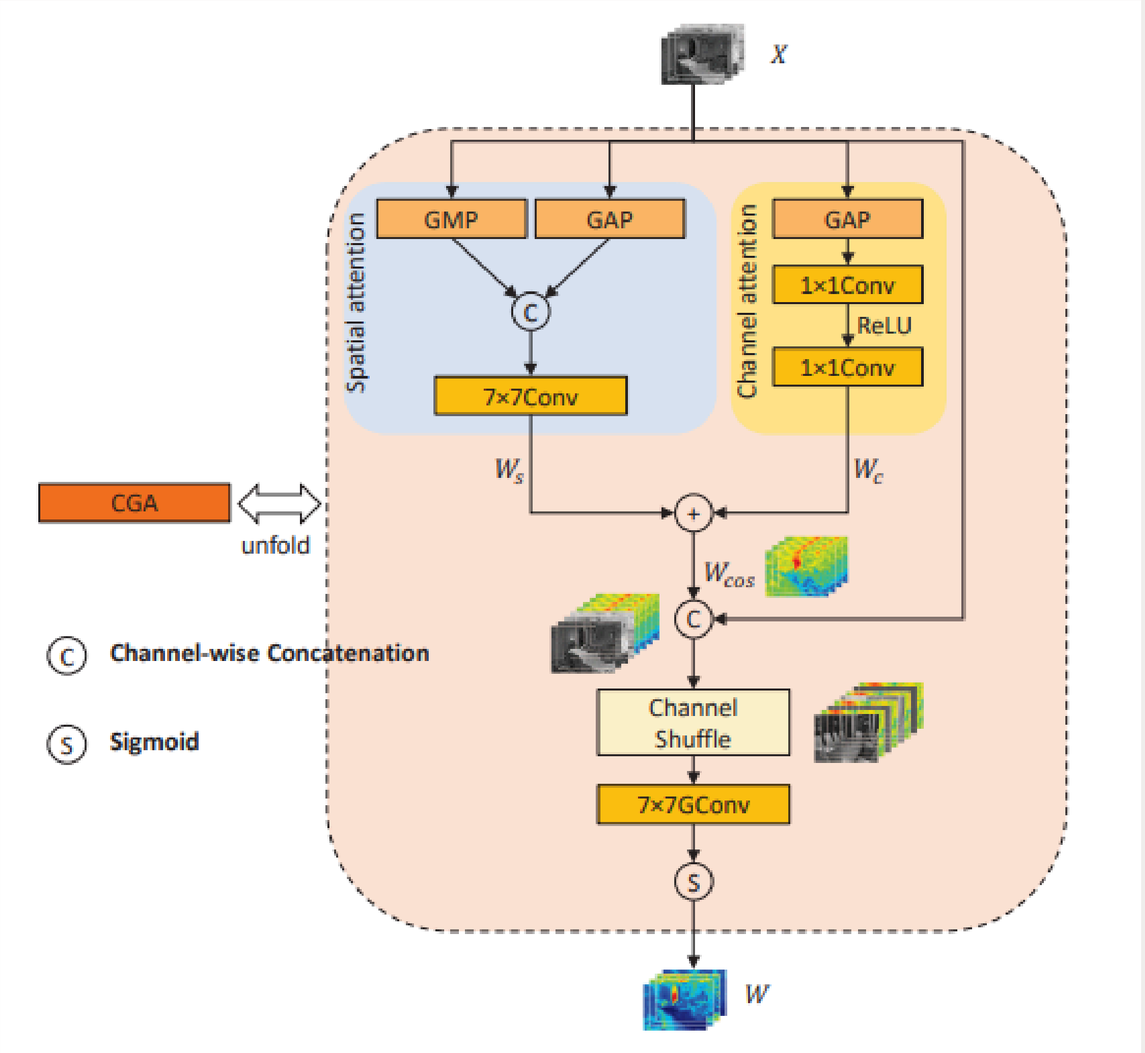
Mục tiêu: học kênh nào quan trọng trong feature map.

(a) Global Average Pooling theo không gian (GAP)

Nén H×W → 1 giá trị/kênh → vector độ quan trọng.

(b) 1×1 Conv → ReLU → 1×1 Conv

Ý nghĩa: Wc cho biết kênh nào mang nhiều thông tin **quan trọng** hơn.



CGA (Content-Guided Attention) Học attention dựa trên nội dung → phục hồi, tăng cường chi tiết quan trọng

MÔ TẢ THUẬT TOÁN — CÁC THÀNH PHẦN CHÍNH VÀ PIPELINE

Trong decoder và feature transform, ta có hai loại đặc trưng:

F_low – đặc trưng tầng thấp

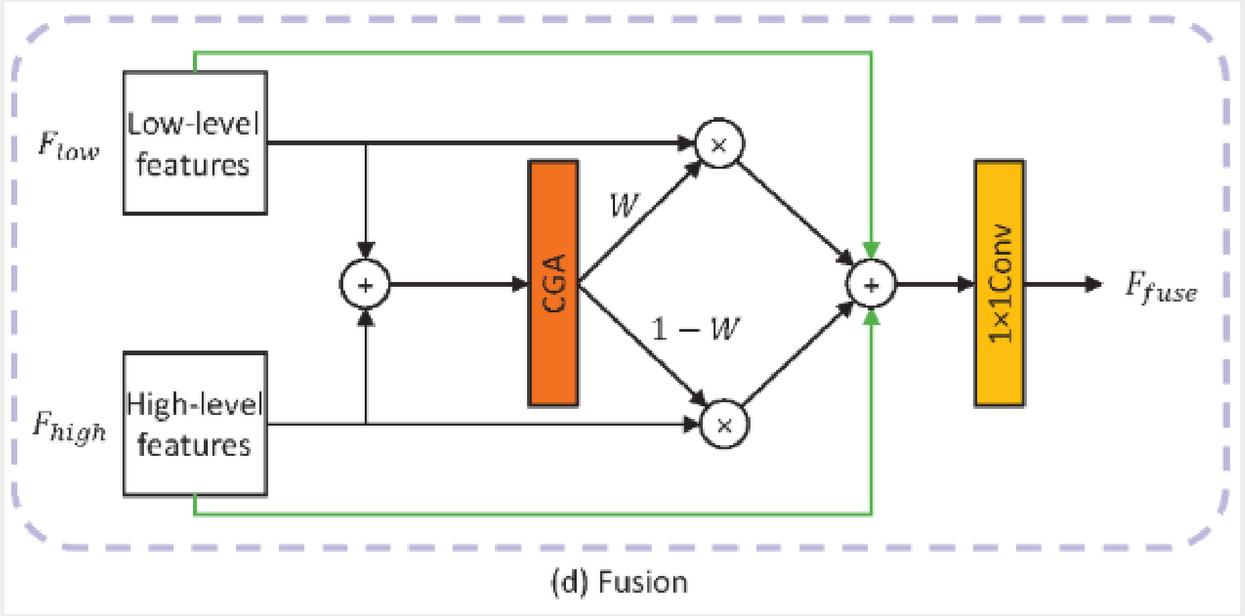
- Giàu texture, cạnh, chi tiết ảnh
- Nhưng nhiễu mạnh, không ổn định

F_high – đặc trưng tầng cao

- Ổn định
- Chứa thông tin ngữ nghĩa (semantic)
- Nhưng thiếu chi tiết (detail loss)

Mục tiêu: Kết hợp F_low + F_high sao cho ảnh khử sương vừa sắc nét vừa đúng cấu trúc.

→ ***CGA-based Mixup Fusion.***



(d) Fusion

Bước 1: Nhập hai dòng đặc trưng

- F_low
- F_high

Bước 2: CGA tính attention map W

- W có giá trị từ 0 → 1
- Kích thước như feature map
- Các vùng có nhiều texture → W cao
- Các vùng "phẳng" → W thấp

Bước 3: Mixup theo trọng số W và 1×1 Convolution

$$F_{fuse} = \mathcal{C}_{1\times1}(F_{low}\cdot W + F_{high}\cdot(1-W) + F_{low} + F_{high}), \quad (5)$$

Nội dung và phương pháp chính

KIẾN TRÚC TỔNG THỂ

- Ba phần chính: encoder (thu nhỏ không gian, tăng kênh), feature-transform (xử lý ở nhiều độ phân giải), decoder (phục hồi kích thước).
- Ba mức (levels): level1 (độ phân giải cao), level2 (độ phân giải trung), level3 (độ phân giải thấp).
- Có 2 lần down-sampling và 2 lần up-sampling (tức từ level1 → level2 → level3 rồi ngược lại).
- Quy ước kích thước:
  - level1: C × H × W
  - level2: 2C × H/2 × W/2
  - level3: 4C × H/4 × W/4
  - Trong phần triển khai của họ, C = 32.

Down-sampling và Up-sampling

- Down-sampling: thực hiện bằng convolution với stride = 2, đồng thời tăng số kênh gấp đôi (như nhiều thiết kế encoder).
- Up-sampling: dùng deconvolution (DEConv) — tức lớp transpose convolution — xem như phép ngược lại của down-sample để phục hồi không gian.
- Ý nghĩa: cách này giữ kiểm soát kích thước không gian và số kênh theo một quy luật đơn giản (mỗi lần giảm kích thước thì tăng độ sâu biểu diễn qua kênh).

Nội dung và phương pháp chính

KIẾN TRÚC TỔNG THỂ

DEB vs DEAB — tại sao khác nhau theo level?

Tác giả không chỉ chuyển đổi (transform) ở không gian thấp (low-resolution) như một vài phương pháp trước đó, mà làm việc từ level1 tới level3. Lý do: với tác vụ nhạy cảm tới chi tiết như khử sương, việc chỉ xử lý ở low-res sẽ mất nhiều chi tiết.

Do đó:

- Ở level1 & level2 (độ phân giải cao và trung): dùng DEB (một block nhẹ/ít phức tạp hơn phù hợp với thông tin chi tiết).
- Ở level3 (độ phân giải thấp, biểu diễn tính tổng quát cao hơn): dùng DEAB (một block có Attention/CGA mạnh hơn để xử lý ngữ cảnh rộng).
- Tức là họ điều chỉnh cấu trúc block theo vai trò và độ phân giải của từng level — giúp vừa giữ chi tiết (levels cao) vừa học quan hệ toàn cục (level thấp).

Feature fusion (kết hợp đặc trưng) — skip connections dạng mixup

- Sau các down-sampling, các feature tương ứng ở encoder được fuse (kết hợp) với feature trước khi up-sampling tại decoder — mũi tên màu xanh lá trong hình.
- Việc kết hợp dùng CGA-based mixup fusion: tức là dùng Cross-Gating Attention để điều khiển (gating) và trộn (mixup) hai nguồn thông tin (encoder và decoder) theo cách có điều kiện/không đơn thuần là concat hoặc add.
- Mục đích: vừa truyền lại chi tiết từ encoder, vừa điều chỉnh bằng attention để không đưa vào noise/haze không mong muốn.

Nội dung và phương pháp chính

KẾT QUẢ

Tiêu chí đánh giá

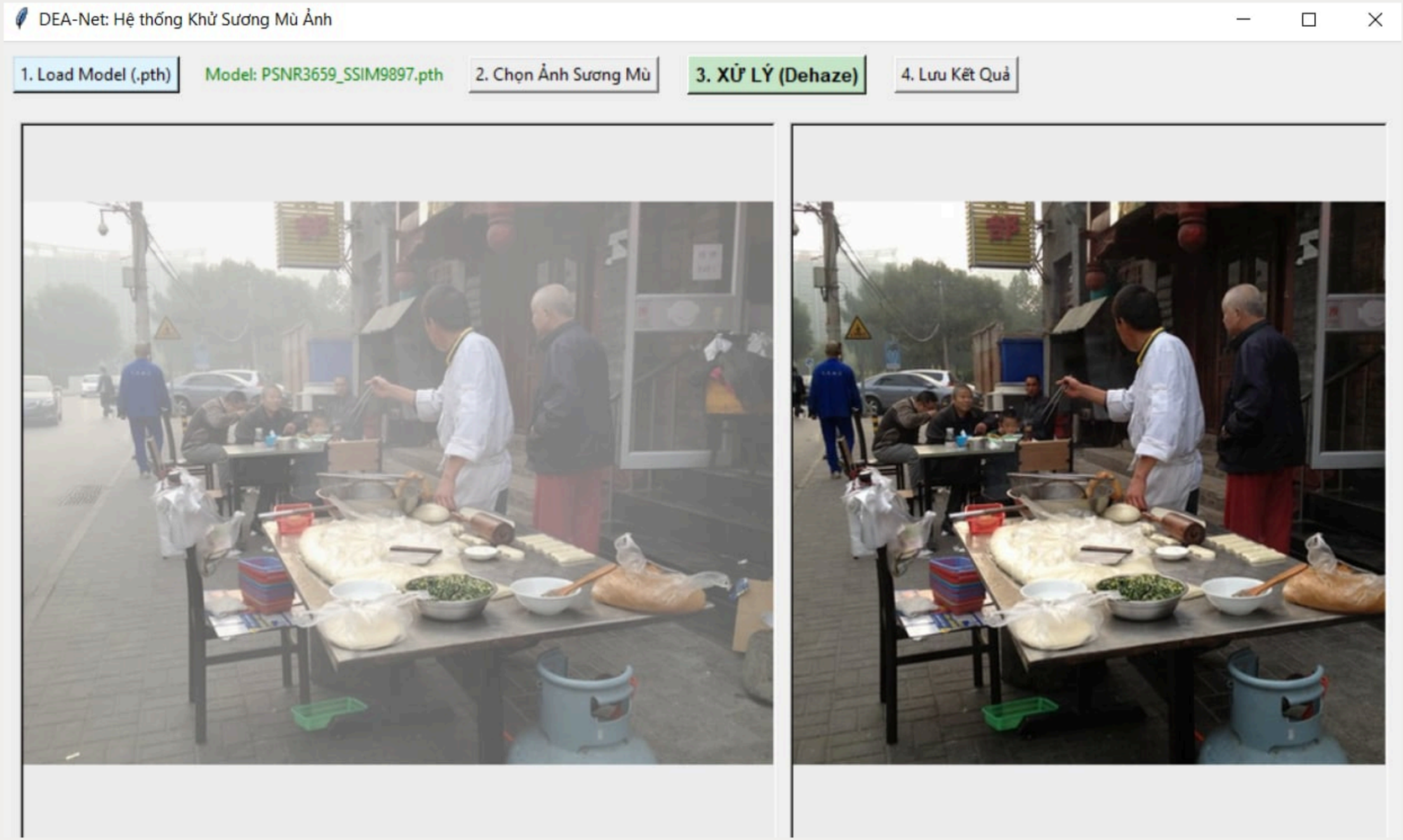
SSIM *(đo khác nhau giữa đầu vào và sinh ra dựa trên độ chói, tương phản và cấu trúc)* có giá trị trong khoảng từ -1 đến 1, đạt giá trị bằng 1 trong trường hợp hai bộ dữ liệu giống hệt nhau. Chỉ số này có giá trị càng lớn thì tương ứng với model càng tốt.

PSNR càng cao thì chất lượng ảnh sinh càng tốt, khi 2 ảnh giống hệt nhau thì MSE=0 và PSNR đi đến vô hạn.

Method	SOTS-indoor [38]		SOTS-outdoor [38]		Haze4K [39]		Overhead		
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	# Param. (M)	# FLOPs (G)	Runtime (ms)
(TPAMI'10) DCP [12]	16.61	0.8546	19.14	0.8605	14.01	0.76	-	-	-
(TIP'16) DehazeNet [2]	19.82	0.8209	27.75	0.9269	19.12	0.84	0.008	0.5409	0.9932
(ICCV'17) AOD-Net [3]	20.51	0.8162	24.14	0.9198	17.15	0.83	0.0018	0.1146	0.3159
(CVPR'18) GFN [23]	22.30	0.8800	21.55	0.8444	-	-	0.4990	14.94	-
(AAAI'20) FFA-Net [5]	36.39	0.9886	33.57	0.9840	26.97	0.95	4.456	287.5	47.98
(CVPR'20) MSBDN [10]	32.77	0.9812	34.81	0.9857	22.99	0.85	31.35	**24.44**	9.826
(ACMMM'21) DMT-Net [39]	-	-	-	-	28.53	0.96	51.79	75.56	26.83
(CVPR'21) AECR-Net [6]	37.17	0.9901	-	-	-	-	**2.611**	52.20	-
(TIP'22) SGID-PFF [21]	38.52	0.9913	30.20	0.9754	-	-	13.87	152.8	20.92
(AAAI'22) UDN [22]	38.62	0.9909	34.92	0.9871	-	-	4.250	-	-
(ECCV'22) PMDNet [11]	38.41	0.9900	34.74	0.9850	33.49	0.98	18.90	-	-
(CVPR'22) Dehamer [17]	36.63	0.9881	35.18	0.9860	-	-	132.4	48.93	14.12
(Ours) DEA-Net-S	39.16	0.9921	-	-	-	-	2.844	24.88	**5.632**
(Ours) DEA-Net	40.20	0.9934	36.03	0.9891	33.19	**0.99**	3.653	32.23	7.093
(Ours) DEA-Net-CR	**41.31**	**0.9945**	**36.59**	**0.9897**	**34.25**	**0.99**	3.653	32.23	7.093

BẢNG SO SÁNH ĐỊNH LƯỢNG CÁC PHƯƠNG PHÁP KHẮC PHỤC HAZING KHÁC NHAU TRÊN SOTS-INDOOR, SOTS-OURDOOR VÀ HAZE4K.

Demo



IEEE TRANSACTIONS ON IMAGE PROCESSING (IEEE TIP)