

Electronic skin pressure map mask generation based on optimized Real-ESRGAN-MSDC

Peng Xu(徐鹏), Jia-He Chen(陈迦禾), Yu-Kang Fu(傅裕康), Tian-En Dai(戴天恩), Xi-Qin Wang(王希沁), Pei-Lin Li(李沛霖), Jian-Wei Shuai(帅建伟), Qing-Hua Xia(夏庆华), Jin-Jin Zhong(钟金金), Qi Zhao(赵琪)

Citation: Chin. Phys. B. 35, 090701 (2026). DOI: 10.1088/1674-1056/ae3071 |
CSTR: 32038.14.CPB.ae3071
http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

M2ANet: Multi-branch and multi-scale attention network for medical image segmentation

Wei Xue(薛伟), Chuanghui Chen(陈创慧), Xuan Qi(戚轩), Jian Qin(秦健), Zhen Tang(唐振), Yongsheng He(何永胜)
Chin. Phys. B. 34, 080703 (2025). DOI: 10.1088/1674-1056/adce96 | CSTR: 32038.14.CPB.adce96

Testing algorithm for the computation of the transverse emittance of the ion beams generated by the ECR mVINIS ion source based on a pepper-pot method

Viktor Jocić, Igor Telečki, Ivan Trajić
Chin. Phys. B. 34, 060701 (2025). DOI: 10.1088/1674-1056/adc674 | CSTR: 32038.14.CPB.adc674

FPGA and computer-vision-based atom tracking technology for scanning probe microscopy

Feng-Du Yu(俞风度), Li Liu(刘利), Su-Ke Wang(王肃珂), Xin-Biao Zhang(张新彪), Le Lei(雷乐), Yuan-Zhi Huang(黄远志), Rui-Song Ma(马瑞松), Qing Huan(郇庆)
Chin. Phys. B. 33, 050705 (2024). DOI: 10.1088/1674-1056/ad34cb

A self-powered and sensitive terahertz photodetection based on PdSe2

Jie Zhou(周洁), Xueyan Wang(王雪妍), Zhiqingzi Chen(陈支庆子), Libo Zhang(张力波), Chenyu Yao(姚晨禹), Weijie Du(杜伟杰), Jiazhen Zhang(张家振), Huaizhong Xing(邢怀中), Nanxin Fu(付南新), Gang Chen(陈刚), Lin Wang(王林)
Chin. Phys. B. 31, 050701 (2022). DOI: 10.1088/1674-1056/ac4908

Visibility enhancement in two-dimensional lensless ghost imaging with true thermal light

Xi-Hao Chen(陈希浩), Ling Yan(燕玲), Wei Wu(吴炜), Shao-Ying Meng(孟少英), Ling-An Wu(吴令安), Zhi-Bin Sun(孙志斌), Chao Wang(王超), Guang-Jie Zhai(翟光杰)
Chin. Phys. B. 26, 060702 (2017). DOI: 10.1088/1674-1056/26/6/060702

Electronic skin pressure map mask generation based on optimized Real-ESRGAN-MSDC

Peng Xu(徐鹏)[1,†], Jia-He Chen(陈迦禾)[2,†], Yu-Kang Fu(傅裕康)[3], Tian-En Dai(戴天恩)[4], Xi-Qin Wang(王希沁)[4], Pei-Lin Li(李沛霖)[3], Jian-Wei Shuai(帅建伟)[5], Qing-Hua Xia (夏庆华)[6,‡], Jin-Jin Zhong(钟金金)[5,§], and Qi Zhao(赵琪)[7,¶]

[1] *School of Electronics and Information Engineering, University of Science and Technology Liaoning, Anshan* 114051, *China*
[2] *The First School of Medicine, School of Information and Engineering, Wenzhou Medical University, Wenzhou* 325000, *China*
[3] *School of Engineering, Zhejiang University, Hangzhou* 310000, *China*
[4] *The Second Clinical Medical College of Wenzhou Medical University, Wenzhou* 325000, *China*
[5] *Wenzhou Institute, University of Chinese Academy of Sciences, Wenzhou* 325001, *China*
[6] *Ningbo Global Innovation Centre, Zhejiang University, Ningbo* 315100, *China*
[7] *School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan* 114051, *China*

(Received 24 November 2025; revised manuscript received 22 December 2025; accepted manuscript online 23 December 2025)

Electronic skin (e-skin) visualizes contact pressure distribution through pressure maps, offering significant potential for robotic tactile sensing and health monitoring. However, insufficient sensor resolution often causes image shape distortion, limiting its use in high-precision applications. We propose an optimized Real-ESRGAN-MSDC model to generate high-fidelity shape masks from distorted pressure maps. Experiments are conducted using 432 image pairs with three rounds of 5-fold cross-validation. Performance differences between traditional computer vision mask generation (TCVMG), Pix2pix, CycleGAN, ESRGAN, Real-ESRGAN, and Real-ESRGAN-MSDC are evaluated using four metrics. The results demonstrate that Real-ESRGAN-MSDC outperforms other models, achieving 1.00% higher IoU, 0.59% higher F1 score, 0.70% higher SSIM, and 0.38% lower LPIPS compared to the second- ranked model, indicating improved structural similarity and edge fidelity. Our study provides novel insights into intelligent applications of e-skin for tactile sensing and health monitoring. Future work could expand the dataset, optimize training strategies, and reduce computational complexity to enable arbitrary shape prediction and real-time applications.

Keywords: electronic skin, pressure map, image enhancement, mask generation

PACS: 07.05.Pj, 07.05.Mh, 07.07.Df, 87.85.fk

DOI: 10.1088/1674-1056/ae3071 **CSTR:** 32038.14.CPB.ae3071

1. Introduction

Electronic skin (e-skin), an artificial integument inspired by human skin, integrates flexible sensor arrays to mimic tactile perception, enabling real-time detection of pressure, temperature, strain, and other stimuli through wearable, biocompatible interfaces.[1,2] This biomimetic technology holds transformative potential in human-machine interaction, robotic tactile sensing, and health monitoring by providing high-resolution, multimodal feedback for precise environmental adaptation and physiological assessment.[2] The rapid development of flexible electronics has advanced these e-skin applications. In recent years, Cao and Cai advanced e-skin applications in 2022 by developing novel material systems,[3] while Yang *et al*. in 2019 and Liu *et al*. in 2022 enhanced e-skin for human-machine interaction, robotic tactile sensing, and health monitoring through innovations in sensor arrays and wireless interfaces.[4,5] In 2024, Fu *et al*. integrated flexible substrate materials, such as polydimethylsiloxane and polyimide, with micro-nanofabrication techniques to enhance e-skin's sensitivity and wearability, leveraging advances in materials science and artificial intelligence.[6] By integrating miniature sensor arrays, such as capacitive, piezoresistive, and piezoelectric sensors, e-skin enables real-time monitoring of physical quantities, including pressure, temperature, humidity, and strain, providing innovative solutions for smart robots, medical diagnosis, and wearable devices.[7–9] These sensor arrays typically employ multilayer structures, such as flexible films based on graphene or carbon nanotubes, maintaining stable performance in complex environments. In 2022, Chen *et al*. developed self-healing materials for e-skin,[10] followed by Zarei *et al*. in 2023 and Liu *et al*. in 2023, who introduced biodegradable polymers, significantly expanding e-skin's potential in biomedical and wearable devices.[11,12]

Its flexible substrate and biocompatibility confer unique

[†]These authors contributed equally to this work.
[‡]Corresponding author. E-mail: xiaqinghua@zju.edu.cn
[§]Corresponding author. E-mail: zjj89@wiucas.ac.cn
[¶]Corresponding author. E-mail: zhaoqi@lnu.edu.cn

© 2026 Chinese Physical Society and IOP Publishing Ltd. All rights, including for text and data mining, AI training, and similar technologies, are reserved.

advantages to e-skin in the fields of bionic robotics, prosthetic control, and telemedicine.[12,13] In bionic robotics, e-skin mimics the tactile function of human skin, enabling precise grasping and environmental adaptation through high-resolution pressure sensing. In 2025, Sankar *et al*. demonstrated that e-skin equipped with flexible pressure sensors enables dynamic manipulation of soft objects, improving robotic grasping precision.[14] In prosthetic control, e-skin enhances natural user–prosthetic interaction through real-time pressure and temperature feedback. In 2025, Patil *et al*. demonstrated that e-skin response times have been reduced to the millisecond range, enabling real-time tactile feedback.[15] In telemedicine, e-skin, integrated with wireless transmission modules, enables remote monitoring of patients' vital signs. In 2023, Tian *et al*. utilized flexible sensor arrays to monitor heart rate and skin hydration changes, providing a novel tool for chronic disease management.[16] Furthermore, e-skin's multimodal sensing capability enables it to perform exceptionally in complex scenarios. For example, in surgical robots, e-skin, integrating visual and tactile data, enhances surgical precision.[16,17]

However, dynamic pressure variations and non-uniform surface contact often cause sensor data ambiguity, particularly in generating pressure maps, making it challenging to capture pressure peaks and edge details due to limited resolution and acquisition conditions.[18–21] This ambiguity primarily arises from the nonlinear response of sensor arrays under dynamic loading and local signal distortion due to non-uniform contact, causing significant distortion in the shape of acquired pressure maps relative to the actual pressure profile.[22] For example, non-uniform contact may cause contour distortion in pressure maps, and dynamic pressure variations worsen edge information distortion, particularly in recognizing complex geometric shapes (e.g., letters and numbers).[23] Such shape distortion in pressure maps limits high-precision applications. In robotic grasping tasks, the current millimeter-level spatial resolution struggles to capture the fine contact contours of irregular objects, leading to pose estimation errors. In health monitoring, large-area flexible arrays often misjudge high-pressure zone areas due to folds.[24–26]

Furthermore, multimodal sensing requirements further emphasize the importance of pressure map-based shape extraction techniques in e-skin data processing.[16] As e-skin evolves toward intelligent systems, single pressure sensing is insufficient for complex application scenarios, and multimodal data integrating temperature, humidity, and tactile feedback has become a trend.[16,24,27,28] However, this challenge is compounded by limitations in traditional image processing methods, such as bilinear and bicubic interpolation, for detail recovery.[29] These methods interpolate by assuming local pixel smoothness, making it challenging to reconstruct high-frequency edge information and performing particularly poorly on small-sample datasets.[30] In recent years, deep learning-based shape extraction techniques demonstrate significant advantages in handling complex contours, offering new approaches for shape recovery in e-skin pressure maps. For example, U-Net excels at capturing multi-scale image features through its encoder-decoder structure and skip-connection design, and is widely applied in medical image segmentation and contour extraction tasks.[31] E-SegNet introduces a novel E-shaped architecture that optimizes the U-Net framework for medical image segmentation, emphasizing multi-scale feature aggregation, parameter efficiency, and complex boundary processing.[32] Generative adversarial networks (GANs) generate realistic images through adversarial training of generators and discriminators, making them particularly suitable for recovering high-resolution details from ambiguous inputs.[33]

In 2017, Isola *et al*. developed Pix2pix, a conditional GAN model that learns image-to-image mappings using paired input-output data, demonstrating superior structural preservation in pressure map-to-mask conversion.[34] In the same year, Zhu *et al*. proposed CycleGAN, which achieves unpaired image conversion through cyclic consistency loss, suitable for scenarios lacking paired data, though its training stability requires further optimization.[35] In 2018, Wang *et al*. developed ESRGAN, which enhances detail recovery in super-resolution images through residual dense blocks and perceptual loss, making it suitable for enhancing the clarity of pressure maps.[36] However, these models still face bottlenecks in small-sample optimization and complex contour extraction for e-skin pressure maps, necessitating the development of dedicated high-precision shape extraction techniques.[37] Therefore, we develop dedicated techniques for generating masks closely resembling the actual pressure profile from distorted pressure maps, addressing a critical challenge.

To address these challenges, we propose an optimized Real-ESRGAN-MSDC model designed to generate high-quality masks closely resembling target shapes from e-skin pressure maps. The ground truth (GT) mask, generated from the actual shape of pressed e-skin, serves as a reference for evaluating the accuracy of model predictions. We integrate MultiScaleConv and DetailConv modules into the residual-in-residual dense block (RRDB) architecture of Real-ESRGAN, enhancing multi-scale feature extraction and edge detail reconstruction capabilities, respectively. This approach optimizes model performance, particularly for small-sample datasets.

Our study utilizes a dataset provided by a partner, derived from low-resolution pressure maps and their corresponding GT masks acquired by an array of flexible pressure sensors,

to validate the model's mask generation capability and lay the foundation for future extensions. The mask generation method TCVMG, based on threshold segmentation and morphological operations, exhibits limited generalization ability in complex pressure distributions and small-sample scenarios. In contrast, existing deep learning methods, such as Pix2pix and CycleGAN, improve performance but face limitations in data requirements and training stability, respectively. Meanwhile, ESRGAN and Real-ESRGAN optimize image processing but remain inadequate for shape extraction in small-sample e-skin pressure maps. Therefore, the innovation of Real-ESRGAN-MSDC lies in addressing the gap in shape mask generation for small-sample pressure maps, providing a novel approach for the intelligent application of e-skin in tactile sensing and health monitoring.

This paper is organized as follows. Section 1 introduces the background and significance of the study. Section 2 presents the methodology, including model architecture and module design. Section 3 details the experimental results, covering datasets, evaluation metrics, and experimental setup. And Section 4 analyzes the results and discusses future directions.

2. Materials and methods

2.1. Overall structure

Our model comprises a generator and a discriminator, with the generator producing high-quality masks closely resembling the GT shape, and the discriminator evaluating the realism of the generated masks. We adopt the UNetDiscriminatorSN discriminator from Real-ESRGAN, with a focus on optimizing the generator and leveraging the official pre-trained Real-ESRGAN model to accelerate convergence in small-sample scenarios. The RRDB architecture is adapted to address complex feature extraction requirements. Figure 1(a) depicts the overall workflow of Real-ESRGAN-MSDC. It outlines the end-to-end process from input e-skin pressure maps and GT masks through data augmentation, including horizontal flipping and rotation, to mask generation and authenticity assessment via the discriminator's probability distribution. This visualization underscores the adversarial feedback loop that refines mask quality for complex e-skin contours. The input pressure map and paired GT masks are first augmented with data (including horizontal flipping and rotation) before being fed into the generator. The generator employs the RRDB$\times$23 structure and residual dense blocks for feature extraction, integrating MultiScaleConv2d and DetailConv2d modules to enhance multi-scale details and edge information, ultimately generating probability maps through two upsampling layers. The generator produces a mask image, which is evaluated by the discriminator to generate a probability map reflecting the difference between the generated mask and the ground truth mask. This probability map is used to compute the loss function, including the adversarial loss function and the total loss function, optimizing the generator's and discriminator's parameters to better align the mask's boundaries and shapes with the ground truth, with the final mask obtained through binarization to enhance edge details for e-skin pressure maps. The generator substantially enhances the structural details and quality of the mask by optimizing the total loss, combining content and adversarial losses. During this process, the discriminator adopts the UNetDiscriminatorSN architecture, employing multilayer convolution and spectral normalization to assess the realism of generated and GT masks, outputting probability distributions. Through adversarial loss optimization, the discriminator accurately distinguishes between GT and generated masks, refining discriminant boundaries and enhancing the realism of the generator's outputs.

The generator employs the RRDBNet structure, based on RRDBs, comprising 23 RRDB layers. Each layer consists of five densely connected 3×3 convolutions with 64 feature channels, 32 growth channels, and 3 input and output channels, matching RGB images and using a scaling factor of 1 for the mask generation task. The input undergoes a pixel unshuffle operation to expand the channel count by 16 times, with features extracted by an initial 3×3 convolutional layer. The main network comprises 23 RRDBs, each integrating DetailConv modules to enhance detail processing. The main network's output is fused by a 3×3 convolutional layer before entering the upsampling stage. This stage employs MultiScaleConv modules for two upsampling operations, combined with nearest-neighbor interpolation and convolution, ultimately producing a 128×128 pixel mask via a final 3×3 convolutional layer. The dense connectivity of RRDBNet optimizes gradient flow and facilitates deep network convergence, making it suitable for small-sample scenarios. As illustrated in Fig. 1(b), the generator network highlights the integration of 23 RRDB modules with DetailConv and MultiScaleConv in the upsampling stage, leveraging residual dense blocks and multi-scale convolution to enhance feature extraction and edge detail recovery for distorted e-skin pressure maps.

The discriminator, based on UNetDiscriminatorSN within the U-Net architecture, includes four downsampling layers with channel counts of 64, 128, 256, and 512, and four upsampling layers with channel counts of 256, 128, 64, and 1. Spectral normalization is employed to stabilize training, with skip connections enabled to integrate multi-scale features. Presented in Fig. 1(c), the discriminator's architecture emphasizes spectral normalization and skip connections across downsampling and upsampling stages, enabling per-pixel scoring that distinguishes real GT masks from generated ones, thereby stabilizing training for e-skin mask realism.

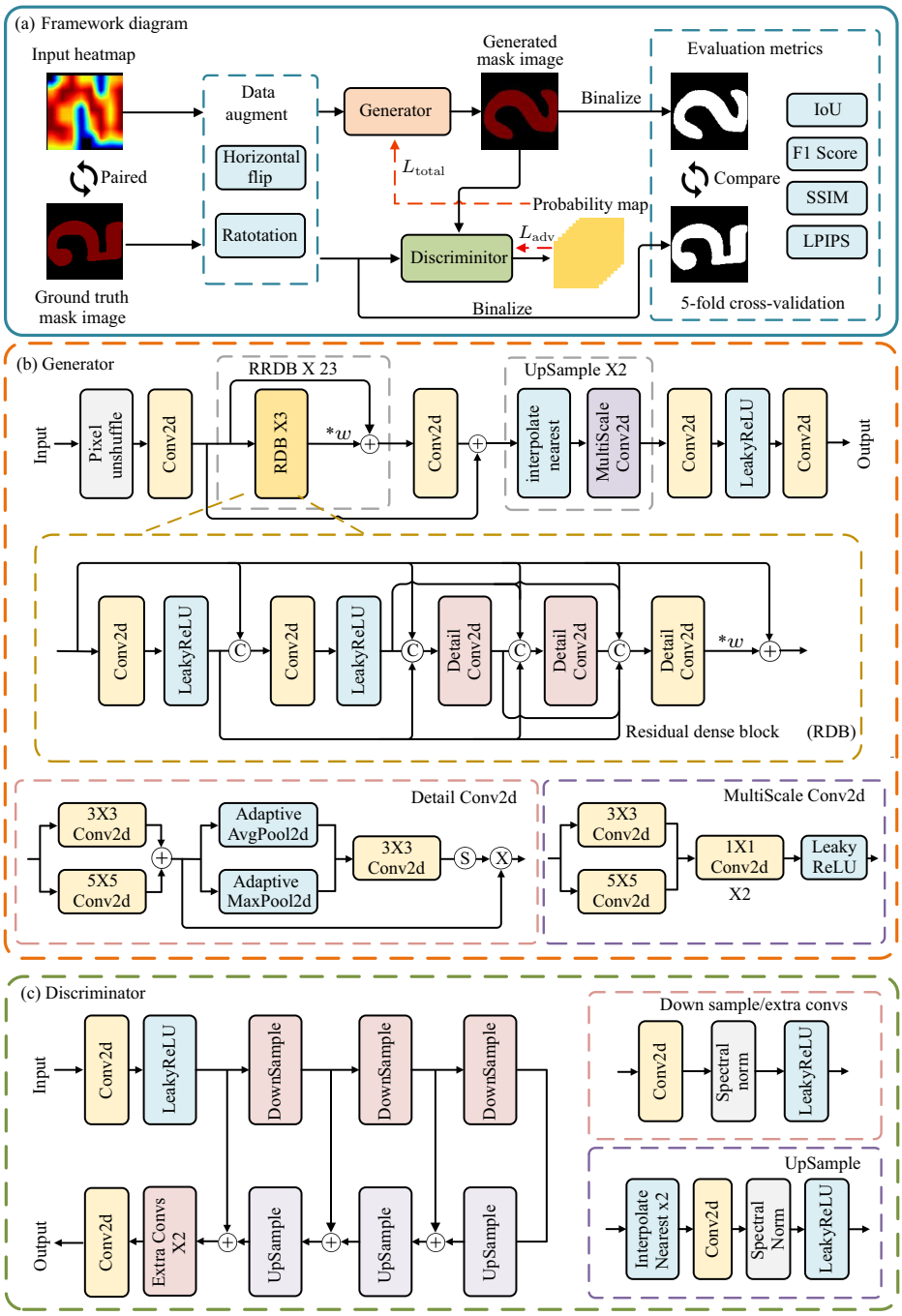


Fig. 1. Architectural diagram of Real-ESRGAN-MSDC. (a) Workflow from input heatmap and ground truth mask to generator and discriminator, including data augmentation and evaluation metrics. (b) Generator details with 23 RRDB modules, upsampling, DetailConv, and MultiScaleConv ($w = 0.2$). (c) Discriminator details using UNetDiscriminatorSN with spectral normalization and Leaky ReLU.

2.2. DetailConv module

DetailConv modules are integrated into RRDBs to enhance detail recovery for 26 English alphabets and digits 0–9. The module includes 3×3 and 5×5 convolutional branches with a 7:3 output channel ratio, processing features from different receptive fields independently and fusing the outputs through addition. The spatial attention mechanism extracts global statistical information through adaptive mean pooling and maximum pooling, compressing it to one channel via a 3×3 convolutional layer. Sigmoid activation generates the attention weight map, and bilinear interpolation restores the resolution, which is then multiplied with the convolutional output to enhance critical regions. The attention mechanism leverages information theory to optimize feature selection and reduce redundant information. Residual connections adjust dimensions via a 1×1 convolutional layer, with weights initialized using Kaiming initialization and the output activated by Leaky ReLU with a negative slope of 0.1. The overall output is expressed by

$$A = \text{Sigmoid}(W_{\text{conv}} \cdot (\text{AvgPool}(F) + \text{MaxPool}(F))), \quad (1)$$

$$F_{\text{out}} = \text{Leaky ReLU}(W_{1\times1} \cdot (A \cdot (0.7 \cdot F_{3\times3} + 0.3 \cdot F_{5\times5}) + F_{\text{in}})), \quad (2)$$

where A denotes the attention weight computation, F denotes the input feature map from the RRDB layer, F_{in} represents the module input for the residual connection to preserve low-level features, $F_{3\times3}$ and $F_{5\times5}$ are the outputs of 3×3 and 5×5 convolutional branches capturing local and global details, respectively, A is the generated spatial attention weight map in the range (0, 1), W_{conv} is the 3×3 convolution kernel for attention computation, and $W_{1\times1}$ is the 1×1 kernel for dimension adjustment. Illustrated in Fig. 1(b), the DetailConv module's schematic layout visualizes the integration of convolutional branches, spatial attention, and residual connections. The 7:3 channel ratio follows the common lightweight design practice adopted in many efficient multi-branch architectures, such as Inception-v3, MobileNetV2/V3, EfficientNet, and the ShuffleNet series. Given the small training set, we intentionally keep this empirically effective default fixed throughout all experiments to ensure training stability.

2.3. MultiScaleConv module

The MultiScaleConv module is integrated into the upsampling stage of RRDBNet to enhance multi-scale feature extraction and adapt to the pressure maps of 26 English alphabets and digits 0–9. The module employs parallel 3×3 and 5×5 convolutional layers with a 7:3 output channel ratio, allocating 70% of channels to the 3×3 convolution and 30% to the 5×5 convolution. The channel count is reduced to half the output channels via a 1×1 convolutional layer, and features are merged via another 1×1 convolutional layer, followed by Leaky ReLU activation with a negative slope of 0.2. The multi-scale design, based on convolutional receptive field theory, optimizes both large-scale and local feature extraction. The module employs nearest-neighbor interpolation for two upsampling operations to capture the global and local patterns of the pressure map, enhancing the structural integrity of the mask. The overall output is expressed by

$$F_{\text{inter}} = \text{ReLU}(W^{(1)}_{1\times1} \cdot (0.7 \cdot F_{3\times3} + 0.3 \cdot F_{5\times5})), \tag{3}$$

$$F_{\text{out}} = \text{Leaky ReLU}(W^{(2)}_{1\times1} \cdot F_{\text{inter}} + \text{Resize}(F_{\text{in}}, \text{scale} = 2)), \tag{4}$$

where $F_{3\times3}$ and $F_{5\times5}$ denote the outputs of the parallel 3×3 and 5×5 convolutional layers for local and global pattern capture, F_{inter} is the intermediate fused feature after initial channel reduction, F_{in} is the input feature for upsampling, F_{out} is the final output feature map, while $W^{(1)}_{1\times1}$ and $W^{(2)}_{1\times1}$ are the first and second 1×1 convolution kernels for fusion and merging. Equation (3) describes multi-scale feature extraction and initial fusion, and Eq. (4) describes feature merging and output. Depicted in Fig. 1(b), the MultiScaleConv module's schematic diagram elucidates the arrangement of parallel convolutional layers and upsampling operations.

3. Results

3.1. Dataset

This study utilizes a 432-pair dataset provided by a partner, comprising low-resolution pressure maps of 128×128 pixels and their corresponding GT masks. The dataset includes 26 English alphabets and digits 0–9, with 12 pairs per category, totaling 432 image pairs. Alphabet and digit shapes are chosen as the validation dataset for their complex contour features to assess the model's mask generation capability, enabling future extension to other shapes to improve generalization. GT masks derived from physical contours provide accurate contour features of alphabets and digits. Specifically, the GT masks are generated using templates of the 26 letters and 10 digits. High-resolution photographs of the templates are captured directly with a camera, then filled and binarized to create masks with sharp physical boundaries. No morphological operations are applied given the templates' inherent precision. Subsequently, data preprocessing involves normalization, scaling pixel values to the (0, 1) interval, and data augmentation, including random horizontal flipping, while maintaining the original resolution.

To ensure experimental reproducibility, three rounds of 5-fold cross-validation are employed. The dataset is divided by category, with the first four folds each containing 7 categories (336 images total) and the fifth fold containing 8 categories (96 images total).[38] Each fold's training set comprises 26 categories (312 images total), and the validation set comprises 3 categories (36 images total). The test set consists of 7 or 8 categories for the corresponding fold.[39] Grouping by category prevents data leakage, with the validation set used for hyperparameter tuning.[40] Three rounds of 5-fold cross-validation enhance result robustness through varied random splits, with random seeds set to 42, 123, and 2023 to ensure experimental reproducibility and data distribution diversity.

3.2. Assessment of indicators

Model performance is evaluated using Intersection over Union (IoU), F1 score, SSIM, and LPIPS, with the similarity of generated masks to GT masks assessed by computing the mean of three rounds of 5-fold cross-validation.

IoU is a widely applied metric in image segmentation tasks, assessing segmentation accuracy through the proportion of region overlap between the predicted mask and the GT mask.[41,42] Values closer to 1 indicate that the predicted mask aligns more closely with the GT mask. IoU is defined by

$$\text{IoU} = \frac{|M \cap G|}{|M \cup G|}, \tag{5}$$

where M is the generated mask, G is the GT mask, and values lie in the range (0, 1). In this study, IoU quantifies the overlap between the generated mask and the GT mask, reflecting the

model's accuracy in recovering complex contours (e.g., alphabet and digit edges).[31]

F1 score, the harmonic mean of precision and recall, comprehensively assesses segmentation task performance, particularly in class imbalance scenarios.[41] In this study, the F1 score balances the precision and recall of generated masks, adapting to complex contour patterns.[43] The F1 score is defined by Eq. (6), with precision and recall defined by Eqs. (7) and (8), respectively:

$$\text{F1 score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{6}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{7}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{8}$$

where TP, FP, and FN denote true positives, false positives, and false negatives. In application scenarios, the F1 score effectively evaluates the model's capability to accurately capture edge details in distorted pressure maps, ensuring neither over-segmentation nor omission of critical contour regions occurs.[44]

SSIM, based on human visual perception, assesses image similarity by comparing luminance, contrast, and structural features, making it suitable for evaluating the structural consistency of masks.[45] In this study, SSIM emphasizes the spatial structural integrity of masks. SSIM is defined by

$$\text{SSIM}(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \tag{9}$$

where μ_x and μ_y denote the means of images x and y, σ_x and σ_y denote the standard deviations, σ_{xy} denotes the covariance, and c_1 and c_2 are stability constants, with SSIM values in the range (0, 1). In this study, SSIM assesses the spatial structural consistency of generated masks relative to the GT mask.[46]

The LPIPS metric employs the VGG19 neural network to extract features and evaluate perceptual similarity between the generated mask and the GT mask, and it is widely applied in image generation tasks.[47] LPIPS quantifies the visual quality of the mask, with smaller values indicating higher perceptual similarity. In shape recovery scenarios, LPIPS evaluates the naturalness of the generated mask in terms of human visual perception and its alignment with the GT mask.[46,47]

3.3. Training strategies

Training is conducted using the PyTorch framework on an NVIDIA RTX 4090 GPU with 24 GB of memory, running on Windows Server 2022. The input is a 128×128 pixel three-channel pressure map, and the target output is a 128×128 pixel mask. Training parameters include a generator learning rate of 1×10^{-4}, a discriminator learning rate of 1×10^{-5}, and Adam optimizer parameters $\beta_1 = 0.9$ and $\beta_2 = 0.99$. The batch size is 8, with training conducted for 200 epochs, totaling 7800 iterations. The learning rate difference balances feature learning and training stability, suitable for a small-sample dataset of 432 pairs. Figure 2 presents the performance metrics under various learning rate combinations for the generator and discriminator, with Fig. 2(a) highlighting IoU, F1 score, SSIM, and LPIPS values across combinations such as $(1 \times 10^{-3}, 1 \times 10^{-5})$ to $(1 \times 10^{-4}, 1 \times 10^{-5})$, supported by error bars indicating standard deviation, while Fig. 2(b) illustrates the metric evolution over training epochs from 50 to 300, with shaded areas reflecting standard deviation ranges. These visualizations confirm that the four combinations exhibit highly similar performance (differences within one standard deviation). Given this relatively flat performance landscape, no exhaustive search over intermediate values is performed, and the combination yielding the numerically best results (generator: 1×10^{-4}, discriminator: 1×10^{-5}) is selected to optimize convergence and ensure high mask generation accuracy for e-skin pressure maps.

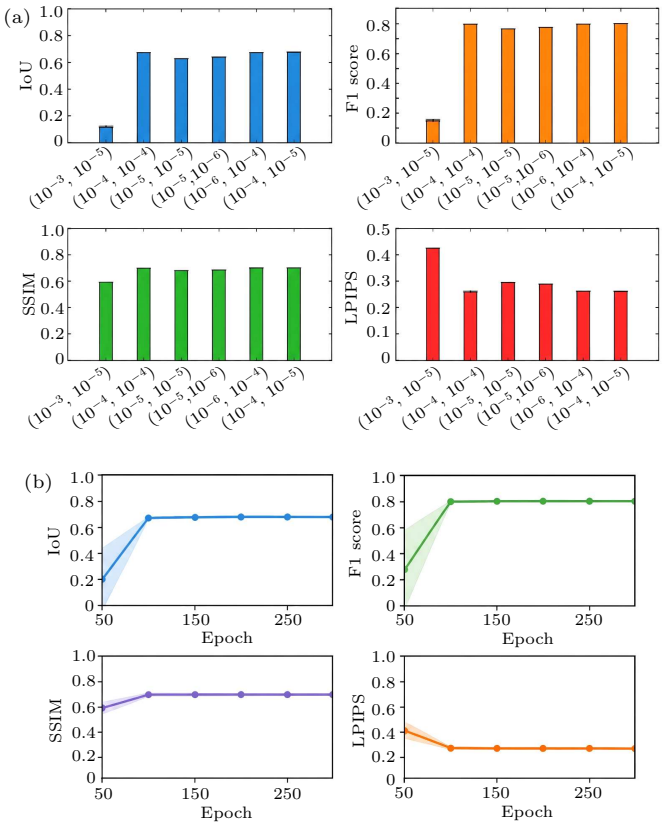


Fig. 2. (a) Performance metrics under different learning rate combinations. (b) Performance evolution across 50–300 epochs on test set, with shaded areas for standard deviation.

The loss function combines L_1 loss, perceptual loss, and adversarial loss. The perceptual loss is derived from the VGG19 network, using layers conv1_2 to conv5_4 with weights of 0.1, 0.1, 1, 1, 1, and 1, while the adversarial loss, based on a vanilla GAN, has a weight of 0.1. The L_1 loss is

defined by

$$L_1 = \frac{1}{N}\sum_i |y_i - \hat{y}_i|. \tag{10}$$

The perceptual loss is defined by

$$L_{\text{perceptual}} = \sum_l w_l \parallel \phi_l(y) - \phi_l(\hat{y}) \parallel_2^2 . \tag{11}$$

The adversarial loss is defined by

$$L_{\text{adv}} = E[\log D(y)] + E[\log(1 - D(\hat{y}))]. \tag{12}$$

The combined loss is defined by

$$L_{\text{total}} = \lambda_1 L_1 + \lambda_2 L_{\text{perceptual}} + \lambda_3 L_{\text{adv}}. \tag{13}$$

Here, y denotes the GT mask, $\hat{y}$ denotes the generated mask, N is the total number of pixels, ϕ_l represents VGG19 layer features, w_l are weights (0.1, 0.1, 1, 1, 1, 1), E represents the expectation operator, D is the discriminator, and $\lambda_1 = 1.0$, $\lambda_2 = 1.0$, $\lambda_3 = 0.1$. The loss weights $\lambda_1 = 1.0$, $\lambda_2 = 1.0$, and $\lambda_3 = 0.1$ follow the official Real-ESRGAN repository and are kept fixed across all experiments. Figure 3(a) depicts the training loss trends of Real-ESRGAN-MSDC, illustrating the dynamic evolution of L_1, perceptual, and adversarial losses across training iterations, which collectively indicate gradual convergence and validate the effectiveness of the multi-loss strategy in optimizing e-skin mask generation. The discriminator is not used during the pre-training phase, with the discriminator updated once per iteration thereafter. ESRGAN, Real-ESRGAN, and Real-ESRGAN-MSDC leverage the official Real-ESRGAN pre-trained model to accelerate convergence for small-sample datasets, adapted to meet the demands of complex feature extraction in the RRDB architecture. As shown in Fig. 3(b), the training performance of Real-ESRGAN-MSDC features a side-by-side comparison of generated mask images with GT counterparts across iterations, based on 5-fold cross-validation, highlighting the progressive improvement in mask quality and structural fidelity for e-skin pressure maps.

To ensure a practical and meaningful comparison under our small-sample conditions, each model family is initialized following its established best practices. For the image translation models (Pix2pix and CycleGAN), domain-specific pretrained weights are unavailable, and ImageNet initialization would impose a generic natural-image prior that is misaligned with our specialized task. Therefore, they are trained from scratch. In contrast, the real-world super-resolution models (ESRGAN, Real-ESRGAN, and Real-ESRGAN-MSDC) rely on pretraining that captures the complex degradation–restoration mapping essential for their performance. Hence, they are initialized from the same official Real-ESRGAN ×4 pretrained checkpoint, namely the only publicly released foundation for this architecture, which provides the necessary prior for stable optimization and effective generalization on limited data. This differentiated approach allows each model to perform under its most representative and effective configuration, enabling a fair comparison of their achievable performance rather than enforcing an artificial parity in initialization.

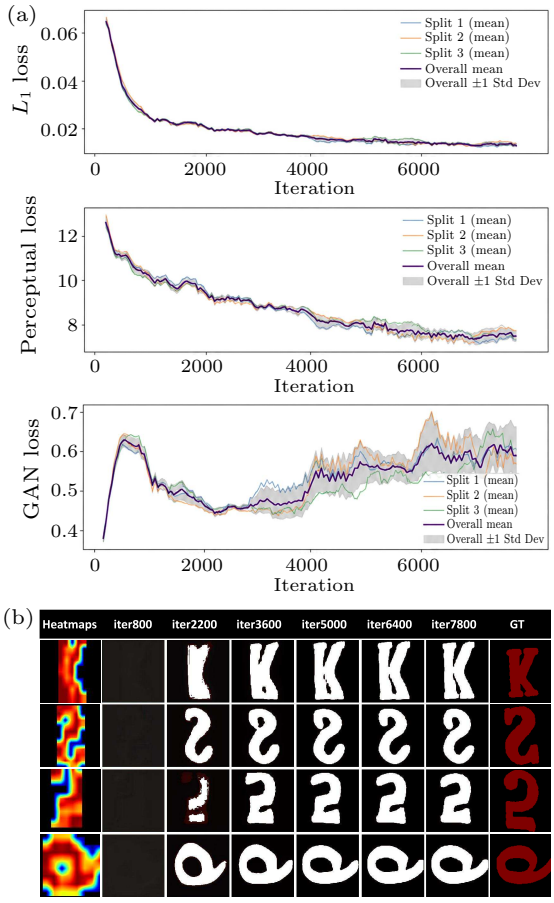


Fig. 3. (a) Training loss trends (L_1, perceptual, adversarial) of Real-ESRGAN-MSDC over iterations. (b) Visualization of generated mask evolution and ground truth comparison across training iterations.

3.4. Comparison with other methods

Comparison experiments evaluate TCVMG, based on classical image processing techniques, alongside Pix2pix, CycleGAN, ESRGAN, Real-ESRGAN, and Real-ESRGAN-MSDC in generating masks for e-skin pressure maps, using three rounds of 5-fold cross-validation.[34–36,48] The experimental design utilizes 432 image pairs from the aforementioned dataset. Three rounds of 5-fold cross-validation ensure that the training set (312 images), validation set (36 images), and test set (48 or 96 images) are uniformly distributed by category, with random seeds set to 42, 123, and 2023 to enhance result robustness and reproducibility. TCVMG generates masks for e-skin pressure maps by applying JET color mapping to convert RGB pressure maps to grayscale images. A 7×7 kernel Gaussian blur (standard deviation = 3) smooths noise, followed by binarization with a threshold of 128 and

a single morphological expansion and erosion operation to optimize mask contours, aiming to separate high-pressure and low-pressure regions. Pix2pix, CycleGAN, ESRGAN, and Real-ESRGAN leverage the U-Net generator, cyclic consistency loss, and super-resolution optimization frameworks. Real-ESRGAN-MSDC further integrates MultiScaleConv and DetailConv modules. Performance is evaluated using IoU, F1 score, SSIM, and LPIPS metrics, with the mean and standard deviation of each metric computed from the test set.[49] Figure 4(a) compares IoU, F1 score, SSIM, and LPIPS metrics across TCVMG, Pix2pix, CycleGAN, ESRGAN, Real-ESRGAN, and Real-ESRGAN-MSDC, based on 5-fold cross-validation, underscoring our model's superior structural similarity and edge fidelity in generating e-skin pressure map masks.

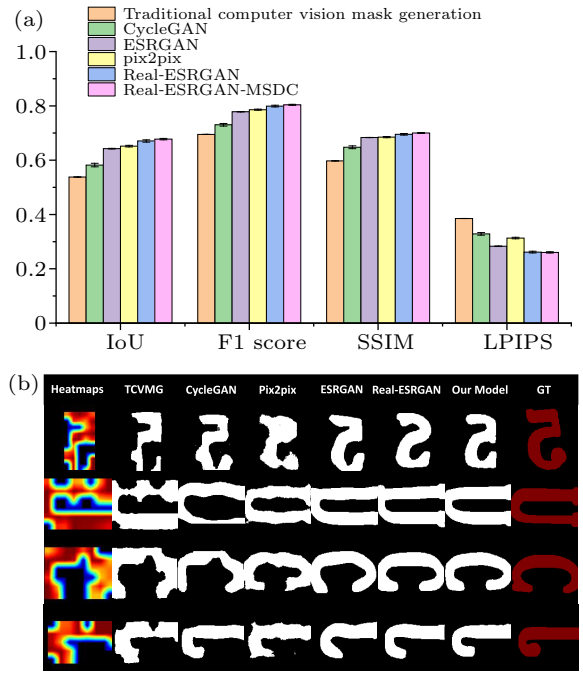


Fig. 4. (a) Comparative performance of Real-ESRGAN-MSDC across models via 5-fold cross-validation. (b) Generated mask images by TCVMG, CycleGAN, ESRGAN, pix2pix, Real-ESRGAN, and Real-ESRGAN-MSDC after 200 epochs compared with ground truth.

As shown in Table 1, Real ESRGAN MSDC performs well on all evaluation metrics, with an IoU of 0.6775 ± 0.0026, an F1 score of 0.8039 ± 0.0018, an SSIM of 0.7003 ± 0.0016, and an LPIPS of 0.2605 ± 0.0023. Quantitative analysis shows that, compared to the second-best-performing Real-ESRGAN, our model achieves improvements of approximately 1.0%, 0.6%, and 0.7% in IoU, F1 score, and SSIM, respectively, while reducing LPIPS by about 0.4%. Compared to Pix2pix and ESRGAN, the improvements are more significant, with IoU increasing by approximately 3.9% and 5.4%, respectively. In contrast to TCVMG, the performance improvements are substantial, with IoU and F1 score leading by more than 25% and 15%, respectively. These results verify the synergistic effect of MultiScaleConv and DetailConv modules in enhancing structural similarity and edge fidelity. In contrast, TCVMG exhibits the poorest performance, with an IoU of 0.5380 ± 0.0001, an F1 score of 0.6921 ± 0.0002, an SSIM of 0.5812 ± 0.0003, and an LPIPS of 0.4123 ± 0.0004. These results indicate that traditional image processing methods struggle to effectively capture high-frequency contour details in complex pressure patterns and small-sample scenarios, particularly in recovering alphabet and digit edges. Pix2pix outperforms ESRGAN, achieving an F1 score of 0.7812 ± 0.0021 and an SSIM of 0.6723 ± 0.0019 compared to ESRGAN's 0.7654 ± 0.0024 and 0.6598 ± 0.0020. However, its higher LPIPS value (0.2934 ± 0.0027) indicates reduced perceptual quality, likely due to its high dependence on paired data and limited capability to generate complex textures. CycleGAN exhibits greater metric variability due to its unpaired training nature (IoU: 0.6123 ± 0.0065, F1 score: 0.7345 ± 0.0058), with limited training stability restricting its application in high-precision mask generation. Real-ESRGAN achieves performance close to Real-ESRGAN-MSDC (IoU: 0.6643 ± 0.0022, F1 score: 0.7921 ± 0.0019), but its lack of MultiScaleConv and DetailConv modules results in slightly inferior edge details and global structure. The standard deviation range (0.0001–0.0065) indicates high consistency in experimental results, reflecting the stability of data splitting and training strategies. Significance analysis (Table 2, $p < 0.05$) confirms Real-ESRGAN-MSDC's statistical superiority across all four metrics, particularly in edge prediction for complex contours. As shown in Fig. 4(b), the comparison of generated mask images by TCVMG, CycleGAN, ESRGAN, Pix2pix, Real-ESRGAN, and Real-ESRGAN-MSDC after 200 training epochs, presented side-by-side with their GT counterparts, demonstrates the more precise boundaries and enhanced alignment of our model's masks, confirming its robustness and accuracy for e-skin applications in small-sample settings.

Table 1. Comparative performance of models on e-skin heatmap-to-mask conversion. The best and second-best results in each column are highlighted in bold and underlined, respectively.

Methods	IoU↑	F1 score↑	SSIM↑	LPIPS↓
TCVMG	0.5380 ± 0.0001	0.6950 ± 0.0001	0.5976 ± 0.0003	0.3853 ± 0.0002
CycleGAN	0.5821 ± 0.0065	0.7304 ± 0.0048	0.6476 ± 0.0051	0.3288 ± 0.0048
ESRGAN	0.6426 ± 0.0010	0.7784 ± 0.0007	0.6834 ± 0.0004	0.2832 ± 0.0009
Pix2pix	0.6518 ± 0.0030	0.7860 ± 0.0020	0.6851 ± 0.0022	0.3133 ± 0.0028
Real-ESRGAN	$\underline{0.6708 \pm 0.0043}$	$\underline{0.7992 \pm 0.0032}$	$\underline{0.6954 \pm 0.0030}$	$\underline{0.2615 \pm 0.0031}$
Real-ESRGAN-MSDC	$\mathbf{0.6775 \pm 0.0026}$	$\mathbf{0.8039 \pm 0.0018}$	$\mathbf{0.7003 \pm 0.0016}$	$\mathbf{0.2605 \pm 0.0023}$

Table 2. Significance analysis tables present p-values from t-tests, evaluating statistical differences in the four metrics between Real-ESRGAN-MSDC and other models.

Model Comparison	IoU (t/p-value)	F1 score (t/p-value)	SSIM (t/p-value)	LPIPS (t/p-value)
Real-ESRGAN-MSDC vs. TCVMG	$25.8875/3.18\times10^{-13*}$	$23.8708/9.65\times10^{-13*}$	$23.7801/1.02\times10^{-12*}$	$-24.4538/6.94\times10^{-13*}$
Real-ESRGAN-MSDC vs. CycleGAN	$17.3787/7.16\times10^{-11*}$	$16.6987/1.22\times10^{-10*}$	$10.2753/6.67\times10^{-8*}$	$-13.8689/1.43\times10^{-9*}$
Real-ESRGAN-MSDC vs. ESRGAN	$17.2064/8.18\times10^{-11*}$	$16.7088/1.21\times10^{-10*}$	$11.2988/2.02\times10^{-8*}$	$-12.6685/4.65\times10^{-9*}$
Real-ESRGAN-MSDC vs. Pix2pix	$8.9557/3.60\times10^{-7*}$	$8.3135/8.73\times10^{-7*}$	$4.5845/0.0004^{*}$	$-14.3086/9.49\times10^{-10*}$
Real-ESRGAN-MSDC vs. Real-ESRGAN	$2.4868/0.0261^{*}$	$2.4635/0.0273^{*}$	$3.6299/0.0027^{*}$	$-0.6692/0.5143$

In summary, Real-ESRGAN-MSDC substantially enhances the shape recovery of e-skin pressure maps by integrating multi-scale feature extraction and detail enhancement mechanisms, particularly excelling in handling complex contour patterns. The significant improvements in key performance metrics compared to baseline models quantitatively demonstrate the effectiveness of this architecture. The poor performance of TCVMG highlights the limitations of traditional methods in handling nonlinear pressure patterns. Similarly, Pix2pix and CycleGAN are constrained by data requirements and training stability, while ESRGAN and Real-ESRGAN, although optimized for super-resolution tasks, require further improvement in small-sample optimization and edge detail restoration. The superiority of Real-ESRGAN-MSDC establishes a foundation for high-precision applications in robotic tactile sensing and health monitoring, but its generalization ability requires further validation in diverse datasets and real-time scenarios.

3.5. Ablation experiments

The ablation experiment evaluates the contributions of the MultiScaleConv and DetailConv modules to Real-ESRGAN-MSDC's performance. It compares the baseline Real-ESRGAN, variants with MultiScaleConv removed, variants with DetailConv removed, and the full Real-ESRGAN-MSDC, using three rounds of 5-fold cross-validation consistent with the comparison experiments. During the experiments, the baseline Real-ESRGAN serves as the reference. The experiment compares variants by removing either the MultiScaleConv module (retaining only DetailConv) or the DetailConv module (retaining only MultiScaleConv), alongside the full Real-ESRGAN-MSDC model. As illustrated in Fig. 5, the ablation study plot visualizes IoU, F1 score, SSIM, and LPIPS metrics for Real-ESRGAN, models lacking MultiScaleConv or DetailConv, and the complete Real-ESRGAN-MSDC, based on 5-fold cross-validation, indicating the critical synergy of these modules in boosting mask generation accuracy for e-skin pressure maps.

Experiments demonstrate that Real-ESRGAN-MSDC achieves the best performance across all metrics, with an IoU of 0.6775 ± 0.0026, an F1 score of 0.8039 ± 0.0018, an SSIM of 0.7003 ± 0.0016, and an LPIPS of 0.2605 ± 0.0023. These results verify the synergistic effect of MultiScaleConv and DetailConv modules. Removing MultiScaleConv significantly reduced model performance, with IoU dropping to 0.6521 ± 0.0032, F1 score to 0.7845 ± 0.0025, SSIM to 0.6856 ± 0.0021, and LPIPS rising to 0.2754 ± 0.0024. This indicates that multi-scale feature extraction is critical for capturing global pressure patterns and maintaining the structural integrity of complex shapes, particularly in closed contour recovery for alphabets and digits. Removing DetailConv reduced IoU to 0.6608 ± 0.0029, F1 score to 0.7902 ± 0.0023, SSIM to 0.6894 ± 0.0019, and increased LPIPS to 0.2754 ± 0.0026, indicating that reduced localized edge details significantly impact the sharpness of high-contrast regions, such as alphabet 'B' and digit '8' edges. The baseline Real-ESRGAN (IoU: 0.6643 ± 0.0022, F1 score: 0.7921 ± 0.0019, SSIM: 0.6598 ± 0.0020, LPIPS: 0.2934 ± 0.0027) outperforms variants with either module removed but underperforms the full model. This demonstrates that the joint application of MultiScaleConv and DetailConv significantly enhances the co-optimization of global structure and local details.

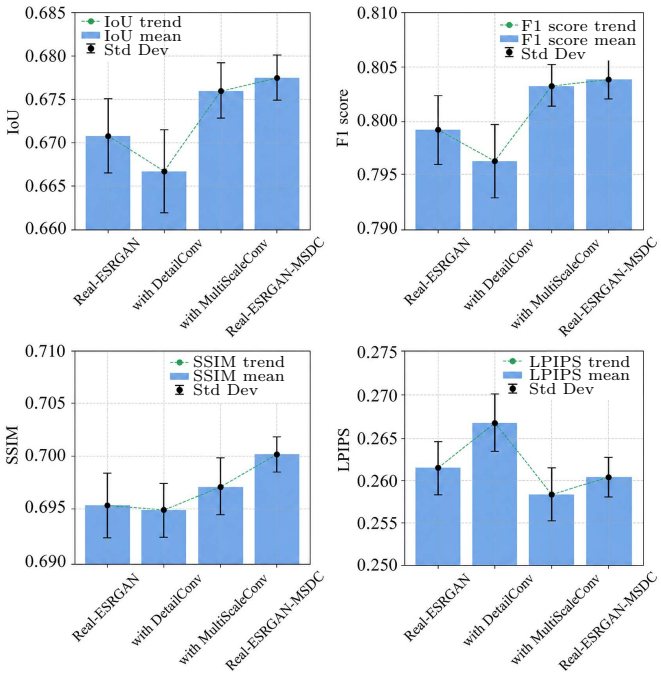


Fig. 5. Ablation study of IoU, F1 score, SSIM, and LPIPS metrics for Real-ESRGAN, models without MultiScaleConv or DetailConv, and Real-ESRGAN-MSDC via 5-fold cross-validation.

In summary, the ablation results underscore that Real-ESRGAN-MSDC substantially outperforms the baseline and single-module variants across all evaluated metrics, confirming the essential and complementary roles of MultiScaleConv

in preserving global structures and DetailConv in refining local edges. This provides strong evidence for the design robustness of Real-ESRGAN-MSDC, while suggesting that the inter-module feature fusion strategy could be further optimized to enhance performance.

3.6. Robustness experiments

Robustness experiments systematically evaluate the performance stability of Real-ESRGAN-MSDC under various noise conditions. These conditions are tested by adding mild Gaussian noise (mean = 0, variance = 0.01) and high-density salt-and-pepper noise (density = 0.1) to the low-resolution pressure maps of the test set, using three rounds of 5-fold cross-validation with random seeds set to 42, 123, and 2023 to ensure statistical reliability. Mild Gaussian noise simulates natural disturbances in sensor acquisition, while high-density salt-and-pepper noise simulates data distortion in extreme environments. For each noise condition, the mean and variation of IoU, F1 score, SSIM, and LPIPS are computed, with results summarized in Table 3. Depicted in Fig. 6, the noise robustness table outlines the mean and standard deviation of IoU, F1 score, SSIM, and LPIPS metrics under Gaussian noise (variances 0.01 and 0.1) and salt-and-pepper noise (densities 0.05 and 0.2), providing a comprehensive evaluation of Real-ESRGAN-MSDC's tolerance to diverse e-skin sensor disturbances. Under mild Gaussian noise conditions, Real-ESRGAN-MSDC demonstrates robust performance, with IoU decreasing by only 0.24% to 0.75% (from 0.6775 to 0.6734–0.6748) and F1 score decreasing by 0.21% to 0.46% (from 0.8039 to 0.8002–0.8015). Similarly, SSIM decreases by 0.24% to 0.26% (from 0.7003 to 0.6987–0.6990), and LPIPS increases by 0.27% to 0.34% (from 0.2605 to 0.2613–0.2620). These minimal changes demonstrate that Real-ESRGAN-MSDC exhibits strong adaptability to common noise in sensor acquisition. The MultiScaleConv and DetailConv modules effectively mitigate the impact of mild noise on mask contour recovery through multi-scale feature extraction and edge detail enhancement, maintaining high structural similarity and visual quality. Under high-density salt-and-pepper noise conditions, performance degrades significantly, with IoU decreasing by 10.4% (from 0.6775 to 0.6071) and F1 score by 9.8% (from 0.8039 to 0.7250). Similarly, SSIM decreases by 8.6% (from 0.7003 to 0.6402), and LPIPS increases by 12.3% (from 0.2605 to 0.2925). This result highlights the model's sensitivity to discrete noise, likely due to the severe disruption of local pixels in the pressure map by salt-and-pepper noise. This leads to the loss of high-frequency information during feature extraction, particularly impacting fine edge detail recovery for alphabets and digits. More specifically, the substantial performance drop under high-density salt-and-pepper noise can be attributed to the discrete nature of such outliers. They severely disrupt local pixel continuity and exceed the corrective capacity of the purely convolutional receptive fields, and limit spatial attention mechanism employed in the current architecture. This suggests a fundamental limitation in the existing design when confronting extreme discrete noise perturbations.

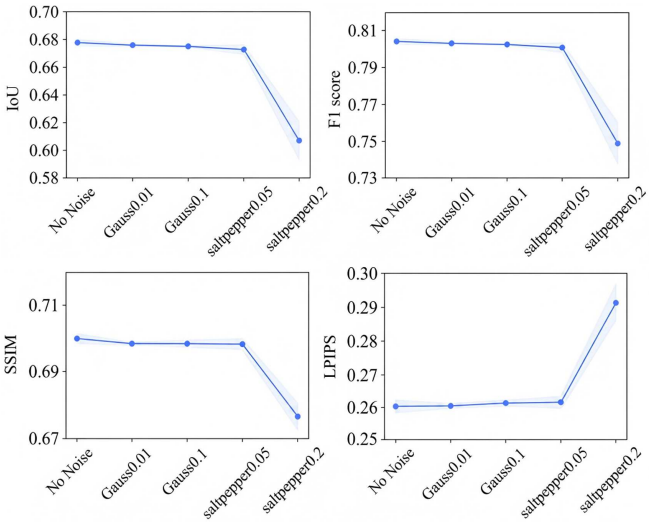


Fig. 6. Noise robustness table presents the mean and standard deviation of the four metrics under Gaussian (variances 0.01, 0.1) and saltpepper (densities 0.05, 0.2) noise conditions, evaluating model tolerance.

Table 3. Noise robustness analysis of Real-ESRGAN-MSDC. The best result in each column is marked in bold.

Condition	IoU (Mean ± SD)	F1 score (Mean ± SD)	SSIM (Mean ± SD)	LPIPS (Mean ± SD)
No noise	**0.6775 ± 0.0026**	**0.8039 ± 0.0018**	**0.7003 ± 0.0016**	0.2605 ± 0.0023
Saltpepper (0.05)	0.6724 ± 0.0041	0.8002 ± 0.0029	0.6985 ± 0.0017	0.2614 ± 0.0020
Saltpepper (0.2)	0.6068 ± 0.0134	0.7491 ± 0.0111	0.6766 ± 0.0044	0.2912 ± 0.0056
Gaussian (0.01)	0.6759 ± 0.0013	0.8027 ± 0.0010	0.6985 ± 0.0006	**0.2604 ± 0.0011**
Gaussian (0.1)	0.6750 ± 0.0006	0.8022 ± 0.0002	0.6986 ± 0.0012	0.2612 ± 0.0009

In summary, Real-ESRGAN-MSDC maintains high generation quality under mild noise conditions, making it suitable for common interference in real sensor acquisition scenarios. However, high-density salt-and-pepper noise significantly degrades model performance. This suggests the need for future improvements, such as introducing a preprocessing module for noise suppression or enhancing the model's anti-noise training strategy by incorporating noisy samples via data augmentation or optimizing the loss function to improve robustness to discrete noise. These findings provide guidance for the model's practical application in complex environments, while indicating directions for optimization.

4. Discussion and conclusion

In this study, we address the challenge of generating high-fidelity masks from e-skin pressure maps by proposing Real-ESRGAN-MSDC, a novel deep learning framework that enhances contour prediction for 26 English alphabets and digits 0–9. By integrating MultiScaleConv and DetailConv modules into the RRDB architecture of Real-ESRGAN, our model achieves superior structural similarity and edge fidelity compared to baseline methods like TCVMG, Pix2pix, CycleGAN, and ESRGAN. Evaluated on a 432-pair dataset using three rounds of 5-fold cross-validation, Real-ESRGAN-MSDC demonstrates robust performance, particularly in small-sample scenarios, and supports applications in robotic tactile perception and health monitoring. Additionally, robustness tests reveal minimal performance degradation under mild Gaussian noise, underscoring the model's suitability for real-world e-skin applications.

The superior performance of Real-ESRGAN-MSDC stems from several key factors. First, the integration of MultiScaleConv and DetailConv modules enables comprehensive feature extraction. The MultiScaleConv module captures both global and local patterns in pressure maps through parallel 3×3 and 5×5 convolutional layers, while the DetailConv module enhances edge details via a spatial attention mechanism, ensuring precise contour recovery for complex shapes. Second, leveraging the pre-trained Real-ESRGAN model accelerates convergence on small-sample datasets, addressing the challenge of limited training data in e-skin applications. Third, the adversarial training framework, combining L_1, perceptual, and adversarial losses, optimizes mask quality by balancing structural fidelity and visual realism. This multi-loss approach ensures that generated masks closely align with GT, as evidenced by high IoU (0.6775 ± 0.0026) and F1 score (0.8039 ± 0.0018) metrics.

Despite its advancements, Real-ESRGAN-MSDC faces several limitations. First, the dataset of 432 image pairs restricts generalization across diverse pressure map patterns, potentially limiting applicability to varied e-skin scenarios. Second, the fixed learning rate of 1×10^{-4} may hinder optimal convergence, particularly for larger or more complex datasets. Third, the model's computational complexity poses challenges for real-time applications, such as dynamic robotic control. Finally, its adaptability to diverse sensor types, such as capacitive or piezoresistive sensors, remains unverified, which could affect performance in heterogeneous e-skin systems. These limitations are not unique to our model. For instance, Wang and colleagues in 2025 noted similar challenges in dataset scarcity for deep learning in digital breast tomosynthesis.[50] In summary, Real-ESRGAN-MSDC is an effective tool for high-precision e-skin mask generation, offering significant potential for robotic and medical applications.

To address these limitations, future work can explore several strategies. Enriching the dataset with diverse pressure map patterns, as suggested by Cao and colleagues in 2024, could enhance generalization.[51] Incorporating multi-sensor data and arbitrary shapes, as explored by Huang and colleagues in 2023, may improve cross-domain robustness.[52] Adopting adaptive learning rates or transfer learning could optimize training efficiency, while model pruning, as proposed by Guo and colleagues in 2022, could reduce computational complexity for real-time applications.[53] Additionally, extending the model to applications like gesture recognition and tactile feedback optimization could broaden its impact. Yang and colleagues in 2025 demonstrated that multi-task models with feature fusion enhance small-sample learning, providing a promising direction for optimizing Real-ESRGAN-MSDC.[54–56] Ultimately, these advancements will strengthen the model's precision and applicability, making it a valuable tool for e-skin technology in human-computer interaction, health monitoring, and tactile sensing.

Data availability

The source code and datasets are available online at https://github.com/zhaoqi106/Real-ESRGAN-MSDC.

Acknowledgements

This study is supported by the Ministry of Science and Technology of China (STI2030-Major Projects2021ZD0201900), the National Natural Science Foundation of China (Grant No. U24A2014), Wenzhou Science and Technology Bureau's Project (Grant No. ZY2024002), Science and Technology Plan Project of Liaoning Province (Grant No. 2025-MSLH-351), Fundamental Research Funds for the Liaoning Universities (Grant No. LJ212410146026), and Ningbo Scientific and Technological Innovation 2035 Major Project (Grant Nos. 2024Z148 and 2024Z234).

References

[1] Leogrande E, Filosa M, Ballanti S, De Cicco L, Mazzoleni S, Ackerley R, Oddo C M and Dell'Olio F 2025 *Sens. Actuators Rep.* **9** 100312
[2] Yadav N, Zhang Q Q, Qi D F, Yadav A and Zheng H Y 2025 *Nanoscale Adv.* **7** 4803
[3] Zhang H Y, Hong J L, Zhu J X, *et al*. 2025 *Matter* **8** 102136
[4] Jassim H S, Akhter Y Aalwahab D Z and Neamah H A 2025 *Biosens. Bioelectron. X* **26** 100669
[5] Liu Y M, Yiu C, Song Z, Huang Y, Yao K M, Wong T, Zhou J K, Zhao L, Huang X C, Nejad S K, Wu M G, Li D F, He J H, Guo X, Yu J S, Feng X, Xie Z Q and Yu X G 2022 *Sci. Adv.* **8** eabl6700
[6] Fu X M, Cheng W, Wan G X, Yang Z J and Tee B C K 2024 *Chem. Rev.* **124** 9899
[7] Duan Y H, He S X, Wu J, Su B L and Wang Y S 2022 *Nanomaterials* **12** 2495
[8] Xu F L, Li X Y, Shi Y, Li L H, Wang W, He L and Liu R P 2018 *Micromachines* **9** 580

[9] Zhu P Y, Li Z H, Pang J B, He P and Zhang S Y 2024 *Soft Sci.* **4** 17

[10] Chen J J, Wang L, Xu X O, Liu G M, Liu H Y, Qiao Y X and Chen J L 2022 *Gels* **8** 356

[11] Zarei M, Lee G, Lee S G and Cho K 2023 *Adv. Mater.* **35** 2203193

[12] Liu C Y, Peng K X, Wu Y L and Fu F F 2023 *Smart Med.* **2** e20230024

[13] Zhang L, Xing S C, Yin H F, Weisbecker H, Tran H T, Guo Z H, Han T H, Wang Y H, Liu Y H, Wu Y Z, Xie W R, Huang C Q, Luo W, Demaesschalck M, McKinney C, Hankley S, Huang A, Brusseau B, Messenger J, Zou Y C and Bai W B 2024 *Nat. Commun.* **15** 4777

[14] Sankar S, Cheng W Y, Zhang J H, Slepyan A, Iskarous M M, Greene R J, DeBrabander R, Chen J J, Gupta A and Thakor N V 2025 *Sci. Adv.* **11** eadr9300

[15] Patil C S, Ghode S B, Kim J, Kamble G U, Kundale S S, Mannan A, Ko Y, Noman M, Saqib Q M, Patil S R, Bae S Y, Kim J H, Park J H and Bae J 2025 *Mater. Horiz.* **12** 2045

[16] Tu J Q, Wang M, Li W L, Su J T, Li Y Z, Lv Z S, Li H C, Feng X and Chen X D 2023 *Soft Sci.* **3** 25

[17] Wei D Y, Guo J J, Qiu Y Q, Liu S Y, Mao J Y, Liu Y T, Chen Z B, Wu H and Yin Z Q 2022 *Natl. Sci. Rev.* **9** nwac227

[18] Wang J X, Wei X Y, Shi J L, Bai N N, Wan X, Li B, Chen Y C, Jiang Z D and Guo C F 2024 *Nat. Commun.* **15** 7094

[19] Wang Z Y, Li Q, Wang Z R and Yan H 2015 *Sensors* **15** 17748

[20] Zheng Q Q, Dai X Y, Wu Y H, Liang Q H, Wu Y P, Yang J K, Dong B Q, Gao G J, Qin Q and Huang L B 2023 *BMEMat* **1** e12008

[21] Vitolo P, Liguori R, Di Benedetto L, Rubino A, Pau D and Licciardo G D 2025 *Microprocess. Microsyst.* **117** 105183

[22] Huang Y, Hu S X, Li Y, Wang R, Yang Y C, Zhu W and Deng Y 2025 *npj Flex. Electron.* **9** 41

[23] Duan X, Taurand S and Soleimani M 2019 *Sci. Rep.* **9** 8831

[24] Li X Y, Wu Z R, Chen Z D and Shi J D 2025 *Soft Sci.* **5** 16

[25] Sun J F, Zhang C Y, Yang C X, Ren Y H, Ma B H, Yuan W Z and Ye T 2025 *Soft Sci.* **5** 18

[26] Zhu L F, Wang Y C, Mei D Q and Jiang C P 2020 *Micromachines* **11** 770

[27] Wang Y, Guo J H, Cao X Y and Zhao Y J 2024 *Smart Med.* **3** e20230023

[28] Lei Z Y, Xu W T and Zhang G G 2023 *Smart Med.* **2** e20220026

[29] Adil A 2021 *J. Ilmu Komput.* **14** 20

[30] Parsania P and Virparia P V 2014 *Int. J. Innov. Res. Comput. Commun. Eng.* **2** 7409

[31] Ronneberger O, Fischer P and Brox T 2015 *Medical Image Computing and Computer-Assisted Intervention – MICCAI* 2015, October 5–9, 2015, Munich, Germany, pp. 234–241

[32] Wu W, Yang X, Yao C G, Liu O, Zhao Q and Shuai J W 2025 *Research* **8** 0869

[33] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y 2020 *Commun. ACM* **63** 139

[34] Isola P, Zhu J Y, Zhou T H and Efros A A 2017 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, July 21–26, 2017, Honolulu, HI, USA, p. 1125

[35] Zhu J Y, Park T, Isola P and Efros A A 2017 *Proceedings of the IEEE International Conference on Computer Vision*, October 22–29, 2017, Venice, Italy, p. 2223

[36] Wang X T, Yu K, Wu S X, Gu J J, Liu Y H, Dong C, Qiao Y and Loy C C 2018 *Proceedings of the European Conference on Computer Vision Workshops*, September 8–14, 2018, Munich, Germany, pp. 1–16

[37] Guo Y C, Sun X D, Li L L, Shi Y, Cheng W and Pan L J 2025 *Sensors* **25** 1615

[38] Yates L A, Aandahl Z, Richards S A and Brook B W 2023 *Ecol. Monogr.* **93** e1557

[39] Bayle P, Bayle A, Janson L and Mackey L 2020 *Advances in Neural Information Processing Systems* **33** 16339

[40] Varma S and Simon R 2006 *BMC Bioinform.* **7** 91

[41] Cho Y J 2024 *Pattern Recognit. Lett.* **185** 101

[42] Rahman M A and Wang Y 2016 *Proceedings of the International Symposium on Visual Computing*, December 12–14, 2016, Las Vegas, NV, USA, p. 234

[43] Wang G N, Li Z, Weng G R and Chen Y Y 2025 *Intell. Robot.* **5** 143

[44] Liu T J, Wang H W, Christian M, Chang C W, Lai F P and Tai H C 2023 *Sci. Rep.* **13** 680

[45] Wang Z, Bovik A C, Sheikh H R and Simoncelli E P 2004 *IEEE Trans. Image Process.* **13** 600

[46] Dohmen M, Klemens M, Baltruschat I, Truong T and Lenga M 2025 *Sci. Rep.* **15** 3853

[47] Zhang R, Isola P, Efros A A, Shechtman E and Wang O 2018 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 18–22, 2018, Salt Lake City, UT, USA, p. 586

[48] Wang X T, Xie L B, Dong C and Shan Y 2021 *Proceedings of the IEEE/CVF International Conference on Computer Vision*, October 11–17, 2021, Montreal, QC, Canada, p. 1905

[49] Steinhauser C, Reis P, Padusinski H, Langner J and Sax E 2025 *IEEE Intelligent Vehicles Symposium (IV), Cluj-Napoca, Romania* 1564

[50] Wang R Y, Chen F X, Chen H M, Lin C X, Shuai J C, Wu Y T, Ma L X, Hu X Q, Wu M, Wang J, Zhao Q, Shuai J W and Pan J Y 2025 *MedComm* **6** e70247

[51] Cao G Q, Jiang J Q, Bollegala D, Li M and Luo S 2024 *Robot. Auton. Syst.* **176** 104688

[52] Huang Y, Zhou J K, Ke P C, *et al.* 2023 *Nat. Electron.* **6** 1020

[53] Guo J H, Luo Z Q, Wang F Y, Gu H C and Li M L 2022 *Smart Med.* **1** e20220003

[54] Yang X, Sun J Q, Jin B Y, Lu Y E, Cheng J Y, Jiang J J, Zhao Q and Shuai J W 2025 *J. Adv. Res.* **68** 477

[55] Yang X, Wang Y, Lin Y, Zhang M X, Liu O, Shuai J W and Zhao Q 2025 *Adv. Sci.* **12** 2412987

[56] Hu J P, Yang X, Yao C G, Zhang M X, Shen S H, Na L J and Zhao Q 2026 *Interdiscip. Sci. Comput. Life Sci.* **18** 471