frontiers | Frontiers in Physics

TYPE Original Research
PUBLISHED 14 February 2024
DOI 10.3389/fphy.2024.1364372

OPEN ACCESS

EDITED BY
Bo Xiao,
Imperial College London, United Kingdom

REVIEWED BY
Huafeng Li,
Kunming University of Science and Technology, China
Puhong Duan,
Hunan University, China
Youlin Wang,
University of Montreal, Canada

*CORRESPONDENCE
Dongli Fan,
fdltmmu@sina.com
Zeyuan Lei,
leizeyuan0854@163.com

RECEIVED 02 January 2024
ACCEPTED 02 February 2024
PUBLISHED 14 February 2024

CITATION
Xiong Y, Yu K, Lan Y, Lei Z and Fan D (2024), Hair cluster detection model based on dermoscopic images.
Front. Phys. 12:1364372.
doi: 10.3389/fphy.2024.1364372

Hair cluster detection model based on dermoscopic images

Ya Xiong[1], Kun Yu[2], Yujie Lan[1], Zeyuan Lei[1]* and Dongli Fan[1]*

[1]Department of Plastic and Cosmetic Surgery, Xinqiao Hospital, The Army Medical University, Chongqing, China, [2]College of Automation, Chongqing University of Posts and Telecommunications, Chongqing, China

Introduction: Hair loss has always bothered many people, with numerous individuals potentially facing the issue of sparse hair.

Methods: Due to a scarcity of accurate research on detecting sparse hair, this paper proposes a sparse hair cluster detection model based on improved object detection neural network and medical images of sparse hair under dermatoscope to optimize the evaluation of treatment outcomes for hair loss patients. A new Multi-Level Feature Fusion Module is designed to extract and fuse features at different levels. Additionally, a new Channel-Space Dual Attention Module is proposed to consider both channel and spatial dimensions simultaneously, thereby further enhancing the model's representational capacity and the precision of sparse hair cluster detection.

Results: After testing on self-annotated data, the proposed method is proven capable of accurately identifying and counting sparse hair clusters, surpassing existing methods in terms of accuracy and efficiency.

Discussion: Therefore, it can work as an effective tool for early detection and treatment of sparse hair, and offer greater convenience for medical professionals in diagnosis and treatment.

KEYWORDS

hair loss, dermatoscope, hair cluster detection, feature fusion, dual attention module

1 Introduction

As a common issue, sparse hair [1] brothers many people, affecting both males and females alike [2], [3]. Hair loss or thinning primarily attributed to genetic factors, hormonal changes, environmental conditions, or medical conditions is a prevalent problem affecting millions worldwide [4]. Regardless of gender or age, it impacts an individual's self-esteem, personal aesthetics, and overall mental health. Traditional solutions such as drug treatments, hair transplants, or wearing wigs have achieved varying degrees of success and affordability, but they do not fundamentally resolve the problem or prevent its recurrence. Therefore, early detection and predictive analysis of sparse hair conditions are vital for implementing preventative measures and more effective treatments [5].

Over the past few decades, both domestic and international researchers have been exploring how to accurately detect sparse hair. The earliest research primarily relies on manual feature extraction and traditional image processing techniques [6]. However, due to the limitations on the selection and representational power of features, these methods are difficult to adapt to the complex and diverse forms of hair clusters. Therefore, with the rapid development of computer vision and deep learning [7], researchers introduce neural network into the field of sparse hair target detection. In recent years, with the advent of

artificial intelligence (AI) and deep learning technologies, their application in the healthcare sector grows exponentially, providing promising results in different fields like diagnosis, prognosis, treatment planning, and public health [8]. In light of this, the development of AI-driven sparse hair detection models [9], especially those based on neural network, offers a promising research pathway.

Based on the strong learning capability and adaptability, neural network is able to learn effective feature representations from a large amount of data and train and optimize through the backpropagation algorithm. This provides new opportunities and challenges for the target detection of sparse hair [10]. Researchers design and improve hair cluster target detection models based on neural network to enhance detection accuracy and robustness.

At present, domestic and international research in the field of sparse hair detection is still in the exploratory stage [11]. Some studies have utilized traditional Convolutional Neural Network (CNN) to detect hair clusters, improving detection performance by constructing deep-level feature representations and using effective loss functions. Other studies have explored more advanced network structures, such as Recurrent Neural Network (RNN) and Attention Mechanisms, to capture the temporal information and local details of hair clusters. In summary, using neural network in hair cluster target detection models for sparse hair detection has enormous potential to thoroughly transform hair care and treatment [12].

However, the target detection of sparse hair still faces some challenges. Hair clusters exhibit diverse morphologies with differences in color, texture, and shape [13], posing difficulties for detection algorithms. Additionally, due to the sparse distribution of hair, hair cluster targets unevenly occupy proportions in images, making target detection more challenging. Currently, dermatoscopy is a non-invasive diagnostic technique that allows the observation of hair shafts, follicles, and capillaries, providing a visual representation of inflammation around the scalp and changes in hair shaft diameter and shape [14]. It is widely used in the diagnosis and treatment of hair diseases, as well as in the assessment and follow-up of prognosis [15], [16], [17], [18]. Digital intelligent analysis of dermatoscopy is still in the developmental stage, and research on dermoscopy for androgenetic alopecia is limited. For the daily management and assessment of treatment outcomes for patients with hair loss, hair counting plays a crucial role. However, there are currently no clear standards for a comprehensive evaluation of hair loss across the entire scalp.

In response to these challenges, this study utilizes hair images obtained by dermoscopy, combined with existing advanced target detection techniques, to propose an efficient and accurate sparse hair cluster target detection model. This model sets the hair cluster as the detection target (in this paper, the sparse hair or hair loss area) and predicts the number of hair clusters. This paper has three main contributions as follows.

1. Based on the advanced existing object detection networks, a dermoscopy image hair detection network structure based on an improved object detection neural network is proposed to better adapt to sparse hair detection. Through experiments, it proves that the proposed method surpasses the existing methods in terms of accuracy and efficiency, providing an effective tool for early detection and treatment of sparse hair.
2. Multi-Level Feature Fusion Module: A new multi-level feature fusion Module (MLFF) is designed to extract and fuse features at different levels. The MLFF structure can obtain features from different convolutional layers, then integrate these features through a specific fusion strategy to produce a richer, more representative feature expression.
3. Channel-Space Dual Attention Module: A new attention mechanism, the Channel-Space Dual Attention Module, is proposed to consider both channel and spatial dimensions' information simultaneously. The CSDA module can handle channel and spatial correlation in a unified framework, thereby further enhancing the model's expressive capacity and accuracy of sparse hair detection.

2 Related work

With the rapid development of computer technology and computer-assisted medical diagnostic systems, the continuous growth of computational power and data, deep learning has experienced tremendous development, becoming one of the powerful tools in the medical field. The technology of feature extraction and classification from medical images [19], [20] using maturing deep learning models is increasingly mature.

The field of object detection has always been a research hotspot. For instance, one study proposed a safety helmet detection method based on the YOLOv5 algorithm [21]. This research involved annotating a collected dataset of 6,045, training, and testing the YOLOv5 model with different parameters. In another study, YOLOv4 was employed for small object detection and anti-complex background interference in remote sensing images [22]. With the use of deep learning-based algorithms, ship detection technology has greatly enriched, allowing monitoring of large, distant seas. Through the use of a custom dataset with four types of ship targets, Kmeans++ clustering algorithm for prior box framework selection, and transfer learning method, the study enhanced YOLOv4's detection ability. Further improvements were introduced by replacing Spatial Pyramid Pooling (SPP) with a Receptive Field Block with dilated convolution and adding a Convolutional Block Attention Module (CBAM). These modifications have improved the detection performance of small vessels and enhanced the model's resistance to complex backgrounds. Due to the relatively large size and distinct features of vessels, the detection results are satisfactory. However, it remains a challenge for densely packed, small targets.

In recent years, there has been an emergence of research utilizing deep learning methods in skin imaging analysis, particularly in studies related to hair. Researchers have explored the application of deep learning-based object detection [23], [24], segmentation [25], and other algorithms in hair detection and segmentation. These studies primarily focus on aspects such as hair detection, removal, segmentation, and even reconstruction, but there is room for improvement in terms of accuracy.

Various deep learning structures and techniques are introduced in multiple studies to address the challenges related to hair recognition and removal in dermoscopic images. One such

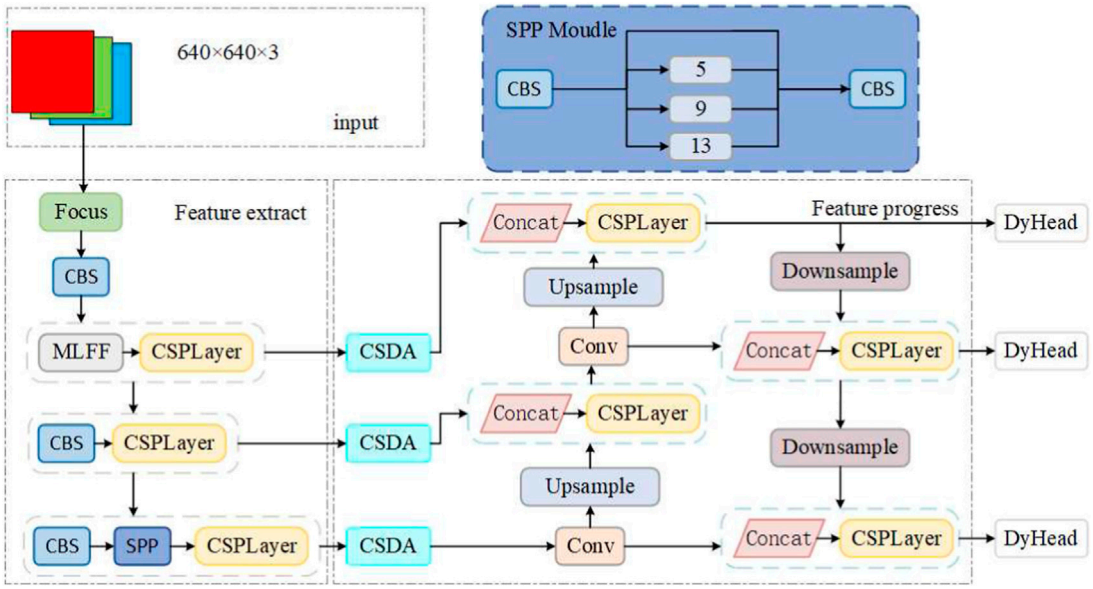

FIGURE 1
The method proposed in this paper.

study proposed a novel deep learning technique, Chimera Net [26], an encoder-decoder architecture that uses a pretrained EfficientNet and squeeze-and-excitation residual (SERes) structure. This method exhibited superior performance over well-known deep learning methods like U-Net and ResUNet-a. Additionally, other research explored difficulties and solutions related to hair reconstruction. A novel method was proposed to capture high-fidelity hair geometry with strand-level accuracy [13]. The multi-stage approach includes a new multiview stereo method and a novel cost function for reconstructing each hair pixel into a 3D line. The task of Digital Hair Removal (DHR) also received ample research. One study proposed a DHR deep learning method using U-Net and free-form image restoration architecture [9]. It outperforms other state-of-the-art methods on the ISIC2018 dataset. Another study explored a similar theme Attia et al. [10], highlighting the challenges associated with hair segmentation and its impact on subsequent skin lesion diagnosis. Moreover, one paper delved into an important metric for determining the number of hairs on the scalp [27]. It stressed the need for an automated method to increase speed and throughput while lowering the cost of counting and measuring hair in trichogram images. The proposed deep learning-based, enables rapid, fully automatic hair counting and length measurement. Another study described a real-time hair segmentation method based on a fully convolutional network, the basic structure of which is an encoder-decoder [28]. This method uses Mobile-Unet, a variant of the U-Net segmentation model, which combines the optimization techniques of MobileNetV2.

In summary, the above studies emphasize the enormous potential of deep learning techniques in advancing hair-related dermoscopy research. However, deep learning-based sparse hair detection is still in the exploratory stage. To address these challenges, this paper, based on sparse hair dermoscopic medical images, proposes a dermoscopic image hair detection network structure based on an improved object detection neural network to achieve the detection of sparse hair clusters (sparse hair or hair loss areas in this paper) and predict the number of hair clusters.

3 Materials and methods

In this section, we will provide a detailed introduction to the proposed sparse hair detection network structure, which is based on the object detection network [29]. Firstly, we will describe the overall structure of the network in Section 3.1. Subsequently, we will highlight the novel contributions of this paper in Sections 3.2, 3.3, namely, the MLFF Module and the CSDA Module, respectively.

3.1 Overall structure

The overall framework proposed for sparse hair detection in this article is illustrated in Figure 1, primarily based on enhancements to classical object detection network architectures. Given the crucial significance of the accuracy of the sparse hair detection model for hair target recognition and assisting doctors in obtaining diagnostic results, the model proposed in this article is intended for application in sparse hair target detection models.

It can be divided into three parts: the feature extraction backbone network, the feature enhancement and processing network, and the detection network. Specifically, the feature extraction backbone network is a convolutional neural network that incorporates the concept of a feature pyramid architecture, capable of extracting image features at different levels and reducing model computation while speeding up training. As shallow features contain more semantic information, a MLFF Module is proposed to handle them, preventing the loss of semantic information. At the end of the feature extraction backbone network, there is a Spatial Pyramid Pooling (SPP) module aimed at improving the network's

receptive field by transforming feature maps of arbitrary sizes into fixed-size feature vectors. Three main backbone features can be obtained through the feature extraction backbone network.

In the feature enhancement and processing network, the Channel-Spatial Dual Attention module (CSDA) is introduced. The three feature layers obtained from the backbone network undergo processing through this module to generate enhanced features. Subsequently, processing is carried out based on the YOLOv5 network model. This network segment primarily consists of a series of feature aggregation layers that mix and combine image features to generate a Feature Pyramid Network (FPN). The output feature maps are then transferred to the detection network. With the adoption of a novel FPN structure, this design strengthens the bottom-up pathway, improving the transfer of low-level features and enhancing the detection of objects at different scales. Consequently, it enables the accurate identification of the same target object with varying sizes and proportions.

The detection network is primarily employed for the final detection phase of the model. It applies anchor boxes to the feature maps output from the preceding layer and outputs a vector containing the class probability, object score, and position of the bounding box around the object. The detection network of the proposed architecture consists of three detection layers, with inputs being feature maps of sizes 80 × 80, 40 × 40, and 20 × 20, respectively, used for detecting objects of different sizes in the image. Each detection layer ultimately outputs an 18-dimensional vector ((4 + 1+1)×3 anchor boxes). The first four parameters are used for determining the regression parameters for each feature point, and adjusting these regression parameters yields the predicted box. The fifth parameter is utilized to determine whether each feature point contains an object, and the last parameter is employed to identify the category of the object contained in each feature point. Subsequently, the predicted bounding boxes and categories of the targets in the original image are generated and labeled, enabling the detection of clusters of hair targets in the image.

Algorithm 1 describes the training process of the hair detection model in dermoscopic images. The computation time increases linearly with the increase of training sample, batch size, and training epochs. The time complexity of the training algorithm is $O\left[E \times (n/B) \times 2 \times (M-1)\right]$.

Input: Training dataset D, segmentation model M, number of epochs E, learning rate η, n training samples, loss function L, batch size B
Output: Trained segmentation model $\hat{M}$.
Initialize segmentation model M
for $e \in [1, E]$ **do**
  for $b \in [1, n/B]$(mini-batch b in D with size B) **do**
    Perform forward pass on M with mini-batch b
    Calculate detection loss according to the loss function L
    Perform backward pass and update model weights and model according to the gradient
  end for
  Save the trained model $\hat{M}$
end for

Algorithm 1. A dermoscopy-image hair detection model based on improved object detection neural network.

3.2 Multi-level feature fusion structure

The main task of the MLFF (Multi-Level Feature Fusion) structure is to process a large amount of semantic information contained in shallow layers. Its structure is shown in Figure 2. The purpose of this module is to extract and fuse semantic information from shallow features, so that the resulting feature information is more detailed and more suitable for subsequent object detection tasks. Semantic feature information reflects a global feature of homogeneous phenomena in the image, depicting the surface organization and arrangement rules of slow-changing or cyclically-changing structures in the image. However, the low-level information extracted by the original backbone network (such as pixel values or local region attributes) is often of low quality and contrast, making it difficult to obtain and utilize this low-level information effectively. This paper proposes the MLFF module to address this problem.

As shown in Figure 2, in this module, a feature X_1 Eq. 1 before the output of this module serves as the input. It undergoes two consecutive CBS modules, resulting in two feature layers X_2 and X_3 Eq. 1, represented as follows:

$$\begin{aligned} X_1 &\in \mathbb{R}^{H\times W\times C} \\ X_2 &\in \mathbb{R}^{H\times W\times C} \\ X_3 &\in \mathbb{R}^{H\times W\times C} \end{aligned} \tag{1}$$

The CBS module represents a sequence of convolution operation, batch normalization operation, and activation function operation. This sequence is designed to capture local relationships within the input data, facilitating effective feature learning in images. Simultaneously, it helps mitigate the vanishing gradient problem and enhances the model's adaptability to changes in the distribution of input data. The CBS module can be expressed as follows:

$$X_{out} = SiLU\{BN\left[Conv\left(X_{in}, c_{in}, c_{out}\right)\right]\} \tag{2}$$

Where Conv represents the convolution operation, BN represents batch normalization operation, and SiLU represents the activation function operation. X_{out} represents the output feature of the CBS module, X_{in} represents the input feature of the CBS module, c_{in} represents the number of channels in the input feature, and c_{out}represents the number of channels in the output feature.

After the three features obtained through stacking and fusion, two feature layers are obtained. They will undergo another CBS module (where $c_{in} = c_{out}$) for feature processing. Finally, these features will be stacked together, achieving feature integration. With the depth of feature processing and fusion, the dimension of the image feature vector continuously increases, and the size of each slice changes accordingly. Finally, after passing through a CBS module (where $c_{in} = c_{out}$), as in Eq. 2, the output feature Eq. 3 is:

$$X_{MLFF} \in \mathbb{R}^{\frac{H}{4}\times\frac{W}{4}\times 4C} \tag{3}$$

The obtained features will be inputted into the feature enhancement and processing network for further processing, where the abundant semantic information contained in the shallow layers will be fully utilized to achieve better detection performance. The first three branches actually correspond to dense residual structures, which take into account the easy-to-

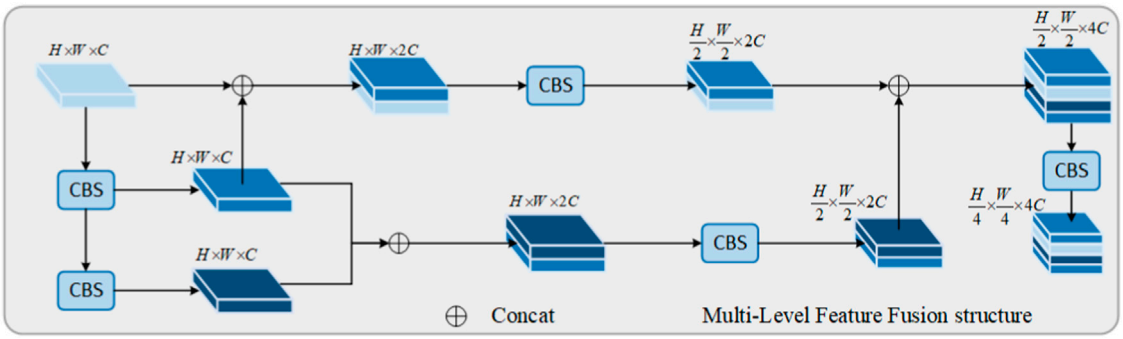

FIGURE 2
Multi-level feature fusion structure.

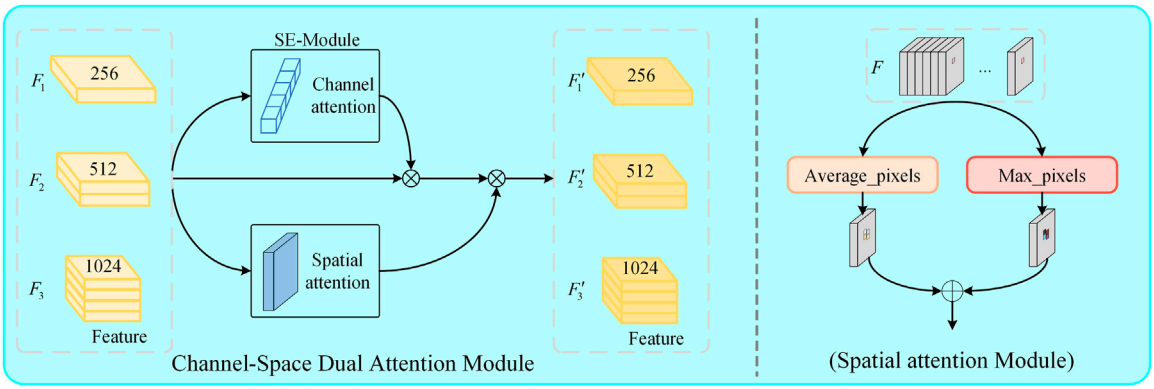

FIGURE 3
Channel spatial dual attention module.

optimize characteristics of residual networks, and the ability of residual networks to improve the overall accuracy of the network by adding a considerable depth. In addition, skip connections are used to alleviate the problem of gradient disappearance caused by the depth of the neural network.

For the CBS module, the SiLU activation function is used, which is an improved version based on the Sigmoid activation function and ReLU activation function. SiLU has the characteristics of no upper bound and a lower bound, smoothness, and non-monotonicity. SiLU performs better than ReLU in deep models and can be considered as a smoothed ReLU activation function. Its specific implementation is shown in the equation below Eq. 4:

$$f(x) = x \cdot sigmoid(x) \tag{4}$$

3.3 Channel-space dual attention module

After obtaining feature information at different depths, it is necessary to further process these features to capture the target information in them. Therefore, this paper proposes a Channel-Space Dual Attention Module (CSDA) for feature inference, as shown in Figure 3. Finally, the inferred information is passed through the second part of the object detection model architecture to obtain three types of feature maps.

The module proposed in this article takes the feature layers obtained from the feature extraction backbone network, namely, $F_1 \in \mathbb{R}^{80\times80\times256}$, $F_2 \in \mathbb{R}^{40\times40\times512}$ and $F_3 \in \mathbb{R}^{20\times20\times1024}$, and infers attention maps along two different dimensions. One dimension is the channel attention mechanism, which is based on the SE module [30] and uses global average pooling to calculate channel attention. The other dimension is the spatial attention mechanism, which focuses on which pixels in different feature maps are important and require significant attention. Then, the channel attention map and the spatial attention map are multiplied successively with the feature maps on the backbone to perform adaptive feature focusing, resulting in corresponding feature maps F'_1, F'_2 and F'_3.

For the Squeeze-and-Excitation module, it can be viewed as a computational unit that mainly embeds the dependency factors of feature map channels into variable v. This is to ensure that the network can enhance its sensitivity to information features and suppress less useful features. In the channel-wise optimization process, squeezing and excitation steps are applied to optimize the response of the convolutional kernel, in order to capture the correlation of channel information. The specific implementation is shown in the following equation:

$$C_{tran}: x \rightarrow y;\ x, y \in \mathbb{R}^{H \times W \times C} \tag{5}$$

In the equation, C_{tran} is the convolutional operator, $v = [v_1, v_2, \ldots, v_n]$ represents the learned weights in the network, and n denotes the parameters of the $n - th$ convolutional kernel. Therefore, the output of the convolutional operator is $Y = [y_1, y_2, \ldots, y_n]$, which is implemented as shown in Eq. 5 and Eq. 6. In the proposed attention module, after the channel attention, we can obtain the feature $F_{channel}$.

$$Y = v * X = \sum_{n=1}^{n} v_n * x_n \tag{6}$$

Regarding the spatial attention module, as shown in the right half of Figure 3, the feature map obtained by the feature extraction network is understood as a three-dimensional space, where each slice corresponds to a channel. Firstly, the values at the same position on different channels are subjected to average pooling and max pooling operations to obtain the features $F_{\max}$, $F_{average}$ Eq. 7.

$$\begin{aligned} F_{\max} &= MaxPool(F) \\ F_{average} &= AvgPool(F) \end{aligned} \tag{7}$$

Finally, convolution and normalization operations are applied to generate a 2D spatial attention map $F_{spatial}$, which is computed as follows Eq. 8:

$$F_{spatial} = sigmoid\left(f^{7\times7}\left(F_{\max}, F_{average}\right)\right) \tag{8}$$

The symbol $f^{7 \times 7}$ represents a convolution operation with a kernel size of 7×7. After obtaining the channel attention map, it is multiplied with the input feature map F to obtain a new feature map F'. This new feature map F' is then multiplied with the spatial attention map to obtain the final feature map F''. The overall process can be described as follows Eq. 9:

$$\begin{aligned} F' &= F_{channel} \otimes F \\ F'' &= F' \otimes F_{saptial} \end{aligned} \tag{9}$$

Finally, three feature maps, denoted as F'_1, F'_2 and F'_3, can be obtained. The obtained new features are then processed and enhanced using feature processing networks and detection networks to obtain the final object detection results. The experimental results of the proposed network will be discussed in Section 3 of this paper.

3.4 Attention dynamic head

Introducing dynamic heads [31], based on three feature maps F'_1, F'_2 and F'_3, the general formula for applying self-attention is as follows Eq. 10:

$$W(\mathcal{F}) = \pi(\mathcal{F}) \cdot \mathcal{F} \tag{10}$$

Where $\pi(\cdot)$ is an attention function. A simple solution to this attention function is achieved through fully connected layers. However, due to the high dimensionality of tensors, directly learning attention functions across all dimensions is computationally expensive and practically unaffordable. Therefore, transforming the attention function into attention along three directions, with each attention focusing on a single direction, is proposed Eq. 11.

$$W'(\mathcal{F}) = \pi_C(\pi_S(\pi_L(\mathcal{F}) \cdot \mathcal{F}) \cdot \mathcal{F}) \cdot \mathcal{F} \tag{11}$$

Where $\pi_L(\cdot)$, $\pi_S(\cdot)$, $\pi_C(\cdot)$ are three different attention functions applied respectively to dimensions L, S, and C.

4 Experimental results and analysis

4.1 Datasets

In the experiment described in this paper, both the training and testing datasets are sourced entirely from hospitals and collected based on different patients, each with varying degrees of hair sparsity. The original dataset is devoid of any annotations, and labeling is used to annotate it, generating XML-format files to store the labeled tags. Each image corresponds to one XML file, containing multiple hair cluster labels, primarily annotating each hair cluster. In the experiment, each hair cluster does not exceed three strands. A total of 200 images were annotated for the dataset. As neural network-based object detection models are developed on the basis of extensive image data, the dataset is expanded and divided through data augmentation, resulting in 500 images. From these, 50 images are randomly selected as the validation set, and another 50 images are chosen as the test set. This is done to enrich the dataset size, better extract features of hair belonging to different labeled categories, and prevent the trained model from overfitting. The objective of this dataset is to achieve hair detection in populations with sparse hair, identifying the number of hair clusters.

4.2 Experimental details

During the preprocessing stage, the source dataset had a size of 1,920 × 1,080. In this study, all hair datasets underwent image enhancement and partitioning, resulting in a final size of 640 × 640 for each slice.

In the experiment, all programs were implemented in the PyTorch framework under the Windows 10 operating system. The training process used one GeForce RTX 3090 GPU and was written in Python language, calling CUDA, CuDNN, OpenCV, and other required libraries. The optimizer used in the experiment was SGD, with a momentum of 0.937 and default parameters for other settings. The initial learning rate, weight decay, and batch size were set to 0.01, 5e-4, and 8, respectively, and the epoch was set to 500. The trained model's weight file was saved, and the model's performance was evaluated using the test set.

The model evaluation metrics adopted include commonly used object detection metrics such as Precision, Recall, mAP (mean average precision), and F1 score, which are used to assess the performance of the trained model. Visual comparison was also conducted. The implementation of these metrics is as follows Eq. 12:

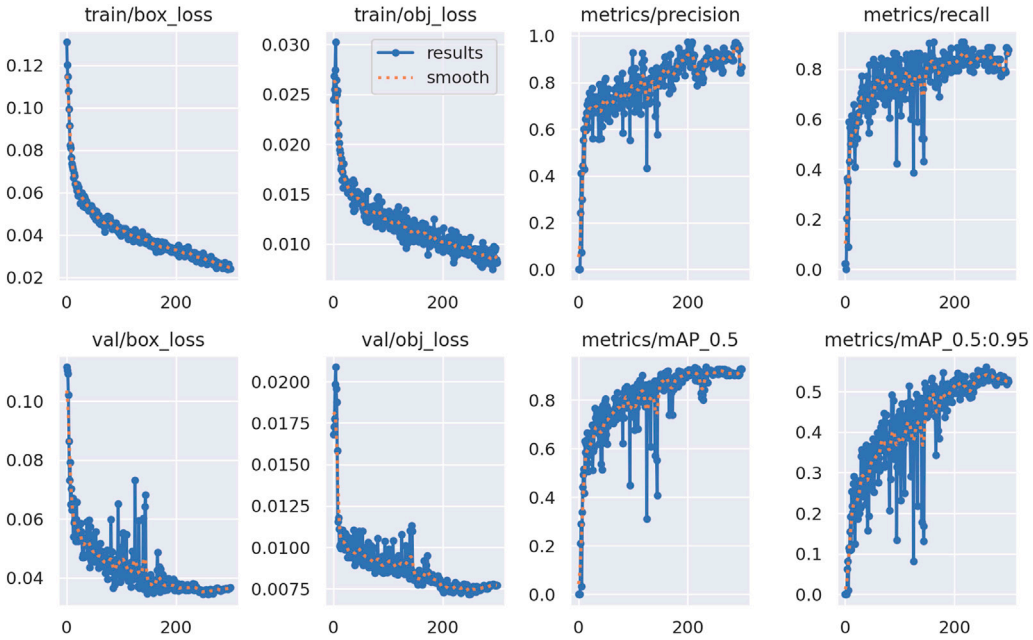

FIGURE 4
Network training situation.

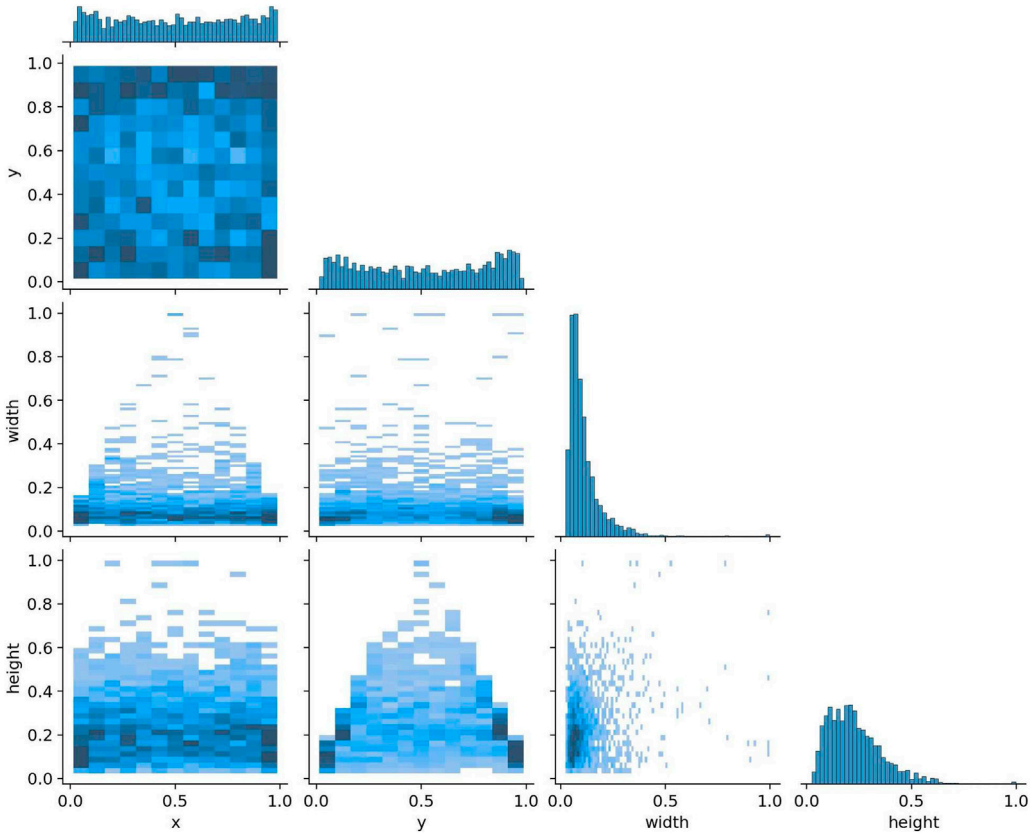

FIGURE 5
Correlation between predicted labels during network training.

TABLE 1 Comparison with different detection networks (Bold numbers represent best results).

Networks	year	Precision	mAP	F1 score	Recall
YOLOv3	2018	0.733	0.500	0.35	0.471
YOLOv4	2020	0.768	0.561	0.58	0.434
Mobilenet YOLOv4	2020	0.792	0.406	0.21	0.245
YOLOv5	2020	0.865	0.706	0.63	0.677
Detr	2020	0.822	0.717	0.65	0.854
FastestV2	2021	0.479	0.458	0.52	0.564
YOLOv7	2022	0.816	0.697	0.66	0.691
FastestDet	2022	0.609	0.524	0.47	0.593
YOLOv8	2023	0.820	0.658	0.63	0.712
Our network	-	**0.898**	**0.734**	**0.72**	**0.873**

TABLE 2 Comparison of ablation experiments of target detection indicators on data sets (Bold numbers represent best results).

Networks	Precision	mAP	F1 score	Recall
Without MLFF	0.817	0.680	0.57	0.712
Without CSDA	0.762	0.599	0.33	0.588
Our network	**0.898**	**0.734**	**0.72**	**0.873**

$$\begin{aligned} Precision &= \frac{TP}{TP+FP} \\ Recall &= \frac{TP}{TP+FN} \\ mAP &= \frac{1}{C}\sum_{k=0}^{C} AP_k \\ F1 &= \frac{2}{\frac{1}{Precision}+\frac{1}{Recall}} \end{aligned} \tag{12}$$

Among them, *TP* represents the number of correctly identified clusters of hair; *FP* represents the number of clusters mistakenly identified as hair; *FN* represents the number of hair cluster targets that were not successfully identified; *C* represents the number of categories of hair cluster targets; *AP* represents the area enclosed by the precision-recall curve and the coordinate axis.

Figure 4 displays the training and validation loss curves, as well as precision, recall, and mAP curves for the entire training process. The model is trained from scratch, and from the curves in the figure, it is evident that the network model descends rapidly in the first 50 epochs and gradually stabilizes thereafter. In the figure, a smaller box_loss indicates more accurate bounding boxes, and a smaller obj_loss indicates more accurate predictions of targets. Precision, recall, and mAP curves stabilize later, indicating a good training outcome. In summary, the figure demonstrates that the model for hair cluster detection is well-trained and does not exhibit overfitting. Figure 5 shows the correlation between predicted labels during the training process of the hair cluster object detection model. Figure 5 is a set of 2D histograms, illustrating the contrast between each axis of the data. Labels in the image are located in the xywh space, where x and y represent the center values of the label box, and w and h represent the length and width of the label box. The histograms of x and y in Figure 5 indicate that the size variation of detected targets is small. Additionally, the distribution plots of x and width, as well as y and height, show that their relationships have a linear correlation. Combined with Figure 4, this suggests that the proposed model for the hair cluster object detection task is trainable.

4.3 Comparative experiments

In the comparative experiments, to validate the performance of the proposed hair cluster detection model based on sparse hair, experiments and analyses were conducted on test set images using publicly available source code of classical object detection models. The object detection network developed in this study was compared with YOLOv3 [32], YOLOv4 [33], MobileNet YOLOv4, YOLOv5, Detr, FastestV2, YOLOv7, FastestDet, and YOLOv8 on test set images. Table 1 presents the performance of the proposed method and other methods on the test set.

The comparative experimental results in Table 1 indicate that the hair cluster detection model proposed in this study achieves the highest mAP value, surpassing the classical YOLOv5 network model by 2.8%. Additionally, it outperforms the latest YOLOv8 by 7.6%. This suggests that the proposed algorithm has advantages in the task of hair cluster target recognition. Moreover, the proposed model achieves the highest Precision, F1, and Recall scores, demonstrating the superior performance of the sparse hair cluster model proposed in this study. Therefore, the results indicate that the proposed model can ensure accurate identification of sparse hair clusters, comparable to the best methods in terms of metrics, and surpassing most other methods.

To more clearly illustrate the performance of the proposed method, visual experiments were conducted on six images selected from the test set, as shown in Figure 6. Figure 6 displays the visual comparison of hair cluster detection results obtained by the proposed method and five other methods (YOLOv8, YOLOv7, Detr, FastestDet, FastestV2) under the same experimental conditions. It is evident that the proposed method achieves more accurate hair cluster detection results compared to other methods.

As evident from the obtained detection results above, the proposed hair cluster detection model for sparse hair in this study has achieved significant results. Simultaneously, the algorithm accomplishes counting and visualizing the detected clusters. A comparison reveals that the method developed in this study exhibits the best performance in hair cluster detection. In Figure 6, it can be observed that other methods show instances of hair cluster omission. In summary, the method investigated in this study demonstrates commendable hair cluster detection performance. Finally, for a more comprehensive comparison of the advantages of the proposed method against different approaches, Figure 7 depicts bar charts representing the hair cluster detection performance of various methods across different metrics. The performance on four metrics is illustrated separately. It is evident that the proposed method holds a significant advantage in hair cluster detection tasks.

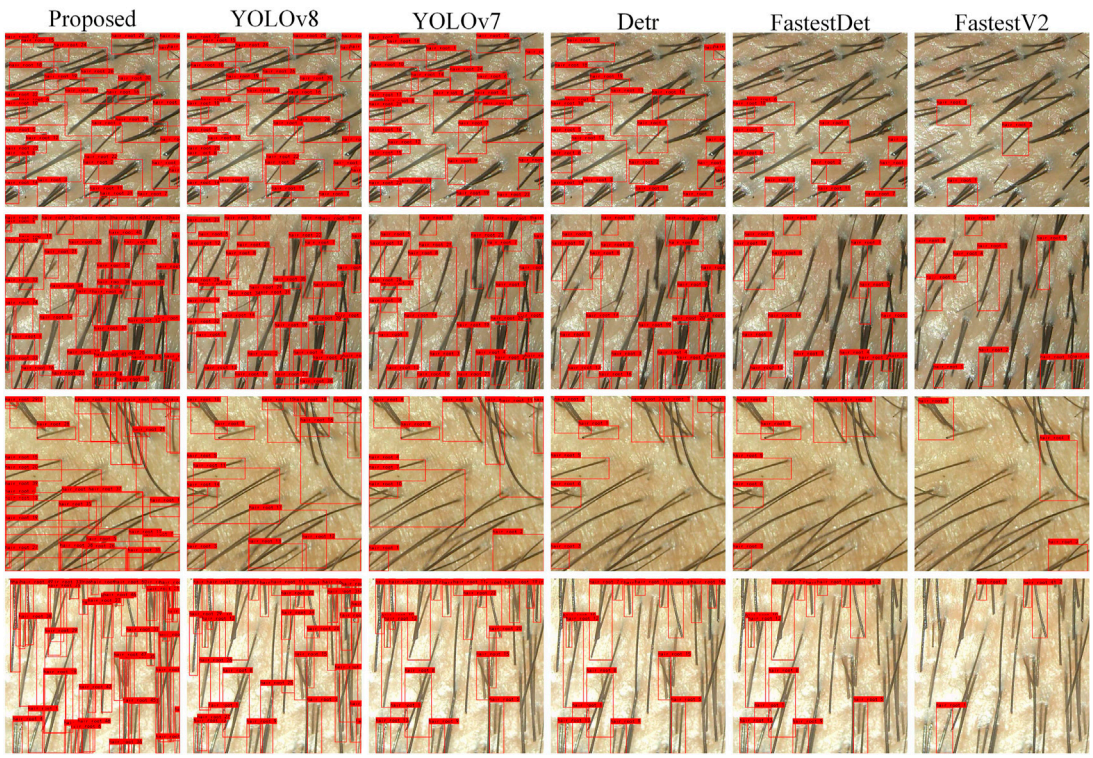

FIGURE 6
Visual comparison of hair cluster detection results.

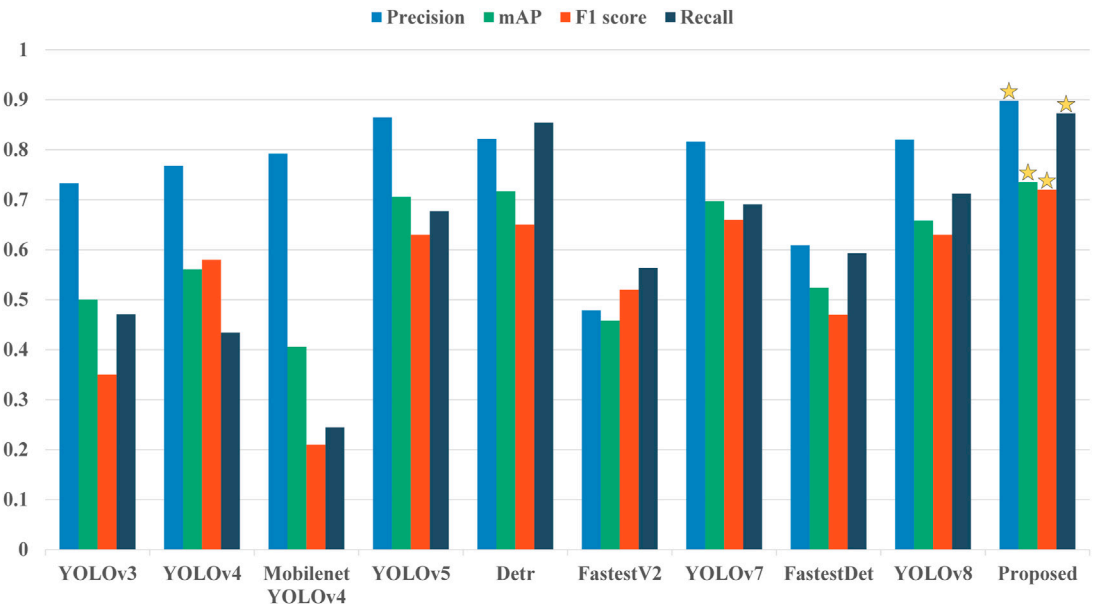

FIGURE 7
Performance comparison of different detection methods on the four indicators of Precision, Recall, mAP (mean average precision), and F1 score. The method that performs best in each case is marked with an asterisk.

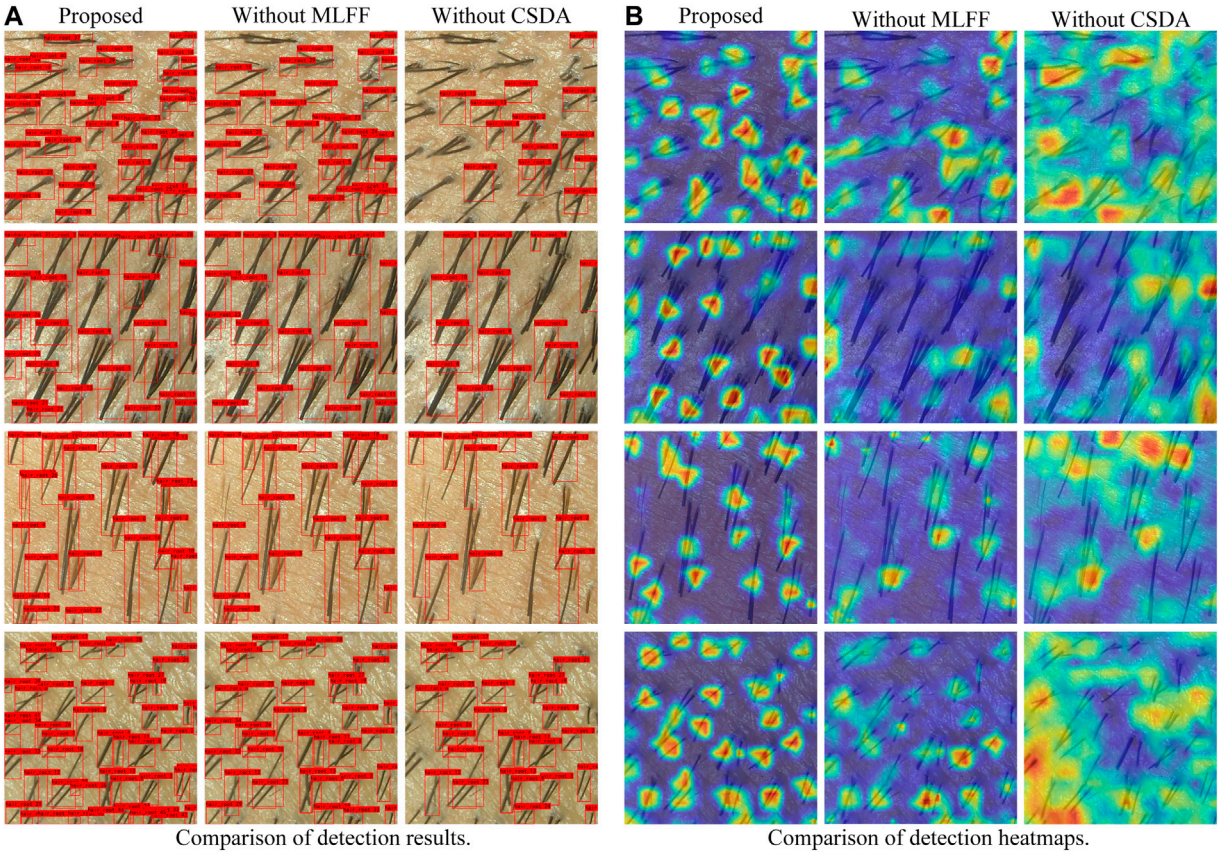

FIGURE 8
Visual comparison of ablation experiment results. **(A)**: Comparison of detection results; **(B)**: Comparison of detection heatmaps.

4.4 Ablation experiment

This study utilizes the developed model as the network for sparse hair target detection (Ours) in hair cluster detection. Experiments were conducted by removing the designed modules from this model. Specifically, the MLFF module was removed from the feature extraction network to assess the extraction of image features, and the CSDA module was removed from the feature enhancement and processing network to examine feature inference and fusion. As shown in the performance metrics results in Table 2, removing the corresponding modules leads to a decrease in the model's detection performance. Additionally, as depicted in Figure 8A, it is apparent that some smaller and overlapping hair clusters are missed when certain modules are removed, while the detection results proposed in this study remain superior.

To further explore the differences between different modules and their reasons, a heatmap analysis was conducted. Figure 8B visualizes the objective performance of different modules. It can be observed that removing the CSDA module generates regions of interest extending beyond the actual target area, focusing on some irrelevant background information. While focusing on certain background regions might not significantly impact normal target detection, it proves detrimental for densely distributed small targets, exacerbating background interference and the difficulty of instance recognition. Without the MLFF module, the situation of missed detections is more severe, indicating that the inclusion of the MLFF module in the network brings more information about the target. In conclusion, the proposed modules in this study contribute to improving the model's detection performance to a certain extent, significantly enhancing the overall performance of the target detection network.

5 Conclusion

In this study, we have proposed and implemented an efficient and accurate detection model specifically designed for sparse hair clusters. This model is based on an improved neural network for object detection. The construction of this model introduces three innovative aspects: firstly, we designed a new neural network structure based on existing advanced object detection networks to optimize the detection of sparse hair. Secondly, a novel multi-level feature fusion structure was devised to better extract and fuse features at different levels. Lastly, a new attention mechanism, the Channel-Spatial Bi-Attention Module, was introduced to simultaneously consider information in both channel and spatial dimensions, further enhancing the model's expressive power and the accuracy of sparse hair detection.

The model primarily consists of three parts: a feature extraction backbone network, a feature enhancement and processing network, and a detection network. It effectively achieves the detection of hair

clusters, predicting the number of hair clusters with promising results in experiments. Despite the application of dermoscopy in hair detection being in an exploratory and developing stage, and related research being incomplete, our study provides a new and effective tool for the precise detection of sparse hair clusters. It opens up new avenues for research and applications in hair detection, contributing to the advancement of dermoscopy in hair detection. This, in turn, assists healthcare professionals in diagnosing conditions and selecting treatment plans, while also providing convenience for daily management and condition monitoring for individuals with hair loss.

If the decisions made by the model are not interpretable, they may not be accepted by individuals. In future research, our project team will explore the interpretability of the hair cluster object detection network, applying these advancements to help healthcare professionals understand the processes in image analysis. Additionally, in order to bring the detection model to edge devices for user convenience, we will explore the development of lightweight hair cluster object detection models in the future.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Medical Ethics Committee of the Second Affiliated Hospital of Army Medical University of Chinese People's Liberation Army. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

YX: Data curation, Software, Supervision, Visualization, Writing–original draft. KY: Resources, Software, Validation, Writing–original draft. YL: Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Software, Writing–review and editing. ZL: Data curation, Formal Analysis, Project administration, Software, Supervision, Writing–review and editing. DF: Conceptualization, Data curation, Resources, Writing–original draft.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Sperling LC, Mezebish DS. Hair diseases. *Med Clin North America* (1998) 82: 1155–69. doi:10.1016/s0025-7125(05)70408-9

2. Franzoi SL, Anderson J, Frommelt S. Individual differences in men's perceptions of and reactions to thinning hair. *J Soc Psychol* (1990) 130:209–18. doi:10.1080/00224545.1990.9924571

3. Shapiro J. Hair loss in women. *New Engl J Med* (2007) 357:1620–30. doi:10.1056/nejmcp072110

4. Ahmed A, Almohanna H, Griggs J, Tosti A. Genetic hair disorders: a review. *Dermatol Ther* (2019) 9:421–48. doi:10.1007/s13555-019-0313-2

5. York K, Meah N, Bhoyrul B, Sinclair R. A review of the treatment of male pattern hair loss. *Expert Opin Pharmacother* (2020) 21:603–12. doi:10.1080/14656566.2020.1721463

6. O'Mahony N, Campbell S, Carvalho A, Harapanahalli S, Hernandez GV, Krpalkova L, et al. Deep learning vs traditional computer vision. In Advances in Computer Vision: Proceedings of the 2019 Computer Vision Conference (CVC), Volume 1 1; 25-26 April 2019; Las Vegas, Nevada, USA. Springer (2020). 128–44.

7. Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E, et al. Deep learning for computer vision: a brief review. *Comput intelligence Neurosci* (2018) 2018:1–13. doi:10.1155/2018/7068349

8. Esteva A, Chou K, Yeung S, Naik N, Madani A, Mottaghi A, et al. Deep learning-enabled medical computer vision. *NPJ digital Med* (2021) 4:5. doi:10.1038/s41746-020-00376-2

9. Li W, Raj ANJ, Tjahjadi T, Zhuang Z. Digital hair removal by deep learning for skin lesion segmentation. *Pattern Recognition* (2021) 117:107994. doi:10.1016/j.patcog.2021.107994

10. Attia M, Hossny M, Zhou H, Nahavandi S, Asadi H, Yazdabadi A. Digital hair segmentation using hybrid convolutional and recurrent neural networks architecture. *Comp Methods Programs Biomed* (2019) 177:17–30. doi:10.1016/j.cmpb.2019.05.010

11. Kim M, Gil Y, Kim Y, Kim J. Deep-learning-based scalp image analysis using limited data. *Electronics* (2023) 12:1380. doi:10.3390/electronics12061380

12. Hosny KM, Elshora D, Mohamed ER, Vrochidou E, Papakostas GA. Deep learning and optimization-based methods for skin lesions segmentation: a review. *IEEE Access* (2023) 11:85467–88. doi:10.1109/access.2023.3303961

13. Nam G, Wu C, Kim MH, Sheikh Y. Strand-accurate multi-view hair capture. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; June 16 2019 to June 17 2019; Long Beach, CA, USA (2019). p. 155–64.

14. Cuéllar F, Puig S, Kolm I, Puig-Butille J, Zaballos P, Martí-Laborda R, et al. Dermoscopic features of melanomas associated with mc1r variants in Spanish cdkn2a mutation carriers. *Br J Dermatol* (2009) 160:48–53. doi:10.1111/j.1365-2133.2008.08826.x

15. Tosti A, Torres F. Dermoscopy in the diagnosis of hair and scalp disorders. *Actas dermo-sifiliográficas* (2009) 100:114–9. doi:10.1016/s0001-7310(09)73176-x

16. Pirmez R, Tosti A. Trichoscopy tips. *Dermatol Clin* (2018) 36:413–20. doi:10.1016/j.det.2018.05.008

17. Van Camp YP, Van Rompaey B, Elseviers MM. Nurse-led interventions to enhance adherence to chronic medication: systematic review and meta-analysis of randomised controlled trials. *Eur J Clin Pharmacol* (2013) 69:761–70. doi:10.1007/s00228-012-1419-y

18. Shen X, Yu RX, Shen CB, Li CX, Jing Y, Zheng YJ, et al. Dermoscopy in China: current status and future prospective. *Chin Med J* (2019) 132:2096–104. doi:10.1097/cm9.0000000000000396

19. Zhu Z, He X, Qi G, Li Y, Cong B, Liu Y. Brain tumor segmentation based on the fusion of deep semantics and edge information in multimodal mri. *Inf Fusion* (2023) 91:376–87. doi:10.1016/j.inffus.2022.10.022

20. He X, Qi G, Zhu Z, Li Y, Cong B, Bai L. Medical image segmentation method based on multi-feature interaction and fusion over cloud computing. *Simulation Model Pract Theor* (2023) 126:102769. doi:10.1016/j.simpat.2023.102769

21. Zhou F, Zhao H, Nie Z. Safety helmet detection based on yolov5. In: 2021 IEEE International conference on power electronics, computer applications (ICPECA) (IEEE); January 22-24, 2021; Shenyang, China (2021). p. 6–11.

22. Huang Z, Jiang X, Wu F, Fu Y, Zhang Y, Fu T, et al. An improved method for ship target detection based on yolov4. *Appl Sci* (2023) 13:1302. doi:10.3390/app13031302

23. Qi G, Wang H, Haner M, Weng C, Chen S, Zhu Z. Convolutional neural network based detection and judgement of environmental obstacle in vehicle operation. *CAAI Trans Intelligence Tech* (2019) 4:80–91. doi:10.1049/trit.2018.1045

24. Qi G, Zhang Q, Zeng F, Wang J, Zhu Z. Multi-focus image fusion via morphological similarity-based dictionary construction and sparse representation. *CAAI Trans Intelligence Tech* (2018) 3:83–94. doi:10.1049/trit.2018.0011

25. Li Y, Wang Z, Yin L, Zhu Z, Qi G, Liu Y. X-net: a dual encoding–decoding method in medical image segmentation. *Vis Comp* (2021) 39:2223–33. doi:10.1007/s00371-021-02328-7

26. Lama N, Kasmi R, Hagerty JR, Stanley RJ, Young R, Miinch J, et al. Chimeranet: U-net for hair detection in dermoscopic skin lesion images. *J Digital Imaging* (2023) 36:526–35. doi:10.1007/s10278-022-00740-6

27. Sacha JP, Caterino TL, Fisher BK, Carr GJ, Youngquist RS, D'Alessandro BM, et al. Development and qualification of a machine learning algorithm for automated hair counting. *Int J Cosmet Sci* (2021) 43:S34–S41. S34–S41. doi:10.1111/ics.12735

28. Yoon HS, Park SW, Yoo JH. Real-time hair segmentation using mobile-unet. *Electronics* (2021) 10:99. doi:10.3390/electronics10020099

29. Wu W, Liu H, Li L, Long Y, Wang X, Wang Z, et al. Application of local fully convolutional neural network combined with yolo v5 algorithm in small target detection of remote sensing image. *PloS one* (2021) 16:e0259283. doi:10.1371/journal.pone.0259283

30. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; June 18 2018 to June 23 2018; Salt Lake City, UT, USA (2018). 7132–41.

31. Dai X, Chen Y, Xiao B, Chen D, Liu M, Yuan L, et al. Dynamic head: unifying object detection heads with attentions. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; June 20 2021 to June 25 2021; Nashville, TN, USA (2021). 7373–82.

32. Redmon J, Farhadi A. Yolov3: an incremental improvement[J] (2018). arXiv preprint arXiv:1804.02767 Available at: https://arxiv.org/pdf/1804.02767.pdf.

33. Bochkovskiy A, Wang CY, Liao HYM. Yolov4: optimal speed and accuracy of object detection[J] (2020). arXiv preprint arXiv:2004.10934 Available at: https://arxiv.org/abs/2004.10934.