

Article

A Skin Cancer Classification Method Based on Discrete Wavelet Down-Sampling Feature Reconstruction

Qing-e Wu ^{1,*}, Yao Yu ¹ and Xinyang Zhang ^{2,3}

¹ School of Electrical and Information Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, China; 332101030027@email.zzuli.edu.cn

² School of Mechanical Engineering, University of Science and Technology Beijing, Beijing 100083, China

³ Shunde Innovation School, University of Science and Technology Beijing, Foshan 528399, China

* Correspondence: wqe2008@zzuli.edu.cn or wqe969699@163.com

Abstract: Aiming at the problems of feature information loss during down-sampling, insufficient characterization ability and low utilization of channel information in skin cancer diagnosis of melanoma, a skin pathological mirror classification method based on discrete wavelet down-sampling feature reconstruction is proposed in this paper. The wavelet down-sampling method is given first, and the multichannel attention mechanism is introduced to realize the pathological feature reconstruction of high-frequency and low-frequency components, which reduces the loss of pathological feature information due to down-sampling and effectively utilizes the channel information. A skin cancer classification model is given, using a combination of depth-separable convolution and 3×3 standard convolution and wavelet down-sampling as the input backbone of the model to ensure the perceptual field while reducing the number of parameters; the residual module of the model is optimized using wavelet down-sampling and Hard-Swish activation function to enhance the feature representation capability of the model. The network weight parameters are initialized on ImageNet using transfer learning and then debugged on the augmentation HAM10000 dataset. The experimental results show that the accuracy of the proposed method for skin cancer pathological mirror classification is significantly improved, reaching 95.84%. Compared with the existing skin cancer classification methods, the proposed method not only has higher classification accuracy but also accelerates the classification speed and enhances the noise immunity. The method proposed in this paper provides a new classification method for skin cancer classification and has some practical value.

Keywords: melanoma; down-sampling; wavelet transform; image classification; transfer learning; data augmentation



Citation: Wu, Q.-e.; Yu, Y.; Zhang, X. A Skin Cancer Classification Method Based on Discrete Wavelet Down-Sampling Feature Reconstruction. *Electronics* **2023**, *12*, 2103. <https://doi.org/10.3390/electronics12092103>

Academic Editor: Hyunjin Park

Received: 11 April 2023

Revised: 30 April 2023

Accepted: 1 May 2023

Published: 4 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Skin cancer is one of the most widespread threats to human health worldwide, and one malignant skin tumor, also known as melanoma, is one of the swift-growing cancers in the world. This disease can be easily treated if detected early [1]. A report has shown that the five-year survival rate of localized malignant melanoma is 99% when diagnosed and treated early, whereas the survival rate of advanced melanoma is only 25%; therefore, whether the disease can be diagnosed early has a crucial impact on whether patients can improve their survival rate [2]. The traditional examination method is to first go through a doctor's visual inspection and then use dermoscopic imaging to aid the diagnosis [3]. However, benign and malignant skin cancers are extremely similar in characteristics, and the varying levels of professional doctors and the uncertainty of subjective decision-making have led to numerous skin cancer patients not being diagnosed early and treated timely. With the development of artificial intelligence in the medical field, deep learning has been widely used for the detection and classification of medical images over the past few years [4]. The application of artificial intelligence in medical image-assisted diagnosis to provide a

second opinion to help professional doctors diagnose skin cancer has become an inevitable trend [5]. Deep convolutional neural networks (CNNs), an important research topic in deep learning, have been widely used in the medical field [6].

Codella et al. [7] used features from sparse coding features, low-level manual features and deep residual networks to adjust the output of the AlexNet network and use a support vector machine (SVM) to perform classification experiments on the ISIC database and finally obtained 93.1% accuracy, 92.8% specificity, and 94.9% sensitivity. Pomponiu et al. [8] used a migration learning-based AlexNet convolutional neural network and data augmentation to classify skin lesion data after feature extraction using the K nearest neighbor (KNN) algorithm and finally obtained 93.64% accuracy, 95.18% specificity, and 92.1% sensitivity. Esteva et al. [9] used a skin lesion classification method based on a pre-trained GoogleNet architecture and performed classification experiments on clinical images and finally obtained 72.1% accuracy. Khan et al. [10] improved a novel deep CNN based on VGG and AlexNet models to classify skin lesions, and classification tests that were performed on ISIC 2016 and ISIC 2017 datasets finally obtained 92.1% and 96.5% accuracy, respectively. Ahmed et al. [11] used three models based on deep CNN (InceptionResNetV2, Xception and NASNetLarge) ensembles to test the pre-trained models on the ISIC2019 dataset, and finally, the CNN ensembles obtained 93.70% accuracy and 0.931 AUC and exceeded the top-2 rankings on the ISIC2019 classification challenge leaderboard for that year. Abayomi-Alli et al. [12] improved a new data augmentation model for skin cancer classification, and a migration learning based pre-trained squeeze net architecture was used to classify melanoma skin cancer. Classification experiments were performed on the PH2 dataset and finally obtained 92.18% accuracy. Although convolutional neural networks can classify dermoscopy images well, the issue of partial feature information loss caused by down-sampling will be difficult to avoid as the depth and breadth of the network increase. As the performance characteristics of malignant and benign skin diseases are too similar, the problem of partial loss of feature information due to down-sampling may affect the accuracy of the classification of dermoscopy images. Pooling operation is widely used in convolutional neural networks as the main down-sampling method, which often ignores some useful feature information and causes poor generalization ability, which affects the performance of the model. To deal with this problem, Zeiler et al. [13] proposed a stochastic pooling strategy that replaced the traditional deterministic pooling operation with a stochastic process and was able to combine it with other arbitrary forms of regularization, abandoning the dropout-like approach of discarding information in favor of selecting from the network's existing information, obtaining state-of-the-art performance at the time. He et al. [14] designed a stride convolution as a learnable down-sampling method, which was widely used in the model but introduced more parameters and overhead at the same time. Zhao et al. [15] embedded random shifting in the down-sampling layer during the training process and dynamically adjusted receptive fields by shifting the kernel centers on the feature maps in different directions, effectively mitigating the loss of feature information during down-sampling. Jiang et al. [16] introduced parameters into the pooling layer and used parametric pooling to replace the traditional pooling process, which preserves the features that are desired to be preserved in the convolutional neural network with a small increase in network parameters. Although the above methods alleviate the problem of losing feature information due to pooling operations to a certain extent, they are still traditional pooling operations in nature, and no matter what pooling methods are used, they will cause feature information loss and thus affect the performance of convolutional neural networks.

Aiming at the problems of feature information loss, low utilization of channel information and insufficient characterization capability in traditional down-sampling, this paper proposes a discrete wavelet-based down-sampling feature reconstruction method applied to skin cancer mirror classification. The method uses discrete wavelet transform as the down-sampling operation and adds a multichannel attention mechanism to effectively combine the high-frequency components and low-frequency components after wavelet

decomposition and also effectively utilizes the channel information to enhance the characterization ability of the model; a skin cancer classification model is given, based on the ResNet50 network model, using a combination of 3×3 convolutional kernel and depth-separable convolution and wavelet down-sampling as the input backbone of the model, which reduces the number of parameters while maintaining the perceptual field of the network. Finally, the network weight parameters are initialized on ImageNet using transfer learning, and fine-tuned training is performed on the augmentation HAM10000 dataset to further improve the model classification performance.

The main contributions of the proposed method to the classification of melanoma skin cancer are as follows:

- In the current study, we developed a wavelet down-sampling feature reconstruction method to address the problem of traditional down-sampling feature information loss, introducing a multichannel attention mechanism to effectively combine low-frequency components with high-frequency components, fully utilizing channel information to reveal finer details.
- This paper developed a wavelet down-sampling feature reconstruction method-based convolutional neural network as a feature extractor to classify melanoma skin cancer using skin pathological mirror.
- A data augmentation and hair removal algorithm is used to pre-process the skin pathological mirror dataset HAM1000, and the pre-processing method is tested on the pre-processing HAM10000 dataset using transfer learning. The experimental results show that the proposed method has high accuracy for melanoma skin cancer classification.

2. Related Work

The wavelet transform is capable of focusing on arbitrary details of the signal by multiscale refinement through telescopic translation operations and is widely used in signal processing and pattern recognition [17–19]. In recent years, wavelets have been analyzed by many scholars in combination with convolutional neural networks as emerging down-sampling methods, such as Mallat et al. [20] by linking wavelet transform with average pooling cascade, which preserves the detail information of images to a great extent with shift-invariance features. Zhang et al. [21] introduced multi-level wavelets into convolutional neural networks, which greatly preserves the texture information of images by using the inverse transform of wavelets. Li et al. [22] used the low-frequency component of the discrete wavelet transform output to replace the traditional commonly used down-sampling method and obtained strong robustness and better network performance while confirming the effectiveness of using the low-frequency component of the two-dimensional discrete wavelet to replace the down-sampling layer output in the convolutional neural network. The low-frequency component of the wavelet transform output contains the main feature information of the image, and the high-frequency component contains detailed information such as the texture features of the image [23]. In the existing forms of combining wavelets and convolutional neural networks, mainly the high-frequency components are discarded, and only the low-frequency components are used, or the high and low-frequency components are used in equal proportions. However, both of them have certain disadvantages: if the high-frequency components are discarded, it may cause the loss of important texture information, leading to the problem of poor results in detail recovery and insufficient characterization ability, while the high- and low-frequency components contain very different amounts of information and have different effects on the loss function, so if they are used in equal proportions, it is not reasonable enough and there is also the problem of low utilization of channel information [24].

3. Wavelet Down-Sampling Reconstruction

The input skin mirror image (X) is feature decomposed using a 2D discrete wavelet transform to obtain a low-frequency component (X_{LL}) containing information on the main pathology feature and three high-frequency components containing information on

pathological details: the horizontal component (X_{LH}), the vertical component (X_{HL}) and the diagonal component (X_{HH}). 2D discrete wavelet transform can be formulated as

$$[X_{LL}, X_{LH}, X_{HL}, X_{HH}] = 2D_{DWT}(X) \tag{1}$$

where $(X_{LL}, X_{LH}, X_{HL}, X_{HH}) \in c \times \frac{h}{2} \times \frac{w}{2}$; c is the number of channels of the input skin mirror image; and h and w are the height and width, respectively. The new high-frequency components $X'_{LH}, X'_{HL}, X'_{HH}$, are obtained by summing the low-frequency components with each high-frequency component separately and can be formulated as

$$\begin{cases} X'_{LH} = \alpha_1 X_{LH} + X_{LL} \\ X'_{HL} = \alpha_2 X_{HL} + X_{LL} \\ X'_{HH} = \alpha_3 X_{HH} + X_{LL} \end{cases} \tag{2}$$

where α_i is the absolute value of the high-frequency component to which it belongs, reflecting the severity of the pathology, $i = 1, 2, 3$.

Based on SKNet [25] multi-branch structure selective channel attention mechanism, each new high-frequency component in Equation (2) is integrated to obtain the texture feature F of pathology; then, global average pooling (GAP) operation is performed to obtain the global information s on each channel, and then the fully connected (FC) operation is performed on s to obtain the size of the weight z accounted for by each channel; the normalization function softmax is used to reaggregate the weights Q after normalizing z to obtain the new pathological texture features F' and finally obtain the integrated pathology comprehensive feature M by jumping connection X_{LL} to improve the characterization ability. The whole process is shown in Figure 1. The integrated pathology comprehensive feature M can be formulated as

$$Q = \text{Softmax}(\sigma_{\text{relu}}(\text{BN}(W \cdot \text{GAP}(F_{(i,j)})))) \tag{3}$$

$$F' = Q_1 \times X'_{LH} + Q_2 \times X'_{HL} + Q_3 \times X'_{HH} \tag{4}$$

$$M = \text{concat}([F', X_{LL}]) \tag{5}$$

where $(i, j) \in (h, w)$; $W \in d \times c$; $d = \max(\frac{c}{r}, L)$; r is the scaling ratio; and L is the minimum value of d to control the dimensionality of the output. FC is composed of the batch normalized BN and σ_{relu} activation function; in turn, $M \in 2c \times \frac{h}{2} \times \frac{w}{2}$.

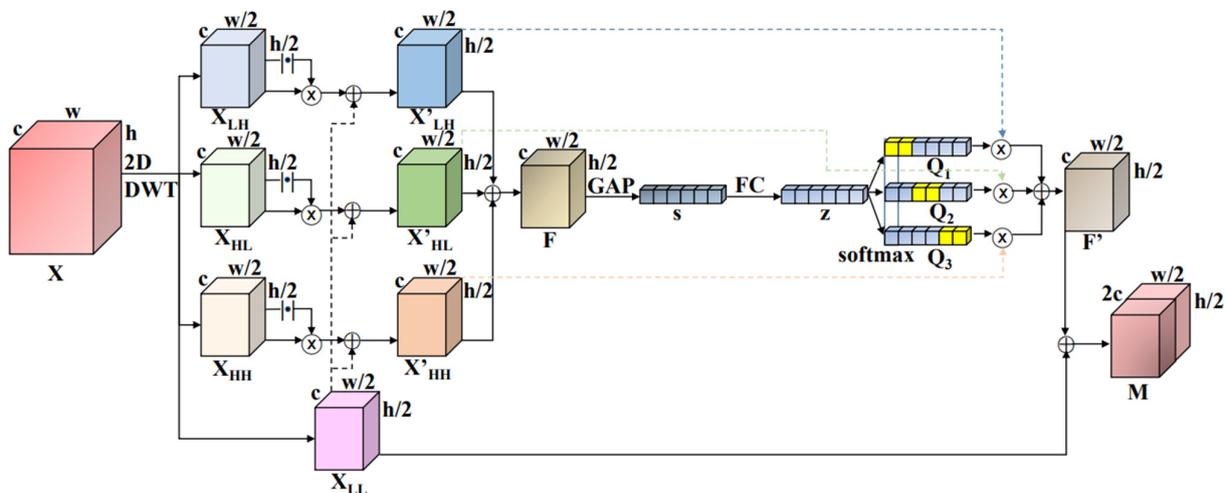


Figure 1. 2D discrete wavelet down-sampling reconstruction.

Finally, the pathology comprehensive feature M is reconstructed by using 1D discrete wavelet for feature decomposition to output the final pathology feature map. Firstly, the comprehensive feature M is downsampled using the reshape operation to obtain a feature map of the form $1 \times n$, and then the 1D discrete wavelet decomposition is performed to obtain the low-frequency component X_L and the high-frequency component X_H . Finally, the weights of the low-frequency component and the high-frequency component are calculated using the selective channel attention mechanism, and the final reconstructed feature rec_X is composed according to the weights; the whole process is shown in Figure 2. The final reconstructed feature rec_X can be formulated as

$$[X_L, X_H] = 1D_{DWT}(M) \tag{6}$$

$$q = \text{Softmax}(\sigma_{\text{relu}}(\text{BN}(W \cdot \text{GAP}(X_L \oplus X_H)))) \tag{7}$$

$$rec_X = q_1 \times X_L + q_2 \times X_H \tag{8}$$

where q_1 and q_2 are the weights of the low-frequency component and the high-frequency component features within each channel, $q_1 + q_2 = 1$, $rec_X \in c \times \frac{h}{2} \times \frac{w}{2}$.

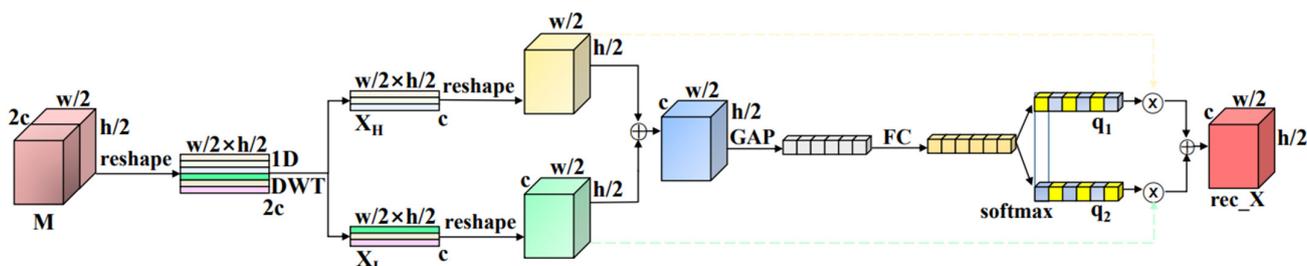


Figure 2. 1D discrete wavelet down-sampling reconstruction.

4. Skin Cancer Classification Model

In this paper, the skin cancer classification model is based on the ResNet50 network model. The input backbone of the ResNet50 network is mainly composed of a 7×7 convolutional kernel and a maximum pooling layer. In this paper, a combination of a 3×3 standard convolution and a depth-separable convolution [26] is used to replace the 7×7 convolutional kernel, and wavelet down-sampling is used to replace the maximum pooling operation, which reduces the computational effort while deepening the network depth during convolution. The depth-separable convolution and input backbone structures are shown in Figures 3 and 4.

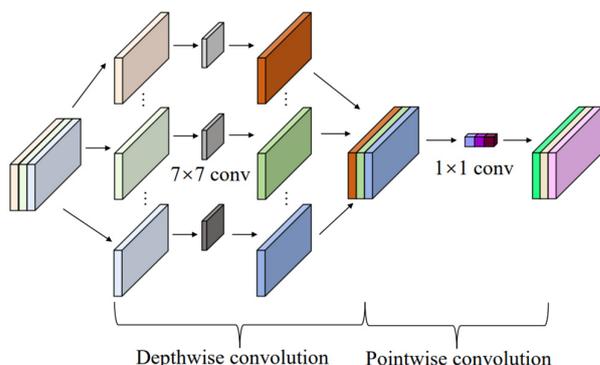


Figure 3. Depth-separable convolution structure.

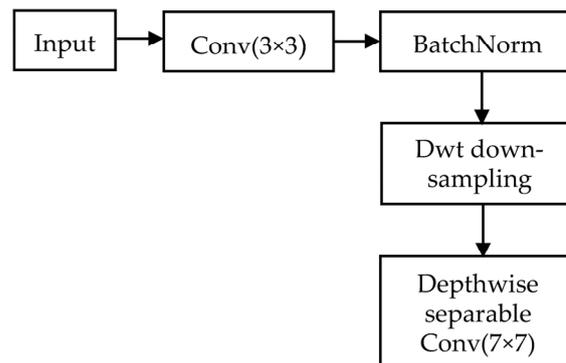


Figure 4. Input backbone structure.

As shown in Figures 3 and 4, the wavelet down-sampling output feature map is fed into the depth-separable convolution, and three convolution kernels of size $7 \times 7 \times 1$ are used. Each convolution kernel does convolution for one input channel only, and the output feature map of size $224 \times 224 \times 3$ is obtained. Then, using 64 point-by-point convolutions with $1 \times 1 \times 3$ convolution kernels, each convolution kernel will obtain a mapping of size $224 \times 224 \times 1$, resulting in an output feature map of size $224 \times 224 \times 64$. The improved input backbone not only maintains the original perceptual field but also reduces the number of parameters by 32.4% and deepens the depth of the network.

The Hard-Swish function [27] is used instead of the ReLU function in the backbone network to avoid the “dead ReLU” problem and the problem that the output is reduced to zero when the input is used. The Hard-Swish function is unbounded, smooth and non-monotonic and can be formulated as

$$\text{Hard-Swish}(x) = \begin{cases} 2x, & x \geq \frac{5}{2\lambda} \\ 0.4\beta x^2 + x, & -\frac{5}{2\lambda} < x < \frac{5}{2\lambda} \\ 0, & x \leq -\frac{5}{2\lambda} \end{cases} \quad (9)$$

In Equation (9), λ is a trainable parameter or a custom parameter, and the output is more likely to be set to 0 when x is smaller rather than directly equal to 0 as in the ReLU function, which not only increases the randomness but also increases the dependence on the input, effectively improving the robustness of the model. To avoid excessive activation and BN layers that degrade the model performance [28], the ReLU function in the residual block is replaced by the Hard-Swish function; the activation and BN layers after the first convolution are discarded; the activation function before the output of the residual block is moved to the input; and then the wavelet down-sampling method is used to improve the learning ability of the model. The improved residual block is shown in Figure 5.

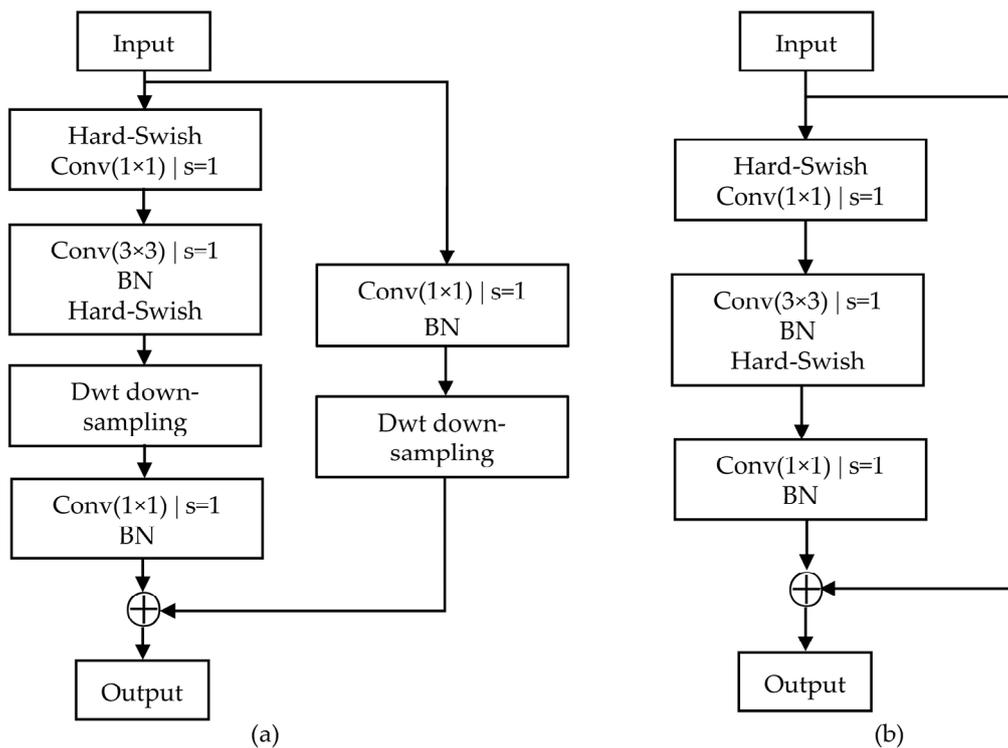


Figure 5. 1D discrete wavelet down-sampling reconstruction: (a) Conv Block; (b) Identity Block.

5. Experiments and Analysis of Results

5.1. Dataset and Pre-Processing

The publicly available skin lesion dataset HAM10000 [29] was used to perform experiments on skin cancer classification models. The dataset consists of seven skin diseases, namely benign keratosis (BKL), actinic keratosis and intra-epithelial carcinoma (AKIEC), dermatofibrosarcoma (DF), melanocytic nevus (NV), vascular lesion (VASC), melanoma (MEL), and basal cell carcinoma (BCC), with a total of 10,015 images, as shown in Figure 6.

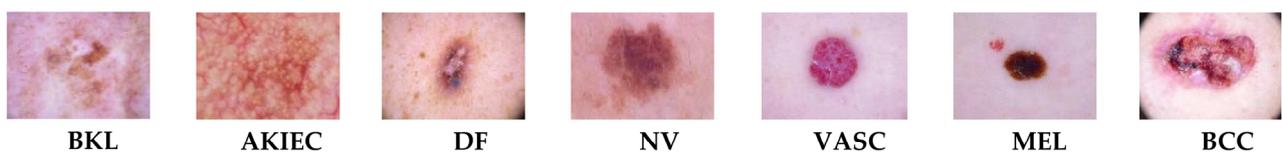


Figure 6. Example of HAM10000 dataset.

Aiming at the problem that hair occlusion in the skin mirror image can affect the diagnosis results, this paper gives a hair removal algorithm [30] to automatically remove the hair occlusion in the skin mirror image and restore the information of the hair occlusion area, the Algorithm 1 shows the hair removal algorithm and Figure 7 shows the comparison before and after hair removal in the skin mirror image.

Algorithm 1: hair removal algorithm

Input: skin mirror image (img)

Output: restoration skin mirror image (img1)

1. use morphological closed top-hat operator to enhance the hairs with varying strength in the skin mirror image img.
 2. extract the effective measure of hairs based on the defined extension characteristic function to describe the extension state of the stripe-like connected region.
 3. the image restoration technique using partial differential equations is used to restore the hair-obscured areas to obtain the skin mirror image img1 after hair removal.
-



Figure 7. Comparison of before and after hair masking restoration. (a) Before restoration. (b) After restoration.

Since the sample distribution in the HAM10000 dataset is extremely uneven, this paper uses data augmentation methods including rotation, cropping and random flipping [31] to ensure the data category balance and enhance the generalization ability of the model. The augmentation dataset consists of 16,002 samples, which are divided into two independent and non-intersecting datasets, the training set and the test set, with a division ratio of 8:2.

5.2. Experimental Environment and Evaluation Indicators

The experiments were run on a Windows 10 64-bit platform environment with a 12th generation i7-12700H, 32G RAM, NVIDIA GTX3080Ti discrete graphics card with 12G video memory and Pytorch as the deep learning framework.

AC, F1-Score was used as the evaluation indicators of the model and defined as follows:

$$AC = \frac{TP + TN}{TP + FP + TN + FN} \quad (10)$$

$$F1 = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (11)$$

where AC is the classification accuracy; F1-Score is the summed average of accuracy and completeness; and the higher the F1-Score value is, the better the classification performance of individual categories is. TP, TN, FP and FN represent the number of samples with true positives, true negatives, false positives and false negatives, respectively.

5.3. Comparative Experiments

To verify the effectiveness of the proposed method and to also determine that the improved input backbone and wavelet down-sampling methods have improved the model performance, ablation experiments were conducted on the test set, and the experimental results are shown in Table 1.

Table 1. Ablation experiments of the proposed method.

Base Model	Input Backbone	Residual Module	Data Augmentation	AC/%	F1/%
ResNet50			✓	90.48	90.69
	✓		✓	93.45	93.67
		✓	✓	94.51	94.64
	✓	✓		93.94	94.03
	✓	✓	✓	95.28	95.39

As can be seen from Table 1, the model classification performance is improved by using the improved input backbone in this paper, and the classification accuracy is improved by 2.97%. However, the classification performance of the model decreases significantly when data augmentation is not used to balance the data categories. It is verified that the data augmentation method can effectively improve the classification performance of the network and the effectiveness of the method proposed in this paper.

The performance of the improved ResNet50 model in the paper is compared with five mainstream deep learning network models, AlexNet [32], VGG19 [33], MobileNet-V2 [34], DenseNet-121 [35] and EfficientNet-B0 [36], all of which are pre-trained on the ImageNet dataset. Pre-training is performed to initialize some of the weight parameters, and then experiments are conducted on the augmentation HAM10000 dataset, and the experimental results are shown in Table 2.

Table 2. Performance comparison of different classification models on the HAM1000 dataset.

Model	AC/%	F1/%
AlexNet	87.81	88.13
VGG19	90.34	90.61
MobileNet-V2	88.23	88.86
DenseNet-121	92.93	93.14
EfficientNet-B0	93.62	93.78
Proposed model	95.84	95.96

From Table 2, the classification accuracy and F1 values of the improved model in this paper on the HAM10000 dataset are better than those of other network models, which are 95.84% and 95.96%, respectively, among which AlexNet has the lowest accuracy and F1 values, which are 87.81% and 88.13%, respectively, and EfficientNet-B0 has the highest accuracy and F1 values except for the model in this paper, which are 93.62% and 93.78%, respectively.

For the HAM10000 dataset, many scholars have proposed research methods, and the research methods proposed in this paper were compared with the existing methods, and the comparison results are shown in Tables 3 and 4.

Table 3. Comparison with existing methods on the HAM1000 dataset (AC, F1 and Parameters).

References	AC/%	F1/%	Parameters/MB
ResNet50 + SA [37]	91.55	91.30	91.2
CNN ensembles [11]	93.09	-	-
IRv2-SA [37]	93.47	93.65	181.3
Loss balance + ensemble [38]	92.60	-	-
Low-cost augmentation + CNN [39]	95.79	86.14	42.0
FixCaps [40]	96.49	-	1.86
Our method	95.84	95.96	60.9

Table 4. Comparison with existing methods on the HAM1000 dataset (AUC).

References	AUC	Category						
		AKIEC	BCC	BKL	DF	MEL	NV	VASC
ResNet50 + SA [37]	0.980	0.981	0.996	0.964	0.971	0.973	0.979	0.999
CNN ensembles [11]	0.929	0.902	0.934	0.885	0.968	0.925	0.951	0.941
IRv2-SA [37]	0.985	0.981	0.998	0.982	0.973	0.974	0.984	1.000
Loss balance + ensemble [38]	0.941	0.919	0.947	0.908	0.980	0.931	0.960	0.942
Low-cost augmentation + CNN [39]	0.976	0.988	0.989	0.968	0.972	0.943	0.970	0.995
FixCaps [40]	-	-	-	-	-	-	-	-
Our method	0.982	0.983	0.992	0.975	0.976	0.971	0.983	0.992

In Table 3, our method was compared with six previously reported melanoma skin cancer classification methods, among which FixCaps method achieved the highest classification accuracy of 96.49% because its method discarded the down-sampling operation and used the dynamic routing algorithm, which not only directly avoided the problem of feature information loss caused by traditional down-sampling but also could read useful feature information to the maximum extent. In addition, the other five methods and our

method all contain down-sampling operation, among which our method achieves the highest classification accuracy and F1 value of 95.84% and 95.96%, respectively. In terms of the number of model parameters, FixCaps has a minimum of 1.86 MB, and IRv2-SA has a maximum of 181.3 MB, while the model in our method is in the middle with 60.9 MB.

In Table 4, we compared the AUC scores of our method with previous melanoma skin cancer classification methods, and the AUC score is the golden indicator of classification model performance; the higher the AUC score is, the better the model performance is, and the AUC score is also the ranking indicator of melanoma skin cancer classification ranking. From Table 4, we can see that the highest AUC score of the IRv2-SA method is 0.985; the lowest AUC score of CNN ensembles method is 0.929; and the AUC score of our method is 0.982, which is in the second place.

Combined with Tables 3 and 4, the proposed method in this paper has excellent performance in classifying melanoma skin cancer, which exceeds most of the previously reported melanoma skin cancer methods. Figure 8 shows the confusion matrix at the best classification case. Figure 9 shows the ROC curve that is plotted with true positive rate against the false positive rate of each lesion category, individually, and Table 5 shows the corresponding classification metrics. As seen in Table 5, the highest classification accuracy of 99.16% was achieved for dermatofibrosarcoma (DF), and the lowest classification accuracy of 93.31% was achieved for benign keratosis (BKL).

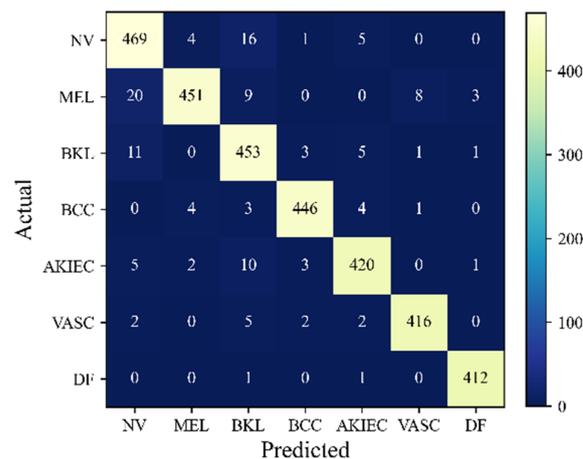


Figure 8. Best classification confusion matrix.

Table 5. Classification indicators.

Category	Precision/%	Recall/%	F1-Score/%
NV	92.50	94.75	93.61
MEL	97.83	91.85	94.75
BKL	91.15	95.57	93.31
BCC	98.02	97.38	97.70
AKIEC	96.11	95.24	95.67
VASC	97.65	97.42	97.54
DF	98.80	99.52	99.16

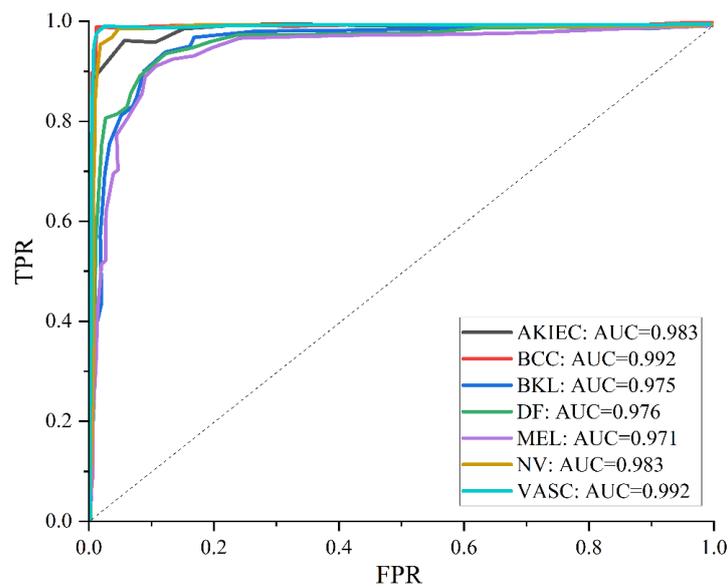


Figure 9. ROC curve of the best classification.

6. Conclusions

This paper introduces the implementation and application of wavelet analysis as down-sampling in convolutional neural networks to solve the problem of insufficient network characterization ability due to the loss of feature information by down-sampling in skin cancer classification and to solve the problem of low channel information utilization and inability to effectively combine low-frequency and high-frequency components by introducing the multichannel attention mechanism into wavelet analysis. Firstly, a combination of depth-separable convolution and 3×3 standard convolution and wavelet down-sampling is used as the input backbone of the ResNet50 network, and then the Hard-Swish function and wavelet down-sampling are used to improve the residual module of the ResNet50 network, which not only ensures the perceptual field of the original network but also reduces the number of parameters and the feature loss of down-sampling and improves the expression ability of the network. The experimental results show that the method achieves excellent classification results on the augmentation HAM10000 dataset, and the classification accuracy reaches 95.84% after combining with migration learning, which effectively improves the skin cancer diagnosis accuracy.

Author Contributions: Methodology, Q.-e.W., Y.Y., X.Z.; writing—original draft, Q.-e.W., Y.Y.; writing—review and editing, Q.-e.W., X.Z.; funding acquisition, Q.-e.W. All authors have read and agreed to published version of the manuscript.

Funding: Qing-e Wu for funding this work through the Key Science and Technology Program of Henan Province number (222102210084), Key Science and Technology Project of Henan Province University number (23A413007), respectively.

Data Availability Statement: Dataset used in the experiment is available on web publicly at <https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000> (accessed on 17 January 2023).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer statistics, 2019. *CA A Cancer J. Clin.* **2019**, *69*, 7–34. [[CrossRef](#)] [[PubMed](#)]
2. Koh, H.K. Melanoma screening: Focusing the public health journey. *Arch. Dermatol.* **2007**, *143*, 101–103. [[CrossRef](#)] [[PubMed](#)]
3. Zunair, H.; Hamza, A.B. Melanoma detection using adversarial training and deep transfer learning. *Phys. Med. Biol.* **2020**, *65*, 135005. [[CrossRef](#)] [[PubMed](#)]
4. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)]

5. Qiu, C.H.; Huang, C.F.; Xia, S.R.; Kong, D.X. Application review of artificial intelligence in medical images aided diagnosis. *Space Med. Med. Eng.* **2021**, *34*, 407–414.
6. Chen, Y.D.; Zhang, Q.; Lan, L.; Peng, L.; Yin, J. A Review of Deep Convolutional Neural Networks in Medical Image Segmentation. *Chin. J. Health Inform. Manag.* **2021**, *18*, 278–284.
7. Codella, N.; Cai, J.; Abedini, M.; Garnavi, R.; Halpern, A.; Smith, J.R. Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images. In *Machine Learning in Medical Imaging: 6th International Workshop, MLMI 2015, Proceedings of the MICCAI 2015, Munich, Germany, 5 October 2015; Held in Conjunction with MICCAI 2015; Springer International Publishing: Cham, Switzerland, 2015; pp. 118–126.*
8. Pomponiu, V.; Nejati, H.; Cheung, N.M. Deepmole: Deep neural networks for skin mole lesion classification. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 2623–2627.
9. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **2017**, *542*, 115–118. [[CrossRef](#)]
10. Khan, M.A.; Sharif, M.I.; Raza, M.; Anjum, A.; Saba, T.; Shad, S.A. Skin lesion segmentation and classification: A unified framework of deep neural network features fusion and selection. *Expert Syst.* **2022**, *39*, e12497. [[CrossRef](#)]
11. Ahmed, S.A.A.; Yanikoğlu, B.; Göksu, Ö.; Aptoula, E. Skin lesion classification with deep CNN ensembles. In Proceedings of the 2020 28th Signal Processing and Communications Applications Conference (SIU), Gaziantep, Turkey, 5–7 October 2020; pp. 1–4.
12. Abayomi-Alli, O.O.; Damasevicius, R.; Misra, S.; Maskeliunas, R.; Abayomi-Alli, A. Malignant skin melanoma detection using image augmentation by oversampling in nonlinear lower-dimensional embedding manifold. *Turk. J. Electr. Eng. Comput. Sci.* **2021**, *29*, 2600–2614. [[CrossRef](#)]
13. Zeiler, M.D.; Fergus, R. Stochastic pooling for regularization of deep convolutional neural networks. In Proceedings of the 1st International Conference on Learning Representations, ICLR 2013, Scottsdale, AZ, USA, 2–4 May 2013.
14. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
15. Zhao, G.; Wang, J.; Zhang, Z. Random Shifting for CNN: A Solution to Reduce Information Loss in Down-Sampling Layers. In Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), Melbourne, Australia, 25 August 2017; pp. 3476–3482.
16. Jiang, Z.T.; Qin, J.Q.; Zhang, S.Q. Parameterized pooling convolution neural network for image classification. *Acta Electron. Sin.* **2020**, *48*, 1729.
17. Daubechies, I. Ten Lectures on Wavelets. *Comput. Phys.* **1992**, *6*, 697. [[CrossRef](#)]
18. Hong, J.; Wang, Z.; Qu, C.; Zhou, Y.; Shan, T.; Zhang, J.; Hou, Y. Investigation on overcharge-caused thermal runaway of lithium-ion batteries in real-world electric vehicles. *Appl. Energy* **2022**, *321*, 119229. [[CrossRef](#)]
19. Hong, J.; Zhang, H.; Xu, X. Thermal fault prognosis of lithium-ion batteries in real-world electric vehicles using self-attention mechanism networks. *Appl. Therm. Eng.* **2023**, *226*, 120304. [[CrossRef](#)]
20. Bruna, J.; Mallat, S. Invariant scattering convolution networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1872–1886. [[CrossRef](#)]
21. Liu, P.; Zhang, H.; Zhang, K.; Lin, L.; Zuo, W. Multi-level wavelet-CNN for image restoration. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 773–782.
22. Li, Q.; Shen, L.; Guo, S.; Lai, Z. Wavelet integrated CNNs for noise-robust image classification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 7245–7254.
23. Xu, D. Application of wavelet transform-based image processing techniques. *J. Soochow Univ. (Nat. Sci.)* **2002**, *1*, 45–49.
24. Xu, K.; Qin, M.; Sun, F.; Wang, Y.; Chen, Y.K.; Ren, F. Learning in the frequency domain. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 1740–1749.
25. Li, X.; Wang, W.; Hu, X.; Yang, J. Selective kernel networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 510–519.
26. Dang, L.; Pang, P.; Lee, J. Depth-wise separable convolution neural network with residual connection for hyperspectral image classification. *Remote Sens.* **2020**, *12*, 3408. [[CrossRef](#)]
27. Avenash, R.; Viswanath, P. Semantic Segmentation of Satellite Images using a Modified CNN with Hard-Swish Activation Function. In Proceedings of the International Joint Conference on Computer Vision, Imaging and Computer Graphics (VISIGRAPP), Prague, Czech Republic, 25–27 February 2019; pp. 413–420.
28. Bronskill, J.; Gordon, J.; Requeima, J.; Nowozin, S.; Turner, R. Tasknorm: Rethinking batch normalization for meta-learning. In Proceedings of the International Conference on Machine Learning, Virtual, 13–18 July 2020; pp. 1153–1164.
29. Tschandl, P.; Rosendahl, C.; Kittler, H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci. Data* **2018**, *5*, 180161. [[CrossRef](#)]
30. Xie, F.; Qin, S.; Jiang, Z.; Meng, R. Unsupervised repair of hair-occluded information for skin melanoma image. *Chin. J. Sci. Instrum.* **2009**, *30*, 699–705.
31. Youwen, G.; Benjun, Z.; Xiaofei, H. Research on image recognition of convolution neural network based on data augmentation. *Comput. Technol. Dev.* **2018**, *28*, 62–65.
32. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [[CrossRef](#)]

33. Bansal, M.; Kumar, M.; Sachdeva, M.; Mittal, A. Transfer learning for image classification using VGG19: Caltech-101 image data set. *J. Ambient. Intell. Humaniz. Comput.* **2021**, *14*, 3609–3620. [[CrossRef](#)] [[PubMed](#)]
34. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4510–4520.
35. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
36. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 10–15 June 2019; Volume 97, pp. 6105–6114.
37. Gao, M. Soft Attention Improves Skin Cancer Classification Performance. In *Interpretability of Machine Intelligence in Medical Image Computing, and Topological Data Analysis and Its Applications for Medical Data: 4th International Workshop, Proceedings of the iMIMIC 2021, and 1st International Workshop, Strasbourg, France, 27 September 2021*; TDA4MedicalData 2021, Held in Conjunction with MICCAI 2021; Springer Nature: Berlin, Germany, 2021; Volume 12929.
38. Gessert, N.; Nielsen, M.; Shaikh, M.; Werner, R.; Schlaefer, A. Skin lesion classification using ensembles of multi-resolution EfficientNets with meta data. *MethodsX* **2020**, *2020*, 100864. [[CrossRef](#)] [[PubMed](#)]
39. Shen, S.; Xu, M.; Zhang, F.; Shao, P.; Liu, H.; Xu, L.; Zhang, C.; Liu, P.; Yao, P.; Xu, R.X. Erratum to “A Low-Cost High-Performance Data Augmentation for Deep Learning-Based Skin Lesion Classification”. *BME Front.* **2023**, *4*, 0011. [[CrossRef](#)]
40. Lan, Z.; Cai, S.; He, X.; Wen, X. FixCaps: An improved capsules network for diagnosis of skin cancer. *IEEE Access* **2022**, *10*, 76261–76267. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.