Applied Soft Computing Journal 99 (2021) 106881

Contents lists available at ScienceDirect

Applied Soft Computing Journal

journal homepage: www.elsevier.com/locate/asoc

Cascade knowledge diffusion network for skin lesion diagnosis and segmentation

Qiangguo Jin [a,c], Hui Cui [b], Changming Sun [c], Zhaopeng Meng [a,d], Ran Su [a,*]

[a] School of Computer Software, College of Intelligence and Computing, Tianjin University, Tianjin, China
[b] Department of Computer Science and Information Technology, La Trobe University, Melbourne, Australia
[c] CSIRO Data61, Sydney, Australia
[d] Tianjin University of Traditional Chinese Medicine, Tianjin, China

ARTICLE INFO

Article history:
Received 23 July 2020
Received in revised form 29 September 2020
Accepted 1 November 2020
Available online 6 November 2020

Keywords:
Skin lesion diagnosis
Skin lesion segmentation
Cascade feature aggregation
Knowledge diffusion

ABSTRACT

Accurate diagnosis and segmentation of skin lesion is critical for early detection and diagnosis of skin cancer. Recent multi-task learning methods require expensive annotations for skin lesion analysis while single-task driven models cannot fully utilize the potential knowledge. The aim of this study is to utilize the neglected knowledge by a flexible architecture in dermoscopy skin lesion classification and segmentation. In this work, we propose a cascade knowledge diffusion network (CKDNet) to transfer and aggregate knowledge learnt from different tasks to simultaneously boost the performances of classification and segmentation. CKDNet consists of a sequence of coarse-level segmentation, classification, and fine-level segmentation networks. We design two novel feature entanglement modules, Entangle-Cls and Entangle-Seg, for classification and segmentation. The Entangle-Cls module aggregates the diffused features from initial segmentation to drive the classification network's attention to image regions relevant to the disease. The Entangle-Seg module integrates the cascaded context knowledge learnt from classification to benefit fine-level segmentation, especially at uncertain boundaries. The entanglement modules can adaptively control the knowledge that can be diffused from one task to another, which avoids the empirical selection of weights for different learning tasks compared to other multi-task methods. We perform extensive evaluations and comparisons with state-of-the-art methods on skin lesion classification and segmentation with challenge datasets, ISIC2017 and ISIC2018. Our CKDNet demonstrated superior performance without using any ensemble approaches or any external datasets. The effectiveness of each component and loss functions are demonstrated by interpretable results using class activation maps (CAM), t-SNE, and classification and segmentation results.

1. Introduction

Skin cancer is one of the most common cancer worldwide [1]. For instance, there are over 5 million new cases reported annually in the United States [1]. The survival rate from melanoma, a severe type of skin cancer, can be increased dramatically from 14% to over 99% if this type of skin cancer can be detected and diagnosed in its early stage [2]. Dermoscopy images are widely used in skin cancer diagnosis. The typical tasks related to computer-aided skin lesion analytics are classification and segmentation. The classification task aims to predict different types of skin lesions such as melanoma, melanocytic nevus, and basal cell carcinoma by using image features. The segmentation task aims at detecting and delineating the lesion boundaries automatically. Early-stage skin cancer diagnosis, however, is a challenging task. The primary reason is that there are large variations of appearances in size, shape, and color as well as multiple types of textures mixed together in melanoma. For instance, melanoma and non-melanoma skin lesions have high visual similarities. Besides, the boundaries between the lesion of interest and its surrounding tissues on dermoscopy images can be indistinct, which is especially the case with the presence of artifacts and noise such as hair, blood, and veins [3]. Advanced image processing and learning methods are of great importance in automatic skin lesion classification and segmentation from images. Thus, the aim of this study is to build models to automatically classify and segment skin lesions in dermoscopy images.

1.1. Related work

Deep learning models, especially the convolutional neural networks (CNN), have attracted intensive research interests in skin lesion classification and segmentation [4–7]. Existing methods

[*] Corresponding author.
E-mail address: ran.su@tju.edu.cn (R. Su).

https://doi.org/10.1016/j.asoc.2020.106881
1568-4946/© 2020 Elsevier B.V. All rights reserved.

for skin lesion classification and segmentation include two main categories: models designed for a single task or multi-tasks.

1.1.1. CNN for single task skin lesion classification or segmentation

To classify the type of skin lesions using dermoscopy images, Barata et al. identified relevant regions in dermoscopy images by an attention module to classify skin lesions [8]. Gessert et al. [4] proposed a patch-based attention architecture with a diagnosis-guided loss function. The proposed method improved the skin lesion classification by extracting the global contexts between low- and high-resolution patches. Zhang et al. [5] proposed an attention residual learning CNN model (ARL-CNN) to leverage multiple ARL blocks to tackle the challenge of data insufficiency, inter-class similarities, and intra-class variations. Harangi et al. [9] used an ensemble approach where they fused and averaged the prediction results from four different deep neural networks as the final diagnosis result. This approach, however, is complicated and may not be suitable for a general real-world application. There are also methods exploring to boost the performance of classification by segmentation. For instance, Yu et al. [10] proposed to firstly segment lesions from the whole dermoscopy images and then use the cropped regions as the input to a classification network. By doing such, the insufficiency of training data is alleviated. Although there are methods [10,11] investigating to boost classification by segmentation, the models using classification results to enhance segmentation performance are under-developed.

For skin lesion segmentation tasks, Xue et al. [12] proposed SegAN, a generative adversarial networks (GAN) based framework, for skin lesion segmentation. They captured the global and local features by long- and short-range connections and introduced a multi-scale L_1 loss function to optimize the network. Yuan et al. [6] combined a 19-layer deep fully-connected neural network (FCN) using a set of strategies for efficient and effective segmentation of skin lesions from raw images. The model designed by Sarker et al. [13] integrated skip-connections, dilated residual and pyramid pooling modules for skin lesion segmentation. Goyal et al. [14] proposed a model with ensembled Mask R-CNN and DeepLabV3+ [15], and demonstrated improved performance over the individually trained FrCN, FCNs, U-Net, and SegNet. In order to track the evolutions of skin regions of interests, Navarro et al. [7] presented an architecture for both automatic skin lesion segmentation and registration based on superpixel techniques. There are also methods which embedded multi-step segmentation networks in one framework. For instance, González-Díaz [11] demonstrated that segmentation results could be improved when two separate segmentation networks for lesion and dermoscopic structures are embedded together in one framework

1.1.2. Multi-task model for skin lesion analysis

Recently, there are increasing research interests in multi-task learning. The primary reason is that skin lesion classification and segmentation are correlated tasks. Single-task architectures cannot leverage the features learnt from different tasks to benefit each other.

Multi-task learning models formulate universal knowledge from the branches of different tasks. The universal knowledge could conversely improve the generalization of each branch by sharing the common information across various tasks. For instance, Chen et al. [16] proposed a gate function to control the message transmission in a classification branch and a segmentation branch. The experimental results showed that the performances of both of the branches were improved. Song et al. [3] introduced a three-phase joint training strategy to enhance the skin lesion diagnosis, detection, and segmentation. The proposed approach achieved promising segmentation performance on the ISBI 2016 and ISIC 2017 datasets. Liao et al. [17] constructed a deep multi-task learning framework to jointly optimize skin lesion classification and body location classification. Their experimental results demonstrated that the joint learning model was more robust than a standalone network. Murabayashi et al. [18] proposed a hybrid semi-supervised and multi-task learning method for melanoma diagnosis. They demonstrated that the combined semi-supervised and multi-task learning could achieve competitive results even with a limited amount of training data. Chen et al. [19] combined lesion attribute classification and segmentation in a multi-task U-Net model to detect lesion attributes of melanoma automatically.

Even though multi-task learning has attracted increased attention in skin lesion analysis, there are ongoing challenging problems to be solved. Most of the multi-task based lesion analysis networks use a shared encoder to extract common features and different decoders for multiple tasks. The fixed encoder, however, lacks of learning efficiency and may limit the extendibility of the model. Firstly, it is difficult to set the weights to balance the loss functions of different tasks in such an architecture. A common approach in existing work is to choose empirical values based on trials and experiments. The experiments guided parameter selection approach, however, requires tremendous time for training and testing. More importantly, hand-crafted weight values cannot be applied to other work generally. Secondly, most of the supervised multi-task deep learning methods [3,16,20] rely on large datasets and expensive annotations to support the training of multi-tasks. Such datasets are challenging to acquire, especially for medical images, which limit the development of multi-task learning algorithms. Besides, imbalanced classification and segmentation datasets, and class labels may also affect the learning performance.

Knowledge transfer and diffusion strategies demonstrated the capacity to utilize the neglected knowledge [21–25]. For instance, Li et al. [21] took advantage of the information from a station-intensive source and proposed a knowledge adaption module to transfer the relevant knowledge to station-sparse sources. Aldieri et al. [22,23] developed the uncentered correlation index between different technology sectors to control the knowledge diffusion process between the sectors. Kuzina et al. [24] addressed the problem of knowledge transfer between medical datasets via the Bayesian approach with implicit generative prior in the space of the convolution filters. Fernando et al. [25] proposed a two-step transfer learning based training process for skin lesion classification.

Although the knowledge transfer and diffusion strategies have attracted attention in some research applications, the direct adaption of the conventional strategy of knowledge diffusion between multi-data sources to skin lesion analysis is a challenging issue because skin lesion classification and segmentation are two independent tasks.

1.2. Contributions

To address the challenging issues of architecture extendibility and weighting balance in multi-task learning, and to fully utilize the potential knowledge, we propose a novel cascade knowledge diffusion network (CKDNet). Different from approaches in [3,16,20], our model uses different sub-networks for various tasks instead of a shared encoder. To transfer and aggregate knowledge learnt from different tasks to benefit the mutual goal, we propose the concept of feature entanglement, which is achieved by the novel entanglement modules.

As shown in Fig. 1, CKDNet consists of a sequence of three sub-networks for initial segmentation, classification, and fine-level segmentation. The first segmentation network is to obtain

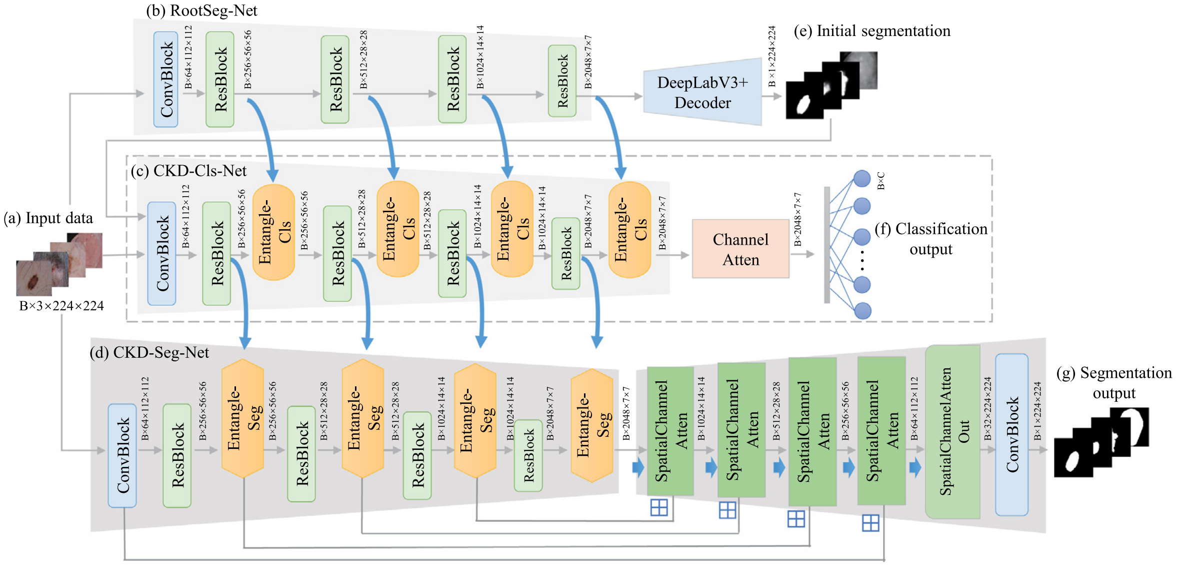


Fig. 1. Illustration of the proposed cascade knowledge diffusion network (CKDNet). CKDNet consists of a sequence of (b) rooted segmentation network (RootSeg-Net) for initial coarse-level segmentation, (c) CKD classification network (CKD-Cls-Net) for lesion classification and (d) CKD segmentation network (CKD-Seg-Net) for fine-level segmentation. We use the pre-trained ResNet101 as the backbone for the encoders in the three sub-networks. RootSeg-Net consists of an encoder and a DeepLabV3+ decoder. CKD-Cls-Net consists of an encoder with novel entanglement module (Entangle-Cls) to aggregate the context knowledge cascaded from RootSeg-Net, and a channel attention (ChannelAtten) block before final classification. CKD-Seg-Net is composed of an encoder with entanglement modules (Entangle-Seg) to integrate the knowledge learnt from classification, a decoder with spatial and channel attention (SpatialChannelAtten) to boost fine-level segmentation, and a SpatialChannelAtten Out block. The details of the Entangle-Cls, ChannelAtten, Entangle-Seg, SpatialChannelAtten, and SpatialChannelAtten Out blocks are given in Figs. 2, 3, and 4. The output feature size of each block is given in batch size × channel size × height × width (B × C × H × W) format.

coarse lesion segmentation and acquire coarsely-located features for later feature entanglement in classification. The classification network takes an image and its initial segmentation as input to emphasis the attention to image regions relevant to disease. The two major components in the classification network include a novel entanglement module, Entangle-Cls, to aggregate the diffused coarsely-located features from initial segmentation, and a channel attention block to enhance the semantic feature representation for classification. Finally, the informative context knowledge learnt by the encoder during classification is passed on to the final segmentation network using another entanglement module, Entangle-Seg, to benefit the fine-level segmentation.

The contributions on the proposed CKDNet are as below: We propose entanglement modules to cascade knowledge in a framework with sub-networks to simultaneously boost the performance of each task. The entanglement modules can adaptively control the knowledge that can be diffused from one task to another, which avoids the empirical selection of weighting factors for different learning tasks compared to other multi-task methods. In addition, the entanglement modules and the diffusion strategy can be applied to other task-specific learning modules, while the encoder of each sub-network can be flexibly replaced by other backbones for various applications. We also design an effective loss function based on the Dice loss and the Focal loss to alleviate the class imbalance problem for skin image analysis. Lastly, experimental results suggest that our model improves classification and segmentation performances without using external datasets or ensembled approaches. We report extensive evaluations and comparisons with state-of-the-art (SOTA) methods on datasets from skin lesion classification and segmentation challenges, ISIC2017 and ISIC2018. We further perform comprehensive experiments to analyze the effectiveness of the proposed entanglement components, attention module, and loss function using class activation map (CAM), t-SNE, and classification and segmentation results.

2. Materials and methodology

2.1. Materials

We use the 2017 and 2018 skin lesion challenge datasets to validate the performance of the proposed knowledge diffusion learning model. The 2017 dataset is used to evaluate the performance of classification and segmentation when an image sample has both classification and segmentation annotations. The 2018 dataset is used to demonstrate the generability of our model when there is only one type of annotation available for an image sample.

2.1.1. ISIC2017 skin lesion challenge dataset (ISIC2017)

Each of the images in ISIC2017 [26] is associated with a disease type and manual segmentation of the lesion. There are three types of lesions including melanoma, nevus, and seborrheic keratosis. Manual segmentation is used as the ground truth (GT). There are 2000 training samples, 150 validation samples, and 600 testing samples. The size of the images in the dataset varies from 453 × 679 to 4499 × 6748 pixels. The lesion diagnosis task has two independent binary image classification tasks. The first binary classification task aims to distinguish between (a) melanoma and (b) nevus and seborrheic keratosis, while the second task classifies (a) seborrheic keratosis and (b) nevus and melanoma. The proportions of the lesion types in the training dataset are 374 melanomas, 254 seborrheic keratosis, and 1372 nevus. Similar proportions are found in the test dataset.

2.1.2. ISIC2018 skin lesion challenge dataset (ISIC2018)

ISIC2018 [27,28] provides skin images for two independent tasks, classification and segmentation. For the classification task, there are 10,015 training images. All the images are classified into one of the following 7 categories: melanoma, melanocytic nevus, basal cell carcinoma, actinic keratosis/bowen's disease, benign keratosis, dermatofibroma, and vascular lesion. The proportions of aforementioned lesion classes are 1113, 6705, 514, 327, 1099, 115, 142 respectively in the training dataset. For the lesion segmentation task, there are 2594 images and the corresponding GT. The testing dataset is composed of 1000 images. The labels of the test data in both tasks are not publicly available. It is worth noting that the ISIC2018 dataset is used for validating the generability of our proposed model.

2.2. Hypothesis and CKDNet framework

We hypothesize that aggregating knowledge learnt from classification and segmentation can simultaneously boost the performances of the two tasks. Initial segmentation of the lesion could drive a classification network's attention to image regions relevant to the disease, and the high-level context knowledge learnt during the classification process could benefit a fine-level segmentation. To cascade knowledge effectively, we propose the concept of feature entanglement, which is achieved by new entanglement modules, to diffuse knowledge in the following sub-networks.

The overview of the proposed CKDNet model is given in Fig. 1. The CKDNet consists of a sequence of three sub-networks for initial coarse-level segmentation, lesion classification, and fine-level segmentation. We define the coarse-level segmentation branch as a rooted segmentation network (RootSeg-Net), and propose two novel architectures, CKD-Cls-Net and CKD-Seg-Net, for classification and fine-level segmentations. As RootSeg-Net is trained for initial segmentation, the encoder in RootSeg-Net can pay attention to the lesion region of interest (ROI), and extract context information. We cascade the context information to CKD-Cls-Net by a unique entanglement module, Entangle-Cls, to benefit the classification task. Similarly, the knowledge learnt by the encoder in CKD-Cls-Net is passed on to the final CKD-Seg-Net by another entanglement component, Entangle-Seg. In this work, we use pre-trained ResNet101 [29] as encoders for all the three sub-networks. Other backbones such as VGG [30] can be easily integrated into our model. As shown in Fig. 1, for CKD-Cls-Net, we introduce a channel attention (ChannelAtten) block to enhance the semantic feature representation before final classification. For CKD-Seg-Net, we introduce spatial and channel attention (SpatialChannelAtten) blocks in the decoder to improve the segmentation of uncertain boundaries and exploit skip connections from U-Net [31] to enable the long-range transfer of fine-level information from shallower layers of the encoder to the decoder.

Let (x_i, y_i) be a sample pair belonging to data $(\mathcal{X}_c, \mathcal{Y}_c)$ for skin lesion diagnosis, where y_i is in $\{1, \ldots, C\}$ and C denotes the number of lesion types. Similarly, let (x_j, y_j) be a sample pair in data $(\mathcal{X}_s, \mathcal{Y}_s)$ for skin lesion segmentation, where y_j denotes the pixel-wise binary label of segmentation. Given $(\mathcal{X}_c, \mathcal{Y}_c)$ and $(\mathcal{X}_s, \mathcal{Y}_s)$, the training process for CKDNet is to improve the overall accuracy for the complex nonlinear mapping functions including the encoders and decoders in RootSeg-Net ($\phi_{enc}(\cdot)$ and $\phi_{dec}(\cdot)$), CKD-Cls-Net ($f_{enc}(\cdot)$ and $f_{dec}(\cdot)$), and CKD-Seg-Net ($g_{enc}(\cdot)$ and $g_{dec}(\cdot)$) by training on the available $(\mathcal{X}_c, \mathcal{Y}_c)$ and $(\mathcal{X}_s, \mathcal{Y}_s)$ datasets. Thus, the classification procedure is formulated as follows:

$$\begin{aligned} &{y_i}' = \phi_{dec}\left(\phi_{enc}\left(x_i\right)\right), \\ &\forall x_i \in \mathcal{X}_c, y_i \in \mathcal{Y}_c : f_{dec}\left(f_{enc}\left(x_i, {y_i}', \phi_{enc}\left(x_i\right)\right)\right) \rightarrow y_i, \end{aligned} \tag{1}$$

and the segmentation procedure is formulated as follows:

$$\begin{aligned} &{y_j}' = \phi_{dec}\left(\phi_{enc}\left(x_j\right)\right), \\ &{y_j}'' = f_{enc}\left(x_j, {y_j}', \phi_{enc}\left(x_j\right)\right), \\ &\forall x_j \in \mathcal{X}_s, y_j \in \mathcal{Y}_s : g_{dec}\left(g_{enc}\left(x_j, {y_j}''\right)\right) \rightarrow y_j. \end{aligned} \tag{2}$$

During the training process, the knowledge is transferred from RootSeg-Net to CKD-Cls-Net and finally diffused to CKD-Seg-Net.

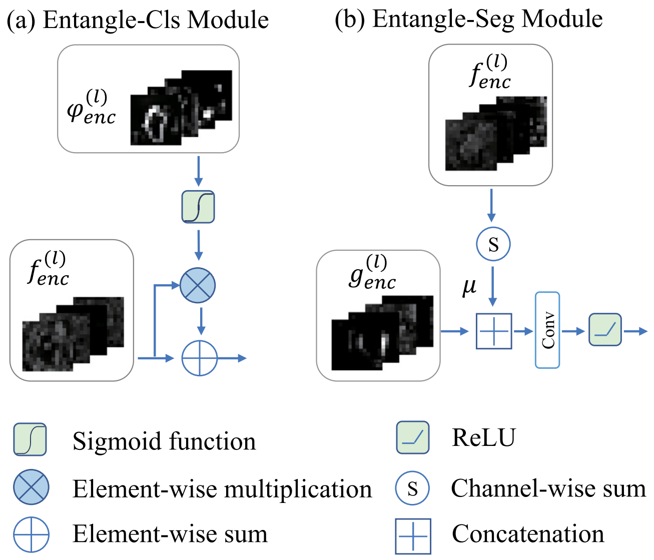


Fig. 2. The proposed entanglement modules to enable and control automated cascade knowledge diffusion in classification and segmentation. Given the output from the lth ResBlock in the encoders as an example; (a) The Entangle-Cls module integrates the cascaded features from RootSeg-Net with CKD-Cls-Net; (b) The Entangle-Seg module aggregates the cascaded features from CKD-Cls-Net with CKD-Seg-Net. The Entangle-Cls module consists of Sigmoid, element-wise multiplication, and element-wise sum operations. The Entangle-Seg module is composed of channel-wise sum, feature concatenation, a convolutional layer, batch normalization, and ReLU operations.

2.3. RootSeg-Net

The aim for the first RootSeg-Net is two-fold: firstly, to obtain the initial lesion segmentation to guide the network to pay attention to lesion areas; and secondly, to acquire a set of high-level dermoscopic features corresponding to global and local structures that have turned out to be of special interest in classification. The RootSeg-Net could be any traditional segmentation networks. In this work, we use the layers before the final average pooling layer of ResNet101 as the encoder, where there are a sequence of convolutional blocks (ConvBlock) and residual blocks (ResBlock) as shown in Fig. 1(b). The encoder is pre-trained on ImageNet [32]. The decoder is from DeepLabV3+ [15]. The detailed architecture of DeepLabV3+ can be found in [15].

2.4. CKD-Cls-Net

The CKD-Cls-Net is designed towards a two-fold goal: to aggregate the diffused features from RootSeg-Net for boosted classification, and to derive context information for the following fine-level segmentation. The architecture of the CKD-Cls-Net is given in Fig. 1(c). The CKD-Cls-Net takes an image and its initial segmentation probability map as input to emphasis the ROI during the learning process. The major components in the CKD-Cls-Net include an encoder with Entangle-Cls modules, a ChannelAtten block and a final predictive layer. The encoder is the same as that in RootSeg-Net, which is a pre-trained ResNet101 on ImageNet. The Entangle-Cls is designed to integrate and control the diffused knowledge from RootSeg-Net with the encoder in CKD-Cls-Net. The ChannelAtten block is used to enhance the semantic feature representation before a global average pooling (GAP) layer and a fully connected (FC) layer for final prediction of C classes. The detailed operations in Entangle-Cls and the ChannelAtten block are given in Figs. 2(a) and 3.

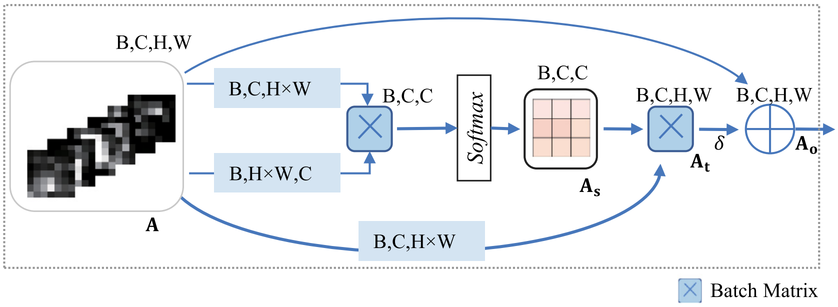


Fig. 3. Illustration of the channel attention (ChannelAtten) block. **A**, $\mathbf{A_s}$, $\mathbf{A_t}$, and $\mathbf{A_o}$ denote the input features, attention maps, attentive features, and the final output features respectively.

2.4.1. Entangle-Cls module

As shown in Fig. 1(c), the Entangle-Cls module is used to integrate the context features $\phi_{enc}^{(l)}(\cdot)$ derived from the lth ResBlock of RootSeg-Net with the features $f_{enc}^{(l)}(\cdot)$ from a corresponding lth ResBlock in CKD-Cls-Net. As shown in Fig. 2(a), $\phi_{enc}^{(l)}(\cdot)$ is firstly smoothed by a Sigmoid activation function to magnify the impact of the correlated features. The magnified features are then served as attention maps by element-wise multiplication and sum operations to enhance feature extraction in the classification encoder.

Let $o^{(l)}$ denotes the entangled feature output of the lth Entangle-Cls modules, $o^{(l)}$ is obtained as:

$$o^{(l)} = \left(1 + \sigma\left(\phi_{enc}^{(l)}(x_i)\right)\right) f_{enc}^{(l)}(x_i) \tag{3}$$

where $\sigma(\cdot)$ denotes the Sigmoid operation. The value of $\sigma\left(\phi_{enc}^{(l)}(x_i)\right)$ is within [0,1]. When $\sigma\left(\phi_{enc}^{(l)}(x_i)\right)$ is 0, the entangled feature is the same as $f_{enc}^{(l)}(x_i)$ which means that there is no knowledge cascaded from RootSeg-Net. The larger $\sigma(\cdot)$ is, the more knowledge can be diffused from RootSeg-Net. Thus, the $\sigma(\cdot)$ can automatically control and adaptively adjust to strengthen the impact of correlated regions and weaken the inferences from irrelevant features which are learned from the initial segmentation.

2.4.2. ChannelAtten block

Given high-level entangled features extracted by the encoder, each channel map before the last decision layer can be regarded as a class-specific response, and different semantic decisions are associated with each other [33]. By exploiting the interdependencies between channel maps, we could improve the feature representation of specific semantics. Therefore, we use the ChannelAtten block to model feature interdependencies between channel maps. Fig. 3 illustrates the structure of the ChannelAtten block.

Given the input features $\mathbf{A} \in \mathbb{R}^{B\times C\times H\times W}$, we reshape $\mathbf{A}$ to the size of $B \times C \times T$, where T is equal to $H \times W$. Afterward, a matrix multiplication between $\mathbf{A} \in \mathbb{R}^{B\times C\times T}$ and $\mathbf{A}' \in \mathbb{R}^{B\times T\times C}$ is performed, where $\mathbf{A}'$ is the transpose of $\mathbf{A}$. Finally, a softmax layer is used to obtain the channel attention map $\mathbf{A_s} \in \mathbb{R}^{B\times C\times C}$. Moreover, a matrix multiplication and reshape between the transpose of $\mathbf{A_s}$ and $\mathbf{A}$ is applied, which gains the attentive feature result to $\mathbf{A_t} \in \mathbb{R}^{B\times C\times H\times W}$. We perform an element-wise sum operation with $\mathbf{A}$ to obtain the final output $\mathbf{A_o}$, which is calculated as:

$$\mathbf{A_o} = \delta\mathbf{A_t} + \mathbf{A}, \tag{4}$$

where δ is a learnable weight. The ChannelAtten block models the semantic dependencies between feature maps, and it enhances the feature representation.

2.5. CKD-Seg-Net

Previous research found that a classification model tends to focus on specific regions for decision making instead of the entire image [34]. A common approach to identify and visualize these salient regions is by the classification activation map (CAM) [34]. Even though a straightforward approach in classification-enhanced segmentation is using CAM, the highlighted image regions by CAM may not always be the exact lesion to be segmented. Thus, directly using CAM as saliency maps may mislead the segmentation results. Nevertheless, the context information learnt from the classification task can reveal informative features of the disease. Hence, instead of using CAM, we propose to entangle the context knowledge extracted from the classification encoder to the segmentation encoder.

As shown in Fig. 1(d), the CKD-Seg-Net also uses pre-trained ResNet101 as the backbone and takes an image as input. The Entangle-Seg module controls the cascaded context knowledge from the CKD-Cls-Net and integrates them with the features extracted from the segmentation encoder. The decoder has the SpatialChannelAtten blocks and the SpatialChannelAtten Out block to enhance the network's attention to both spatial and channel-wise information for lesion boundary definition. The skip connections transfer features from shallower layers of the encoder to the decoder. The details of the Entangle-Seg, the SpatialChannelAtten block, and the SpatialChannelAtten Out block are shown in Figs. 2(b) and 4.

2.5.1. Entangle-Seg module

As shown in Fig. 2(b), an Entangle-Seg module integrates the context features $f_{enc}^{(l)}(\cdot)$ of the lth ResBlock of CKD-Cls-Net with features $g_{enc}^{(l)}(\cdot)$ from the lth ResBlock in CKD-Seg-Net. We firstly fuse all channel-wise features in $f_{enc}^{(l)}(\cdot)$ by sum operations. These context features are then multiplied by a learnable balance factor μ which is tuned during the training process to obtain adjusted context features from classification. Afterward, the adjusted classification features $\mu f_{enc}^{(l)}(\cdot)$ are concatenated to the segmentation features $g_{enc}^{(l)}(\cdot)$. Finally, the concatenated features are fed to a sequence of a 3 × 3 convolutional layer with 1 padding, batch normalization, and a ReLU function for final embedding.

2.5.2. SpatialChannelAtten block

The segmentation decoder is composed of three SpatialChannelAtten blocks, one SpatialChannelAtten Out block, and a convolutional layer. As shown in Fig. 4(b), the SpatialChannelAtten block consists of a sequence of a Concurrent Spatial and Channel Squeeze and Excitation (scSE) module, two convolutional layers, and an scSE module, where the scSE module is exploited from [35]. It has been proved in [35] that the scSE module can enhance the extraction and modeling of spatial and channel-wise relevant information. The SpatialChannelAtten Out block is

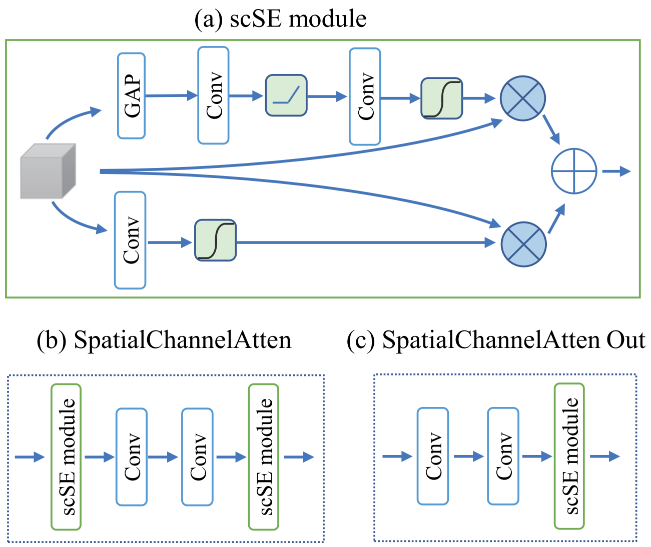


Fig. 4. The spatial channel attention (SpatialChannelAtten) block and SpatialChannelAtten Out block in the segmentation decoder. The major components include (a) scSE module and convolutional layers; (b) The SpatialChannelAtten block consisting of a sequence of one scSE module, two convolutional layers, and one scSE attention module; (c) The SpatialChannelAtten Out block with two convolutional layers and one scSE module.

Table 1
Architectures of the SpatialChannelAtten block and the SpatialChannelAtten Out block. The symbol [,] denotes the long-range summation connections.

Module/Layer	SpatialChannelAtten	SpatialChannelAtten Out
Upsampling	2 × 2	2 × 2
Concatenation	[,]	None
scSE module	global average pooling 1 × 1 convolution ReLU 1 × 1 convolution Sigmoid 1 × 1 convolution Sigmoid	None
Conv	3 × 3, padding 1 batch normalization ReLU	3 × 3, padding 1 batch normalization ReLU
Conv	3 × 3, padding 1 batch normalization ReLU	3 × 3, padding 1 batch normalization ReLU
scSE module	global average pooling 1 × 1 convolution ReLU 1 × 1 convolution Sigmoid 1 × 1 convolution Sigmoid	global average pooling 1 × 1 convolution ReLU 1 × 1 convolution Sigmoid 1 × 1 convolution Sigmoid
dropout	0.2	0.2

composed of two convolutional layers and an scSE module. Each of the convolutional layers in Fig. 4(a) is with a 1 × 1 kernel. Meanwhile, for all the convolutional layers in Fig. 4(b)(c), each of them is with a 3 × 3 kernel, padding 1, followed by a batch normalization and a ReLU function.

By using skip connections in Fig. 1(d), we first upsample the output from the 4th Entangle-Seg block and concatenate it with the output of the 3rd Entangle-Seg in the encoder through long-range transfer. Then, the concatenation is decoded using the 1st SpatialChannelAtten block for fine-level segmentation. Similarly, we carry out the upsampling and concatenation procedures for the 2nd Entangle-Seg and 2nd SpatialChannelAtten block, the 1st Entangle-Seg and 3rd SpatialChannelAtten block, and the 1st ConvBlock in Entangle-Seg and 4th SpatialChannelAtten block. Since there is not concatenation after the 5th SpatialChannelAtten, feature fusion by scSE is not needed. Thus, we delete the first scSE block of the SpatialChannelAtten block to be a new block, termed as SpatialChannelAtten Out block. Finally, features are sent to the SpatialChannelAtten Out block before a 1 × 1 convolutional layer for fine-level segmentation.

2.6. Loss function

A weighted cross-entropy (WCE) loss function is used for the skin lesion classification task. The weight is calculated based on the number of samples for each lesion type in the training data. The WCE is defined as:

$$\mathrm{WCE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C}\beta_i^c p_i^c \log \hat{p}_i^c, \tag{5}$$

where β_i^c denotes the weight belonging to class c, $\hat{p}_i^c$ denotes the probability of a sample i, p_i^c denotes the ground-truth class, N indicates the total number of samples, and C denotes the number of lesion types.

For the segmentation task, we propose a combined soft Dice loss (DL) and Focal loss (FL) [36] to emphasis the segmentation performance at uncertain lesion boundaries to improve the results. The combined segmentation loss (SL) is defined as:

$$\mathrm{SL} = \mathrm{DL} + \lambda_1 \mathrm{FL}, \tag{6}$$

where DL is formulated as:

$$\mathrm{DL} = 1 - \frac{2\sum_{j=1}^{M} q_j \hat{q}_j}{\sum_{j=1}^{M} q_j^2 + \sum_{j=1}^{M} \hat{q}_j^2}. \tag{7}$$

where $\hat{q}_j$ denotes the probability of a pixel j, q_j denotes the groundtruth class, and M indicates the total number of pixles. FL is defined as follows:

$$q_t = \begin{cases} \hat{q}_j & \text{if } q_j = 1 \\ 1 - \hat{q}_j & \text{otherwise} \end{cases} \tag{8}$$

$$\mathrm{FL} = \frac{1}{M}\sum_{j=1}^{M}\sum_{c=1}^{C} -\alpha_t \left(1 - q_t\right)^{\gamma} \log\left(q_t\right). \tag{9}$$

2.7. Evaluation methods

The official evaluation metrics for the classification task in ISIC2017 include area under the receiver operating characteristic curve (AUC), accuracy (ACC), sensitivity (SEN), specificity (SPC), and positive predictive value (PPV). The evaluation metrics for the segmentation task include Jaccard (JA), ACC, SEN, SPC, and Dice coefficient (Dice).

For the classification task in ISIC2018, the overall classification performance is evaluated by multi-class accuracy (BMA). Apart from BMA, the classification performance is measured by SEN, SPC, ACC, AUC, Dice, PPV, and negative predictive value (NPV). The segmentation task in ISIC 2018 is evaluated by a thresholded Jaccard index (Thres_Jaccard) metric as:

$$\mathrm{Thres\_Jaccard} = \begin{cases} 0, & J(\mathrm{SR}, \mathrm{GT}) \le 0.65 \\ J(\mathrm{SR}, \mathrm{GT}), & \text{otherwise} \end{cases} \tag{10}$$

where $J(\mathrm{SR}, \mathrm{GT})$ denotes the JA between the segmentation result (SR) and groundtruth (GT). The $J(\mathrm{SR}, \mathrm{GT})$ is defined as follows:

$$J(\mathrm{SR}, \mathrm{GT}) = \frac{|\mathrm{SR} \cap \mathrm{GT}|}{|\mathrm{SR}| + |\mathrm{GT}| - |\mathrm{SR} \cap \mathrm{GT}|}. \tag{11}$$

Table 2
Results on ISIC2017 diagnosis test dataset for CKDNet and top-ranked SOTA methods.

Method	Ensemble	External data	Task 1					Task 2					Avg
			AUC	ACC	SEN	SPC	PPV	AUC	ACC	SEN	SPC	PPV	AUC
Harangi [9]	Yes	No	0.851	0.852	0.402	0.719	–	0.930	0.880	0.711	0.851	–	0.891
GP-CNN-DTEL [37]	Yes	No	0.891	0.850	0.376	**0.965**	–	0.960	**0.935**	0.722	0.973	–	0.926
SDL [38]	No	Yes	0.868	**0.888**	–	–	0.720	0.958	0.925	–	–	0.840	0.913
DeVries et al.'s [39]	Yes	Yes	0.836	0.845	0.350	**0.965**	–	0.935	0.913	0.556	0.976	–	0.886
Bi et al.'s [40]	Yes	Yes	0.870	0.858	0.427	0.963	–	0.921	0.918	0.589	0.976	–	0.896
Menegola et al.'s [41]	Yes	Yes	0.874	0.872	0.547	0.950	–	0.943	0.895	0.356	**0.990**	–	0.908
DermaKNet [11]	No	Yes	0.873	–	–	0.460	–	**0.962**	–	–	0.843	–	0.917
ARL-CNN [5]	No	Yes	0.875	0.850	0.658	0.896	–	0.958	0.868	**0.878**	0.867	–	0.917
Barata et al.'s [8]	Yes	No	0.855	–	**0.735**	0.838	–	0.932	–	0.611	0.972	–	0.894
CKDNet	**No**	**No**	**0.905**	0.881	0.700	0.925	**0.738**	0.959	0.923	0.689	0.965	**0.854**	**0.932**

Finally, the overall metric value for the entire dataset takes the average JA of each image. To be consistent with the evaluations of ISIC2017, we also calculate SEN, SPC, ACC, JA, and Dice for the segmentation using ISIC2018.

The performance of the proposed CKDNet is validated by the comparison with SOTA methods using the ISIC2017 and ISIC2018 datasets. The effectiveness of the proposed entanglement modules and other major components are evaluated by ablation study.

2.8. Implementation, initialization, and parameter settings

2.8.1. Network architecture

The overall architecture and the output feature size of each block are illustrated in Fig. 1. Specifically, the structure of the backbone is identical to the layers before the final average pooling layer of ResNet101 [29], which consists of a ConvBlock and four ResBlocks. The ConvBlock is composed of a 7 × 7 convolutional layer with stride 2 and padding 3, a batch normalization layer, a ReLU layer, and a max-pooling layer. It is noted that the input channel of ConvBlock for CKD-Cls-Net is set to 4 for the additional coarse mask from RootSeg-Net. The weights of the 4th channel are initialized by averaging the weights of the other three channels. The parameters of each ResBlock can be found in [29]. For RootSeg-Net, we exploited the decoder of DeepLabV3+ [15]. The remaining modules of CKD-Cls-Net contain a ChannelAtten block (with settings given in Fig. 3), a global average pooling (GAP) layer, and an FC layer for final prediction of C types of lesions. For the decoder of CKD-Seg-Net, the overall structure is illustrated in Section 2.5. Moreover, the detailed layers of the SpatialChannelAtten blocks and the SpatialChannelAtten Out block are shown in Table 1. The codes will be released publicly at GitHub.[1]

2.8.2. Pre-processing and data augmentation

Augmentation transformations during the training process included random flipping in vertical and horizontal directions, affine transformations using translation within [0.01, 0.1], scaling within [0.9, 1.1], and rotating by up to 90°. The transformed images were center cropped and resized to 224 × 224 pixels for training. Finally, all these training images were normalized by ImageNet's standard deviation and mean value.

2.8.3. Training details

During training, the batch size was set as 32 for classification and 16 for segmentation. We used Adam optimizer with a weight decay of 0.0001. The initial learning rate was 0.0001 and this value was divided by 10 if the performance did not improve in 20 epochs. The training epochs of RootSeg-Net, CKD-Cls-Net, and CKD-Seg-Net were set as 20, 100, and 100 respectively. For the loss function in Eq. (6), λ_1 was set as 1, α and γ in Eq. (9) were set the same as [36]. The framework was implemented using PyTorch with a Tesla P100 graphics card.

3. Results and discussion

3.1. Comparison with SOTA methods for skin lesion classification

To evaluate the classification performance of the proposed CKDNet, we compare with SOTA methods using ISIC2017 and ISIC2018.

3.1.1. Results over ISIC2017

To demonstrate the performance of our CKDNet in terms of the classification task for ISIC2017, we retrieve the published SOTA results. The classification results of Task 1 (melanoma versus nevus and seborrheic keratosis) and Task 2 (seborrheic keratosis versus nevus and melanoma) are given in Table 2. It is noted that the top-ranked SOTA methods used either an ensemble strategy or external data (collected from ISIC achieve[2]), or both of them to boost the performance. Our method, however, is trained without using any external data or any ensembled approaches.

Overall, our model achieved the best average AUC of 0.932, which outperformed the methods with ensembles [9,37,8], the methods trained using external data [38,11,5], and those using both ensemble and external data [39–41].

For classification Task 1, our model achieved the highest AUC of 0.905, which outperformed the second-best model with ensemble [37] by 1.4% and the third-ranked model using external data [5] by 3.0%. We obtained the second-best ACC of 0.881, which was slightly lower than the best model with external data [38] by 0.7%. When it comes to classification Task 2, our CKDNet achieved the third-best AUC of 0.959 with a marginal difference of 0.3% when compared with the best model [11]. Our method, however, had less data for training.

3.1.2. Results over ISIC2018

The ISIC2018 has seven different types of skin diseases in the classification task. For a fair comparison, the pre-processings and loss functions of all the other compared methods were set the same as our model. The evaluation results are given in Table 3. As shown in this table, our model outperformed the other methods in all the evaluation metrics.

Our primary finding is that our model achieved improved classification performance, especially for the classification task of melanoma versus nevus and seborrheic keratosis in ISIC2017 and the seven-class classification task in ISIC2018. Our second finding is that the classification performance is improved even though there is no external data or ensembled architectures. The context knowledge diffused from initial segmentation by the proposed Entangle-Cls module could explain the improved performance.

[1] https://github.com/qgking/CKDNet

[2] https://www.isic-archive.com

Table 3
Results on ISIC2018 diagnosis test dataset for CKDNet and SOTA methods which were trained from scratch.

Method	AUC	ACC	SEN	SPC	Dice	PPV	NPV	BMA
ResNet34 [29]	0.943	0.945	0.707	0.963	0.718	0.737	0.953	0.727
DenseNet121 [42]	0.944	0.949	0.715	0.962	0.732	0.756	0.959	0.724
ARL [5]	0.934	0.934	0.689	0.959	0.665	0.657	0.944	0.711
Barata et al.'s [8]	0.936	–	0.637	0.956	–	–	–	0.641
CKDNet	**0.975**	**0.963**	**0.802**	**0.976**	**0.810**	**0.836**	**0.966**	**0.816**

Table 4
Performance comparison of the proposed CKDNet with some SOTA methods on ISIC2017 segmentation test dataset.

Method	JA	Dice	ACC	SEN	SPC
DDN [43]	0.765	0.866	0.939	0.825	0.984
CDNN [44]	0.765	0.849	0.934	0.825	0.975
SLSDeep [13]	0.782	**0.878**	0.936	0.816	0.983
FCN + SSP [45]	0.773	0.857	0.938	0.855	0.973
Yuan et al.'s [46]	0.765	0.849	0.934	0.825	0.975
Berseth et al.'s [47]	0.762	0.847	0.932	0.820	0.978
Chen et al.'s [16]	0.787	0.868	0.944	–	–
SegAN [12]	0.785	0.867	0.941	–	–
Bi et al.'s [48]	0.771	0.851	–	–	–
Bi et al.'s [40]	0.760	0.844	0.934	0.802	**0.985**
Goyal et al.'s [14]	0.793	0.871	0.941	**0.899**	0.950
CKDNet	**0.800**	0.877	**0.946**	0.887	0.961

3.2. Comparison with SOTA methods on skin lesion segmentation

To validate the segmentation capacity of our CKDNet, we compare it with SOTA methods using ISIC2017 and ISIC2018 datasets.

3.2.1. Results over ISIC2017

The segmentation results on the ISIC2017 test dataset are given in Table 4. As shown in this table, our model achieved the best JA of 0.800 and ACC of 0.946 among all the methods in comparison. Our Dice metric was second-best which was slightly lower than the best model [13] with a marginal difference of 0.1%.

Eighteen (18) examples of the segmentation results achieved by the proposed CKDNet with various sizes, lesion types, and surrounding tissues are given in Fig. 5, As shown by 'ISIC_0014503', 'ISIC_0014278', 'ISIC_0016055', and 'ISIC_0015031', our model can segment lesions of various sizes and textures even though there are surrounding hairs, veins, and other artifacts. It is also shown by 'ISIC_0013242', 'ISIC_0013242', 'ISIC_0014666', and 'ISIC_0015203' that the lesion boundaries obtained by our model were adherent to the salient lesion areas when compared with GT annotations.

3.2.2. Results over ISIC2018

The evaluation results are given in Table 5. Our model achieved the best results out of 5 evaluation measures except SPC. Our SPC was ranked in the fifth place among all the methods in comparison.

In summary, our model achieved the best segmentation results on JA and ACC over both ISIC 2017 and 2018 datasets, best Dice and SEN using ISIC2018, competitive Dice and second-best SEN on ISIC 2017. Without external dataset or ensembled strategy, the experimental results demonstrated the capacity of our segmentation method by using the entangled knowledge from the classification task. It is notable that the proposed CKDNet is not a multi-task based method, which means that the CKDNet does not need adequate annotations of one sample for both classification and segmentation.

Table 5
Performance comparison of the proposed CKDNet and some SOTA methods which were trained from scratch on the ISIC2018 skin lesion segmentation dataset.

Method	SEN	SPC	ACC	JA	Dice	Thres_Jaccard
FCN8s [49]	0.949	0.905	0.918	0.753	0.848	0.666
U-Net [31]	0.914	0.942	0.918	0.761	0.851	0.680
DeepLab [50]	0.944	0.913	0.924	0.768	0.858	0.698
R2U-Net [51]	0.894	**0.967**	0.920	0.782	0.861	0.713
BCDU-Net [52]	0.885	0.947	0.911	0.752	0.838	0.681
CKDNet	**0.967**	0.904	**0.934**	**0.794**	**0.877**	**0.742**

Table 6
Effectiveness analysis of the Entangle-Cls module and the ChannelAtten block in CKDNet using the ISIC2017 skin lesion diagnosis test dataset for classification Task 1.

Method	AUC	ACC	SEN	SPC	PPV
ResNet101	0.849	0.847	0.556	0.917	0.616
ResNet101 + ChannelAtten	0.879	0.841	0.675	0.882	0.662
ResNet101 + Entangle-Cls	0.895	0.878	0.667	**0.930**	0.719
CKDNet	**0.905**	**0.881**	**0.700**	0.925	**0.738**

3.3. Effectiveness analysis of knowledge diffusion and entanglement modules in CKD-Cls-Net

We investigate the contributions of the proposed knowledge diffusion strategy and the entanglement component in the classification network CKD-Cls-Net in our CKDNet. Ablation studies are performed using the ISIC2017 dataset for classification Task 1.

We firstly perform ablation study to validate the contributions of the Entangle-Cls module for cascaded knowledge integration and the ChannelAtten block for channel-wise feature representation enhancement. The results are given in Table 6 where ResNet101 is the baseline model, ResNet101 + ChannelAtten denotes our CKD-Cls-Net without Entangle-Cls, and ResNet101 + Entangle-Cls denotes our classification network without the ChannelAtten block. As shown in Table 6, the ChannelAtten module improved AUC, SEN, and PPV by 3.0%, 11.9%, 4.6% respectively when compared with ResNet101. The Entangle-Cls module boosted the classification results for all the evaluation metrics. Our CKD-Cls-Net with both Entangle-Cls and ChannelAtten further improved the classification results. To intuitively show the improved performance, AUC-ROC and precision–recall curve are illustrated in Fig. 6. Comparing to the baseline model, the curves show that both of the ChannelAtten and Entangle-Cls modules contributed to the improved performance. Specifically, the ChannelAtten module improved the average precision (AP) by 4.6% when compared with ResNet101, while the Entangle-Cls module significantly improved the AP by 10.2%, which demonstrated the effectiveness of the proposed Entangle-Cls module. The statistical results revealed that the CKDNet achieved AP of 73.8%, which outperformed other methods.

We also use t-SNE [53] to interpret the distributions of the image samples in the high-level semantic latent feature space constructed by different models in classification. We extracted 2048-dimensional features (2048 × 1) from the penultimate layer and embedded them by t-SNE. The results using baseline ResNet101, ResNet101 + ChannelAtten, and our CKD-Cls-Net are given in Fig. 7. It can be observed that although CKD-Cls-Net still has some issues in fully discriminating high-level features, samples of different classes are better clustered in the high-level semantic space than the other two methods. The results further demonstrated the contribution of the aggregation of cascaded knowledge by Entangle-Cls in our classification networks.

Finally, we interpret the network's attention during the classification process using CAM. The CAMs of four examples using different classification networks are given in Fig. 8. Using ResNet101

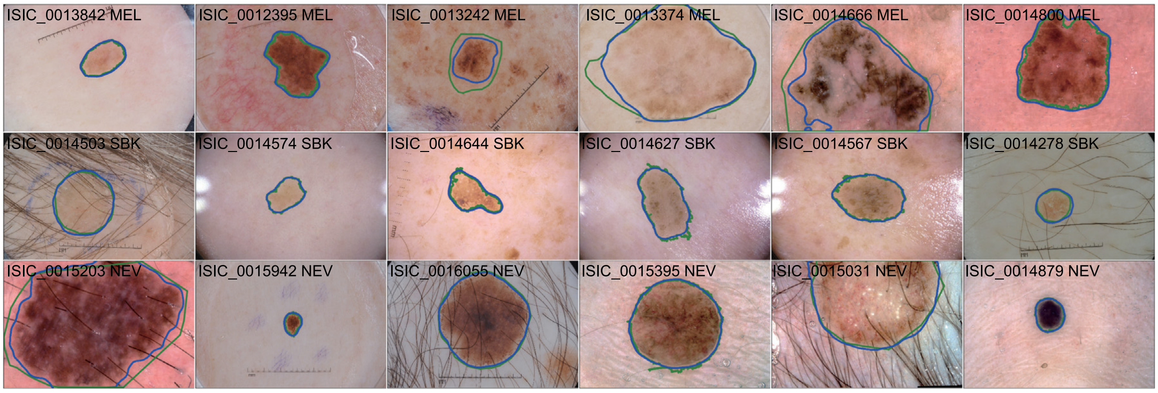


Fig. 5. Segmentation results with contours on the ISIC2017 test dataset. We choose six images from each lesion type, i.e., melanoma (MEL), seborrheic keratosis (SBK), and nevus (NEV). Image name and lesion type are listed at the top of each image. Ground truth and our segmentation results are shown by green and blue contours.

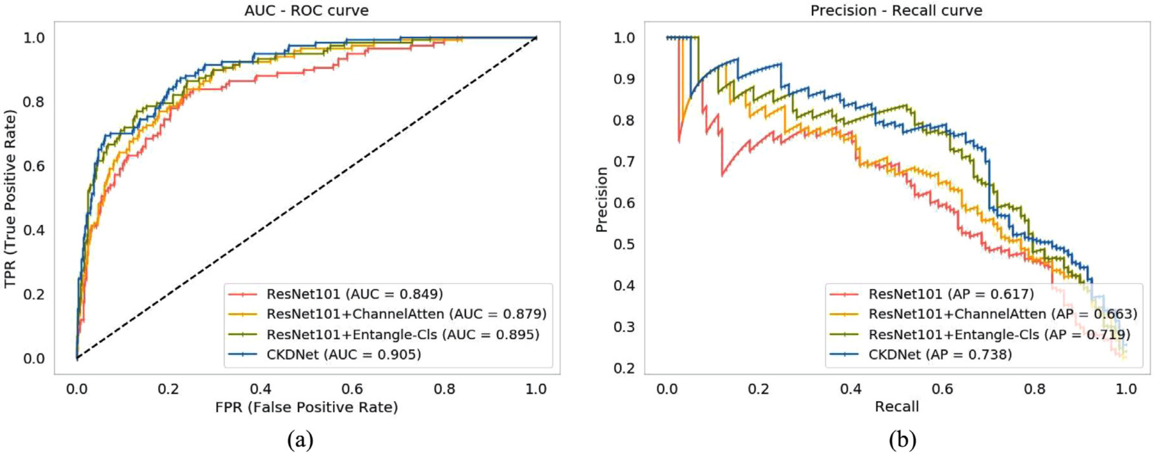


Fig. 6. AUC curve (a) and precision–recall curve (b) for effectiveness analysis of the Entangle-Cls module and the ChannelAtten block in CKDNet.

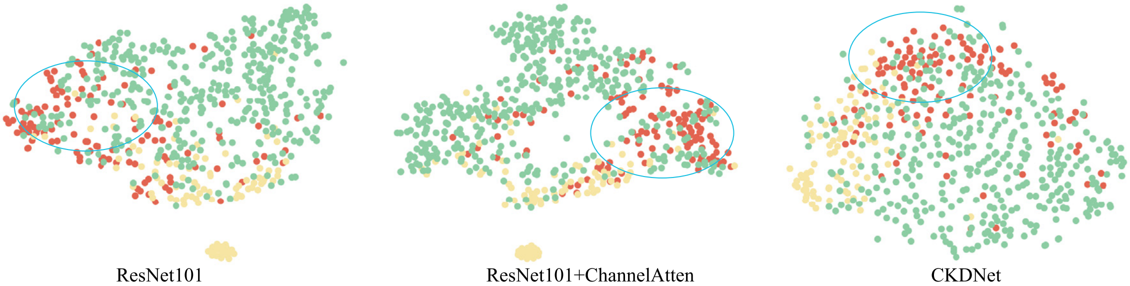


Fig. 7. Visual interpretation of high-level CNN features of three classification methods using t-SNE on the ISIC2017 test dataset. The red, yellow, and green dots are MEL, SBK, and NEV data, respectively. Blue circled areas indicate that those high-level features are not highly discriminative.

Table 7
Effectiveness analysis (mean ± std) of the Entangle-Seg module and different loss functions in CKDNet using the ISIC2017 skin lesion segmentation test dataset.

Method	JA	Dice	ACC	SEN	SPC
BCE	0.777 ± 0.197	0.856 ± 0.175	0.938 ± 0.096	0.860 ± 0.198	0.959 ± 0.107
Dice	0.785 ± 0.185	0.864 ± 0.157	0.941 ± 0.090	0.853 ± 0.189	**0.969 ± 0.077**
Dice + BCE	0.789 ± 0.180	0.867 ± 0.150	0.941 ± 0.090	0.864 ± 0.176	0.966 ± 0.085
Dice + Focal	0.795 ± 0.174	0.872 ± 0.146	0.941 ± 0.092	0.883 ± 0.157	0.958 ± 0.102
CKDNet	**0.800 ± 0.166**	**0.877 ± 0.138**	**0.946 ± 0.082**	**0.887 ± 0.164**	0.961 ± 0.095

alone, the network failed to pay attention to the lesion region. When adding the ChannelAtten module to ResNet101, the classification model started raising attention to the ROIs. Compared with ResNet101 and ResNet101 + ChannelAtten, our CKD-Cls-Net demonstrated the concentration on the regions which are most relevant to the diagnosis. The CAM reveals the power of our CKD-Cls-Net in capturing contextual information related to the disease for classification.

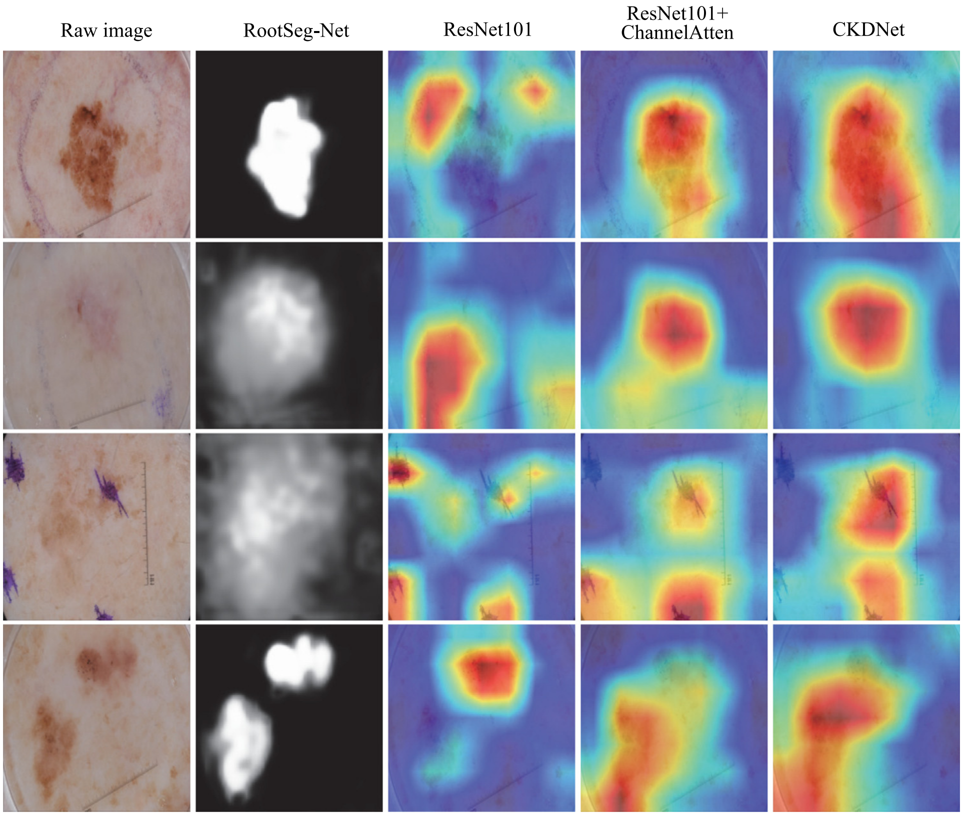


Fig. 8. Visualization and comparison of initial segmentation by RootSeg-Net and the CAMs obtained by three classification models. In each row, there are original image, initial segmentation by RootSeg-Net, CAM by baseline ResNet, CAM by ResNet with ChannelAtten block only, and CAM by our CKDNet.

3.4. *Effectiveness analysis of entanglement modules in CKD-Seg-Net and loss function*

We evaluate the contributions of the Entangle-Seg module in the segmentation network and diffusion loss functions. Table 7 presents the experimental results using the ISIC2017 segmentation test dataset with different loss functions. The baseline model is our CKD-Seg-Net without the Entangle-Seg module. We tested the binary cross-entropy (BCE), Dice, and Focal losses. As shown in Table 7, the Dice loss substantially improved segmentation performance in all the five evaluation measures when compared with the BCE loss. Using combined Dice and BCE losses further improved JA, Dice, and SEN results but disadvantaged SPC. When the combined Dice and Focal losses are used, the values of JA, Dice, and SEN increased while that of SPC further decreased and ACC remained the same. Finally, our CKDNet (i.e., the CKD-Seg-Net of the CKDNet) with the Entangle-Seg module and the combined Focal and Dice losses improved all the evaluation measures of JA (0.800 ± 0.166), Dice (0.877 ± 0.138), ACC (0.946 ± 0.082), SEN (0.887 ± 0.164), and SPC (0.961 ± 0.095). Furthermore, we use box plots to statistically analyze the performance of the Entangle-Seg module and diffusion loss functions. As shown in Fig. 9, we found that for the BCE and Dice losses, the *first quartile* (Q_1) and *minimum* values of the Dice loss were inferior to those of BCE loss, which were worse than any other methods. However, the *median* (Q_2) and mean values were better than those of BCE loss. With gradually substituting the loss function, Q_1, Q_2, *minimum*, and the mean values were improved, which demonstrated the effectiveness and contribution of the combined loss function. Besides, it also reveals that the performance is substantially improved with the integration of the Entangle-Seg module. Conclusions can be drawn from the comparison. Firstly, when dealing with the class imbalance issue, the combination of Dice and Focal losses achieves better BCE loss and Dice loss. Secondly, the proposed Entangle-Seg module has the capacity to boost the segmentation performance based on the knowledge learnt from the classification task.

We also investigate lesion boundary delineation results using different combinations of loss functions and the Entangle-Seg module. Three examples are given in Fig. 10. As shown in the figure, our CKDNet model obtained the best segmentation results, especially for the areas with uncertain boundaries. We explain this finding by the learnable weight balance factor μ in the Entangle-Seg module, which can adjust the contributions of context information from classification for different cases. For instance, it adjusts the impact of supervisions from the classification task to a confident region in the segmentation, and preserves informative guidance to benefit the definition of uncertain boundaries.

3.5. *Impact of parameter settings*

As discussed in the previous sections, the tuning and selection of hyper-parameters are time-consuming and labor-intensive in multi-task learning algorithms. On the contrary, our proposed model contains a smaller number of hyper-parameters, which is consequently convenient for optimizing the model. Firstly, the purpose of network design is to transfer the coarse lesion masks to the CKD-Cls-Net and drive the classification network's attention to image regions relevant to the disease. Thus, it is not necessary to carefully select the best possible hyper-parameters in RootSeg-Net. Secondly, in the proposed CKD-Cls-Net, there are two important hyper-parameters, i.e., the learning rate (lr) of the pre-trained backbone and the remaining blocks. We investigated the lr settings of the backbone with 1e−3, 1e−4, and 1e−5, and that of the remaining blocks to 1e−3, 1e−4, and 1e−5.

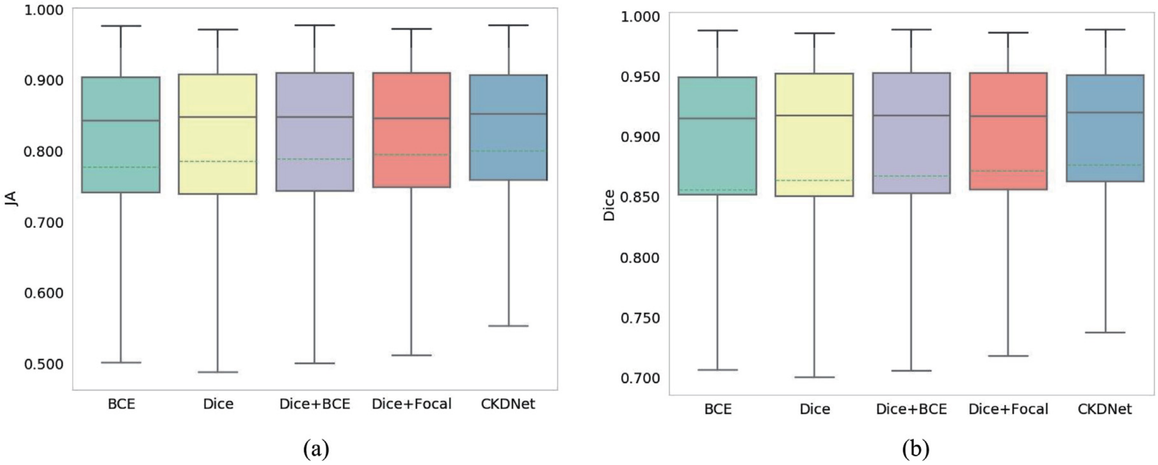


Fig. 9. Box plots of the typical metrics, i.e., JA (a) and Dice (b), for effectiveness analysis of the Entangle-Seg module and different loss functions in CKDNet. The mean value of each metric is in the green dashed line.

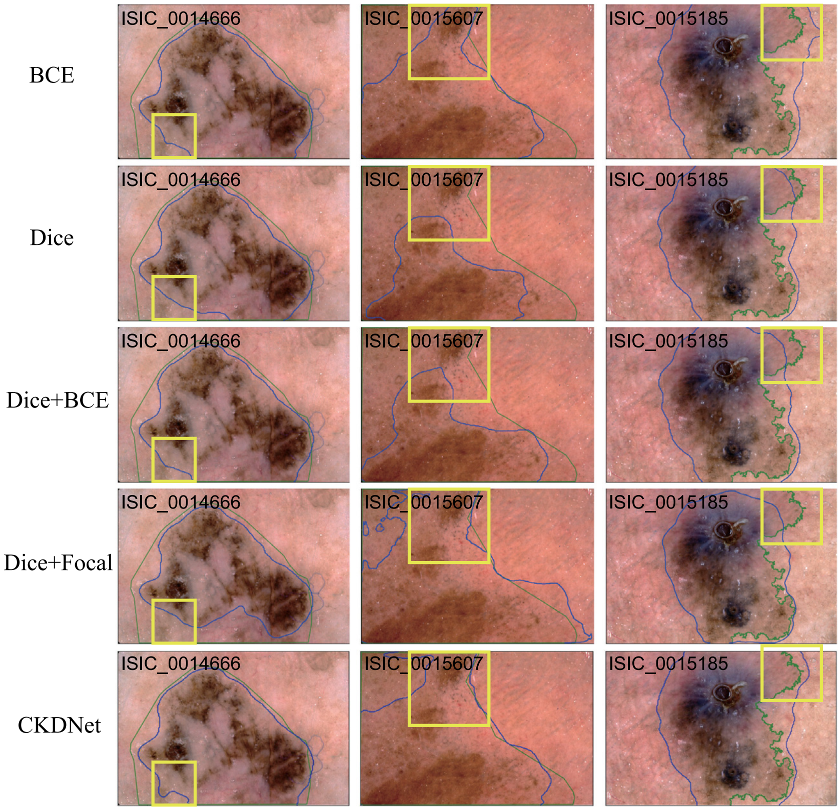
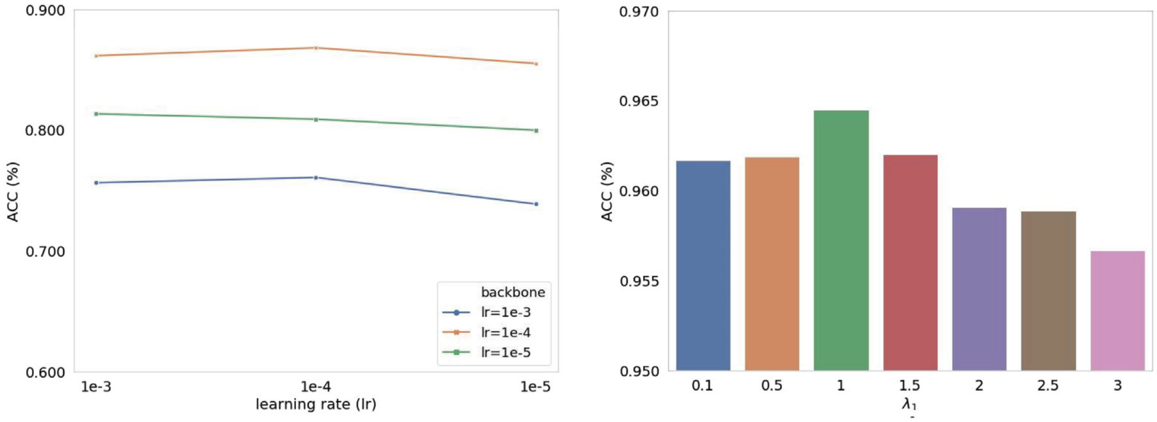


Fig. 10. Comparison of the segmentation results obtained by different loss functions and CKDNet. Challenging and uncertain lesion boundaries are highlighted by yellow boxes.

The experimental results over the ISIC2017 validation dataset are illustrated in Fig. 11(a). It suggests that 1e−4 is the best lr for the pre-trained backbone given the same lr of the remaining blocks. This is reasonable since the backbone has been pre-trained on the ImageNet, and it is relatively easy to train the images using pre-trained weights than training from scratch. When the lr of the backbone is too large, the CKD-Cls-Net turns out to be more unstable to converge. On the contrary, when the lr of the backbone is too small, the CKD-Cls-Net would take much time to converge. On the other hand, 1e−4 is also the best lr for the remaining blocks of CKD-Cls-Net. Hence, we empirically set lr to 1e−4 in this study. Finally, for CKD-Seg-Net, the critical parameter λ_1, which controls the contribution of the combined loss, is investigated by repeating the skin lesion segmentation training procedure with different values of λ_1 between 0.1 and 3. Our initial empirical observation is that the DL and the FL were on the same scale when the model reaches convergence. Hence, we set the values of λ_1 to 0.1, 0.5, 1, 1.5, 2, 2.5, and 3. Fig. 11(b) shows that CKD-Seg-Net achieved the highest ACC on the ISIC2017 validation dataset when λ_1 is set to 1. Therefore, we



(a) learning rate settings for the backbone and the rest of CKDNet in skin lesion diagnosis

(b) different $\lambda_1$settings of combined loss in CKDNet for skin lesion segmentation

Fig. 11. Comparison of different learning rates (a) and λ_1 values (b) in CKDNet on the ISIC2017 validation dataset.

use 1 as the default weighting factor in the proposed combined segmentation loss.

3.6. Computational efficiency

In our experiments, it took approximately 23 min for training RootSeg-Net, 21 h for training CKD-Cls-Net, and 5.5 h for training CKD-Seg-Net on the ISIC2017 dataset. When applying the trained model to testing images, it took less than 0.5 s to classify a skin lesion image, and 1.3 s for segmentation on a server with 1 T P100 graphics card. The computational efficiency shows the potential of using our model for computer aided clinical skin lesion diagnosis to be applied in a routine clinical workflow.

4. Conclusion

We propose a new concept of cascade knowledge diffusion in different sub-networks, with applications on dermoscopic skin lesion images. To fully leverage both the context features from diagnosis and segmentation tasks, we propose to use task-specific entanglement modules simultaneously. Those two modules boost learning and greatly improve performances. Furthermore, an effective loss function based on the Dice loss and the Focal loss to alleviate the class imbalance problem for skin image segmentation is developed. The proposed CKDNet architecture achieves competitive performances compared with the state-of-the-art skin lesion diagnosis and segmentation methods without using any external data or in any ensemble manner. Our future work includes providing an end-to-end platform for clinical research and reduce the complexity of the training procedure. We will also generalize our CKDNet and cascade knowledge diffusion strategy to other image classification and segmentation datasets.

CRediT authorship contribution statement

Qiangguo Jin: Conceptualization, Methodology, Software, Writing - original draft. **Hui Cui:** Validation, Writing - review & editing, Visualization. **Changming Sun:** Formal analysis, Investigation. **Zhaopeng Meng:** Supervision, Funding acquisition, Supervision. **Ran Su:** Resources, Writing - review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by the National Natural Science Foundation of China [Grant No. 61702361], the Science and Technology Program of Tianjin, China [Grant No. 16ZXHLGX00170], the National Key Technology R&D Program of China [Grant No. 2018YFB1701700], and the program of China Scholarship Council.

References

[1] H.W. Rogers, M.A. Weinstock, S.R. Feldman, B.M. Coldiron, Incidence estimate of nonmelanoma skin cancer (keratinocyte carcinomas) in the US population, 2012, JAMA Dermatol. 151 (10) (2015) 1081–1086.
[2] A. Esteva, B. Kuprel, R.A. Novoa, J. Ko, S.M. Swetter, H.M. Blau, S. Thrun, Dermatologist-level classification of skin cancer with deep neural networks, Nature 542 (7639) (2017) 115–118.
[3] L. Song, J.P. Lin, Z.J. Wang, H. Wang, An end-to-end multi-task deep learning framework for skin lesion analysis, IEEE J. Biomed. Health Inf. (2020) 1.
[4] N. Gessert, T. Sentker, F. Madesta, R. Schmitz, H. Kniep, I. Baltruschat, R. Werner, A. Schlaefer, Skin lesion classification using CNNs with patch-based attention and diagnosis-guided loss weighting, IEEE Trans. Biomed. Eng. 67 (2) (2020) 495–503.
[5] J. Zhang, Y. Xie, Y. Xia, C. Shen, Attention residual learning for skin lesion classification, IEEE Trans. Med. Imaging 38 (9) (2019) 2092–2103, http://dx.doi.org/10.1109/TMI.2019.2893944.
[6] Y. Yuan, M. Chao, Y. Lo, Automatic skin lesion segmentation using deep fully convolutional networks with Jaccard distance, IEEE Trans. Med. Imaging 36 (9) (2017) 1876–1886, http://dx.doi.org/10.1109/TMI.2017.2695227.
[7] F. Navarro, M. Escudero-Viñolo, J. Bescós, Accurate segmentation and registration of skin lesion images to evaluate lesion change, IEEE J. Biomed. Health Inf. 23 (2) (2019) 501–508, http://dx.doi.org/10.1109/JBHI.2018.2825251.
[8] C. Barata, J.S. Marques, M. Emre Celebi, Deep attention model for the hierarchical diagnosis of skin lesions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2019.
[9] B. Harangi, Skin lesion classification with ensembles of deep convolutional neural networks, J. Biomed. Inform. 86 (2018) 25–32.
[10] L. Yu, H. Chen, Q. Dou, J. Qin, P.-A. Heng, Automated melanoma recognition in dermoscopy images via very deep residual networks, IEEE Trans. Med. Imaging 36 (4) (2016) 994–1004.
[11] I. González-Díaz, Dermaknet: Incorporating the knowledge of dermatologists to convolutional neural networks for skin lesion diagnosis, IEEE J. Biomed. Health Inf. 23 (2) (2018) 547–559.

[12] Y. Xue, T. Xu, X. Huang, Adversarial learning with multi-scale loss for skin lesion segmentation, in: 2018 IEEE 15th International Symposium on Biomedical Imaging, ISBI 2018, IEEE, 2018, pp. 859–863.

[13] M.M.K. Sarker, H.A. Rashwan, F. Akram, S.F. Banu, A. Saleh, V.K. Singh, F.U. Chowdhury, S. Abdulwahab, S. Romani, P. Radeva, et al., SLSDeep: Skin lesion segmentation based on dilated residual and pyramid pooling networks, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2018, pp. 21–29.

[14] M. Goyal, A. Oakley, P. Bansal, D. Dancey, M.H. Yap, Skin lesion segmentation in dermoscopic images with ensemble deep learning methods, IEEE Access 8 (2020) 4171–4181.

[15] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 801–818.

[16] S. Chen, Z. Wang, J. Shi, B. Liu, N. Yu, A multi-task framework with feature passing module for skin lesion classification and segmentation, in: 2018 IEEE 15th International Symposium on Biomedical Imaging, ISBI 2018, IEEE, 2018, pp. 1126–1129.

[17] H. Liao, J. Luo, A deep multi-task learning approach to skin lesion classification, 2018, arXiv preprint arXiv:1812.03527.

[18] S. Murabayashi, H. Iyatomi, Towards explainable melanoma diagnosis: Prediction of clinical indicators using semi-supervised and multi-task learning, in: 2019 IEEE International Conference on Big Data, Big Data, IEEE, 2019, pp. 4853–4857.

[19] E.Z. Chen, X. Dong, X. Li, H. Jiang, R. Rong, J. Wu, Lesion attributes segmentation for melanoma detection with multi-task U-net, in: 2019 IEEE 16th International Symposium on Biomedical Imaging, ISBI 2019, IEEE, 2019, pp. 485–488.

[20] X. Yang, Z. Zeng, S.Y. Yeo, C. Tan, H.L. Tey, Y. Su, A novel multi-task deep learning model for skin lesion segmentation and classification, 2017, arXiv preprint arXiv:1703.01025.

[21] C. Li, L. Bai, W. Liu, L. Yao, S. Waller, Knowledge adaption for demand prediction based on multi-task memory neural network, 2020, arXiv preprint arXiv:2009.05777.

[22] L. Aldieri, B. Bruno, L. Senatore, C.P. Vinci, The future of pharmaceuticals industry within the triad: The role of knowledge spillovers in innovation process, Futures 122 (2020) 102600.

[23] L. Aldieri, C.P. Vinci, Climate change and knowledge spillovers for cleaner production: New insights, J. Cleaner Prod. 271 (2020) 122729.

[24] A. Kuzina, E. Egorov, E. Burnaev, Bayesian generative models for knowledge transfer in MRI semantic segmentation problems, Front. Neurosci. 13 (2019) 844.

[25] F. Navarro, S. Conjeti, F. Tombari, N. Navab, Webly supervised learning for skin lesion classification, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2018, pp. 398–406.

[26] N.C.F. Codella, D. Gutman, M.E. Celebi, B. Helba, M.A. Marchetti, S.W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, A. Halpern, Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC), arXiv preprint arXiv:1710.05006.

[27] N.C. Codella, D. Gutman, M.E. Celebi, B. Helba, M.A. Marchetti, S.W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, et al., Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (isic), in: 2018 IEEE 15th International Symposium on Biomedical Imaging, ISBI 2018, IEEE, 2018, pp. 168–172.

[28] P. Tschandl, C. Rosendahl, H. Kittler, The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions, Sci. Data 5 (2018) 180161.

[29] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[30] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014, arXiv preprint arXiv:1409.1556.

[31] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.

[32] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, ImageNet: A large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2009, pp. 248–255.

[33] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, H. Lu, Dual attention network for scene segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 3146–3154.

[34] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, Learning deep features for discriminative localization, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2921–2929.

[35] A.G. Roy, N. Navab, C. Wachinger, Concurrent spatial and channel 'squeeze & excitation'in fully convolutional networks, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2018, pp. 421–429.

[36] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2980–2988.

[37] P. Tang, Q. Liang, X. Yan, S. Xiang, D. Zhang, GP-CNN-DTEL: Global-part CNN model with data-transformed ensemble learning for skin lesion classification, IEEE J. Biomed. Health Inf. (2020) 1.

[38] J. Zhang, Y. Xie, Q. Wu, Y. Xia, Medical image classification using synergic deep learning, Med. Image Anal. 54 (2019) 10–19.

[39] T. DeVries, D. Ramachandram, Skin lesion classification using deep multi-scale convolutional neural networks, 2017, arXiv preprint arXiv:1703.01402.

[40] L. Bi, J. Kim, E. Ahn, D. Feng, Automatic skin lesion analysis using large-scale dermoscopy images and deep residual networks, 2017, arXiv preprint arXiv:1703.04197.

[41] A. Menegola, J. Tavares, M. Fornaciali, L.T. Li, S. Avila, E. Valle, RECOD titans at ISIC challenge 2017, 2017, arXiv preprint arXiv:1703.04819.

[42] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.

[43] H. Li, X. He, F. Zhou, Z. Yu, D. Ni, S. Chen, T. Wang, B. Lei, Dense deconvolutional network for skin lesion segmentation, IEEE J. Biomed. Health Inf. 23 (2) (2018) 527–537.

[44] Y. Yuan, Y.-C. Lo, Improving dermoscopic image segmentation with enhanced convolutional-deconvolutional networks, IEEE J. Biomed. Health Inf. 23 (2) (2017) 519–526.

[45] Z. Mirikharaji, G. Hamarneh, Star shape prior in fully convolutional networks for skin lesion segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2018, pp. 737–745.

[46] Y. Yuan, Automatic skin lesion segmentation with fully convolutional-deconvolutional networks, 2017, arXiv preprint arXiv:1703.05165.

[47] M. Berseth, ISIC 2017-skin lesion analysis towards melanoma detection, 2017, arXiv preprint arXiv:1703.00523.

[48] L. Bi, D. Feng, M. Fulham, J. Kim, Improving skin lesion segmentation via stacked adversarial learning, in: 2019 IEEE 16th International Symposium on Biomedical Imaging, ISBI 2019, IEEE, 2019, pp. 1100–1103.

[49] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.

[50] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs, IEEE Trans. Pattern Anal. Mach. Intell. 40 (4) (2017) 834–848.

[51] M.Z. Alom, M. Hasan, C. Yakopcic, T.M. Taha, V.K. Asari, Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation, 2018, arXiv preprint arXiv:1802.06955.

[52] R. Azad, M. Asadi-Aghbolaghi, M. Fathy, S. Escalera, Bi-Directional ConvLSTM U-Net with densley connected convolutions, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2019.

[53] L. Van Der Maaten, Accelerating t-SNE using tree-based algorithms, J. Mach. Learn. Res. 15 (1) (2014) 3221–3245.