Deep Multi-Task Learning for Diabetic Retinopathy Grading in Fundus Images

Xiaofei Wang,[1] Mai Xu,[1]* Jicong Zhang,[2]* Lai Jiang,[1] Liu Li [3]

[1]School of Electronic and Information Engineering, Beihang University, Beijing, China
[2]School of Biological Science and Medical Engineering, Beihang University, Beijing, China
[3]Department of Computing, Imperial College London, London, UK
{xfwang, maixu, jicongzhang, jianglai.china}@buaa.edu.cn, liu.li20@imperial.ac.uk

Abstract

Recent years have witnessed the growing interest in disease severity grading, especially for ocular diseases based on fundus images. The existing grading methods are usually trained with high resolution (HR) images. However, the grading performance decreases a lot given low resolution (LR) images, which are common in practice. In this paper, we mainly focus on diabetic retinopathy (DR) grading with LR fundus images. According to our analysis on the DR task, we find that: 1) image super-resolution (ISR) can boost the performance of DR grading and lesion segmentation; 2) the lesion segmentation regions of fundus images are highly consistent with pathological regions for DR grading. Thus, we propose a deep multi-task learning based DR grading (DeepMT-DR) method for LR fundus images, which simultaneously handles the auxiliary tasks of ISR and lesion segmentation. Specifically, based on our findings, we propose a hierarchical deep learning structure that simultaneously processes the low-level task of ISR, the mid-level task of lesion segmentation and the high-level task of DR grading. Moreover, a novel task-aware loss is developed to encourage ISR to focus on the pathological regions for its subsequent tasks: lesion segmentation and DR grading. Extensive experimental results show that our DeepMT-DR method significantly outperforms other state-of-the-art methods for DR grading over two public datasets. In addition, our method achieves comparable performance in two auxiliary tasks of ISR and lesion segmentation.

Introduction

The past few years have witnessed the great success in the interdisciplinary research of artificial intelligence and medical image analysis (Zhou et al. 2019; Li et al. 2019a), especially for the task of severity grading of diabetic retinopathy (DR) on fundus images (Gulshan et al. 2016; Gargeya and Leng 2017), which is one of the leading cause of blindness and visual disability among working-age people. In clinic, the fundus images are sometimes at low resolution (LR) (Mahapatra et al. 2017), mainly due to the following reasons. 1) Many LR retinal imaging devices currently used in clinic are with only around 550×550 resolution (Singh et al. 2019), e.g., Topcon TRV-50 and Canon CR6–45NM (Weber and Mertz 2018). 2) The acquired fundus images in clinic are often down-sampled to LR before transmitting to the medical systems, e.g., PACS, due to the limited storage capacity and high cost of data migration (Alhajeri and Shah 2019). However, the existing DR grading methods focus on relatively high resolution (HR) images, which ignores the practical clinical condition. In this paper, we propose a hierarchical deep learning structure for DR grading on LR fundus images, which offers great potential in practical applications.

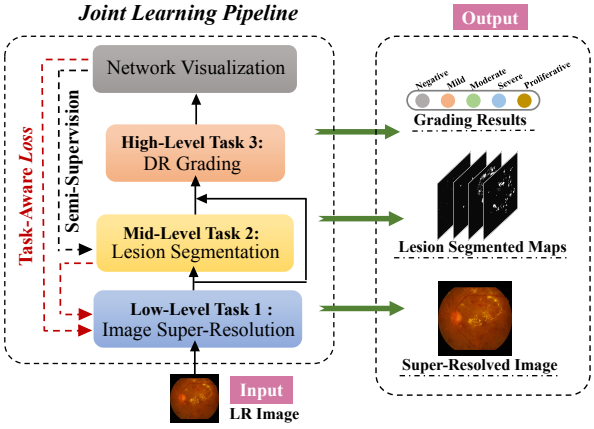

Figure 1: Brief framework of the proposed DeepMT-DR for DR grading with multiple tasks at low, mid and high-levels.

On the other hand, DR diagnosis in clinic highly depends on the detected retinal pathologies, such as microaneurysm. However, in LR fundus images, the small lesions are usually inconspicuous, which hinders the early detection of the disease. It is therefore intuitive to apply lesion segmentation and image super resolution (ISR) to boost the performance of DR grading with LR images. Accordingly, we thoroughly analyze the correlation among the three tasks of ISR, lesion segmentation and DR grading, and then find that: 1) ISR can actually boost the performance of DR grading and lesion segmentation and 2) the lesion segmentation regions of fundus images are highly consistent with pathological regions for DR grading. Thus, the above three tasks are closely correlated with each other, indicating the potential gain of DR grading with the lesion segmentation and ISR tasks.

*Corresponding Authors.

In this paper, we propose a deep multi-task learning based DR grading method for LR fundus images, called DeepMT-DR. Our DeepMT-DR consists of three tasks, i.e., ISR, lesion segmentation and DR grading. These three tasks can be regarded as low, mid and high-level vision tasks (Wang et al. 2018; Nathan 2018), respectively. It is therefore inappropriate to simply share the hidden layers across tasks as the conventional multi-task learning methods (Kendall, Gal, and Cipolla 2018) do. In this paper, we adopt an efficient hierarchy of these three tasks in our DeepMT-DR framework, consistent with our findings of the task correlation. We also propose a novel task-aware loss, to help the ISR process focus more on the pathological regions of both lesion segmentation and DR grading. To obtain the pathological regions for DR grading in fundus images, we propose a gradient-based multi-scale network visualization (GMSV) algorithm. Different from the traditional multi-scale visualization algorithm (Feng et al. 2017a), our GMSV leverages the gradient mechanism to obtain the importance weights of features. Besides, our GMSV is more flexibility and can be directly used in most CNN-based networks without modifying the network.

Contributions. To the best of our knowledge, our work is the first attempt to perform multiple medical tasks at low, mid and high-levels simultaneously. The main contributions of this paper are as follows. (1) We analyze the correlation among the tasks of ISR, lesion segmentation and DR grading, indicating that these three tasks can benefit from each other. (2) We propose a deep multi-task learning method for the main task of DR grading and the auxiliary tasks of both ISR and lesion segmentation. (3) Extensive experiments verify that our method not only achieves excellent performance in DR grading, but also obtains comparable results in the ISR and lesion segmentation.

Related Work

Medical Image Analysis in Fundus Images

In the field of MIA, disease grading, lesion segmentation and ISR are three important tasks, which are able to largely advance the progress of computer-aided diagnosis.

Disease Grading. Recently, deep-learning models have been widely used in the disease grading tasks (Litjens et al. 2017). Among them, many convolutional neural network (CNN) architectures (Wang et al. 2017; Zhao et al. 2019) have been proposed to diagnose DR for classifying its severity. For example, Gulshan *et al.* utilized Inception-V3 for DR grading (Gulshan et al. 2016), while Zhao *et al.* designed a pathology-aware network (Zhao et al. 2019). In practice, many fundus images are captured in LR; however, none of the existing works has used ISR for the precise DR grading on LR fundus images.

Lesion Segmentation. Lesion segmentation is an important task to provide pathological guidance to the diagnosis process. (Tan et al. 2017; Asiri et al. 2019). Most recently, taking fundus images as inputs, several U-shaped structures (Feng et al. 2017b; Kou et al. 2019) have been proposed to generate lesion segmentation with corresponding probability distribution. Specifically, Feng *et al.* (Feng et al. 2017b) proposed an U-shaped structure with short and long skip connections for exudates segmentation. Kou *et al.* designed a recurrent residual U-shaped structure for segmenting microaneurysms lesions (Kou et al. 2019). However, only a few works (Fan et al. 2018) have associated lesion segmentation with other tasks.

Image Super-Resolution. Recently, medical ISR methods have been widely used to improve the quality of indistinct pathological areas, especially for ocular diseases based on fundus images (Mahapatra et al. 2017; Ren et al. 2019). The existing ISR methods are mainly on the top of generative adversarial network (GAN)-based (Ledig et al. 2017; Mahapatra et al. 2017) or CNN-based (Ren et al. 2019) structures. Specifically, Mahapatra *et al.* (Mahapatra et al. 2017) proposed a saliency-based GAN network for ISR in retinal fundus images, while Ren *et al.* (Ren et al. 2019) designed a multi-scale ISR structure with deep residual connections.

Unfortunately, none of the previous works has studied the relationship of the above three tasks.

Multi-Task Learning

Multi-task learning (MTL) is a learning paradigm in machine learning, to leverage the shared information in related tasks (Anwar, Khan, and Barnes 2019). Most recently, the effectiveness of MTL with deep architecture has been verified in many vision tasks, such as ISR (Kim, Oh, and Kim 2019), object detection (He et al. 2017) and pose estimation (Ranjan, Patel, and Chellappa 2017). In general, the existing multi-task frameworks can be divided into two categories: hard parameter sharing (Kendall, Gal, and Cipolla 2018) and soft parameter sharing (Sun et al. 2019). Specifically, hard parameter sharing is applied by sharing the hidden layers among all tasks, while keeping several task-specific output layers independent. For example, Li *et al.* (Li et al. 2019c) designed a cross-disease attention network (CANet) for simultaneously grading the DR and diabetic macular edema. In soft parameter sharing, each task has its own parameters, but are encouraged to be similar with each other through certain regularization. For example, Sun *et al.* (Sun et al. 2019) proposed a joint framework of image reconstruction and segmentation for compressed sensing magnetic resonance imaging.

However, the existing methods with parameter sharing neglect the possible information feedback from the high-levels to low-levels. Thus, we propose a deep multi-task learning based DR grading method, which combines the three tasks in a hierarchical way, instead of simply sharing the parameters.

Task Correlation Analysis

In this section, we thoroughly analyze the correlation among the tasks of disease grading, lesion segmentation and ISR for DR problem, through which we demonstrate the potential gain of multi-task learning. According to the analysis, two findings are investigated as follows.

Finding 1: The performance of DR grading and lesion segmentation can be improved, when the resolution of input fundus images is increased.

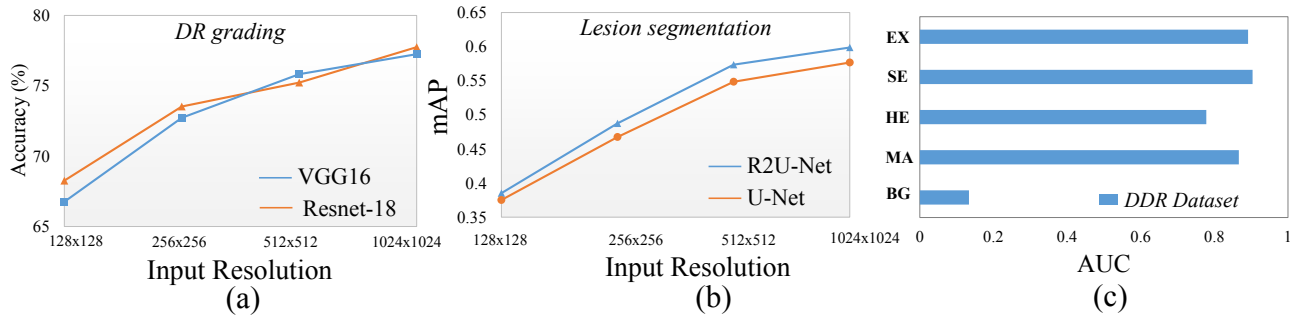

Figure 2: Task correlation analysis. (a) Accuracy of DR grading *vs.* input resolution ranging from 128 × 128 to 1024 × 1024. (b) Mean average precision (mAP) of retinal lesion segmentation *vs.* varying input resolutions. (c) AUC of DR grading network visualization maps in the background (BG) and lesion segmented regions of MA, HM, SE and HE.

Analysis: Figure 2 (a) shows the results of DR grading by 2 commonly used algorithms, at different input resolutions, over DDR database (Li et al. 2019b). Specifically, the fundus images are cropped and downsampled to 1024×1024 for unifying the size of input images. Then, these 1024×1024 images are further downsampled by 2, 4 and 8 scales. Consequently, each fundus image has four resolutions varying from 128×128 to 1024×1024, as the input to DR grading. Note that for fair comparison of DR grading at different resolutions, we use spatial pyramid pooling (SPP) (He et al. 2015) before the dense-connection layers in the algorithms. As can be seen in the Figure 2 (a), the grading accuracy obviously improves along with the increase of input resolution. This indicates the potential improvement of DR grading after applying ISR to fundus images.

Similarly, we also conduct lesion segmentation on fundus images at varying resolutions. For fair comparison, the low-resolution images are upsampled to 1024 × 1024 by the bicubic SR algorithm, and then supervised with the high-resolution (i.e., 1024 × 1024) segmentation labels. As can be seen in Figure 2 (b), the segmentation performance improves along with increased input resolution. This implies the lesion segmentation task can benefit from the ISR task.

Finding 2: The lesion segmentation regions of fundus images are highly consistent with pathological regions for DR grading.

Analysis: Here, we further study the correlation between lesion segmentation and DR grading. To be specific, we apply the proposed GMSV algorithm to generate the evidence map of the final decision from the DR grading network. We then calculate the AUC values between the evidence map and different lesion segmented regions (including microaneurysms (MA), haemorrhages (HM), soft exudates (SE) and hard exudates (HE)) as well as background (BG) regions. Figure 2 (c) shows that the AUC results of lesion segmented regions are significantly higher than those of background. This means that the lesion segmentation regions of fundus images are highly consistent with pathological regions for DR grading.

Methodology

Framework

According to the task correlation analysis, the tasks of ISR, lesion segmentation and DR grading are closely related with each other. Therefore, it is intuitive to jointly learn these three tasks in a unified framework. In this paper, we propose a novel DeepMT-DR framework for the main task of DR grading, simultaneously handling the auxiliary tasks of ISR and lesion segmentation. The framework of DeepMT-DR is shown in Figure 3. As seen in this figure, given an LR fundus image $\mathbf{X} \in \mathbb{R}^{H\times W\times 3}$ (with height H, width W and 3 channels of RGB) as the input, our DeepMT-DR model can output the DR severity grade $\hat{l} \in \mathbb{R}^5$, the super-resolved image $\hat{\mathbf{Y}}^1 \in \mathbb{R}^{4H\times 4W\times 3}$ and the segmented maps $\hat{\mathbf{S}} \in \mathbb{R}^{4H\times 4W\times 4}$ of 4 lesion types: MA, HM, SE and HE. Note that in our 5-class DR grading task, grade 0 to 4 refer to negative, mild, moderate, severe and proliferative DR, respectively.

Hierarchical Structure of DeepMT-DR

As shown in Figure 3, our DeepMT-DR consists of three subnets, i.e., ISR, lesion segmentation and DR grading subnets. The detailed structures are described as follows.

ISR Subnet. A schematic diagram of the ISR subnet is shown in the yellow part of Figure 3. As shown, the ISR subnet consists of several cascaded components, i.e., 5 feature extraction layers and 2 up-scaling layers. Specifically, the feature extraction layers are developed to extract the pathological information from LR fundus images. Then, for generating the super-resolved fundus image, the output of the feature extraction layers is processed with 2 up-scaling layers with upscale factor of 2 in each layer.

Lesion Segmentation Subnet. In the lesion segmentation subnet, we extend the U-Net (Ronneberger, Fischer, and Brox 2015) structure with the multi-scale residual module proposed in (Khened, Kollerathu, and Krishnamurthi 2019),

[1]Note that the upscale factor for ISR is set to 4 in this paper, and it can be easily extended to other values in practice.

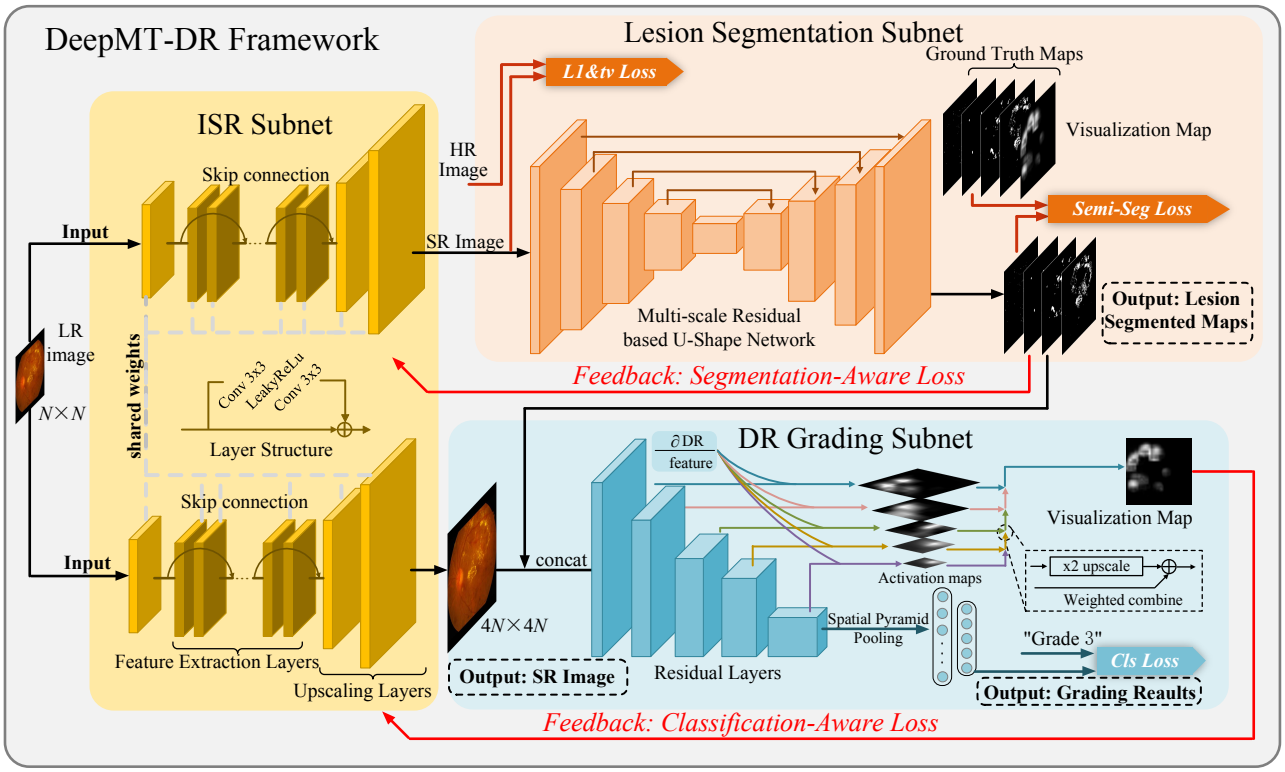

Figure 3: Framework of the proposed DeepMT-DR for DR grading with multiple tasks.

Algorithm 1: Gradient-weighted feature combination in GMSV.

Input: The input fundus image $\mathbf{X}$; the feature maps $\{\mathbf{F}^{i,j}\}_{i=1,j=1}^{K^j,J}$ of the DR grading network, and their corresponding weights $\{w^{i,j}\}_{i=1,j=1}^{K^j,J}$, where J is the total number of layers and K^j is the number of channels at the j-th layer.

Output: The visualization map $\mathbf{V}$ of the grading network.

for $j = J \rightarrow 1$ **do**
  $\mathbf{V}^j \leftarrow \sum_{i=1}^{K^j} w^{i,j}\mathbf{F}^{i,j}$, where $\mathbf{V}^j$ indicates the visualization map of the j-th layer.
  if $j < J$ **then**
    $w^j \leftarrow \sum_{i=1}^{K^j} w^{i,j}$, where weight w^j represents the importance of the j-th layer for DR grading.
    Upsample the $\mathbf{V}^{j+1}$ by a factor of 2, which is denoted as $\hat{\mathbf{V}}^{j+1}$.
    $\mathbf{V}^j \leftarrow \hat{\mathbf{V}}^{j+1} + w^j\mathbf{V}^j$.
  end
  $j \leftarrow j - 1$.
end
return $\mathbf{V} \leftarrow \mathbf{V}^1$

to capture the multi-scale information of retinal lesions. A schematic diagram of the lesion segmentation subnet is shown in the orange part of Figure 3. As shown, we modify the U-shape structure to be a multi-lesion generator to obtain the masks for different lesions, i.e., MA, HM, SE and HE. To be more specific, the encoder and decoder for the mask generator include nine feature mapping tuples with multi-scale residual modules.

DR Grading Subnet. Given the super-resolved image $\hat{\mathbf{Y}}$ and the lesion segmented map $\hat{\mathbf{S}}$ as the inputs, the DR grading subnet is developed to generate the DR severity grade. The structure of the DR grading subnet is illustrated in the blue part of Fig. 3. As illustrated, our DR grading subnet is developed on the top of ResNet-18 (He et al. 2016). Besides, an spatial pyramid pooling (SPP) (He et al. 2015) is added before the dense-connection layers, for utilizing different input resolutions. In addition to the grading result, we also generate the visualization map for the grading network with the proposed GMSV algorithm. Thus, the visualization map can be further used as information flow of feedback from the high-level tasks to the low-level task.

GMSV Algorithm for Information Feedback

In addition to the hierarchical structure, we also propose an information flow of feedback from the high-level tasks to the low-level task. Specifically, for obtaining the pathological regions of DR grading, we design a GMSV algorithm to generate the fine-grained evidence map of the final grading decision. In GMSV, let $\mathbf{F}^{i,j} \in \mathbb{R}^{M^j \times N^j}$ denote the i-th feature map in the j-th convolutional layer of the DR grading network, where M^j and N^j refer to the height and width of

the feature map. Then, the global-average-pooling (GAP) is applied to obtain the importance weights $w^{i,j}$ for $\mathbf{F}^{i,j}$:

$$w^{i,j} = \frac{1}{N^j \times M^j} \sum_{m=1}^{M^j} \sum_{n=1}^{N^j} A^{i,j}_{m,n}, \tag{1}$$

$$\text{where}[A^{i,j}_{m,n}]^{M_j,N_j}_{m=1,n=1} = \mathbf{A}^{i,j}, \mathbf{A}^{i,j} = \frac{\partial y_{\text{DR}}}{\partial \mathbf{F}^{i,j}}. \tag{2}$$

In (2), $\mathbf{A}^{i,j}$ indicates the gradients of the final positive DR score[2] y_{DR} with respect to the feature map $\mathbf{F}^{i,j}$. Finally, given the feature maps $\{\mathbf{F}^{i,j}\}^{K^j,J}_{i=1,j=1}$ and their corresponding importance weights $\{w^{i,j}\}^{K^j,J}_{i=1,j=1}$, we perform a gradient-weighted feature combination to obtain the final visualization map $\mathbf{V}$ as the input to the grading network, the process of which is summarized in Algorithm 1. Specifically, different from the traditional visualization algorithms that only focus on the last layer, we utilize the features of all layers in the network, which can extract fine-grained pathological information in multiple scales.

Loss Functions

Here, we introduce the loss functions for training our DeepMT-DR network, including the task-aware loss to guide the low-level task with information from the high-level tasks, and the intra-task loss for each single task. Details of the proposed loss functions are introduced as follows.

Task-Aware Loss. The task-aware loss is developed to guide ISR to focus more on the most relevant regions of its follow-up tasks. Specifically, the segmentation-aware loss and classification-aware loss are developed for utilizing the feedback information from lesion segmentation and DR grading, respectively. To be specific, let $\hat{\mathbf{S}}_{\mathbf{sr}}$ and $\hat{\mathbf{S}}_{\mathbf{hr}}$ denote the segmentation results of the super-resolved image and its corresponding HR image. The segmentation-aware loss is obtained by penalizing the mean square error (MSE) between $\hat{\mathbf{S}}_{\mathbf{sr}}$ and $\hat{\mathbf{S}}_{\mathbf{hr}}$:

$$\mathcal{L}_{\text{seg}-\text{aware}} = \|\hat{\mathbf{S}}_{\mathbf{sr}} - \hat{\mathbf{S}}_{\mathbf{hr}}\|_2^2. \tag{3}$$

With the guidance of segmentation-aware loss, the ISR subnet is able to focus more on the lesion segmented regions, leading to better segmentation results of the super-resolved image.

Similarly, the classification-aware loss encourages ISR to concentrate on the pathological regions of DR grading. Specifically, we use our GMSV algorithm to visualize the pathology-areas of DR grading from both super-resolved image $\hat{\mathbf{Y}}$ and corresponding HR image $\mathbf{Y}$. Given the GMSV algorithm, the classification-aware loss can be defined as follows:

$$\mathcal{L}_{\text{cls}-\text{aware}} = \left\| \left(\text{Vis}(\hat{\mathbf{Y}}) > \phi_{\text{vis}}\right) - \left(\text{Vis}(\mathbf{Y}) > \phi_{\text{vis}}\right) \right\|_2^2, \tag{4}$$

where $\text{Vis}(\cdot)$ refers to the visualization process of our GMSV algorithm; $\oplus$ denotes channel-wise concatenation; and ϕ_{vis} is the threshold to generate the binary mask. Benefiting from the classification-aware loss, the super-resolved image tends to highlight pathological regions for DR grading, leading to a better grading result.

Intra-Task Loss. In addition to the above task-aware loss, we further propose the intra-task loss to encourage our network to perform better in each single task, i.e, ISR, lesion segmentation and DR grading.

- For the ISR task, we measure the ℓ_1-norm difference between the super-resolved image $\hat{\mathbf{Y}}$ and the ground-truth HR image $\mathbf{Y}$ as follows:

  $$\mathcal{L}^{\text{img}}_{\text{isr}} = \|\hat{\mathbf{Y}} - \mathbf{Y}\|_1. \tag{5}$$

  Besides, to consider the spatial smoothness in the generated image, we calculate the total variation loss to $\hat{\mathbf{Y}}$:

  $$\mathcal{L}^{\text{tv}}_{\text{isr}} = \sum_{w=1}^{W} \sum_{h=1}^{H} (\|\hat{\mathbf{Y}}_{w,h+1} - \hat{\mathbf{Y}}_{w,h}\|_2^2 + \|\hat{\mathbf{Y}}_{w+1,h} - \hat{\mathbf{Y}}_{w,h}\|_2^2), \tag{6}$$

  where W and H are the width and height of the image $\hat{Y}$.

- For lesion segmentation, we develop a *semi-supervised* training method, which utilizes the fine-grained visualization maps as pseudo segmentation ground-truth maps to co-train the lesion segmentation subnet.

  First, for those data with segmentation annotation, we adopt dice loss (Fidon et al. 2017) to measure the overlap area between the segmentation result $\hat{\mathbf{S}}$ and its ground-truth lesion mask $\mathbf{S}$:

  $$\mathcal{L}^{\text{fully}}_{\text{seg}} = 1 - \frac{2\|\hat{\mathbf{S}} \circ \mathbf{S}\|_1}{\|\hat{\mathbf{S}}\|_1 + \|\mathbf{S}\|_1}, \tag{7}$$

  where $\circ$ denotes the hadamard product. Then, for those data without segmentation annotation, the semi-supervised segmentation loss is defined as follows:

  $$\mathcal{L}^{\text{un}}_{\text{seg}} = \|\hat{\mathbf{S}} - (\text{Vis}(\mathbf{Y}) > \phi_{\text{vis}})\|_2^2, \tag{8}$$

  where $\text{Vis}(\cdot)$ refers to the visualization process of our GMSV algorithm, and ϕ_{vis} is a shared parameter in (4).

- For the DR grading task, we develop a weighted cross-entropy loss to consider not only the classification accuracy but also the importance of different misclassified cases. Specifically, for each training sample, the classification loss is penalized by the distance between the predicted label $\hat{l}$ and its groundtruth label l. Mathematically, our loss function for the DR grading can be formulated as

  $$\mathcal{L}_{\text{cls}} = -\frac{|\hat{l} - l|}{\sum_{i=0}^{C-1}(|i-l|+1)} \sum_{j=0}^{C-1} 1\{j = l\} \log \hat{p_j} \tag{9}$$

  In (9), C is the total number of the DR severity grades; $\hat{p_j}$ indicates the predicted probability of the j-th grade and $1\{\cdot\}$ denotes the indicator function.

Overall Loss for ISR. Finally, by combining the task-aware and the intra-task losses for ISR, the overall loss function for our ISR subnet can be formulated as

$$\mathcal{L}_{\text{isr}} = \lambda_{\text{img}}\mathcal{L}^{\text{img}}_{\text{isr}} + \lambda_{\text{tv}}\mathcal{L}^{\text{tv}}_{\text{isr}} + \lambda_{\text{sa}}\mathcal{L}_{\text{seg}-\text{aware}} + \lambda_{\text{ca}}\mathcal{L}_{\text{cls}-\text{aware}}. \tag{10}$$

where $\lambda_{\text{img}}, \lambda_{\text{tv}}, \lambda_{\text{sa}}$ and λ_{ca} are hyper-parameters to balance the intra-task, segmentation-aware and classification-aware losses.

[2]We define the positive DR score by the sum of grades 1 to 4.

Experiments

Implementation Details

The scheme for training the DepepMT-DR model consists of two stages. In the first stage, we jointly train the subnets of the two auxiliary tasks, i.e., ISR and lesion segmentation, in order to extract sufficient pathological features for the main DR grading task. Besides, we also pre-train the DR grading subnet with the task of DR severity classification. In the second stage, we simultaneously fine-tune the subnets of ISR, lesion segmentation and DR grading, given the models pre-trained in the first stage. Besides, in both stages, the parameters are updated using the Adam (Kingma and Ba 2014) optimizer, together with the weight decay. The values of key hyper-parameters in the training stages are listed in Table 1. Note that all hyper-parameters are tuned over the validation set. All experiments are conducted on a computer with an Intel(R) Core(TM) i7-4770 CPU@3.40GHz, 32GB RAM and 4 Nvidia GeForce GTX 1080 Ti GPUs.

Stage I	Initial learning rate	1×10^{-4}
	λ_{img} for $\mathcal{L}_{\text{isr}}$ in equation (10)	1
	λ_{tv} for $\mathcal{L}_{\text{isr}}$ in equation (10)	1×10^{-6}
	λ_{sa} for $\mathcal{L}_{\text{isr}}$ in equation (10)	10
Stage II	Initial learning rate	5×10^{-5}
	Threshold ϕ_{vis} in equation (4) and (8)	0.5
	λ_{img} for $\mathcal{L}_{\text{isr}}$ in equation (10)	1
	λ_{tv} for $\mathcal{L}_{\text{isr}}$ in equation (10)	1×10^{-6}
	λ_{sa} for $\mathcal{L}_{\text{isr}}$ in equation (10)	1
	λ_{ca} for $\mathcal{L}_{\text{isr}}$ in equation (10)	10

Table 1: Values of some key hyper-parameters in the two training stages.

Datasets

In our experiments, we evaluate the performance of our DeepMT-DR method on two public DR datasets, i.e., the DDR dataset (Li et al. 2019b) and EyePACS dataset (Graham 2015). These two datasets have 13,673 and 88,702 retinal fundus images for DR grading, respectively. We use the default data split setting of these two datasets. Additionally, 757 fundus images of the DDR dataset are annotated with the pixel-wise segmentation for four retinal lesions, including MA, HM, SE and HE. All images are downsampled to 1024×1024, as HR images with the same resolution. Then, to obtain LR-HR pairs for training and test, all HR images are downsampled by a factor of 4 to generate LR images at resolution of 256×256.

Evaluation on DR Grading

We evaluate the DR grading performance of our DeepMT-DR method over DDR and EyePACS datasets, compared with 7 other state-of-the-art methods of CKML (Vo and Verma 2016), VNXK (Vo and Verma 2016), M-Net (Wang et al. 2017), AFN (Lin et al. 2018), MMCNN (Zhou et al. 2018), Adly (Adly, Ghoneim, and Youssif 2019) and Zhou (Zhou et al. 2019). Note that all compared methods are conducted on the fundus images super-resolved (using the SOTA ISR method (Ren et al. 2019)) or downsampled to the required resolutions of these methods. In our experiments, we apply 4 metrics to evaluate the performance of DR grading: accuracy, precision, F_1-score and the Cohen's kappa coefficient (Cohen 1960). Note that the larger values of these 4 metrics indicate more accurate DR grading. Table 2 tabulates the results for our and other methods. As shown, our DeepMT-DR method performs considerably better than all other methods over both datasets, in terms of all 4 metrics. Specifically, on the DDR dataset, our method achieves at least 3.0%, 0.6%, 2.4 % and 2.1% improvements in accuracy, precision, F_1-score and kappa, respectively. Similar results can be found in the EyePACS dataset.

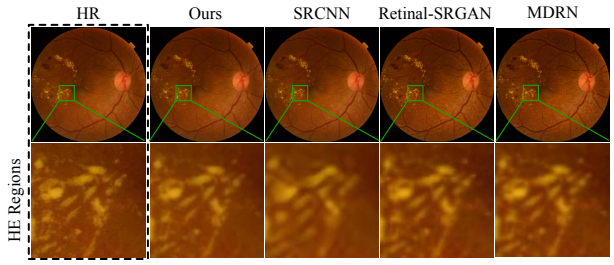

Figure 4: Qualitative ISR results of our DeepMT-DR and other methods. We crop and zoom-in the pathological region to compare the performance of ISR.

Evaluation on Auxiliary Tasks

In the following, we evaluate the performance of our DeepMT-DR method in these two auxiliary tasks of ISR and lesion segmentation, over DDR dateset.

Evaluation on ISR. We compare the ISR performance of our and 3 state-of-the-art ISR methods, i.e., SRCNN (Dong et al. 2014), Retinal-SRGAN (Mahapatra et al. 2017) and MDRN (Ren et al. 2019), over the DDR dataset, with the upscale factor of 4. Experimental results are shown in Table 3, in terms of Peak Signal-to-Noise Ratio (PSNR) and structural similarity (SSIM). As shown, our DeepMT-DR method outperforms all compared methods. Moreover, Figure 4 shows the super-resolved images generated by our DeeMT-DR and 3 compared methods. As shown, our method can clearly super-resolve the LR fundus images, in particular the pathological areas. The above results verify the effectiveness of our DeepMT-DR method in the task of ISR.

Evaluation on Lesion Segmentation. We further validate the performance of DeepMT-DR method in the auxiliary task of lesion segmentation, via comparing with 4 state-of-the-art methods: U-Net (Ronneberger, Fischer, and Brox 2015), DRU-Net (Kou et al. 2019), HGN (Sarhan et al. 2019) and CE-Net (Gu et al. 2019). Table 4 reports the results of average precision (AP) and area under the receiver operating characteristic curve (AUC). As shown, our DeepMT-DR method achieves the best performance for the lesions of MA, HM and HE. As such, we can conclude that our DeepMT-DR method is effective in the task of lesion segmentation.

	DDR						
	Ours	CKML	VNXK	M-Net	AFN	MMCNN	Adly
Accuracy	**82.5**(0.8)	77.5(0.6)	78.3(0.8)	78.9(0.5)	79.2(0.4)	79.5(0.3)	78.6(0.5)
Precision	**82.2**(0.5)	80.0(0.7)	81.6(1.2)	80.1(0.8)	80.6(0.6)	81.6(0.8)	81.2(0.7)
F_1-score	**81.8**(0.6)	77.0(0.6)	78.2(0.7)	78.4(0.3)	78.7(1.0)	79.4(0.4)	78.3(1.1)
Kappa	**77.8**(0.8)	73.3(0.4)	74.7(0.6)	74.2(0.5)	74.9(0.3)	75.7(0.8)	73.9(0.4)
	EyePACS						
	Ours	CKML	VNXK	M-Net	AFN	MMCNN	Adly
Accuracy	**85.7**(0.7)	81.1(0.6)	81.8(0.7)	81.8(0.6)	82.8(0.5)	83.1(0.6)	81.3(0.6)
Precision	**86.0**(0.8)	81.3(0.4)	82.1(0.5)	82.7(0.4)	83.3(0.2)	83.7(0.2)	82.3(0.4)
F_1-score	**83.9**(0.5)	80.7(0.5)	80.8(0.6)	82.5(0.6)	82.2(0.5)	82.9(0.7)	81.6(0.4)
Kappa	**83.7**(0.6)	80.1(0.5)	80.6(0.3)	81.3(0.4)	81.8(0.7)	82.6(0.8)	80.9(0.5)

Table 2: Mean (standard deviation) values in terms of percentage for DR grading metrics by our and other methods over the DDR and EyePACS datasets.

	Ours	SRCNN	Retinal-SRGAN	MDRN
PSNR	**39.7**	36.8	38.1	38.9
SSIM	**0.883**	0.772	0.836	0.856

Table 3: Mean values in terms of percentage for PSNR (dB) and SSIM by our and other ISR methods over DDR datasets.

Leiosns	MA		HM		HE		SE	
Methods	AUC	AP	AUC	AP	AUC	AP	AUC	AP
U-Net	95.7	45.5	95.3	49.4	96.5	65.5	97.2	58.4
DRU-Net	98.1	49.3	95.8	50.4	98.1	67.0	98.5	62.8
HGN	97.4	48.1	96.1	51.8	98.4	67.7	98.6	63.1
CE-Net	97.6	48.5	96.4	52.3	98.2	67.6	**99.4**	**65.2**
Ours	98.7	50.1	97.1	55.4	99.0	71.8	99.2	64.7

Table 4: Mean values in terms of percentage for AUC and AP of our and other lesion segmentation methods over DDR dataset.

Ablation Study

We ablate different components of our DeepMT-DR method to thoroughly analyze their effects on DR grading.

Ablation on ISR. We further conduct the ablation experiments on ISR in our DeepMT-DR method with the following 3 experimental settings: (1) Removing the ISR subnet and the related task-aware loss; (2) training ISR independently, i.e., the ISR subnet is fixed in the second training stage; (3) training the 3 tasks together, i.e., our DeepMT-DR method. Figure 5 (a) shows the DR grading results with above settings. As shown, the DR grading performance of DeepMT-DR greatly degrades, when ISR is removed or trained independently. This indicates the effectiveness of our framework with ISR task.

Ablation on Lesion Segmentation. Then, we conduct ablation experiments to evaluate the impact of lesion segmentation (LS) on DR grading. Specifically, we compare the performance of DR grading under 3 experimental settings: (1) Removing the LS subnet and related segmentation-aware loss; (2) training LS independently, i.e., jointly train the ISR and DR grading subnets with segmented masks generated from the pre-trained LS subnet; (3) training the 3 tasks together, i.e., our DeepMT-DR method. DR grading results with above ablation settings are shown in Figure 5 (b). As shown, our DeepMT-DR method performs worse, when LS is removed or trained independently. This indicates that LS has positive impact as an auxiliary task.

Ablation on Task-Aware Loss. Moreover, we conduct ablation experiments to evaluate the effect of our task-aware loss, by removing the segmentation-aware loss (SAL) and classification-aware loss (CAL). Figure 5 (c) presents the DR grading results of our DeepMT-DR method under 4 experimental settings. As shown, both the SAL and CAL contribute to the DR grading performance. To conclude, the above results verify the effectiveness of the task-aware loss.

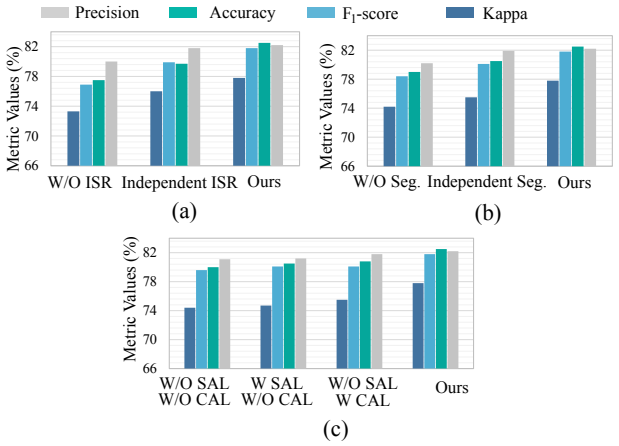

Figure 5: DR grading results of ablation study on ISR (a), lesion segmentation (b) and task-aware loss (c) over DDR dataset.

	GBP	Grad-CAM	SmoothBP	Ours
Accuracy	81.4	81.5	81.7	**82.5**

Table 5: Mean values in terms of percentage for DR grading accuracy by our and other network visualization algorithms over DDR datasets.

Ablation on GMSV algorithm. Finally, we evaluate the impact of our GMSV algorithm on DR grading. Specifically,

we conduct ablation experiments by replacing the GMSV algorithm with other 3 commonly used network visualization methods, including GBP (Springenberg et al. 2014), Grad-CAM (Selvaraju et al. 2017) and SmoothBP (Smilkov et al. 2017). The DR grading results are shown in Table 5. As shown, the DR grading performance degrades when using other visualization algorithms. This indicates the advantage of our GMSV algorithm in extracting pathological information for DR grading.

Conclusions

This paper has proposed a multi-task learning method, called DeepMT-DR, for the main task of DR grading and the auxiliary tasks of ISR and lesion segmentation. Based on the task correlation analysis, we proposed our DeepMT-DR method with a hierarchical structure and task-aware loss. Besides, the GMSV algorithm was proposed to supervise the ISR task with feedback information from the high-level tasks. Finally, experimental results verified that our DeepMT-DR method greatly outperforms other state-of-the-art methods in DR grading, and also achieves excellent performance in the ISR and lesion segmetation tasks.

Acknowledgments

This work was supported by the National Nature Science Foundation of China (Grant Number: 62050175, 61922009, 61876013), by Beijing Natural Science Foundation (Grant Number: JQ20020, Z200024), by the National Key Research and Development Program of China (Grant Number: 2016YFF0201002) and by the University Synergy Innovation Program of Anhui Province (Grant Number: GXXT-2019-044).

References

Adly, M. M.; Ghoneim, A. S.; and Youssif, A. A. 2019. On the grading of diabetic retinopathies using a binary-tree-based multiclass classifier of cnns. *International Journal of Computer Science and Information Security* 17(1).

Alhajeri, M.; and Shah, S. G. S. 2019. Limitations in and solutions for improving the functionality of Picture Archiving and Communication System: An exploratory study of PACS professionals' perspectives. *Journal of digital imaging* 32(1): 54–67.

Anwar, S.; Khan, S.; and Barnes, N. 2019. A deep journey into super-resolution: A survey. *arXiv preprint arXiv:1904.07523* .

Asiri, N.; Hussain, M.; Al Adel, F.; and Alzaidi, N. 2019. Deep learning based computer-aided diagnosis systems for diabetic retinopathy: A survey. *Artificial intelligence in medicine* .

Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement* 20(1): 37–46.

Dong, C.; Loy, C. C.; He, K.; and Tang, X. 2014. Learning a deep convolutional network for image super-resolution. In *ECCV*, 184–199.

Fan, Z.; Sun, L.; Ding, X.; Huang, Y.; Cai, C.; and Paisley, J. 2018. A segmentation-aware deep fusion network for compressed sensing MRI. In *ECCV*, 55–70.

Feng, X.; Yang, J.; Laine, A. F.; and Angelini, E. D. 2017a. Discriminative localization in CNNs for weakly-supervised segmentation of pulmonary nodules. In *MICCAI*, 568–576. Springer.

Feng, Z.; Yang, J.; Yao, L.; Qiao, Y.; Yu, Q.; and Xu, X. 2017b. Deep retinal image segmentation: A fcn-based architecture with short and long skip connections for retinal image segmentation. In *ICNIP*, 713–722.

Fidon, L.; Li, W.; Garcia-Peraza-Herrera, L. C.; Ekanayake, J.; Kitchen, N.; Ourselin, S.; and Vercauteren, T. 2017. Generalised wasserstein dice score for imbalanced multi-class segmentation using holistic convolutional networks. In *MICCAI Workshop*, 64–76.

Gargeya, R.; and Leng, T. 2017. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology* 124(7): 962–969.

Graham, B. 2015. Kaggle diabetic retinopathy detection competition report. *University of Warwick* .

Gu, Z.; Cheng, J.; Fu, H.; Zhou, K.; Hao, H.; Zhao, Y.; Zhang, T.; Gao, S.; and Liu, J. 2019. CE-Net: Context Encoder Network for 2D Medical Image Segmentation. *TMI* .

Gulshan, V.; Peng, L.; Coram, M.; Stumpe, M. C.; Wu, D.; Narayanaswamy, A.; ; et al. 2016. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama* 316(22): 2402–2410.

He, K.; Gkioxari, G.; Dollár, P.; and Girshick, R. 2017. Mask r-cnn. In *CVPR*, 2961–2969.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition. *TPAMI* 37(9): 1904–1916.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *CVPR*, 770–778.

Kendall, A.; Gal, Y.; and Cipolla, R. 2018. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *CVPR*, 7482–7491.

Khened, M.; Kollerathu, V. A.; and Krishnamurthi, G. 2019. Fully convolutional multi-scale residual DenseNets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers. *MIA* 51: 21–45.

Kim, S. Y.; Oh, J.; and Kim, M. 2019. Deep SR-ITM: Joint Learning of Super-resolution and Inverse Tone-Mapping for 4K UHD HDR Applications. In *CVPR*, 3116–3125.

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* .

Kou, C.; Li, W.; Liang, W.; Yu, Z.; and Hao, J. 2019. Microaneurysms segmentation with a U-Net based on recurrent residual convolutional neural network. *Journal of Medical Imaging* 6(2): 025008.

Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. 2017. Photo-realistic single image super-resolution using a generative adversarial network. In *CVPR*, 4681–4690.

Li, L.; Xu, M.; Wang, X.; Jiang, L.; and Liu, H. 2019a. Attention based glaucoma detection: A large-scale database and CNN Model. In *CVPR*, 10571–10580.

Li, T.; Gao, Y.; Wang, K.; Guo, S.; Liu, H.; and Kang, H. 2019b. Diagnostic Assessment of Deep Learning Algorithms for Diabetic Retinopathy Screening. *Information Sciences* .

Li, X.; Hu, X.; Yu, L.; Zhu, L.; Fu, C.-W.; and Heng, P.-A. 2019c. CANet: Cross-disease Attention Network for Joint Diabetic Retinopathy and Diabetic Macular Edema Grading. *TMI* .

Lin, Z.; Guo, R.; Wang, Y.; Wu, B.; Chen, T.; Wang, W.; Chen, D. Z.; and Wu, J. 2018. A Framework for Identifying Diabetic Retinopathy Based on Anti-noise Detection and Attention-Based Fusion. In *MICCAI*, 74–82.

Litjens, G.; Kooi, T.; Bejnordi, B. E.; Setio, A. A. A.; Ciompi, F.; Ghafoorian, M.; Van Der Laak, J. A.; Van Ginneken, B.; and Sánchez, C. I. 2017. A survey on deep learning in medical image analysis. *MIA* 42: 60–88.

Mahapatra, D.; Bozorgtabar, B.; Hewavitharanage, S.; and Garnavi, R. 2017. Image super resolution using generative adversarial networks and local saliency maps for retinal image analysis. In *MICCAI*, 382–390.

Nathan, S. S. 2018. A Review: Image Analysis Techniques to Improve Labeling Accuracy of Medical Image Classification. In *SCDM*, volume 700, 298.

Ranjan, R.; Patel, V. M.; and Chellappa, R. 2017. Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. *TPAMI* 41(1): 121–135.

Ren, S.; Jain, D. K.; Guo, K.; Xu, T.; and Chi, T. 2019. Towards efficient medical lesion image super-resolution based on deep residual networks. *SPIC* 75: 1–10.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, 234–241.

Sarhan, M. H.; Albarqouni, S.; Yigitsoy, M.; Navab, N.; and Eslami, A. 2019. Multi-scale Microaneurysms Segmentation Using Embedding Triplet Loss. In *MICCAI*, 174–182.

Selvaraju, R. R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; and Batra, D. 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *ICCV*, 618–626.

Singh, J.; Kabbara, S.; Conway, M.; Peyman, G.; and Ross, R. D. 2019. Innovative Diagnostic Tools for Ophthalmology in Low-Income Countries. In *Novel Diagnostic Methods in Ophthalmology*. IntechOpen.

Smilkov, D.; Thorat, N.; Kim, B.; Viégas, F.; and Wattenberg, M. 2017. Smoothgrad: removing noise by adding noise. *arXiv preprint arXiv:1706.03825* .

Springenberg, J. T.; Dosovitskiy, A.; Brox, T.; and Riedmiller, M. 2014. Striving for simplicity: The all convolutional net. *arXiv preprint arXiv:1412.6806* .

Sun, L.; Fan, Z.; Ding, X.; Huang, Y.; and Paisley, J. 2019. Joint CS-MRI reconstruction and segmentation with a unified deep network. In *IPMI*, 492–504.

Tan, J. H.; Fujita, H.; Sivaprasad, S.; Bhandary, S. V.; Rao, A. K.; Chua, K. C.; and Acharya, U. R. 2017. Automated segmentation of exudates, haemorrhages, microaneurysms using single convolutional neural network. *Information sciences* 420: 66–76.

Vo, H. H.; and Verma, A. 2016. New deep neural nets for fine-grained diabetic retinopathy recognition on hybrid color space. In *ISM*, 209–215.

Wang, X.; Yu, K.; Dong, C.; and Change Loy, C. 2018. Recovering realistic texture in image super-resolution by deep spatial feature transform. In *CVPR*, 606–615.

Wang, Z.; Yin, Y.; Shi, J.; Fang, W.; Li, H.; and Wang, X. 2017. Zoom-in-net: Deep mining lesions for diabetic retinopathy detection. In *MICCAI*, 267–275.

Weber, T. D.; and Mertz, J. 2018. Retina and choroid imaging with transcranial back-illumination. In *Clinical and Translational Biophotonics*, CF3B–8. Optical Society of America.

Zhao, Z.; Zhang, K.; Hao, X.; Tian, J.; Chua, M. C. H.; Chen, L.; and Xu, X. 2019. BiRA-Net: Bilinear Attention Net for Diabetic Retinopathy Grading. In *ICIP*, 1385–1389.

Zhou, K.; Gu, Z.; Liu, W.; Luo, W.; Cheng, J.; Gao, S.; and Liu, J. 2018. Multi-Cell Multi-Task Convolutional Neural Networks for Diabetic Retinopathy Grading. In *EMBC*, 2724–2727.

Zhou, Y.; He, X.; Huang, L.; Liu, L.; Zhu, F.; Cui, S.; and Shao, L. 2019. Collaborative Learning of Semi-Supervised Segmentation and Classification for Medical Images. In *CVPR*, 2079–2088.