Medical Image Analysis 84 (2023) 102703

Contents lists available at ScienceDirect

Medical Image Analysis

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

A generalizable and robust deep learning algorithm for mitosis detection in multicenter breast histopathological images

Xiyue Wang [a,b], Jun Zhang [c], Sen Yang [c], Jingxi Xiang [c], Feng Luo [d], Minghui Wang [a,b], Jing Zhang [a,*], Wei Yang [c], Junzhou Huang [c], Xiao Han [c,*]

[a] College of Biomedical Engineering, Sichuan University, Chengdu 610065, China
[b] College of Computer Science, Sichuan University, Chengdu 610065, China
[c] Tencent AI Lab, Shenzhen 518057, China
[d] Shenzhen International Graduate School, Tsinghua University, Shenzhen 518057, China

ARTICLE INFO

Keywords:
Histopathology
Mitosis detection
Domain shift
Feature extraction
Deep learning

ABSTRACT

Mitosis counting of biopsies is an important biomarker for breast cancer patients, which supports disease prognostication and treatment planning. Developing a robust mitotic cell detection model is highly challenging due to its complex growth pattern and high similarities with non-mitotic cells. Most mitosis detection algorithms have poor generalizability across image domains and lack reproducibility and validation in multicenter settings. To overcome these issues, we propose a generalizable and robust mitosis detection algorithm (called FMDet), which is independently tested on multicenter breast histopathological images. To capture more refined morphological features of cells, we convert the object detection task as a semantic segmentation problem. The pixel-level annotations for mitotic nuclei are obtained by taking the intersection of the masks generated from a well-trained nuclear segmentation model and the bounding boxes provided by the MIDOG 2021 challenge. In our segmentation framework, a robust feature extractor is developed to capture the appearance variations of mitotic cells, which is constructed by integrating a channel-wise multi-scale attention mechanism into a fully convolutional network structure. Benefiting from the fact that the changes in the low-level spectrum do not affect the high-level semantic perception, we employ a Fourier-based data augmentation method to reduce domain discrepancies by exchanging the low-frequency spectrum between two domains. Our FMDet algorithm has been tested in the MIDOG 2021 challenge and ranked first place. Further, our algorithm is also externally validated on four independent datasets for mitosis detection, which exhibits state-of-the-art performance in comparison with previously published results. These results demonstrate that our algorithm has the potential to be deployed as an assistant decision support tool in clinical practice. Our code has been released at https://github.com/Xiyue-Wang/1st-in-MICCAI-MIDOG-2021-challenge.

1. Introduction

According to 2020 American cancer statistics, breast cancer is one of the most common cancers in women, accounting for 30% of the female cancers (Siegel et al., 2020). Nottingham grading system is adopted as the clinical standard for breast cancer grading, where mitosis counting, nuclear pleomorphism, and tubule formation are three factors to be considered (Elston and Ellis, 1991). Among them, mitosis counting is a key biomarker for patient prognostication since mitotic density is highly correlated with cell proliferation activity. In the clinic, the number of mitotic cells is obtained by manually counting the nuclei undergoing mitosis in 8–10 microscope high power fields (HPFs) (Balkenhol et al., 2019; Sohail et al., 2021a). This repetitive task is very time-consuming and labor-intensive for pathologists. In addition, there is a high intra- and inter-observer variability due to the heterogeneous appearances of mitotic cells (Yousif et al., 2021; Pan et al., 2021). Automatic mitosis detection technology is urgently required to alleviate the tedious manual counting and the wide variability between observers, helping provide pathologists with fast, accurate, and reproducible aided diagnosis.

Nevertheless, automatic mitosis detection is quite challenging due to the highly varied characteristics of mitotic cells and different appearances of histopathological images caused by domain discrepancies.

* Corresponding authors.
E-mail addresses: jing_zhang@scu.edu.cn (J. Zhang), haroldhan@tencent.com (X. Han).

https://doi.org/10.1016/j.media.2022.102703
Received 20 March 2022; Received in revised form 16 September 2022; Accepted 21 November 2022
Available online 23 November 2022

First, there is extensive appearance variability across mitotic cells and high similarity between mitotic and non-mitotic cells. Mitosis (i.e., cell division) has a complex growth pattern that consists of four stages (prophase, metaphase, anaphase, and telophase), with a quite different cell morphology for each stage (Wilkins and Holliday, 2009). Especially in telophase, a nucleus is divided into two parts, but it can only be counted as one mitotic cell. In addition, some non-mitosis cells (e.g., apoptotic or necrotic nuclei, dense nuclei, compressed nuclei, "junk" particles, and other artifacts) differ little from mitotic cells (Veta et al., 2015), which complicates the process of mitosis counting. Second, histopathological images from different laboratories exhibit heterogeneous appearances due to differences in their slide preparation and image acquisition processes (Madabhushi and Lee, 2016), which is known as a domain shift problem. In the slide preparation process, variations may be introduced by different stains, reagents, manufacturers, protocol, room conditions, etc (Schömig-Markiefka et al., 2021; Madabhushi and Lee, 2016). During slide scanning, different scanner types and parameter settings (e.g., image compression level, color scheme, brightness, and contrast) can also cause domain shifts (Schömig-Markiefka et al., 2021). As highlighted by the MIDOG 2021 challenge (Aubreville et al., 2022), different scanners can lead to large variations in the image appearance of mitoses. In fact, as also pointed out in previous studies (Leo et al., 2016; Aubreville et al., 2020, 2022), the effect of scanner-induced domain shift may be stronger than that caused by differences in slide preparation. Fig. 1 shows some diverse breast histopathological images with mitotic cells from different laboratories and scanners. These serious domain variations make it very challenging to develop a generalizable and robust mitosis detection algorithm across diverse cohorts.

There have been some studies on automatic mitosis detection, which can be divided into patch-classification-based (Maroof et al., 2020; Saha et al., 2018; Sigirci et al., 2021), object-detection-based (Sebai et al., 2020), semantic-segmentation-based (Li et al., 2019a), and hybrid methods (Li et al., 2018; Mahmood et al., 2020; Sohail et al., 2021a; Lei et al., 2020; Nateghi et al., 2021; Tellez et al., 2018). Despite the promising results of these methods, there are still three aspects that could be further improved. First, a simple and effective feature extractor is required. Among these approaches, various feature extraction methods have been proposed to cover the complex appearance of mitotic figures, including traditional features (e.g., morphological, statistics, and color features) (Maroof et al., 2020; Saha et al., 2018; Sigirci et al., 2021) and deep-learning-based features (Li et al., 2019a, 2018; Sebai et al., 2020; Mahmood et al., 2020; Sohail et al., 2021a; Lei et al., 2020; Nateghi et al., 2021). To further improve the robustness of deep features, a series of strategies have been proposed, including two-step learning (Li et al., 2018; Mahmood et al., 2020; Sohail et al., 2021a; Lei et al., 2020; Nateghi et al., 2021) and model ensemble (Mahmood et al., 2020; Tellez et al., 2018; Sohail et al., 2021a). However, both two strategies greatly increase the complexity and the training time of the network. Second, current mitosis detection algorithms lack reproducibility and independent validation on multicenter cohorts. These methods do not sufficiently consider the domain shift that is widespread in the field of histopathological images (Madabhushi and Lee, 2016), which makes the performance of mitosis detection algorithms degrade dramatically on out-of-domain data, resulting in impractical clinical trials. Third, the current publicly available datasets used for mitosis detection are very small and have limited sample diversity. They are publicly available challenge datasets (e.g., AMIDA13 Veta et al., 2015, MITOSIS14 MITOS-ATYPIA-14, 2014, and TUPAC-auxiliary Veta et al., 2019), where TUPAC-auxiliary is the largest dataset with 73 WSIs.

To address the above-mentioned problems, this work proposes a novel mitosis detection algorithm (called FMDet) on breast histopathological images, which is performed by converting the detection problem into a semantic segmentation one to capture more refined morphological features of cells. To generalize to unseen domain data, we adopt a Fourier-based data augmentation method, which exchanges low-frequency and high-frequency components between image pairs of two domains. To better extract robust feature representations, we integrate a channel-wise attention mechanism into the U-Net framework, which helps extract multi-scale features. Our algorithm is developed on much larger training data (200 WSIs with 8000 × 8000 pixels) of the MIDOG 2021 challenge. And then, it is evaluated by the data in unseen domains (e.g., the independent test data in the MIDOG 2021 challenge and four external datasets), which demonstrates state-of-the-art performance. Our main contributions are summarized as follows.

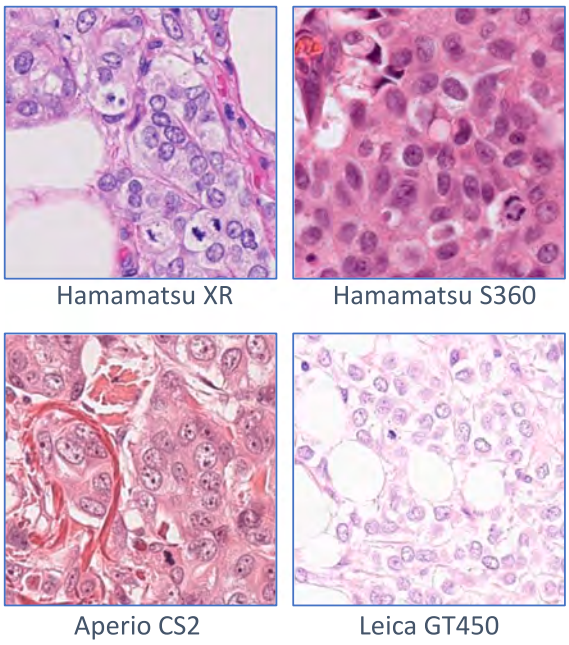

Fig. 1. The heterogeneous appearance caused by different scanners. These four examples are from breast cancer patients in the MIDOG 2021 challenge.

1. We propose a generalizable and effective mitosis detection algorithm, which is one of the pioneer studies to fully consider the mismatch across domains.
2. We transform the mitosis detection task into a semantic segmentation one to achieve precise mitotic figure identification. An attention mechanism is utilized to generate channel-wise multi-scale features, which helps the network to extract more robust features.
3. To reduce the domain gaps, we adopt an unsupervised method to align the domain discrepancies by exchanging the low-frequency spectrum between domains.
4. Our method ranked first place at the MIDOG 2021 challenge that aims to drive solutions regardless of different scanners. Moreover, our algorithm also shows a successful generalization to four unseen independent multicenter datasets. These results fully verify the effectiveness of our mitosis detection algorithm.

2. Related works

This section first briefly reviews recent mitosis detection methods. Then, domain adaptation/generalization strategies used in histopathological images are also reviewed since domain shift is a key factor to affect model generalization.

2.1. Mitotic cell detection

With the emergence of publicly available datasets (e.g., MITOSIS14, AMIDA13, and TUPAC16), a series of automatic mitosis detection algorithms have emerged. Early algorithms typically combined handcrafted features (e.g., shape, texture, color, etc.) and machine learning (or simple deep learning) techniques for mitotic figure classification (Maroof et al., 2020; Tellez et al., 2018; Beevi et al., 2019; Hwang et al., 2020; Saha et al., 2018; Sigirci et al., 2021). These methods usually adopted some threshold constraints to find candidate mitotic cells and then applied a support vector machine (SVM), linear classifier, or CNN architecture on these selected regions for further mitosis verification. Maroof et al. (2020) applied threshold techniques on blue ratio images to select candidate nuclei and extracted nuclear features (e.g., color, morphology, and texture features). Then, SVM was used to classify these candidate nuclei into mitosis and non-mitosis. Tellez et al. (2018) found candidate mitotic cells by iteratively detecting all circular areas with a threshold constraint for their diameters, and then applied CNN architecture to further verify these candidate mitoses. Beevi et al. (2019) utilized a localized active contour model and krill herd algorithm to localize nuclei and then applied a pre-trained VGG to classify these corresponding cells into mitosis and non-mitosis. Hwang et al. (2020) used a fuzzy thresholding scheme to segment potential mitotic nuclei and a linear attention-based classifier for further mitosis identification. Saha et al. (2018) and Sigirci et al. (2021) incorporated handcrafted traditional features and high-level deep-learning-based features for mitosis detection. In the above-mentioned methods, their candidate mitosis detection process involves various thresholding techniques that may produce high false-positive and false-negative regions.

The recently emerged deep learning techniques for mitosis detection can be divided into three categories: (1) object detection, (2) combination of object detection and binary classification, and (3) semantic segmentation. The first type of method performed end-to-end learning and directly adopted these detection results as the final mitotic cells, where Faster-RCNN was used as the main framework (Sebai et al., 2020). The second type of approach employed a two-stage strategy, where candidate mitotic cells were first detected and then further classified by a classification model. The utilized detection methods included Faster-RCNN (Li et al., 2018; Mahmood et al., 2020; Sohail et al., 2021a), customized region proposal network (RPN) (Lei et al., 2020), and patch-based classification (Nateghi et al., 2021). The binary classification process aims to further identify mitosis regions, where a ResNet50 (Li et al., 2018), a multi-branch classification (Lei et al., 2020), or an ensemble of two or more CNN frameworks (Mahmood et al., 2020; Tellez et al., 2018; Sohail et al., 2021a) were used. The last semantic segmentation method transformed the detection problem as a pixel-level classification one (Li et al., 2019a), which directly inferred the precise location of mitosis from the segmentation maps. However, these above-mentioned studies do not sufficiently take into account the differences between domains.

2.2. Domain shift in histopathology

Domain shift is defined as the data heterogeneity between the source and target domains (Challen et al., 2019), leading to a poor generalization of models trained on the source data to the unseen target data. For histopathological images, domain shift is mainly caused by different slide preparation processes and microscope scanning, and even within the same institution, appearance discrepancies could occur over time. Several methods have been presented to mitigate the domain shift problem, which can be divided into three categories: data augmentation, stain normalization, and domain-invariant representation.

Data augmentation methods empower models to address various image variations by increasing the data diversity, which includes two techniques: rule-based transformation and generative adversarial networks (GAN)-based data generation. (1) For the rule-based transformation, the commonly used data generation methods include rotation, contrast, sharpness, color distortion, etc (Faryna et al., 2021). Among them, color-based augmentation is the main method used in histopathological images since color change is the main manifestation of stain variations. Some studies applied a color transformation to training data by adding color distortions on each RGB channel (Lafarge et al., 2017) or on hematoxylin, eosin, and residual channels (Tellez et al., 2018). Recently, Chang et al. (2021) proposed a stain mix-up augmentation method by mixing the stain color matrix between the source and target domains. Although the stain mix-up yields a more realistic stain color compared to the above-mentioned color augmentations, it requires access to the information of the target domain. In realistic scenes, the data of the target domain are often invisible and unpredictable. More recently, several CNN-based image transformation algorithms have emerged, including the neural style transfer method that aligns the mean and variance of content features across styles (Yamashita et al., 2021) and the vector-multiplication-based kernel adaptation method that aligns the distribution of convolutional kernels across domains (Huang et al., 2017). (2) For GAN-based data generation methods, cycleGAN and conditional GAN have been widely used for data augmentation (Mahmood et al., 2019; de Haan et al., 2021; Xue et al., 2021). The cycleGAN aims to learn the mapping between domains, which has been used for synthesizing training data for the nuclear segmentation task (Mahmood et al., 2019), generating realistic samples for histopathological image classification (Xue et al., 2021), or transferring stain style between different stains (de Haan et al., 2021).

Stain normalization has been widely used in the field of digital pathology (Chen et al., 2021), which aims to reduce the appearance variation (Stacke et al., 2020). The traditional stain normalization method requires first selecting an image as the reference and then matching the color spectrum of all images to that of the reference in color channels or staining channels (Stacke et al., 2020). However, manual selection of the reference image requires specialized skills. In recent years, GAN has eliminated the need for the reference image, which has also been used to unify color variations (Shaban et al., 2019; Cho et al., 2017). Shaban et al. (2019) employed CycleGAN to map images into different color patterns but preserved the same tissue structure. Cho et al. (2017) applied a conditional GAN to learn certain stain-style distribution of the training data.

Several studies have been dedicated to learning domain-invariant representations, which can be categorized into two kinds of methods: feature alignment (Alirezazadeh et al., 2018) and domain adversarial learning (Lafarge et al., 2017; Zhang et al., 2019). To align features across different data sources, Alirezazadeh et al. (2018) learned a projection matrix to map feature vectors extracted from training and test sets to an invariant space. However, this feature mapping process relies on handcrafted features, which is difficult to achieve stable and robust performance. Domain adversarial network is a commonly used technique to learn domain-invariant features with an addition of a domain classifier (Lafarge et al., 2017; Zhang et al., 2019). Lafarge et al. (2017) developed a classification model with both mitosis and domain identification branches to ensure that the original domain cannot be recovered from the learned feature representation. Zhang et al. (2019) used adversarial learning to learn domain-invariant features, reducing the variations between domains. However, such adversarial training is computationally demanding and difficult to achieve stability (Chen et al., 2020).

3. Methods

We develop a Fourier-based mitosis detection algorithm (named FMDet) as shown in Fig. 2, which alleviates the domain shift problem to guarantee better generalization over unseen domains. Our FMDet algorithm can be divided into three parts: Fourier-based data augmentation (Yang and Soatto, 2020), pixel-level annotation generation, and segmentation-based mitosis detection. The proposed Fourier-based data augmentation aims to reduce the domain gaps arising from different

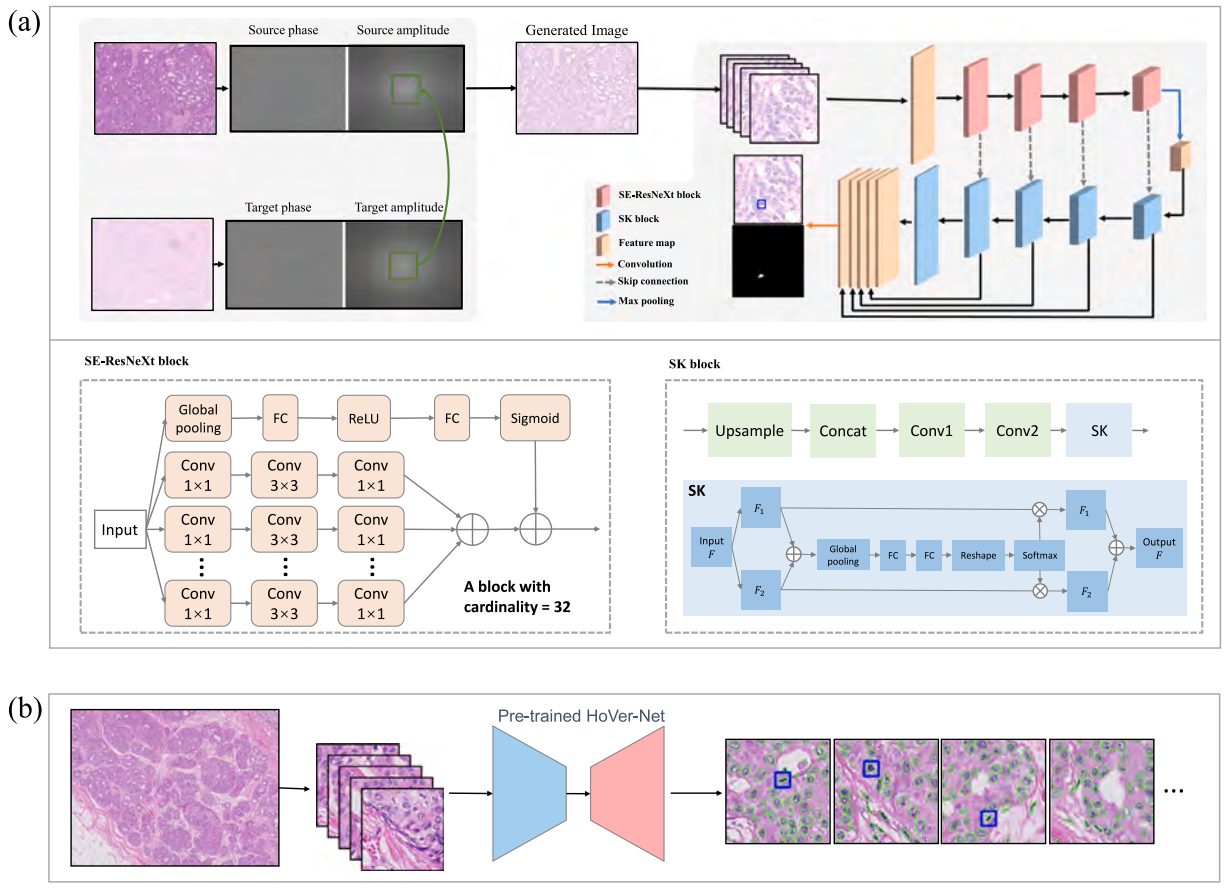

Fig. 2. Pipeline of our proposed mitosis detection algorithm. (a) An overview of our Fourier-based mitosis detection procedure, which contains three parts: Fourier-based data augmentation, segmentation mask generation, and segmentation-based mitosis detection. The Fourier-based data augmentation replaces the low-frequency spectrum of the image in the source domain with that in the reference domain. These generated images combine with original images to form the final diverse training samples. As shown in (b), the segmentation mask is obtained by modifying the bounding box labels into pixel-wise annotations. We directly infer a well-trained HoVer-Net network to generate precise nuclear boundaries. The interactions between the nuclear prediction results and bounding boxes are used as the final segmentation masks. In our proposed mitosis segmentation framework, a SE-ResNeXt50 encoder and an SK-based decoder are used to capture multi-scale features.

conditions (e.g., different centers, scanners, and staining processes), which is performed by replacing the low-frequency part of the image in the source domain using that in the reference domain. This procedure is conducted at the image level, which does not require any training process and can be implemented very fast. Afterward, these images are cropped into small patches to conduct the detailed mitosis detection process. The organizer of the MIDOG 2021 challenge provides only the bounding boxes of mitotic regions, which are not strictly labeled at the pixel level. To generate more accurate pixel-level annotations for mitotic nuclei, we first use a well-trained nuclear segmentation network (HoVer-Net) to obtain the masks of nuclei (Graham et al., 2019), and then the new mitosis masks are obtained by intersecting the nuclear masks and the bounding boxes provided by the organizer. Finally, the generated new mitosis masks are used to develop our segmentation network. In the following, we will present a detailed introduction to our FMDet algorithm.

3.1. Fourier-based data augmentation

To address the domain shift problem, we propose a Fourier-based mitosis detection method with promising domain generalization capability. The motivation behind this is that the variations in the low-frequency amplitude spectrum do not affect the high-level semantic perception (Yang and Soatto, 2020). Our Fourier-based data augmentation first performs a fast Fourier transformation (FFT) on all the training data provided in the MIDOG 2021 challenge. Then, images from two different domains are adopted as the source and reference images, respectively. We replace the low-frequency spectrum of the image in the source domain with that in the reference one, and then the new frequency spectrum in the source domain is transformed with an inverse FFT to generate the corresponding new image. The generated image keeps the same label as the original source image. This method transfers the style of the reference domain to the source domain without changing any image details or object boundaries. Fourier-based data augmentation can be regarded as an image-to-image transformation without any training process, which is formulated as follows.

Assuming that there are M domains for algorithm development, we randomly choose two domains as source and reference, respectively, to perform the Fourier-based data augmentation. For example, given images in a source domain $I^s = \{\mathbf{x}_i^s\}_{i=1}^{N^s}$ and images in an reference domain $I^r = \{\mathbf{x}_i^r\}_{i=1}^{N^r}$, where $\mathbf{x}_i \in \mathbb{R}^{H\times W\times 3}$ is an image, N^s and N^r represent the number of images in the source and reference domains, respectively. By applying Fourier-based data augmentation, the generated data can be represented as $I^g = \{\mathbf{x}_i^g\}_{i=1}^{N^g}$. The detailed computation process will be shown below.

For an image $\mathbf{x}$, its fast FFT can be represented by

$$\mathcal{F}(\mathbf{x}) = F(u,v) = \sum_{x=0}^{H-1}\sum_{y=0}^{W-1}\mathbf{x}(x,y)e^{-j2\pi(ux/H+vy/W)}, j^2 = -1, \tag{1}$$

here $\mathcal{F}^A = |F(u,v)|$ and $\mathcal{F}^P = \angle F(u,v)$ represent the amplitude and phase spectrum, respectively, which keep the same size as image $\mathbf{x}$.

To generate the new data $\mathbf{x}^g$, the Fourier-based data augmentation algorithm is implemented by replacing the low-frequency amplitude spectrum of the source images using that of the reference images. Then

the inverse FFT represented by $\mathcal{F}^{-1}$ is applied to recover the source images as new images $\mathbf{x}^g$, which are calculated as follows.

$$\mathbf{x}^g = \mathcal{F}^{-1}\left(\left[\left(M_\beta \cdot \mathcal{F}^A(\mathbf{x}^r) + \left(1 - M_\beta\right) \cdot \mathcal{F}^A(\mathbf{x}^s)\right), \mathcal{F}^P(\mathbf{x}^s)\right]\right), \tag{2}$$

where $M_\beta \in \mathbb{R}^{H \times W \times 3}$ represents a mask to obtain the low-frequency part, which has a value of 1 in the central region and 0 in the remaining regions. $\beta \in (0, 1)$ is utilized to control the range of the mask by the following settings.

$$M_\beta(h, w) = \begin{cases} 1, \text{ if } (h, w) \in (-\beta H : \beta H, -\beta W : \beta W) \\ 0, \text{ else }, \end{cases} \tag{3}$$

3.2. Segmentation mask generation

For the mitosis detection task, the labels provided by the MIDOG 2021 challenge are bounding boxes, as shown in Fig. 2(b). We transform these bounding box labels as pixel-level annotations for segmenting these nuclei with mitosis. We use a well-trained nuclear segmentation network called HoVer-Net to obtain the pixel-level boundaries of nuclei, which has demonstrated excellent performance on multiple independent multi-tissue haematoxylin & eosin stained histopathological image datasets (Graham et al., 2019). HoVer-Net adopts multi-task learning to simultaneously implement nuclear segmentation and classification, which facilitates a better feature extraction procedure. Also, to achieve more precise segmentation, it utilizes horizontal and vertical distance maps to separate overlapped nuclei. Due to its superior segmentation capability, we directly infer the trained HoVer-Net for nuclear segmentation, which eliminates the requirement of nuclear annotations. As shown in Fig. 2(b), the regions surrounded by green lines are the predicted nuclear areas by HoVer-Net. The blue bounding boxes are the mitotic cell annotations provided by the challenge organizers. The intersections of the green and blue regions are taken as the final annotations for our mitotic nuclear segmentation.

3.3. Basic network architecture

Our basic network structure adopts the encoder–decoder U-Net framework, which is shown in Fig. 2(a). In our segmentation network, two key modules with attention mechanism are used for better feature mining, i.e., the squeeze-and-excitation (SE) (Hu et al., 2018) and selective kernel (SK) (Li et al., 2019b) modules. The SE module squeezes the feature information into a global feature descriptor that is then integrated with the input descriptor to dynamically model the channel-wise feature dependencies, which allows the model to selectively and adaptively emphasize feature maps that contribute more to the prediction results, similar to human visual perception (Li, 2014). The SK module uses a channel-wise attention mechanism to dynamically fuse feature maps with different convolution kernel sizes, which enables the network to automatically aggregate appropriate feature scales for the prediction task. Overall, both attention modules help the network to recognize mitoses with large variations in cell size, color, texture, morphology, etc.

The encoder part uses a SE-ResNeXt50 model, which learns fused features from multiple subspaces by aggregating residual transformations. It includes four main layers and each layer is constructed by repeating a block with C paths (cardinalities) that perform $C = 32$ groups of convolutions. The number of repeated blocks in the four encoder layers is 3, 4, 6, and 3, respectively. Each block is constructed by combining a ResNeXt50 network and a SE attention module. For an input feature map $\mathbf{x}$, the attention feature can be calculated by

$$\mathbf{x}' = \mathbf{x} + f_{SE}\left(\sum_{i=1}^{C} \mathcal{T}_i(\mathbf{x})\right), \tag{4}$$

$$f_{SE} = f_{gp}(f_{fc}(f_{fc}(\cdot))), \tag{5}$$

where the input feature map $\mathbf{x}$ and the output one $\mathbf{x}'$ keep the same dimensions. $\mathcal{T}_i$ represents the transformation functions in each path, which keeps the same topology in each block. For example, in the first block, $\mathcal{T}_i$ denotes a sequential of 1×1 convolution, 3×3 convolution, and 1×1 convolution as shown in Fig. 2(a). f_{SE} represents the utilized SE module, which is implemented by a sequence of global pooling and dense connection operations to capture channel-wise attention features.

The decoder part consists of five upsampling blocks that recover these feature maps into the original image size. As shown in Fig. 2(a), in each upsampling block, TransConv is first performed to increase the image size of the feature maps, and then these feature maps are concatenated with those in the corresponding encoder layer. Next, two convolutional layers and an SK module are incorporated into each upsampling block. As shown in Fig. 2(a), the SK module aims to make the network adaptively adjust its respective field sizes by non-linearly aggregating multiple branches with different kernel sizes (Li et al., 2019b). The detailed computation process of the SK attention feature is shown as follows.

Given an input feature $\mathbf{x}'$, it is split into two new features ($\mathbf{x}'_k, k = 1, 2$) using 3×3 atrous convolution with different dilation values. Then, the two feature maps are summed and go through a global pooling and two fully connected layers to obtain an aggregated feature $\mathbf{z} \in \mathbb{R}^{2C}$. Then, $\mathbf{z}$ is reshaped as $\mathbf{z}' \in \mathbb{R}^{2 \times C}$ in terms of channel-wise to generate soft attention as follows.

$$m_k = e^{\mathbf{W}_k \mathbf{z}'} / \sum_{k=1}^{2} e^{\mathbf{W}_k \mathbf{z}'} \quad (k \in \{1, 2\}), \tag{6}$$

The final feature maps $\hat{\mathbf{x}}$ can be calculated by integrating the adaptive features from the two branches.

$$\hat{\mathbf{x}} = \mathbf{x}' + \sum_{k=1}^{2} m_k \times \mathbf{x}'_k, \tag{7}$$

Following the decoder, we fuse multi-scale features at five different resolution levels, which is performed by upsampling the features at each block of decoder (denoted as $\hat{\mathbf{x}}_0$, $\hat{\mathbf{x}}_1$, $\hat{\mathbf{x}}_2$, $\hat{\mathbf{x}}_3$, and $\hat{\mathbf{x}}_4$) to the same size as the input image, and then these features are concatenated together as $[\hat{\mathbf{x}}_0, \hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \hat{\mathbf{x}}_3, \hat{\mathbf{x}}_4]$. The obtained multi-scale feature maps pass through a convolution, batch normalization, and rectified linear unit to generate the final segmentation map $out = conv([\hat{\mathbf{x}}_0, \hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \hat{\mathbf{x}}_3, \hat{\mathbf{x}}_4])$.

3.4. Loss functions for segmentation

Our FMDet architecture is trained with a combination of focal loss (L_{FL}) and Dice loss (L_{DICE}). Focal loss has been successfully used to address data imbalance problems in various segmentation and detection tasks (Yang et al., 2021; Lin et al., 2017). In this work, we use focal loss to alleviate the imbalance between the number of mitotic and non-mitotic cells. Dice loss has been proposed to measure the degree of overlap between two images, which has been widely used as a metric in image segmentation tasks. For an individual image, our loss function L_{seg} is calculated as follows.

$$L_{seg} = L_{FL} + L_{DICE}, \tag{8}$$

$$L_{FL} = -\sum_i \left[\left(1 - \hat{y}_i\right)^\gamma y_i \log \hat{y}_i + \left(\hat{y}_i\right)^\gamma \left(1 - y_i\right) \log\left(1 - \hat{y}_i\right)\right], \tag{9}$$

$$L_{DICE} = -\frac{2 \sum_i y_i \hat{y}_i + \epsilon}{\sum_i y_i^2 + \sum_i \hat{y}_i^2}, \tag{10}$$

where y_i and $\hat{y}_i$ denote the ground truth and predicted label of the ith pixels, respectively. The $\gamma \geq 0$ in the focal loss is a tunable focus parameter. When γ is equal to 0, the focal loss degenerates to cross-entropy. The ϵ in Dice loss is set as a small positive value for training stability consideration.

Table 1
Details of datasets used for model development and cross-center validation.

Data split	Datasets	No. of cases	No. of images	Scanners	Institutions
Training set	MIDOG21 training	200	200 WSIs	Hamamatsu XR nanozoomer 2.0, Hamamatsu S360 (0.5 NA), Aperio ScanScope CS2, and Leica GT450	UMC Utrecht in Netherlands
Unseen test set	MIDOG21 test	80	80 WSIs	Hamamatsu XR nanozoomer 2.0, Hamamatsu RS nanozoomer 2.0, 3D Histech Panoramic 1000, and Leica GT450	UMC Utrecht in Netherlands
External test sets	AMIDA13	23	606 HPFs	Aperio ScanScope XT	UMC Utrecht in Netherlands
	MITOSIS14	11	2400 HFPs	Aperio Scanscope XT and Hamamatsu Nanozoomer 2.0-HT	Pathology Department at Pitié-Salpêtrière Hospital in Paris
	TUPAC-auxiliary	73	73 WSIs	Aperio Scanscope XT and Leica SCN400	(1). UMC Utrecht in Netherlands (2). Symbiant Pathology Expert Center in Alkmaar (SPEC-Alkmaar) (3). Symbiant Pathology Expert Center in Zaandam (SPEC-Zaandam)
	MIDOG22	50	50 WSIs	Aperio ScanScope CS2	FU Berlin in Germany

4. Experimental results and discussions

This section first describes the datasets utilized for model development and external evaluation. Then, the experimental setups and evaluation metrics are depicted in detail. Next, we conduct a comparison with top-performing methods in the MIDOG 2021 challenge and comparison with state-of-the-art mitosis detection methods. Then, ablation experiments are performed to verify the effectiveness of the key strategies we used. Next, we show the comparison with related methods on external datasets. Finally, a comparative experiment is conducted to compare our strategy for mitigating domain shifts (i.e., the FFT-based data augmentation) with other stain normalization methods.

4.1. Datasets

This work utilizes five datasets: MIDOG21, AMIDA13, MITOSIS14, TUPAC-auxiliary, and MIDOG22, which are released in public challenges. Our proposed mitosis detection method is developed on the training set in the MIDOG 2021 challenge and tested on the MIDOG21 test, AMIDA13, MITOSIS14, TUPAC-auxiliary, and MIDOG22 datasets, as shown in Table 1. The MIDOG21 dataset aims to develop algorithms for solving the domain discrepancies caused by different WSI scanners. We participated in the MIDOG 2021 challenge and secured first place. The remaining four datasets are adopted as independent multicenter test sets to further validate the generalization capability of our proposed algorithm. Note that although multiple datasets came from the UMC Utrecht, they were collected in different years. For example, samples in the AMIDA13 were collected before 2013 (Veta et al., 2015) and samples in the MIDOG21 were selected sequentially from 2017 and 2018 (Aubreville et al., 2021a). Therefore, there are no overlapping samples among these datasets.

MIDOG21[1]: The 2021 MIDOG challenge released large-scale data for mitosis detection, which includes 280 breast cancer WSIs acquired from UMC Utrecht in Netherlands. The challenge aims to develop robust algorithms to reduce the domain discrepancies caused by different WSI scanners. In the MIDOG 2021 challenge, tissues with an area of 2 mm^2 are scanned at a magnification of 40×, yielding WSIs with a size of around 8000 × 8000 pixels (depending on the specific scanners). These scanned WSIs are divided into a training set and a test set, with no overlapped cases in each set. The training dataset includes a total of 200 WSIs from four scanners (No. 1: Hamamatsu XR nanozoomer 2.0, No. 2: Hamamatsu S360 (0.5 NA), No. 3: Aperio ScanScope CS2, and No. 4: Leica GT450), where each scanner contains 50 WSIs. It is noted that the WSIs from the fourth scanner (Leica GT450) are not provided with mitosis annotations. The test set contains a total of 80 WSIs from four scanners (20 WSIs from each scanner), where two scanners (No. 1 and No. 4) are the same as the training set and the other two new scanners are Hamamatsu RS nanozoomer 2.0 and 3D Histech Panoramic 1000. Since the data in the test set and their annotations are not visible to the users, the evaluation is performed by submitting the trained model to the organizer.

AMIDA13: The AMIDA13 dataset was released in the *Assessment of Mitosis Detection Algorithms 2013* challenge (Veta et al., 2015), which contains 311 HPF images in the training set and 295 HPF images in the test set collected from a total of 23 subjects. These images have a size of 2000 × 2000 pixels (corresponding to an area of 0.25 mm^2). This dataset was later incorporated as part of the TUPAC-auxiliary. Different from the data split process of the challenge organizer, we adopt all these training and test images as an independent test set to validate the generalization capability of our model.

MITOSIS14[2]: The MITOSIS14 dataset was presented in the *MITOS-ATYPIA-14 Grand* challenge, which comprises a total of 2400 HPF images acquired from 11 breast cancer patients. Each image has been scanned by two scanners, resulting in 1200 HPFs with the image size of 1539 × 1376 pixels from scanner 1 (Aperio Scanscope XT) and 1200 HPFs with the image size of 1663 × 1485 pixels from scanner 2 (Hamamatsu Nanozoomer 2.0-HT) at 40× magnification. All 2400 images are employed as an external test set.

TUPAC-auxiliary[3]: TUPAC-auxiliary was released in the *2016 Tumor Proliferation Assessment Challenge* (TUPAC), which provides 73 well-annotated breast cancer WSIs from 73 cases for mitosis detection. These WSIs are acquired from three pathology centers and scanned with two different scanners. Patients 1–23 are from the AMIDA13 challenge and their WSIs are scanned by the Aperio ScanScope XT. Patients 24–48 and 49–73 are from two different pathology centers but are scanned with the same Leica SCN400. WSIs of patients 24–73 are scanned from tissues with an area of 2 mm^2 at a magnification of 40×, with a spatial resolution of 0.25 microns/pixel. In our experiment, all or part images in the TUPAC-auxiliary dataset are adopted as our external test set.

MIDOG22[4]: Following MIDOG 2021, the MIDOG 2022 challenge released new data for mitosis detection. We use 50 canine cases with mast cell tumor as another external test set, which were collected from FU Berlin and scanned by Aperio ScanScope CS2.

[1] https://midog2021.grand-challenge.org/.

[2] https://mitos-atypia-14.grand-challenge.org/Dataset/.

[3] https://tupac.grand-challenge.org/Dataset/.

[4] https://imig.science/midog/download-dataset/.

4.2. Experimental setups

The images are first cropped into patches with a size of 512 × 512 pixels. Our segmentation model is initialized by ImageNet weights and trained using the training data from the MIDOG 2021 challenge. The data distribution between mitotic and non-mitotic patches is extremely unbalanced. We ensure that the ratio of mitosis-positive to mitosis-negative samples within a mini-batch is 6:4 using a random sampling strategy. Real-time data augmentation is applied to increase the diversity of the training data, which includes random horizontal, vertical, and 90-degree flipping, random scaling, and random color distortion in HSV space (Tellez et al., 2019). The β for the Fourier-based data augmentation is set as 0.01 as recommended in the Ref. Yang and Soatto (2020).

Our model is optimized using the Adam optimizer (Kingma and Ba, 2014) with a weight decay of 0.0001. The learning rate is updated based on the cosine annealing scheme with an initial learning rate of 0.0003. The mini-batch size for model training is set to 24. Our experiments are carried out on the PyTorch framework and a workstation equipped with four 32 GB memory NVIDIA Tesla V100 GPU cards.

4.3. Evaluation metrics

In the current mitosis detection studies (e.g., in the MIDOG 2021 challenge and the 2016 TUPAC challenge), F1 score is a widely used evaluation criterion, which is a weighted average of precision and recall. Thus, this work adopts precision, recall, and F1 score as metrics, which are defined as follows.

$$\text{Precision} = \frac{N_{TP}}{N_{TP} + N_{FP}}, \tag{11}$$

$$\text{Recall} = \frac{N_{TP}}{N_{TP} + N_{FN}}, \tag{12}$$

$$\begin{aligned} \text{F1 score} &= 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \\ &= \frac{2 \times N_{TP}}{2 \times N_{TP} + N_{FN} + N_{FP}}. \end{aligned} \tag{13}$$

where N_{TP} represents the number of mitotic cells that are correctly classified as mitosis. N_{FP} and N_{FN} denote the number of normal images that are misclassified as mitosis and the number of mitosis images that are mislabeled as normal, respectively. We transform the mitosis detection problem as a semantic segmentation task. In the metric computation, if the Euclidian distance between the ground truth center and the corresponding center of the segmentation map is less than 7.5 μm, the detected object will be considered as a true positive (TP). All detection results outside 7.5 μm of the ground truth annotation are considered as false positive (FP) samples. All ground truth objects not detected within 7.5 μm are considered as false negative (FN) samples.

4.4. Comparison with top-performing methods in the MIDOG 2021 challenge

In the MIDOG 2021 challenge, the final leaderboard[5] is obtained by ranking the model performance evaluated in the unseen test set in terms of F1 score. Table 2 lists the detailed results of the top 13 teams in the challenge, where our solution ranked first place. These top solutions can be divided into three main lines: object detection, semantic segmentation, and patch-level image classification. To overcome domain variations, data augmentation, adversarial training, and stain normalization strategies are widely used among these teams. To ensure more robust performance, multi-model ensemble and two-stage methods are also popularly adopted schemes.

As shown in Table 2, it can be seen that detection and classification methods are the most widely used. Among them, eight teams (e.g., Team 3–9 and Team 11) adopt detection methods as the mainline. This may be due to the fact that the challenge organizer sets the task as an object detection one and provides the bounding box annotations for mitosis detection. The classification strategy is usually combined with the detection model to form a two-stage scheme, which acts as a binary classifier in the second stage to further verify the mitosis detection results (e.g., Team 2–3, Team 6, and Team 11). In addition, the remaining three classification schemes used by Team 5 and Team 12–13 are an end-to-end detection framework and two two-stage patch-level classification approaches, respectively. Team 5 uses the RetinaNet model with an addition of a domain classification loss for the domain adversarial training. For Team 12 and 13, their two-step methods first obtain the candidate mitosis figures by setting a threshold on the blue ratio images or employing a rotation-invariant convolutional network, and then identify mitotic figures by a binary classifier.

The segmentation scheme is adopted by four teams (Team 1–3 and Team 10), which can be categorized into a two-stage method (Team 2–3) and an end-to-end semantic segmentation method (Team 1 and Team 10). The two-stage strategy is similar to the one in the detection method, which is achieved by adding a classifier to further identify mitosis. Team 2 uses Efficient-UNet to obtain mitotic candidates and then applies Efficient-Net-B7 to classify these mitotic candidates into mitosis and non-mitosis. Team 3 uses mitosis segmentation as an auxiliary task to obtain mitosis-specific bounding boxes based on cascaded Mask-RCNN (with a mask prediction branch), and then trains an ensemble classifier (ResNet50 + DenseNet201) to further identify mitosis. For the end-to-end training methods (Team 1 and Team 10), U-Net is adopted as a segmentation model in both teams. The main difference lies in the generation of segmentation annotations. Team 10 directly transfers the full bounding box as a pixel-level annotation for segmentation, which introduces noisy annotations and therefore decreases the model performance. Our method (Team 1) generates more precise annotations by combining nuclear segmentation results with the bounding boxes, which makes a considerable performance improvement as discussed in our ablation study.

To address the domain shift problem, data augmentation, adversarial learning, and stain normalization are widely used by these teams. Among these, data augmentation is the most used (7 out of 13 teams), which can be categorized into regular and cross-domain-based augmentations. These regular augmentation methods are used by Team 9 and 13, including transposition, color shift, Gamma correction, additive Gaussian noise, etc., which do not consider the interaction between domains. Cross-domain-based augmentation methods either focus on color variations (e.g., stain appearance shift in Team 11 and statistical-analysis-based color transformation and a structure-preserved color normalization scheme in Team 8), or use deep-learning-based image style transfer (e.g., neural style transfer in Team 10, residual cycle-GAN in Team 3, starGAN in Team 4, and CycleGAN in Team 10). In addition to data augmentation, adversarial learning is the second most widely used method (6 out of 13 teams), and stain normalization is the last one that has only been used by three teams (Team 2, 6, and 7). The typical adversarial learning method aims to add a domain classifier to distinguish images from different domains (Team 5, 8, and 12), which forces the network to learn a domain-agnostic feature but has limited training stability under the adversarial loss. Different from these methods, our solution (Team 1) can be regarded as another type of traditional method that does not require a complex training process. We apply Fourier-based data augmentation to exchange the frequency information between domains, which is performed at the image level and therefore considers global information.

To further improve the model generalizability, some teams use multi-model ensemble and two-stage refinement strategies. For model ensemble, Team 3 trains two models (ResNet50 and DenseNet201) and

[5] https://midog2021.grand-challenge.org/evaluation/midog-final-test-phase/leaderboard/.

Table 2
Comparison of the top-13 teams in the MIDOG 2021 challenge. Note that each team in the "Rank" column has a hyperlink corresponding to the official leaderboard.

Rank	F1 score	Recall	Precision	Scheme							
				Det.	Seg.	Cla.	Aug.	AL	SN	Ensemble	Multi-stage
1	0.7476	0.7416	0.7537		✓		✓				
2	0.7474	0.7624	0.7330		✓	✓			✓		✓
3	0.7361	0.7095	0.7647	✓	✓	✓	✓	✓		✓	✓
4	0.7243	0.6790	0.7761	✓			✓	✓			
5	0.7108	0.7319	0.6909	✓		✓		✓			
6	0.7069	0.6870	0.7279	✓		✓			✓	✓	✓
7	0.7064	0.7047	0.7081	✓					✓		✓
8	0.7010	0.6437	0.7697	✓			✓	✓			✓
9	0.6963	0.6533	0.7454	✓			✓				
10	0.6859	0.6854	0.6865		✓		✓	✓			
11	0.6764	0.6677	0.6853	✓		✓	✓				✓
12	0.6708	0.7030	0.6413			✓		✓			✓
13	0.6319	0.6035	0.6631			✓	✓			✓	

Detection (Det.), segmentation (Seg.), classification (Cla.), data augmentation (Aug.), adversarial learning (AL), stain normalization (SN).

Table 3
Comparison with state-of-the-art methods for mitosis detection (evaluated on out-of-domain validation set from the MIDOG21 data).

Methods	Techniques	F1 score	Recall	Precision
RetinaNet (Aubreville et al., 2021b)	RetinaNet detector	0.6236	0.6687	0.5842
DeepMitosis (Li et al., 2018)	Two stages: Faster-RCNN + ResNet50 classifier	0.6440	0.6265	0.6624
MaskMitosis (Sebai et al., 2020)	Mask R-CNN detector	0.6601	0.6024	0.7299
SegMitos (Li et al., 2019a)	FCN segmentation + a threshold filtering mechanism	0.6791	0.6271	0.7405
DHE-Mit (Sohail et al., 2021a)	Two stages: Mask R-CNN + five classifier ensemble	0.6996	0.7200	0.6804
FMDet (ours)	End-to-end segmentation	**0.7773**	**0.8146**	**0.7440**

Table 4
Effect of different model architectures and annotation schemes on the mitosis detection accuracy. The results are computed on the out-of-domain validation data from MIDOG21.

Methods	F1 score	Recall	Precision
U-Net + Bbox mask	0.6521	0.7169	0.5980
U-Net + instance mask	0.6757	0.7519	0.6103
U-Net (SEResNeXt50) + Bbox mask	0.6822	0.7070	0.6592
U-Net (SEResNeXt50) + instance mask	0.7184	0.7801	0.6657
SK-Unet (SEResNeXt50 + SK) + Bbox mask	0.7003	0.7375	0.6667
SK-Unet (SEResNeXt50 + SK) + instance mask	0.7343	0.8096	0.6719

Team 6 trains five models (ASTMNet, ATTENNet, DSTMNet, ResidualNet, and RHINet) for mitosis classification, respectively. They adopt the average of these models as the final prediction of mitosis. The ensemble strategy provides considerable performance gains by balancing the errors between different models, but it is computationally intensive. In the two-stage scheme, as mentioned above, a classifier is usually combined with a mitosis candidate selection method (detection or segmentation model). This approach requires training different models, which is more complex compared to our end-to-end scheme.

4.5. Comparison with state-of-the-art methods for mitosis detection

To provide a fair comparison with state-of-the-art mitosis detection methods, we compare five algorithms that have released their codes to the public, including RetinaNet (Aubreville et al., 2021b), DeepMitosis (Li et al., 2018), MaskMitosis (Sebai et al., 2020), SegMitos (Li et al., 2019a), and DHE-Mit (Sohail et al., 2021a). Detailed results are shown in Table 3. The RetinaNet and MaskMitosis methods directly use object detection frameworks (e.g., RetinaNet and Mask R-CNN) to detect mitotic figures. The DeepMitosis and DHE-Mit methods adopt a two-stage technique by combining an object detector and a CNN classifier. SegMitos and our method adopt semantic segmentation to obtain mitotic regions, where SegMitos also adds another threshold filtering mechanism to further identify the target mitotic regions. Notably, since the test set of MIDOG21 is not visible, we divided a separate validation set from the training set by regarding images from each of the three scanners as an out-of-domain validation set.

As shown in Table 3, it is seen that our method achieves the best performance, outperforming DHE-Mit by around +8% in terms of F1 score. Comparing our end-to-end segmentation scheme with SegMitos, our method outperforms by around +10%. The main reason for such a large performance improvement is that all previous methods have not sufficiently considered the domain shifts between the training data and unseen test data. According to their released codes, only the MaskMitosis method applies a stain normalization strategy to guarantee a better generalization capability. By comparing the MaskMitosis with RetinaNet, it is seen that the MaskMitosis method has a higher performance of +2%. This can be attributed to the contribution of multi-task learning techniques in Mask R-CNN and the stain normalization for the domain shift problem. By comparing the two two-stage methods (DeepMitosis and DHE-Mit), it is seen that DHE-Mit has a higher F1 score of about +5%. The huge performance difference may be caused by the model ensemble technique used in DHE-Mit.

4.6. Ablation study

We conduct ablation experiments to verify the effectiveness of our used strategies, which include adopting SEResNeXt50 block as the encoder, adding SK attention module to the decoder, modifying semantic segmentation masks, and adding Fourier-based data augmentation for diverse image generation. This experiment is also performed on the out-of-domain validation set from the MIDOG21 data as mentioned above. Detailed results are shown in Tables 4 and 5.

A comparison of the bounding box (Bbox) masks and instance masks for mitosis detection are shown in Table 4. The Bbox masks are obtained by finding the inner tangent circles of the bounding box annotations provided by the organizers of the MIDOG 2021 challenge. In contrast to the coarse BBox masks, the instance masks correspond to detailed pixel-level annotations for mitotic nuclei, which are obtained by intersecting the ground truth bounding boxes and the nucleus segmentation results from a well-trained HoVer-Net (Graham et al., 2019). As shown in Table 4, the instance masks offer consistent performance gains over the Bbox masks, no matter which segmentation network is

Table 5
Effect of the Fourier-based data augmentation method on the mitosis detection accuracy. The results are computed on the out-of-domain validation set from MIDOG21. (aug. means data augmentation).

Methods	F1 score	Recall	Precision
U-Net + instance mask	0.6757	0.7519	0.6103
U-Net + instance mask + Fourier-based aug.	0.6982	0.7573	0.6477
U-Net (SEResNeXt50) + instance mask	0.7184	0.7801	0.6657
U-Net (SEResNeXt50) + instance mask + Fourier-based aug.	0.7488	0.7866	0.7145
SK-Unet (SEResNeXt50 + SK) + instance mask	0.7343	0.8096	0.6719
SK-Unet (SEResNeXt50 + SK) + instance mask + Fourier-based aug. (ours)	0.7773	0.8146	0.7440

Table 6
Comparison of the proposed method with prior works on external test sets (AMIDA13, MOTOSIS14, TUPAC-auxiliary, and MIDOG22 datasets). Center 1, 2, and 3 denote UMC Utrecht, SPEC-Alkmaar, and SPEC-Zaandam, respectively. MITOSIS14 and MIDOG22 data are collected from the Pathology Department at Pitié-Salpêtrière Hospital in Paris and FU Berlin, respectively.

Centers	Methods	F1 score	Recall	Precision
AMIDA13	SegMitos (Li et al., 2019a)	0.6728	0.6773	**0.6685**
(also Center 1 in	Deep consensus network (Wollmann and Rohr, 2017)	0.6090	0.6860	0.5470
TUPAC-auxiliary)	FMDet (ours)	**0.6791**	**0.7405**	0.6271
	SegMitos (Li et al., 2019a)	0.6690	0.7000	0.6400
TUPAC-auxiliary	Domain-adversarial neural networks (Lafarge et al., 2017)	0.6200	/	/
(Center 2 and 3)	Detection based on paired staining (Tellez et al., 2018)	0.4800	/	/
	Semi-supervised mitosis detection (Akram et al., 2018)	0.6400	0.6710	0.6130
	FMDet (ours)	**0.7458**	**0.8018**	**0.6971**
TUPAC-auxiliary	Deep consensus network (Wollmann and Rohr, 2021)	0.4700	/	/
(Center 1, 2, and 3)	Two-stage mitosis detection (Mahmood et al., 2020)	0.6420	0.6420	**0.6410**
	FMDet (ours)	**0.6946**	**0.7738**	0.6301
	MitosisDetection (Lei et al., 2020)	0.4000	/	/
MITOSIS14	DeepMitosis (Li et al., 2018)	0.4370	0.4430	0.4310
(Scanner 1 and 2)	MaskMitosis (Sebai et al., 2020)	0.4750	0.4530	**0.5000**
	FMDet (ours)	**0.4901**	**0.5564**	0.4380
	RetinaNet* (Aubreville et al., 2021b)	0.6145	0.7325	0.5292
	DeepMitosis* (Li et al., 2018)	0.6375	0.7656	0.5462
MIDOG22	SegMitos* (Li et al., 2019a)	0.6797	0.7603	0.6146
	DHE-Mit* (Sohail et al., 2021a)	0.6863	0.7524	0.6309
	FMDet (ours)	**0.7389**	**0.7796**	**0.7022**

used. In particular, for the SK-Unet, using the instance masks brings a performance gain of about 3% in the F1 score, 7% in recall, and 0.5% in precision. These results suggest that using the finer annotations provided by the instance masks can guide the network to focus more on the key parts of mitosis figures, which helps the network to better learn the features of mitotic nuclei and reduce false positives.

To improve the model generalization capability, we use unsupervised Fourier-based data augmentation to increase the diversity of the training samples by varying their frequency contents. The effect of the Fourier-based data augmentation method on the mitosis detection accuracy is summarized in Table 5. As seen from Table 5, Fourier-based data augmentation also brings consistent gains in all comparison setups. For example, used together with the full SK-Unet network and instance masks, the Fourier-based data augmentation yields a substantial increase of around 4% in the F1 score, 0.5% in recall, and 7% in precision. The major improvement in precision is due to the large reduction of false positives, which can be attributed to the abundant and diverse training samples generated by the Fourier-based augmentation method. These diverse and often hard samples increase the ability of the trained network to distinguish true mitoses from closely resembled false positives.

Results of comparing different network designs can also be seen in both Tables 4 and 5. Using the classical U-Net model (Ronneberger et al., 2015; Siddique et al., 2021) produces the lowest detection performance as shown at the top of both tables. Replacing the encoder of the original U-Net with the SEResNeXt50 module leads to a performance gain of at least 3% in terms of F1 score, as seen in the middle of both tables. To automatically aggregate features of different scales, we further added the SK attention module to the U-Net decoder, which improves the F1 score by another 2%, as shown in the comparison of the bottom and middle results in Tables 4 and 5.

Fig. 3(a)–(c) show some representative mitosis detection results on the out-of-domain validation data from MIDOG21. It can be seen that the yellow bounding boxes produced by our method match well with the ground truth green bounding boxes provided by the organizers. False positives do exist, but they mainly correspond to hard negatives (annotated with blue colors). These results are acceptable since the decision of these hard negatives is highly subjective even among experienced pathologists.

4.7. Comparison with related methods on external datasets

To further validate the generalization capability of our model, we directly infer the trained model on four external multicenter datasets, namely, AMIDA13, MITOSIS14, TUPAC-auxiliary, and MIDOG22, which are collected from a total of five different centers. It is noted that TUPAC-auxiliary covers the full range of AMIDA13 data. Table 6 compares our results obtained on the four datasets with that of several related methods. Visualization of mitosis detection results by our method is also shown in Fig. 3(d)–(h). The results of these compared methods are obtained from their corresponding publications, except for the methods marked with an asterisk, which are re-implemented by us based on their published codes.

As shown in Table 6, it is seen our method produces the highest F1 score among all methods on all test datasets. Even though SegMitos (Li et al., 2019a) uses a similar semantic segmentation scheme, the method was trained using weak annotations and its F1 score is 6%–7% lower than ours on the TUPAC-auxiliary and MIDOG22 data. Our proposed method utilizes a trained nucleus segmentation network to generate accurate masks for mitotic nuclei, which helps guide the model to learn important mitotic features and leads to much higher

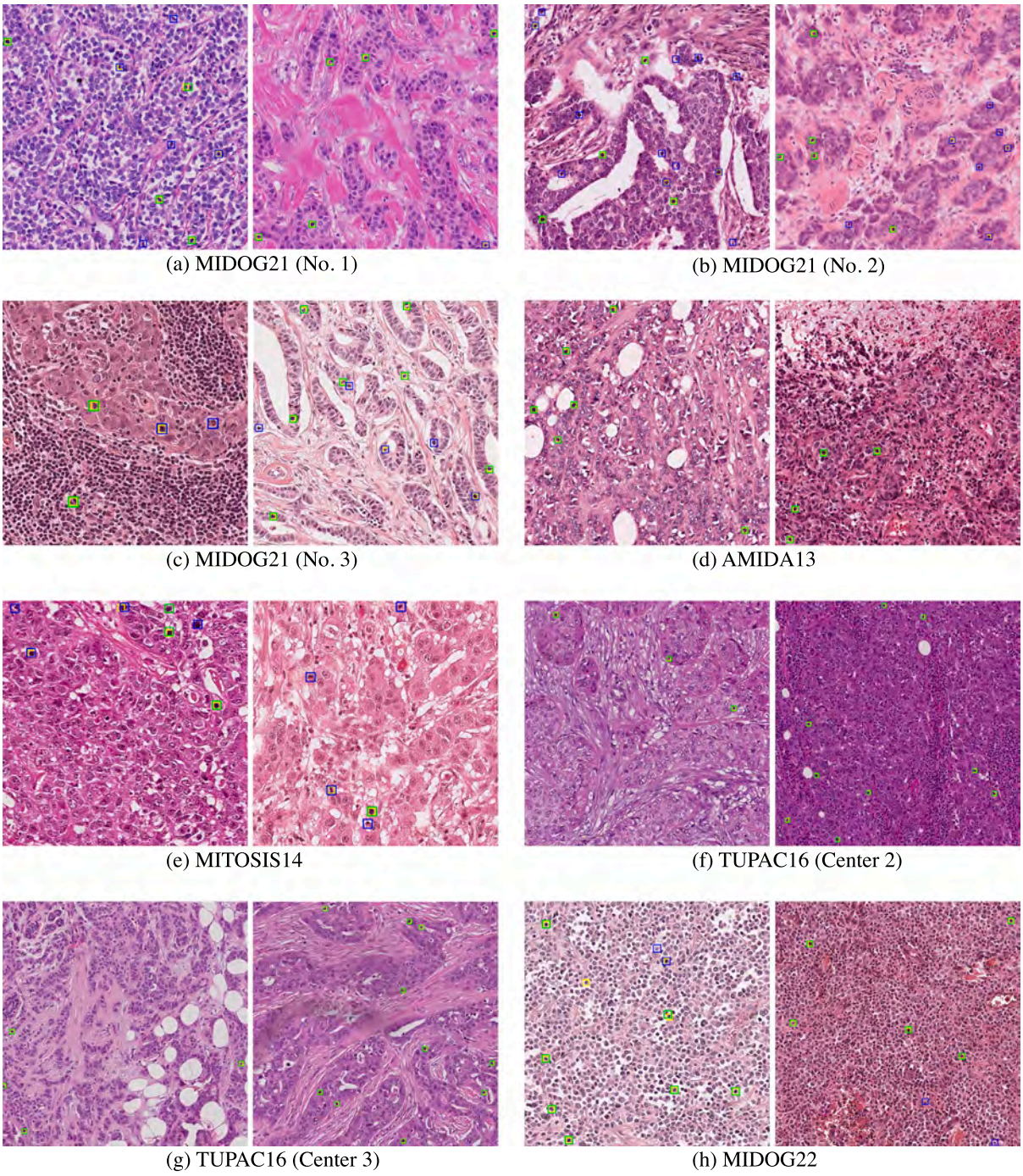

Fig. 3. Visualization of mitosis detection results from the proposed FMDet method. The predicted semantic segmentation results are converted into bounding boxes by drawing a rectangle of fixed size around the centroid of each segmentation mask. The green and blue bounding boxes are ground truth annotated by the corresponding data providers, which represent the true mitoses and hard negative samples (mitosis-like), respectively. The yellow bounding boxes are the predicted results by our method. The sub-figures correspond to sub-regions cropped from WSIs from different data sources for better visualization. Sub-figures (a)–(c) show examples from the MIDOG21, which are collected from No. 1: Hamamatsu XR nanozoomer 2.0, No. 2: Hamamatsu S360 (0.5 NA), and No. 3: Aperio ScanScope CS2 scanners, respectively. Sub-figures (d) and (e) show examples from the AMIDA 13 and MITOSIS14, respectively. Sub-figures (f) and (g) show examples from SPEC-Alkmaar (Center 2) and SPEC-Zaandam (Center 3) in the TUPAC16 dataset. Sub-figure (h) shows examples from MIDOG22, which are collected from a new center (FU Berlin).

accuracy. In the full TUPAC-auxiliary data, the two-stage mitosis detection method (Mahmood et al., 2020) is lower than ours by about 5% in terms of F1 score, which may be due to the fact that it does not consider the domain shift problem. Our method employs Fourier-based data augmentation to reduce domain discrepancies, which has shown robust performance in all experiments. Even though one study has reported higher performance on the TUPAC-auxiliary data (Sohail et al., 2021b), the same dataset is utilized during the model training. In other words, the reported results are not independently validated. In contrast, our results are computed on completely independent test data, which is more objective and can better indicate the generalization ability of the proposed method in real practice.

Although our method obtains the highest F1 and recall scores, the precision metric is not optimal. For example, the SegMitos (Li et al., 2019a) on AMIDA13, the two-stage mitosis detection method (Mahmood et al., 2020) on TUPAC-auxiliary, and the MaskMitosis (Sebai

Table 7
Comparison with other stain normalization methods on the out-of-domain validation set of MIDOG21. The baseline method is our mitosis segmentation method with the combination of SK-Unet (SEResNeXt50 + SK) and instance masks. Nor. and Aug. denote normalization and Fourier-based data augmentation, respectively.

Methods	F1 score	Recall	Precision
Baseline	0.7343	0.8096	0.6719
Baseline + SVD-based Nor.	0.7375	0.7946	0.6881
Baseline + SPCN	0.7390	0.7756	0.7056
Baseline + WSICS	0.7428	0.7938	0.6980
Baseline + Aug. (Ours)	0.7773	0.8146	0.7440

et al., 2020) on MITOSIS14 produce higher precision scores than our FMDet, which indicates that the results predicted by our method contain more false-positive images. The possible reasons can be summarized as follows. (1) Some non-mitotic cells (e.g., apoptotic or necrotic nuclei, dense nuclei, compressed nuclei, and "junk" particles) have a similar appearance to that of mitotic cells, which may be mistakenly detected as mitosis by our algorithm, as shown in Fig. 3(d)–(g). But it is able to improve the rate of missed mitosis detections, which is significantly important in clinical practice. (2) In some works (e.g., SegMitos and MaskMitosis methods (Li et al., 2019a; Sebai et al., 2020)), no domain discrepancy exists in their data, where the training and test images are from the same labs and scanners. Different from their settings, we adopt these datasets listed in Table 6 as unseen domains for independent algorithm validation, which can imitate the actual clinical scenario. (3) Some works employ complex computations to improve performance. For example, Mahmood et al. (2020) adopt a model ensemble to infer mitotic regions, which can be a valid strategy to reduce false predictions. SegMitos (Li et al., 2019a) requires complex post-processing to obtain accurate mitotic regions. In contrast, our end-to-end model does not require any post-processing step, which ensures a faster inference speed and more reliable predictions.

From Table 6, it is seen that the accuracy of our method on the MITOSIS14 data is the lowest. There are a couple of explanations. First, the MITOSIS14 dataset is known to be more challenging due to high variations in tissue appearance (Mahmood et al., 2020; Beevi et al., 2019; Sebai et al., 2020). Second, there are annotation inconsistencies across different datasets, since each dataset was labeled by different human observers and thus subject to inter-rater variations. We note that the reported results of other methods on MITOSIS14 were computed using only data from one scanner type, but our method was tested on all MITOSIS14 data. If tested using only data from the same scanner (1200 images in total), the F1 score will be 0.5218 instead of 0.4901, which is about 5% higher than the previously best-reported result.

4.8. Comparison with stain normalization methods

In this subsection, we conduct a set of experiments to compare our strategy for mitigating domain shift (i.e., FFT-based data augmentation) with existing stain normalization methods. A stain normalization method aims to remove stain variation so as to get consistent color distributions for both training and testing data, which is often adopted as a preprocessing step in histopathological image analysis. The most widely used stain normalization scheme is color deconvolution, where separating stains from the RGB images is a key step (Tosta et al., 2019). Three popular stain normalization methods are compared in Table 7. The SVD-based method of Macenko et al. (2009) uses singular value decomposition (SVD) to obtain the stain vectors. The structure-preserving color normalization (SPCN) method uses sparse non-negative matrix factorization to separate stains (Vahadane et al., 2016). The whole-slide image color standardizer (WSICS) method derives the stain components using the image color and spatial information (Bejnordi et al., 2015). From Table 7, it is seen that these stain normalization methods have little performance gains for the mitosis detection task, while our Fourier-based data augmentation method offers a much larger performance improvement. These results are consistent with that of a previous study (Tellez et al., 2019), where data augmentation was found to provide higher accuracy in a pathology classification task than using stain normalization. The reason is that it is difficult to accurately estimate the stain from a single image, and traditional stain normalization methods could create image artifacts. When used with deep learning methods, the reduced sample diversity due to stain normalization may increase the risk of model overfitting, whereas our data augmentation method increases data variation and promotes more robust network training.

5. Conclusion

In this work, we propose a mitosis detection method (FMDet) on breast histopathological images. To alleviate the problem of domain shift, we employ Fourier-based data augmentation to generate diverse domain-agnostic images. The importance of our data augmentation method has been verified in our ablation experiments. And then, we transform the detection task as a semantic segmentation one based on a fully convolutional network, where the course bounding box annotations are converted into fine pixel-level boundaries of mitotic nuclei. The instance-based semantic segmentation with precise supervision produces higher performance compared with the bounding-box-based segmentation. To achieve more robust feature extraction, we integrate a channel-wise multi-scale attention mechanism into the U-Net framework, which has also been verified to achieve superior performance in our ablation study. Our FMDet algorithm has been evaluated in the MIDOG 2021 challenge and secured first place. The comparison experiments with state-of-the-art mitosis detection methods also demonstrate the effectiveness of our solution. Further, we externally validate our generalization capability on four independent datasets for mitosis detection. The corresponding experimental results provide evidence for model generalization and reproducibility across multi-center cohorts. In our future work, we will attempt to improve the model generalizability by combining self-supervised learning on large-scale unlabeled data.

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.

Data availability

I have shared the link to my data/code at the manuscript.

Acknowledgments

This research was in part funded by the National Natural Science Foundation of China (No. 61571314), Science & technology department of Sichuan Province, China (No. 2020YFG0081), and the Innovative Youth Projects of Ocean Remote Sensing Engineering Technology Research Center of State Oceanic Administration of China (No. 2015001).

References

Akram, S.U., Qaiser, T., Graham, S., Kannala, J., Heikkilä, J., Rajpoot, N., 2018. Leveraging unlabeled whole-slide-images for mitosis detection. In: OMIA COMPAY. Springer, pp. 69–77.

Alirezazadeh, P., Hejrati, B., Monsef-Esfahani, A., Fathi, A., 2018. Representation learning-based unsupervised domain adaptation for classification of breast cancer histopathology images. Biocybern. Biomed. Eng. 38 (3), 671–683.

Aubreville, M., Bertram, C.A., Donovan, T.A., Marzahl, C., Maier, A., Klopfleisch, R., 2020. A completely annotated whole slide image dataset of canine breast cancer to aid human breast cancer research. Sci. Data 7 (1), 1–10.

Aubreville, M., Bertram, C., Veta, M., Klopfleisch, R., Stathonikos, N., Breininger, K., ter Hoeve, N., Ciompi, F., Maier, A., 2021a. Mitosis domain generalization challenge: Structured description of the challenge design (MIDOG 2021). http://dx.doi.org/10.5281/zenodo.4573978.

Aubreville, M., Bertram, C., Veta, M., Klopfleisch, R., Stathonikos, N., Breininger, K., ter Hoeve, N., Ciompi, F., Maier, A., 2021b. Quantifying the scanner-induced domain gap in mitosis detection. arXiv preprint arXiv:2103.16515.

Aubreville, M., Stathonikos, N., Bertram, C.A., Klopleisch, R., ter Hoeve, N., Ciompi, F., Wilm, F., Marzahl, C., Donovan, T.A., Maier, A., et al., 2022. Mitosis domain generalization in histopathology images–the MIDOG challenge. arXiv preprint arXiv:2204.03742.

Balkenhol, M.C., Tellez, D., Vreuls, W., Clahsen, P.C., Pinckaers, H., Ciompi, F., Bult, P., van der Laak, J.A., 2019. Deep learning assisted mitotic counting for breast cancer. Lab. Invest. 99 (11), 1596–1606.

Beevi, K.S., Nair, M.S., Bindu, G., 2019. Automatic mitosis detection in breast histopathology images using Convolutional Neural Network based deep transfer learning. Biocybern. Biomed. Eng. 39 (1), 214–223.

Bejnordi, B.E., Litjens, G., Timofeeva, N., Otte-Höller, I., Homeyer, A., Karssemeijer, N., Van Der Laak, J.A., 2015. Stain specific standardization of whole-slide histopathological images. IEEE Trans. Med. Imaging 35 (2), 404–415.

Challen, R., Denny, J., Pitt, M., Gompels, L., Edwards, T., Tsaneva-Atanasova, K., 2019. Artificial intelligence, bias and clinical safety. BMJ Qual. Saf. 28 (3), 231–237.

Chang, J.-R., Wu, M.-S., Yu, W.-H., Chen, C.-C., Yang, C.-K., Lin, Y.-Y., Yeh, C.-Y., 2021. Stain mix-up: Unsupervised domain generalization for histopathology images. In: MICCAI. Springer, pp. 117–126.

Chen, C.-L., Chen, C.-C., Yu, W.-H., Chen, S.-H., Chang, Y.-C., Hsu, T.-I., Hsiao, M., Yeh, C.-Y., Chen, C.-Y., 2021. An annotation-free whole-slide training approach to pathological classification of lung cancer types using deep learning. Nature Commun. 12 (1), 1–13.

Chen, M., Zhao, S., Liu, H., Cai, D., 2020. Adversarial-learned loss for domain adaptation. In: AAAI, Vol. 34. pp. 3521–3528.

Cho, H., Lim, S., Choi, G., Min, H., 2017. Neural stain-style transfer learning using GAN for histopathological images. arXiv preprint arXiv:1710.08543.

Elston, C.W., Ellis, I.O., 1991. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology 19 (5), 403–410.

Faryna, K., van der Laak, J., Litjens, G., 2021. Tailoring automated data augmentation to H&E-stained histopathology. In: Medical Imaging with Deep Learning.

Graham, S., Vu, Q.D., Raza, S.E.A., Azam, A., Tsang, Y.W., Kwak, J.T., Rajpoot, N., 2019. Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. Med. Image Anal. 58, 1–15.

de Haan, K., Zhang, Y., Zuckerman, J.E., Liu, T., Sisk, A.E., Diaz, M.F., Jen, K.-Y., Nobori, A., Liou, S., Zhang, S., et al., 2021. Deep learning-based transformation of H&E stained tissues into special stains. Nature Commun. 12 (1), 1–13.

Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-excitation networks. In: CVPR. pp. 7132–7141.

Huang, Y., Zheng, H., Liu, C., Ding, X., Rohde, G.K., 2017. Epithelium-stroma classification via convolutional neural networks and unsupervised domain adaptation in histopathological images. IEEE J. Biomed. Health Inf. 21 (6), 1625–1632.

Hwang, M., Wang, D., Wu, C., Jiang, W.-C., Kong, X.-X., Hwang, K.-S., Ding, K., 2020. A fuzzy segmentation method to learn classification of mitosis. Int. J. Fuzzy Syst. 22, 1653–1664.

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

Lafarge, M.W., Pluim, J.P., Eppenhof, K.A., Moeskops, P., Veta, M., 2017. Domain-adversarial neural networks to address the appearance variability of histopathology images. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Dcision Support. Springer, pp. 83–91.

Lei, H., Liu, S., Elazab, A., Gong, X., Lei, B., 2020. Attention-guided multi-branch convolutional neural network for mitosis detection from histopathological images. IEEE J. Biomed. Health Inf. 25 (2), 358–370.

Leo, P., Lee, G., Shih, N.N., Elliott, R., Feldman, M.D., Madabhushi, A., 2016. Evaluating stability of histomorphometric features across scanner and staining variations: prostate cancer diagnosis from whole slide images. J. Med. Imaging 3 (4), 138–150.

Li, Z., 2014. Understanding Vision: Theory, Models, and Data. Oxford University Press, USA.

Li, X., Wang, W., Hu, X., Yang, J., 2019b. Selective kernel networks. In: CVPR. pp. 510–519.

Li, C., Wang, X., Liu, W., Latecki, L.J., 2018. DeepMitosis: Mitosis detection via deep detection, verification and segmentation networks. Med. Image Anal. 45, 121–133.

Li, C., Wang, X., Liu, W., Latecki, L.J., Wang, B., Huang, J., 2019a. Weakly supervised mitosis detection in breast histopathology images using concentric loss. Med. Image Anal. 53, 165–178.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2017. Focal loss for dense object detection. In: CVPR. pp. 2980–2988.

Macenko, M., Niethammer, M., Marron, J.S., Borland, D., Woosley, J.T., Guan, X., Schmitt, C., Thomas, N.E., 2009. A method for normalizing histology slides for quantitative analysis. In: ISBI. IEEE, pp. 1107–1110.

Madabhushi, A., Lee, G., 2016. Image analysis and machine learning in digital pathology: Challenges and opportunities. Med. Image Anal. 33, 170–175.

Mahmood, T., Arsalan, M., Owais, M., Lee, M.B., Park, K.R., 2020. Artificial intelligence-based mitosis detection in breast cancer histopathology images using faster R-CNN and deep CNNs. J. Clin. Med. 9 (3), 749–769.

Mahmood, F., Borders, D., Chen, R.J., McKay, G.N., Salimian, K.J., Baras, A., Durr, N.J., 2019. Deep adversarial training for multi-organ nuclei segmentation in histopathology images. IEEE Trans. Med. Imaging 39 (11), 3257–3267.

Maroof, N., Khan, A., Qureshi, S.A., ul Rehman, A., Khalil, R.K., Shim, S.-O., 2020. Mitosis detection in breast cancer histopathology images using hybrid feature space. Photodiagn. Photodyn. Ther. 31, 101885.

MITOS-ATYPIA-14, 2014. Mitos-atypia-14-dataset. https://mitos-atypia-14.grand-challenge.org/dataset/.

Nateghi, R., Danyali, H., Helfroush, M.S., 2021. A deep learning approach for mitosis detection: Application in tumor proliferation prediction from whole slide images. Artif. Intell. Med. 114, 102048.

Pan, X., Lu, Y., Lan, R., Liu, Z., Qin, Z., Wang, H., Liu, Z., 2021. Mitosis detection techniques in H&E stained breast cancer pathological images: A comprehensive review. Comput. Electr. Eng. 91, 107038.

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

Saha, M., Chakraborty, C., Racoceanu, D., 2018. Efficient deep learning model for mitosis detection using breast histopathology images. Comput. Med. Imaging Graph. 64, 29–40.

Schömig-Markiefka, B., Pryalukhin, A., Hulla, W., Bychkov, A., Fukuoka, J., Madabhushi, A., Achter, V., Nieroda, L., Büttner, R., Quaas, A., et al., 2021. Quality control stress test for deep learning-based diagnostic model in digital pathology. Mod. Pathol. 34 (12), 2098–2108.

Sebai, M., Wang, X., Wang, T., 2020. MaskMitosis: a deep learning framework for fully supervised, weakly supervised, and unsupervised mitosis detection in histopathology images. Med. Biol. Eng. Comput. 58, 1603–1623.

Shaban, M.T., Baur, C., Navab, N., Albarqouni, S., 2019. StainGAN: Stain style transfer for digital histological images. In: ISBI. IEEE, pp. 953–956.

Siddique, N., Paheding, S., Elkin, C.P., Devabhaktuni, V., 2021. U-net and its variants for medical image segmentation: A review of theory and applications. IEEE Access 9, 82031–82057.

Siegel, R.L., Miller, K.D., Jemal, A., 2020. Cancer statistics, 2020. CA: Cancer J. Clin. 70 (1), 7–30.

Sigirci, I.O., Albayrak, A., Bilgin, G., 2021. Detection of mitotic cells in breast cancer histopathological images using deep versus handcrafted features. Multimedia Tools Appl. 1–24.

Sohail, A., Khan, A., Nisar, H., Tabassum, S., Zameer, A., 2021a. Mitotic nuclei analysis in breast cancer histopathology images using deep ensemble classifier. Med. Image Anal. 72, 102121.

Sohail, A., Khan, A., Wahab, N., Zameer, A., Khan, S., 2021b. A multi-phase deep CNN based mitosis detection framework for breast cancer histopathological images. Sci. Rep. 11 (1), 1–18.

Stacke, K., Eilertsen, G., Unger, J., Lundström, C., 2020. Measuring domain shift for deep learning in histopathology. IEEE J. Biomed. Health Inf. 25 (2), 325–336.

Tellez, D., Balkenhol, M., Otte-Höller, I., van de Loo, R., Vogels, R., Bult, P., Wauters, C., Vreuls, W., Mol, S., Karssemeijer, N., et al., 2018. Whole-slide mitosis detection in H&E breast histology using PHH3 as a reference to train distilled stain-invariant convolutional networks. IEEE Trans. Med. Imaging 37 (9), 2126–2136.

Tellez, D., Litjens, G., Bándi, P., Bulten, W., Bokhorst, J.-M., Ciompi, F., Van Der Laak, J., 2019. Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. Med. Image Anal. 58, 101544.

Tosta, T.A.A., de Faria, P.R., Neves, L.A., do Nascimento, M.Z., 2019. Computational normalization of H&E-stained histological images: Progress, challenges and future potential. Artif. Intell. Med. 95, 118–132.

Vahadane, A., Peng, T., Sethi, A., Albarqouni, S., Wang, L., Baust, M., Steiger, K., Schlitter, A.M., Esposito, I., Navab, N., 2016. Structure-preserving color normalization and sparse stain separation for histological images. IEEE Trans. Med. Imaging 35 (8), 1962–1971.

Veta, M., Heng, Y.J., Stathonikos, N., Bejnordi, B.E., Beca, F., Wollmann, T., Rohr, K., Shah, M.A., Wang, D., Rousson, M., et al., 2019. Predicting breast tumor proliferation from whole-slide images: the TUPAC16 challenge. Med. Image Anal. 54, 111–121.

Veta, M., Van Diest, P.J., Willems, S.M., Wang, H., Madabhushi, A., Cruz-Roa, A., Gonzalez, F., Larsen, A.B., Vestergaard, J.S., Dahl, A.B., et al., 2015. Assessment of algorithms for mitosis detection in breast cancer histopathology images. Med. Image Anal. 20 (1), 237–248.

Wilkins, A.S., Holliday, R., 2009. The evolution of meiosis from mitosis. Genetics 181 (1), 3–12.

Wollmann, T., Rohr, K., 2017. Deep residual hough voting for mitotic cell detection in histopathology images. In: ISBI. IEEE, pp. 341–344.

Wollmann, T., Rohr, K., 2021. Deep Consensus Network: Aggregating predictions to improve object detection in microscopy images. Med. Image Anal. 70, 102019.

Xue, Y., Ye, J., Zhou, Q., Long, L.R., Antani, S., Xue, Z., Cornwell, C., Zaino, R., Cheng, K.C., Huang, X., 2021. Selective synthetic augmentation with HistoGAN for improved histopathology image classification. Med. Image Anal. 67, 101816.

Yamashita, R., Long, J., Banda, S., Shen, J., Rubin, D.L., 2021. Learning domain-agnostic visual representation for computational pathology using medically-irrelevant style transfer augmentation. arXiv preprint arXiv:2102.01678.

Yang, D., Myronenko, A., Wang, X., Xu, Z., Roth, H.R., Xu, D., 2021. T-AutoML: Automated machine learning for lesion segmentation using transformers in 3D medical imaging. In: CVPR. pp. 3962–3974.

Yang, Y., Soatto, S., 2020. FDA: Fourier domain adaptation for semantic segmentation. In: CVPR. pp. 4085–4095.

Yousif, M., van Diest, P.J., Laurinavicius, A., Rimm, D., van der Laak, J., Madabhushi, A., Schnitt, S., Pantanowitz, L., 2021. Artificial intelligence applied to breast pathology. Virchows Arch. 1–19.

Zhang, Y., Chen, H., Wei, Y., Zhao, P., Cao, J., Fan, X., Lou, X., Liu, H., Hou, J., Han, X., et al., 2019. From whole slide imaging to microscopy: Deep microscopy adaptation network for histopathology cancer image classification. In: MICCAI. Springer, pp. 360–368.