HHS Public Access

Author manuscript

Proc COMPSAC. Author manuscript; available in PMC 2020 June 18.

Published in final edited form as:

Proc COMPSAC. 2019 July ; 2019: 696–703. doi:10.1109/compsac.2019.00105.

Improving Classification of Breast Cancer by Utilizing the Image Pyramids of Whole-Slide Imaging and Multi-Scale Convolutional Neural Networks

Li Tong[1], **Ying Sha**[2], **May D. Wang**[1,*]

[1]Dept. of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, GA 30332

[2]School of Biology, Georgia Institute of Technology, Atlanta, GA 30332

Abstract

Whole-slide imaging (WSI) is the digitization of conventional glass slides. Automatic computer-aided diagnosis (CAD) based on WSI enables digital pathology and the integration of pathology with other data like genomic biomarkers. Numerous computational algorithms have been developed for WSI, with most of them taking the image patches cropped from the highest resolution as the input. However, these models exploit only the local information within each patch and lost the connections between the neighboring patches, which may contain important context information. In this paper, we propose a novel multi-scale convolutional network (ConvNet) to utilize the built-in image pyramids of WSI. For the concentric image patches cropped at the same location of different resolution levels, we hypothesize the extra input images from lower magnifications will provide context information to enhance the prediction of patch images. We build corresponding ConvNets for feature representation and then combine the extracted features by 1) late fusion: concatenation or averaging the feature vectors before performing classification, 2) early fusion: merge the ConvNet feature maps. We have applied the multi-scale networks to a benchmark breast cancer WSI dataset. Extensive experiments have demonstrated that our multiscale networks utilizing the WSI image pyramids can achieve higher accuracy for the classification of breast cancer. The late fusion method by taking the average of feature vectors reaches the highest accuracy (81.50%), which is promising for the application of multi-scale analysis of WSI.

Keywords

Whole-Slide Imaging; Breast Cancer; Image Pyramid; Multi-Scale Convolutional Neural Network

I. INTRODUCTION

Whole-slide imaging (WSI) generates digital slides by scanning conventional glass slides, which enables the computer-aided diagnosis (CAD) for pathological images and further the integration of digital pathology with high-dimensional genomics features [1]. The virtual

*Corresponding author contact: maywang@bme.gatech.edu.

slide with resolution almost at the optical resolution limits can be generated in sixty seconds for an entire glass slide [1]. The resulting virtual slides consist of digital images of histological sections over a complete range of standard magnifications. Extensive validations by the College of American Pathologists Pathology have demonstrated the effectiveness of WSI for diagnostic interpretation [2]. Manual examination of slides and diagnosis by pathologists can be subjective and time-consuming because of the large fields-of-view to be reviewed. With the adoption of WSI systems in clinical settings, researchers have built numerous CAD systems for various clinical endpoints. The development of WSI and the corresponding computational algorithms forms two major components of digital pathology, which has paved the way towards precision medicine.

Due to the large size of image data contained in a virtual slide, it's impractical to process the entire digital image of a slide as a whole. Most CAD systems first select the region of interest (ROI) using low-resolution thumbnails and then tile the image within ROIs into small image patches at the highest magnification level. After the tiling process, the image patches are processed with feature extraction, feature selection, and predictive modeling for conventional digital image processing pipelines. With the huge success of deep learning in natural images, models like deep convolutional networks (ConvNets) have also been applied to these image patches for efficient feature representation. The features generated by ConvNets are then fed into extra networks for segmentation and classification. Recent works have demonstrated promising results for applying deep learning to WSI [3], [4].

However, current CAD models typically process the image patches independently without considering the geographical relationships between each patch. This paradigm results in two major drawbacks. First, the image patches are processed without the support from neighboring image patches, which may contain useful context information. For example, specific structures may be partially cropped in a specific image patch and reduce the power of feature representation. Second, only the finest images at the highest magnification are used for feature extraction while the coarser images at lower magnifications are discarded completely. These coarser images also carry useful information and should have been utilized for the feature representation. To address these drawbacks of current CAD models for pathological images, we propose to exploit the context information of each image patch with multi-scale ConvNets. Besides the basic image patch tiled from the finest resolution, we also extract image patches from two lower resolutions at the same geographic center to provide context information (Fig. 1). These two extra image patches are captured from coarser levels of the WSI image pyramid and thus have larger Field of Views (FOVs) compared with the original image patches. We build three ConvNets for three concentrical image patches respectively and then combine the extracted features for the final classification of the basic image patch. We have applied our multi-scale ConvNets to a benchmark breast cancer dataset. Extensive experiments have demonstrated that our multi-scale ConvNets can significantly improve the classification performance of breast cancer.

The rest of this paper is organized as follows. We first present the related works in Section II. We then present the material and methods in detail in Section III. In section IV, we will present the results on benchmark breast cancer dataset with ablations. Finally, we will draw a conclusion and discuss the further directions in Section V.

II. RELATED WORK

A. Computer-aided diagnosis for WSI

The development of CAD systems for digital pathology has become one of the fastest growing fields of research due to the growing popularity of WSI. A typical CAD pipeline for WSI consists of four major components: 1) image quality control, 2) feature extraction at the pixel, object and semantic levels, 3) predictive modeling using imaging features, and 4) model visualization for interactive discovery [5].

Image quality control is an essential step for digital pathology because of the heterogeneities of WSIs collected between different clinical sites with various platforms and slide preparation protocols, which is the so-called batch effect. On the other hand, the WSI may also have artifacts including tissue folds, blurred regions, pen marks, and shadows. The batch effects and image artifacts have unpredictable effects on image segmentation, classification, and other quantitative image analysis tasks. Researchers have developed multiple techniques including color normalization, scale normalization, and blur detection to eliminate or correct the batch-effect and image artifacts of WSIs.

Feature extraction is another essential step to represent the WSI data quantitatively. Conventional digital imaging processing techniques extract features from pathological images at pixel and object levels to capture the morphological properties [6]. Pixel-level feature extraction identifies the properties of color and texture for all image pixels. Color features are typically expressed with the color spread, prominence, and cooccurrence using statistics and frequencies of color histograms in different color spaces. Texture features quantify image sharpness, contrast, changes in intensity, and discontinuities or edges by measuring properties from gray-level intensity profiles. Object-level feature extraction requires the segmentation of cellular structures and captures the shape, texture, and spatial distribution of cellular structures in a WSI. Besides the features extracted from WSI, researchers also proposed to integrate the pathological features with clinical features and genomic features for improved diagnosis [7].

However, the conventional feature extraction relies heavily on hand-crafted features, which limits the generalizability of the features. With the development of deep learning, the human-designed feature extraction has been replaced by feature representation with deep neural networks. For imaging data, the most popular feature representation network is the convolutional neural network (ConvNets). Deep ConvNets can learn efficient feature representation from a large amount of training data. By combining ConvNets with fully connected (FC) layers and a softmax layer for classification, the deep networks can be trained end-to-end and thus can learn both feature representation and classification from the training data. With the success of deep learning in natural images, deep neural networks like ConvNets have been applied to medical images including MRI (brain tumor [8]), CT (lung nodule [9]), endomicroscopy (Barrett’s esophagus [10]), and WSI (breast cancer [3], lung cancer [11], aglioma [4], heart rejection [12], etc.).

B. Image pyramids and multi-scale models for natural images

Our multi-scale ConvNets for WSI are inspired by the image pyramids in object recognition. Image pyramids consist of multi-scale representations of the same image. Feature pyramids built upon image pyramids are one of the most standard solutions for object recognition at various scales in computer vision. The object's scale change is offset by its level within the pyramid so that the objects can be recognized in a scale-invariant fashion. By scanning a trained model over both positions and pyramid levels, objects across a large range of scales can be robustly detected. Before the era of deep learning, dense sampling on image pyramids along with handcrafted features is critical for accurate object detection. With the success of deep learning for natural images, the handcrafted feature extraction step has been replaced by the powerful feature representation using deep ConvNets. Although ConvNets are much more robust to scale variances compared to hand-crafted features, image pyramids are utilized to ensure the most accurate performance. For example, multiple top entries in the ImageNet [13] and COCO [14] detection challenges exploit multi-scale inference with image pyramids.

Besides the direct application of image pyramids for multiscale inference, recent works have also utilized the built-in feature hierarchy of deep ConvNets for multi-scale feature representations. For example, feature pyramid network (FPN) makes use of the pyramid shape of a ConvNet's feature hierarchy and builds semantically strong multi-scale representations with the bottom-up pathway, top-down pathway, and lateral connections [15]. The FPN gets rid of image pyramids and creates in-network feature pyramids, which significantly reduces the speed and memory without sacrificing the multiscale representation power. Multiple works including Mask RCNN [16] has utilized the FPN framework to achieve improved performance.

One major difference between our work and the multiscale object detection is that we aim to make use of the built-in image pyramids of WSI and improve the prediction performance with the support of context information. Thus, we improve the current model by enlarging the FOVs using images of coarser levels on the WSI image pyramids.

C. Multi-scale features for medical images

Xu et al. have applied multi-scale context-aware networks for colorectal liver metastases [17]. They utilized the context information from low magnification levels by concatenating the feature maps generated by deep convolutional neural networks at early and late stages respectively. The combined features improved the performance for image segmentation and classification tasks.

III. MULTI-SCALE CONVOLUTIONAL NETWORKS

A. Late Fusion

One intuitive way to integrate images from different scale is to combine the hidden feature vectors extracted from the three concentric images. Since the integration happens before the last FC layer, we call this family of methods as late fusion (Fig. 2). We first use the five layers of ConvNets and two FC layers to extract features from input images respectively.

After the five layers of convolution, each image is represented with a feature map with size 256 × 6 × 6. We then flatten the feature maps and get a vector with length 4, 096 for each image after the two FC layers. Finally, we combine the feature vectors before feeding the last classification layer by either concatenation or taking average. The combined feature vectors are then go through the another FC layer for classification.

B. Early Fusion

Another way to integrate these multi-scale images is to combine the feature maps at a relatively early stage, which is within the five layers of ConvNets. We call these methods as early fusion. Based on the methods we use to fuse the feature maps, we can further classify them into full concatenation and partial fusion.

Full concatenation is to directly concatenate the feature maps of different concentric images (Fig. 3). Since the images are scaled to the same dimension (3 × 224 × 224) and parsed through ConvNets of the same structure, the feature maps are also of the same dimension. We can directly concatenate these feature maps along the depth dimension and then feed the combined feature maps to the rest of the networks. Since the concatenation will increase the depth by three times, the following ConvnNet is modified correspondingly. To simplify the model, all network parameters are shared for three levels of images.

Partial fusion aims to take the various FOVs into consideration when integrate the multi-scale feature maps (Fig. 4). The feature maps are merged sequentially. One image with smaller FOVs is firstly passed through two ConvNets to get a smaller feature maps. While the other image with larger FOVs is passed through one ConvNets and get a larger feature maps. The depth of smaller feature maps are reduced to match that of larger feature maps by 1 × 1 convolution (Fig. 5a). Then they are partially merged by aligning them at the center and take means of the overlapped elements (Fig. 5b). After two partial fusions to combine the three multi-scale images, the integrated feature maps are processed by the remaining three layers of ConvNets and three FC layers for classification.

IV. EXPERIMENTS

A. Datasets

The dataset we use in this study is Part A – microscopy images from ICIAR 2018 Grand Challenge on Breast Cancer Histology images. The dataset contains 400 microscopy images, labeled as one of the following classes: normal, benign, *in situ* carcinoma and invasive carcinoma. Each class contains 100 images.The size of each image is 2048×1536 pixels. The labeling of images was performed by two medical experts, and label disagreements have been excluded from the dataset.

B. Data Normalization

Because of the variations in preparing each tissue sample and in scanning for microscopic examination, we found significant visual differences in the histology images. To prevent this artifact affecting our classification results, we use a color normalization method based on non-negative matrix factorization (NMF) [18]. Specifically, for a given image, the method

calculates stain color bases and stain densities in an unsupervised manner. We rebase the color distribution of each image by multiplying reference color bases and normalized stain densities. Our study uses Image b027 as the reference image.

C. Data Augmentation

Given that we have only limited image samples compared to typical DL applications, we augment our data to increase variations of the data and to reduce over-fitting when training our models. Specifically, we rotate the original image and crop it to a desired size (either 256×256, 512×512, or 1024×1024 pixels) (Fig. 1). We make sure that each cropped image will not contain more than 30% of white background. We resize each cropped image to 224 × 224 pixels to fit the input size of AlexNet, the backbone network architecture we use.

D. Settings of hyperparameters

We use four-fold cross validation and set 300 images as the training set and the remaining 100 images as the test set. Both the training and test sets have even distribution of four class labels. During training, the batch size is set to 60. We use stochastic gradient descent (SGD) and set momentum to 0.9 for optimization. We set a relatively large learning rate 0.01 when we do not use ImageNet-pretrained parameters, and a relatively small learning rate 0.0002 when we use ImageNet-pretrained parameters. We record the best accuracy for the test set during 200 training iterations for each fold and report the mean and standard deviation best accuracy for the four folds.

V. RESULTS

We list the prediction accuracy in (Table I). Below we talk about our major findings.

A. ImageNet Pretraining

We found out that using ImageNet-pretrained parameters to initialize our model significantly improved the prediction accuracy (Fig. 6). In Table I, the increase is 33.25%, 49.75%, and 43.75% under the FOV as 256 × 256, 512 × 512, and 1024×1024, respectively. ImageNet is a large image database that contains 1.28 million images that belong to 1, 000 classes. The classes cover a wide variety of common objects such as hen and cat. Although general image classification is very different from pathological image classification, the large and comprehensive ImageNet data set enables the convolutional filters of AlexNet to capture the heterogeneous and basic visual components in the world. Therefore, ImageNet pretrained parameters set the stage for training DL models for pathology image classification, given limited number of pathology images in the Challenge.

B. Field of View

From Table I, we see that choices of FOV affect prediction accuracy. Specifically, when we set a small FOV (256 × 256 pixels), the prediction accuracy is lowest with ImageNet pretraining among three sizes of FOV. With ImageNet pretraining, the medium FOV (512×512 pixels) achieves the best accuracy 78.75% among single-scale inputs. To better understand the effect of FOV, we visualize an example image in three scales and highlight important regions for AlexNet prediction with Grad-CAM [19]. We argue that although

large FOV include the most context information for classification, they lose the most details when we resize the input to fit the specified input size of AlexNet (224 × 224). On the other hand, small FOVs (256 × 256) are most probable to miss distinctive regions for identifying cancers. The medium FOV (512 × 512) maintains a good balance between context and detailed information (Fig 7).

C. Multi-scale Input

Table I shows that multi-scale input increases prediction accuracy up to 81.5% when we combine the features at a late stage by concatenating feature vectors before the last fully connected layers and share the parameters of AlexNet backbone for each scale of input (Fig. 8). The difference in accuracy between the two late fusion methods we use, concatenation and averaging of the feature vector before the last fully connected layers, is minor (81.5% and 79.75%). The difference between late fusion and the two early fusion methods by concatenating feature maps at Conv layer 3 and Conv layer 2 is also small (within 1.25%). However, the accuracy of partial fusion is only 72%, which is less than that of single scale input at FOV of 512 and 1024. We reason that the inferior performance of partial fusion may be caused by the interruption of the original AlexNet structures. With limited number of training data, the original AlexNet structure with pertained parameters contributes most to the classification performance.

VI. CONCLUSIONS AND DISCUSSIONS

In this paper, we have built a multi-scale ConvNets for late fusion and early fusion of the WSI image pyramids to improve the classification of breast cancer. Based on the extensive ablation experiments, we found that the late fusion method by taking average of feature vectors and sharing AlexNet pretrained parameters reaches the highest accuracy (80%), which beats the other methods including late fusion by concatenation, early fusion by concatenation, and early partial fusion. We conclude that 1) utilizing the AlexNet pretrained parameters, 2) multi-scale image pyramid inputs, and 3) sharing network parameters contribute to better classification performance. We hope the work can inspire more studies for multi-scale analysis of WSI pathological images.

ACKNOWLEDGEMENT

The work was supported in part by grants from the National Center for Advancing Translational Sciences of the National Institute of Health (NIH) under Award UL1TR000454, the National Science Foundation EAGER Award NSF1651360, Children's Healthcare of Atlanta and Georgia Tech Partnership Grant, Giglio Breast Cancer Research Fund, and Carol Ann and David D. Flanagan Faculty Fellow Research Fund. This work was also supported in part by the scholarship from China Scholarship Council (CSC) under the Grant CSC NO. 201406010343. The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

REFERENCE

[1]. Ghaznavi F, Evans A, Madabhushi A, and Feldman M. Digital imaging in pathology: whole-slide imaging and beyond. Annual Review of Pathology: Mechanisms of Disease, 8:331–359, 2013.

[2]. Pantanowitz L, Sinard JH, Henricks WH, Fatheree LA, Carter AB, Contis L, Beckwith BA, Evans AJ, Lal A, and Parwani AV. Validating whole slide imaging for diagnostic purposes in pathology: guideline from the college of american pathologists pathology and laboratory quality center.

Archives of Pathology and Laboratory Medicine, 137(12):1710–1722, 2013. [PubMed: 23634907]

[3]. Araujo T, Aresta G, Castro E, Rouco J, Aguiar P, Eloy C, Polonia A, and Campilho A. Classification of breast cancer histology images using convolutional neural networks. PloS one, 12(6):e0177544, 2017. [PubMed: 28570557]

[4]. Hou L, Samaras D, Kurc TM, Gao Y, Davis JE, and Saltz JH. Patch-based convolutional neural network for whole slide tissue image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2424–2433, 2016.

[5]. Kothari S, Phan JH, Stokes TH, and Wang MD. Pathology imaging informatics for quantitative analysis of whole-slide images. Journal of the American Medical Informatics Association, 20(6):1099–1108, 2013. [PubMed: 23959844]

[6]. Kothari Sonal, Phan John H, Osunkoya Adeboye O, and Wang May D. Biological interpretation of morphological patterns in histopathological whole-slide images. In Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine, pages 218–225. ACM, 2012.

[7]. Phan John H, Quo Chang F, Cheng Chihwen, and Wang May Dongmei. Multiscale integration of-omic, imaging, and clinical data in biomedical informatics. IEEE reviews in biomedical engineering, 5:74–87, 2012. [PubMed: 23231990]

[8]. Pereira Sergio, Pinto Adriano, Alves Victor, and Silva Carlos A. Brain tumor segmentation using convolutional neural networks in mri images. IEEE transactions on medical imaging, 35(5):1240–1251, 2016. [PubMed: 26960222]

[9]. Kumar Devinder, Wong Alexander, and Clausi David A. Lung nodule classification using deep features in ct images. In 2015 12th Conference on Computer and Robot Vision, pages 133–138. IEEE, 2015.

[10]. Wu Hang, Tong Li, and Wang May D. Improving multi-class classification for endomicroscopic images by semi-supervised learning. In 2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), pages 5–8. IEEE, 2017.

[11]. Yu Kun-Hsing, Zhang Ce, Berry Gerald J, Altman Russ B, Christopher R e, Rubin Daniel L, and Snyder Michael. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. Nature communications, 7:12474, 2016.

[12]. Tong Li, Hoffman Ryan, Deshpande Shriprasad R, and Wang May D. Predicting heart rejection using histopathological whole-slide imaging and deep neural network with dropout. In 2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), pages 1–4. IEEE, 2017.

[13]. Deng J, Dong W, Socher R, Li L-J, Li K, and Fei-Fei L. Imagenet: A large-scale hierarchical image database. In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pages 248–255. IEEE, 2009.

[14]. Lin T-Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollar P, and Zitnick CL. Microsoft coco: Common objects in context. In European conference on computer vision, pages 740–755. Springer, 2014.

[15]. Lin T-Y, Dollar P, Girshick R, He K, Hariharan B, and Belongie S. Feature pyramid networks for object detection In CVPR, volume 1, page 4, 2017.

[16]. He K, Gkioxari G, Dollar P, and Girshick R. Mask r-cnn. In Computer Vision (ICCV), 2017 IEEE International Conference on, pages 2980–2988. IEEE, 2017.

[17]. Xu Z and Zhang Q. Multi-scale context-aware networks for quantitative assessment of colorectal liver metastases. In Biomedical & Health Informatics (BHI), 2018 IEEE EMBS International Conference on, pages 369–372. IEEE, 2018.

[18]. Vahadane A, Peng T, Sethi A, Albarqouni S, Wang L, Baust M, Steiger K, Schlitter AM, Esposito I, and Navab N. Structure preserving color normalization and sparse stain separation for histological images. IEEE transactions on medical imaging, 35(8):1962–1971, 2016. [PubMed: 27164577]

[19]. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, and Batra D. Grad-cam: Visual explanations from deep networks via gradient-based localization. See https://arxiv.org/abs/1610.02391v3, 7(8), 2016.703

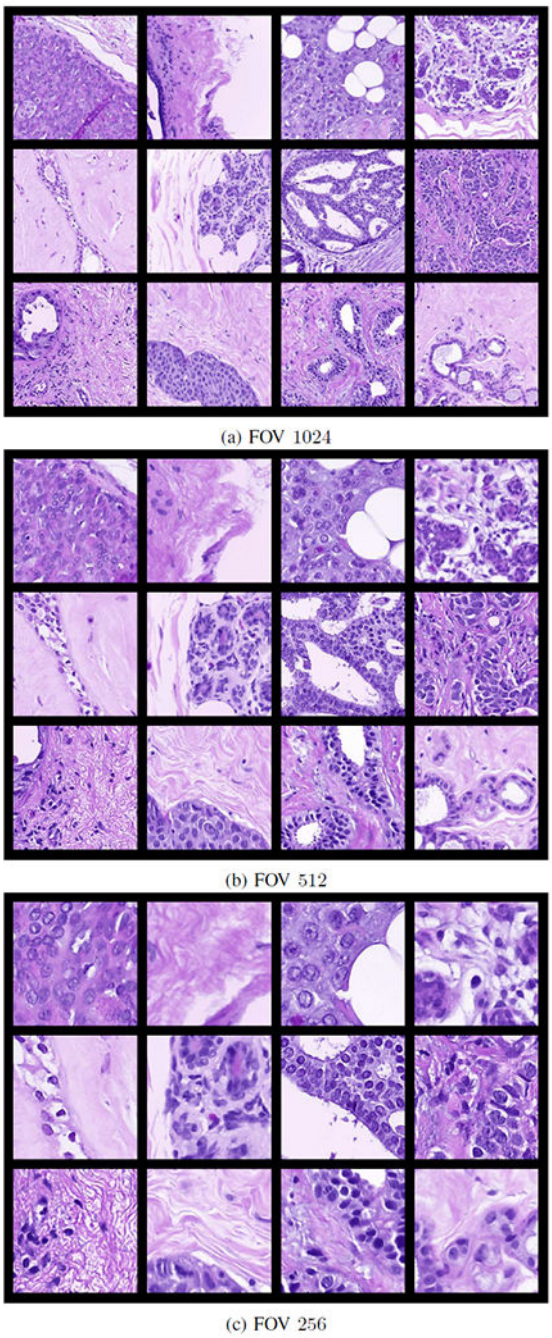


Fig. 1:
Visualization of concentric images with different FOVs. From (a) to (b), the FOV is getting smaller and smaller, with more zoomed in view of the same image patch.

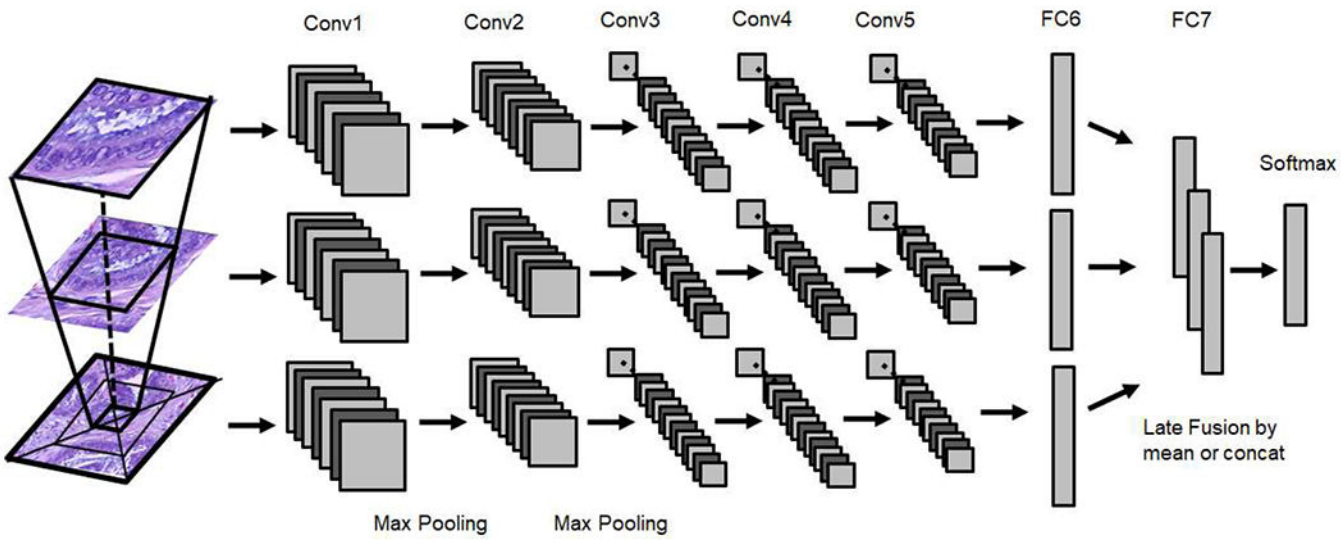


Fig. 2:

Visualization of multi-scale patch integration by late fusion. The information extracted from each scale is not combined until feeding into classification layer. The feature vectors represented from each scale of concentric image patches are integrated by either concatenation or taking average. The combined feature vectors are then fed into last FC layer for classification.

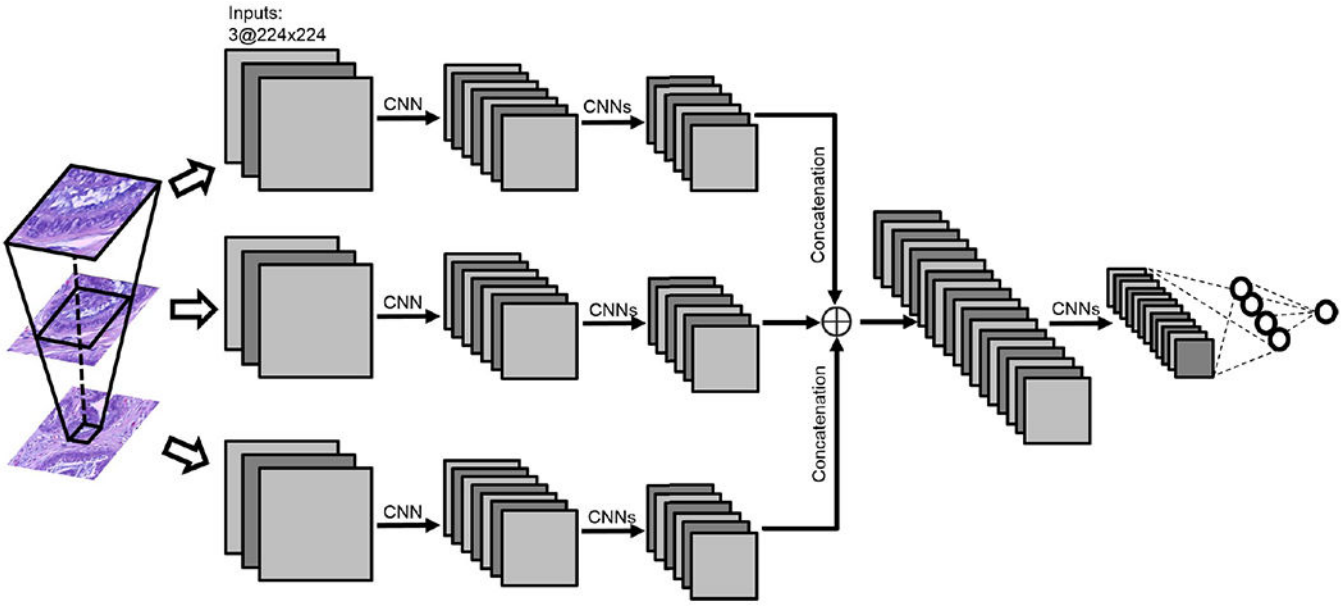


Fig. 3:

Visualization of multi-scale patch integration by early fusion and full concatenation. The feature maps generated by convolutional neural networks are fully concatenated. We have tried to concatenate the feature maps at the third and fourth ConvNet layers respectively.

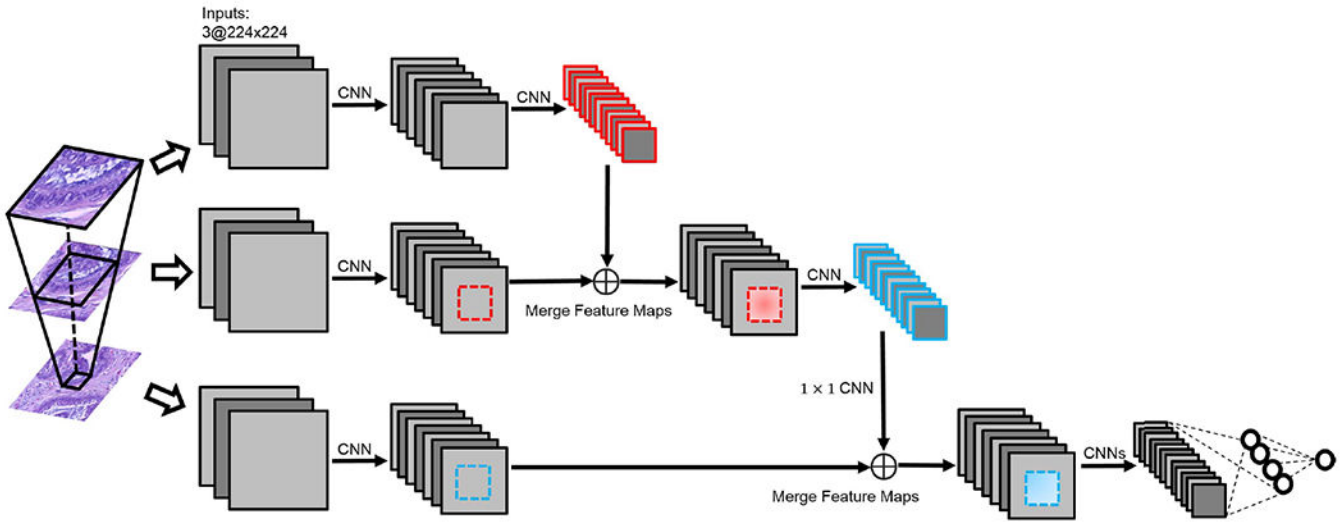
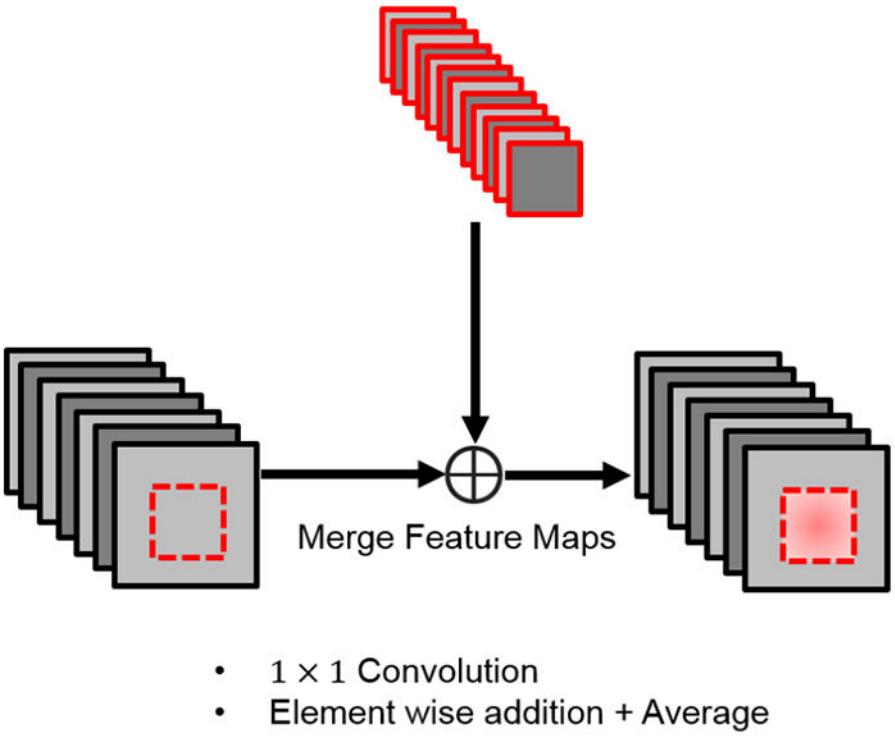
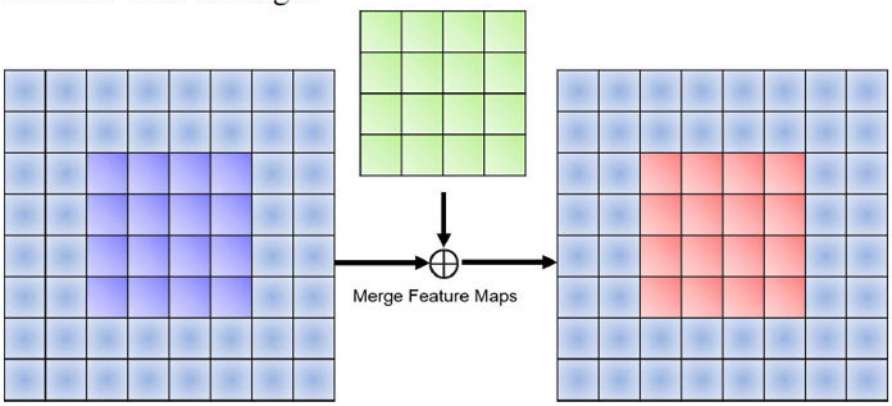


Fig. 4:

Visualization of multi-scale patch integration early partial fusion. We fuse the feature maps of multi-scale images sequentially. The image with smaller field of view is processed with two ConvNet layers and then fused with the image processed with one ConvNet layer.



(a) Partially merge the feature maps by 1×1 convolution and element wise average.



(b) The illustration of partially overlapped feature maps based on field of views.

Fig. 5:
Details of partial fusion methods.

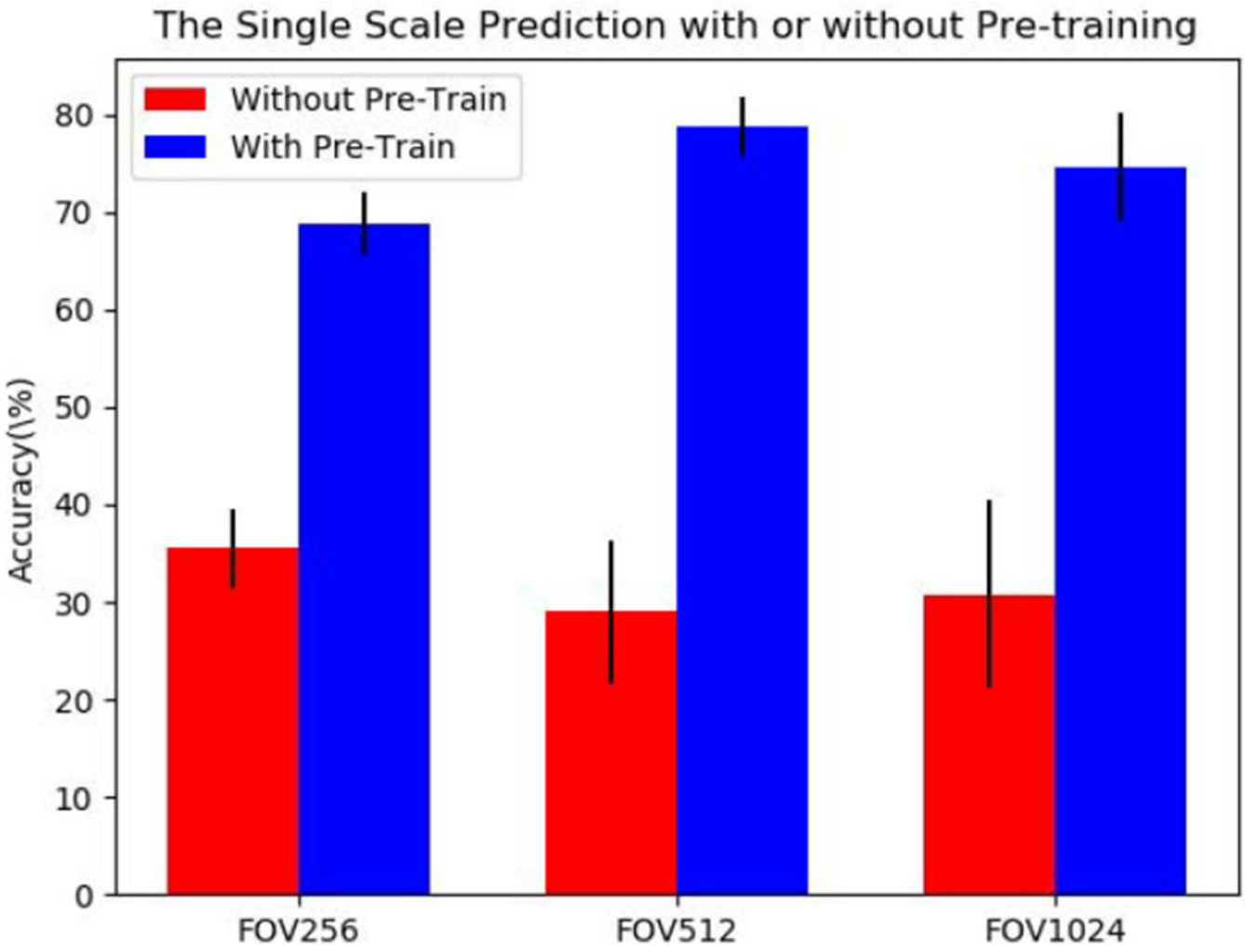


Fig. 6:

Bar plot for prediction performance of single-scale images with or without ImageNet pretraining. For all single-scale inputs, ImageNet pretraining significantly improved the prediction performance.

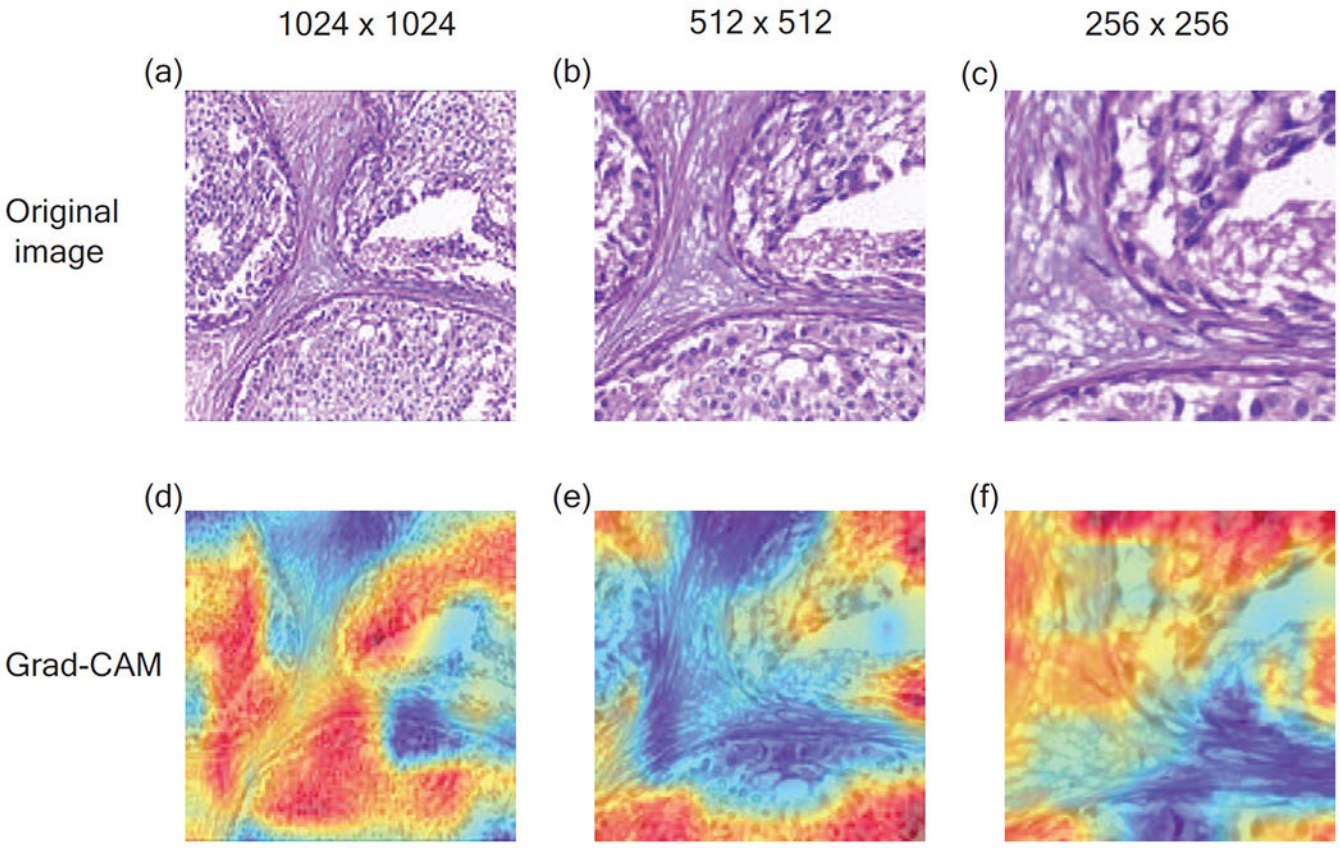


Fig. 7:
Visualization of important regions for predictions in different scales of an example image using Grad-CAM [19].

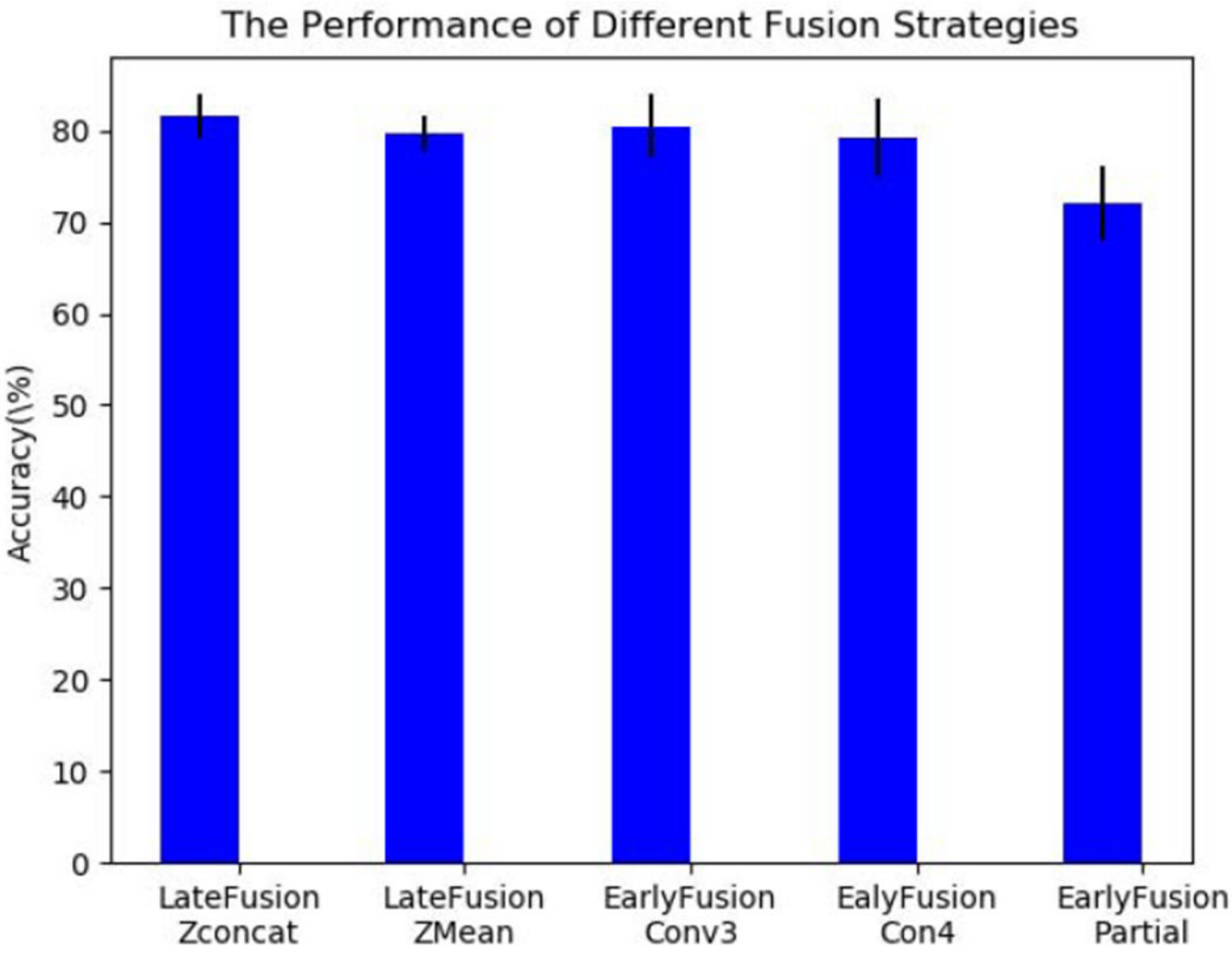


Fig. 8:

Bar plot for prediction performance of five multi-scale fusion methods. All networks are pretrained with ImageNet. LateFusion by ZConcat achieves the highest performance.

TABLE I:

Experiment Results for Multi-Scale ConvNets

Multi-Scale	FOV-256	FOV-512	FOV-1024	Fusion	ImageNet-Pretrained	Accuracy
Single Scale	+	–	–	–	–	35.50 ± 4.04 %
	+	–	–	–	+	68.75 ± 3.20%
	–	+	–	–	–	29.00 ± 7.35%
	–	+	–	–	+	78.75 ± 2.99%
	–	–	+	–	–	30.75 ± 9.60%
	–	–	+	–	+	74.50 ± 5.57%
Late Fusion	+	+	+	Z concat	+	81.50 ± 2.38%
	+	+	+	Z mean	+	79.75 ± 1.89%
Early Fusion	+	+	+	Conv3 concat	+	80.50 ± 3.41%
	+	+	+	Conv4 concat	+	79.25 ± 4.27%
	+	+	+	Partial fusion	+	72.00 ± 4.08%