HH-Net: Image driven microscope fast auto-focus with deep neural network

Yunhao Ge
Robotics Institute
Shanghai Jiao Tong University
Shanghai, China
gyhandy@sjtu.edu.cn

Bin Li
Robotics Institute
Shanghai Jiao Tong University
Shanghai, China
lbin_sjtu@sjtu.edu.cn

Yanzheng Zhao, Weixin Yan
State Key Lab of Mechanical System and Vibration
Shanghai, China
xiaogu4524@sjtu.edu.cn

ABSTRACT

Computer aid auto-focus system is necessary for accurate microscope diagnosis, especially for the high precision microscope, which leaves little physical distance for focus adjusting manually. We proposed an image-driven microscope fast auto-focus system with a deep neural network. There are two main contributions. First, combining the high-level feature learning ability advantages of convolution neural network (CNN) and the handcraft feature selection ability of statistical learning, we proposed a High-level-Handcraft Neural Network (HH-Net) to accurately determine the distance index between microscope lens and cell smear by evaluating the image focus quality. It deployed 13 layers CNN for the high-level feature extraction from image patches. While the handcraft features which provide global information from the raw image were extracted by statistical algorithms and merged into CNN features. Finally, the combined features are utilized by the fully connected layers in the network to obtain the final distance index by classifying the biomedical image focus quality. Second, cooperated with the HH-Net, we propose an end to end image driven microscope fast auto-focus system, that can learn auto-focus policies from visual input and finish at a clear spot automatically. The accuracy of our patch level focus quality prediction is 92.4% with HH-Net, while the real-time image level focus quality predication can be 99.99% with 0.025s cost time by certainty voting strategy. Our auto-focus system can also cooperate with the X-Y Micro platform to automatically scan the whole cell smear and get the real-time best-in-focus image of a microscope with fast response, accuracy, and robustness.

CCS Concepts

• **Computing methodologies → Object recognition.**

Keywords

Fast auto-focus, Image focus quality, Deep neural network, High-level features, Handcraft features, Convolutional neural network, Microscope.

1. INTRODUCTION

A large number of microbiological samples are analyzed by researchers in biological areas every day. For some of the high-resolution microscopes, the adjusting range is so limited that can hardly achieve focusing process manually. The need for fast, powerful, and reliable automated systems increases as these analyses deal with higher-resolution images. One step in an automatic system to capture microbiological images is to obtain the best-in-focus image from a biological sample, which is a challenging task. Many autofocusing methods have been developed. Automatic focus is widely used in curve estimation [1-2]. Some researchers adopt mechanical structures and laser sensor [3-4] for the distance detection while in high-resolution detection part, the move of the Z axis is tiny. Hardware adjust can easily be influenced by manufacture errors and assembly errors. Another direction is to use image-based approaches to obtain the best in-focus image from a set of captured microscopic images and use it to finish the focus step, for example, P. Jiao et, al developed an auto-focusing algorithm based on GLCM features which contain the texture information of the image [5]. Image-guided autofocus is necessary. Some of the algorithms developed for this purpose are the analysis of the global and local variance of the images' gray levels to get a measure of their contrast [1,2]. Machine learning methods such as decision trees [6] are also adopted in this field. Besides, fresnel transformation was also used for the feasibility of automatic focusing [7]. Nonlinear correlation [8] and Prewitt edge detection operator [9] were used to evaluate the image quality. All these algorithms have proven to be effective in obtaining the best-in-focus image; however, utilizing more features means more computation and more time, they require considerable time when the images have high resolution.

With the development of the convolutional neural network (CNN), which can automatically extract high-level features like semantic, big structure, image style, CNN based algorithm can perform great on image classification task. Some expert tried to use CNN to design deep neural network (DNN) to make image quality classification, which may be helpful for the auto-focus task [10], however, the accuracy is not high enough because most of the DNN is the end to end network which makes the final classification by multi convolutional layers. The big image may lose some spatial details in the convolution layers and pooling layers to make final classification in CNN.

In this paper, we proposed a High-level-Handcraft Neural Network (HH-Net) for large biomedical image in-focus quality classification, which can be used to fast autofocus for the microscope. We take advantage of CNN's high-level feature extraction and the powerful nonlinear fitting ability of large scale data. We crop the raw large 2048*1536 biomedical cell smear slides image into 300*300 patches and use the convolutional layers to extract high-level features from them. To complement the lost global information of the whole image, we use DLCM features, statistical features, correlation coefficient features and contrast location features, extracted from the raw image, to help the patch level image quality classification. Finally, we use the certainty combination of each patch to make the final image level classification of image focus quality. The transform of distance index from the image focus quality can be used for the auto-focus system. The real-time distance output can be used to adjust the velocity of focus motor along the Z axis to help make autofocus accurately. Section 3 provides the hardware and the whole autofocus system of the microscope. Section 4 shows the computational experiments and provides the graphical results of those experiments, where we illustrate the performance of the

proposed algorithm. And finally, Section 5 summarized our conclusions and planned future work.

2. Biomedical Image Focus quality classification based on High-level-Handcraft Neural Network (HH-Net)

The contribution of our High-level-Handcraft neural network (HH-Net) is that, first, it combines the advantage of the convolution neural network's high-level feature extraction ability and the strength of statistical and handcraft feature selection ability based on prior knowledge. Second, it fused the global information from row image and details information in image patches to make patch level classification decision. Third, we use a certainty-based voting method to guarantee the final prediction accuracy on image level focus quality classification.

2.1 Dataset

In this study, we create the Stained Circulating Tumor Cell (CTC) Samples dataset and get the biomedical image under the microscope. We captured a set of several images at different Z distances from the Stained CTC samples. Depending on the distance between the microscope lens and sample smear slides and the out-focus image quality of the captured image. We manually separate the image into seven classes or defocus levels (Totally blur, out of focus degree III, out of focus degree II, out of focus degree I, in-focus, too close I, too close II), as shown in Fig.1. Each class contains 350 raw RGB images with 2048*1536 resolution, the training set includes 300 raw images, and the testing set contains 50 raw images. We use multiple data augmentation when we do training of HH-net.

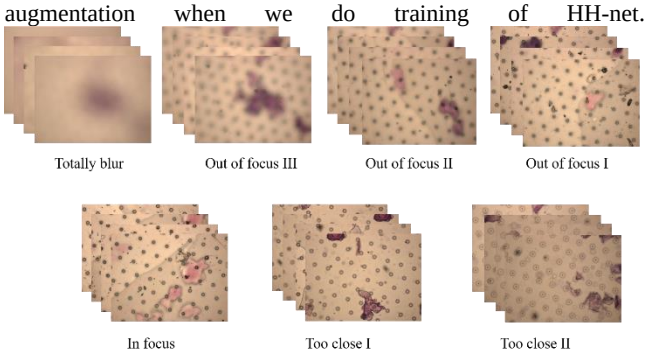

Fig.1 Stained Circulating Tumor Cell (CTC) Samples dataset, the images are separated into seven classes or defocus levels depending on the distance between microscope lens and sample smear slides (Totally blur, out of focus degree III, out of focus degree II, out of focus degree I, in-focus, too close I, too close II)

2.2 Implementation

As shown in Fig.2, the framework of our High-level-Handcraft Neural Network (HH-Net) includes four key components (1) Raw image crop and data augmentation (2) High level features extraction by CNN and Patch level classification by network; (3) Handcraft features extracted by statistical learning and prior experience; (4) Image level focus quality prediction by certainty-based voting algorithm.

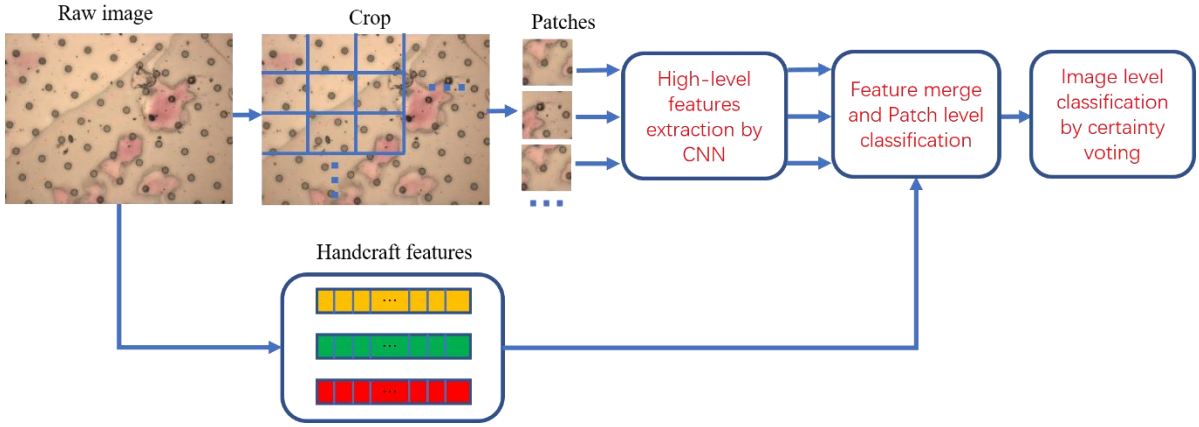

Fig.2 The skeleton of proposed High-level-Handcraft neural network (HH-Net).

2.2.1 Raw image crop and data augmentation

Different from the semantic information which needs a large field of vision, the distance classification task concentrates on the regain pixel blur details which reflect the image quality. On the contrary, if we use CNN to make distance classification with the whole image as input, we need more layers to make convolution or pooling which shrink the size of images to make a final classification. However, there is a significant shortage of CNN about details information lost while making convolution and pooling. To solve the problem of the large size input image, we crop the raw 2048*1536 images into 300*300 patches. The patches images are treated as CNN's input which can both shrink the network structure and decrease the pooling layers in CNN to eliminate the information loss.

Data augmentation is often used in the context of deep learning and refers to the process of generating new samples from existing data, which is used to ameliorate data scarcity and prevent overfitting (Kooi et al., 2017) [11]. Transformations include rotations, translations, horizontal and vertical reflections, crops, zooms, and jittering. For tasks such as optical character recognition, (Simard et al., 2003) [12] showed that elastic deformations could greatly improve performance. The primary sources of variation in our CTC images are rotation, scale, translation, and amount of occluding tissue. By combining different argumentation method, every lesion can be shown in any specific orientation, up to 32 times, and argumentation-realistic lesion candidates can be obtained to eliminate the overfitting problem.

2.2.2 High level features extraction by CNN and Patch level classification by network

A deep convolutional neural network was trained on the patch level image focus quality classification task. Given training examples of 300 × 300 pixel, input image patches cropped from

raw images and the corresponding labels of image focus quality. The first 13 layers of CNN can extract the high-level features from input patches image. Fig.3 and table 1 show the network architectural details. The stride of the kernel is fixed as 1 and Rectified Linear Units (ReLU) are used as the activation function for each convolutional layers. In the first fully connected layers, 224 nodes connected with the feature map of last convolution layers, other 32 nodes connected with the extracted handcraft features from raw images which contains the global information. Then the second fully connected layers combined the whole features from last layers. Moreover, a dropout layer with probability 0.7 and 0.5 respectively were connected with the first two fully connected layers. The model was implemented and trained with TensorFlow.

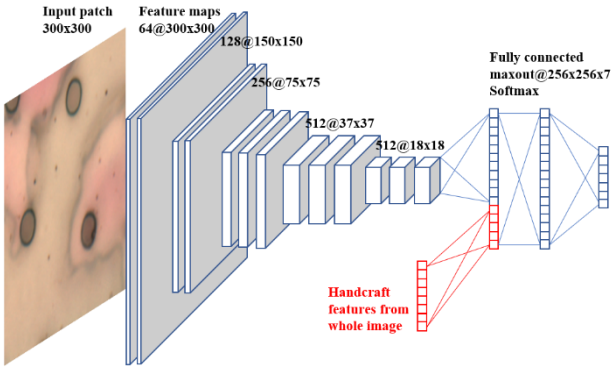

Fig.3 The framework of the neural network model architecture of HH-net

Table 1 An illustration of the HH-net Configurations

HH-net Configurations (16 weight layers**)**
Input(300×300 RGB patch images)
2 times @ Conv5-64
Max pooling
2 times @ Conv3-128
Max pooling
3 times @ Conv3-256
Max pooling
3 times @ Conv3-512
Max pooling
3 times @ Conv3-512
FC-256(combined handcraft features) (Dropout rate-0.7)
FC-256 (Dropout -0.5)
FC-7
soft-max

2.2.3 Handcraft features extraction by statistical learning and prior experience

To make global information complementary, we add the statistical and handcraft feature extracted from the whole raw image for the classification layers of HH-net. The handcraft features and statistical features consist of statistical parameters, GLCM features, location features, and correlation coefficient features. Below is a list of eighteen features extracted from the mask area of the original image; they were selected to be combined with the output of the final convolutional layer to train and test our method.

Statistical parameters: Four statistical parameters of the whole raw image were extracted: mean, variance, skewness, and kurtosis. For each individual pixel value f_k, $p(f_k)$ is the probability of this unique pixel value in the whole ROI. The four parameters are calculated as follows:

$$mean: \mu = \sum_{k=1}^{N} f_k p_f(f_k) \tag{1}$$

$$\text{var}\,iance: \sigma^2 = \sum_{k=1}^{N} (f_k - \mu)^2 p_f(f_k) \tag{2}$$

$$skewness: ske = \sum_{k=1}^{N} [(f_k - \mu)^3 p_f(f_k)] / \sigma^3 \tag{3}$$

$$kurtosis: kur = \sum_{k=1}^{N} [(f_k - \mu)^4 p_f(f_k)] / \sigma^4 \tag{4}$$

GLCM features: GLCM features consists of sum entropy (SE), sum average (SA), difference variance (DV), and difference entropy (DE). SE is a logarithmic of ROI in consideration. SA is calculated from the ROI and the size of gray scale. DV is a variance measure between the ROI intensities calculated as a function of the SE calculated previously. DE is an entropy measure which provides a measure of no uniformity while taking into consideration a difference measure obtained from the original image. And these four parameters are calculated as follows:

$$SE = -\sum_{i=2}^{2N_g} p_{x+y}(i) \log\{p_{x+y}(i)\} \tag{5}$$

$$SA = \sum_{i=2}^{2N_g} i p_{x+y}(i) \tag{6}$$

$$DV = \sum_{i=2}^{2N_g} (i - SE)^2 p_{x-y}(i) \tag{7}$$

$$DE = -\sum_{i=2}^{2N_g} p_{x-y}(i) \log\{p_{x-y}(i)\} \tag{8}$$

Location features: Four parameters about ROI location and shape were extracted: convexity (C(S)), compactness (C), aspect ratio (AR), area ratio (*R_Area*). The four parameters are calculated as follows:

$$C(S) = \frac{A}{Area(CH(S))} \tag{9}$$

$$C = \frac{P^2}{4\pi A} \tag{10}$$

$$AR = \frac{D_y}{D_x} \tag{11}$$

$$R\_Area = \frac{Area\_ROI(in\_pixels)}{Area\_window(in\_pixels)} \tag{12}$$

Where: S is a ROI cropped from raw image, CH(S) is its convex hull and A is the ROI's area, P is the ROI's perimeter, and $Area\_window = D_x * D_y$, D_x is the width's ROI and D_y is the height's ROI.

Correlation coefficient features: randomly crop four 100*100 patches from raw images and calculated the correlation coefficient values between any two of them. The correlation coefficient defined in Eq.13 is used as the metric to describe how the pixels from one patch relate to pixels on a second patch.

$$L_{cc}(X,Y)=\frac{Cov(X,Y)}{\sigma_X \sigma_Y} \tag{13}$$

where *Cov* denotes the covariance. σ denotes the variance, X, Y denote the two patches.

The 18 handcraft features are connected with the fully connected layers in HH-Net and make global information complement for patch level image quality classification.

2.2.4 Image level classification by certainty-based voting

The final classification is depended on the certainty-based voting algorithm which makes more accurate classification of image focus quality. The certainty-based voting algorithm can aggregate the independent predictions on non-overlapping patches within a single image to make the final image level image focus prediction. For each 300*300 image patch, the predicted probability distribution output, for i ∈ {1,⋯, C} in our case C = 7 quality levels, yields a measure of certainty in the range [0.0, 1.0]. The output is computed by normalizing the information entropy of the distribution [13]:

$$C_{er}=1-\frac{\sum_{i=1}^{C} P_i \log p_i}{\log C} \tag{14}$$

To obtain the whole-image predicted probability distribution. We take a certainty-weighted average of the distributions predicted for the individual patches. The final image level quality class will have the biggest certainty-weighted average probability.

3. Microscope Autofocus system Based on High-level-Handcraft neural network (HH-Net)

3.1 Hardware and devices

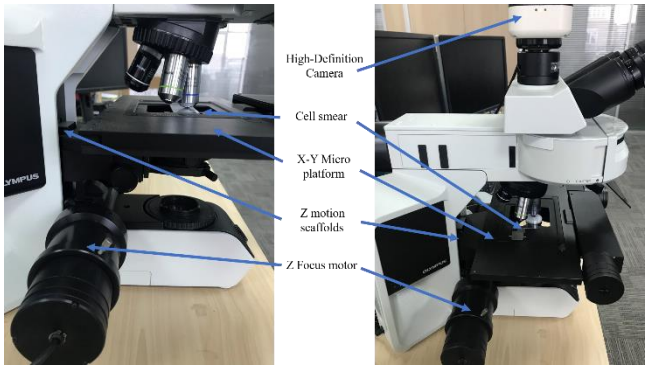

Fig.4 Hardware and devices for image driven auto-focus system

As is shown in Fig.4, the hardware of our auto-focus system mainly consists of High-Definition camera, X-Y Micro platform, Z motion scaffolds, Focus motor and Computer Controller. All of the computing including the HH-net prediction is finished on the Computer controller, and the focus motor can adjust the movement of the X-Y micro platform which achieve the auto-focus task. We can see from the Fig.4 that the distance between the objective of the microscope and cell smear is quite small which can hardly be adjusted manually.

3.2 Auto-focus algorithm based on High-level-Handcraft neural network (HH-Net)

After obtaining the image focus quality, we need to convert the quality class to distance index which reflects the physical distance between the objective of microscope and cell smear. We turn the focus status to distance index (from far to close, from "totally bure" to "Too close II" is "4 to -2", and 0 reflect the in-focus status). Then we proposed an auto-focus control algorithm to adjust the movement of the X-Y micro platform to get the real-time best-in-focus image of the microscope with fast response, accuracy, and robustness.

Algorithm1 shows the skeleton about our control algorithm; the focus status is the image level focus quality. The converted distance index is not strictly linear, but the value can reflect the approximate relationship of size. The default speed is a safe speed of focus motor, nearly 0.2237 mm/s

Algorithm 1 Autofocus algorithm based on High-level-Handcraft neural network (HH-Net)

Input (HH-Net**):** Raw CTC image

Output (HH-Net**):** focus status

Turn focus status to distance index (from far to close: 4 to -2)

While distance index != in focus (0)

If distance index == 4

Move the focus motor on Z axis (speed = default speed)

Else

Move the focus motor on Z axis (speed = 0.2*default speed * distance index)

4. Computational Experiments and Results

4.1 HH-net training and testing performance

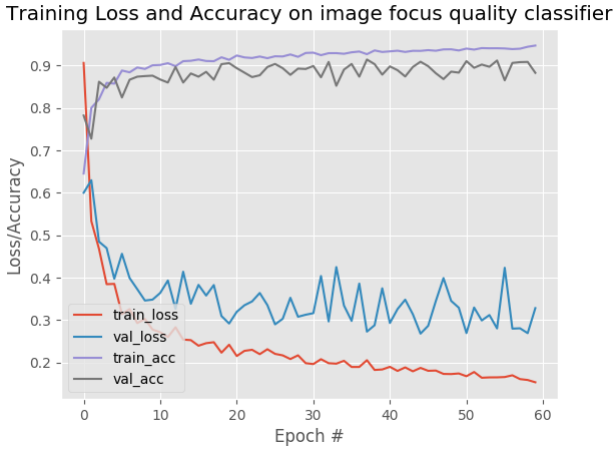

Fig.5 Traing and testing log of HH-net

The validation loss has fluctuation mainly because the number of testing data may not be big enough. Fig.5 can reflect the accuracy of patch level classification with an accuracy of 92.4%. The high patch level prediction is crucial for our final image level quality classification task.

Table 2. Training Methods of the HH-Net in Our Experiment

Training Methods	Optimization Algorithm	Learning Rate (α)	Dropout (p)
Option/value	Adam	0.001	0.7 and 0.5
Function	Adjust the learning rate	Speed up the training procedure	Reduce the overfitting substantially

Table 2 shows the training methods of HH-net in our experiment. Table 3 shows the performance of different voting methods. The image level prediction time shows the response time in our auto-focus system. Even though the hard voting method has smaller prediction time, the certainty based voting method performs better in the final system due to the high accuracy. The prediction time for our HH-net for each image is 0.025 s, which can achieve the real-time adjust for the movement of focus motor under certain safety velocity.

Table 3 Comparation of different voting methods

Voting Methods	Hard voting	Linear weighted voting	Certainty based voting
Image level Accuracy (%)	97.23	97.67	**99.99**
Image level prediction time (s)	0.022	0.024	0.025

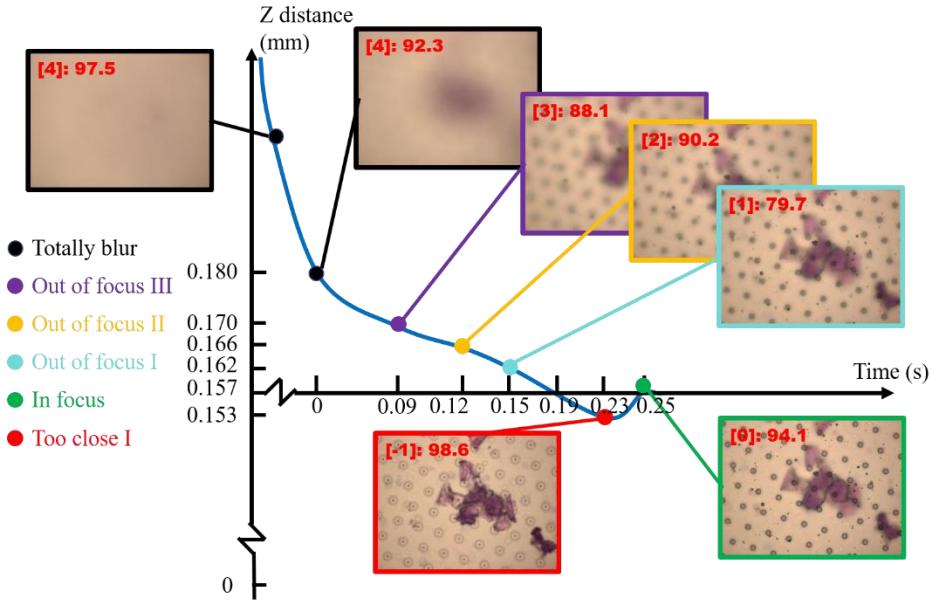

Fig.6 Dynamic response curve of our auto-focus system

Fig. 6 showed the dynamic response curve of our auto-focus system on practical test. The x-axis denotes response time and the y-axis represent the Z direction relative distance between microscope lens and cell smear. Due to the different starting point of the auto-focus system, we set the starting time as the last time of "Totally blur" class, which reflect the "4" distance index and calculate the relationship between the response time and distance. The crossing point responses the in-focus position which has 0.09 mm Z distance. The values on status image represent the predicted distance index and probability by HH-Net. The average auto-focus time is 0.25s and the error between final in-focus status and ideal status is less than 0.001mm.

5. Conclusions and Future Work

Our proposed High-level-Handcraft neural network (HH-Net) enables classification of focus quality microscope images with both higher accuracy and precision. The crop of the raw image can avoid the information loss in multi-layers convolution and pooling. The CNN layers can automatically extract the high-level features from patches images. The handcraft features extracted from the whole raw image by statistical learning method can provide global information complement for network classification. Fully connected layers merge the extracted high-level features and handcraft features in the network and make a final classification. To further improve the accuracy, we proposed a certainty-based voting algorithm to make image level focus quality classification. The HH-net achieved 92.4% accuracy of patch level classification and 99.99% final image level classification accuracy. Based on the HH-net, we convert the image quality classification output to distance index which reflects the distance about physical focus

distance. The experiment performance showed that our autofocus control algorithm has a tremendous dynamic response with 0.025s prediction time, which achieved real-time auto-focus. And the error of final in-focus status and ideal status is less than 0.001mm which guarantee the precision in use. Our auto-focus system can also cooperate with X-Y micro platform to achieve automatically in-focus cell smear biomedical image collection.

6. REFERENCES

[1] Lee, H.-J. and T.-H. Park. Auto-focusing system for a curved panel using curve estimation. in Information and Automation (ICIA), 2016 IEEE International Conference on. 2016. IEEE.

[2] Park, H.-D., A Study of Edge Detection for Auto Focus of Infrared Camera. 한국컴퓨터정보학회논문지, 2018. 23(1): p. 25-32.

[3] Cao, B.X., et al., Automatic real-time focus control system for laser processing using dynamic focusing optical system. Optics Express, 2017. 25(23): p. 28427-28441.

[4] All-optical microscope autofocus based on an electrically tunable lens and a totally internally reflected IR laser

[5] JIAO, P., et al., Auto-focusing algorithm based on gray level co-occurrence matrix. Optical Technique, 2018(3): p. 4.

[6] Haiyan, J., Z. Shanshan, and Z. Hongmin, An auto-focus algorithm based on machine learning. Microcomputer & Its Applications, 2016. 10: p. 004.

[7] Qiu, P., The feasibility of automatic focusing in digital holography by using Fresnel transform as numerical holographic reconstruction algorithm. Optik-International Journal for Light and Electron Optics, 2017. 137: p. 220-227.

[8] Cabazos-Marín, A.R. and J. Álvarez-Borrego, Automatic focus and fusion image algorithm using nonlinear correlation: Image quality evaluation. Optik, 2018. 164: p. 224-242.

[9] Lofroth, M. and E. Avci. An Auto-Focusing Approach for Micro Objects at Different Focal Planes. in 2018 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM). 2018. IEEE.

[10] Yang SJ, Berndl M, Michael Ando D, Barch M, Narayanaswamy A, Christiansen E, et al. Assessing microscope image focus quality with deep learning. BMC Bioinformatics. 2018. 19: 28962.

[11] Kooi, T., Litjens, G., van Ginneken, B., Gubern-Merida, A., Sanchez, C. I., Mann, R., den Heeten, A., & Karssemeijer, N. 2017. Large scale deep learning for computer aided detection of mammographic lesions. Med Image Anal, 35: 303-312.

[12] Simard, P.Y., Steinkraus, D., Platt, J.C., 2003. Best practices for convolutional neural networks applied to visual document analysis. Document Analysis and Recognition, pp. 958-963.

[13] Shannon CE. The mathematical theory of communication. 1963. MD Comput. 1997;14:306–17.

Columns on Last Page Should Be Made As Close As Possible to Equal Length

Authors' background

Your Name	Title*	Research Field	Personal website
Yunhao, Ge	**Master student**	**Deep learning; semantic segmentation; medical image analysis**	
Li, Bin	**Master student**	**Deep learning; semantic segmentation; medical image analysis**	
Yanzheng Zhao	**Professor**	**Robotics, mechanical design, control**	
Weixin, Yan	**Associate professor**	**Deep learning; medical image analysis; robotics**	

***This form helps us to understand your paper better, the form itself will not be published.**

***Title can be chosen from: master student, Phd candidate, assistant professor, lecturer, senior lecturer, associate professor, full professor, research, senior research**