HQNet: An Efficient Convolutional Neural Network for Cervical Cancer Classification

1st Han YANG
Program of Computer Science and Technology
BNU-HKBU United International College
China
p930026144@mail.uic.edu.cn

2nd Teoh Teik Toe
NTU Business AI Lab
Nanyang Technological University
Singapore
ttteoh@ntu.edu.sg

***Abstract*—With the rapid development of artificial intelligence, cancer cell classification recognition is also gradually intelligent. Although the classification and recognition methods for some common cancers are mature, Cervical Cancer, as a common cancer among middle-aged women, is often neglected in diagnosis because of the low incidence of malignant tumors. In recent years, the incidence of Cervical Cancer has been increasing year by year and has also gained attention. In this paper, an efficient deformable convolutional neural network system(HQNet) was constructed to identify Cervical Cancer cells with different degrees of development.**

***Keywords*—CNN, Deformable Convolution, Cervical Cancer Classification**

I. INTRODUCTION

Deep learning has evolved rapidly in the last decade, allowing people to build neural networks with large parametric quantities and very complex structures. These neural networks are widely used in industry, medicine and everyday life. They have a very efficient performance, can extract the required features very precisely, and are more accurate in delineating decision boundaries. In recent years, many efficient neural network algorithms have emerged in the field of computer vision and image processing. In the last century, Yann LeCun introduced the idea of convolutional neural network to the field of computer vision and constructed LeNet-5 [1], which successfully accomplished the task of handwritten letter recognition. LeNet uses local links and weight sharing to make the network performance more robust, and uses tanh as the activation function, proving that a symmetric activation function can help the network converge faster. This is also the reason why we have paid special attention to the activation function selection in the process of proposing HQNet.

After LeNet, deeper convolutional neural networks were proposed and a number of breakthroughs were made in the field of image recognition and classification. Deep networks stack convolutional layers with rich features in an end-to-end manner. Several studies [2] [3] have demonstrated that deeper convolutional neural networks achieve more leading results. Among them, AlexNet [4] and VGG [5] networks, as representatives of deep networks with 30 layers or less, are widely used in image recognition, image processing, and other fields with perfect results. However, if you want to build a deeper network structure with more than 30 layers, it cannot be achieved by stacking convolutional layers. In the process of stacking convolutional layers, the network accuracy will be saturated or even decreased.In 2015, Kaiming He et al. proposed ResNet [6], which solves the problems of gradient disappearance, gradient explosion and training difficulties of deep networks through the residual structure. Using such a network structure, networks with hundreds or thousands of layers can be constructed.The design process of HQNet also found that the performance of deep-layer networks would be better than shallow-layer networks. HQNet also adopts the residual learning structure to design and build a deep convolutional neural network, which has shown excellent performance.

Cervical Cancer as the second killer of malignant tumors in women. Its incidence rate is second only to breast cancer. In the past three years, the incidence of Cervical Cancer has been increasing year by year, and the WTO draft [7] states that by 2030, the incidence of Cervical Cancer will reach 700,000 and the number of deaths will exceed 400,000. However, in the field of medical image processing and cancer cell detection, only the common cancers with high mortality rate are often focused on, and the detection of Cervical Cancer cells with different levels of development is often neglected. HQNet is the classification and detection of five types of Cervical Cancer cells with different levels of development, and has shown good performance.

II. RELATED WORKS

Convolutional Neural Network. Since Krizhevsky proposed AlexNet in 2012, convolutional neural networks have been very successful in large-scale image recognition and classification. Convolutional neural network consists of convolutional layer, pooling layer, activation function, fully connected layer, etc. Convolution can be achieved by weighting and summing the pixel points contained in the convolution kernel to obtain a new pixel point, which reflects the features of the original image at the image level. AlexNet is a new way to use convolutional neural network to deal with image problems, and the proposed Dropout [8] method can effectively suppress overfitting during network training. In contrast to LeNet, AlexNet uses ReLU [9] as the activation function, which effectively reduces the gradient disappearance in network training and again demonstrates that the selection of activation function can effectively improve the network efficiency. Since then, many excellent network designs have emerged in this area, most of which improve network performance by increasing the number of convolutional layers. Among them, VGG successfully boosted the convolutional neural network to 19 layers. Subsequent studies have shown that increasing the number of convolutional layers can indeed effectively improve the performance of the network, which is the reason why HQNet uses deep convolutional neural networks.

Residual Structure. In the field of computer graphics and low-level vision, residual representation [10] is widely used. The VLAD [11] and Fisher [12] vectors are used as representatives of the coded representation by residual vectors. This coding representation can be effectively applied to the field of image retrieval and image classification. Short connection [13] has been applied to train multi-layer perceptrons (MLPs) [14] in early experiments and has been shown to be effective in suppressing the gradient explosion and gradient disappearance problems encountered in network training. In order to effectively improve the performance by increasing the number of convolutional layers of a convolutional neural network, Kaiming He combined residual representation and short connection to construct a residual structure, which was referred to a convolutional neural network to construct a deep neural network, ResNet. ResNet is composed of a series of residual blocks, and each residual block is divided into a direct mapping and a residual part. During the training process, because of the residual structure, the network does not have the problem of gradient disappearance, and the information can be passed smoothly between the higher and lower layers of the network.

Deformable Convolution. A problem in traditional image processing networks, convolutional neural networks, is that when pooling layers are used to reduce the image to increase the perceptual field or when up-sampling is used to increase the image size. By reducing and then increasing the size, some important information in the image will be lost. Dai proposed Deformable Convolution [15] in 2017. It adds displacement variables to the convolution or ROI sampling layer, and offsets different positions in the convolution, so that the convolution kernel can be scaled and changed to expand the range of the perceptual field and learn some global information in the image better.

III. Methodology

A. Data Preprocessing

Data. The Cervical Cancer cell data at different stages of development are from the Kaggle open source dataset, Multi Cancer Dataset, which contains images of eight different types of cancer cells, including Cervical Cancer, Brain Cancer, and others. Cervical Cancer contains five categories: cervix_dyk, cervix_koc, cervix_mep, cervix_pab, and cervix_sfi. Each category contains 5000 images. We split the dataset by using 70% of the data as training samples and 30% of the data as test samples and validation samples. Fig. 1 shows the example images in the dataset. The dataset can be downloaded from the following link https://www.kaggle.com/datasets/obulisainaren/multi-cancer.

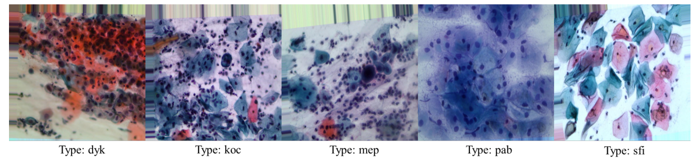


Fig. 1. **Example Images**. This figure shows example images of five kinds of Cervical Cancer cell data.

Data Augmentation. Image augmentation produces training samples that are similar but inconsistent with the original image by making a series of randomly changes to the image. We firstly crop all the data to obtain 224 x 224 size data. Then perform random up/down/left/right rotation, random position cropping and color change operations on the training dataset. Fig. 2 shows the results of the above operations. After this operation, we expand the training sample and reduce the sensitivity of the model to the target location.

B. Models

HQNet is a deep convolutional neural network. The beginning of the network uses three 3 x 3 convolutional kernels to extract the image features. Compared with other classical deep neural networks [6] that use a 7 x 7 convolutional kernel as the first convolutional layer to extract features, the number of parameters brought by three 3 x 3 kernels is much smaller than that of a 7 x 7 kernel. As shown in Table I, the number of parameters and computation of a 7 x 7 convolutional kernel is almost twice the number of parameters and computation of three 3 x 3 convolutional kernels when the number of input and output channels are both 1. And as shown in the Table II, when experimenting on an RTX 2080 GPU with channels of 8, 16, 32, 64, and 128, the time required for three 3 x 3 convolutional cores is also less than that of one 7 x 7 convolutional core. The middle part of the network uses a

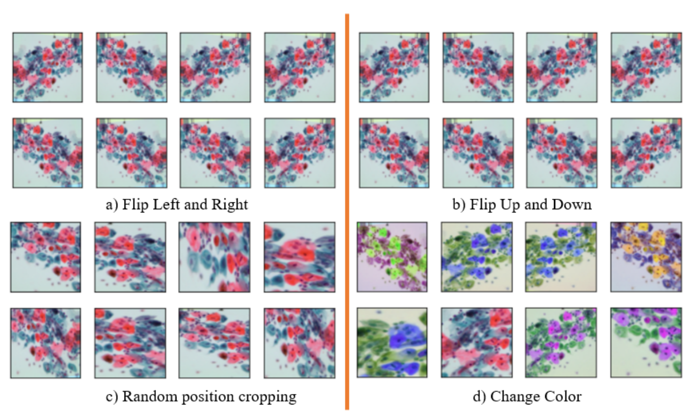


Fig. 2. **Data Augmentation**. a) The figure shows the result of flipping the image left and right. b) The figure shows the result of flipping the image up and down. c) The figure shows the result of cropping the image at a random position. d) The figure shows the result of changing the image color.

combination of four large convolutional layers, each of which contains 3, 4, 10, and 3 residual modules, respectively. Each residual module consists of two 1 x 1 convolutions and one 3 x 3 convolution and batch normalization. Each convolution is the deformable convolution [15]. Each large convolutional layer is connected by a GELU [16] activation function. The final part of the network uses Average Pooling, Dropout, and linear layers for the final classification. Fig. 4 shows the model of HQNet. Fig. 5 shows the partial HQNet convolutional layers parameters and the total parameters number.

TABLE I
PARAMETERS & COMPUTATION OF CHANNEL IS 1

Model	Parameters	Computation
One 7 x 7 conv	50.23	2.522
Three 3 x 3 conv	**27.85**	**1.394**

TABLE II
TIME OF DIFFERENT CHANNELS

Number of Channel	One 7 x 7 conv (s)	Three 1 x 1 conv (s)
8	0.834	0.572
16	0.042	0.061
32	0.076	0.072
64	0.244	0.142
128	1.783	0.436

Residual structure. Each residual structure consists of a combination of two 1 x 1 convolutions and a 3 x 3 convolution. If the residual structure is not added, the deep neural network training will result in network degradation. Then it is necessary to map the low-level network features to the high-level network. So the residual network is designed. The residual block can be expressed as

$$x_{l+1} = h(x_l) + F(x_l, W_l) \tag{1}$$

The residual block consists of two parts. The $h(x_l)$ is the direct mapping and the $F(x_l, W_l)$ residual part. When the dimensions of x_l and x_{l+1} are not the same, it is necessary to use 1 x 1 convolution to perform the dimension raising or lowering operation.

Deformable Convolution. The main idea of deformable convolution is to add displacement variables to the convolution or ROI sampling. Since traditional convolution kernels are fixed in size and step size, the sampled features do not represent the information of the whole graph very well. With deformable convolution, the convolution kernel has a displacement, and the convolution scales in every direction. This makes the field of perception of the convolution kernel larger, and it is no longer a square, but a polygon. The overall information in the graph can be better expressed, and there is a certain degree of improvement for the degree of information learning. As shown in the Fig. 3, adding a proper offset to the original convolution kernel becomes deformable convolution, which can improve the size of the perceptual field and extract the image features better.

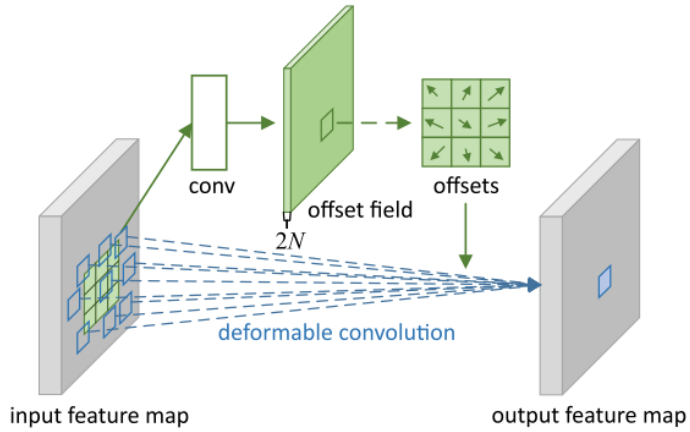


Fig. 3. **Deformable Convolution**. This figure shows Deformable Convolution.

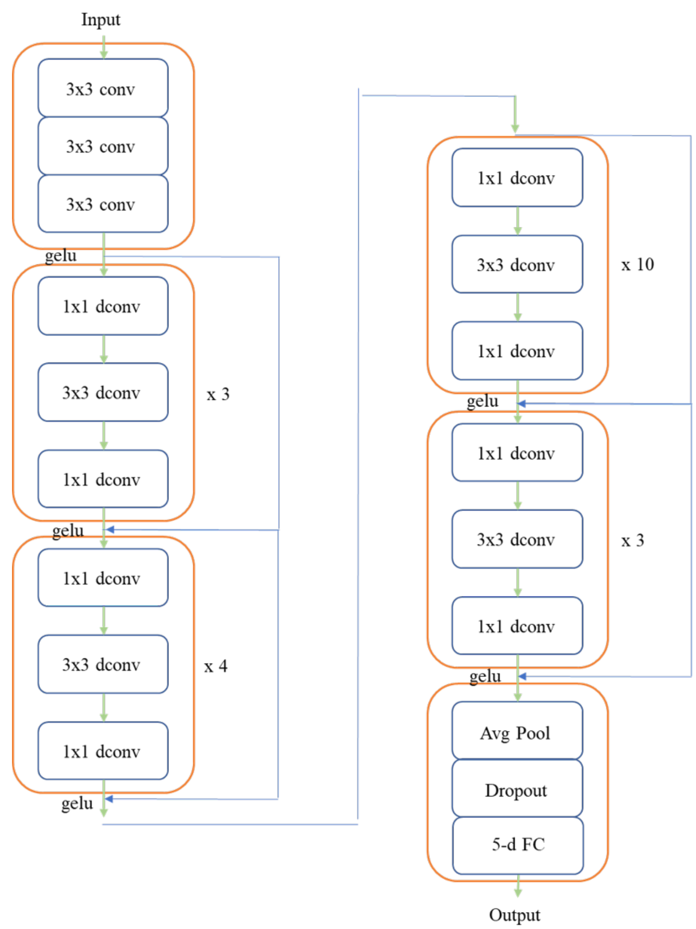


Fig. 4. **HQNet Model**. This figure shows HQNet Model.

Layer (type)	Output Shape	Param #
Conv2d-1	[-1, 64, 112, 112]	1,728
Conv2d-2	[-1, 64, 56, 56]	36,864
Conv2d-3	[-1, 64, 28, 28]	36,864
BatchNorm2d-4	[-1, 64, 28, 28]	128
GELU-5	[-1, 64, 28, 28]	0
MaxPool2d-6	[-1, 64, 14, 14]	0
ConvOffset2D-7	[-1, 64, 14, 14]	73,728
Conv2d-8	[-1, 64, 14, 14]	4,096
BatchNorm2d-9	[-1, 64, 14, 14]	128
GELU-10	[-1, 64, 14, 14]	0
ConvOffset2D-11	[-1, 64, 14, 14]	73,728
Conv2d-12	[-1, 64, 14, 14]	36,864
BatchNorm2d-13	[-1, 64, 14, 14]	128
GELU-14	[-1, 64, 14, 14]	0
ConvOffset2D-15	[-1, 64, 14, 14]	73,728
Conv2d-16	[-1, 256, 14, 14]	16,384
BatchNorm2d-17	[-1, 256, 14, 14]	512
ConvOffset2D-18	[-1, 64, 14, 14]	73,728
Conv2d-19	[-1, 256, 14, 14]	16,384
BatchNorm2d-20	[-1, 256, 14, 14]	512
GELU-21	[-1, 256, 14, 14]	0
Bottleneck-22	[-1, 256, 14, 14]	0
ConvOffset2D-23	[-1, 256, 14, 14]	1,179,648
Conv2d-24	[-1, 64, 14, 14]	16,384
BatchNorm2d-25	[-1, 64, 14, 14]	128
GELU-26	[-1, 64, 14, 14]	0
ConvOffset2D-27	[-1, 64, 14, 14]	73,728
Conv2d-28	[-1, 64, 14, 14]	36,864
BatchNorm2d-29	[-1, 64, 14, 14]	128
GELU-30	[-1, 64, 14, 14]	0

Total params: 469,831,237
Trainable params: 469,831,237
Non-trainable params: 0

Fig. 5. **Partial HQNet Convolutional Layers Parameters**. This image shows the parameters of some of the convolutional layers of HQNet. And the total parameters number of HQNet.

GELU Activation Function. The GELU activation function was chosen as the activation function for HQNet instead of the traditional activation functions such as tanh [1], Sigmoid [17] or ReLU [4]. The GELU combines the advantages of dropout, zoneout and ReLU. the mathematical expression of GELU is as follows

$$GELU(x) = xP(X <= x) = x\Phi(x) \tag{2}$$

where $\Phi(x)$ is a function of the normal distribution. the figure of the function of GELU is Fig. 6

C. Optimization

HQNet chose Nadam [18] as the optimizer instead of the traditional Adam [19] optimizer, which consists of two main components, the momentum component and the adaptive learning rate component. However, it was found that the regular momentum used by Adam is inferior to Nesterov's accelerated gradient (NAG) [20] in terms of performance and experience. the Nadam optimizer is the one that introduces the insights of the NAG algorithm into the momentum component

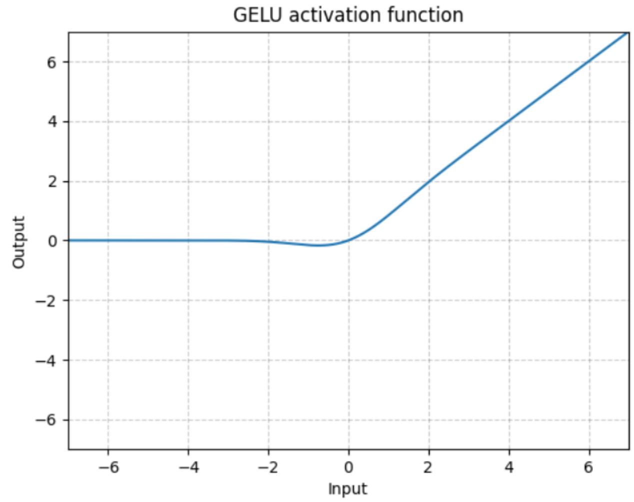


Fig. 6. **GELU Function**. This figure shows GELU activation function.

of the Adam optimizer. This improves the convergence speed and the quality of the learned model. Before showing changes to the Adam optimizer update rules, the following equation shows how to rewrite the NAG algorithm to make it easier to integrate the NAG algorithm with the Adam optimizer. The original step of calculating the network gradient using the momentum step to update the parameters is undone, entering with the original parameters in their original state and later using the momentum step during the actual gradient update. This means that the momentum step is used at the time t before the update instead of at t + 1:

$$\mathbf{g}_t \leftarrow \nabla_{\theta_{t-1}} f_t(\theta_{t-1}) \tag{3}$$

$$\mathbf{m}_t \leftarrow \mu_t \mathbf{m}_{t-1} + \alpha_t \mathbf{g}_t \tag{4}$$

$$\theta_t \leftarrow \theta_{t-1} - (\mu_{t+1}\mathbf{m}_t + \alpha_t \mathbf{g}_t) \tag{5}$$

where $\mathbf{g}_t$ is the current gradient, $f_t(\theta)$ is the Stochastic objective function which parameter is θ and the index is the time step t. $\mathbf{m}$ is the momentum vector at time t. The Eq.5 shows the update process.

After rewriting the NAG algorithm, the same improvement is used for the Adam optimizer, first rewriting the update step of the Adam optimizer using $\mathbf{m}_{t-1}$ and $\mathbf{g}_t$, and then replacing the current momentum step with the next one. This form of momentum can be combined with the adaptive learning rate to form the Nadam optimizer.

IV. Experiments

A. Experiment setup

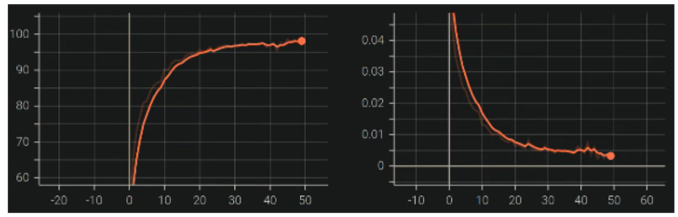
To test the performance of HQNet. First, we compare HQNet with some classic shallow neural networks such as AlexNet, VGG16 and VGG19. Then we compare HQNet with deeper and wider neural networks such as ResNet34, ResNet50, DenseNet, MobileNet. An RTX 2080 GPU was used for training until the network fully converged. The performance of HQNet was evaluated by recording the loss, accuracy, and other metrics of the different networks. To test the performance of HQNet. First, we compare HQNet with some classical shallow neural networks such as AlexNet, VGG16, VGG19, and Deformable Convolutional Neural Network. Then we compare HQNet with deeper and wider neural networks such as ResNet34, ResNet50, DenseNet [21], MobileNet [22]. An RTX 2080 GPU was used for training until the network fully converged. The performance of HQNet is evaluated by recording the loss, accuracy, and other metrics of the different networks.

B. Experiment Result

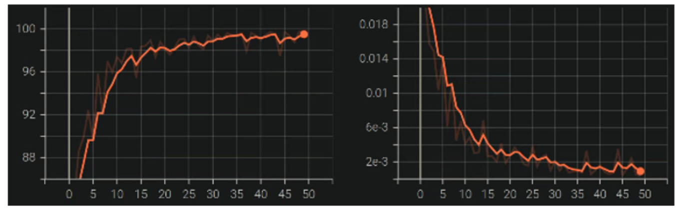
The Table III shows the correctness of the network in the Cervical Cancer cell classification task at different levels of development obtained in our comparison experiments. Figure 7, **??** shows the Accuracy and Loss curves of the different networks obtained during training. HQNet achieved **99.9%** correctness in the training set in this task, which is the best among all networks. The test set correct rate of **98.1%** is also the best among all networks. The performance of the sample neural network is generally slightly worse than that of the deep neural network. The VGG19 network, the deepest of the sample networks, performed relatively well, showing a 99.4% correct rate. In the comparison experiments of deep networks, it can be seen that when the number of layers of the network is smaller than HQNet, the performance of the network is not as good as HQNet. 99.6% is achieved by ResNet34 and 98.6% is achieved by ResNet50. The performance of MobileNet is almost similar to that of HQNet when the number of layers is higher than HQNet, but the training time is longer than that of HQNet, while the performance of DenseNet is not as good as that of HQNet. It is demonstrated that HQNet performs best in the task of classifying Cervical Cancer cells at different levels of development.

TABLE III
TRAINING & TEST ACCURACY COMPARISON

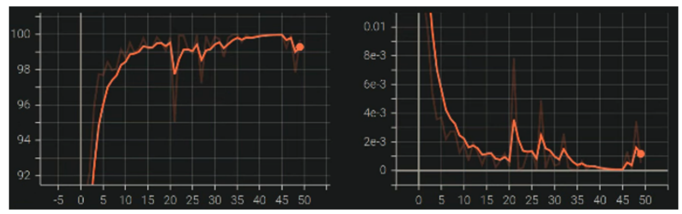
Name	Training Accuracy	Test Accuracy
AlexNet	98.4%	95.2%
VGG16	98.9%	95.9%
VGG19	99.4%	96.3%
ResNet34	99.6%	96.9%
ResNet50	98.6%	96.2%
DenseNet	99.5%	97.3%
MobileNet	99.8%	97.6%
HQNet	**99.9%**	**98.1%**



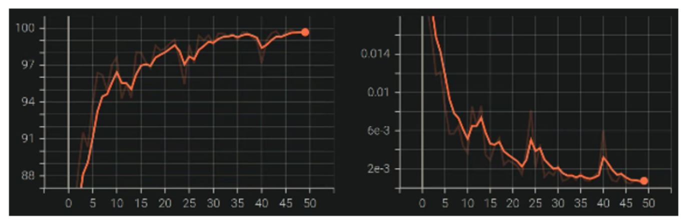
a) AlexNet Accuracy & Loss



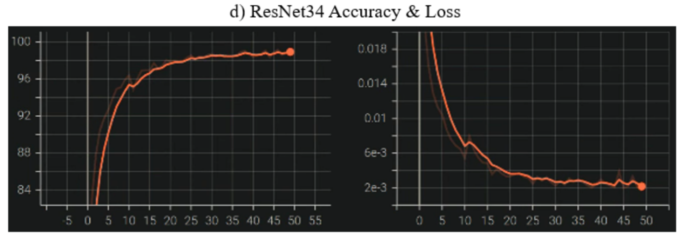
b) VGG16 Accuracy & Loss



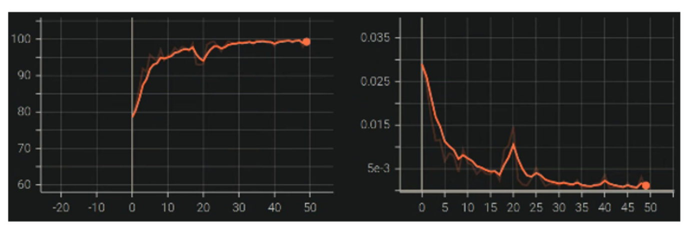
c) VGG19 Accuracy & Loss



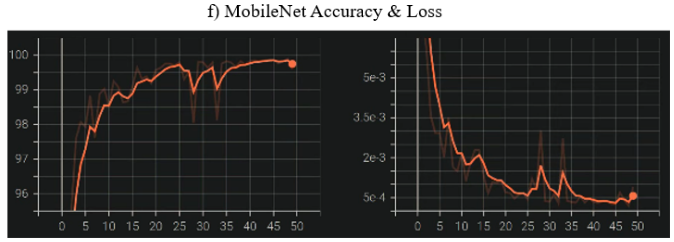
d) ResNet34 Accuracy & Loss



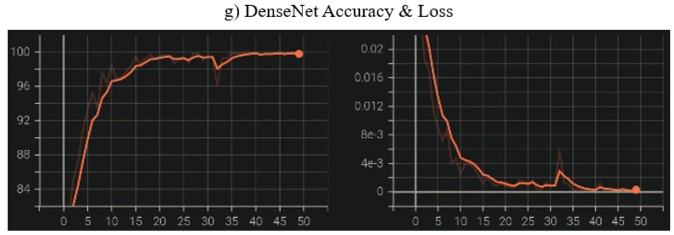
e) ResNet50 Accuracy & Loss



f) MobileNet Accuracy & Loss



g) DenseNet Accuracy & Loss



h) HQNet Accuracy & Loss

V. Conclusion

This paper presents HQNet, an Efficient Convolutional Neural Network for Cervical Cancer Classification, which achieves excellent results on the task of classifying Cervical Cancer cell images at different levels of development. HQNet is designed using a deep residual deformable convolutional neural network structure. Deformable convolution provides a more comprehensive view of the overall image features. GELU activation function helps the deep network to converge better. And use Nadam as the network optimizer to make the network converge better and improve the quality of the model by more flexible momentum update.

Acknowledgments

Han YANG would like to express his deepest gratitude to Prof. Teoh Teik Toe, the director of the AI lab in Nanyang Technological University, for his guidance and encouragement. Han YANG would also like to thank Ms. Qun HE for her great support to this article, without her continuous spiritual support, this article would not have been presented.

References

[1] Yann LeCun et al. Lenet-5, convolutional neural networks. *URL: http://yann. lecun. com/exdb/lenet*, 20(5):14, 2015.

[2] Ammarah Farooq, SyedMuhammad Anwar, Muhammad Awais, and Saad Rehman. A deep cnn based multi-class classification of alzheimer's disease using mri. In *2017 IEEE International Conference on Imaging systems and techniques (IST)*, pages 1–6. IEEE, 2017.

[3] S Deepak and PM Ameer. Brain tumor classification using deep cnn features via transfer learning. *Computers in biology and medicine*, 111:103345, 2019.

[4] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.

[5] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[7] Xue-jiao FENG, Hai-long HOU, Qiong YU, and Jun-shu WANG. Market analysis and countermeasures of cervical cancer vaccine in china. *China Biotechnology*, 40(11):96–101, 2020.

[8] Haibing Wu and Xiaodong Gu. Towards dropout training for convolutional neural networks. *Neural Networks*, 71:1–10, 2015.

[9] Abien Fred Agarap. Deep learning using rectified linear units (relu). *arXiv preprint arXiv:1803.08375*, 2018.

[10] Herve Jegou, Matthijs Douze, and Cordelia Schmid. Product quantization for nearest neighbor search. *IEEE transactions on pattern analysis and machine intelligence*, 33(1):117–128, 2010.

[11] Hervé Jégou, Florent Perronnin, Matthijs Douze, Jorge Sánchez, Patrick Pérez, and Cordelia Schmid. Aggregating local image descriptors into compact codes. *IEEE transactions on pattern analysis and machine intelligence*, 34(9):1704–1716, 2011.

[12] Florent Perronnin and Christopher Dance. Fisher kernels on visual vocabularies for image categorization. In *2007 IEEE conference on computer vision and pattern recognition*, pages 1–8. IEEE, 2007.

[13] Christopher M Bishop et al. *Neural networks for pattern recognition*. Oxford university press, 1995.

[14] Brian D Ripley. *Pattern recognition and neural networks*. Cambridge university press, 2007.

[15] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable convolutional networks. In *Proceedings of the IEEE international conference on computer vision*, pages 764–773, 2017.

[16] Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*, 2016.

[17] André C Marreiros, Jean Daunizeau, Stefan J Kiebel, and Karl J Friston. Population dynamics: variance and the sigmoid activation function. *Neuroimage*, 42(1):147–157, 2008.

[18] Timothy Dozat. Incorporating nesterov momentum into adam. 2016.

[19] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[20] Weijie Su, Stephen Boyd, and Emmanuel Candes. A differential equation for modeling nesterov's accelerated gradient method: theory and insights. *Advances in neural information processing systems*, 27, 2014.

[21] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[22] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.