Classification models for HEp-2 cell dataset (ICPR2012)

Weidi Luo*2035862 Bin Hu*2034729
Xi'an Jiaotong-Liverpool University, School of Advanced Technology
{weidi.luo20,bin.hu20}@student.xjtlu.edu.cn

Abstract

To improve the efficiency of HEp-2 cells classification and to reflect the significance of deep learning in medical applications, this report presents the process of developing two different classification models (based on CNN and Transformer respectively) on the HEp-2 cell dataset (ICPR 2012) and reached the average classification accuracy of 73.9% and 75.6%, which are close to the current state-of-the-art of 81.2%. The evaluation and comparison of the result of two models are also analyzed.

1 Background

As an important means to detect many diseases, indirect immunofluorescence (IIF) is widely used in medical examinations. Traditionally, the IIF sample would be classified under the microscope by professional examiners, hence the costs for time and labor are high. With the development of computer vision in the past decade, different models are carried out to classify IIF images. To have a uniform criterion on these models' performance, HEp-2 cell dataset (ICPR 2012) was released in 2012, containing 1455 images of Hep-2 cells, which are typical as IIF images.

Many classification models have been developed to fit the HEp-2 cell dataset since its release. Most of them are based on CNN and adopted two main stages of feature extraction and classification. Data augmentation is widely applied and the train-valid splitting ratio is usually 5:5 in the previous works, the reason should be the small size of the dataset for training. Average classification accuracy is the widely accepted criteria for the model's performance(10), and the current state-of-the-art was achieved by Liu et al(6). of 81.2% ACA with a deep autoencoding-classification network in 2017.

2 Proposed Method

In this section, the basic information of the dataset will be introduced, and the proposed pre-processing methods and networks will be explained.

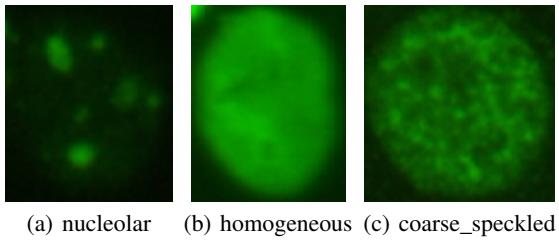
2.1 Data Pre-processing

The original dataset contains 6 patterns of Hep-2 cells (categories) as Figure 1 indicates and is originally divided into training and testing groups (721 and 734 images for each).The pixels and proportions of the images are inconsistent and the distribution of the patterns are not equal as Figure 2 shows. All the images come with a label indicating the pattern stored in two csv files and a black-white mask image which could help to remove the background of the raw image.

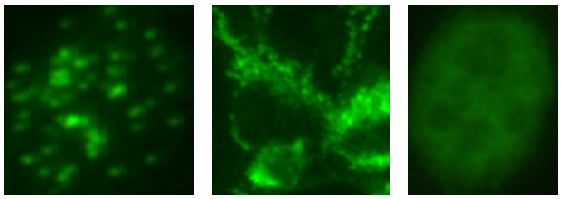
Since the mask is given for images both in the train set and the test set, its role in highlighting the main body of the cell should be important for accuracy increment. The small size of the sample and

* Equal Contribution

the unequal distribution of patterns in the train set are expected to cause overfitting when training, hence data augmentation techniques should be necessary for training.



(a) nucleolar (b) homogeneous (c) coarse_speckled



(d) centromere (e) cytoplasmatic (f) fine_speckled

Figure 1: Example of Patterns

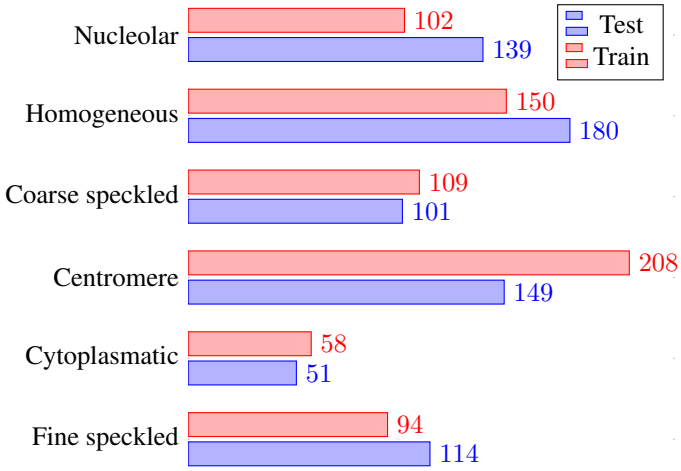


Figure 2: Distribution of Patterns

2.2 Network

The two team members selected networks based on CNN and Transformer(11) respectively. The networks are ConvNeXt (CNN) and Custom CoAtNet (Transformer).

ConvNeXt: ConvNeXt(8) is a pure convolutional network developed by FaceBook and UC Berkeley in 2022. It consists entirely of convolutional blocks like ResNet but borrowed some strategy from Swin-Transformer(7), such as fewer activation layers and separate downsampling layers, which granted it better performance than ResNet(5) and even than Swin-Transformer. Pretrained with ImageNet-22K dataset, ConvNeXt-T achieved an accuracy of 82.1% on ImageNet, exceeding the Swin Transformer (Swin-T, 81.3%) on the similar parameter scale. In the following experiments, the specific network that will be applied is ConvNeXt-B.

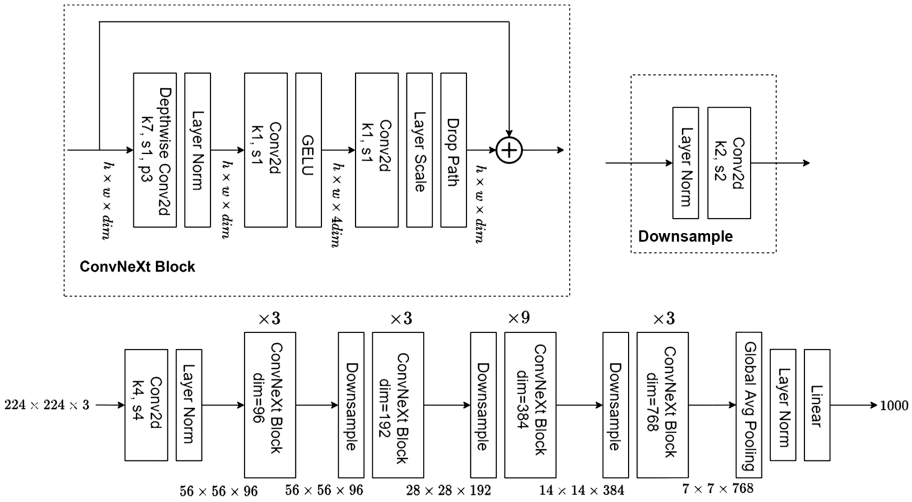


Figure 3: Structure of ConvNext

Custom CoAtNet: CoAtNet(3) is a special transformer network developed by Google in 2021, which adopts both convolution and self-attention layers via simple relative attention. Pretrained with JFT-3B dataset, it could achieve 90.88% accuracy on ImageNet. Benefitting from the adoption of convolutional layers, it has greater generalization performance over other Transformer networks on small datasets, which makes it suitable for the current dataset. Due to the scale of parameters of the original network is large, a simplified version provided by timm package is applied. The parameters could be referred in timm's GitHub repository as the name of "coat-mini".

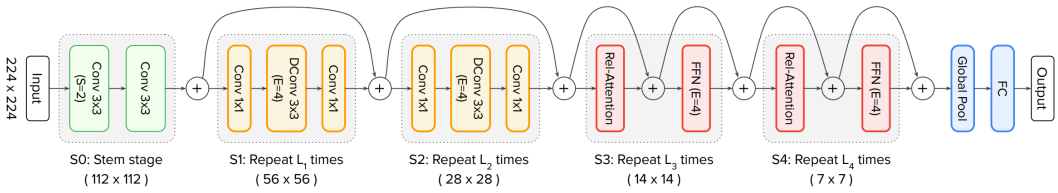


Figure 4: Structure of CoAtNet(3)

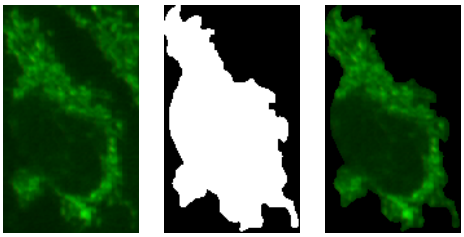
3 Plan & Experiment

In this section, the detailed plan of the experiment would be described.

3.1 Datasets

3.1.1 Mask processing

With the mask, the main body of the Hep-2 cells can be easily extracted and the background can be removed. In most examinations, models trained and tested on masked images could achieve 0.3%-0.5% higher accuracy than those on unmasked images under the same configuration, hence the two models obtained in this report are all based on masked image sets. When adding the mask, whether each pixel's color of the raw image is preserved depends on the corresponding mask pixel's color. If the corresponding mask pixel is white, the pixel's color in the raw image remains, or the pixel color is set to black (RGB=0). The process is illustrated in Figure 5.



(a) raw image (b) mask (c) after

Figure 5: Process of adding mask

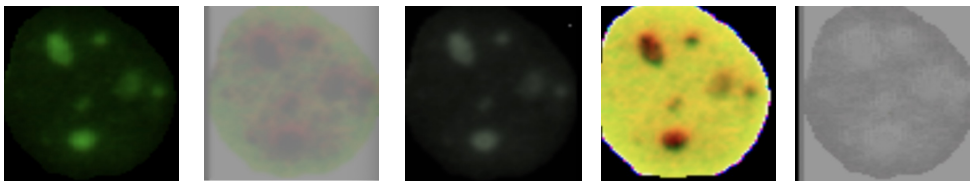
3.1.2 Data augmentation

As a common means to improve the performance of classification models, data augmentation can effectively improve the generalization ability of the model (reducing over-fitting), which is crucial especially for training sets with a small number of samples. After several trials, both models applied a combination of data augmentation methods which could achieve best result. The detailed approaches are list in Table 1.

	ConvNeXt-B ImageNet 22K 224^2	**Custom CoAtNet JFT-3B 224^2**
RomdomErasing	None	p=0.5, scale = (0.02,0.33), ratio=(0.3,3.3)
RomdomHorizontal	p=0.5	p=0.5
RandomVerticalFilp	p=0.5	p=0.5
Colorjitter	None	brightness =(0.3,1.5), contrast=0.5, saturation=0.3, hue=0,01
Normalized	(0.5,0.25)	(0.5,0.25)
Cutout(4)	YES	None
RadomResizedCrop	None	scale=(0.25,1.0), ratio=(0.75,1.33)
RandomRotation	None	None
Rand Augment	YES	None

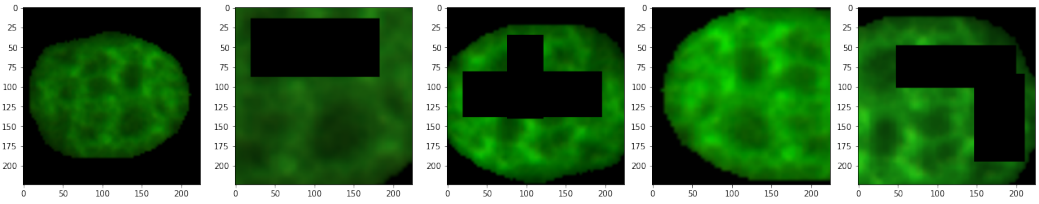
Table 1: Data augmentation methods

The example of image transform in data augmentation is listed in Figure 6 and Figure 7. The listed data augmentation steps take place when fetching batches and only affects the images in the train set. The images in the valid set and test set will only be normalized in the same way as those in the train set.



(a) before (b) after (c) after (d) after (e) after

Figure 6: Example of augmented images in ConvNeXt pathway



(a) before (b) after (c) after (d) after (e) after

Figure 7: Example of augmented images in CoAtNet pathway

3.2 Training configures

The dataset split is necessary since the test set should be kept invisible before the end of training, and the configure of optimization methods plays a key role in the models' loss convergence when training. The original train set was split to train and valid set on the ratio of 5:5. The weights of the two networks are initialized with pre-trained models. In each epoch, the model with the highest accuracy on the valid set will be saved. If the accuracy is the same, the model with lower loss will be saved. Adam and AdamP are picked as the optimizers of the two models respectively. Cosine learning rate schedulers are used by two models. Besides, learning rate warm-up, Cutmix(12), Mixup(13), Randaugment(2), and Label Smoothing(9) are used to reduce overfitting, which is important for datasets in small size. Test Time Augmentation(1) is also used in the ConvNeXt model to improve accuracy. The detailed training configures are listed in Table 2.

	ConvNeXt-B ImageNet-22K 224^2	**Custom CoAtNet JFT-3B 224^2**
Dataset split	5:5 predefined	5:5 predefined
Optimizer	Adam	AdamP
Loss function	Soft Target Cross Entropy	Binary Cross Entropy
Base learning rate	4e-6	1e-5
Weight decay	0	0.1
Optimizer momentum	$\beta1, \beta2 = 0.9, 0.999$	$\beta1, \beta2 = 0.9, 0.999$
Batch size	32	16
Training epochs	30	50
Learning rate schedule	cosine decay	cosine decay
Warm-up epochs	0	10
Warm-up schedule	None	linear
Label smoothing	0.1	None
Mixup	0.8	None
Cutmix	1.0	None
TTA	YES	None

Table 2: Training configures of models

4 Results and Evaluation

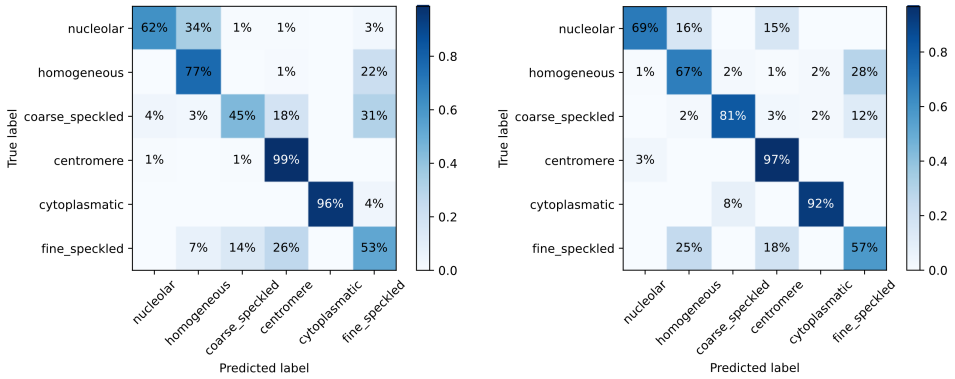
In this section, the result of 2 classification models on the test set and the critical evaluations will be explained. Possible way to improve the models will also be discussed.

4.1 Classification Result

F1-scores and average accuracy are widely accepted to evaluate the results of classification models. The result of the models is shown in Table 3. Confusion Matrix is also an important method to visualize classification results of supervised learning models. The F1-scores and average accuracy of the 2 models are listed in Table 3, and the confusion matrices are illustrated in Figure 8.

	ConvNeXt-B ImageNet-22K 224^2	**Custom CoAtNet JFT-3B 224^2**
f1 score for nucleolar	0.78	0.80
f1 score for homogeneous	0.70	0.69
f1 score for coarse_speckled	0.75	0.86
f1 score for centromere	0.88	0.85
f1 score for cytoplasmatic	0.98	0.91
f1 score for fine_speckled	0.47	0.54
Average classification accuracy	0.739	0.756

Table 3: F1 scores and average accuracy of the classification results



(a) ConvNeXt-B (b) Custom CoAtNet

Figure 8: Confusion Matrices of the classification results

4.2 Evaluation and Discussion

Regarding the results, there exists a slight gap between ConvNeXt-B and the custom CoAtNet on their performance on the HEp-2 cell dataset (ICPR2012). The custom CoAtNet model is 2% higher than the CovNeXt-B model on average accuracy. It could be attributed to the self-attention mechanism that the CoAtNet adopts, which can have a better performance by pre-trained and rigorous data augmentation(8). As a representative of pure convolutional networks, CovNeXt-B's performance exceeds many CNN models developed in previous works on the same dataset, which also shows the potential of the convolutional networks.

Besides, it is believed that these two models have much potential for improvement on this dataset since the pre-trained models adopted to initialize the weights are trained with photographs of common objects in daily life (ImageNet and JFT-3B) instead of IIF images under the microscope, which makes the extracted feathers could not fit the HEp-2 images well. One possible way to improve the performance is to initialize the weights with a model pre-trained with large datasets of IIF images. For example, the Hep-2 cell dataset (ICPR2014) could be a good candidate.

Looking into the F1 score and the confusion matrices, the two models' accuracy is similarly low on 'nucleolar' and 'fine_speckled', but similarly high on 'centromere' and 'cytoplasmatic'. Since the latter two categories respectively have the most and the least images in the train set, and the number of images in the remaining four categories is relatively close, the reason for the low accuracy on 'nucleolar' and 'fine_speckled' should simply be the small size of the training set instead of unequal distribution of images in categories.

For further experiments, suggestions are: Using a pre-trained model with similar datasets, freezing the feature extraction layers and applying fine-tuning with lower learning rate, applying learning rate annealing. Undersampling and oversampling may help to unify the distribution of data, though it

should not the main problem as explained above. For more advanced training methods, the radius of the cell could even be considered as a feature to learn.

5 Conclusions

The Hep-2 cell dataset (ICPR2012) is a small dataset of IIF images which could be a criterion for the accuracy of classification models and could contribute to the development of computer vision on medical imaging analysis. Data augmentation plays a key role in the training process on this dataset. Though the accuracy of the two models is 73.9% (ConvNeXt-B) and 75.6% (Custom CoAtNet) which are lower than the current state-of-the-art (81.2%), the results are still excellent since the data pre-processing for the two models is relatively simpler than these for the state-of-the-art. The close accuracy between the two models has proofed that convolutional networks are still competitive in computer vision. The two models are expected to achieve higher accuracy if the proposed methods given in the discussion section are applied.

References

[1] AYHAN, M. S., AND BERENS, P. Test-time data augmentation for estimation of heteroscedastic aleatoric uncertainty in deep neural networks.

[2] CUBUK, E. D., ZOPH, B., SHLENS, J., AND LE, Q. V. Randaugment: Practical data augmentation with no separate search. *CoRR abs/1909.13719* (2019).

[3] DAI, Z., LIU, H., LE, Q., AND TAN, M. Coatnet: Marrying convolution and attention for all data sizes. *Advances in Neural Information Processing Systems 34* (2021).

[4] DEVRIES, T., AND TAYLOR, G. W. Improved regularization of convolutional neural networks with cutout. *CoRR abs/1708.04552* (2017).

[5] HE, K., ZHANG, X., REN, S., AND SUN, J. Deep residual learning for image recognition. *CoRR abs/1512.03385* (2015).

[6] KASTANIOTIS, D., FOTOPOULOU, F., THEODORAKOPOULOS, I., ECONOMOU, G., AND FOTOPOULOS, S. Hep-2 cell classification with vector of hierarchically aggregated residuals. *Pattern Recognition 65* (2017), 47–57.

[7] LIU, Z., LIN, Y., CAO, Y., HU, H., WEI, Y., ZHANG, Z., LIN, S., AND GUO, B. Swin transformer: Hierarchical vision transformer using shifted windows. *CoRR abs/2103.14030* (2021).

[8] LIU, Z., MAO, H., WU, C.-Y., FEICHTENHOFER, C., DARRELL, T., AND XIE, S. A ConvNet for the 2020s. *arXiv e-prints* (Jan. 2022), arXiv:2201.03545.

[9] MÜLLER, R., KORNBLITH, S., AND HINTON, G. E. When does label smoothing help? *CoRR abs/1906.02629* (2019).

[10] RAHMAN, S., WANG, L., SUN, C., AND ZHOU, L. Deep learning based hep-2 image classification: A comprehensive review. *Medical Image Analysis 65* (2020), 101764.

[11] VASWANI, A., SHAZEER, N., PARMAR, N., USZKOREIT, J., JONES, L., GOMEZ, A. N., KAISER, L., AND POLOSUKHIN, I. Attention is all you need. *CoRR abs/1706.03762* (2017).

[12] YUN, S., HAN, D., OH, S. J., CHUN, S., CHOE, J., AND YOO, Y. Cutmix: Regularization strategy to train strong classifiers with localizable features. *CoRR abs/1905.04899* (2019).

[13] ZHANG, H., CISSÉ, M., DAUPHIN, Y. N., AND LOPEZ-PAZ, D. mixup: Beyond empirical risk minimization. *CoRR abs/1710.09412* (2017).

A Appendix: Whole Train Set for Training

Another way of training is using the whole train set for training and the test set for validation. In this way, the accuracy achieved would usually be higher, but would be considered to be tricky. The result of training the Custom CoAtNet in this way is shown in the Table 4 and Figure 9.

f1 score for nucleolar	0.86
f1 score for homogeneous	0.78
f1 score for coarse_speckled	0.81
f1 score for centromere	0.93
f1 score for cytoplasmatic	0.86
f1 score for fine_speckled	0.52
Average classification accuracy	0.802

Table 4: F1 scores and average accuracy of the classification result

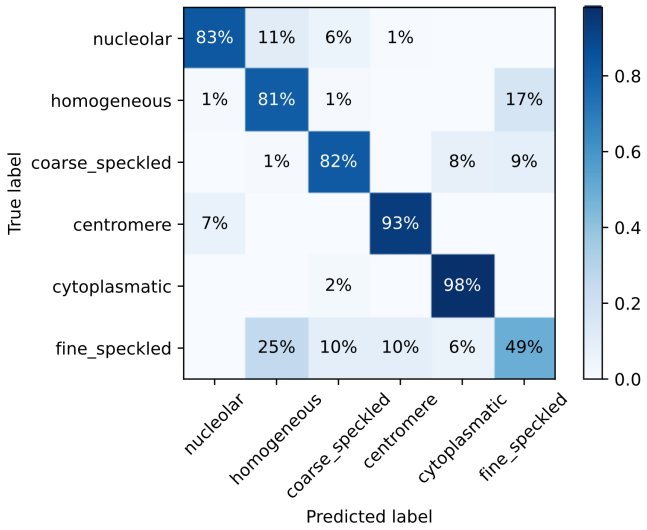


Figure 9: Confusion Matrix of the classification result