Mixture Model-Based Approach for Accurate Finance Table Image Classification

Ho- Jung Kim[1], Yeong- Eun Jeon[1], Won- Seok Jung[2],
Dae- Hyun Bae[2], Yung- Il Park[3] and Dong- Ok Won[1*]

***Abstract*— Financial analysis reports contain a wealth of information, including details about a company's assets and performance, as well as analyst assessments. For subsequent analysis and valuation, information regarding a company's assets and performance is crucial. As a result, numerous securities firms seek to leverage artificial intelligence technology to profit from this data. We intend to resolve the issue of extracting and classifying table data from financial analysis reports by converting them to images. Financial report table images contain less information that can be extracted than typical images used in image classification research, and their sizes vary, making classification difficult with simple models. As a result, we tested a variety of classification models to determine which ones performed the best, and we propose a new model that combines their strengths. This new model achieved 94.5% accuracy and an F1-Score of 0.9414 on our self-collected financial analysis table image dataset, demonstrating its effectiveness in classifying table images in financial analysis reports.**

I. INTRODUCTION

The application of artificial intelligence technology has expanded from simple learning products and basic commercial services to numerous fields and larger-scale services. Finance is a field that is actively adopting and developing AI technology. In the finance industry, where financial analysis reports consisting of images are readily available data, AI technology is used to extract and classify specific data areas. This is due to the fact that accurate classification of existing or new data is essential for the efficient operation of business applications. With the emergence of diverse challenges based on the ImageNet dataset[1], image classification has advanced relatively quickly in comparison to other fields, resulting in rapid progress. Subsequently, the discovery of the significant impact of data relationships on classification through Attention[2] mechanisms paved the way for the development of Transformers[3], which take the form of encoder and decoder architectures, significantly advancing AI technology. Transformers have significant drawbacks, including a high memory requirement for calculating multi-head attention and the need for a large quantity of training data. To address these problems, Vision Transformers[4] have been proposed for image-related tasks, despite the fact that they are not widely used and are primarily employed for experimental research. Therefore, in the field of image classification, several models have been proposed that adapt the technical concepts utilized in Transformers to image data. Similarly, this paper proposes a network for classifying table images in financial analysis reports without the use of Transformers, based on technological advances from previous studies. Financial tables are of varying sizes and shapes, and the image data contains limited information, making classification difficult. The proposed model intends to overcome these constraints and effectively address the issue at hand.

* Dong- Ok Won is corresponding author
[1] Department of Artificial Intelligence Convergence, Hallym University, Chuncheon, Gangwondo, Republic of Korea dongok.won@hallym.ac.kr
[2] Department of Technology research and development, Yonhap Infomax Inc., Seoul, Republic of Korea rca32@yna.co.kr
[3] Department of Contents planning, Yonhap Infomax Inc., Seoul, Republic of Korea ylpark@yna.co.kr

II. METHOD

A. Proposed model for Finance table image classification

Combining the forms of two existing models, the proposed model for classifying table images in financial analysis reports is a new model. The utilized models, HRNet[6] and DenseNet[7], generate feature maps with distinct differences. Therefore, these differences complement one another, resulting in enhanced performance. Fig. 1. demonstrates that, similar to HRNet, the model is divided into multiple stages, with each stage consisting of subnetworks. Through transition layers, the results of each subnetwork are merged and forwarded to the subsequent stage. By configuring distinct input resolutions for each subnetwork, it is possible to generate more precise feature maps. In addition, by using DenseBlocks as the primary blocks, the model increases the feature map density. In addition, prior to classification, UpConvolution is used to increase the size of the feature maps, similar to the structure of a Fully Convolutional Network[8], which aids in precise classification. In addition, weighted addition was implemented to improve the utility of the most abstract feature maps.

B. Training and Validation Dataset

We made efforts to locate diverse data for Finance table image classification, but the majority of available datasets were geared toward resolving detection or structure recognition issues. In order to generate data suitable for addressing our specific issue, we utilized actual financial analysis reports issued by major Korean securities firms. More than 1,400 self-collected financial analysis reports were used in the study, from which approximately 11,872 pages were extracted to obtain table images. There were a total of 34,202 extracted table images, of which 25,873 were used for training and evaluation. The data split for training and

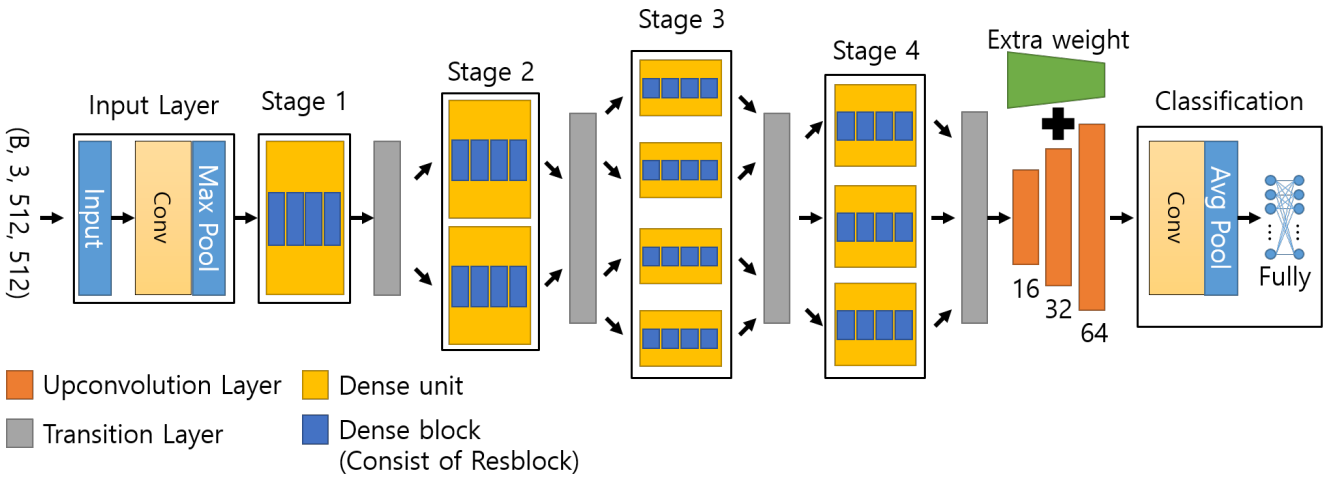

Fig. 1. Overview of the proposed model, Each stage is divided into subnetworks that generate feature maps, and the transition layer's feature maps are integrated. Each subnetwork in each stage is made up of Dense Blocks, and the number of blocks (Depth) can be altered to create a network that is either deeper or shallower

evaluation was conducted with a ratio of 7:3. There are a total of 18 classes, and there is a significant imbalance in the number of data samples across the classes, resulting in a data imbalance issue. During the training process, an Imbalanced Dataset Sampler was utilized to address this issue.

III. RESULT AND CONCLUSION

In order to evaluate the proposed model, it was compared to other models. The accuracy, precision, recall, and F1-score for each model in our dataset are presented in Tab. I. Two comparative foundational models, DenseNet and HRNet, served as the basis for the proposed model. In addition, models that apply attention to image data, such as Squeeze-Excitation Net (SENet)[9] and its enhanced version EfficientNet[10], were used in the comparative analysis. The results demonstrate that HRNet and the other two models, SENet and EfficientNet, perform comparably. However, DenseNet exhibits comparatively superior performance. In spite of this, the proposed model achieves an accuracy of 94% and other evaluation metrics above 0.94, indicating a substantial performance advantage over other models and demonstrated minimal score differences in other evaluation metrics, demonstrating its stability.

In conclusion, we have proposed a new classification model for classifying table images in financial analysis reports accurately. In solving the problem of table image classification, the model has demonstrated high accuracy and stability.

ACKNOWLEDGMENT

This work was supported by Institute of Information and Communications Technology Planning and Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2021-0-02068, Artificial Intelligence Innovation Hub) and partly supported by the Korean National Police Agency (Project Name:AI-Based Crime Investigation Support System/Project Number: PR10-02-000-21)

TABLE I
EVALUATION RESULTS OF EACH MODEL FOR TABLE CLASSIFICATION

Model	Accuracy	Precision	Recall	F1
HR-Net	0.843	0.851	0.828	0.839
SE-Net	0.855	0.859	0.831	0.826
EfficientNet-b2	0.872	0.877	0.857	0.8505
Dense Net	0.89	0.882	0.866	0.865
Proposed model	**0.945**	**0.9424**	**0.9405**	**0.9414**

REFERENCES

[1] J. Deng, W. Dong, R. Socher, L. -J. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248-255

[2] Luong, M. T., Pham, H., and Manning, C. D., "Effective Approaches to Attention-based Neural Machine Translation," Conference on Empirical Methods in Natural Language Processing, 2015, pp. 1412–1421

[3] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, L. Kaiser, and I. Polosukhin, "Attention Is All You Need," Advances in Neural Information Processing Systems, 2017, pp. 5998-6008

[4] A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in 9th Int. Conference on Learning Representations, 2021

[5] Y. Cao, J. Xu, S. Lin, F. Wei, and H. Hu, "GCNet: Non-Local Networks Meet Squeeze-Excitation Networks and Beyond," in Proc. of the IEEE/CVF International Conference on Computer Vision, 2019, Workshops.

[6] K. Sun, B. Xiao, D. Liu, J. Wang, "Deep High-Resolution Representation Learning for Human Pose Estimation," Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 5693-5703

[7] G. Huang, Z. Liu, K. Q. Weinberger, and L. Maaten. "Densely connected convolutional networks," Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700-4708

[8] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.

[9] J. Hu, L. Shen and G. Sun, "Squeeze-and-Excitation Networks," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132-7141

[10] M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in Proc. of the 36th International Conference on Machine Learning, 2019, vol. 97, pp. 6105–6114