SpindleNet: Open-set Chinese Historical Text Recognition via Enhanced Component Representation

First Author[1][0000−1111−2222−3333], Second Author[2,3][1111−2222−3333−4444], and Third Author[3][2222−−3333−4444−5555]

[1] Princeton University, Princeton NJ 08544, USA
[2] Springer Heidelberg, Tiergartenstr. 17, 69121 Heidelberg, Germany
lncs@springer.com
http://www.springer.com/gp/computer-science/lncs
[3] ABC Institute, Rupert-Karls-University Heidelberg, Heidelberg, Germany
{abc,lncs}@uni-heidelberg.de

Abstract. Open-set text recognition remains a persistent challenge in computer vision tasks, particularly in the context of recognizing Chinese historical texts. An important factor causing this problem is that there are many categories of ancient Chinese characters, and they have more significant long-tail distribution characteristics. This severe data imbalance results in a lack of samples for tail classes, often leading to poor performance in recognizing new characters, as new characters are distributed in the tail with a high probability. Consequently, the tail class exhibits a strong correlation with new characters, and mitigating the long tail issue can enhance the recognition performance of these new characters.Chinese characters exhibit structural similarities in their local components (such as parts and strokes), particularly between the head and tail classes. Also, each tail class typically consists of several local parts from various head classes. Thus, enhancing the representation ability of these local parts can improve the performance of both head and tail classes. Based on the above insights, we propose a Character Components Enhanced Spindle Network named Spindle-Net, which improves the ability to represent local parts by increasing the channel numbers of middle layers that model part-level-features. In order to reduce the impact of the number of parameters, we keep the total number of parameters of the model unchanged as much as possible. As a result, compared to the baseline, the network correspondingly reduces deeper layers, yielding a spindle shape. Compared with the mainstream model structure, the spindle network can significantly improve the feature extraction capability, thereby improving the recognition accuracy of tail category characters. Extensive experiments on three challenging Chinese ancient book datasets (TKH, MTH1000, and MTH1200) verify that our method achieves state-of-the-art performances.

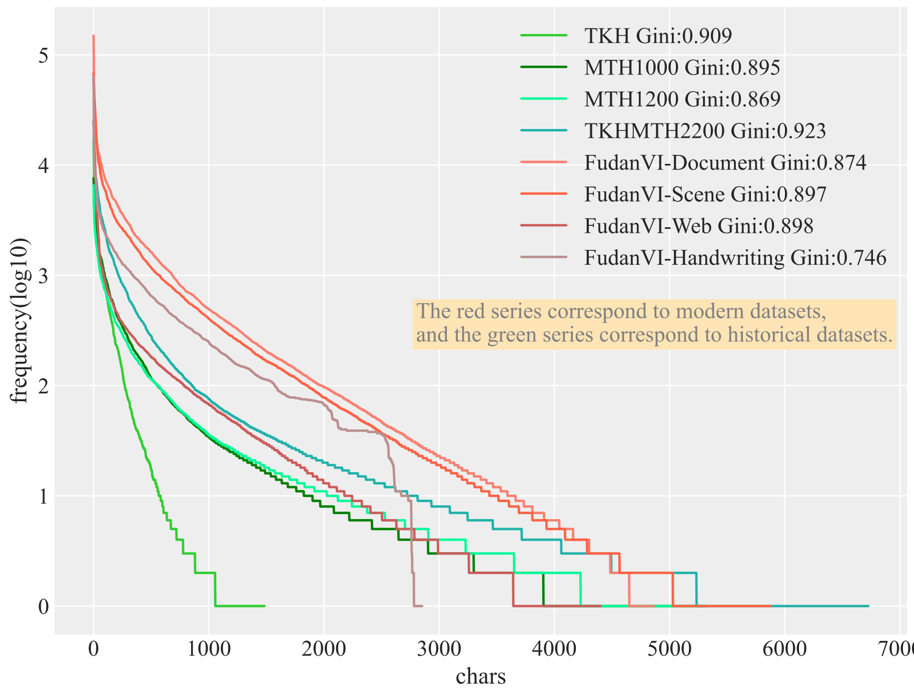

Fig. 1. Comparison on Gini Coefficient [36] between modern text recognition datasets [8] and historical text recognition datasets [35].

1 Introduction

For a long time, China has left behind a large number of historical documents, which have very important academic and artistic value. Therefore, in recent years, the study of historical documents has received widespread attention from researchers [14,3,13].

Different from other text recognition tasks, historical document recognition tasks face unique challenges like complexity and damage to the characters in historical documents, including stains, tears, and ink bleeding. In addition to the complexity of text recognition of historical documents, another major problem comes from the data itself. Specifically, history document recognition suffers from the long-tailed character occurrence distribution [10]. The long-tailed distribution also affects other text recognition tasks like [8], however, historical document recognition suffers a larger “longtailness” measured by the Gini Coefficient [36](See Figure 1). Furthermore, new characters in the test data tend to be distributed in the tail[14], making it a hybrid of open-set recognition and long-tail problems. To address this challenge, existing methods propose to exploit the radical information of each character [30], where individual radicals are often used in both head and tail (including unseen) classes, which are shown to generalize well in recognizing the novel characters [29,39]. On the other hand, Zhang et. al. [40] proved that exploiting the component information can improve the performance of tail classes. However, such methods depend on radical-level annotations to train, yielding expensive annotation costs to deploy.

In this work, we propose to break free from the radical annotation by adopting a visual matching approach [19]. To keep exploiting the similarity of character components, we propose implicitly emphasizing detail feature modeling. Specifically, we propose the spindle backbone network, which increases the number of parameters to layers corresponding to component features [12]. We argue character parts are more similar to texture patterns, which are more modeled in middle layers (conv2 and conv3) according to [2]. On the other hand, high-frequency signals are reported harmful to the generalization [28,41], hence we refrain from making shallow layers wider. The network also reduces the parameters in deeper layers to keep the total parameter to keep a small VRAM footprint and high inference speed.Summarizing the above motivations, we propose a spindle network that has narrower shallow and deep layers but a wider middle layer.

The results show that this design effectively improves the model performance of tail and head classes, while the recognition performance of new characters is also improved. We also conducted architectural ablative experiments, which verified that the spindle design is better than the usual pyramid design and the reverse pyramid design in terms of performance, justifying our motivation.As a structural-knowledge-free approach, the proposed method also possesses decent recognition capability on novel classes, which can reduce the efforts needed for adapting the model for new excavations. In addition, the approach also helps improve the head classes as well.

In summary, the main contributions of this paper can be considered as follows:

- We found that the similarity of character components can be exploited to improve the recognition performance of new characters in Chinese historical books.
- We implemented a spindle network to enhance the ability to extract character component features, leveraging the similarities between character components to improve the recognition performance of new characters.
- We conduct extensive experiments on three challenging Chinese ancient book datasets (TKH, MTH1000, and MTH1200) to validate the superiority of our proposed method. The results show that our approach achieves state-of-the-art performance in this field.

2 Related Works

In this section, we review previous works on layout analysis and text recognition. For document text recognition, we focus on character-based methods.

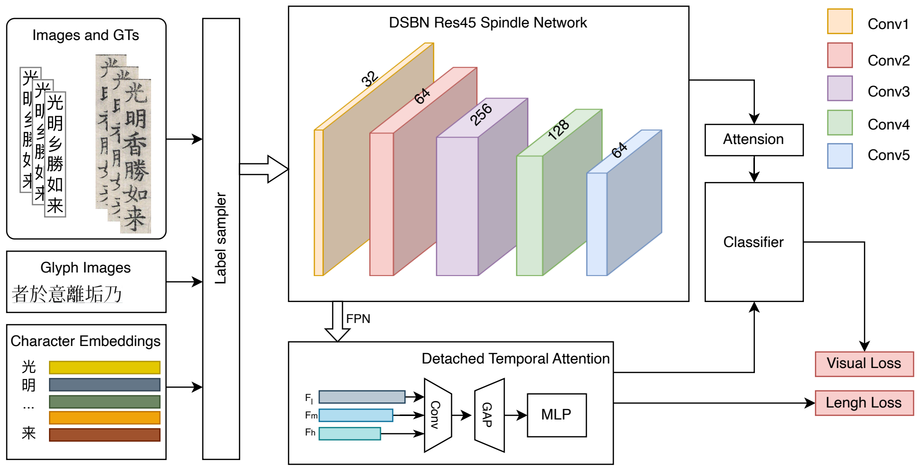

Fig. 2. Overview framework.The architecture consists of three modules: a feature extraction module, a character length and weight prediction module, and a decoder module.

2.1 Historical Text Recognition

Document digitization systems protect printed paper documents from direct manipulation and facilitate consultation, exchange, and remote access. Specifically, text recognition is one of its two main stages together with layout analysis [21].

Historical text recognition methods can be divided into character-based methods and sequence-based methods. Character-based recognition methods typically involve locating individual characters, recognizing them, and grouping them into lines of text [27]. Among the three steps, the single character recognition step is mostly researched, because it faces challenging problems including broken character [1], wild writing styles [13], large class numbers with long-tailed distribution [29], or on the extreme end novel characters that are not covered by the training samples [3,16].

Due to character-level annotations are usually more expensive to obtain, sequence-based methods, which train on line-level images and annotations are proposed [25,14]. Still, they face similar challenges posed to character-level counterparts. In this work, we focus on the long-tailed challenge in the historical text recognition tasks.

2.2 The Long-Tailed Distribution Problem

In real life, data, specifically training data, often have imbalanced occurrence frequency for different labels, whose distributions exhibit broader characteristics than the standard positive land distribution, called long-tail distributions [36]. Specifically, a small number of individuals make significant contributions, resulting in the minority class dominating the data set (called the head class), while the majority class contains only a few data samples (called the tail class). General solutions can be roughly categorized into data-side solutions, optimization-side solutions, and model-side solutions.

Data side like class-balancing sampling [24], data augmentation [38], and data synthetic [6,26]. Optimization-side refers to methods focusing on loss designs or training procedures. Specifically, loss designs can be further categorized into instance-based reweighting [18,15], and regularization terms that enforce priors [8,40,32]. The model-side solution involves alleviating via model design, including ensembling [34,31], classifier modification [33], etc.

Noteworthy, zero-shot learning [23], as an extreme case of long-tailed problem where some classes in testing samples have zero occurrences in the training samples [20].

2.3 Long Tails in Historical Text Recognition

As shown in Fig 1, the long-tail problem is yielding significant challenges in ancient text recognition due to: 1) The long-tail characteristics of human language itself [37]. 2) The number of ancient books is limited. However, this problem has yet to receive wide attention. A few methods propose to alleviate this problem via augmenting [17,22].

The current methods to address this problem are mostly focusing on utilizing component knowledge that is shared between tail and head classes, e.g. radical composition, to address tail performance [10,40,8], or achieve zero-shot recognition capability [14,9,5,4,11]. However, these composition-based methods rely intensively on detailed radical [29] or stroke [7] annotations, which are expensive and bound to specific languages.

To address the dilemma, we propose to implicitly exploit the shared character part by emphasizing the modeling of such features in the backbone network for feature extraction.

3 Methodology

3.1 Overview

Based on vsdf[19], we implemented a spindle architecture for Chinese historical text recognition called SpindleNet. The overview framework is shown in Fig.2. The architecture consists of three modules: a feature extraction module, a character length and weight prediction module, and a decoder module. The input image is first pre-processed by scaling its width to 32 pixels keeping the aspect ratio and then center padding into a 32x320 image I.

The image is then passed to the feature extraction networkNet, resulting output feature maps F: $(F_1, ..., F_5)$,

$$F = Net(I). \tag{1}$$

The fifth layer is then encoded in to the final feature map F^f with a convolution block,

$$F^f = Conv(F_5). \tag{2}$$

Features from the first and the third of the feature extraction network, together with the final feature map, are input to an attention module to predict the sequence length l and the location masked of each individual character A,

$$A, l = LCAM(F_1, F_3, F^f) \tag{3}$$

The image features are then sampled into time-stamp aligned character features F^c,

$$F^c_t = \sum_{i,j}^{w,h} A_{t,i,j} F^f_{i,j} \tag{4}$$

The character features are then input to the decoder for prediction.

$$Y = Pred(F^c). \tag{5}$$

For more details, please refer to vsdf[19].

3.2 Feature extraction network

The feature extraction network consists of multiple ResNet layers. Traditional feature extraction networks have multiple output channels at each layer, with the number of channels increasing for deeper layers. This forms a hierarchical increasing structure, which we denote as a triangle structure. This structure is used in the baseline method. We found that the triangle structure has limited performances on historical document text recognition tasks. Our research show that one factor limits the perfomances is the imbalanced nature of the historical document text dataset. The triangle structure yields poor performance for tail classes.

In this work, we propose to emphasis the component level feature representation, as features at this level are shared by head classes and tail classes alike, providing more generalization capability [40,8]. Since each character is composed by several components, learning at this level is less prone to overfitting, hence can also increase the head class performance as well.

As the implementation, we propose to simply allocate more channels to the corresponding layers. Due to the high frequency of characters, character components should match "pattern" [2] level features, which is mostly modeled by the second and the third layer of the network. Hence, in this work, we allocate the parameter mainly to the third layer, yielding a spindle structure, yielding the spindle network.

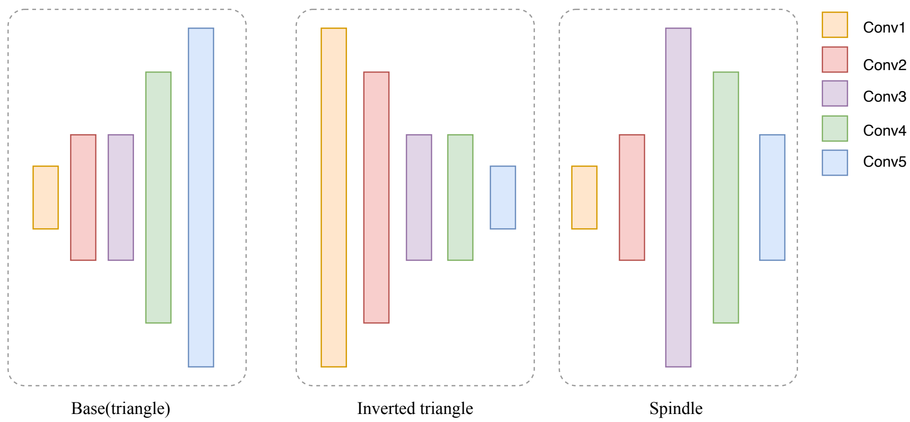

Fig. 3. The three types of feature extraction networks with different strategies are: positive triangle shape, inverted triangle shape, and spindle network shape. Different lengths represent different number of channels, for example, in the Base model, the number of channels in convolutional layers 1-5 is 32, 64, 64, 128, and 256, respectively.

3.3 Baseline network structure and spindle-shaped network structure

The baseline network structure is a triangle structure, with the number of channels increasing for deeper layers. The spindle-shaped network structure is a reversed triangle structure, with the number of channels decreasing for deeper layers. The baseline network structure consists of five ResNet layers, number of channels for each layer is as follows:[32, 64, 64, 128, 256]. The spindle-shaped network structure also consists of five ResNet layers. The number of channels for each layer is as follows: [32, 64, 256, 128, 64]. Their structures are shown in the Fig.3 respectively.

The baseline network structure is a common choice for image classification tasks. However, we found that it performs poorly on ancient document text recognition tasks. This is because the ancient document text dataset is imbalanced, with a large number of rare characters. The baseline network structure has difficulty learning the features of rare characters, which leads to poor performance for those characters. The spindle-shaped network structure addresses this issue by increasing the number of channels in the intermediate layers. This allows the network to learn more complex features, which is beneficial for rare characters. We found that the spindle-shaped network structure achieves the best overall performance on the ancient document text recognition task, including the best performance for rare characters.

The spindle-shaped network structure is a promising architecture for ancient document text recognition. It addresses the issue of imbalanced datasets by increasing the number of channels in the intermediate layers. This allows the network to learn more complex features, which is beneficial for rare characters.

4 Experiments

4.1 Datasets and Protocols

In this work, we measure the model performance on the Tripitaka Koreana in Han (TKH) Dataset and the Multiple Tripitaka in Han (MTH) Dataset [35]. Following [21], we use the combined version of TKH and MTH2200, which is named MTHv2. Specifically, the MTHv2 dataset provides line-level annotation, character-level annotation, and “boundary lines” which include reading order information. In this work, we mainly use the line-level annotation which takes the minimum cost to obtain.

Protocol-wise, we mainly evaluate the overall performance of our method on the full testing set, following the exact split from [21], which randomly split the MTHv2 dataset into the training set and the testing set with the ratio of 3:1. It is important to note that our training and test sets are kept completely consistent with [21]. We did not use any additional data such as synthetic data or pre-trained models Besides the benchmarking, we conduct extensive ablative and behavior analysis to validate the proposed approach.

4.2 Implementation details

The code is implemented based on the OpenCCD code base [19]. The input image is resized to 32 pixels by width and center padded to 32×320 image. The model is trained for 128 epochs with batch size set to 64 from scratch.

The experiments are conducted on a virtual machine with Pytorch-1.12.1, TorchVision-0.13.1 CUDA-11.2, and Ubuntu 22.04. Training from scratch using an Nvidia RTX 4090 GPU would typically take around 10 hours to complete 128 epochs. Depending on the model, when the batch size is set to 64, the GPU memory usage is about 11-21GB. The models, codes, and documents are released on `https://github.com/makaspacex/spindlenet`.

4.3 Open-set Comparison with SOTA

This section discusses the novel characters recognition capability of the proposed model. Since the new characters are mainly distributed in the tail, we discuss the performance change of the tail class and report the main performance indicators of the new characters.

Per-class Performance Analysis This provides per-class performance change analysis to provide more insight on the improving pattern. The overall trend is shown in Fig.4. The blue line is drawn in descending order of character frequency. Under the condition of ensuring character alignment, the red line draws the

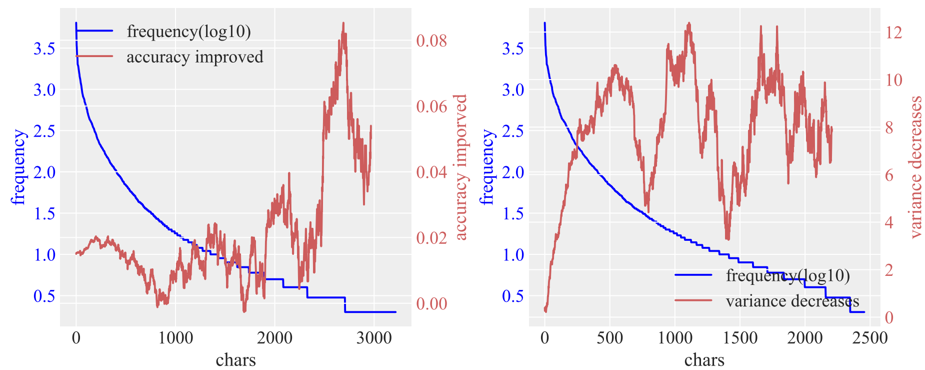

Fig. 4. frequency and accuracy imporved

improvement in accuracy and the reduction in intra-class variance respectively. Since the original data changes greatly and the trend cannot be seen, the red line is smoothed using the moving average method. The window size of the moving average method is 250. It can be seen that the performance improvement of the head class is significantly smaller than that of the tail class, both in terms of accuracy and intra-class variance. This proves that SpindleNet has better feature extraction capabilities, thereby improving the performance of the tail class. Furthermore, the performance of new characters in the tail class has also been improved, and the specific data will be shown next.

Table 1. Novel characters performentce comparison to State of The Art methods on the MTHv2 [21] dataset. LA refers to Line Accuracy and Acc refers to character accuracy.

Name	Venue	AR	CR	LA	Acc
VSDF*[19]	CVPR' 22	77.56	81.65	29.08	51.40
Ours	-	**79.31**	**82.96**	**29.26**	**52.96**

Novel Characters Performance on MTHv2 We report the performances of samples that contain unseen characters to give an estimation of how well the proposed modules handle the unseen characters in Table. 1. In the base model, the accuracy of new characters is **51.40**, while in the spindle network, the accuracy of new characters is **52.96**. When calculating the accuracy of new characters, more stringent standards are used. Specifically, no alignment operation is performed before calculation. Only when the predicted characters are consistent with the characters in GT at the corresponding positions, the characters are judged to be correct. When a line of text contains new characters, the line of text is used to calculate AR, CR, and LA. In Table. 1, SpindleNet achieves SOTA results in new character recognition performance, both in AR, CR, ACC and LA. In summary, we have proven the highly consistent relationship between new characters and tail classes, and the effectiveness of SpindleNet in new character recognition.

4.4 Close-set Comparison with SOTA

This section discusses the close-set recognition capability of the proposed model. We report the performance of the proposed model on mainstream indicators in Table 2. The proposed model achieved SOTA in each indicator of the closed set test. Specifically, the performance of AR, CR and LA are 94.21, 95.27 and 70.77 respectively. According to the observation in Fig.4, the performance improvement mainly comes from the tail class.

Table 2. Comparison to State of The Art methods on the MTHv2 [21] dataset. LA refers to Line Accuracy.

Name	Venue	AR	**CR**	LA
JLA [21]	icfhr' 20	94.08	95.09	-
VSDF*[19]	CVPR' 22	93.14	94.41	67.59
Ours	-	**94.21**	**95.27**	**70.77**

4.5 Ablative Studies

We first conduct module-level ablative experiments to validate the effectiveness of the proposed spindle network. Then, we provide an extended architecture-level ablative analysis, discussing various other possible designs and why they are less feasible than the proposed spindle-shaped network.

Module-level Ablative In this part, we perform ablative studies on the design of the spindle network, the quantitative results are shown in Fig.5. After a period of training, SpindleNet has higher performance than the base model in AR, CR and ACC. This proves that SpindleNet has stronger feature extraction capabilities and stability.

We further perform qualitative analysis to find out how the spindle net affects the character features, shown in Fig.6. In Fig.6, the tsne-cuda tool is used to visualize the last layer of features of the feature extractor. It can be seen that the features extracted by SpindleNet have clearer classification boundaries. For example, the orange character is divided into two in base, while it is gathered together in SpindleNet. This shows that SpindleNet has better feature extraction capabilities.

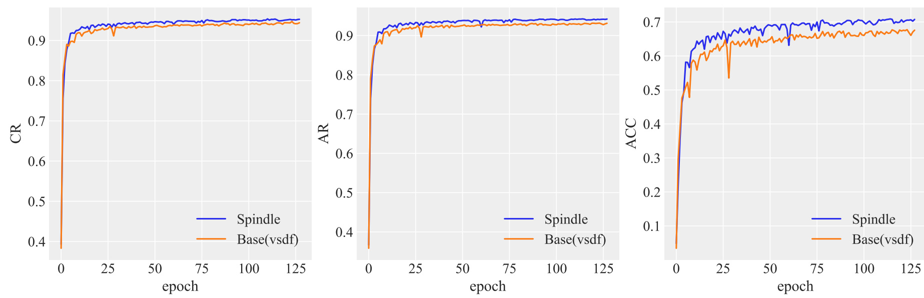

Fig. 5. Changes in CR, AR and ACC during training.

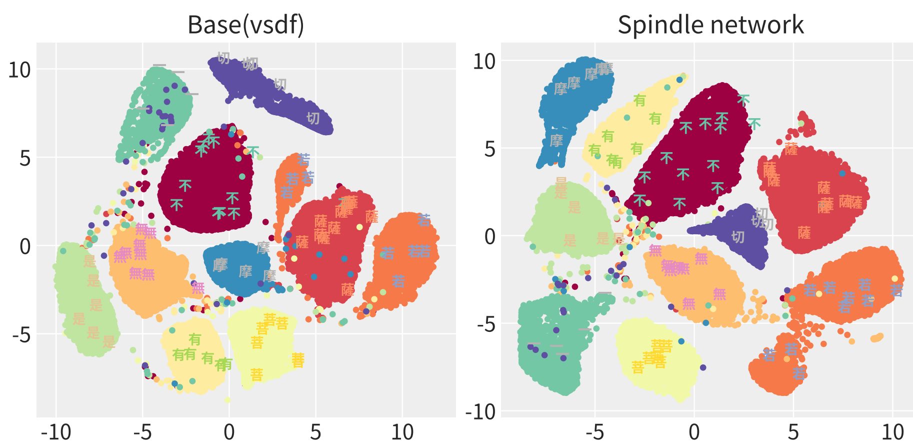

Fig. 6. Visualization of TSNE features from the top 10 character frequencies.

Architecture-level Ablative In this section, we discuss structural sensitivity, i.e. which convolutional layer deserves the most parameters. Fig.7 shows the impact of structures on performance in two extreme cases, one is an equilateral triangle structure used in base, and the other is an inverted triangle structure. In the end, SpindleNet achieved the best results, which proves that appropriately increasing the number of channels or parameters in the head can improve model performance.

5 Conclusion

In this paper, we found that the similarity of character components can be exploited to improve performance in the tail of long-tail distribution data, which can improve the performance of new characters. Furthermore, We implemented a spindle network to enhance the ability to extract character component fea-

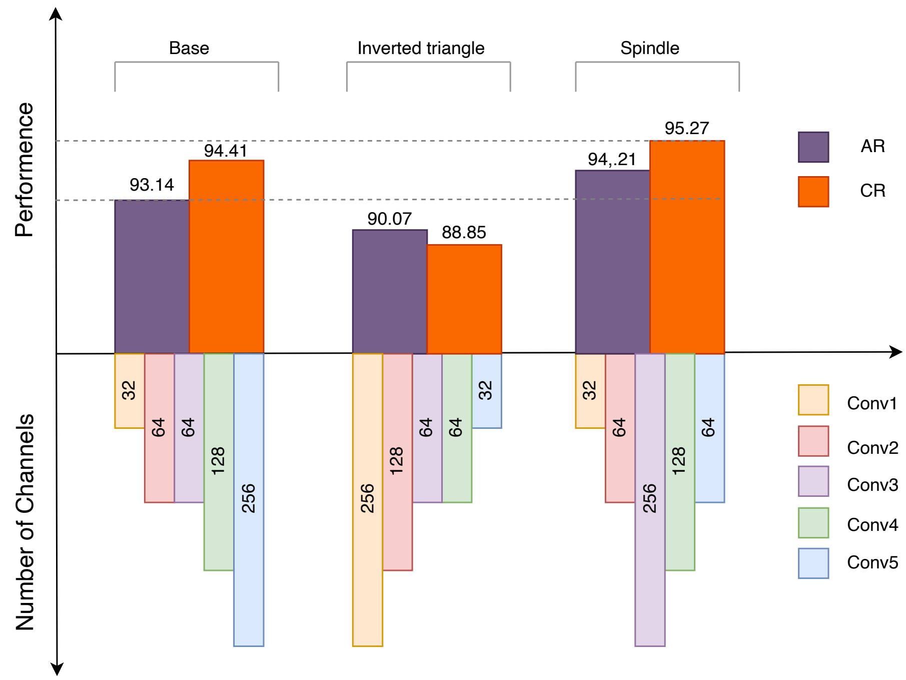

Fig. 7. The ablative on architecture. Note we kept the channel number of the output layer (conv5) fixed to rule out affects from other module to the performance.

tures, leveraging the similarities between character components to improve the recognition performance of new characters. Additional analysis demonstrates the effectiveness of SpindleNet. Experiments on three challenging Chinese ancient book datasets (TKH, MTH1000, and MTH1200) to validate the superiority of our proposed method. The results show that our approach achieves state-of-the-art performance in this field. In the future, we plan to train our framework in a weakly-supervised manner, which has proven to be successful in the scene text field.

6 Acknowledgement

This research is supported in part by NSFC (Grant No.: 61936003), GD-NSF (no.2017A030312006), the National Key Research and Development Program of China (No. 2016YFB1001405), Guangdong Intellectual Property Office Project (2018-10-1), and Fundamental Research Funds for the Central Universities (x2dxD2190570).

References

1. Amin, J., Siddiqi, I., Moetesum, M.: Reconstruction of broken writing strokes in greek papyri. In: International Conference on Document Analysis and Recognition. pp. 253–266. Springer (2023)
2. Bau, D., Zhou, B., Khosla, A., Oliva, A., Torralba, A.: Network dissection: Quantifying interpretability of deep visual representations. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 6541–6549 (2017)
3. Cao, Z., Lu, J., Cui, S., Zhang, C.: Zero-shot handwritten chinese character recognition with hierarchical decomposition embedding. Pattern Recognit. **107**, 107488 (2020)
4. Chanda, S., Baas, J., Haitink, D., Hamel, S., Stutzmann, D., Schomaker, L.: Zero-shot learning based approach for medieval word recognition using deep-learned features. In: 16th International Conference on Frontiers in Handwriting Recognition, ICFHR 2018, Niagara Falls, NY, USA, August 5-8, 2018. pp. 345–350. IEEE (2018)
5. Chanda, S., Haitink, D., Prasad, P.K., Baas, J., Pal, U., Schomaker, L.: Recognizing bengali word images - A zero-shot learning perspective. In: 25th International Conference on Pattern Recognition, ICPR 2020, Virtual Event / Milan, Italy, January 10-15, 2021. pp. 5603–5610. IEEE (2020)
6. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: Smote: synthetic minority over-sampling technique. Journal of artificial intelligence research **16**, 321–357 (2002)
7. Chen, J., Li, B., Xue, X.: Zero-shot chinese character recognition with stroke-level decomposition. In: Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI 2021, Virtual Event / Montreal, Canada, 19-27 August 2021. pp. 615–621. ijcai.org (2021)
8. Chen, J., Yu, H., Ma, J., Guan, M., Xu, X., Wang, X., Qu, S., Li, B., Xue, X.: Benchmarking chinese text recognition: Datasets, baselines, and an empirical study (2021), `https://arxiv.org/abs/2112.15093`
9. Diao, X., Shi, D., Li, J., Shi, L., Yue, M., Qi, R., Li, C., Xu, H.: Toward zero-shot character recognition: A gold standard dataset with radical-level annotations. In: Proceedings of the 31st ACM International Conference on Multimedia. pp. 6869–6877 (2023)
10. Diao, X., Shi, D., Tang, H., Qiang, S., Li, Y., Wu, L., Xu, H.: Rzcr: Zero-shot character recognition via radical-based reasoning. IJCAI (2023)
11. He, S., Schomaker, L.: Open set chinese character recognition using multi-typed attributes (2018), `http://arxiv.org/abs/1808.08993`
12. Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H.: Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861 (2017)
13. Huang, S., Wang, H., Liu, Y., Shi, X., Jin, L.: OBC306: A large-scale oracle bone character recognition dataset. In: 2019 International Conference on Document Analysis and Recognition, ICDAR 2019, Sydney, Australia, September 20-25, 2019. pp. 681–688. IEEE (2019)
14. Huang, Y., Jin, L., Peng, D.: Zero-shot chinese text recognition via matching class embedding. In: 16th International Conference on Document Analysis and Recognition, ICDAR 2021, Lausanne, Switzerland, September 5-10, 2021, Proceedings, Part III. Lecture Notes in Computer Science, vol. 12823, pp. 127–141. Springer (2021)

15. Jiang, C.M., Najibi, M., Qi, C.R., Zhou, Y., Anguelov, D.: Improving the intra-class long-tail in 3d detection via rare example mining. In: European Conference on Computer Vision. pp. 158–175. Springer (2022)
16. Kordon, F., Weichselbaumer, N., Herz, R., Mossman, S., Potten, E., Seuret, M., Mayr, M., Christlein, V.: Classification of incunable glyphs and out-of-distribution detection with joint energy-based models. International Journal on Document Analysis and Recognition (IJDAR) pp. 1–18 (2023)
17. Li, J., Wang, Q.F., Huang, K., Yang, X., Zhang, R., Goulermas, J.Y.: Towards better long-tailed oracle character recognition with adversarial data augmentation. Pattern Recognition **140**, 109534 (2023)
18. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision. pp. 2980–2988 (2017)
19. Liu, C., Yang, C., Yin, X.: Open-set text recognition via character-context decoupling. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022. pp. 4513–4522. IEEE (2022)
20. Liu, Z., Miao, Z., Zhan, X., Wang, J., Gong, B., Stella, X.Y.: Open long-tailed recognition in a dynamic world. IEEE Transactions on Pattern Analysis and Machine Intelligence (2022)
21. Ma, W., Zhang, H., Jin, L., Wu, S., Wang, J., Wang, Y.: Joint layout analysis, character detection and recognition for historical document digitization. In: 2020 17th International Conference on Frontiers in Handwriting Recognition (ICFHR). pp. 31–36. IEEE (2020)
22. Park, S., Chung, S., Lee, J., Choo, J.: Improving scene text recognition for character-level long-tailed distribution. arXiv preprint arXiv:2304.08592 (2023)
23. Pourpanah, F., Abdar, M., Luo, Y., Zhou, X., Wang, R., Lim, C.P., Wang, X., Wu, Q.M.J.: A review of generalized zero-shot learning methods. IEEE Trans. Pattern Anal. Mach. Intell. **45**(4), 4051–4070 (2023)
24. Shen, L., Lin, Z., Huang, Q.: Relay backpropagation for effective learning of deep convolutional neural networks. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VII 14. pp. 467–482. Springer (2016)
25. Souibgui, M.A., Fornés, A., Kessentini, Y., Megyesi, B.: Few shots is all you need: A progressive few shot learning approach for low resource handwriting recognition (2021), `https://arxiv.org/abs/2107.10064`
26. Verma, V.K., Arora, G., Mishra, A., Rai, P.: Generalized zero-shot learning via synthesized examples. In: Proceedings of the IEEE conference on computer vision and pattern recognition (2018)
27. Vu, M.T., Beurton-Aimar, M.: Papytwin net: a twin network for greek letters detection on ancient papyri. In: Proceedings of the 7th International Workshop on Historical Document Imaging and Processing. pp. 43–48 (2023)
28. Wang, H., Wu, X., Huang, Z., Xing, E.P.: High-frequency component helps explain the generalization of convolutional neural networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8684–8694 (2020)
29. Wang, T., Xie, Z., Li, Z., Jin, L., Chen, X.: Radical aggregation network for few-shot offline handwritten chinese character recognition. Pattern Recognit. Lett. **125**, 821–827 (2019)
30. Wang, W., Zhang, J., Du, J., Wang, Z., Zhu, Y.: Denseran for offline handwritten chinese character recognition. In: 16th International Conference on Frontiers in

Handwriting Recognition, ICFHR 2018, Niagara Falls, NY, USA, August 5-8, 2018. pp. 104–109. IEEE Computer Society (2018)
31. Wang, X., Lian, L., Miao, Z., Liu, Z., Yu, S.X.: Long-tailed recognition by routing diverse distribution-aware experts. arXiv preprint arXiv:2010.01809 (2020)
32. Wang, Y., Fei, J., Wang, H., Li, W., Bao, T., Wu, L., Zhao, R., Shen, Y.: Balancing logit variation for long-tailed semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 19561–19573 (2023)
33. Wu, T., Liu, Z., Huang, Q., Wang, Y., Lin, D.: Adversarial robustness under long-tailed distribution. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8659–8668 (2021)
34. Xiang, L., Ding, G., Han, J.: Learning from multiple experts: Self-paced knowledge distillation for long-tailed classification. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16. pp. 247–263. Springer (2020)
35. Yang, H., Jin, L., Huang, W., Yang, Z., Lai, S., Sun, J.: Dense and tight detection of chinese characters in historical documents: Datasets and a recognition guided detector (2018)
36. Yang, L., Jiang, H., Song, Q., Guo, J.: A survey on long-tailed visual recognition. International Journal of Computer Vision **130**(7), 1837–1872 (2022)
37. Yang, L., Jiang, H., Song, Q., Guo, J.: A survey on long-tailed visual recognition. International Journal of Computer Vision **130**(7), 1837–1872 (2022)
38. Yun, S., Han, D., Chun, S., Oh, S.J., Yoo, Y., Choe, J.: Cutmix: Regularization strategy to train strong classifiers with localizable features. In: 2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019. pp. 6022–6031. IEEE (2019)
39. Zhang, J., Du, J., Dai, L.: Radical analysis network for learning hierarchies of chinese characters. Pattern Recognit. **103**, 107305 (2020)
40. Zhang, J., Liu, C., Yang, C.: SAN: structure-aware network for complex and long-tailed chinese text recognition. In: Document Analysis and Recognition - ICDAR 2023 - 17th International Conference, San José, CA, USA, August 21-26, 2023, Proceedings, Part V. Lecture Notes in Computer Science, vol. 14191, pp. 244–258. Springer (2023)
41. Zhang, Z., Meng, D., Zhang, L., Xiao, W., Tian, W.: The range of harmful frequency for dnn corruption robustness. Neurocomputing **481**, 294–309 (2022)