ICDAR 2023 Competition on Recognition of Multi-line Handwritten Mathematical Expressions

Chenyang Gao[1], Yuliang Liu[1], Shiyu Yao[2], Jinfeng Bai[2], Xiang Bai[1](✉), Lianwen Jin[3], and Cheng-Lin Liu[4]

[1] Huazhong University of Science and Technology, Wuhan, China
{m202172425,ylliu,xbai}@hust.edu.cn
[2] Tomorrow Advancing Life Education Group, Beijing, China
yaoshiyu@tal.com, jfbai.bit@gmail.com
[3] South China University of Technology, Guangzhou, China
eelwjin@scut.edu.cn
[4] Institute of Automation, Chinese Academy of Sciences, Beijing, China
liucl@nlpr.ia.ac.cn

Abstract. Mathematical expressions play an essential role in scientific documents and are critical for describing problems and theories in various fields, such as mathematics and physics. Consequently, the automatic recognition of handwritten mathematical expressions in images has received significant attention. While existing datasets have primarily focused on single-line mathematical expressions, multi-line mathematical expressions also appear frequently in our daily lives and are important in the field of handwritten mathematical expression recognition. Additionally, the structure of multi-line mathematical expressions is more complex, making this task even more challenging. Despite this, no benchmarks or methods for multi-line handwritten mathematical expressions have been explored. To address this issue, we present a new challenge dataset that contains multi-line handwritten mathematical expressions, along with a challenging task: recognition of multi-line handwritten mathematical expressions (MLHMER). The competition was held from January 10, 2023 to March 26, 2023 with 16 valid submissions. In this report, we describe the details of this new dataset, the task, the evaluation protocols, and the summaries of the results.

Keywords: Handwritten mathematical expression recognition · Multi-line handwriting recognition

1 Introduction

Mathematical expressions are a crucial component of scientific documents, providing a concise and precise way of describing complex problems and theories in various fields, such as mathematics, physics, engineering, and economics. Consequently, the automatic recognition of handwritten mathematical expressions is

G. A. Fink et al. (Eds.): ICDAR 2023, LNCS 14188, pp. 566–576, 2023.
https://doi.org/10.1007/978-3-031-41679-8_34

an essential task as it can facilitate the digitization and analysis of mathematical texts, assist visually impaired students in learning mathematics, and enable natural handwriting interfaces for mathematical notation.

In recent years, the field of handwritten mathematical expression recognition has received significant attention, with various algorithms and techniques proposed for recognizing single-line mathematical expressions. However, multi-line mathematical expressions are equally important and challenging, appearing frequently in our daily lives and in various scientific documents. More specifically, multi-line expressions are used to represent complex mathematical concepts, such as equations, matrices, systems of equations, and proofs. Moreover, the structure of multi-line expressions is more complex, making their recognition a challenging task.

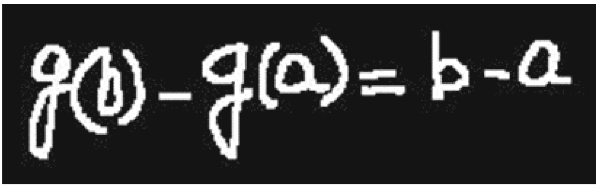
Despite the importance of recognizing multi-line handwritten mathematical expressions, no benchmarks or methods for multi-line expressions have been explored. Therefore, we present a new challenge dataset containing multi-line handwritten mathematical expressions and a task of recognizing them. As shown in Fig. 1, we give a simple comparison between multi-line mathematical expressions in our MLHME-38K dataset and single-line mathematical expressions in CROHME [10] and HME100K [11] datasets. Our goal is to promote research in handwritten mathematical expression recognition and to address the gap in the availability of datasets and methods for multi-line expressions.

Recognizing multi-line handwritten mathematical expressions is a challenging task due to several factors. First, multi-line expressions can be of varying lengths and have complex structures, with nested subexpressions and dependencies between lines. Second, the handwriting styles and fonts used in multi-line expressions can vary significantly, making it difficult to generalize across different writers and domains. Third, the recognition of multi-line expressions requires not only the recognition of individual characters and symbols but also the identification of their spatial and structural relationships.

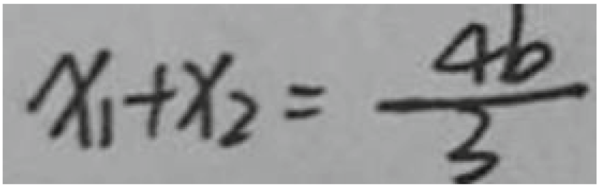
Considering the importance of recognizing multi-line handwritten mathematical expressions and the challenges it faces, we host the ICDAR 2023 Competition on Recognition of Multi-line Handwritten Mathematical Expressions. We hope this competition can attract more researchers to pay attention to this field and promote research in this area.

2 Competition Organization

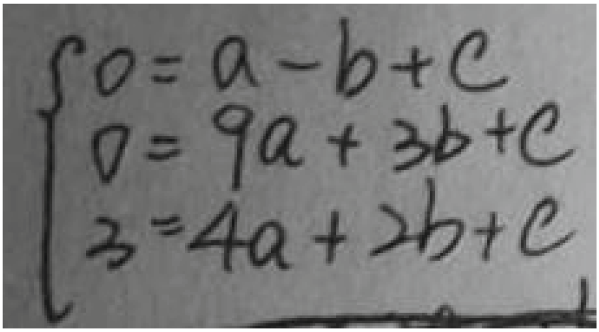
ICDAR 2023 Competition on Recognition of Multi-line Handwritten Mathematical Expressions is organized by a joint team of Tomorrow Advancing Life Education Group, Huazhong University of Science and Technology, South China University of Technology, and Institute of Automation, Chinese Academy of Sciences.



(a) A single-line mathematical expression from CROHME dataset.



(b) A single-line mathematical expression from HME100K dataset.



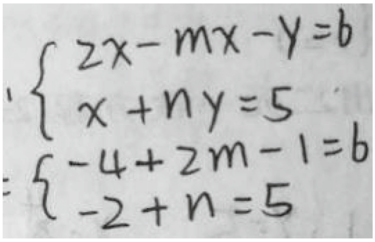
(c) A multi-line mathematical expression from our MLHME-38K dataset.

Fig. 1. Comparison between multi-line mathematical expressions in our MLHME-38K dataset and single-line mathematical expressions in CROHME and HME100K dataset.

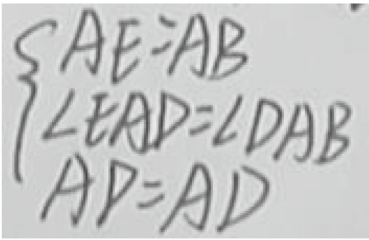
3 Dataset

We name our dataset MLHME-38K, as it focuses on **M**ulti-Line **H**andwritten **M**athematical **E**xpressions. It totally includes 38,000 labeled images, of which 9,971 images are multi-line mathematical expressions and 28,029 images are single-line. All these images were uploaded by users from real-world scenarios. Consequently, our dataset MLHME-38K becomes more authentic and realistic with variations in color, blur, complicated background, twist, illumination, longer length, and complicated structure. Some examples of our dataset are shown in Fig. 2.

The dataset is divided into a training set, a test_A set, and a test_B set. The training set consists of 30,000 images which will be available to the participants

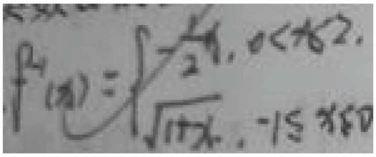


```
\begin{matrix} \left\{ \begin{matrix} 2 x - m x - y = 6 \\ x + n y = 5 \end{matrix} \right. \\ \left\{ \begin{matrix} - 4 + 2 m - 1 = 6 \\ - 2 + n = 5 \end{matrix} \right. \end{matrix}
```



```
\left\{ \begin{matrix} A E = A B \\ \angle E A D = \angle D A B \\ A D = A D \end{matrix} \right.
```

```
\left\{ \begin{matrix} \frac { 5 } { x + y } + \frac { 1 } { x - y } = 7 \textcircled { 1 } \\ \frac { 3 } { x + y } - \frac { 1 } { x - y } = 1 \textcircled { 2 } \end{matrix} \right.
```



```
f ^ { - 1 } ( x ) = \left\{ \begin{matrix} - \frac { 1 } { 2 } x , 0 < x < 2 , \\ \sqrt { 1 + x } , - 1 \leq x \leq 0 \end{matrix} \right.
```

Fig. 2. Some examples from our MLHME-38K dataset

along with their annotations. The test_A set consists of 3,000 images and the test_B set consists of 5,000 images, whose annotations will not be released. In the competition stage, only images are offered. The participants are required to submit their results on the test_A set and test_B set with the specific format. Dividing these two test sets can prevent the model from overfitting on a single test set. During the competition, test_A set will be open first and the top ranked teams on the leaderboard will be selected for the evaluation of test_B set. The final ranking is based on test_B set. Every image in the dataset is annotated with a string of LaTeX sequence denoting the mathematical expression. Annotations for images are stored in a txt file with the format shown in Table 1.

Table 1. The annotation format of our MLHME-38K dataset.

File Name	LaTeX String
train_0.jpg	\begin{matrix} 9 x − y = 3 \\ 2 x + y = 5 \end{matrix}
...	...
train_26.jpg	\begin{matrix} n + 2 m = a \\ 2 n − m = 3 a \end{matrix}

4 Tasks

The competition has only one task: recognition of multi-line handwritten mathematical expressions. The aim of this task is to recognize the multi-line handwritten mathematical expressions in images and output them in LaTeX format.

4.1 Evaluation Protocol

In this task, expression recall and character recall are utilized to evaluate the performance.

– Expression recall: The percentage of predicted LaTeX formula sequences matching ground truth (ignore space).

$$S_{\text{recall}} = \frac{S_{\text{right}}}{S_{\text{sum}}} \tag{1}$$

– Character recall: C_{diff} is the sum of edit distances for all images and C_{sum} is the number of characters for all labels.

$$C_{\text{recall}} = 1 - \frac{C_{\text{diff}}}{C_{\text{sum}}} \tag{2}$$

For example, Suppose there are two predictions: "1+1==2" and "a−b=11", and the corresponding labels are "1+1=2" and "a−b=11". As for the expression recall, S_{sum} is 2 since there are two predictions, and S_{right} is 1 since only "a−b=11" match the target. So the expression recall is 0.5 in this case. For the character recall, C_{sum} is 11 (5+6) since the label "1+1=2" contains 5 characters and "a−b=11" contains 6 characters. C_{diff} is 1 (1+0) because the edit distance between "1+1==2" and "1+1=2" is 1. The edit distance between "a−b=11" and "a−b=11" is 0. So the character recall is 0.909 in this case.

4.2 Evaluation Details

The ranking results on the leaderboard of both test_A and test_B set is based on the expression recall. During the competition, test_A set will be open first and the top-10 teams on the leaderboard will be selected for the evaluation of test_B set. After the submission of test_B set due, the final ranking will be given based on the expression recall (higher priority) and character recall:

$$\text{Better } = \begin{cases} S_{r1}, \text{ if } S_{r1} \geq S_{r2} + 0.001 \\ S_{r1}, \text{ if } |S_{r1} - S_{r2}| < 0.001 \text{ and } 0.9\,(S_{r1} - S_{r2}) + 0.1\,(C_{r1} - C_{r2}) > 0 \\ S_{r1}, \text{ if } S_{r1} > S_{r2} \text{ and } 0.9\,(S_{r1} - S_{r2}) + 0.1\,(C_{r1} - C_{r2}) = 0 \\ S_{r2}, \text{ otherwise} \end{cases}$$

where S_{r1} and C_{r1} denotes the expression recall and the character recall from one team. While S_{r2} and C_{r2} denote the expression recall and the character recall from another team.

The final ranking mainly depends on the expression recall. When the expression recall difference between the two teams does not exceed 0.001, we will utilize character recall as an additional measure. Since tiny expression recall differences indicate that the performance of the two methods is very close. In this case, the character recall can provide a better comparison since it focuses on individual characters. We also give different weights for the expression recall and the character recall. The higher weight is given to the expression recall because it is more commonly used in recognition of handwritten mathematical expressions. The chosen hyperparameters mainly rely on previous experience in the similar competition[1].

5 Submissions

Overall, we received valid submissions from 16 teams from both research communities and industries. The submission results of test_A and test_B set are shown in Table 2 and 3. The final ranking is based on the results of test_B set (Table 3).

5.1 Top 5 Submissions

iFLYTEK-OCR team uses an encoder-decoder architecture that formulates HMER as an image-to-sequence translation problem. Specifically, the Conv2Former [2] is employed as the image encoder, and a bi-directional trained Transformer decoder [13] with Attention Refinement Module [12] is utilized as the latex sequence decoder. A Beam Search Ensemble is proposed to ensemble the models trained with different sizes of characters. Specifically, at each decoding step, probability distributions produced by all member models are averaged by certain weights, and the top-k candidate characters to be output are decided by the averaged probability distribution. As for the data augmentation, blur, random, color jitter, scale [6], and TIA Transform [9] are applied to improve the generalization ability of the model.

100-Gan Car University team utilizes CoMER [12] as the baseline model. To efficiently establish time-series information in the encoder, they use two conformer blocks [1] to extract the sequential information in the image feature map

[1] https://www.heywhale.com/home/competition/5f703ac023f41e002c3ed5e4/content.

Table 2. Test_A set results and rankings on recognition of multi-line handwritten mathematical expressions.

Rank	Team Name	Exp. Recall	Team Members	Affiliations
1	iFLYTEK-OCR	0.6697	Hao Wu, Mingjun Chen, Xuejing Niu, Changpeng Pi	iFLYTEK
2	BOE_AIOT_AIBD	0.6573	Zhanfu An, Guangwei Huang, Ruijiao Shi, Rui Zheng	Boe Technology Group Co., Ltd.
3	MACARON	0.6557	Yamato Okamoto, Baek Youngmin, Ichimura Shuta, Nakao Ryota, Nakagome Yu	LINE Corporation
4	Just do your best	0.6463	Yingnan Fu, Tiandi Ye	East China Normal University
5	TianyuAI	0.6383	Yu Yan	Wuhan Tianyu Information Industry Co., Ltd.
6	100-Gan Car University	0.6330	Zhuoyan Luo, Yinghao Wu, Zihang Xu, Qi Jing, Hui Xue	Southeast University
7	HMEfly	0.6300	Chenyu Liu	University of Science and Technology of China
8	zyyjyvision	0.6217	Hui Zheng, Xiahan Yang, Qianwen Jia	None
9	sysu	0.6210	Qiqiang Lin, Haiyang Xiao	Sun Yat-sen University
10	FM	0.5993	Xiaoyu Zhang, Maojin Xia, Wenhui Dong	Shanghai Yichuang Information Technology Co., Ltd., Huawei Technologies Co., Ltd.
11	C_OCR	0.5937	Haoyang Shen	Guangzhou Shiyuan Electronics Co., Ltd
12	None	0.5687	Dan Luo	None
13	Not ready yet	0.5937	Weiwei Yi	Vivo Communication Technology Co. Ltd.
14	AYYTeam	0.4363	Zhe Wang, Yifan Bian, Mingyi Ma, Zeyu Chen	None
15	BeatMER	0.2717	Jing Xian, Xingran Zhao	Ant Group Co., Ltd.
16	atd_doc_cits	0.2347	Nobukiyo Watanabe, Tadahito Yao	Canon IT Solutions Inc.

from the horizontal and vertical axes, respectively. Then a fusion operation is conducted on the reshaped feature maps from two different perspectives to get the final image representations. Apart from the image encoder, the other parts are the same as CoMER. The scale augmentation [6], distortion, sketch and perspective are employed to augment the images. Moreover, a simple voting scheme is utilized to ensemble the models trained with different settings.

Table 3. Test_B set results and rankings on recognition of multi-line handwritten mathematical expressions. Note that * denotes character recall.

Rank	Team Name	Exp. Recall	Team Members	Affiliations
1	iFLYTEK-OCR	0.6790 (0.9695*)	Hao Wu, Mingjun Chen, Xuejing Niu, Changpeng Pi	iFLYTEK
2	100-Gan Car University	0.6300 (0.9603*)	Zhuoyan Luo, Yinghao Wu, Zihang Xu, Qi Jing, Hui Xue	Southeast University
3	BOE_AIOT_AIBD	0.6244 (0.9618*)	Zhanfu An, Guangwei Huang, Ruijiao Shi, Rui Zheng	Boe Technology Group Co., Ltd.
4	TianyuAI	0.6186 (0.9583*)	Yu Yan	Wuhan Tianyu Information Industry Co., Ltd.
5	MACARON	0.6166 (0.9552*)	Yamato Okamoto, Baek Youngmin, Ichimura Shuta, Nakao Ryota, Nakagome Yu	LINE Corporation
6	Just do your best	0.6036 (0.9492*)	Yingnan Fu, Tiandi Ye	East China Normal University
7	sysu	0.5950 (0.9589*)	Qiqiang Lin, Haiyang Xiao	Sun Yat-sen University
8	zyyjyvision	0.5950 (0.9463*)	Hui Zheng, Xiahan Yang, Qianwen Jia	None
9	FM	0.5456 (0.9381*)	Xiaoyu Zhang, Maojin Xia, Wenhui Dong	Shanghai Yichuang Information Technology Co., Ltd., Huawei Technologies Co., Ltd.
10	HMEfly	0.2272 (0.1714*)	Chenyu Liu	University of Science and Technology of China

BOE_AIOT_AIBD team adopts three steps to solve the problem. First, an adaptive image adjustment is proposed to determine the number of lines of the expression in an image. Second, to solve the problem of different image resolutions, an Image Super Resolution (ISR) module is added to CAN [5] and CoMER [12], resulting in six models with different training strategies. The simple voting scheme is utilized to ensemble these models. Finally, they judge the latex format of the fusion results. If the latex format is incorrect, the SAN network is utilized to further correct these results. As for the data augment methods, color enhancement (adjusting gamma value, adding Gaussian blur, adjusting hue, saturation, and value), scale augmentation (scale ratio value range [0.7,1.4]), and rotation (rotation angle value range [−5,5]) are applied to improve the robustness of the model.

TianyuAI team adopts CAN [5] and CoMER [12] as the baseline models. To enhance the prediction of single-line mathematical expressions, they utilize a CTC branch during the training phase. They also introduce focal loss [7] to address the unbalanced character distribution in the dataset. Similar to most teams, the model ensemble strategy is adopted to improve the performance of their methods. Specifically, a simple voting scheme is utilized to ensemble the CAN and CoMER models trained with different settings. As for the data augmentation, blur, random sharpness, random contrast, color jitter, rotate and stretch are utilized to improve the performance of the models. Moreover, they use Test Time Augmentation (TTA) during the testing phase.

MACARON team use Donut [4] as the baseline framework, which is a method of document understanding that utilizes an OCR-free end-to-end Transformer model. The encoder and decoder of Donut are SwinTransformer [8] and Multilingual BART. They utilize RGB shift, random brightness, random contrast, hue saturation value, channel shuffle, CLAHE, random sun flare, sharpen, gaussian blur, optical distortion, coarse dropout, and rotate as the data augment methods. Different from most teams, no model ensemble strategy is applied to further improve the performance of their method.

5.2 Discussion

In this task, most participants utilize an encoder-decode framework that models the recognition process in a sequence-to-sequence manner. They first employ a powerful backbone as the image encoder to enhance the performance. DenseNet [3], SwinTransformer [8] and Conv2Former [2] are the commonly used image encoder. Additionally, the conformer block [1] and counting sub-task [5] are utilized to reinforce the extraction of image features. As for the LaTeX sequence decoder, counting-based GRU decoder [5], Transformer decoder [13] with Attention Refinement Module [12], and syntax-aware decoder [11] are frequently utilized.

To improve the generalization ability of the model, most participants use various data augmentations, e.g., blur, color jitter, rotation, distortion, sketch, perspective, and so on. Apart from these commonly used data augmentations, the scale [6] and TIA Transform [9] can also effectively improve the performance of different methods. The scale [6] augmentation can address the recognition difficulty caused by symbols of various sizes. The TIA Transform [9] is a learnable geometric augmentation that can bridge the isolated processes of data augmentation and model training.

The model ensemble strategy is also utilized by most teams to improve the performance of their methods. The typical process is first to train a predefined number of models with different training settings of the same method or with different methods. During the inference stage, probability distributions produced by all member models are ensembled to produce the final results at each decoding step. The voting scheme and weighted average are the most widely used ensemble strategies. However, the model ensemble strategy greatly increases the inference time, limiting its application in real life.

Since our MLHME-38K dataset contains not only multi-line mathematical expressions but also single-line mathematical expressions. More specifically, the test B set contains 1,880 multi-line mathematical expressions and 3,120 single-line mathematical expressions. In order to see the performance of different methods on the multi-line mathematical expressions, we conduct experiments on the multi-line subset of the test B set. As shown in Table 4, the iFLYTEK-OCR team still ranks first. However, the MACARON team rises from 5th to 2nd which indicates that document understanding frameworks may have more potential for multi-line mathematical expressions.

From Table 4, it can be observed that the scores for the multi-line subset of Test B are higher than the scores for the single-line subset. The major reason is that the diversity of the multi-line expressions is much smaller than that of single-line expressions in our MLHME-38K dataset. The image sources of our dataset are mainly the test questions and solutions from middle school and high school, which leads to the fact that the multi-line expressions in our dataset are mostly equation sets. We will try to incorporate more diverse data (e.g. university test questions) into our dataset in the future.

Table 4. The results on the multi-line subset and single-line subset of the test_B set. The rankings are based on the multi-line subset.

Rank	Team Name	Multi-line Exp. Recall	Single-line Exp. Recall
1	iFLYTEK-OCR	0.7532	0.6343
2	MACARON	0.7197	0.5545
3	100-Gan Car University	0.7160	0.5782
4	BOE_AIOT_AIBD	0.6968	0.5792
5	Just do your best	0.6952	0.5484
6	zyyjyvision	0.6952	0.5311
7	TianyuAI	0.6947	0.5715
8	sysu	0.6814	0.5429
9	FM	0.6542	0.4801
10	HMEfly	0.3069	0.1792

6 Conclusion

This paper summarizes the organization and results of ICDAR 2023 Competition on Recognition of Multi-line Handwritten Mathematical Expressions. In this competition, we present a new dataset named MLHME-38K, which focuses on multi-Line handwritten mathematical expressions. Compared with single-line mathematical expressions, the structure of multi-line mathematical expressions is more complicated which makes this task more challenging. We received valid submissions from 16 teams from both research communities and industries. These submissions will provide this research area with new insights and potential solutions. We also believe that our dataset will contribute to advancing the field of handwritten mathematical expression recognition.

Acknowledgments. This competition is supported by the National Natural Science Foundation (NSFC#62225603).

References

1. Gulati, A., et al.: Conformer: convolution-augmented transformer for speech recognition. arXiv preprint arXiv:2005.08100 (2020)
2. Hou, Q., Lu, C.Z., Cheng, M.M., Feng, J.: Conv2former: a simple transformer-style convnet for visual recognition. arXiv preprint arXiv:2211.11943 (2022)
3. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4700–4708 (2017)
4. Kim, G., et al.: OCR-free document understanding transformer. In: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T. (eds.) ECCV 2022. LNCS, vol. 13688, pp. 498–517. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-19815-1_29
5. Li, B., et al.: When counting meets HMER: counting-aware network for handwritten mathematical expression recognition. In: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T. (eds.) ECCV 2022. LNCS, vol. 13688, pp. 197–214. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-19815-1_12
6. Li, Z., Jin, L., Lai, S., Zhu, Y.: Improving attention-based handwritten mathematical expression recognition with scale augmentation and drop attention. In: 17th International Conference on Frontiers in Handwriting Recognition, pp. 175–180. IEEE (2020)
7. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2980–2988 (2017)
8. Liu, Z., et al.: Swin transformer: hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10012–10022 (2021)
9. Luo, C., Zhu, Y., Jin, L., Wang, Y.: Learn to augment: joint data augmentation and network optimization for text recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13746–13755 (2020)
10. Mouchere, H., Viard-Gaudin, C., Zanibbi, R., Garain, U.: ICFHR 2014 competition on recognition of on-line handwritten mathematical expressions (CROHME 2014). In: 14th International Conference on Frontiers in Handwriting Recognition, pp. 791–796. IEEE (2014)
11. Yuan, Y., et al.: Syntax-aware network for handwritten mathematical expression recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4553–4562 (2022)
12. Zhao, W., Gao, L.: Comer: modeling coverage for transformer-based handwritten mathematical expression recognition. In: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T. (eds.) ECCV 2022. LNCS, vol. 13688, pp. 392–408. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-19815-1_23
13. Zhao, W., Gao, L., Yan, Z., Peng, S., Du, L., Zhang, Z.: Handwritten mathematical expression recognition with bidirectionally trained transformer. In: Lladós, J., Lopresti, D., Uchida, S. (eds.) ICDAR 2021. LNCS, vol. 12822, pp. 570–584. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-86331-9_37