Supplementary Material: Syntax-Aware Network for Handwritten Mathematical Expression Recognition

Anonymous CVPR submission

Paper ID 4161

1. Dataset Annotation

To develop the HME100K dataset, first, we collect more than one billion expressions from an Internet application. Then we used a pre-trained classifier to filter the data, and we got about 700K handwritten expressions after removing duplicating ones with the MD5 algorithm. Generally, we conducted three-stage labeling processes to finalize the whole database development.

The First Round Labeling. We use an open-sourced DenseWAP model [1] to predict the labels of the 700K images. We then use a crowdsourcing platform to manually correct the labels of the images.

The Second Round Labeling. After the first round of labeling, we randomly split the images into two parts. The first part contains about 600K images, and the second part contains about 100K images. We use the first part to train a new DenseWAP model, by which the labels of the second part are predicted. We then use the crowdsourcing platform to manually correct the labels of the images in the second part. The workers are also asked to remove images that are not MEs from the HME100K dataset.

The Third Round Labeling. After the second round of labeling, we further clean up the dataset by asking new workers to check the labels of the dataset. If the first-round label is different from the second-round label, the image will be sent to the third-round labeling. For those images, we use the third-round labels as the final labels. The obtained dataset is randomly divided into a training set and a testing set with a three-to-one ratio.

2. Most Frequently Observed Symbols

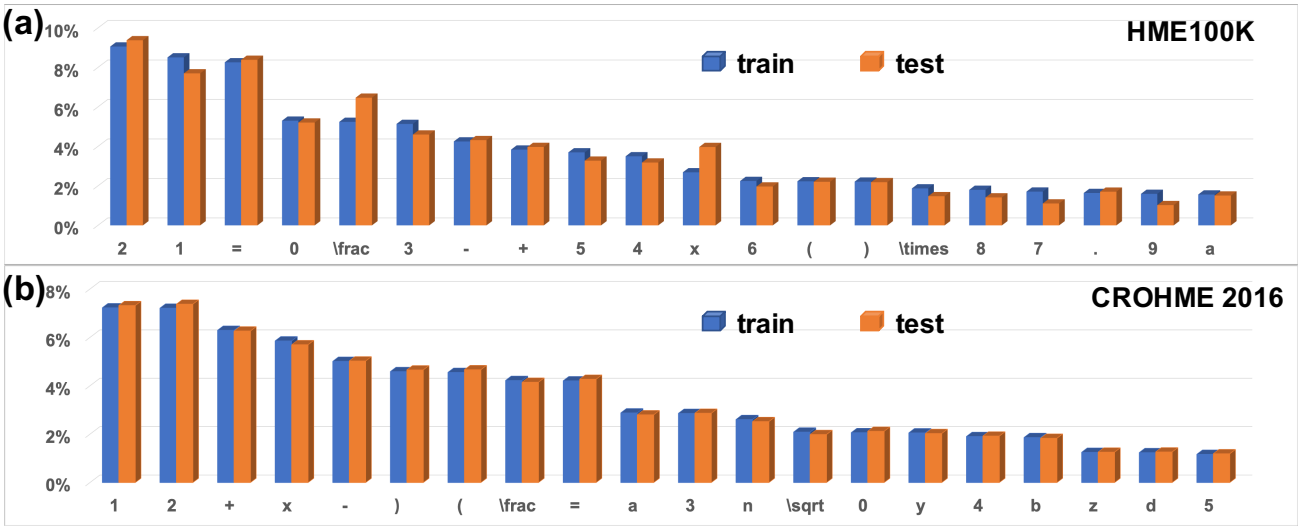

Figure 1. The percentages of 20 most frequently observed symbols in (a) HME100K and (b) CROHME 2016 dataset. The percentages of training and testing datasets are shown in blue and orange, respectively.

3. Distribution of HME100K dataset

Tab. 1 shows the distribution of blur, twist, etc. of HME100K from 1000 randomly selected images.

Table 1. Distributions of blur, twist, etc. of HME100K

Blur	Twist	Color	Complicated BG	Poor Illumination
7.8%	1.3%	12.4%	5.1%	5.3%

4. Effect of the Reverse Decoder and Regularization

As shown in the Tab. 2 on the results with/without the two terms, our method is competitive under both settings.

Table 2. Effects of reverse decoder and regularization

methods	rev. dec. & reg.	2014	2016	2019
DenseWAP	×	43.0	40.1	41.7
Tree Decoder	×	44.5	41.8	45.4
Tree Decoder	✓	49.1	48.5	51.4
Zhao *et al.*, ICDAR'21	×	48.2	44.6	45.0
Zhao *et al.*, ICDAR'21	✓	54.0	52.3	53.0
SAN	×	**53.0**	**52.0**	**52.1**
SAN	✓	**56.2**	**53.6**	**53.5**

References

[1] Jianshu Zhang, Jun Du, and Lirong Dai. Multi-scale attention with dense encoder for handwritten mathematical expression recognition. In *2018 24th international conference on pattern recognition (ICPR)*, pages 2245–2250. IEEE, 2018.