When Counting Meets HMER: Counting-Aware Network for Handwritten Mathematical Expression Recognition

Bohan Li[1,2] Ye Yuan[1] Dingkang Liang[2] Xiao Liu[1] Zhilong Ji[1] Jinfeng Bai[1] Wenyu Liu[2] Xiang Bai[2]

[1] Tomorrow Advancing Life [2] Huazhong University of Science and Technology

TEL AVIV 2022

Introduction

Problems

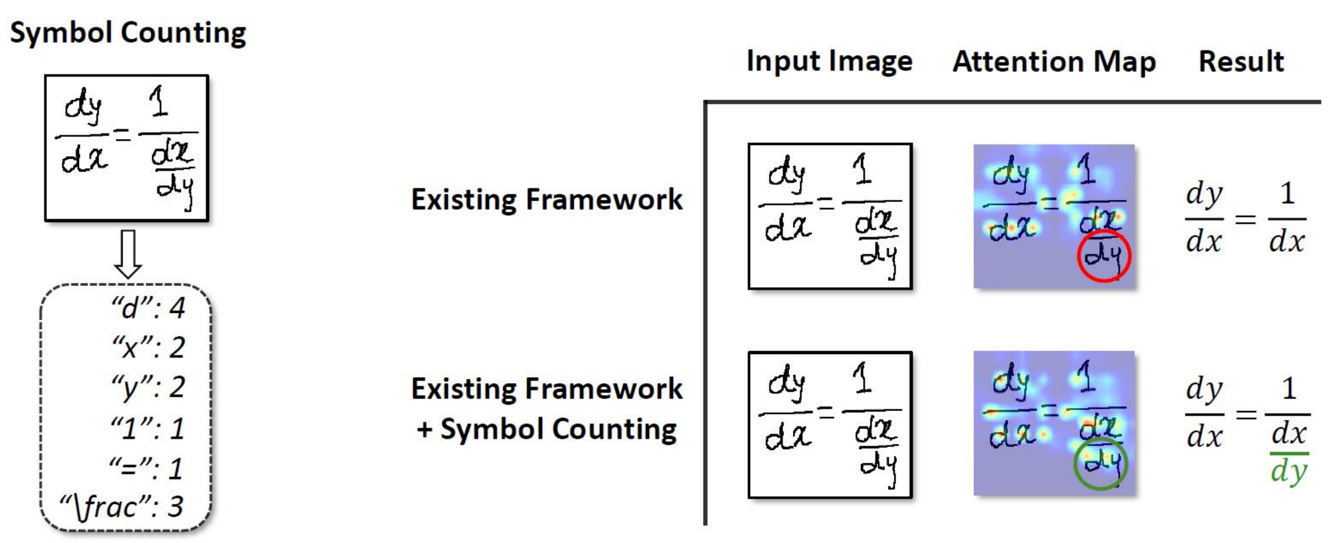
- Most existing handwritten mathematical expression recognition methods adopt the encoder-decoder networks, which directly predict the markup sequences from formula images with the attention mechanism.
- These methods may fail to accurately read formulas with complicated structure or generate long markup sequences, as the attention results are often inaccurate due to the large variance of writing styles or spatial layouts.

Goal

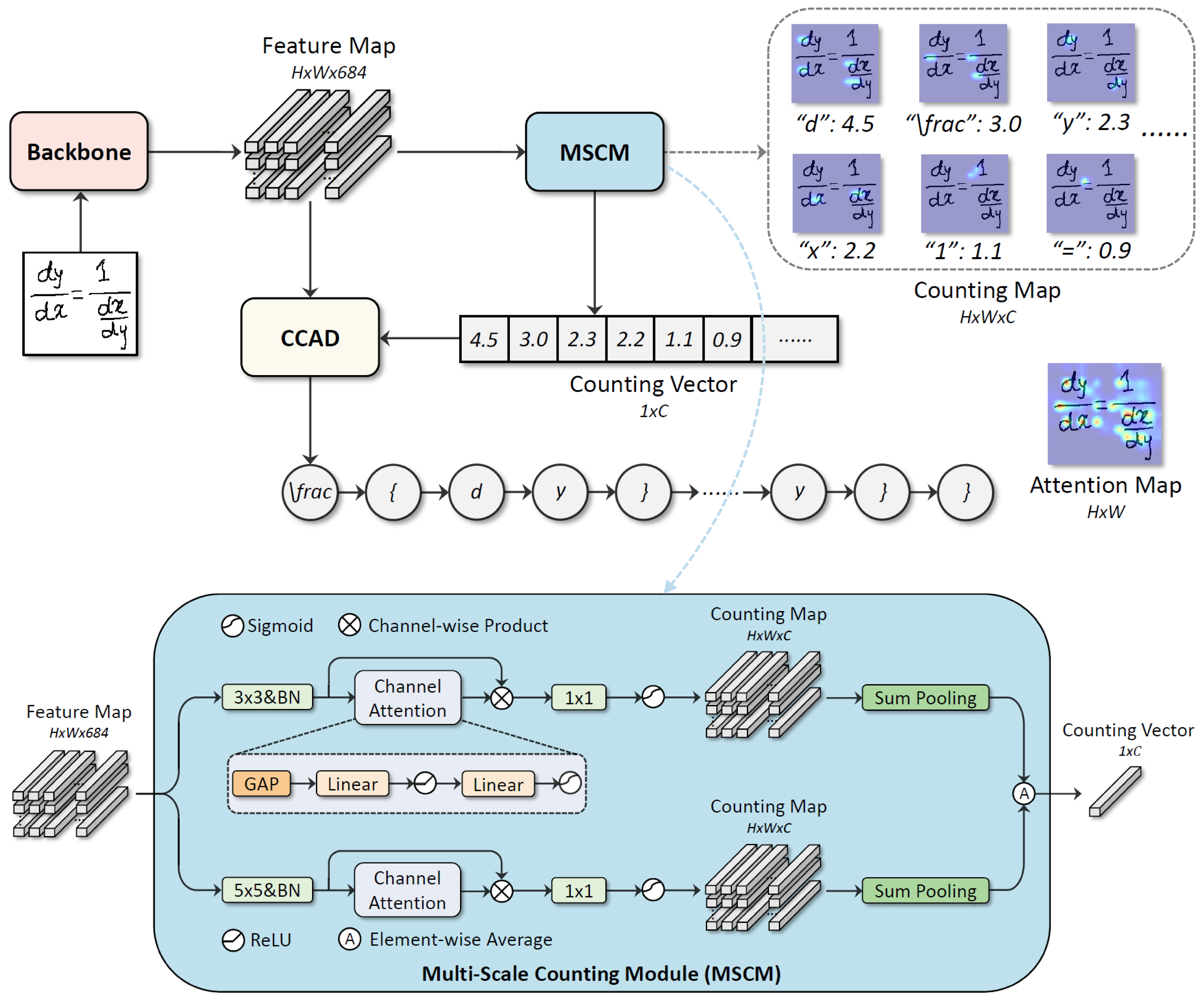
- Propose a new method that can generate more accurate attention results.
- Strengthen model's awareness for the positions and counts of each symbol class.

Contribution

- To the best of our knowledge, we are the first to bring symbol counting into HMER and reveal the relevance and the complementarity of HMER and symbol counting.
- We propose a new method that jointly optimizes symbol counting and HMER, which consistently improves the performance of the encoder-decoder models for HMER.
- We design a weakly-supervised counting module named MSCM, which can be easily plugged into existing encoder-decoder networks and optimized jointly in an end-to-end manner. With this counting module, an encoderdecoder model can be better aware of each symbol's position.

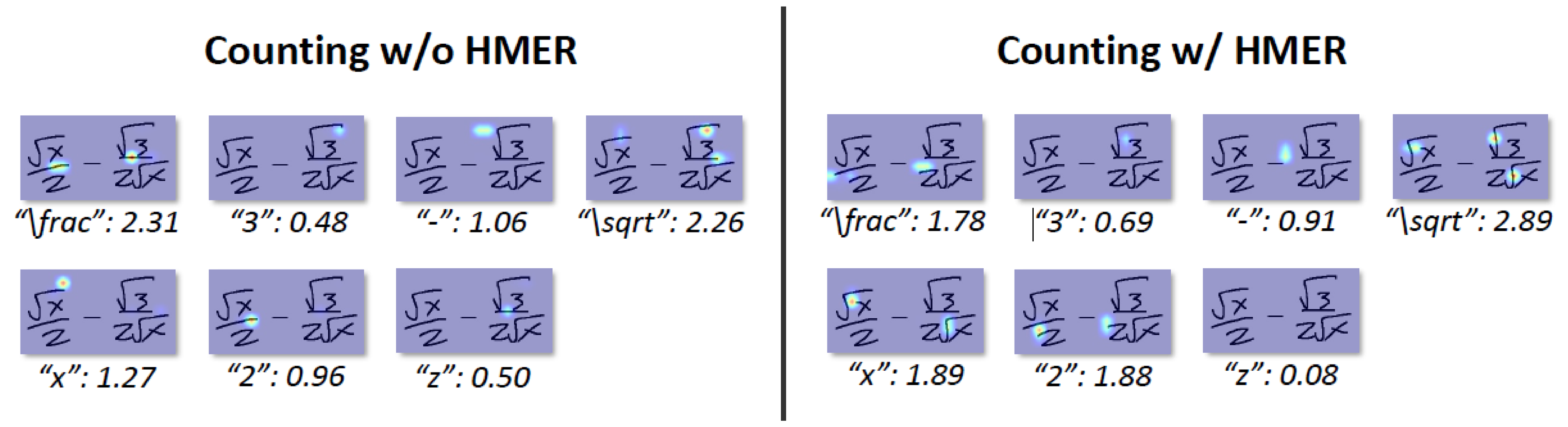


Architecture



HMER Boosts Counting

Method	CROHME 2014	
	$MAE_{Ave}\downarrow$	$MSE_{Ave}\downarrow$
Counting w/o HMER	0.048	0.044
Counting w HMER	**0.033**	**0.037**



Results

- Comparison with SOTA methods:

Method	CROHME 2014			CROHME 2016			CROHME 2019		
	ExpRate↑	$\leq 1\uparrow$	$\leq 2\uparrow$	ExpRate↑	$\leq 1\uparrow$	$\leq 2\uparrow$	ExpRate↑	$\leq 1\uparrow$	$\leq 2\uparrow$
Without data augmentation									
UPV [18]	37.22	44.22	47.26	-	-	-	-	-	-
TOKYO [19]	-	-	-	43.94	50.91	53.70	-	-	-
PAL [30]	39.66	56.80	65.11	-	-	-	-	-	-
WAP [43]	46.55	61.16	65.21	44.55	57.10	61.55	-	-	-
PAL-v2 [31]	48.88	64.50	69.78	49.61	64.08	70.27	-	-	-
TAP [41]*	48.47	63.28	67.34	44.81	59.72	62.77	-	-	-
DLA [14]	49.85	-	-	47.34	-	-	-	-	-
DWAP [40]	50.10	-	-	47.50	-	-	-	-	-
DWAP-TD [42]	49.10	64.20	67.80	48.50	62.30	65.30	51.40	66.10	69.10
DWAP-MSA [40]	52.80	68.10	72.00	50.10	63.80	67.40	47.70	59.50	63.30
WS-WAP [24]	53.65	-	-	51.96	64.34	70.10	-	-	-
MAN [28]*	54.05	68.76	72.21	50.56	64.78	67.13	-	-	-
BTTR [46]	53.96	66.02	70.28	52.31	63.90	68.61	52.96	65.97	69.14
ABM [1]	56.85	73.73	81.24	52.92	69.66	78.73	53.96	71.06	78.65
DWAP (baseline)†	51.48	67.01	73.30	50.65	63.30	70.88	50.04	65.39	69.39
CAN-DWAP (ours)	**57.00**	74.21	80.61	**56.06**	71.49	79.51	**54.88**	71.98	79.40
ABM (baseline)†	56.04	73.10	79.90	53.36	70.01	78.12	53.71	71.23	78.23
CAN-ABM (ours)	**57.26**	74.52	82.03	**56.15**	72.71	80.30	**55.96**	72.73	80.57
With data augmentation									
Li *et al.* [16]	56.59	69.07	75.25	54.58	69.31	73.76	-	-	-
Ding *et al.* [6]	58.72	-	-	57.72	70.01	76.37	61.38	75.15	80.23
DWAP (baseline)†	57.97	73.81	79.19	55.97	71.40	79.86	56.05	72.23	79.15
CAN-DWAP (ours)	**65.58**	77.36	83.35	**62.51**	74.63	82.48	**63.22**	78.07	82.49
ABM (baseline)†	63.76	76.35	83.05	60.86	73.93	81.17	62.22	77.23	81.90
CAN-ABM (ours)	**65.89**	77.97	84.16	**63.12**	75.94	82.74	**64.47**	78.73	82.99

- Qualitative Results :