LREC-COLING 2024

PEKING UNIVERSITY

Restoring Ancient Ideograph: A Multimodal Multitask Neural Network Approach

Siyu Duan[1,3], Jun Wang[1,3,4], Qi Su[2,3,4*]

[1] Department of Information Management, Peking University

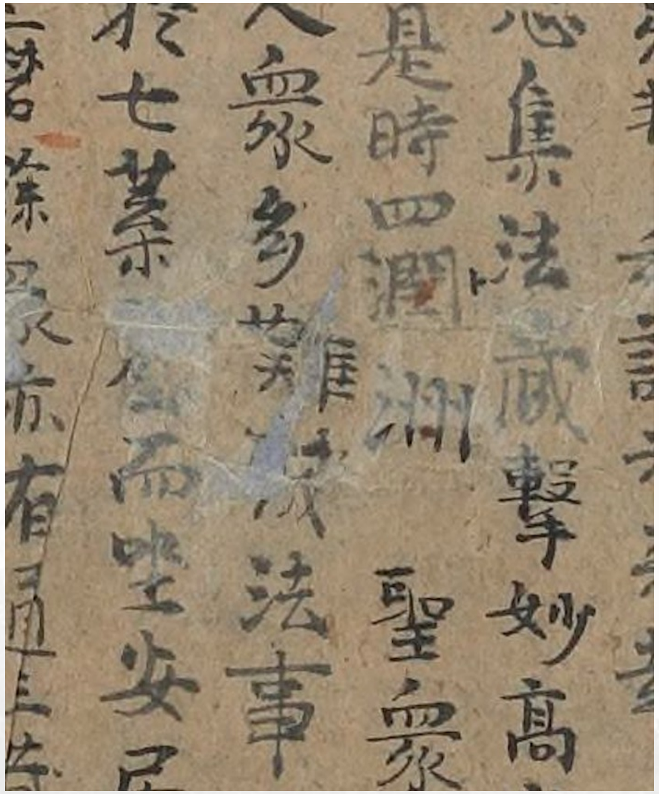
[2]School of Foreign Languages, Peking University

[3]Center for Digital Humanities, Peking University

[4]Institute for Artificial Intelligence, Peking University

{duansiyu, junwang, sukia}@pku.edu.cn

Damaged textual cultural heritage



Ideograph: character shape is linked to semantics

LREC-COLING 2024

Simulated data for training

Damaged context:

classical Chinese corpus

Train	Dev	Test	Max	Avg
575,398	10,000	10,000	50	14.4

$$w_i = \sqrt{\frac{1}{max(f_i, f_{avg})}}$$

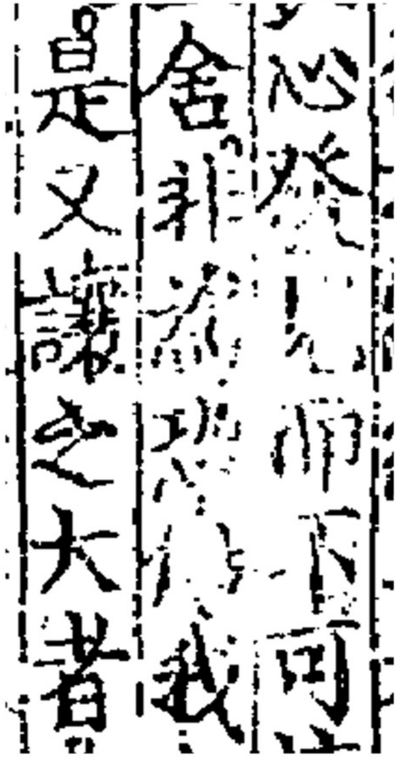
Damaged image:

108 kinds of font files

Font Image

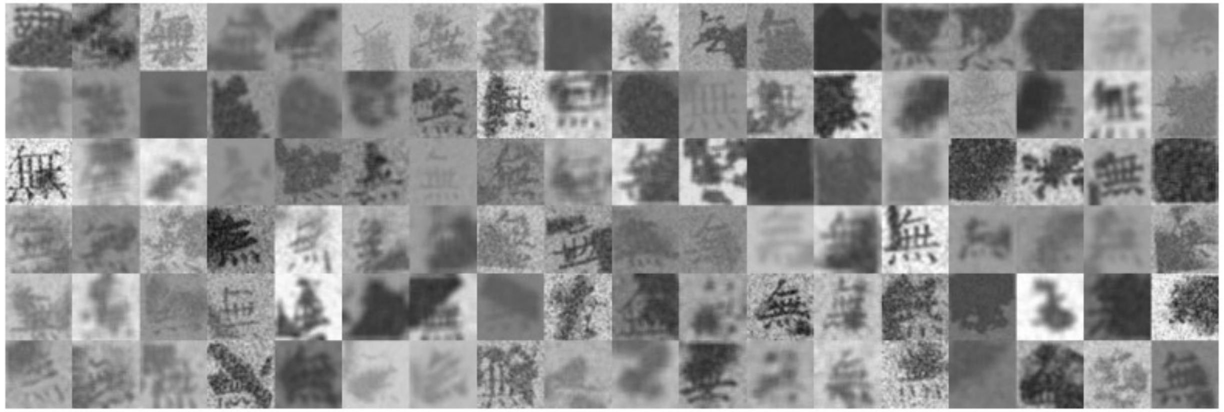
Additive Damage

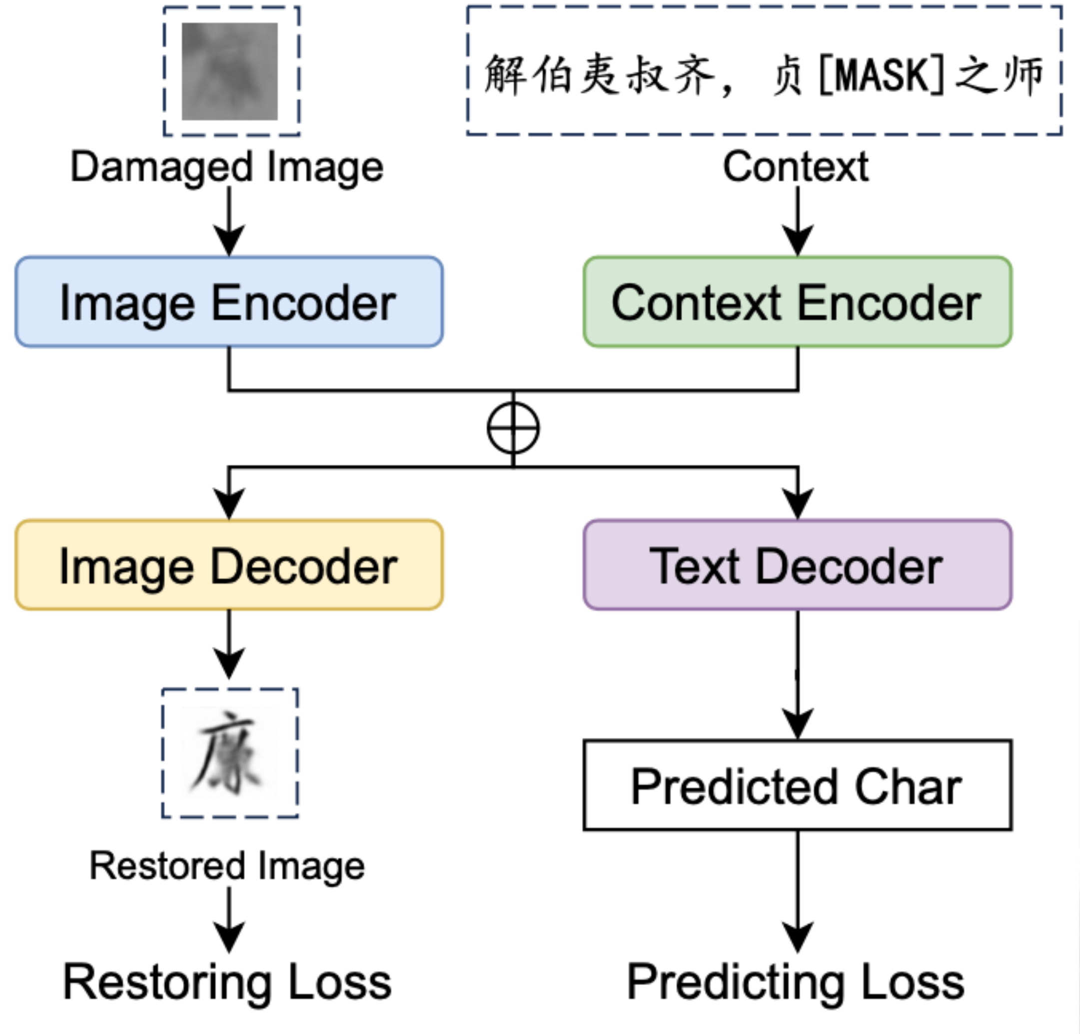
Fading Damage



Simulated Undamaged Image

Simulated Damaged Image





Model:

Input damaged image and context

$$memory = RoBERTa(Context)$$

$$x_1 = memory[i]$$

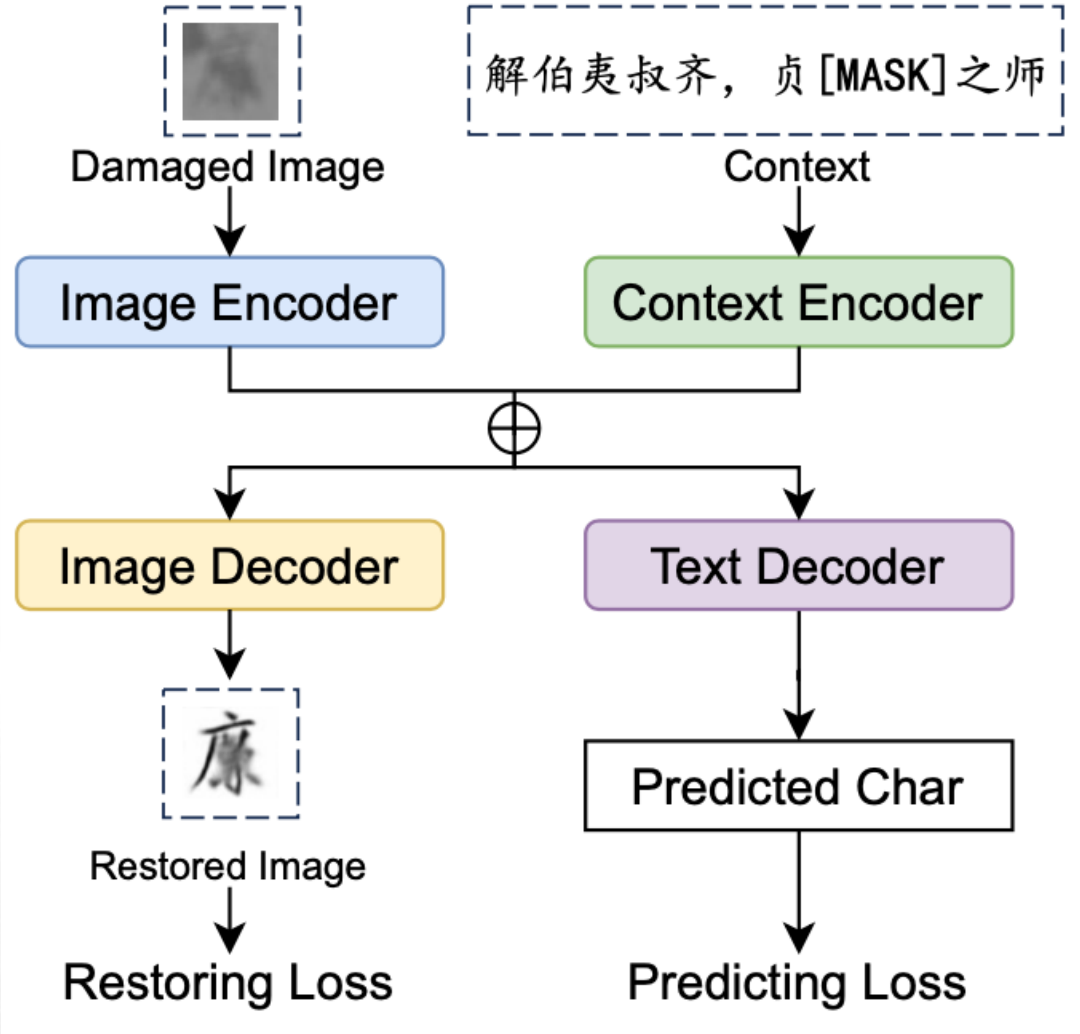
$$x_2 = Resnet50(Img)$$

$$x = x_1 + x_2$$

Output predicted char and restored image

$$Y_{pred} = FC(x)$$

$$Img_{res} = ConvT(x)$$



Two optimization goals

$$Loss_{res} = MSELoss(Img_{res}, Img_{font})$$

$$Loss_{pred} = CELoss(Y_{pred}, Label)$$

$$Loss = \alpha * Loss_{res} + Loss_{pred}$$

Curriculum Learning

incrementally enlarge the damaged area

Single-modal baselines, three metrics

- Average accuracy.
- Hits. The probability of the correct character being in the top k candidates.
- Mean Reciprocal Rank. The reciprocal of the rank of the correct character.

Model	Image	Text	LM ft	MT	CL	Acc	Hit 5	Hit 10	Hit 20	MRR
Img	+					66.00	79.43	83.33	86.55	72.18
LM		+				36.06	56.18	64.10	71.03	45.56
LM ft		+	+			44.75	66.07	73.23	79.48	54.57
MRM	+	+	+			86.74	94.16	95.61	96.87	90.09
MMRM	+	+	+	+		87.34	94.60	96.16	97.29	90.61
MMRM CL	+	+	+	+	+	**87.76**	**95.03**	**96.45**	**97.52**	**91.03**

Multiple characters damaged

Slightly drop, but still work well

Num	Acc	Hit 5	Hit 10	Hit 20	MRR
R	82.83	91.57	93.68	95.30	86.80
1	87.27	94.40	95.99	97.16	90.50
2	85.02	92.95	94.79	96.28	88.62
3	82.67	91.50	93.59	95.23	86.67
4	80.72	90.01	92.45	94.41	84.96
5	78.76	88.87	91.44	93.46	83.38

Cases from simulation experiments

Single-modal baselines:

- Confuse characters with similar shapes
- Suggest other characters that fit the context

Sentence	Proposal	Img	LM	LM ft
弗闕，則與之瑞節而以執之	瑞	端	同	旌
生死與道不相舍離，亦未曾即合	曾	增	尝	必
朝廷以克辰纂嚴，令輿夙駕	纂	篡	戒	告

Real-world Data:

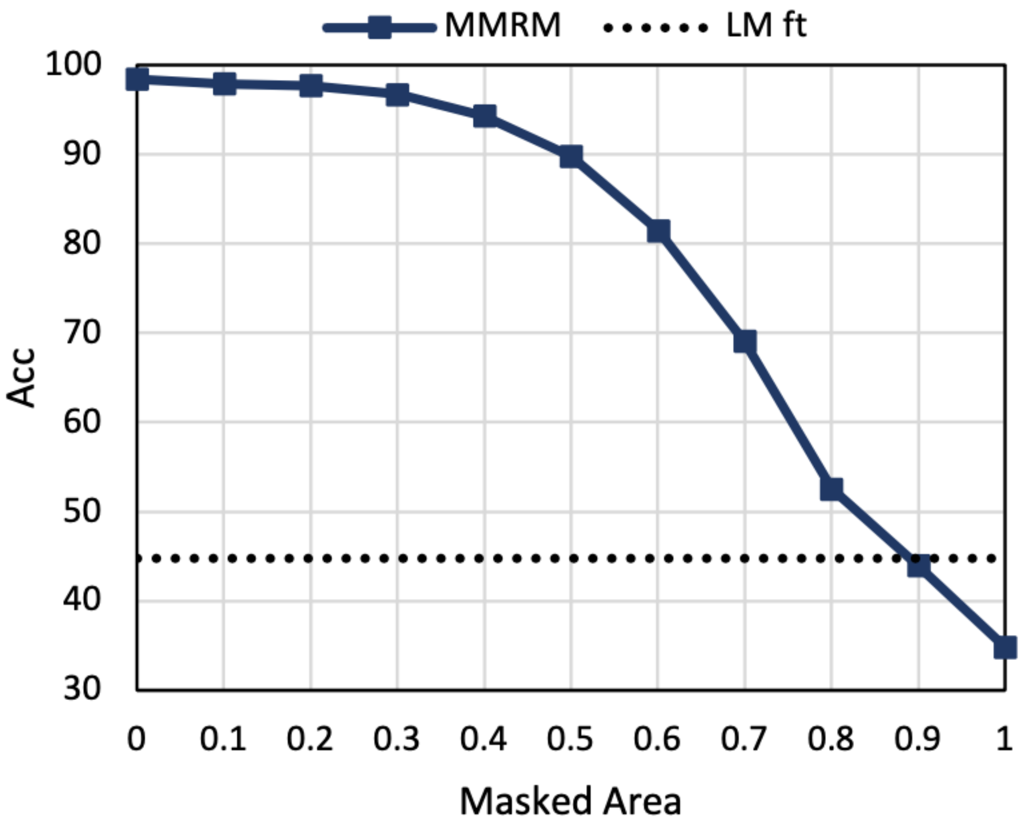
Inscription of Sweet Spring in Jiucheng Palace (九成宫醴泉铭)

Level 1 – 3: Better performance than LM

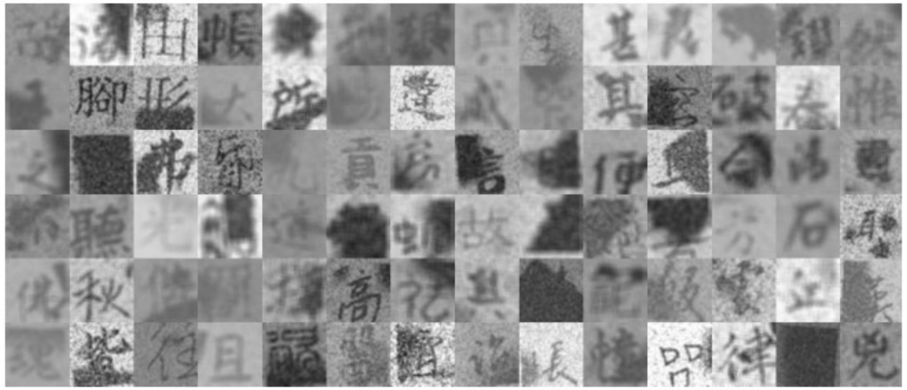
(no visual information) Level 4: Performance is not good

Level	Number	Count	Accuracy	Hit 5	Hit 10	Hit 20	MRR
I	5	5	100.00	100.00	100.00	100.00	100.00
II	17	11	64.70	82.35	82.35	94.11	71.17
III	8	5	62.50	87.50	87.50	100.00	70.83
IV	8	0	0	25.00	37.50	50.00	10.06
Total	38	21	**55.26**	**73.68**	76.31	**86.84**	**62.28**
LM	38	11	28.94	68.42	**84.21**	86.84	46.57
LM ft	38	13	34.21	73.68	78.95	84.21	50.32

LREC-COLING 2024



Damaged Image



Restored Image

LREC-COLING 2024

PEKING UNIVERSITY

Limitations

- generalizing to other ancient languages

Future work

- external database
- low-resource ancient language
- interactive tool