Show Me the World in My Language: Establishing the First Baseline for Scene-Text to Scene-Text Translation

Shreyas Vaidya*[1], Arvind Kumar Sharma*[1], Prajwal Gatti[2], Anand Mishra[1]

[1]Indian Institute of Technology Jodhpur, [2]University of Bristol

(*:equal contribution)

Code & Data Available

① What is in-image translation?

Suppose you are visiting Delhi, India and arrive at the Rithala (Hindi: रिठाला) metro station. If you are not familiar with Hindi, the signboard on the left might be incomprehensible.

② Why it is a hard problem?

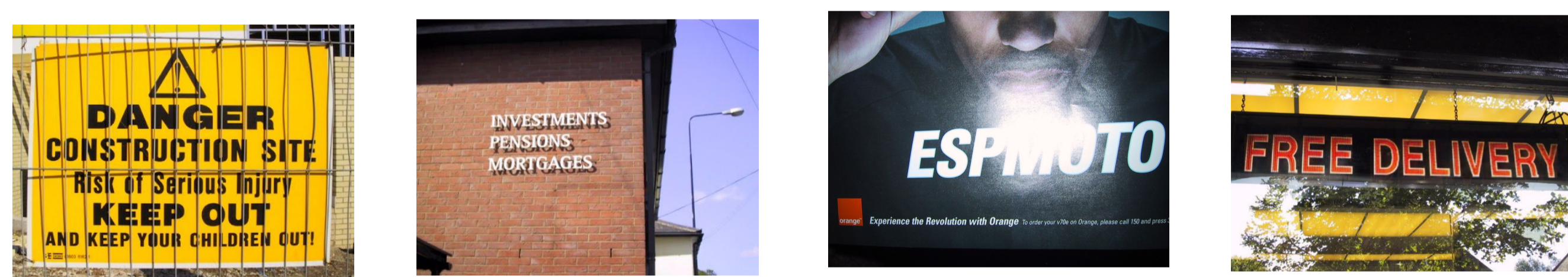### (i) Translate or Transliterate: Limited Context

(ii) Word order in source and target may change

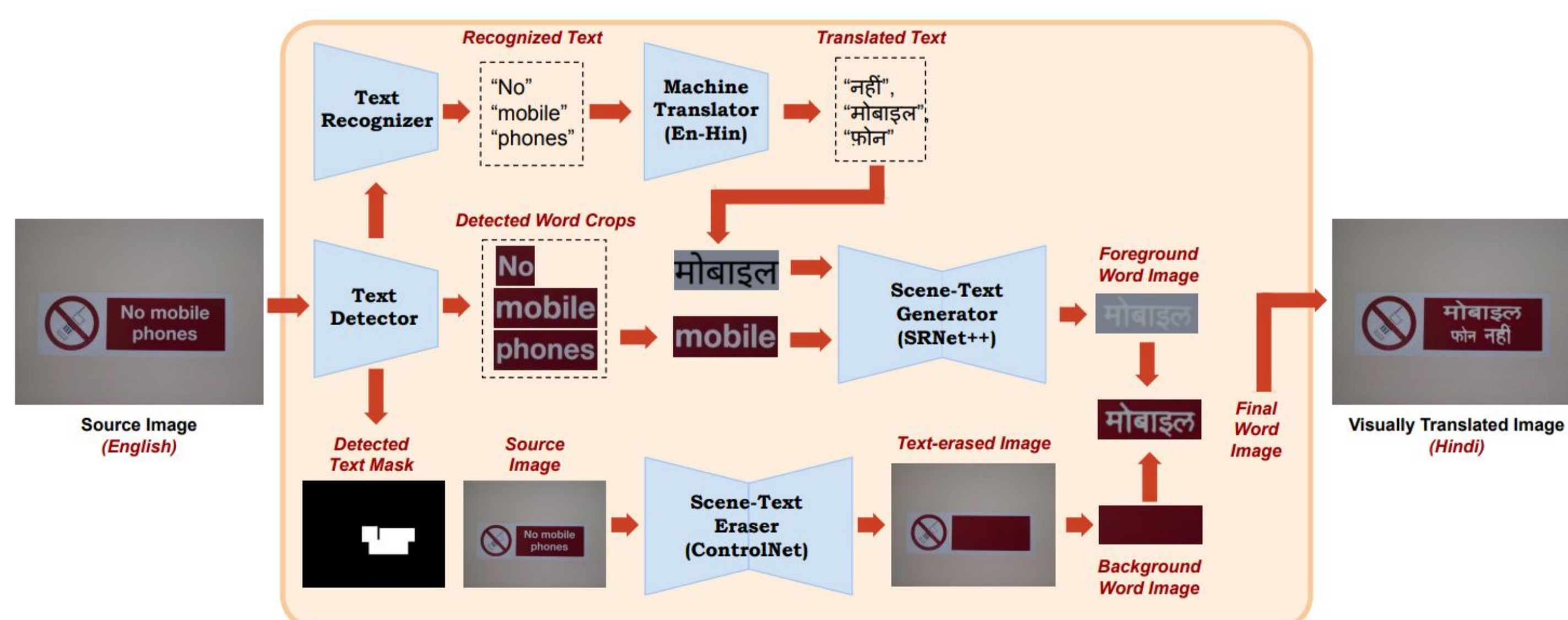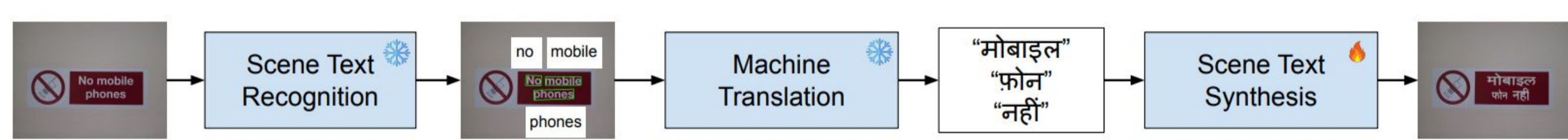### (iii) Source and target length may vary significantly

(iii) Imaging Related Challenges

③ First Open Source Baseline

Commercial tools such as Google Translate exist, but no open source baseline

- Implemented 7 variants of baselines by using different choices in the pipeline
- Modified SRNet by including more robust text erasing technique (SRNet++)
- Implemented several task specific design enhancements

④ Dataset for the Task

Training Data
Scene Text Synthesis Completely trained using synthetic data (600K pair)

Evaluation Dataset: VT-Real

User 1: कोलचेस्टर डाक घर
User 2: कोलचेस्टर पोस्ट ऑफिस
User 3: कौलचेस्टर पोस्ट ऑफिस

269 real scene images translated by three annotators for evaluation

## ⑤ How to evaluate performance?

Automatic Evaluation:
- Translation Quality (BLEU scores)
- Perception Quality (PQ) using image quality evaluators
- VT-score as a combined metric of TQ and PQ

User Evaluation:
Human evaluators rated Translation accuracy, Readability, Perceptual quality, Style preservation

⑥ Results

Method	STR	MT	STS	D.E.	TQ (BL-1)	TQ (BL-2)	PQ	VT-score
English-to-Hindi Translation								
B-7	DBNet+ParSeq	Indic	SRNet++	✓	**25.28**	**20.54**	53.79	**27.51**
B-6	DBNet+ParSeq	Indic	SRNet++	✗	22.57	15.69	53.93	25.59
B-5	Oracle	Indic	SRNet++	✗	22.36	16.90	53.38	23.95
B-4	DBNet+ParSeq	M2M	SRNet++	✗	19.09	14.51	**54.02**	21.52
B-3	Oracle	M2M	SRNet++	✗	19.82	15.33	53.52	22.22
B-2	Oracle	M2M	Mostel	✗	14.13	10.44	46.98	16.58
B-1	Oracle	M2M	SRNet	✗	15.00	12.25	46.71	16.56
Hindi-to-English Translation								
B-7	Oracle	Indic	SRNet++	✓	**38.30**	**29.30**	55.49	**40.08**
B-6	DBNet+ParSeq	Indic	SRNet++	✗	29.10	18.51	**55.77**	28.52
B-5	Oracle	Indic	SRNet++	✗	31.31	19.70	55.62	32.27
B-4	DBNet+ParSeq	M2M	SRNet++	✗	03.22	02.19	55.60	03.81
B-3	Oracle	M2M	SRNet++	✗	04.20	02.89	55.58	04.97
B-2	Oracle	M2M	Mostel	✗	02.03	01.40	53.41	02.46
B-1	Oracle	M2M	SRNet	✗	04.20	02.86	53.82	04.92
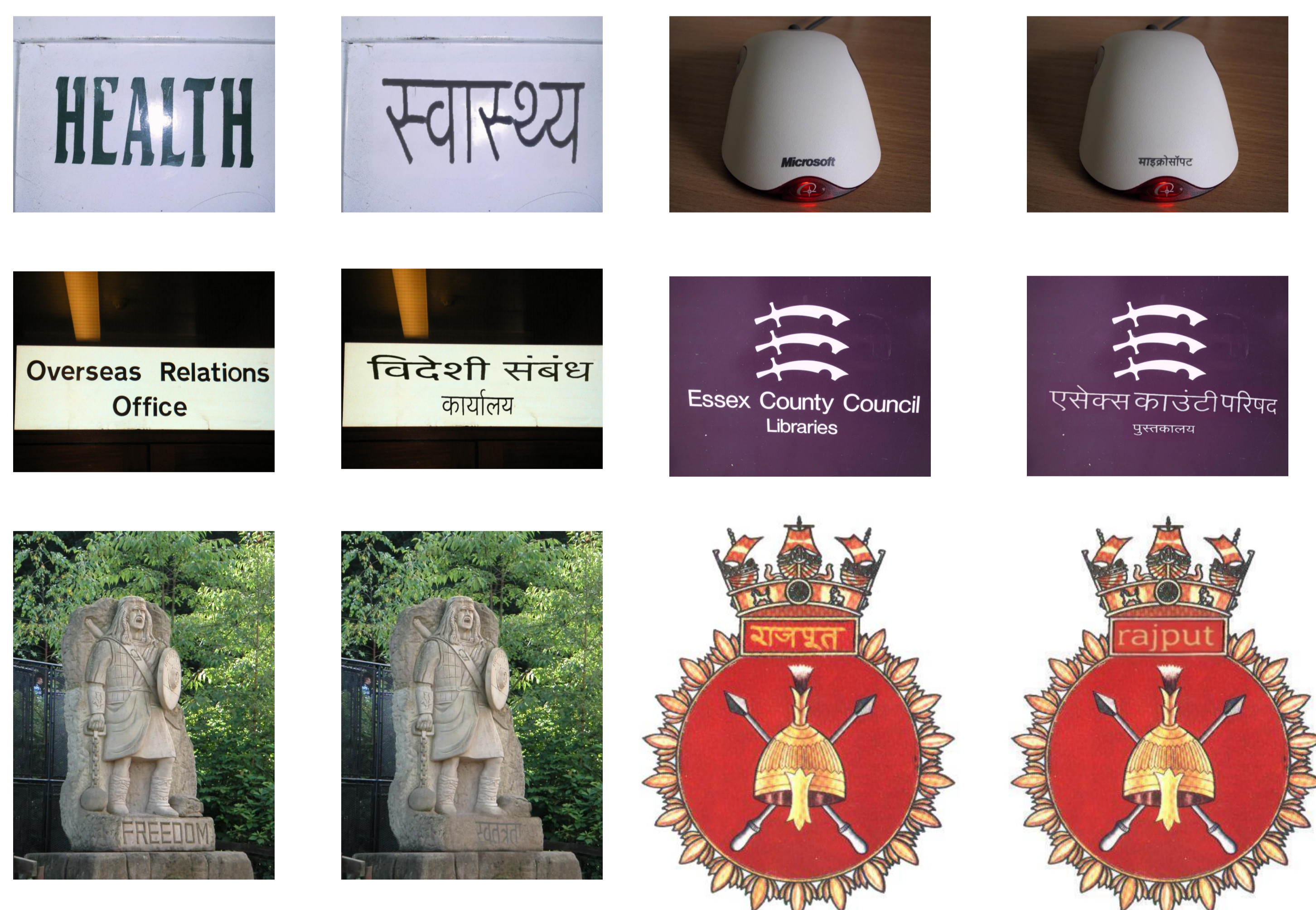
For more details, please visit our project website:
https://vl2g.github.io/projects/visTrans/