#2853

Two-Stage Seamless Text Erasing On Real-World Scene Images

Benjamin Conrad[1] Pei-i Chen[2]
[1] University of Amsterdam [2] Jumio AI Labs

Task:

Text erasing is the task of removing all text found in an image and filling in the background pixels.

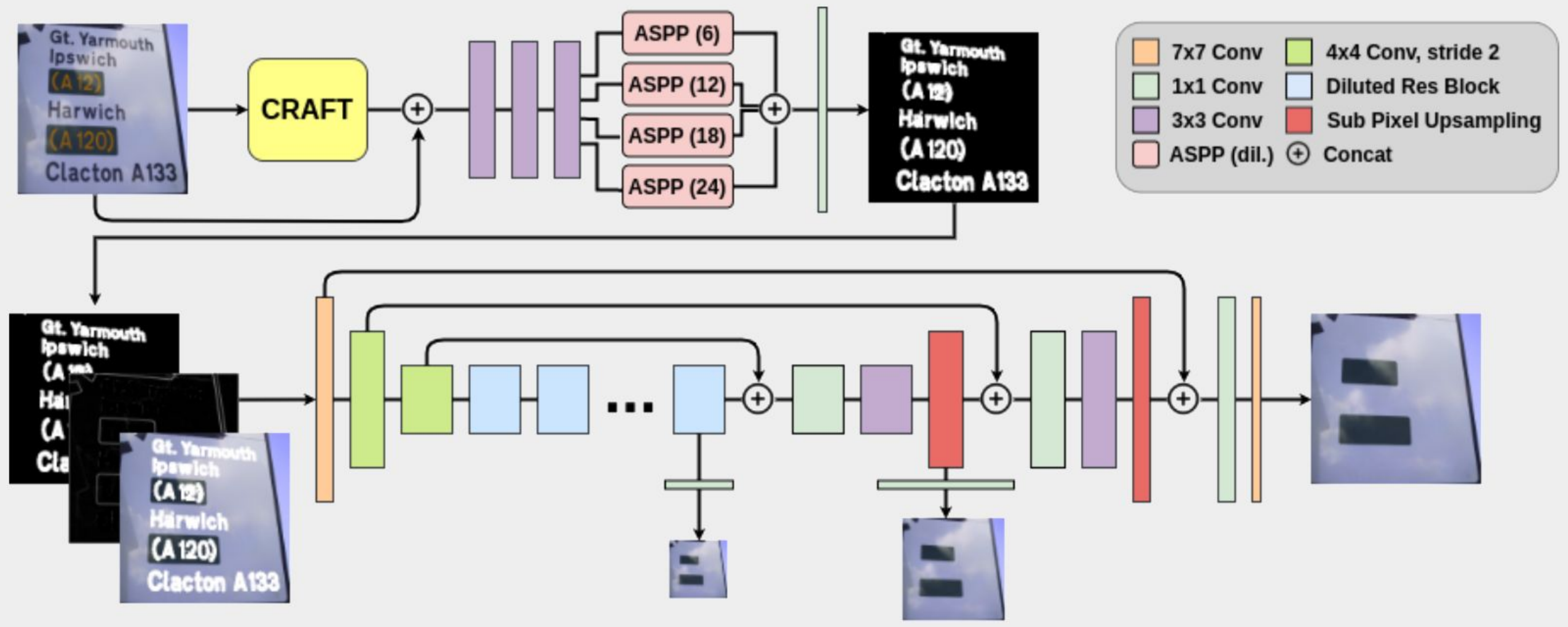
Challenges:

- Previous approaches perform poorly on real-world images:
  - Loss of fine detail
  - Inconsistent background colors
  - Failing to erase all text
- Fine detail inpainting requires computationally expensive models or complicated training procedures.

Contacts:

- benjamin.conrad@student.uva.nl
- pei-i.chen@jumio.com

Methodology



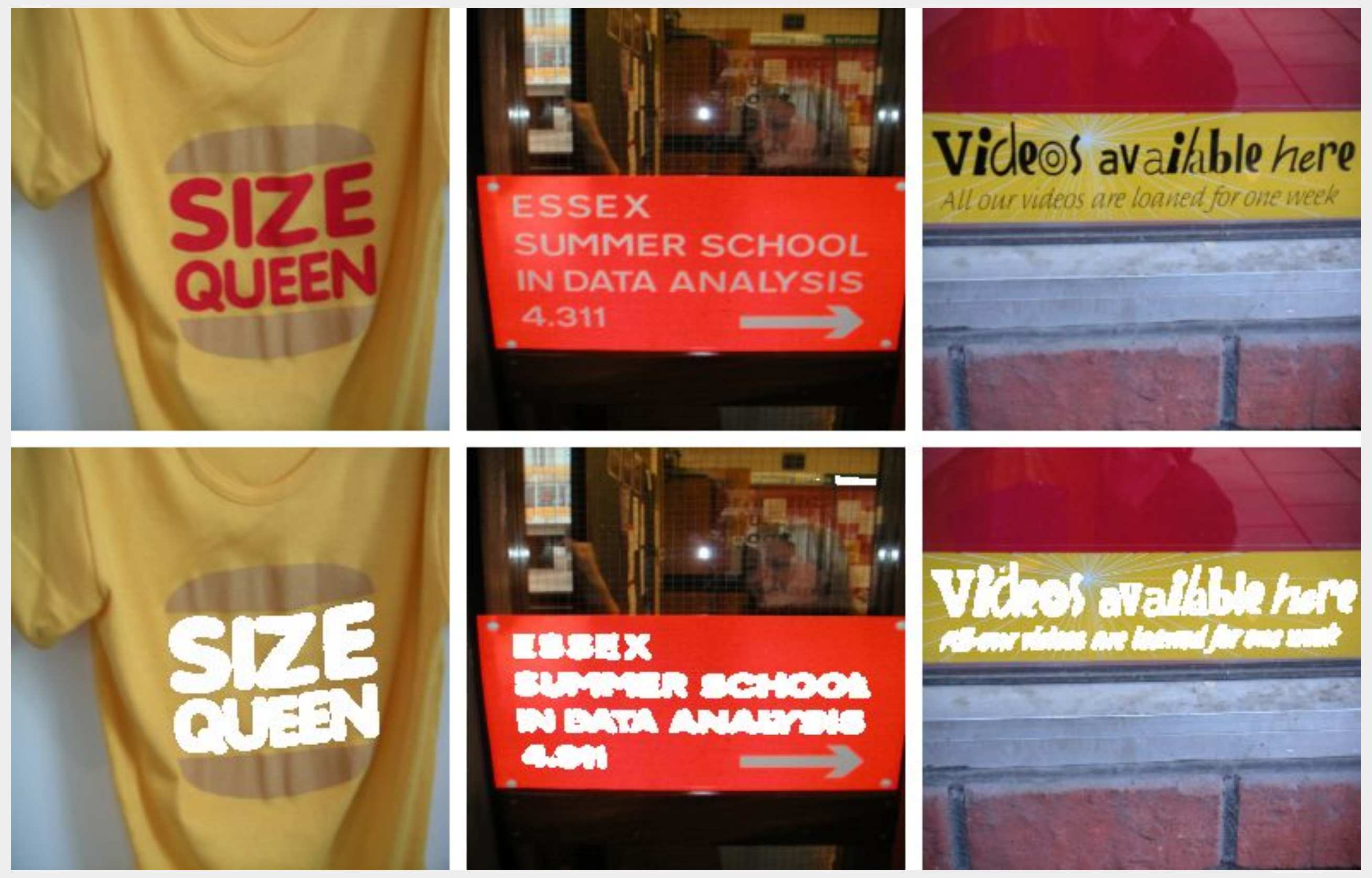
Stage 1: Text Mask Generator

- Generates a binary segmentation mask covering all pixels that contain a character.
- Network consists of a CRAFT text detector with a segmentation network head.
- Trained with Tversky loss to penalized false negatives more than false positives and ensure masks entirely cover each character.

Stage 2: Inpainting Model

- GAN model takes the masked image, image gradients and generated mask as input and produces a text-free version of the original image.
- Builds off the baseline encoder-decoder architecture by incorporating skip connections, sub-pixel upsampling and multiscale inpainting.
- Trained with novel multiscale gradient reconstruction loss to generate fine details and smooth surfaces without any significant computation cost.

$$L_{gr} = \frac{\sum_i \|\nabla I_{pred_i} - \nabla I_i\|_2^2}{S_i}$$


SIZE QUEEN
ESSEX SUMMER SCHOOL IN DATA ANALYSIS 4.311
Videos available here
All our videos are loaned for one week

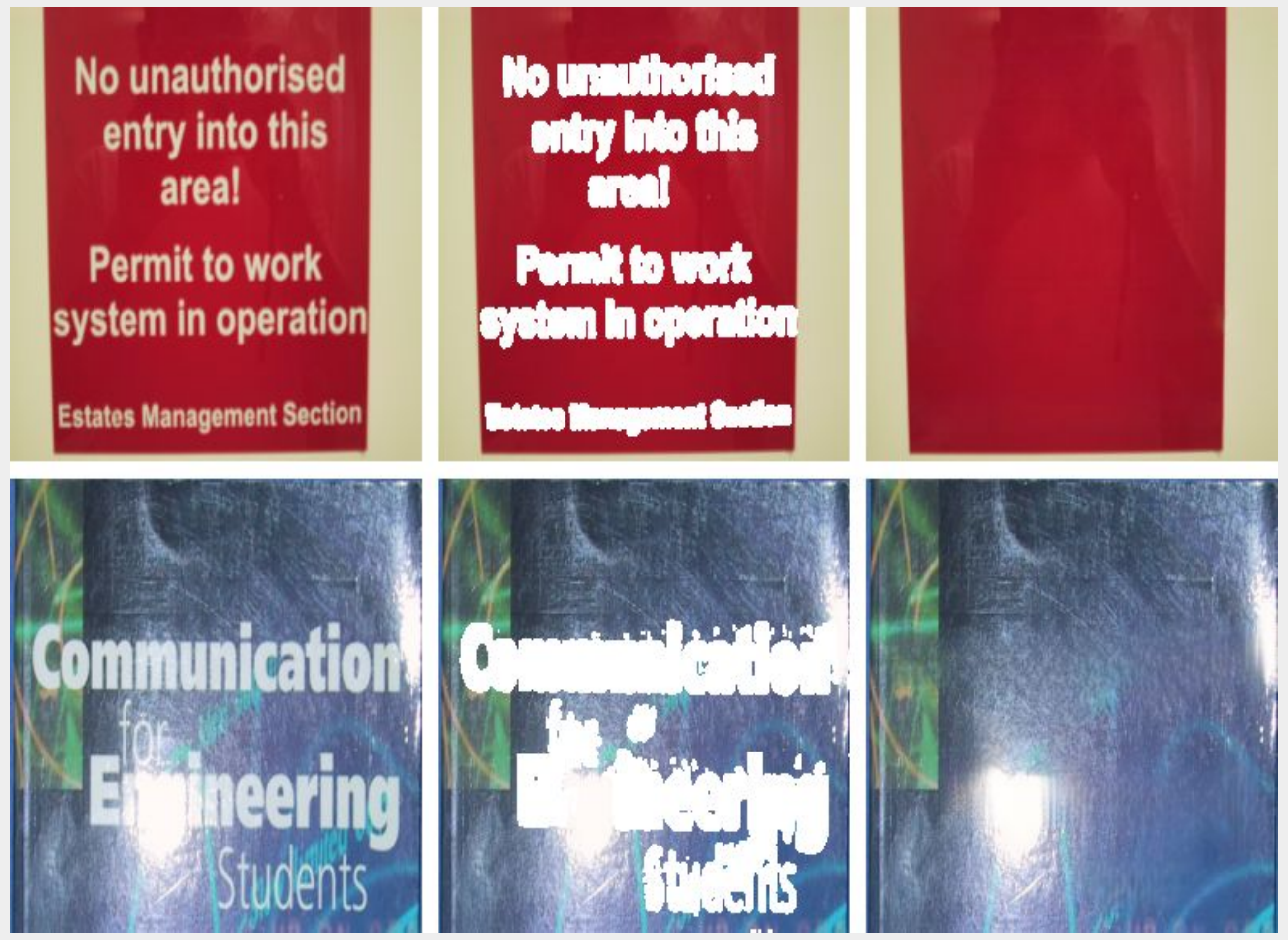
Results

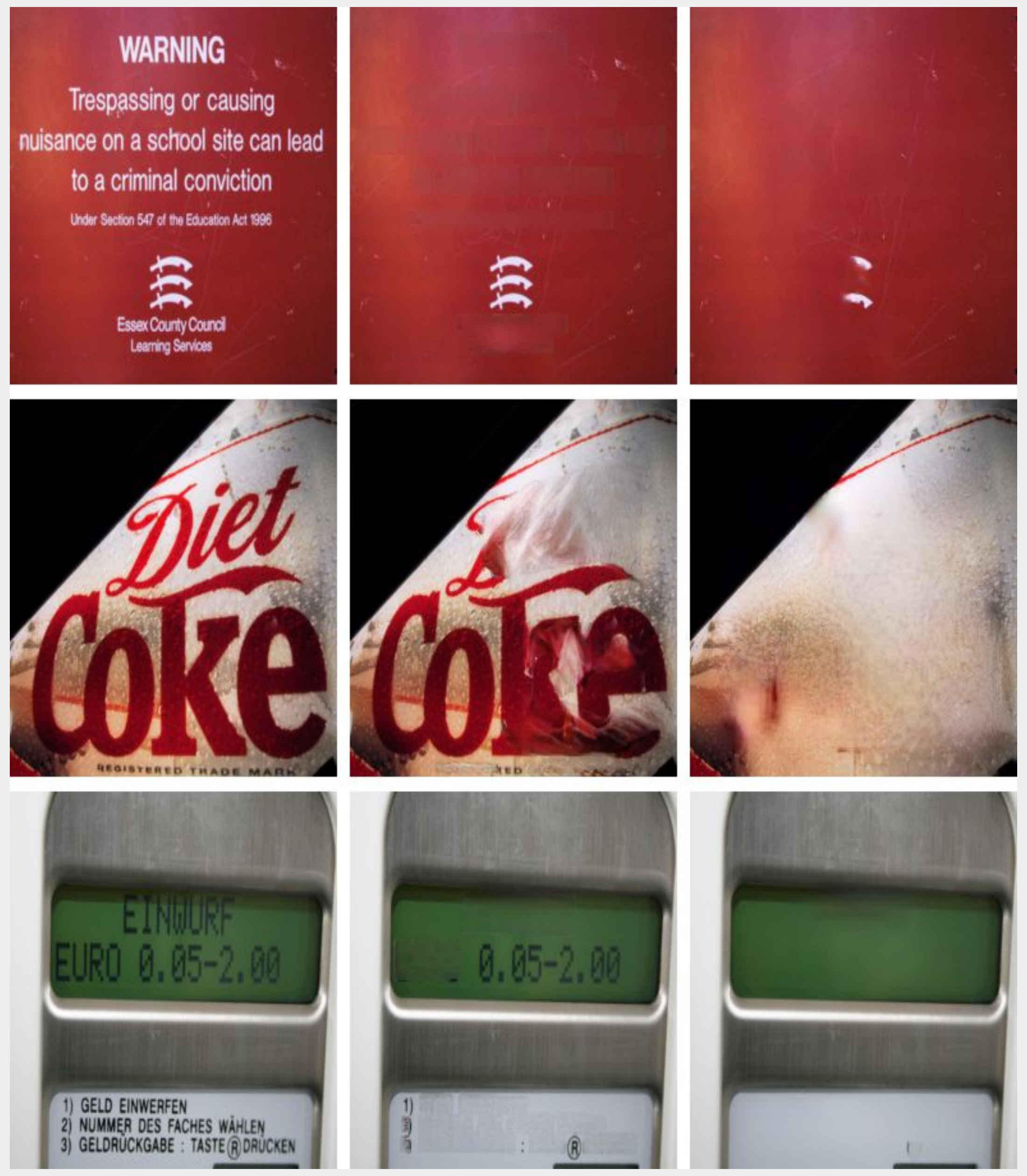
Method	PSNR†	SSIM†	MAE*	Reported PSNR†	Reported SSIM†
EnsNet [2]*	31.18	91.12	0.018	37.36	96.44
MTRNet [3]	30.56	90.14	0.021	29.71	94.43
MTRNet++ [4]	**33.43**	93.10	0.015	34.55	**98.45**
WS-TE (ResNet-50) [5]*	30.73	93.43	0.016	37.44	93.69
WS-TE (ResNet-152) [5]	-	-	-	**37.46**	93.64
Ours	32.97	**94.90**	**0.013**	32.97	94.90

Results on SCUT synthetic text erasing dataset

Method	Recall
Nakamura *et al.*[1]	10.08
EnsNet [2]	5.66
MTRNet [3]	29.11
WS-TE (ResNet-50) [5]	2.47
WS-TE (ResNet-152) [5]	0.64
Ours	**0.55**

Results on ICDAR 2013 text detection benchmark


No unauthorised entry into this area!
Permit to work system in operation
Estates Management Section
Communication for Engineering Students


WARNING
Trespassing or causing nuisance on a school site can lead to a criminal conviction
Diet Coke
EINWURF
EURO 0.05-2.00

Method	# Images	% Votes
WS-TE (ResNet-50) [5]	12	18%
Ours	**213**	**82%**
Tie	8	-

Results of human perceptual study

➔ **Matches SOTA on synthetic datasets.**

➔ **Significantly preferred over previous SOTA on real-world images.**