

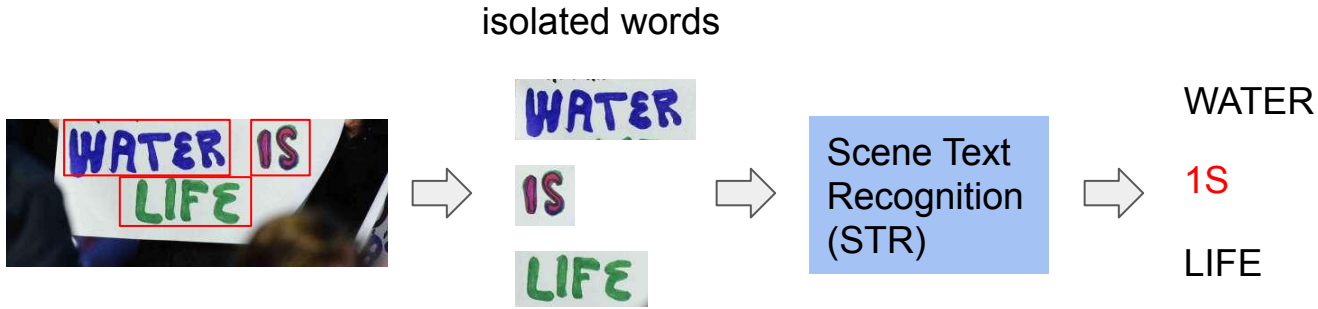
Semantic-based Sentence Recognition in Images Using Bimodal Deep Learning

Yi Zheng¹, Qitong Wang², and Margrit Betke¹

¹Computer Science Department, Boston University

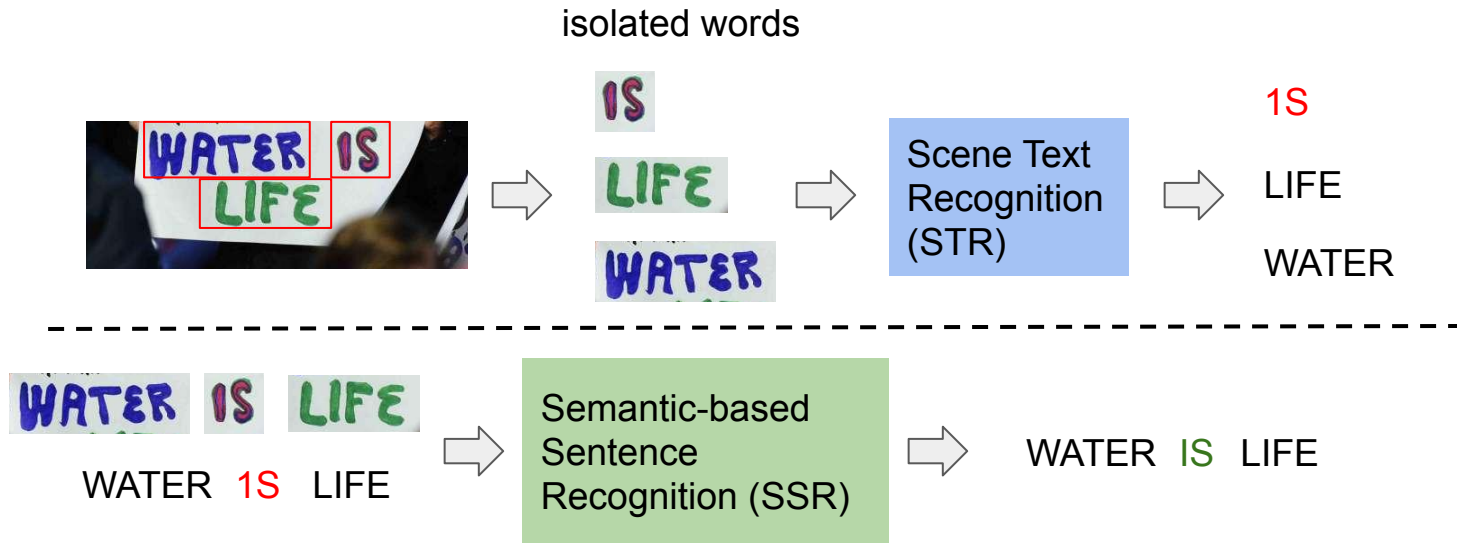
²Department of Electrical and Computer Engineering, Virginia Tech

Background



Scene text recognition is the process of converting text regions into computer readable and editable symbols.

Objective



Semantic coherence within a region of text makes seemingly isolated words easier to recognize.

Datasets

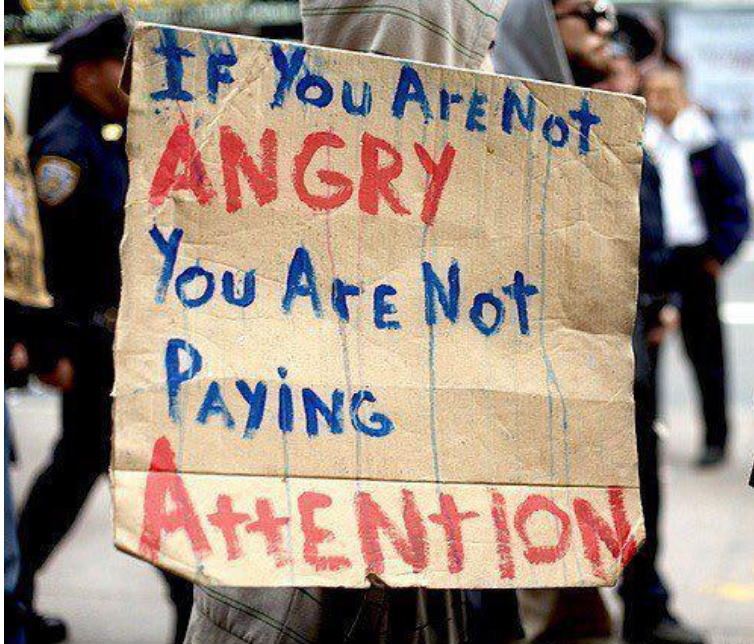
Samples of standard benchmark dataset:

The word "Cure" is written in a stylized, cursive font with a slight shadow effect, appearing to be part of a logo or sign.A close-up of a sign with the letters "HOU" in a bold, sans-serif font, set against a background of green foliage.A close-up of a sign with the word "SPEED" in a bold, sans-serif font, with a yellow background and a red border.A close-up of a sign with the word "FITNESS" in a bold, sans-serif font, with a blue background and a white border.The Starbucks logo, featuring the word "COFFEE" in a bold, sans-serif font, with a green background and a white border.

State-of-the-art STR use standard benchmark datasets such as **III5k-words**, which contains test images collected from the Web, and/or **Street View Text**, collected from Google Street View (and thus mostly contains sparse text regions).

Datasets

Text-containing Protest Image Dataset (TPID)



THE THINGS WE SAY.COM

Interior Design Dataset (IDD)



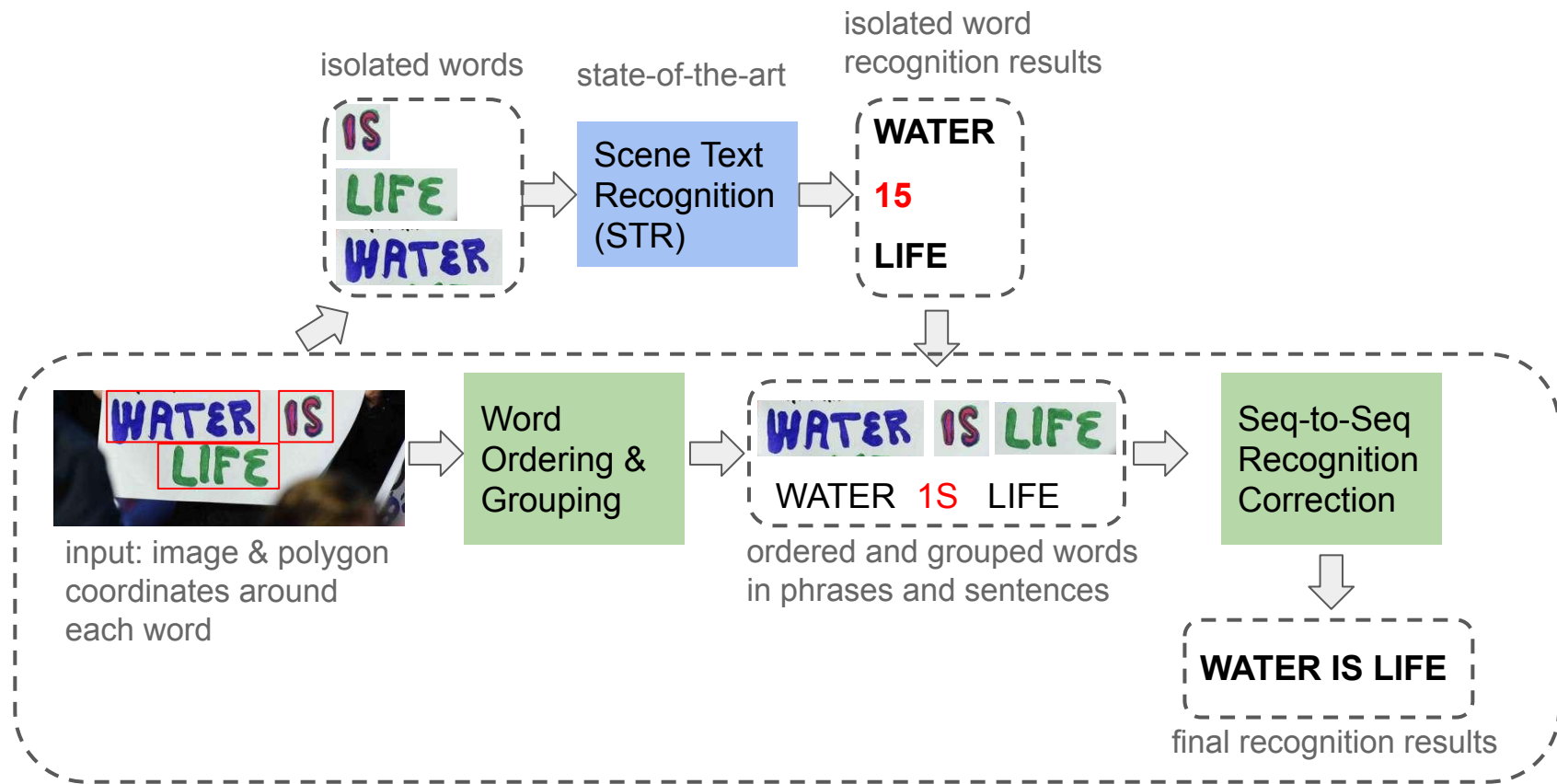
Visit our gift shop Accenten, for beautiful tableware at beautifully low prices. There isn't room to show everything here, so make sure you have a good browse when you get to the store. After all, with our low prices, you can afford a few extras.

1. AVANS candlesticks. White lacquered pine. High \$3. Low \$2.
2. CYLINDER 2. Candlesticks. 10 cm (3 7/8"). White \$1.75. Chrome \$1.75. 15 cm (5"). White \$2.25. Chrome \$2.25. 16 cm (6 1/4"). White \$2.50. Chrome \$2.50. 3. Dish 15 cm (5") \$2.
4. ALLIANS. White glazed feldspar. 4. Bowl \$2.
5. Candleholder/ashtray 4 pack \$3.25.
6. Candlestick \$3.25.
7. Small candlesticks 4 pack \$2.25.
8. Vase, large flat \$4.75.
9. Vase, small flat \$1.75.
10. Duck \$7.
11. Metal candlestick 4 pack Blue, green, red, white or yellow. \$2.
12. Heating candles 6 cm (2 1/4"). 4 pack \$1.
13. Ball candles 4 pack \$3.
14. Heating candles 30 pack \$1.50.

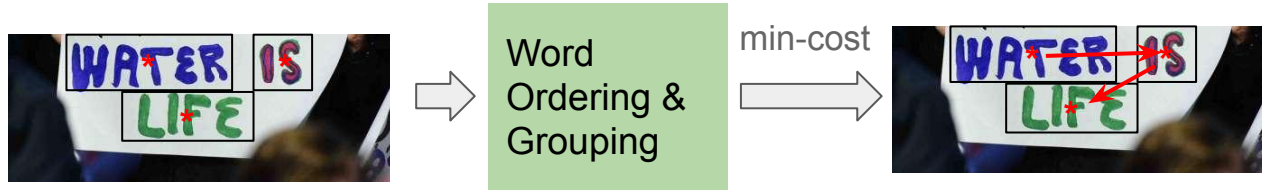
PROPER tablecloth. 150x260 cm (59x102 1/4"). 100% cotton. In beige, blue, grey and pink. Machine wash \$13.
ELT paper serviettes. 20 pack \$1.

PROPER tablecloth.

Methodology

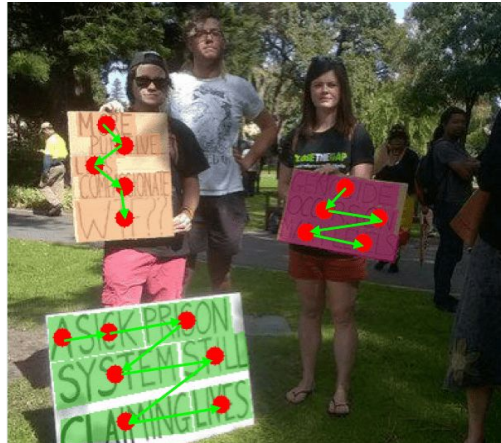
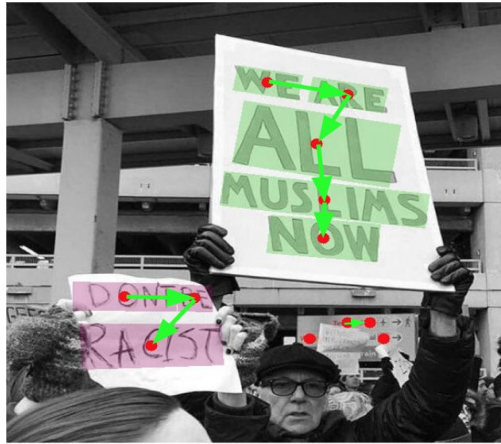


Word Ordering & Grouping

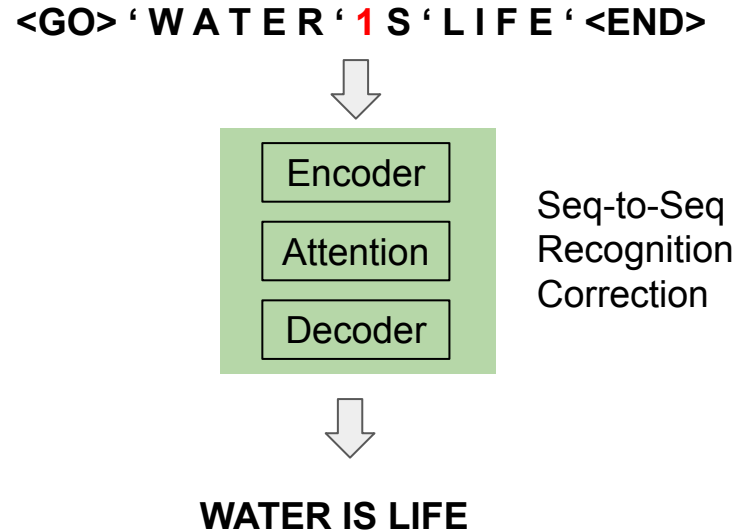


Word Ordering & Grouping arranges isolated words into the common reading order, and groups them into phrases or sentences that belong together.

Word Ordering & Grouping Results



Semantic-based Sentence Recognition



solve the task of correcting the words predicted by STR in the phrases or sentences.

Results

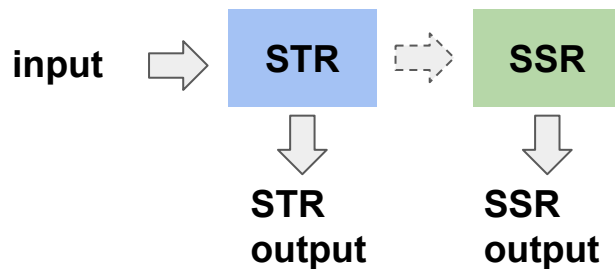
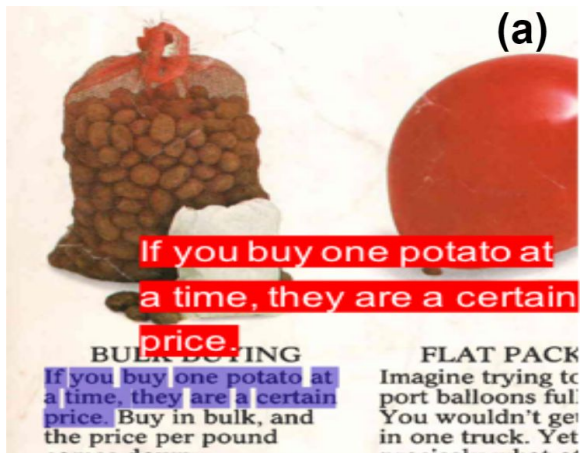


Table 1: Accuracy (%) of the CTC and attention-based versions of our SSR method and the corresponding STR baseline methods when tested on datasets IDD, TPID.

	IDD	TPID
CTC-based STR	85.36	65.63
CTC-based SSR	90.04	71.09
Attention-based STR	93.80	78.65
Attention-based SSR	96.34	89.39