

Text is an important way of communication and comprehension for humans. Text

Incorrect order of

appears everywhere. Text can be viewed in documents, images, videos, etc. Ex-

tracting text from images and videos can provide extra information which helps with

understanding more about them. This idea can be utilized in different computer vi-

sion tasks, such as image-based search [1], [2], robots navigation [3], [2] and industrial

automation [4], [2]. Optical Character Recognition (OCR) is the process of converting

an image of text into a machine-readable text. By having machine-readable text, the

mentioned computer vision problems can become easier to solve. Recognizing text

in natural scenes, also known as Scene Text Recognition (STR), is usually consid-

Long sentence and hard to follow, i

ered as a special form of OCR which is camera-based OCR [5]. Text detection and

recognition in images, which is the main focus of STR, initially involved researchers

designing features by hand which is slow and inaccurate mostly because of character-

istics such as complex text background, varying text, environment noise, etc. Deep

learning methods help profoundly in tackling STR. These methods can automatically

"So",... does not sound acca

extract insightful features and handle the mentioned characteristics much better. So,

the whole process will be faster and text detection/recognition will be more accurate

[2].

There are fundamental problems in the field of STR: text localization, text verifi-

cation, text detection, text segmentation, text recognition, end-to-end systems, text

enhancement and text tracking. Text localization aims to localize text components

and group them into candidate text regions. Text localization focuses on text and

The word "text" is repeated too many times. The sentence

Recursive definition. It doesn't explain why

Same as Text Localizaion abo

tries to include as little background as possible. Text verification focuses on verifying

text candidate regions as text or non-text. Text verification usually follows text lo-

calization to reduce false positives. Text detection determines whether text is present

using text localization and text verification techniques. Text detection focuses on

text presence. Text segmentation is a challenging problem which not only determines

text presence in an image but also tries to find text location in the image. Text

recognition translates a text instance image into a target string sequence which is

the goal of STR. End-to-end systems can directly convert all text regions into target

string sequences. These systems usually include text detection, text recognition, and

postprocessing. Text enhancement recovers degraded text, improves text resolution,

removes the distortions of text, removes the background, etc. Text enhancement can

be used as a preprocessing method and help considerably in text recognition. Text

tracking centers around maintaining text location integrity and tracking text across

adjacent frames in a video [5]. There are proposed solutions to each of these problems

but still each of these problems are active areas of research.

Deep learning methods have achieved the best results when tackling these prob-

lems and are considered the default. However, there is still so much room for im-

provement. STR is still a non-trivial and challenging task and researchers are looking

for possible enhancements.

What does "follows" re

Repetitive, lack of flow, c

uses casual lan

Needs some evidence

# Bibliography

- [1] G. Schroth, S. Hilsenbeck, R. Huitl, F. Schweiger, and E. Steinbach. Exploiting text-related features for content-based image retrieval. In *Proceedings of the 2011 IEEE International Symposium on Multimedia, ISM '11*, page 77–84, USA, 2011. IEEE Computer Society.
- [2] Shangbang Long, Xin He, and Cong Yao. Scene text detection and recognition: The deep learning era. *Int. J. Comput. Vision*, 129(1):161–184, January 2021.
- [3] Ruth Schulz, Benjamin Talbot, Obadiah Lam, Feras Dayoub, Peter Corke, Ben Upcroft, and Gordon Wyeth. Robot navigation using human cues: A robot navigation system for symbolic goal-directed exploration. In N M Amato, editor, *Proceedings of the 2015 IEEE International Conference on Robotics and Automation (ICRA 2015)*, pages 1100–1105. Institute of Electrical and Electronics Engineers Inc., USA, 2015.
- [4] M. AftabChowdhury and Kaushik Deb. Extracting and segmenting container name from container images. *International Journal of Computer Applications*,

74:18–22, 07 2013.

- [5] Xiaoxue Chen, Lianwen Jin, Yuanzhi Zhu, Canjie Luo, and Tianwei Wang. Text recognition in the wild: A survey. *ACM Comput. Surv.*, 54(2), March 2021.