

ADVANCED TECHNOLOGY DAYS

29. i 30. studeni 2022.

What Are Limits of OCR?

Dino Grgić dino.grgic1@gmail.com





ZLATNI SPONZOR









SREBRNI SPONZOR









SPONZOR COFFEE BREAKA







BRONČANI SPONZOR



enna energia naturalis





NESPRESSO



GENERALNI MEDIJSKI POKOROVITELJ

POKROVITELJ

MEDIJSKI POKROVITELJI

GENERALNI MEDIJSKI POKOROVITELJ















ORGANIZATOR









Agenda

- Introduction to Computer Vision and OCR
- How do we do OCR?
- What are the latest trends?
- Some state-of-the-art research examples

Who Am I?

- Junior Software Engineer Student at Unitfly
 - since 2020
- Computer Science masters' student at FER
 - research in AI image/video upscale
 - some experience in NLP research

Put It into Context: Analyzing the Effects of Context Delimiters and Emojis in Emotion Analysis

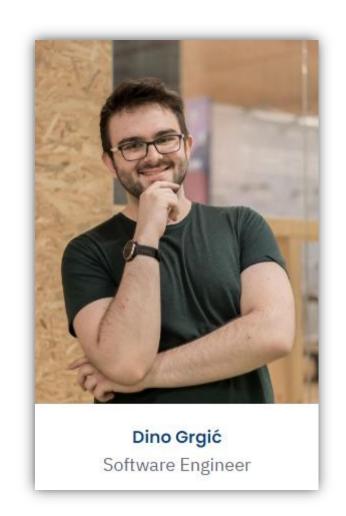
Dino Grgić, Magdalena Mucić, Matej Škrabić

University of Zagreb, Faculty of Electrical Engineering and Computing Unska 3, 10000 Zagreb, Croatia

{dino.grgic, magdalena.mucic, matej.skrabic}@fer.hr

Abstract

A large amount of user-generated data is being created daily due to social media communication. Sentiment and emotion analysis of this data can give us valuable insight, which can be applied in a wide range of situations: enhancing the customer experience, reputation management, market research, analyzing public opinion on different topics, and so on. In this paper, we conduct two experiments. Firstly, we show that using special delimiter tokens to signify the switch between the different utterances in the same dialogue results in improved task performance across three different models. We also experiment with the significance of emojis in emotion classification.

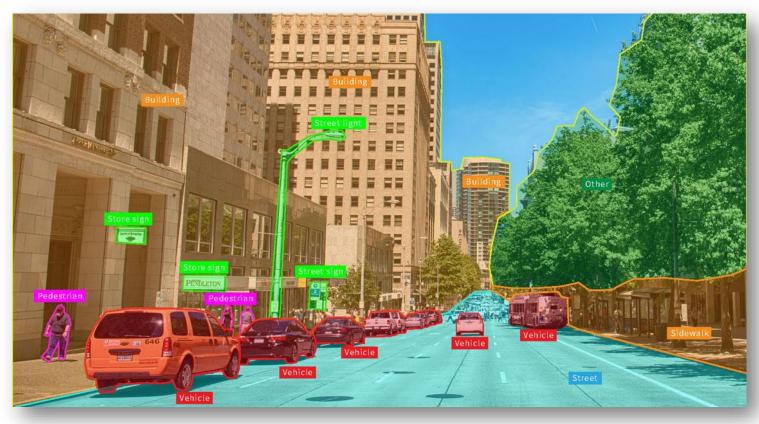


What Is Computer Vision?

• field of computer science

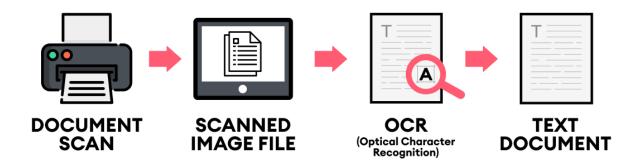
• TASK: enable computers to identify and understand objects in images and

video



What Is OCR?

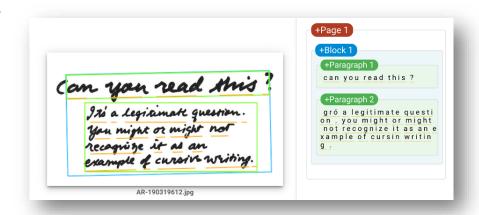
- task of computer vision
- Optical Character Recognition
 - text in image → machine readable text data



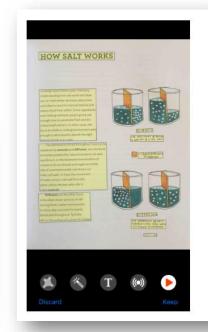
Where Is OCR Used?

- traffic sign recognition
- scanning license plates
- aids for visually impaired
- data entry for business documents
- converting handwritten notes to machine-readable text

• ...









HOW SALT WORKS

Cooking is part artistry,
part chemistry.

Understanding how salt
works will allow you to
make better decisions
about how and when to
use it to improve texture
and season food from
within. Some ingredients

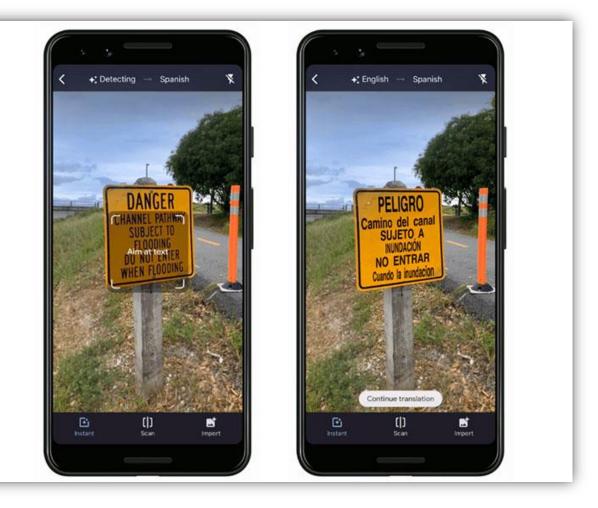
(T) (N) (D)

Cancel Retako Delete Done



Where Is OCR Used?







What Do We Need for OCR?

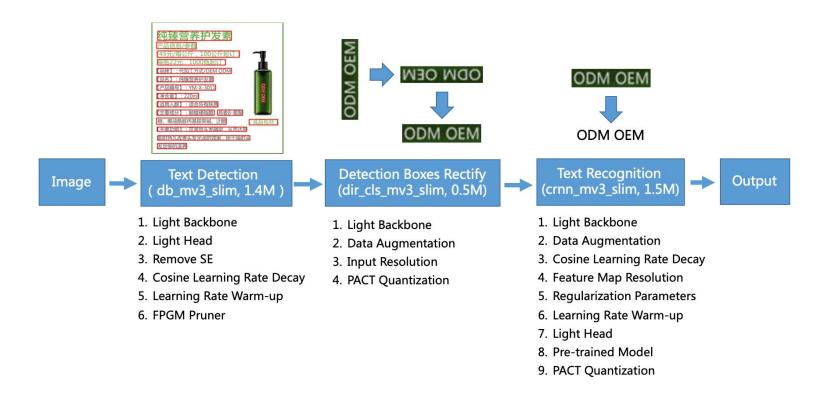
- Model
 - pretrained
 - train on your own
- Dataset
 - Unstructured
 - Structured

No free lunch!



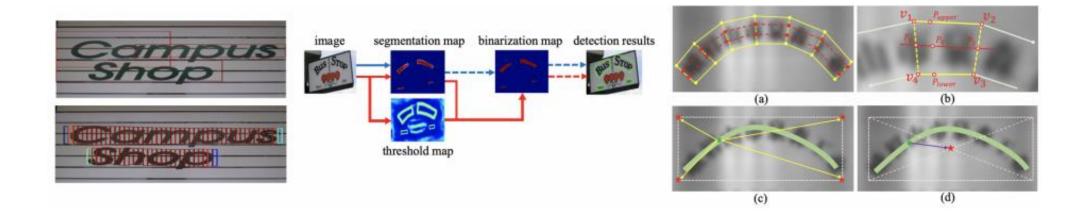
How We Do OCR?

- Approach
 - Computer Vision
 - Deep Learning
- (mostly) 3 step process
 - Pre-processing
 - Text detection + text recognition
 - Post-processing



Text Detection

- locate text regions in the input image
- problem: bounding box of the text



Text Recognition

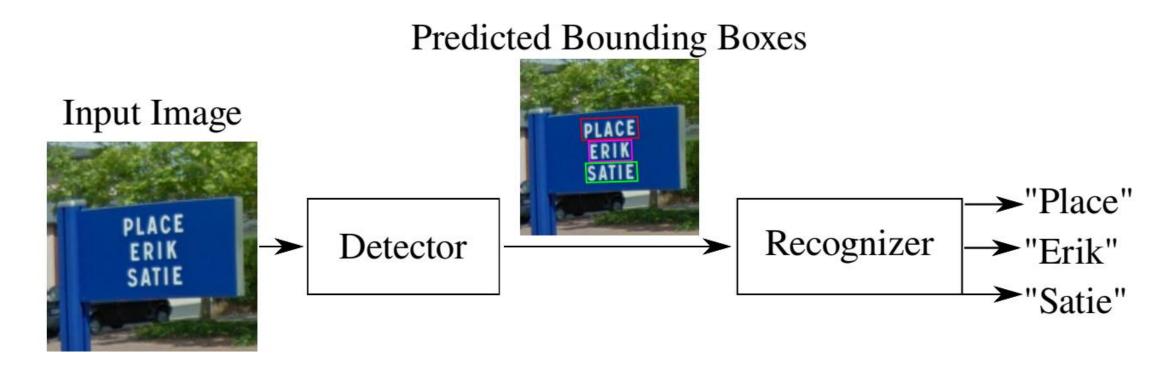
- extract text from input image
 - uses input from text detection
- two categories
 - 1. regular text
 - 2. irregular text research focus



Figure 9: (Left) Regular texts VS. (Right) Irregular texts

Text Detection vs Text Recognition

- text detection = detect placement of the text in image
- text recognition = extract the text



How Good Is OCR Model?

- evaluation metrics
- CER (%) Character Error Rate
 - less is better
 - how are we correct with characters
- WER (%) Word Error Rate
 - less is better
 - how are we correct with words
- Sequence Accuracy (%)
 - more is better
 - the whole sequence must be correct

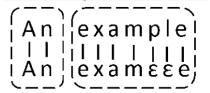
STEAM STEAM STEAM

STEAM STREAM

Substitution Deletion Insertion

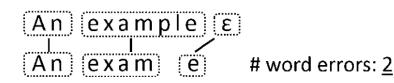
GT: An example OCR: An exame

Char alignment first:



char errors: 2 # word errors: 1

Direct word alignment:



17

Document Structure Recognition

- Layout Analysis
 - structure of the document
- Table Recognition
 - table content extraction
- Key Information Extraction



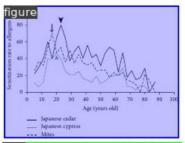
Q1: What's the address of the house?

A1: Room XXX, Building No.X, XX District, Beijing, China

Q2: What is the area of the house?

A2: 90.69 square meters

International Journal of Otolaryngology



1: The rate of sensitization (determined by RAST) is apparese cedar, Japanese cypress, and mites was affected by the patient's age. Black arrow head, gray arrow head, and black arrow show the peaks of each rate, respectively.

table	Total IgE (IU/mL)	Eosinophil ce proportion (%
Only spring pollens	118 ± 16	45±0.4
Only fall pollens	172±93	3.7 ± 1.4
Only perennial allergens	288 ± 51	3.2 ± 0.4
Spring and fall pollens	174±30	5.2 ± 0.9
Spring pollens and perennial allergens	391 ± 67	5.4 ± 0.5
Fall pollens and perennial allergens	-	
Spring and fall pollens and perennial allergens	878 ± 213	6.1 ± 0.6
No sensitization	120 ± 15	3.1 ± 0.2

rage of total serum IgE levels was highest in 8-17-year

text bd Cell Eosinophil Count. The blood cell eosinophicount was also compared between groups. The eosinophical proportion was 4.5 ± 0.4% in patients sensitized only to spring pollens, while it was significantly higher (5.7 ± 0.4% in patients sensitized to both perennial allergens and spring pollens (P = 0.0146, Mann-Whitney U test) (Figure 2(b) Table 2). The blood cell eosinophil count showed the same reductive tendency (Figure 3(b)).

rgic Sensitization in Asthma. Fifty-nine patients (46 idults, 13 children) had been previously diagnosed with sthma. The remaining 599 patients had not been diagnosed with asthma. Sensitization to any allergen was detected in 58% of patients with asthma (34/59). Twenty-six (44%) of 59 patients were sensitized to spring pollens (Table 3) Approximately half of the asthma patients (51%; 30/59) were sensitized to perennial allergens. Seven percent of satients with asthma (4/59) were sensitized only to spring

CABLE 3: Allergic sensitization in asthma

CABLE 3: Pollens 4
Only fall pollens 0
Only perennial allergens 7
Spring and fall pollens 0
Spring pollens and perennial allergens 14
Fall pollens and perennial allergens 14
Spring and fall pollens and perennial allergens 2
Spring and fall pollens and perennial allergens 2
No sensitization 25

while 16% (94/593) in patients without asthma were sensitized exclusively to these allergens. Thirty-seven percent of patients with a previous asthma diagnosis (22/59) were ensitized to both spring and perennial allergens, which was ignificantly higher than that observed in patients without sethma (20% II/593) (P = 0.0017, chi-sourge test)

Text in total serum IgE levels in patients with asthma wer 177 ± 89 IU/mL, while those in patients without asthmere 224 ± 27 IU/mL (P = 0.0001 compared to patien with asthma, Mann-Whitney U test). Blood eosinophil corpoportion in patients with asthma was 5.4 ± 0.6%. If patients without asthma, the proportion was 3.9 ± 0.29 Blood eosinophil cell proportion in patients without asthma was ignificantly higher than those in patients without asthm (P = 0.008, Mann-Whitney U text).

4. Discussion title

text sensitization, as diagnosed by the serum allerger gE level, does not always correspond with th tient's symptoms. We found that approximately twice a any patients were sensitized to both spring pollens ar ring pollens. However, many patients were asymptoma perennial allergens. Exposure to perennial allergens, suc ouse dust mite and cat and dog dandruff, is an importa disposing risk factor for asthma [4]. Previous diagno asthma was largely related to serum IgE levels and blo sinophil counts [5-7]. Even in nonasthmatic patien way responsiveness (assessed using methacholine [8]) acreased in some cases of allergic rhinitis, indicating creased risk for asthma [9-11]. Sensitization to cat dandr ust mite, cockroach, and ragweed is an important predict f airway hyperresponsiveness [12]. Airway hyperresp ess is strongly related to elevated total serum IgE leve ven in asymptomatic patients [5, 13]. In other words, to rum IgE level is considered an indicator of probable airw perresponsiveness or asthma. In our study, total seru levated in patients sensitized to both spring pollens a o spring pollens. Therefore, patients sensitized to both spr lens and perennial allergens might be at greater risk

text spared to adults, fewer children were sensitized on spring pollens. Most children (approximately 80%) h.

Methods	Ext	R	P	F	FPS
TextSnake [18]	Syn	85.3	67.9	75.6	-
CSE [17]	MLT	76.1	78.7	77.4	0.38
LOMO[40]	Syn	76.5	85.7	80.8	4.4
ATRR[35]	Sy-	80.2	80.1	80.1	-
SegLink++ [28]	Syn	79.8	82.8	81.3	-
TextField [37]	Syn	79.8	83.0	81.4	6.0
MSR[38]	Syn	79.0	84.1	81.5	4.3
PSENet-1s [33]	MLT	79.7	84.8	82.2	3.9
DB [12]	Syn	80.2	86.9	83.4	22.0
CRAFT [2]	Syn	81.1	86.0	83.5	-
TextDragon [5]	MLT+	82.8	84.5	83.6	-
PAN [34]	Syn	81.2	86.4	83.7	39.8
ContourNet [36]	-	84.1	83.7	83.9	4.5
DRRG [41]	MLT	83.02	85.93	84.45	-
TextPerception[23]	Syn	81.9	87.5	84.6	-
Ours	-	80.57	87.66	83.97	12.08
Ours	Syn	81.45	87.81	84.51	12.15
Ours	MLT	83.60	86.45	85.00	12.21

Methods	Ext	IR.	P	F	FPS
TextSnake [18]	Syn	85.3	67.9	75.6	-
CSE [17]	MLT	76.1	78.7	77.4	0.38
LOMO[40]	Syn	76.5	85.7	80.8	4.4
ATRR[35]	Sy-	80.2	80.1	80.1	-
SegLink++ [28]	Syn	79.8	82.8	81.3	-
TextField [37]	Syn	79.8	83.0	81.4	6.0
MSR[38]	Syn	79.0	84.1	81.5	4.3
PSENet-1s [33]	MLT	79.7	84.8	82.2	3.9
DB[12]	Syn	80.2	86.9	83.4	22.0
CRAHT [2]	Syn	81.1	86.0	83.5	-
TextDragon 1	MLT+	82.8	84.5	83.6	-
PAN [3I]	Syn	81.2	86.4	83.7	39.8
ContourNet [36]	-0	84.1	83.7	83.9	4.5
DRRG [Δ1]	MLT	83.02	85.93	84.45	-
TextPerception[23	Syn	81.9	87.5	84.6	-
Ours	-	80.57	87.66	83.97	12
Ours	Syn	81.45	87.81	84.51	12.15
Ours	MLT	83.60	86.45	85.00	12.21

# Pre-training Data	# Pre-training Epochs	Precision	Recall	F1	# Pre-training Data	# Pre-training Epochs	Precision	Recall	F1	
	1 epoch	0.5779	0.6955	0.6313		1 epoch	5. 5779	5.6955	5. 6313	
	2 epochs	0.6217	0.705	0.6607		2 epochs	0.6217	0.705	5.6607	
	3 epochs	0.6304	0.718	0.6713		3 epochs	5, 6304	5.718	5, 6713	
500K	4 epochs	0.6383	0.7175	0.6756		4 epochs	5, 6383	5, 7175	5, 6756	
	5 epochs	0.6568	0.734	0.6933		5 epochs	5, 6568	5, 734	5, 6933	
	6 epochs	0.665	0.7355	0.6985	500K	6 epochs	5.665	5. 7355	5. 6985	
	1 epoch	0.6156	0.7005	0.6552		1 epoch	5. 6156	0.7005	5, 6552	
	2 epochs	0.6545	0.737	0.6933		2 epochs	0.6545	5.737	5.6933	
114	3 epochs	0.6794	0.762	0.7184		3 epochs	5. 6794	5.762	5.7184	
1M 5 epochs	0.6812	0.766	0.7211		4 epochs	5.6812	5.766	5.7211		
	5 epochs	0.6863	0.7625	0.7224		5 epochs	0, 6863	0.7625	5, 7224	
	6 epochs	0.6909	0.7735	0.7299	1 M	6 epochs	5, 6909	5, 7735	5, 7299	
	1 epoch	0.6599	0.7355	0.6957		1 epoch	5.6599	5. 7355	5. 6957	
	2 epochs	0.6938	0.759	0.7249		2 epochs	5, 6938	5, 759	5, 7249	
2M	3 epochs	0.6915	0.7655	0.7266		3 epochs	5.6915	5, 7655	5, 7266	
4 epochs 5 epochs		0.7081	0.781	0.7427		4 epochs	5, 7081	5, 781	5, 7427	
		0.7228	0.7875		5 epochs	5.7228	5. 7875	5, 7538		
	6 epochs 0.7377 0.782 0.7592	2W	6 epochs	5.7377	5, 782	5, 7592				
11M	1 epoch	0.7464	0.7815	0.7636	48	1 epoch	5.7464	5. 7815	5, 7636	
TIM	2 epochs	0.7597	0.8155	0.7866	1114	1 epoch	B 2507	5 01FF	5. 7000	-

Figure 13: Table recognition



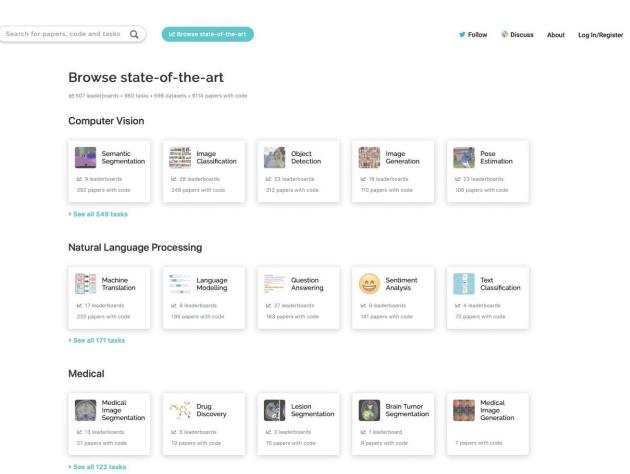
Current OCR Challenges



Figure 4: Technical challenges of OCR algorithms

Following ML Trends

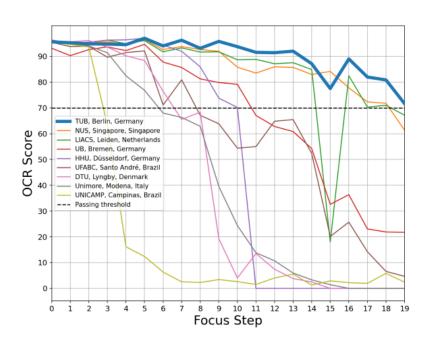
- paperswithcode.com
- public OCR tools
 - Azure Computer Vision
- scientific conference / meetings
 - OCR
 - ICDAR International Conference on Document Analysis and Recognition
 - NLP
 - ACL Association for Computational Linguistics



Research Example (2022)

- Solving problem: Pre-processing Blur
- Let's Enhance: A Deep Learning Approach to Extreme Deblurring of Text Image (Trippe et al., 2022)
 - Technische Universität Berlin & Utrecht University
 - Helsinki Deblur Challenge 2021 (1st place)

level	blurry image	our reconstruction	sharp image (ground truth)
4	fPEb D rPz YXZNvrDpzd BAdUCQvuGm	fPEb D rPz YXZNvrDpzd BAdUCQvuGm	fPEb D rPz YXZNvrDpzd BAdUCQvuGm
9	STRUMENT STATES OF STRUMENT STRUMENT STATES OF STRUMENT STATES OF STRUMENT STATES OF STRUMENT STATES OF STATES OF STRUMENT STATES OF STA	VmqKEBrRGm nXqmrihTPY sfNSXDFKWR	VmqKEBrRGm nXqmrihTPY sfNSXDFKWR
14		FQeMYqKBZj QfsgKuinTr WzQRcALRLW	FQeMYqKBZj QfsgKuinTr WzQRcALRLW
19		BBue FisKT uGNzhTHxHX spTYMuGJzrh	BBue FisKT uGNzhTHxHX spTYMuGJzm





Optical Character Recognition

☑ Edit

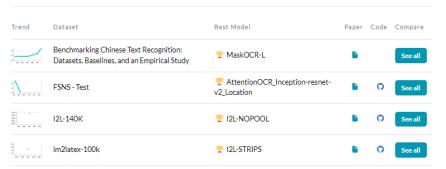
203 papers with code • 4 benchmarks • 49 datasets

Optical character recognition or optical character reader (OCR) is the electronic or mechanical conversion of images of typed, handwritten or printed text into machine-encoded text, whether from a scanned document, a photo of a document, a scene-photo (for example the text on signs and billboards in a landscape photo, license plates in cars...) or from subtitle text superimposed on an image (for example: from a television broadcast)

Benchmarks

Add a Result

These leaderboards are used to track progress in Optical Character Recognition



Libraries 1

Use these libraries to find Optical Character Recognition models and implementations

PaddlePaddle/PaddleOCR	18 papers	26,901★
open-mmlab/mmocr	5 papers	2,977 ★
Media-Smart/vedastr	5 papers	497★
O huggingface/transformers	3 papers	74,915 ★

See all 7 libraries.



Content

- ∠ Benchmarks
- Datasets
- Papers

- Latest

- Introduction
- ஃ Subtasks
- Libraries
- Most implemented
- Social
- No code

Subtasks

Active Learning

Datasets



Handwritten Digit Recognition

[IIII] TextCaps

[III] SciTSR

Irregular Text Recognition

Show all 8 subtasks

See all 49 optical character recognition datasets

Most implemented papers

FUNSD

Most implemented Social Latest No code

Search for a paper, author or keyword



An End-to-End Trainable Neural Network for Imagebased Sequence Recognition and Its Application to Scene Text Recognition

↑ PaddlePaddle/PaddleOCR • → • 21 Jul 2015

In this paper, we investigate the problem of scene text recognition, which is among the most important and challenging tasks in image-based sequence recognition.

79







EAST: An Efficient and Accurate Scene Text Detector

O PaddlePaddle/PaddleOCR • مع فرخ • CVPR 2017

Previous approaches for scene text detection have already achieved promising performances across various benchmarks.





Azure Computer Vision

Use one of your own files or choose from a sample below.



Drag and drop a file here or Browse for a file or Take a photo







Sample form #3



Detected attributes

JSON

Nutrition Facts Amount Per Serving

Serving size: 1 bar (40g)

Serving Per Package: 4

Total Fat 13g

Saturated Fat 1.5g

Amount Per Serving

Trans Fat 0g

alories 190

Cholesterol 0mg

ories from Fat 110

Sodium 20mg

nt Daily Values are based on

Vitamin A 50%

calorie diet.

Optical Character Recognition (OCR)

The Computer Vision Read API supports many languages. The Read API can extract text from images and documents with mixed languages, including from the same text line, without requiring a language parameter.

① Note

Language code optional

Read OCR's deep-learning-based universal models extract all multi-lingual text in your documents, including text lines with mixed languages, and do not require specifying a language code. Do not provide the language code as the parameter unless you are sure about the language and want to force the service to apply only the relevant model. Otherwise, the service may return incomplete and incorrect text.

See How to specify the Read model to use the new languages.

Handwritten text

The following table lists the OCR supported languages for handwritten text by the most recent Read GA model.

Language	Language code (optional)	Language	Language code (optional)
English	en	Japanese	ja
Chinese Simplified	zh-Hans	Korean	ko
French	fr	Portuguese	pt
German	de	Spanish	es
Italian	it		

Print text

The following table lists the OCR supported languages for print text by the most recent Read GA model.

Language	Code (optional)	Language	Code (optional)
Afrikaans	af	Khasi	kha
Albanian	sq	K'iche'	quc
Angika (Devanagiri)	anp	Korean	ko
Arabic	ar	Korku	kfq
Asturian	ast	Koryak	kpy
Awadhi-Hindi (Devanagiri)	awa	Kosraean	kos
Azerbaijani (Latin)	az	Kumyk (Cyrillic)	kum
Bagheli	bfy	Kurdish (Arabic)	ku-arab

Is OCR Solved?

- there are always problems to be solved
- no model is 100% accurate
- OCR is base for many systems
- OCR improvement → CV models improvement



Figure 4: Technical challenges of OCR algorithms

Demo

Useful Links

- https://paperswithcode.com/sota
- https://dl.acm.org/doi/10.1145/1273445.1273458
- https://github.com/PaddlePaddle/PaddleOCR

Digression: Reading Papers

- How to Read a Paper (Keshav, 2007)
- three pass approach
 - Pass one least technical
 - Reading only the title, abstract, and first paragraph of each section
 - 5 10 minutes
 - 2. Pass two
 - Everything but details (formulas, figures, code)
 - 1 hour
 - 3. Pass three most technical
 - Fully understand the paper
 - 4-5 hours

How to Read a Paper

S. Keshav
David R. Cheriton School of Computer Science, University of Waterloo
Waterloo, ON, Canada
keshav@uwaterloo.ca

ABSTRACT

Researchers spend a great deal of time reading research papers. However, this skill is rarely taught, leading to much wasted effort. This article outlines a practical and efficient three-pass method for reading research papers. I also describe how to use this method to do a literature survey.

Categories and Subject Descriptors: A.1 [Introductory and Survey]

General Terms: Documentation. Keywords: Paper, Reading, Hints.

1. INTRODUCTION

Researchers must read papers for several reasons: to review them for a conference or a class, to keep current in their field, or for a literature survey of a new field. A typical researcher will likely spend hundreds of hours every year reading papers.

Learning to efficiently read a paper is a critical but rarely taught skill. Beginning graduate students, therefore, must learn on their own using trial and error. Students waste much effort in the process and are frequently driven to frustration.

For many years I have used a simple approach to efficiently read papers. This paper describes the 'three-pass' approach and its use in doing a literature survey.

2. THE THREE-PASS APPROACH

The key idea is that you should read the paper in up to three passes, instead of starting at the beginning and plowing your way to the end. Each pass accomplishes specific goals and builds upon the previous pass: The first pass gives you a general idea about the paper. The second pass

 Glance over the references, mentally ticking off the ones you've already read

At the end of the first pass, you should be able to answer the $five\ Cs$:

- Category: What type of paper is this? A measurement paper? An analysis of an existing system? A description of a research prototype?
- Context: Which other papers is it related to? Which theoretical bases were used to analyze the problem?
- 3. Correctness: Do the assumptions appear to be valid?
- Contributions: What are the paper's main contributions?
- 5. Clarity: Is the paper well written?

Using this information, you may choose not to read further. This could be because the paper doesn't interest you, or you don't know enough about the area to understand the paper, or that the authors make invalid assumptions. The first pass is adequate for papers that aren't in your research area, but may someday prove relevant.

Incidentally, when you write a paper, you can expect most reviewers (and readers) to make only one pass over it. Take care to choose coherent section and sub-section titles and to write concise and comprehensive abstracts. If a reviewer cannot understand the gist after one pass, the paper will likely be rejected; if a reader cannot understand the highlights of the paper after five minutes, the paper will likely never be read.

2.2 The second pass



Questions?