

Intelligent Automation for AI-Driven Document Understanding

preliminary defense

Jordy Van Landeghem

12/03/2024

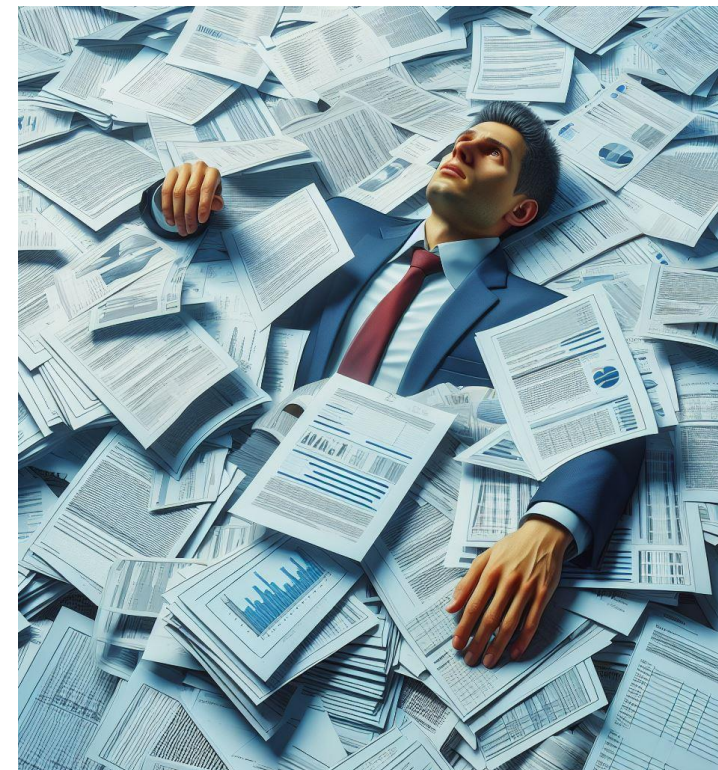
Humans and organizations are *drowning* in visually-rich documents...

- Document-based communication facilitates crucial interactions, decisions and actions

Processed manually
= Inefficient

Why **visually-rich** documents?

- human brain processes visual information 6-600 times faster
 - visual content helps readers retain more information
 - layout provides a physical content organization for better retrieval
- *Technology assistance* is vital for the digitalization quest



...yet organizations lag in adopting **automated document processing** solutions

Two primary challenges:

- I. Need for reliability, robustness and control over associated risks

➔ Intelligent Automation (IA)

- II. Complexity of processing, long multimodal documents algorithmically

➔ Document Understanding (DU)



We focus on **fundamental** enablers and **methodological** advances

- ❑ Investigated the quality of predictive uncertainty quantification (PUQ) on document data and task distributions, devising novel, scalable PUQ methods
- ❑ Promoted more realistic datasets and evaluation, fostering generic DU research
- ❑ Designed practical frameworks for targeting efficient multipage processing

This dissertation innovates in and contributes to:

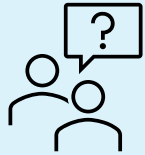
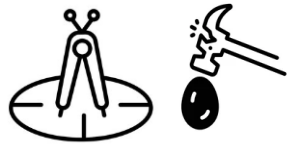
I. Reliable and Robust Deep Learning



II. Realistic and Efficient Document Understanding



Overview: publications and innovation scope



**Predictive
Uncertainty for
Probabilistic Novelty
Detection in Text
Classification**
ICML 2020

**Benchmarking
Scalable Predictive
Uncertainty in
Text Classification**
IEEE Access 2022



**Beyond Document
Page Classification:
Design, Datasets,
and Challenges**
WACV 2024 oral



**Competition on
Document
Understanding of
Everything**
ICDAR 2023 oral

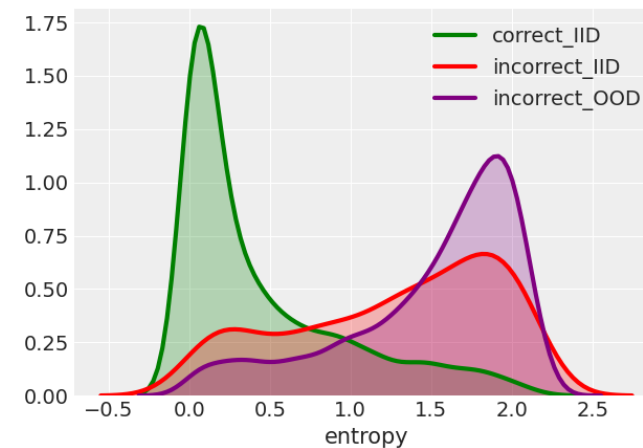
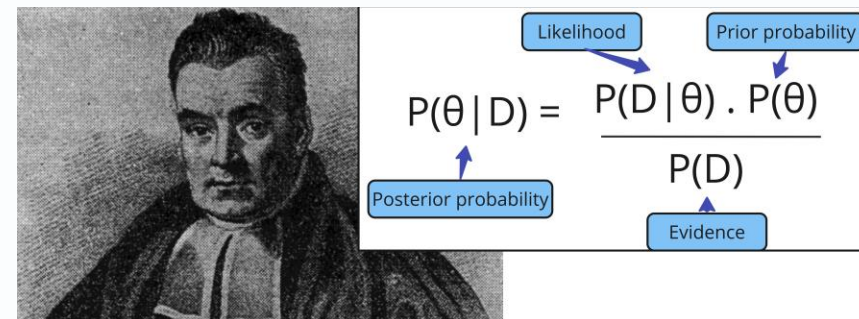
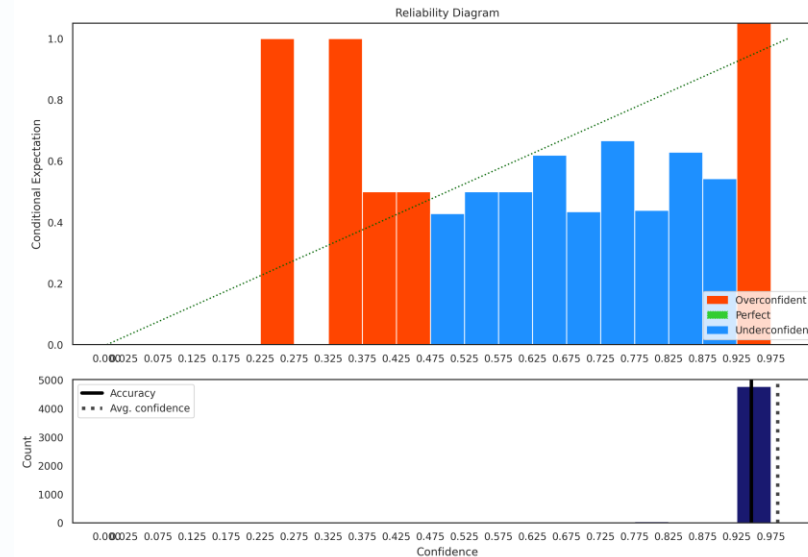
**Document
Understanding
Dataset and
Evaluation**
ICCV 2023



**Knowledge
Distillation for
Visually-Rich
Document
Applications**
*ICDAR 2024**

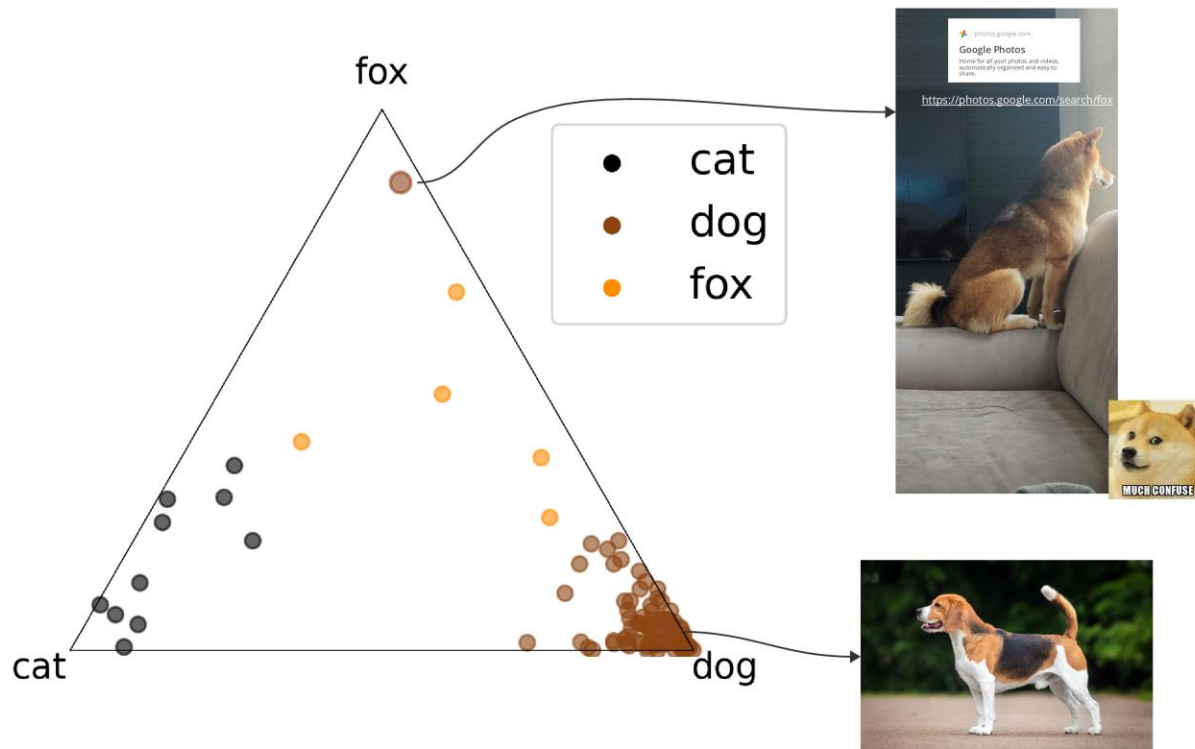


I. Reliable and Robust Deep Learning



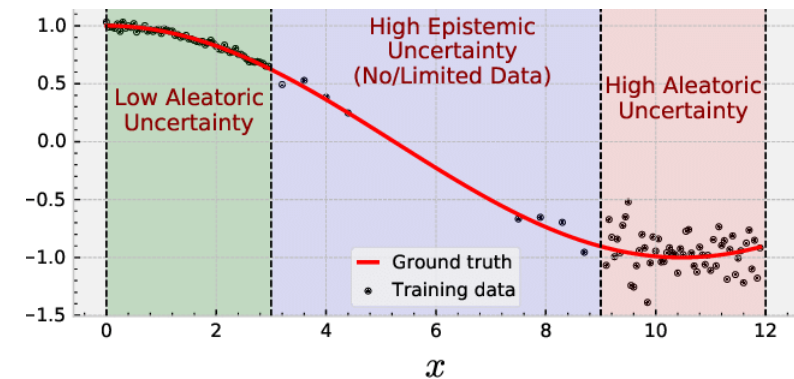


Probabilistic predictors output unreliable uncertainty



Bayesian Deep Learning (BDL):
Investigate parameter uncertainty

Predictive Uncertainty Quantification:
Disentangle sources of uncertainty



Knowledge gaps:

- Missing evaluation of PUQ in NLP
 - Applicability and scalability?
- Architecture, prior and hyperparameter influences on uncertainty quality

→ *Alternative confidence scoring functions?*



Is cross-domain robustness in NLP obtainable?

Rubi
★★★★★ A practical solution for small kitchens
Reviewed in Germany on 25 December 2022
Colour Name: White | **Verified Purchase**
The trash cans fit perfectly in the cupboard. I'm very satisfied.

dafydd
★☆☆☆☆ Rubbish
Reviewed in the United K
Verified Purchase
Rubbish plot and acting

★★★★★ Hajo
Reviewed in Germany
Verified Purchase
Everything ok

Paul
★☆☆☆☆ Bad quality
Beoordeeld in Nederland op 11 juni 2023
Stijlnaam: 17-delige keukenmessenset - Zwart
Rust everywhere after 6 months of use.

Amazon Kunde
★☆☆☆☆ Thin, poorly processed plastic
Reviewed in Germany on 19 September 2023
Colour Name: White | **Verified Purchase**
The bins are made of thin, cheap plastic and are

Trina Sen
★★★★★ Great product
Beoordeeld in Nederland op 4 februari 2024
Kleur: Roze | **Geverifieerde aankoop**
Real good product one should buy

→This is a far different “beast” than detecting a fox in a cats vs. dogs problem

Contributions: **Reliable** and **Robust**



Predictive Uncertainty for Probabilistic
Novelty Detection in Text Classification
ICML UDL 2020

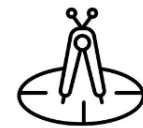
- BDL survey and literature review
- PUQ methods NLP benchmark



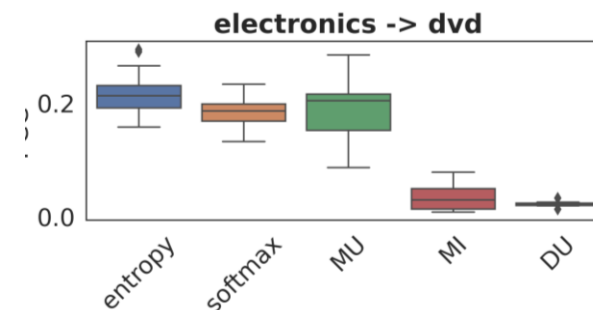
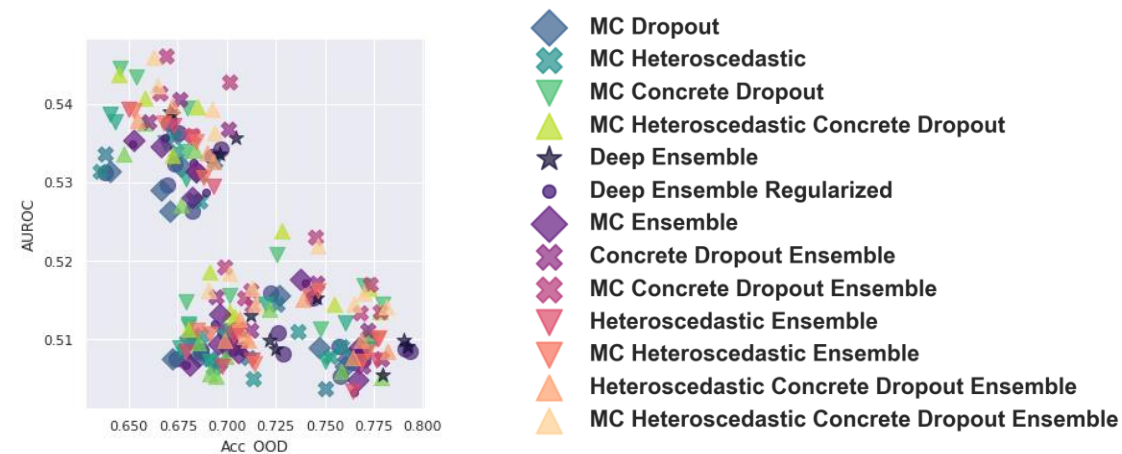
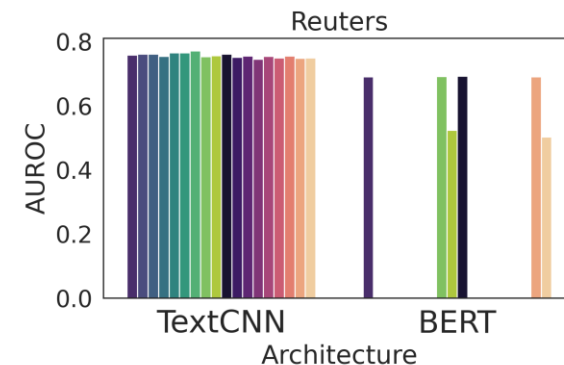
Benchmarking Scalable Predictive
Uncertainty in Text Classification
IEEE Access 2022

- Novel hybrid PUQ methods
- Real-world evaluation setups
- Take-home guidelines for PUQ

Presenting the first, comprehensive benchmark for scalable PUQ in NLP



- ✓ 6 text classification datasets
- ✓ 2 neural network architectures
- ✓ 6 unique, 28 total uncertainty methods
- ✓ 5 uncertainty measures
- ✓ 3 experiment setups
- ✓ 5 random seeds
- ✓ 4 hyperparameter ablations

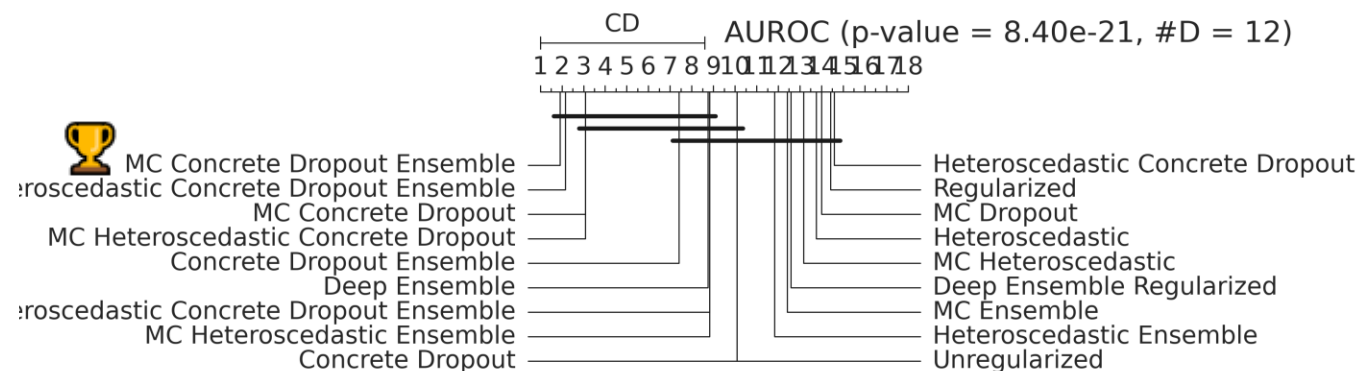
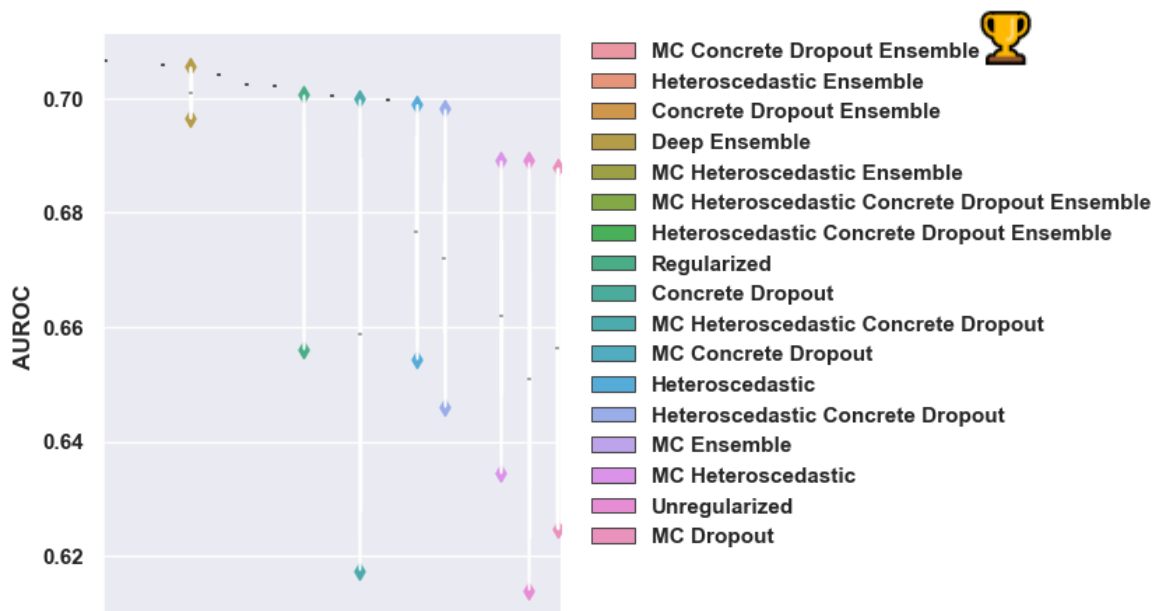
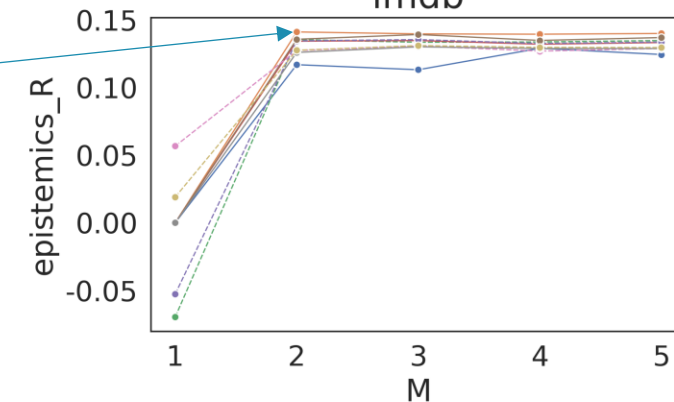
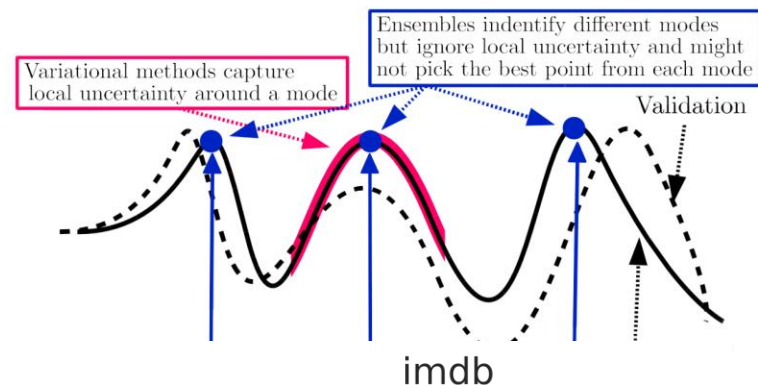


We propose novel, hybrid PUQ methods from complementarity in function space



MC Concrete Dropout Ensemble:

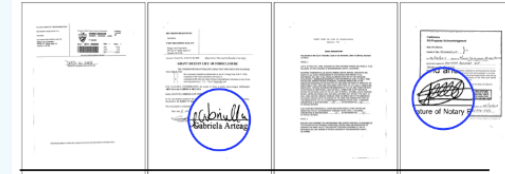
- Presented empirical evidence for theory of complementarity in function space
- Superior at novel class robustness and out-of-domain detection, even at a lower ensemble size



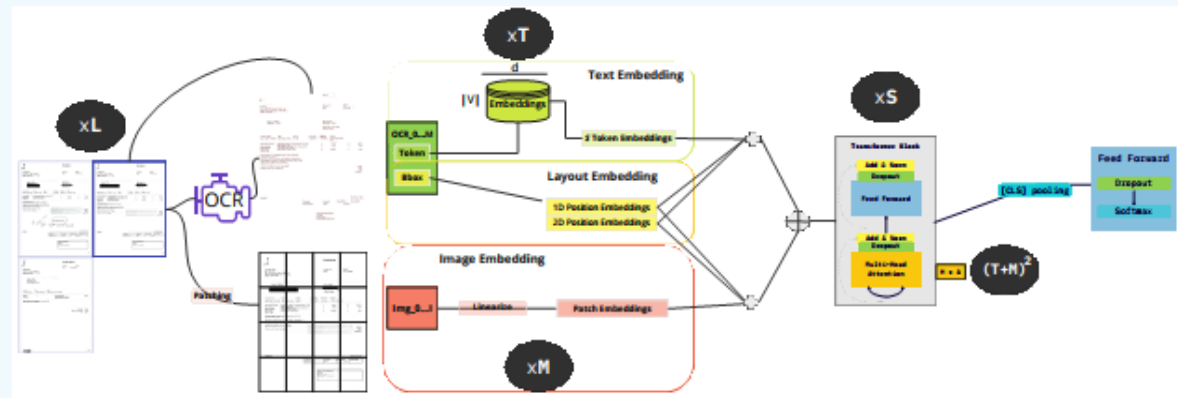
II. Realistic and Efficient Document Understanding



Requires counting. How many pages have a signature?
 The question requires visual comprehension (recognition of signature), knowledge about layout, and counting.



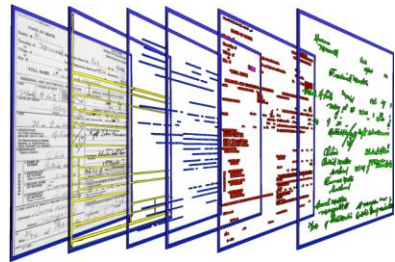
Source	Answer	ANLS	Conf.
Ground truth	2		
Human	2	1.0	—
T5	1	0.0	0.01
ChatGPT	4	0.0	—
GPT3	[Not-answerable]	0.0	—
T5-2D	4	0.0	0.69
HiVT5	4	0.0	0.41





Shifting the focus to Document Understanding

Focus of the field



Optical Character Recognition



Document Understanding



Document data unavailability

- Datasets lacking variety, scale and multipage documents
- Current benchmarks evaluation does not transfer downstream

Pretrain-finetune | Foundation models

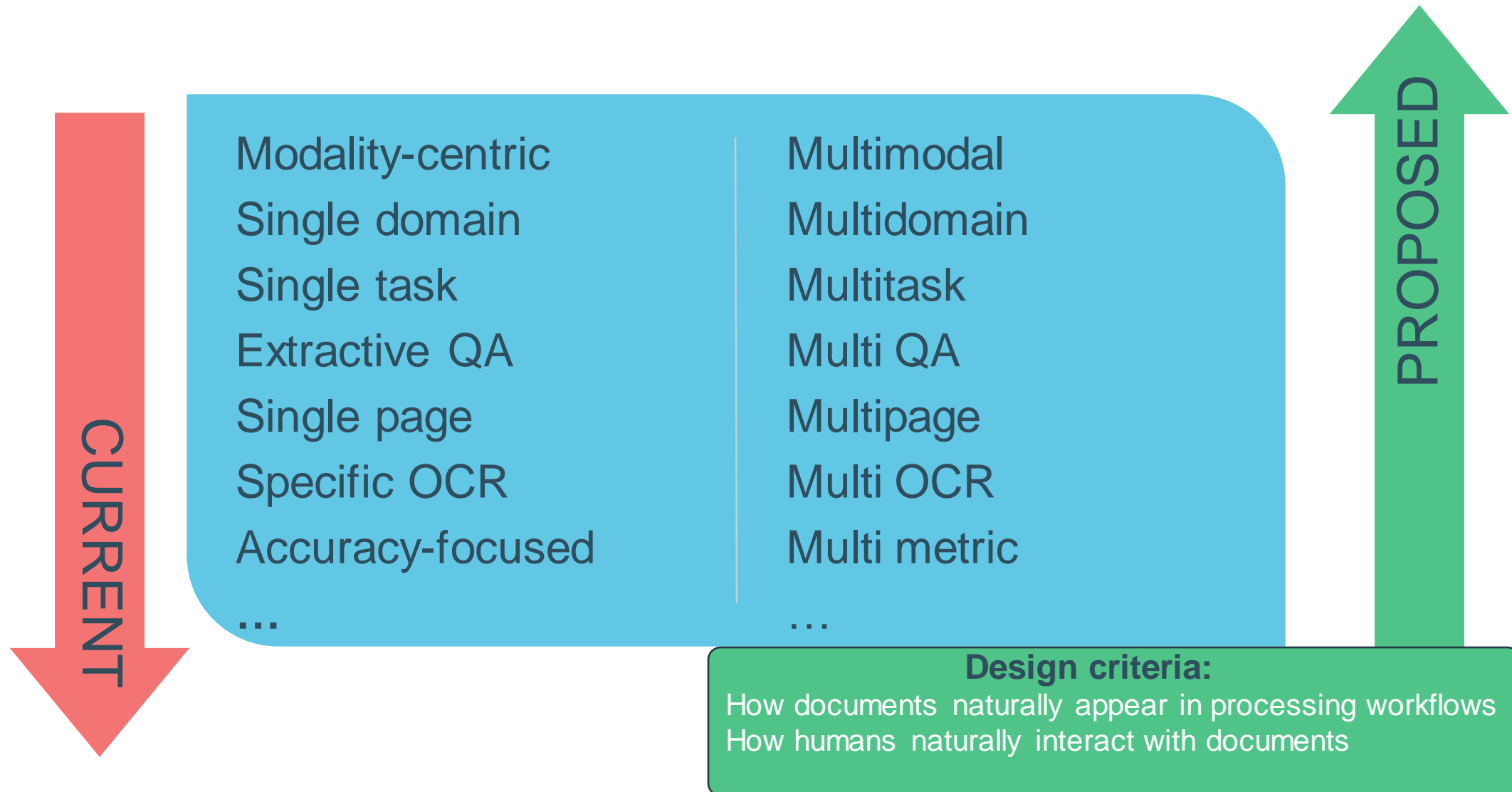
- Text-only LLMs for any document task?
- Foundation models more powerful, yet also more cumbersome

Objectives

- More generally applicable, embrace real-world complexity
- More efficient at modeling the multimodality of documents
- Evaluation more in sync with downstream requirements



What are DU benchmarks missing?



Contributions: Realistic

Beyond Document Page Classification:
Design, Datasets, and Challenges
WACV 2024 *oral



- Formalization of multi-page DC
- Construction of two novel datasets
- Survey and recommendations:
 - Complete DC methodology
 - Dataset construction efforts

Document Understanding Dataset and
Evaluation
ICCV 2023



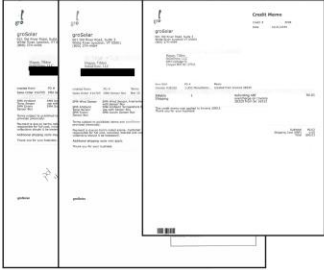
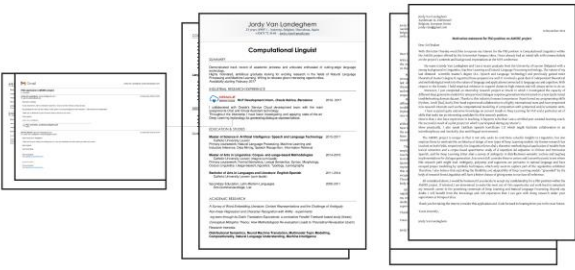


Competition on Document
UnderstandiNG of Everything (DUDE)
ICDAR 2023 *oral

- Design of multi-faceted dataset
- Comprehensive evaluation of SOTA
- Baseline and competition results
- Calibrated, selective generation



Document classification is more complex than reported

Covered in public research benchmarks

INPUT TASK	Document f_d	Document bundle f_b	Page stream f_s	Page splits f_m	
INPUT TASK					
LABELS	collision form	purchase invoice	email; resume; application letter	wage slip, wage_slip; bank statement; id_back, id_front; wage_slip	ticket_1, ticket_2, ..., ticket_9
USE-CASE	<i>Insurance claims</i>	<i>Robotic accounting</i>	<i>HR job screening</i>	<i>Loan application</i>	<i>Expenditure</i>

A multi-faceted benchmark for generic DU challenges the state-of-the-art



Document UnderstanDing of Everything



#non-answerable

Q: In which year does the Net Requirement exceed 25,000?

A: None

#abstractive #counting

Q: How many attorneys are listed for the plaintiffs?

A: Two

#layout-navigating #graphic-intensive

Q: Are the margins of the page uniform on all pages?

A: Yes

#extractive #list

Q: What are the Years mentioned in Chart 1?

A: [2020, 2021, 2022]

Employment Insurance, August 2022
Released at 8:30 a.m. Eastern time in The Daily, Thursday, October 20, 2022

In August, 485,000 Canadians received regular Employment Insurance (EI) benefits, down by 21,000 (-4.1%) from July.

According to the Labour Force Survey (LFS), the unemployment rate in August rose to 5.4%, following a record low of 4.9% observed in June and July.

Chart 1
Fewer regular Employment Insurance beneficiaries in August

Source: Employment Insurance Statistics (EIS), table 14-10-001-01.

In general, variations in the number of beneficiaries can reflect changes in the circumstances of different groups, including those becoming beneficiaries, those going back to work, those exhausting their regular benefits, and those no longer receiving benefits for other reasons.

The number of regular Employment Insurance beneficiaries decreases in eight provinces

Compared with July, the number of people who received regular EI benefits fell in eight provinces in August. The largest decrease occurred in Quebec (-12,000, -9.5%), which accounted for more than half of the national decline.

Other notable proportional decreases in regular EI beneficiaries occurred in Newfoundland and Labrador (-7.1%, -3,000) and New Brunswick (-8.7%, -2,000). British Columbia (+1.0%, +400) was the only province to see an increase in regular EI beneficiaries in August.

On a regional basis, the census metropolitan areas (CMAs) of Trois-Rivières (+16.0%, +400), Moncton (+4.1%, +500), and Montreal (+2.2%, +1,000), as well as areas outside large population centres in Quebec (+8.2%, +4,000) and Newfoundland and Labrador (-7.2%, -2,000), posted the largest proportional decreases in regular EI beneficiaries in August.

Statistics Canada / Statistique Canada

The Daily, Thursday, October 20, 2022

Chart 2
Job search method as a percentage of total job seekers

Source: Statistics Canada, Labour Force Survey (LFS), table 14-10-001-01.

In August, the percentage of women and of men using each job search method was about the same, with only two notable differences. As a percentage of total job seekers, men (51.9%) checked with employers directly more often than women (42.1%), and men (13.3%) were also more likely check with a private employment agency than women (7.8%) (not seasonally adjusted).

The next EIDaily will be released on November 17 with September data.

Sustainable Development Goals

On January 1, 2016, the world officially began implementing the 2030 Agenda for Sustainable Development—the United Nations’ transformative plan of action that addresses urgent global challenges over the following 15 years. The plan is based on 17 specific sustainable development goals.

Employment Insurance statistics are an example of how Statistics Canada supports reporting on global sustainable development goals. This release will be used to help measure the following goal:

Component of Statistics Canada catalogue no. 11-001-X

The Daily, Thursday, October 20, 2022

Table 4 - continued
Beneficiaries receiving regular income benefits,¹ by occupation,² Canada - Seasonally adjusted

Occupation	July 2022	Aug 2022	Aug 2021	Aug 2020	Aug 2019	Aug 2018	Aug 2017	Aug 2016	Aug 2015	Aug 2014	Aug 2013	Aug 2012	Aug 2011	Aug 2010	Aug 2009	Aug 2008	Aug 2007	Aug 2006	Aug 2005	Aug 2004	Aug 2003	Aug 2002	Aug 2001	Aug 2000	
Medical resources, agriculture and related production occupations	46,700	46,576	46,276	46,4	46,488	44	44.4																		
Supervisors, technical occupations in natural sciences, agriculture and related production occupations	12,710	7,503	5,590	80	4,770	1.8	-41.6																		
Professionals	27,600	15,000	15,070	80	15,760	2.4	-60.0																		
Technicians, trade occupations and related production occupations	20,400	13,700	14,400	0	10,200	0.1	-60.0																		
Occupations in manufacturing and other	10,200	43,369	30,700	-1,769	-75,300	-9.9	-44.0																		
Professionals, technicians and related production occupations	1,400	2,400	2,400	0	4,000	-0.1	-60.0																		
Professionals and manufacturing machine operators and related production occupations	20,700	3,960	3,960	0	10,200	-0.1	-60.0																		
Technicians, trade occupations and related production occupations	49,200	4,200	4,200	0	40,000	-0.1	-60.0																		
Occupations in manufacturing, manufacturing and other	48,400	18,170	18,170	400	38,400	-0.1	-60.0																		

¹ Excludes people who receive regular income benefits, whether or not they participate in one of these employment benefit programs, namely the Skills Development Program, the Job Creation Partnership Program and the Entrepreneurship Program.

² Source: Employment Insurance Statistics (EIS).

Available tables: 14-10-0044-01 to 14-10-0011-01, 14-10-0137-01, 14-10-0332-01, 14-10-0323-01, 14-10-0336-01, 14-10-0337-01, 14-10-0343-01, 14-10-0344-01 and 14-10-0348-01.

Definitions, data sources and methods: survey number 3054.

More information about the concepts and use of Employment Insurance statistics is available in the Guide to Employment Insurance Statistics (13-068-0).

For more information, or to enquire about the concepts, methods or data quality of this release, contact us (toll-free 1-800-953-7130; 514-293-8302; info@statcan.ca; or Media Relations (statcan.media@statcan.ca; info@statcan.ca; statcan@statcan.ca).

Component of Statistics Canada catalogue no. 11-001-X

#multi-hop #layout-navigating

Q: From the list of Top 10 Key Recovery Components, which is the last component listed on the second page?

A: Hope

#abstractive #graphic-intensive

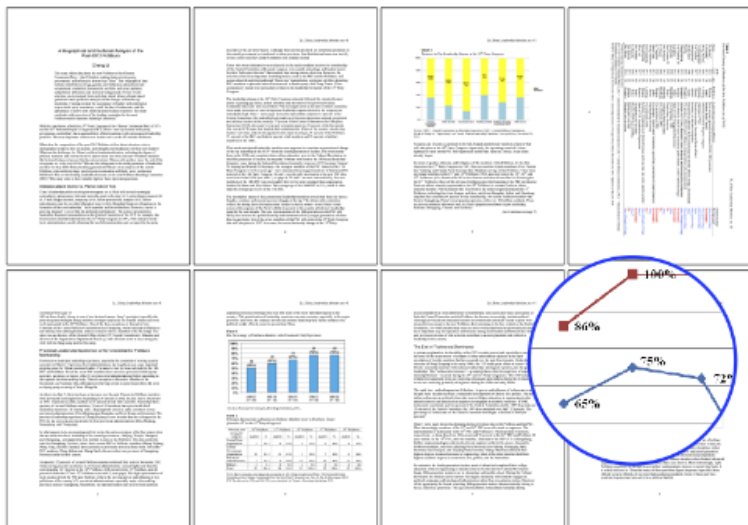
Q: Does this document contain any checkboxes?

A: No



-Everything-, you mean?

Visual evidence (chart). *What is the maximum percentage of the blue graph line on page 8?* A highly demanding question that requires simultaneous competency of visual comprehension (locating chart and line color), navigating through layout (determining adequate page), and numerical comparison (deciding on the highest value).



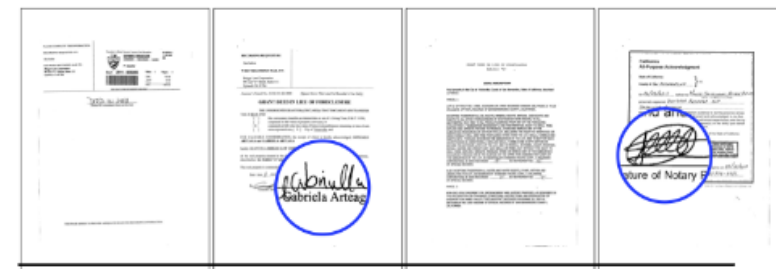
Visual evidence (map), multi-hop. *Which states don't have any marijuana laws?* The multi-hop question requires visually comprehending the map and linking knowledge from its legend with depicted regions.



Requires arithmetic. *What is the difference between how much Operator II and Operator III makes per hour?* The question requires table comprehension, determining relevant values, dividing extracted integers, and correcting the subject-verb agreement.

Operator	Hourly Wage
Operator I	\$22/hr
Operator II	\$17/hr
Operator III	\$17/hr

Requires counting. *How many pages have a signature?* The question requires visual comprehension (recognition of signature), knowledge about layout, and counting.



Source	Answer	ANLS	Conf.
Ground truth	2		
Human	2	1.0	—
T5	1	0.0	0.01
ChatGPT	4	0.0	—
GPT3	[Not-answerable]	0.0	—
T5-2D	4	0.0	0.69
HiVT5	4	0.0	0.41

Contributions: **Efficient**

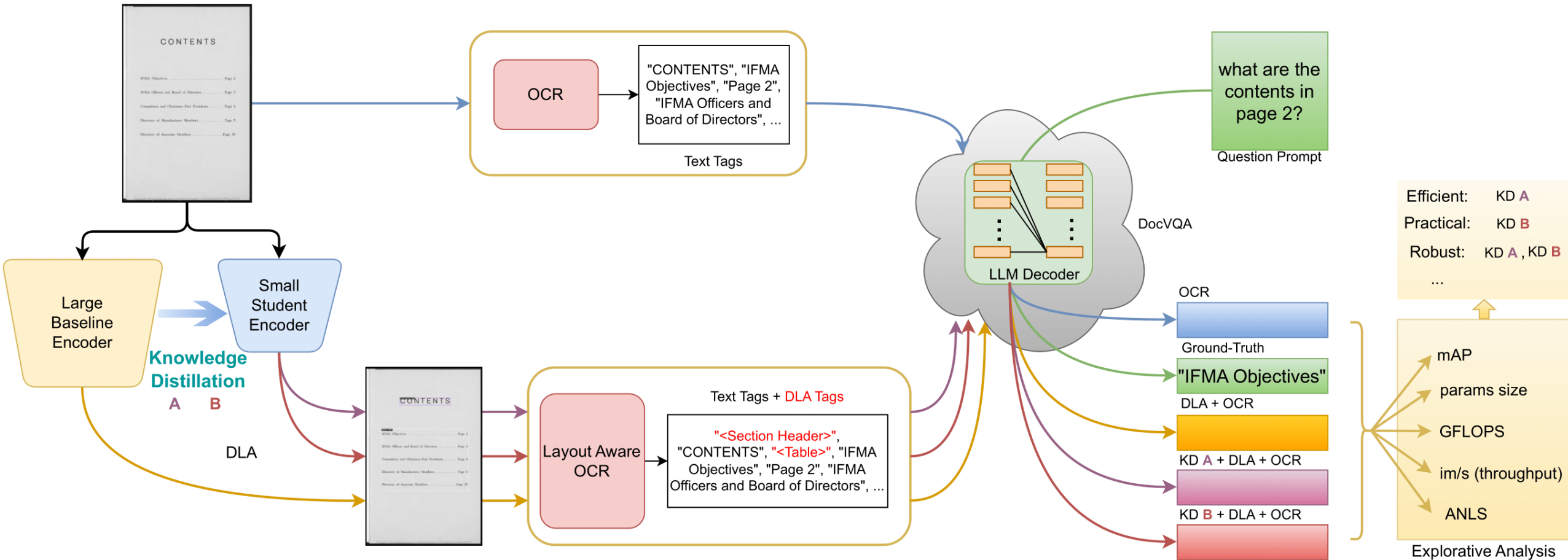


DistilDoc: Knowledge Distillation for
Visually-Rich Document Applications →
ICDAR 2024 (under review)

- KD benchmark on VDU tasks
- Novel downstream evaluation
- Enrich LLMs with semantic layout



Knowledge distillation facilitates small, specialized task modules that enrich downstream representations





DistilDoc streamlines research on compression tailored to VDU tasks

1. Best KD method

- SimKD > vanilla KD, on par with teacher, + under covariate shift

2. Teacher-Student capacity gap

- ViT-Tiny SimKD → 16x smaller model retains 90% rel. accuracy

3. Impact of Pretraining on KD

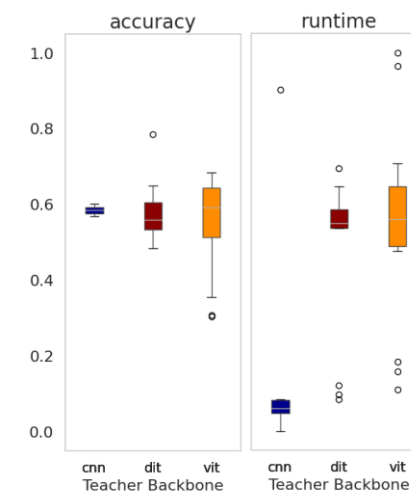
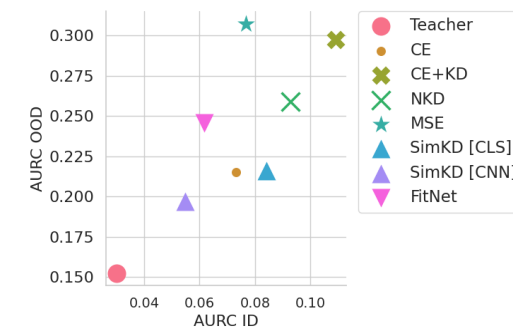
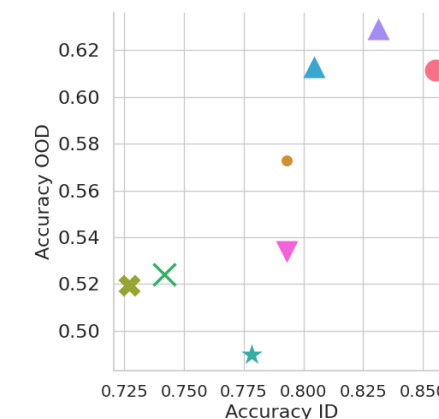
- ViT2ViT > DiT2ViT, - under covariate shift

4. Architecture influence

- Random initialization & DLA-KD: CNN > ViT

5. Applicability for downstream tasks

- DLA-enriched spacing prompting contributes positively to DocVQA



Conclusions

My dissertation addresses gaps, proposes methodologies opening new opportunities:

1. Limited research on scalable uncertainty quantification in NLP



Comprehensive survey and benchmark



Design of hybrid PUQ methods, offering better robustness and scalability

2. Disconnect DU research and applications



Complete redefinition of document classification and methodology

3. Unpredictable performance of SOTA for generic DU




Multi-faceted benchmark and competition incorporating all document modalities



Promote the layout modality and how to obtain it efficiently



A striking, ultra-realistic poster featuring a heartfelt "Thank You" message spelled out in a modern, bold font. The background is a visual representation of a million business documents, with different colors, patterns, and textures, creating a dynamic and visually rich atmosphere., poster @  Ideogram 23