CIVQA

Czech Invoice Visual Question Answering

Šárka Ščavnická

Faculty of Informatics, Masaryk University

November 9, 2023

Visually Rich Documents

VRD

VRD contains such documents whose semantic structure is not determined only by the text but also by the layout and visual elements of the documents

Figure: Example of VRD [1]

Document visual question-answering

- DVQA seeks to obtain knowledge from the documents' visual and textual parts to answer questions
- The asked questions may relate to different parts of the VRDs
  - text
  - inserted images
  - tables
  - forms

Entity	Numeric	Textual	Pattern	Shape
Invoice number	X			
Variable symbol	X			
Specific symbol	X			
Constant symbol	X			
Bank code	X			X
Account number	X			X
ICO	X			X
Total amount	X			
Invoice date	X			X
Due date	X			X
Name of supplier		X		
IBAN	X	X		X
DIC	X	X		X
QR code			X	X
Supplier's address		X		

Table: CIVQA dataset's entities' categories

Tesseract and EasyOCR CIVQA dataset

- Tesseract OCR
  - was developed at HP Research between 1984 and 1994
  - Open-source project since 2005
  - Can recognise more than 100 different languages, including Czech
- EasyOCR
  - Python framework created by Jaded AI
  - Can recognise just over eighty languages, including Czech
- Each type of these dataset has two different versions
  - Readable by human
  - Ready to use

Sliding window

- Maximum Length = 512 tokens
- Special tokens

  [CLS] *question tokens* [SEP] *word tokens* [SEP]
  [PAD] token

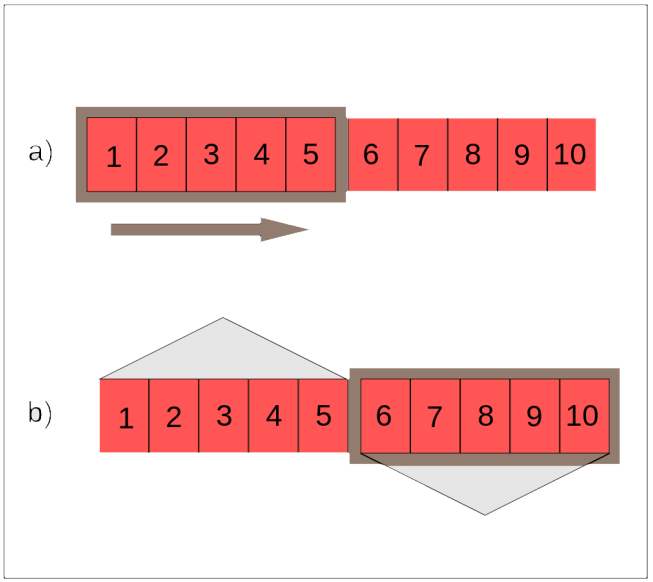

Figure: Sliding window technique [2]

Models

- LayoutLMv2
- LayoutXLM

  Chinese, Japanese, Spanish, French, Italian, German, and Portuguese
- LayoutLMv3
- Impira LayoutLM Invoices

  fine-tuned on the SQuAD and DocVQA datasets plus proprietary dataset of invoices
- Impira LayoutLM Document QA

  fine-tuned on the SQuAD and DocVQA datasets

Tesseract OCR vs EasyOCR

Model	Precision	Recall	F1 score
LayoutXML	**0,7422**	**0,7117**	**0,7079**
LayoutLMv2	0,6917	0,6750	0,6634
LayoutLMv3	0,6989	0,6382	0,6410
Impira QA	0,6773	0,6291	0,6313
Impira Invoice	0,6948	0,6440	0,6434

Table: CIVQA Tesseract OCR results

Model	Precision	Recall	F1 score
LayoutXML	**0,6636**	**0,6633**	**0,6455**
LayoutLMv2	0,6323	0,6129	0,6011
LayoutLMv3	0,6370	0,6164	0,6065
Impira QA	0,6373	0,6015	0,5984
Impira Invoice	0,6345	0,6019	0,5962

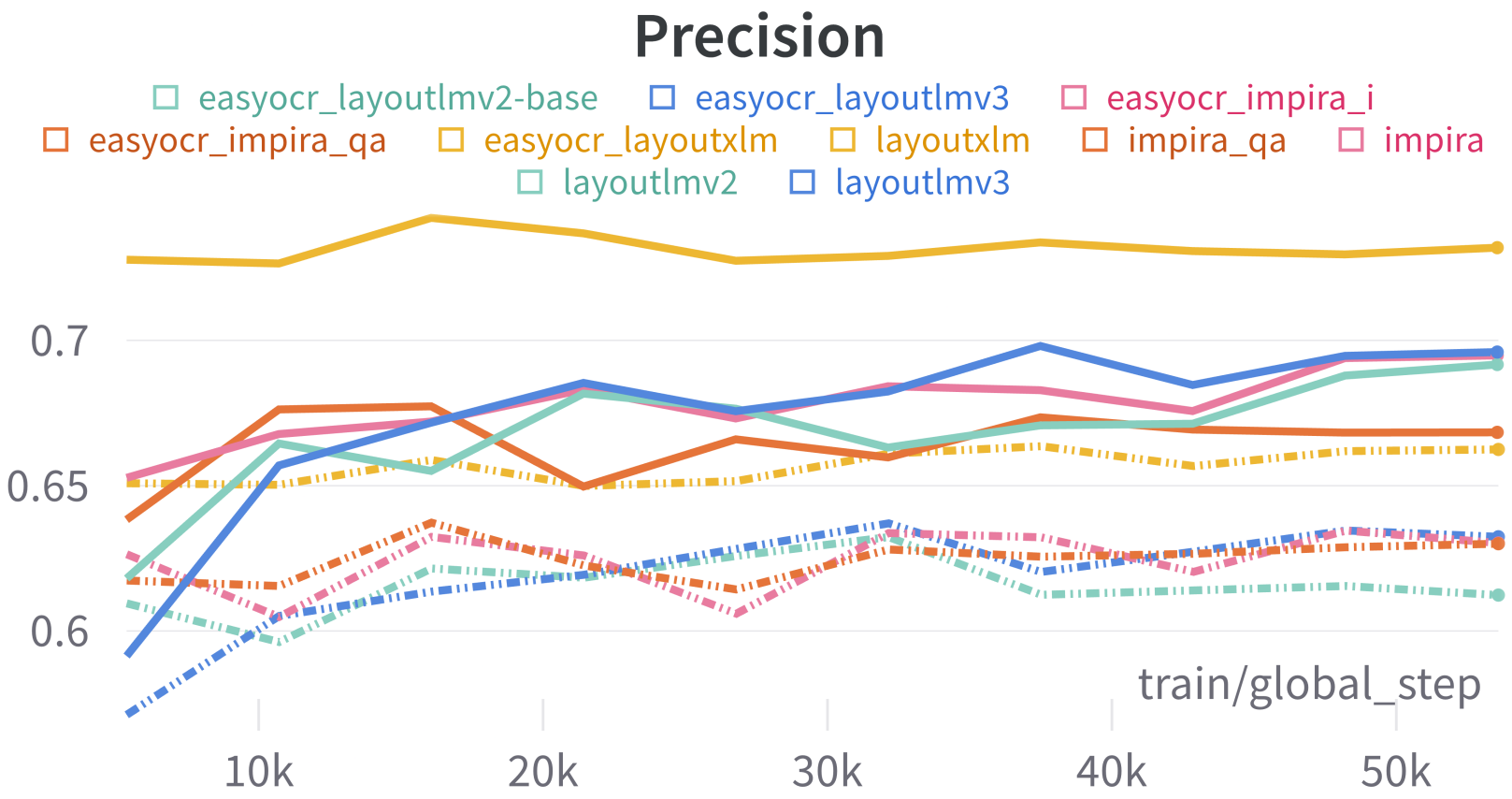

Figure: The precision of the models in the first experiment.

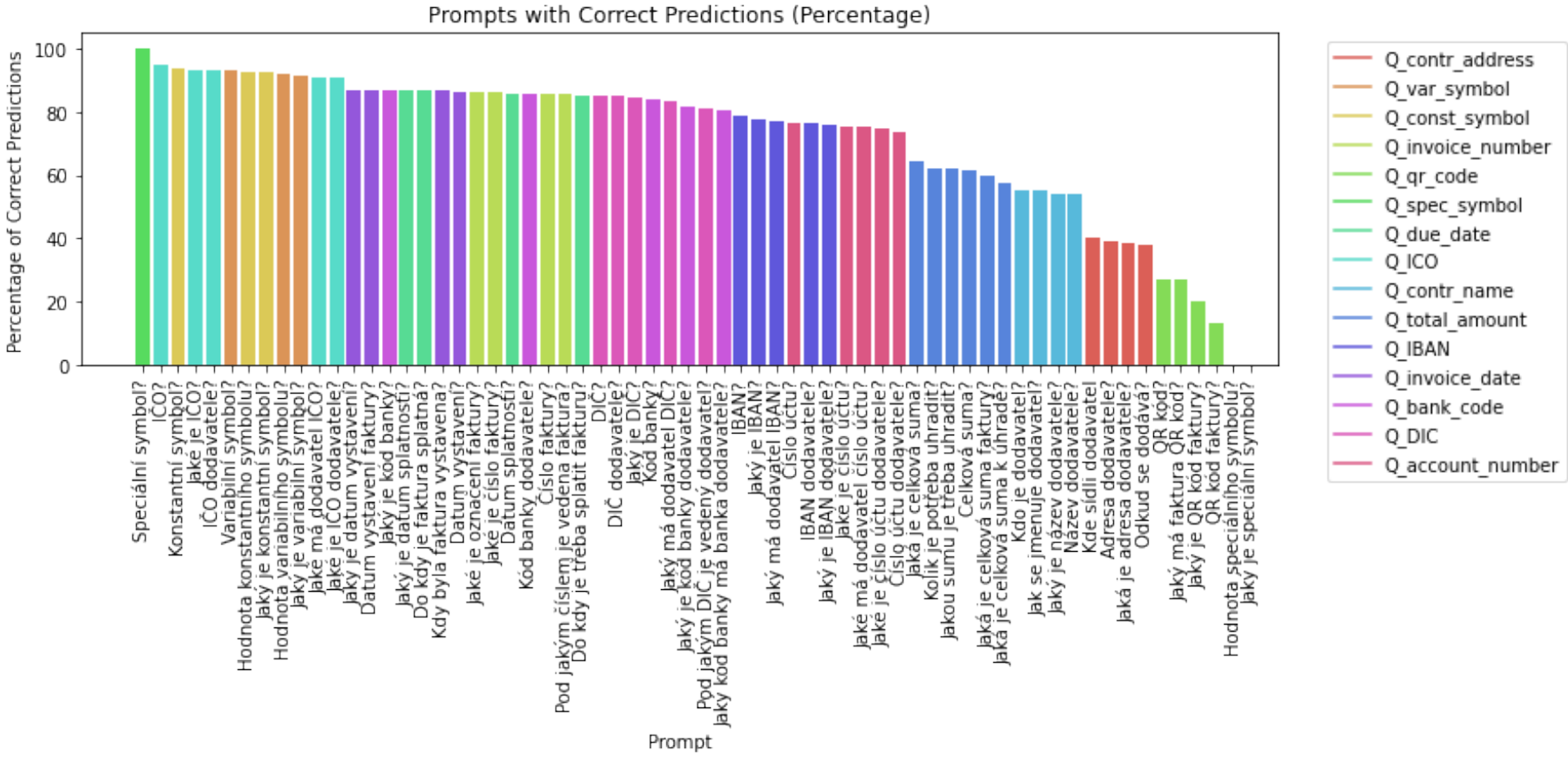

Figure: Validation dataset of CIVQA Tesseract OCR: LayoutXLM model success rate by individual question percentage

	prompt	prediction	true
1	Jaká je adresa dodavatele?	Rosinská cesta 13 010 08 Žilina	Rosinská cesta 13 010 08 Žilina

Figure: The correct answer is on one line.

	prompt	prediction	true
99	Jaká je adresa dodavatele?	Michalovská 62 Variabilný symbol: 1858420000 073 01 Sobrance	073 01 Sobrance

Figure: The correct answer is on multiple lines, so it was split.

CIVQA and unseen types of questions

In this set of experiments, our focus was on developing a practical and robust solution for unseen entities.

- **Invoice number**

  A numerical entity without a fixed shape.
- **ICO**

  A numerical entity with given shape.
- **Supplier's address**

  Textual and numerical entity without a fixed shape.
- **IBAN**

  Textual and numerical entity with a fixed shape.
- **Due date**

  A numerical entity with given shape.

Training with a subset of unknown data

In this experiment, we have tried introducing a different amount of unknown data to the trained models. We choose 5%, 15%, 30% and 50% and compare the results to see how they affect the models.

Model	Precision	Recall	F1 score
LayoutXML	**0,19200**	0,04128	0,05816
LayoutLMv2	0,03427	0,02695	0,02605
LayoutLMv3	0,10220	0,03411	0,04557
Impira QA	0,15120	0,04554	0,06520
Impira Invoice	0,13600	**0,05304**	**0,07235**

Table: CIVQA Tesseract OCR results on unknow entities

Model	Precision	Recall	F1 score
LayoutXML	**0,7002**	**0,6594**	**0,6617**
LayoutLMv2	0,5944	0,5154	0,5192
LayoutLMv3	0,5793	0,5125	0,5254
Impira QA	0,6186	0,5356	0,5466
Impira Invoice	0,5999	0,5255	0,5369

Table: CIVQA Tesseract OCR results on unknow entities: 5%

Model	Precision	Recall	F1 score
LayoutXML	**0,7078**	**0,6911**	**0,6844**
LayoutLMv2	0,6201	0,5717	0,5718
LayoutLMv3	0,6377	0,5755	0,5825
Impira QA	0,6491	0,5907	0,5935
Impira Invoice	0,6410	0,5849	0,5880

Table: CIVQA Tesseract OCR results on unknow entities: 15%

Model	Precision	Recall	F1 score
LayoutXML	**0,7297**	**0,7124**	**0,7069**
LayoutLMv2	0,6852	0,6619	0,6552
LayoutLMv3	0,6751	0,6497	0,6465
Impira QA	0,6815	0,6464	0,6447
Impira Invoice	0,6772	0,6454	0,6421

Table: CIVQA Tesseract OCR results on unknow entities: 30%

Model	Precision	Recall	F1 score
LayoutXML	**0,7360**	**0,7106**	**0,7069**
LayoutLMv2	0,6923	0,6488	0,6508
LayoutLMv3	0,6876	0,6573	0,6566
Impira QA	0,7004	0,6560	0,6566
Impira Invoice	0,6994	0,6720	0,6559

Table: CIVQA Tesseract OCR results on unknow entities: 50%

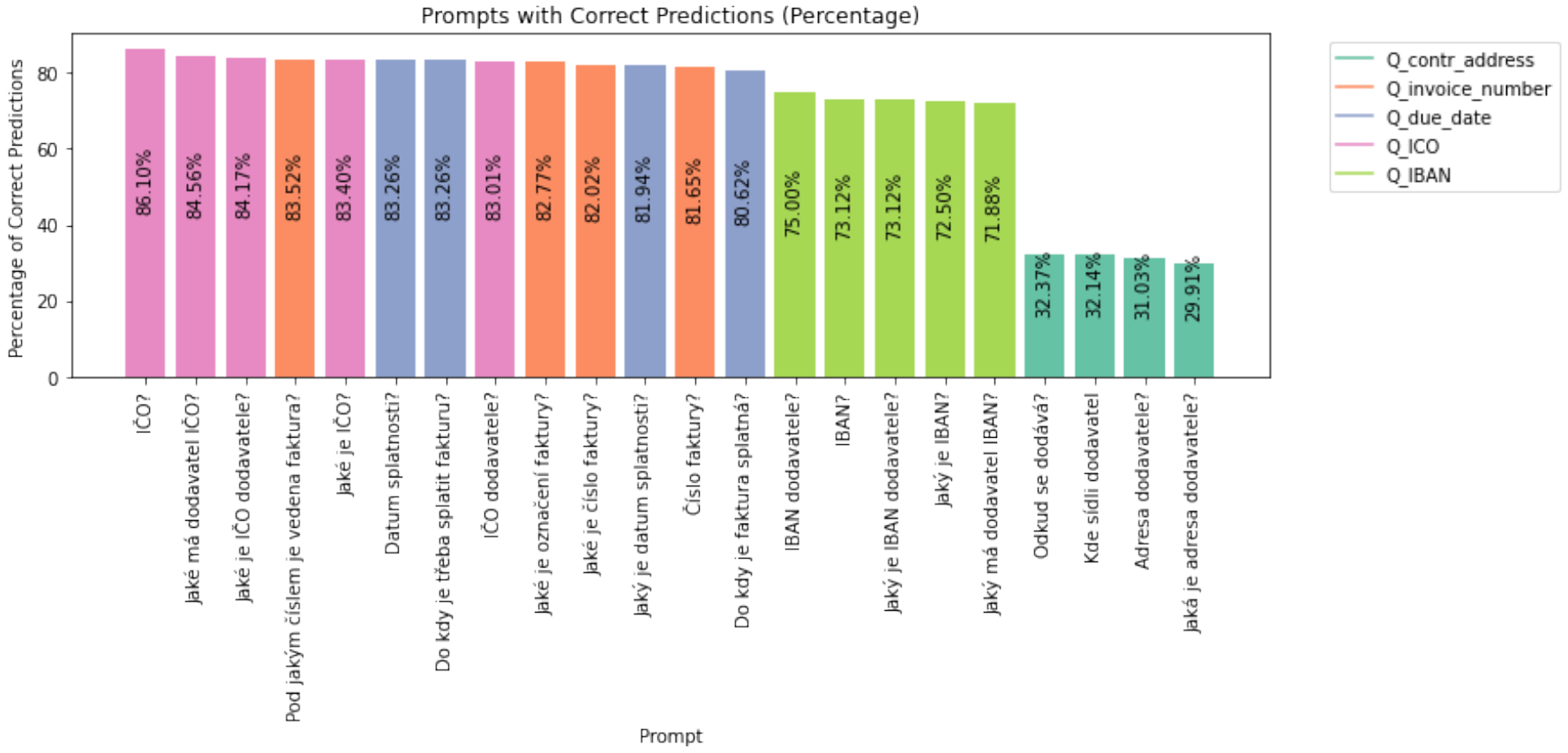

Figure: Validation dataset of CIVQA Tesseract OCR unknown: LayoutXLM model success rate by individual question percentage with 5% training.

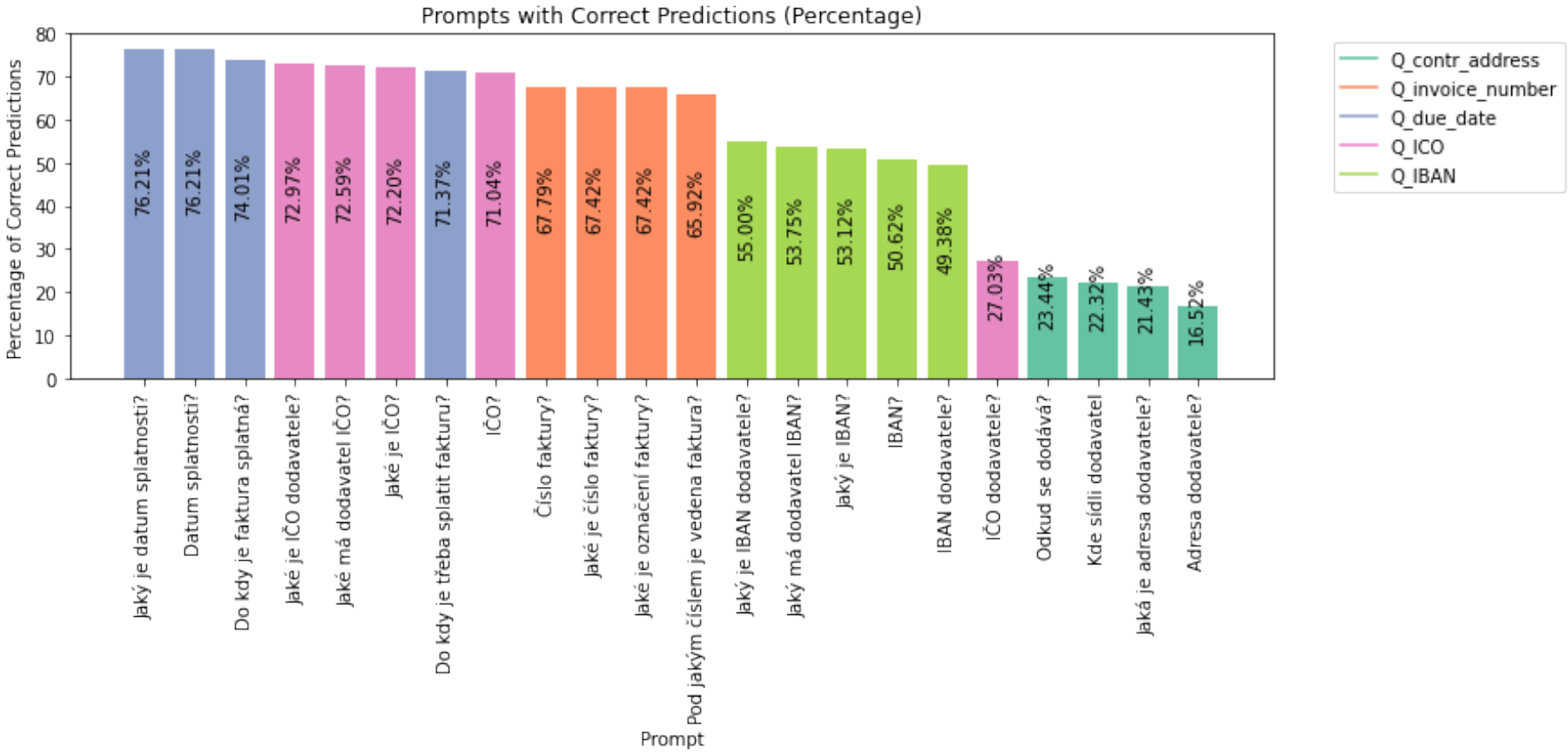

Figure: Validation dataset of CIVQA Tesseract OCR unknown: LayoutLMv3 model success rate by individual question percentage with 5% training.

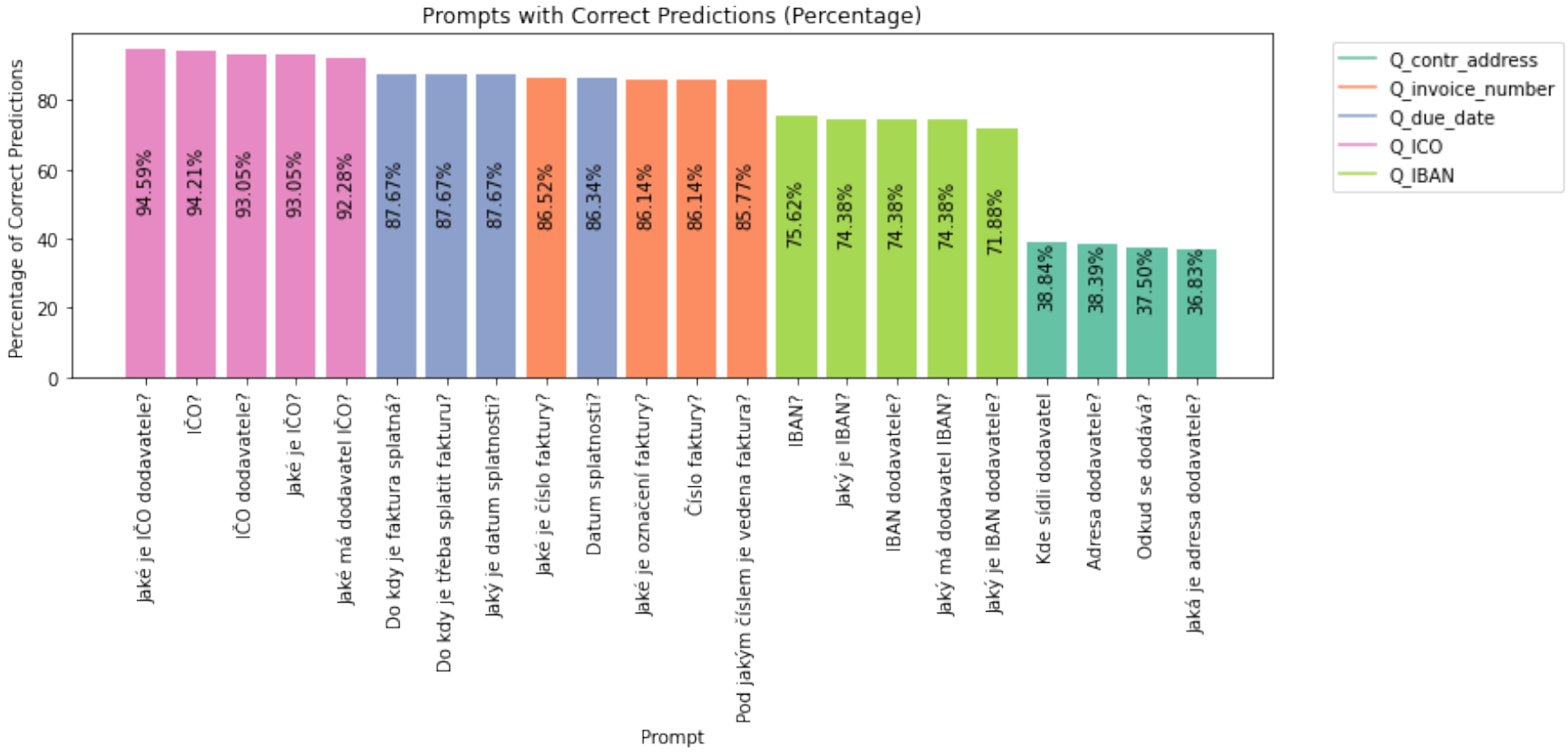

Figure: Validation dataset of CIVQA Tesseract OCR unknown: LayoutXML model success rate by individual question percentage with 50% training.

Training with a subset of unknown data concatenated with the known data dataset

Model	Precision	Recall	F1 score
LayoutXML	**0,7069**	**0,6693**	**0,67**
LayoutLMv2	0,6223	0,5726	0,5755
LayoutLMv3	0,6344	0,5528	0,5631
Impira QA	0,6318	0,5487	0,5670
Impira Invoice	0,6353	0,5577	0,5681

Table: CIVQA Tesseract OCR results on unknow entities concatenated with the known data dataset: 5%

Model	Precision	Recall	F1 score
LayoutXML	**0,7069**	**0,6693**	**0,67**
LayoutLMv2	0,6223	0,5726	0,5755
LayoutLMv3	0,6344	0,5528	0,5631
Impira QA	0,6318	0,5487	0,5670
Impira Invoice	0,6353	0,5577	0,5681

Table: CIVQA Tesseract OCR results on unknow entities concatenated with the known data dataset: 5%

Model	Precision	Recall	F1 score
LayoutXML	**0,7002**	**0,6594**	**0,6617**
LayoutLMv2	0,5944	0,5154	0,5192
LayoutLMv3	0,5793	0,5125	0,5254
Impira QA	0,6186	0,5356	0,5466
Impira Invoice	0,5999	0,5255	0,5369

Model	Precision	Recall	F1 score
LayoutXML	**0,7238**	**0,664**	**0,6919**
LayoutLMv2	0,6428	0,5615	0,5715
LayoutLMv3	0,6591	0,5831	0,5858
Impira QA	0,6629	0,5849	0,5879
Impira Invoice	0,6658	0,6391	0,6359

Table: CIVQA Tesseract OCR results on unknow entities concatenated with the known data dataset: 15%

Model	Precision	Recall	F1 score
LayoutXML	**0,7078**	**0,6911**	**0,6844**
LayoutLMv2	0,6201	0,5717	0,5718
LayoutLMv3	0,6377	0,5755	0,5825
Impira QA	0,6491	0,5907	0,5935
Impira Invoice	0,6410	0,5849	0,5880

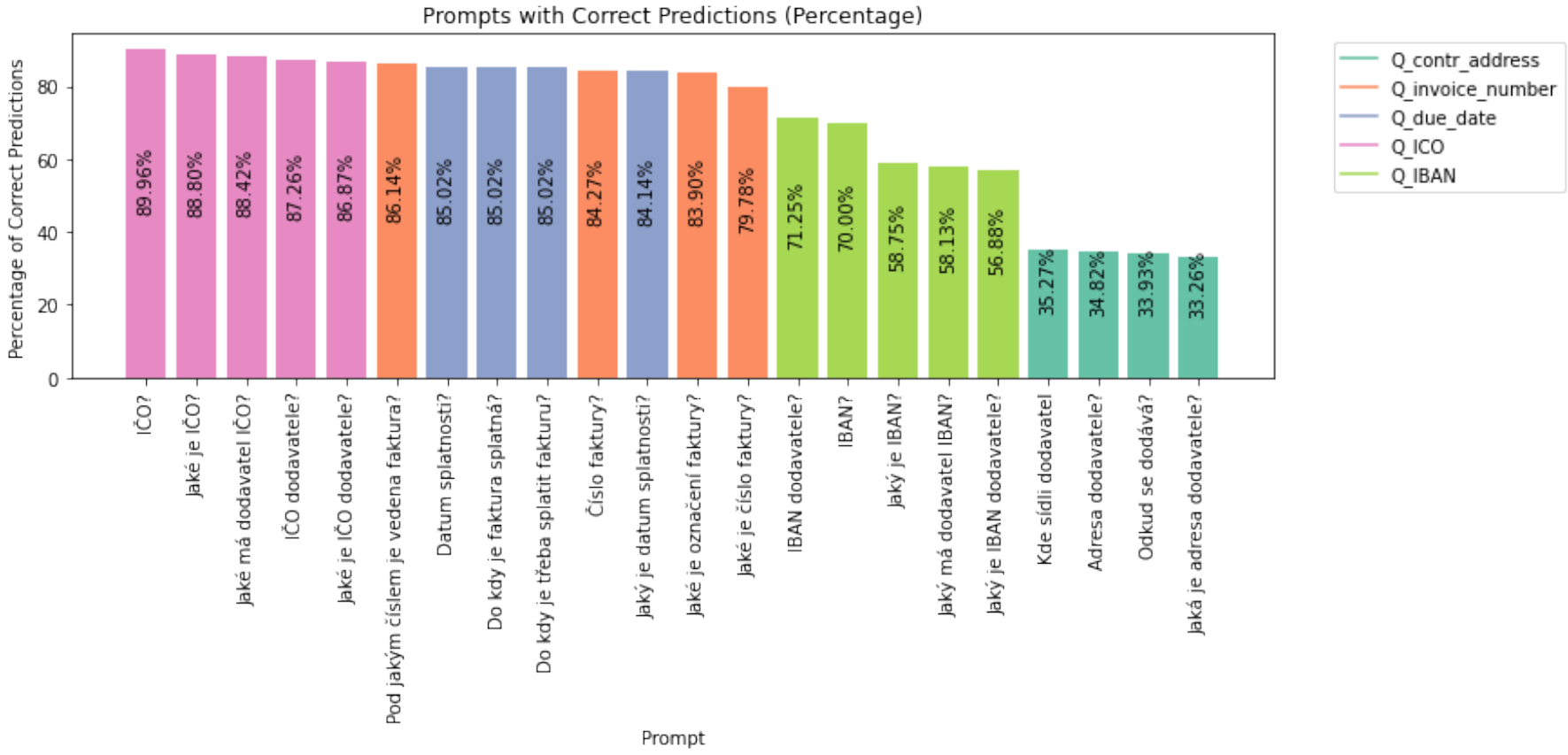

Figure: Validation dataset of CIVQA Tesseract OCR unknown: LayoutXML model success rate by individual question percentage with 5% training.

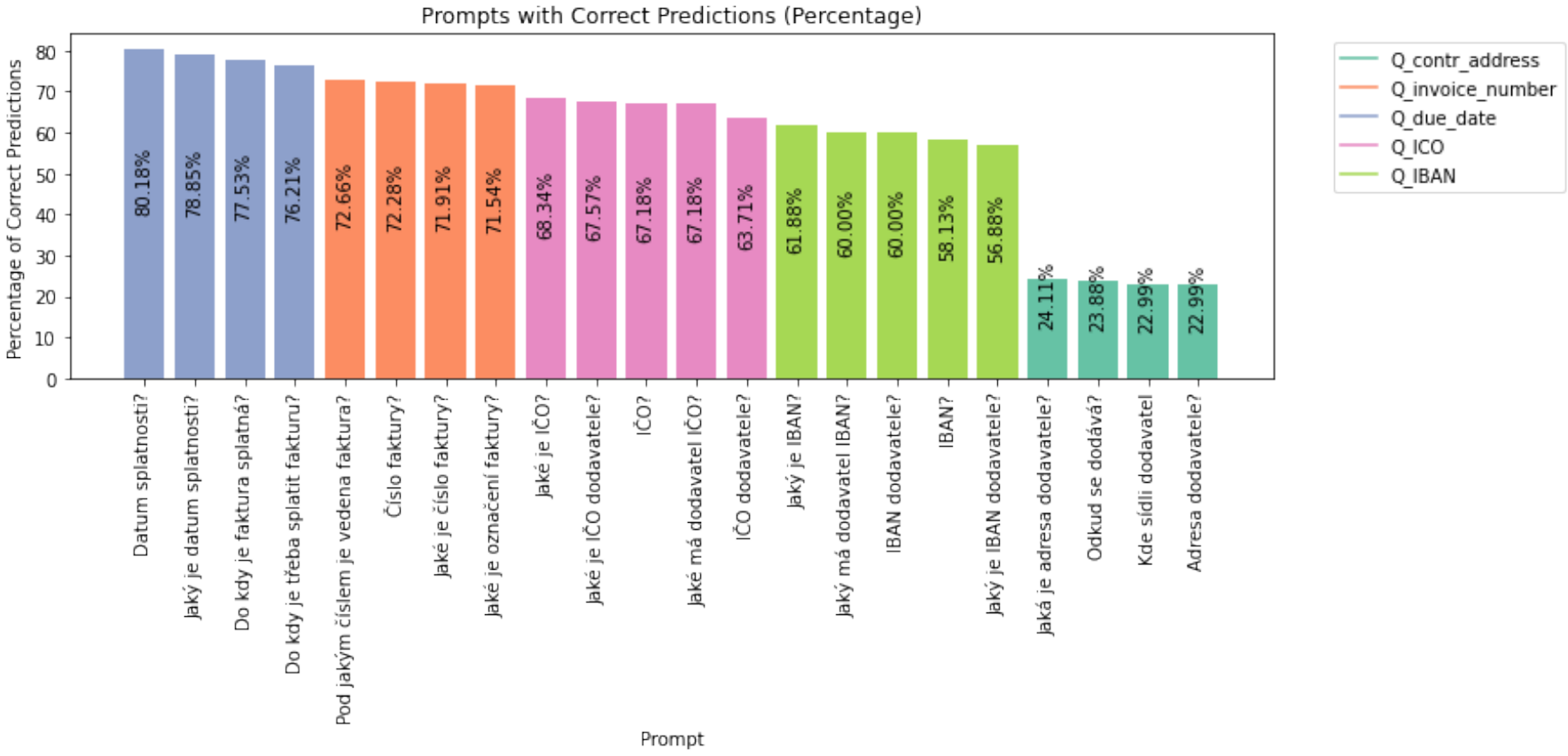

Figure: Validation dataset of CIVQA Tesseract OCR unknown: LayoutXML model success rate by individual question percentage with 5% training.

DocVQA and CIVQA known dataset

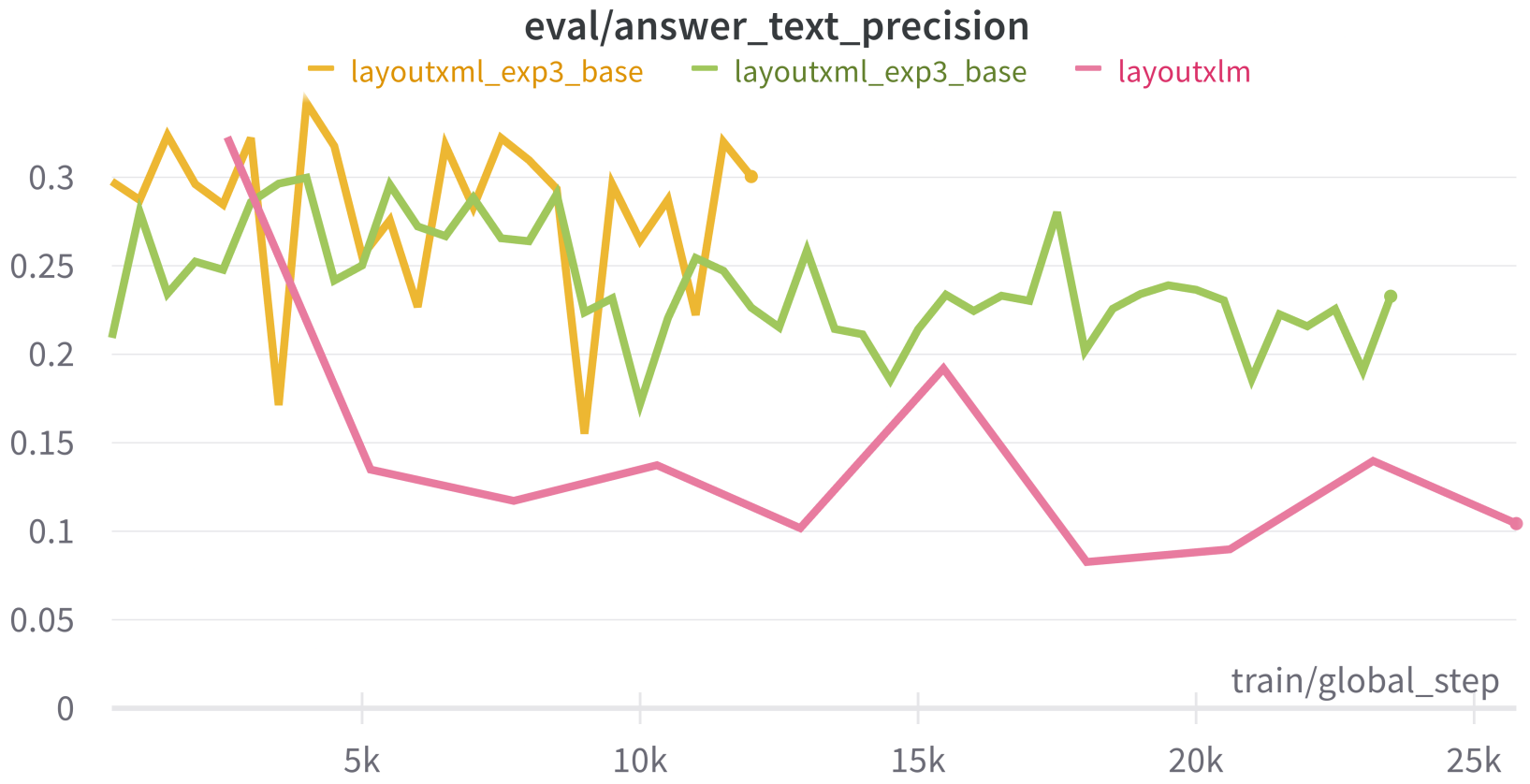

Figure: LayoutXML comparison

Bibliography I

[1] ISHITA JAISWAL ANKUR A. PATEL. *What is Intelligent Document Processing and How LayoutLM's Pre-Trained Model for Text and Image Understanding Works*. Accessed on 08.11. 2023. URL: https://www.ankursnewsletter.com/p/what-is-intelligent-document-processing.

[2] Long Nguyen. *Sliding Window – A common technique to solve algorithmic problems involving String/Array*. Accessed on 08.11. 2023. URL: https://medium.com/swlh/sliding-window-a-common-technique-for-solving-algorithmic-problems-involving-string-array-44adf35e2d5d.

Thank You for Your Attention!

MUNI
FACULTY
OF INFORMATICS