Hypergraph based Understanding for Document Semantic Entity Recognition

Qiwei Li†, Zuchao Li†*, Ping Wang, Haojun Ai* and Hai Zhao

ARCIDS-LAB 档案智能开发与服务重点实验室

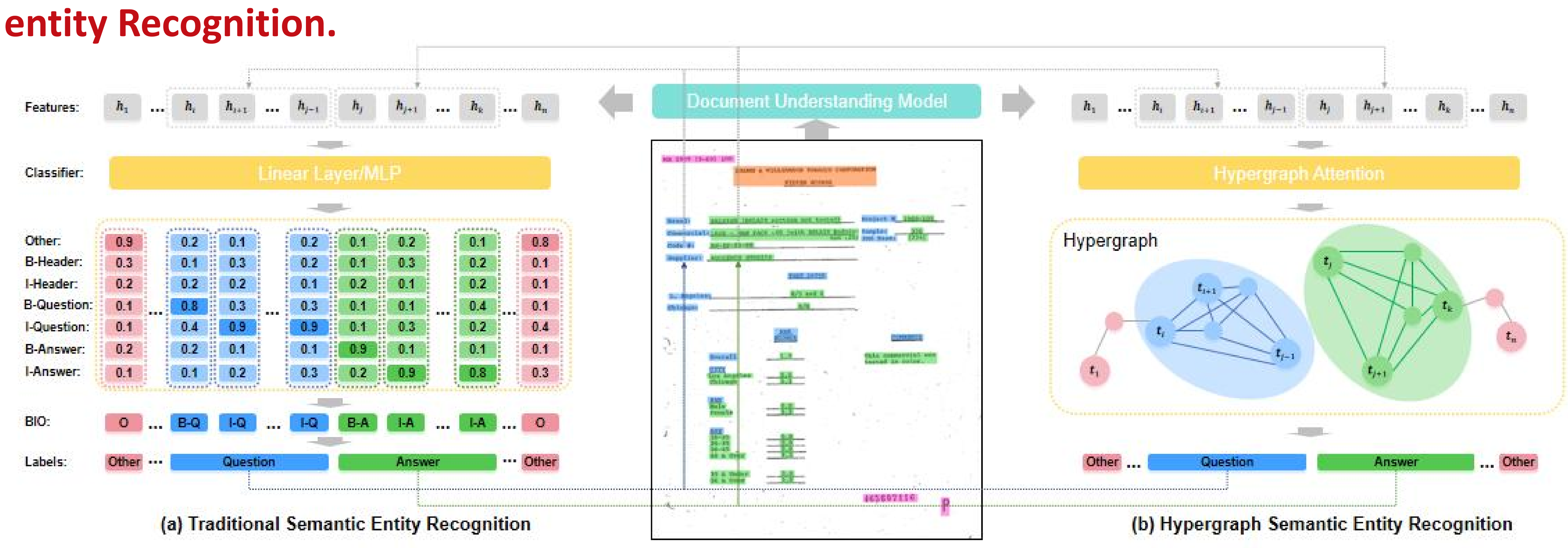
This paper is supported by the special scholarship of Wuhan University Postgraduate Exchange Program.

Document Semantic Entity Recognition

1.Definition of semantic entity recognition.

Semantic entity recognition (SER) of documents refers to the recognition of entities with specific meaning, such as people, objects, signs, etc., from document images. By identifying and analyzing these entities, we can further understand the content in the document.

3.Difference between traditional semantic entity recognition and hypergraph semantic entity Recognition.



(a) Traditional Semantic Entity Recognition

(b) Hypergraph Semantic Entity Recognition

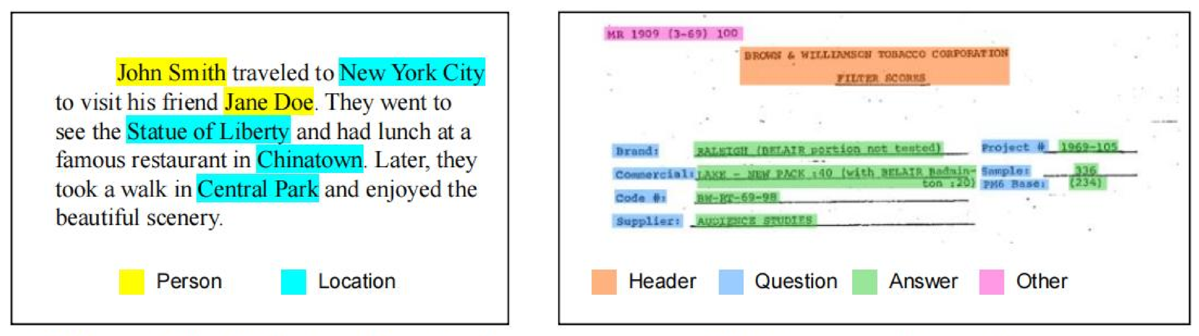
Traditional Semantic Recognition:

- BIO labeling method: For special labels, "Begin" label is used to indicate the beginning position of the entity, and "Inside" is used as the middle and end position of the entity. For text that is not a special entity, label it with "Other".
- The main focus is on the entity category. Entity boundaries and spans of discrete nodes in documents are ignored.

Hypergraph Semantic Entity Recognition:

- Hyperedge labeling method: it is only labeled according to the special entity category, and each type of hyperedge represents a special entity.
- It not only focuses on the entity categories, but also focuses on the boundaries of special entities and the spans of discrete nodes in the document.

2.Difference between named entity recognition and semantic entity recognition.



(a) Named Entity Recognition Task

(b) Document Semantic Entity Recognition Task

Named Entity Recognition (SER):

- The text form of a single modal text task is a fixed text sequence.
- The NER task of a single modal text only needs to consider the semantic relationship between the tokens in the text sequence.
- The span range of entity tags of NER task is flexible.

Semantic Entity Recognition (SER):

- The discrete text in a document is composed of text nodes in different locations.
- The SER task on the document needs to consider not only the semantic relationship between nodes, but also the position relationship between nodes.
- The range of task tags of semantic entity recognition task on document is affected by nodes. Texts of the same node in the document share the same label in most cases.

Hypergraph Attention Method

1.Hypergraph Attention Head.

Assume the document token sequence $x = \{x_1, x_2, ..., x_n\}$ is hypergraph node sets. The understanding document model will convert x into high-dimensional feature representation sequence:

$$h = \{h_1, h_2, ..., h_n\} = \text{Model}(\{x_1, x_2, ..., x_n\})$$

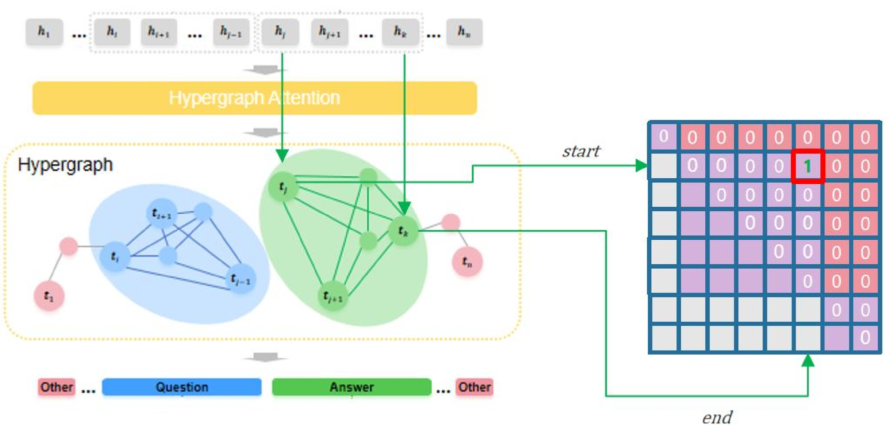
Based on h, we can obtain the query vector q and the key vector k:

$$q = \{q_\alpha : W_{q,\alpha} h + b_{q,\alpha}\}$$
$$k = \{k_\alpha : W_{k,\alpha} h + b_{k,\alpha}\}$$

The hypergraphs can be represented by a self-attention score calculated by q and k:

$$s = q^T k = \{s_\alpha(i,j) : q_{i,\alpha}^T k_{j,\alpha}, i \in \mathbb{Z}^L, j \in \mathbb{Z}^L\}$$

The $s_\alpha(i,j)$ is the attention score at the α type hyperedge span with $[i,j]$. $q_{i,\alpha}$ and $k_{j,\alpha}$ are the start and end of the span with $[i,j]$ in the α type hyperedge matrix.



A Hypergraph Matrix Sample

Token feature sequence $h = \{h_1, h_2, ..., h_n\}$ and text node sequence $N = \{N_1, N_2, ..., N_m\}$ has a surjective relation. This relational mapping can be defined as:

$$f(h_i) = N_j, h_i \in h, N_j \in N$$

Span position can be caculated by f:

$$p_i = Position(f(h_i))$$
$$= Position(N_j)$$
$$= j, h_i \in h, N_j \in N$$

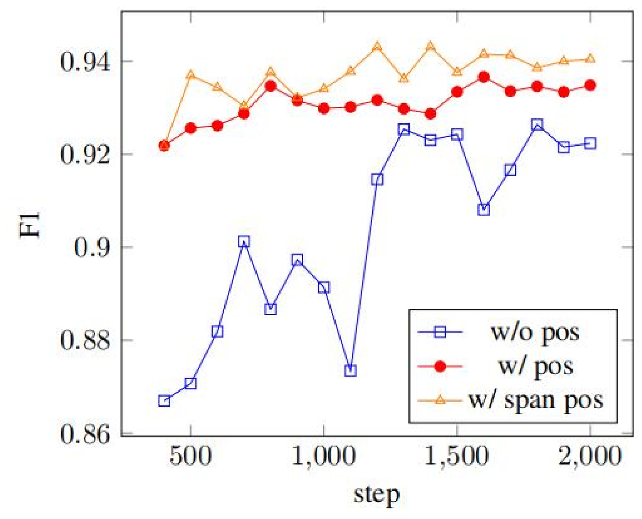
2.Span Position Encoding.

Update the hypergraphs attention score with span position encoding:

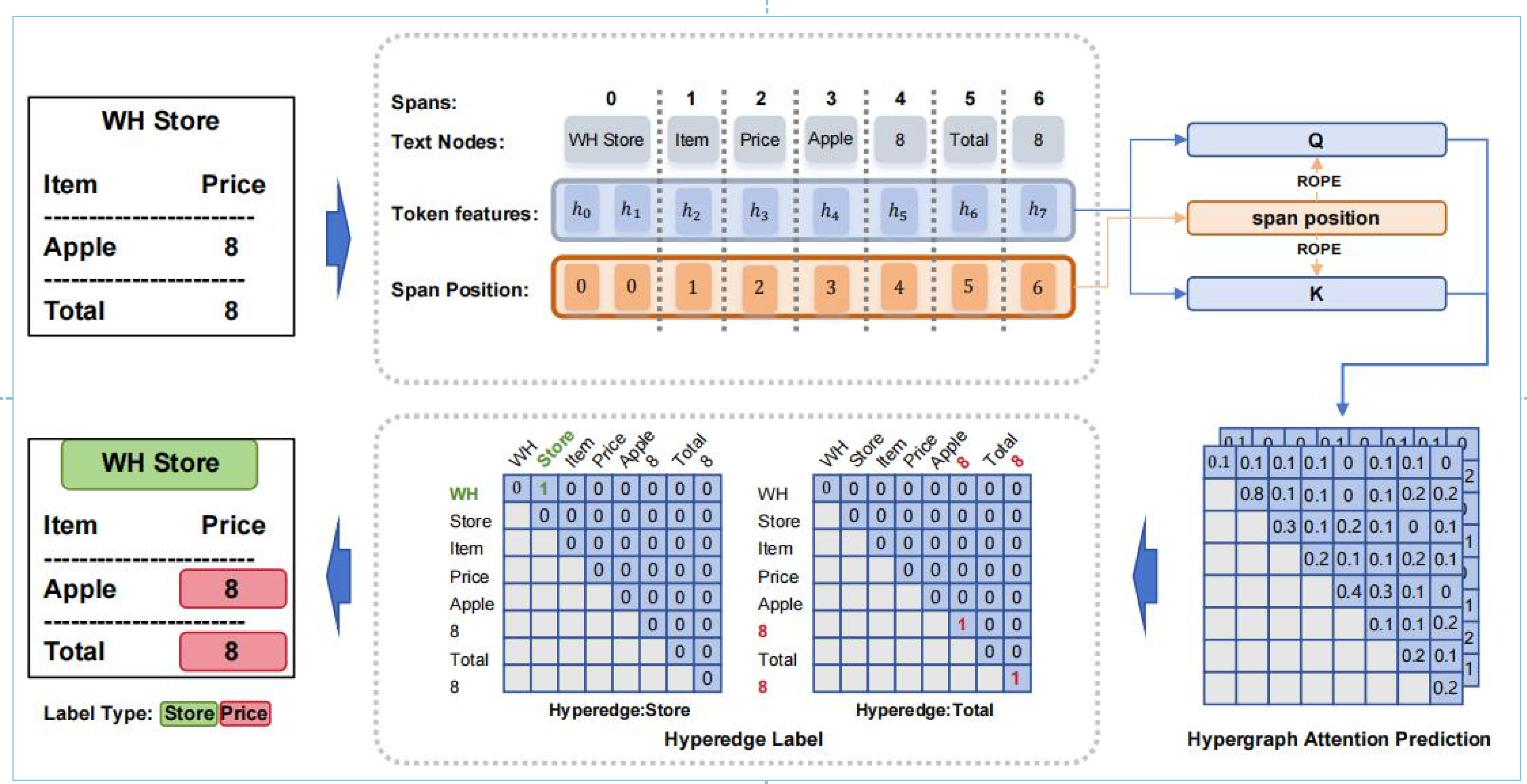
$$s_\alpha(i,j) = (\mathcal{R}_i p_{i,\alpha})^T (\mathcal{R}_j k_{j,\alpha})$$
$$= p_{i,\alpha}^T \mathcal{R}_i^T \mathcal{R}_j k_{j,\alpha}$$
$$= p_{i,\alpha}^T \mathcal{R}_{j-i} k_{j,\alpha}$$

Add lower triangular mask:

$$s_\alpha(i,j) = p_{i,\alpha}^T \mathcal{R}_{j-i} k_{j,\alpha} + m_{tril}(i,j)$$



The Ablation Study of Position Encoding



3.Balanced Hyperedge Loss.

In the process of loss calculation, the positive sample indicates that there is a α type hyperedge span with $[i,j]$ in α type hypergraph, while the reverse is a negative sample.

$$P_\alpha = \{s_\alpha(i,j) | l_\alpha(i,j) = 1\}$$
$$N_\alpha = \{s_\alpha(i,j) | l_\alpha(i,j) = 0\}$$

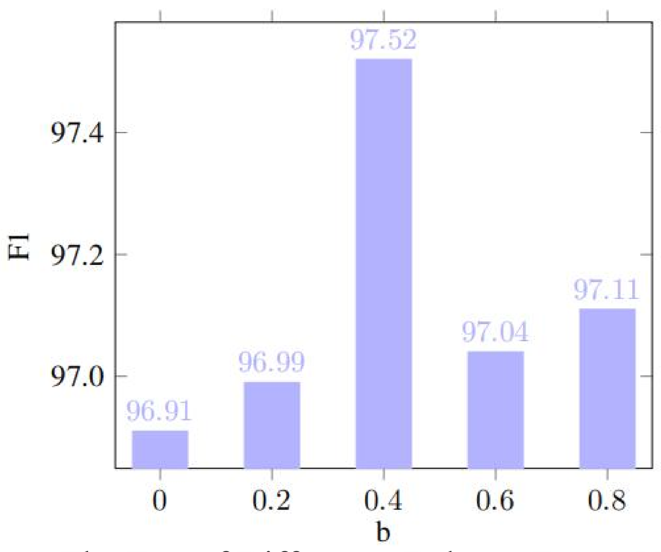
With the sets of positive and negative samples,we can get the positive sample loss L_p and the negative sample loss L_n:

$$\mathcal{L}_p = \log\left(1 + \sum_{(i,j)\in P_\alpha} e^{-s_\alpha(i,j)}\right)$$

$$\mathcal{L}_n = \log\left(1 + \sum_{(i,j)\in N_\alpha} e^{s_\alpha(i,j)}\right)$$

Gain the final loss with a balance factor $b \in [0, 1)$ to avoid the matrix sparsity caused by too many label types:

$$\mathcal{L} = (1+b)\mathcal{L}_p + (1-b)\mathcal{L}_n$$



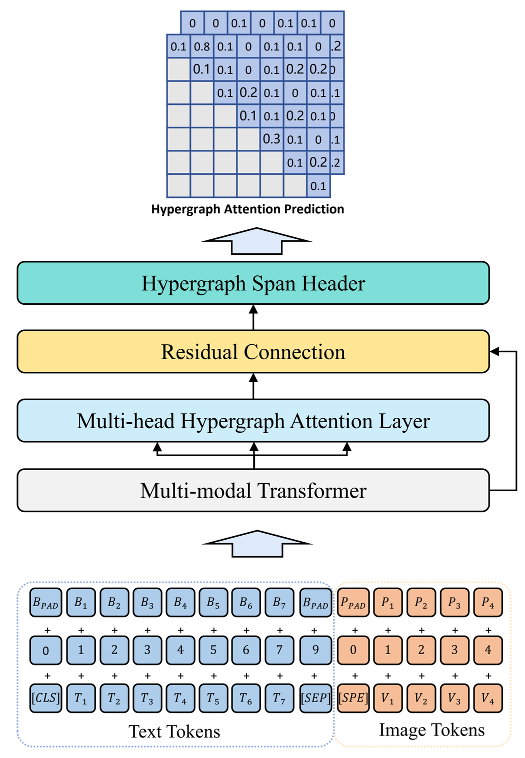
The Test of Difference Balance Factor Values

Head	FUNSD	CORD	SROIE	XFUND
Linear	93.48	96.98	98.99	93.03
MLP	93.58	97.13	99.28	93.48
HGA	94.32	97.52	99.53	94.22

Comparison Results of GraphLayoutLM with Different Types of Heads

4.HGALayoutLM.

GraphLayoutLM is used as the base model for feature encoding. The HGA method is used to help the model extract and classify semantic entities according to the text node span prompts.



Experiments

1.Experiment Settings.

Dataset	Label Num	Train	Dev	Test
FUNSD	3	149	-	50
CORD	30	800	100	100
SROIE	4	626	-	347
XFUND	3	149	-	50

Experiment Datsets

Dataset	Model size	Language	L	M	B	G
FUNSD	BASE	English	1e-5	2000	4	1
	LARGE		1e-5	2000	4	1
CORD	BASE	English	5e-5	2000	4	1
	LARGE		5e-5	3000	4	1
SROIE	BASE	English	1e-5	2000	4	1
	LARGE		1e-5	2000	4	1
XFUND	BASE	CHINESE	7e-5	2000	8	4

Finetuning Hyper-parameters

Model	Head	XFUND P	XFUND R	XFUND F
$\text{LayoutXLM}_{\text{BASE}}$	Linear	-	-	89.24
XYLayoutLM	Linear	-	-	91.76
$\text{LayoutLMv3}_{\text{BASE}}$	Linear	89.80	94.35	92.02
$\text{GraphLayoutLM}_{\text{BASE}}$	Linear	91.80	95.38	93.56
$\text{GraphLayoutLM}^{\dagger}_{\text{BASE}}$	Linear	92.30	94.69	93.48
$\text{HGALayoutLM}_{\text{BASE}}$	HGA	**92.79**	**95.70**	**94.22**

Comparison Results on Chinese Datsets

2.Main Results.

Model	Head	FUNSD P	FUNSD R	FUNSD F	CORD P	CORD R	CORD F	SROIE P	SROIE R	SROIE F
$\text{BERT}_{\text{BASE}}$	Linear	54.69	67.10	60.26	88.33	91.07	89.68	90.99	90.99	90.99
$\text{LayoutLM}_{\text{BASE}}$	Linear	75.97	81.55	78.66	94.37	95.08	94.72	94.38	94.38	94.38
$\text{BROS}_{\text{BASE}}$	Linear	81.16	85.01	83.05	-	-	96.50	-	-	96.28
$\text{LayoutLMv2}_{\text{BASE}}$	Linear	80.29	85.39	82.76	94.53	95.39	94.95	96.25	96.25	96.25
$\text{LayoutXLM}_{\text{BASE}}$	Linear	-	-	79.40	-	-	-	-	-	-
XYLayoutLM	Linear	-	-	83.35	-	-	-	-	-	-
$\text{LayoutLMv3}_{\text{BASE}}$	Linear/MLP	90.82	91.55	91.19	96.35	96.71	96.53	100	100	100
$\text{GraphLayoutLM}_{\text{BASE}}$	Linear/MLP	92.46	**93.85**	93.15	97.02	**97.53**	97.28	-	-	99.30
$\text{GraphLayoutLM}^{\dagger}_{\text{BASE}}$	Linear/MLP	93.62	93.25	93.43	96.87	97.38	97.13	98.40	**99.58**	98.99
$\text{HGALayoutLM}_{\text{BASE}}$	HGA	**94.84**	93.80	**94.32**	**97.89**	97.16	**97.52**	**99.58**	99.48	**99.53**
$\text{BERT}_{\text{LARGE}}$	Linear	61.13	70.85	65.63	88.86	91.68	90.25	92.00	92.00	92.00
$\text{LayoutLM}_{\text{LARGE}}$	Linear	75.69	82.19	78.95	94.32	95.54	94.93	95.24	95.24	95.24
$\text{BROS}_{\text{LARGE}}$	Linear	82.81	86.31	84.52	-	-	97.28	-	-	96.62
$\text{LayoutLMv2}_{\text{LARGE}}$	Linear	83.24	85.19	84.20	95.65	96.37	96.01	99.04	96.61	97.81
$\text{ERNIE-Layout}_{\text{LARGE}}$	Linear	-	-	93.12	-	-	97.21	-	-	97.55
$\text{LayoutLMv3}_{\text{LARGE}}$	Linear/MLP	91.51	92.70	92.10	97.45	97.52	97.49	-	-	-
UDop	Decoder	-	-	92.08	-	-	97.58	-	-	-
GeoLayoutLM	Linear/MLP	-	-	92.86	-	-	**97.97**	-	-	-
$\text{GraphLayoutLM}_{\text{LARGE}}$	Linear/MLP	94.49	94.30	94.39	97.75	**97.75**	97.75	-	-	-
$\text{GraphLayoutLM}^{\dagger}_{\text{LARGE}}$	Linear/MLP	94.37	93.95	94.16	97.32	97.68	97.50	99.27	**99.58**	99.42
$\text{HGALayoutLM}_{\text{LARGE}}$	HGA	**95.67**	**94.95**	**95.31**	**97.97**	97.38	97.67	**99.69**	99.53	**99.61**

Comparison Results on English Datsets

3.Further Study.

Model	Head	Params	Flops
GraphLayoutLM	Linear	88.02M	63.03G
GraphLayoutLM	MLP	88.61M	63.45G
HGALayoutLM	HGA	88.31M	63.24G

Time Complexity and Parameter Scale Comparison

4.Comparison with LLM.

Model	FUNSD	CORD	Params	Flops
$\text{HGALayoutLM}_{\text{LARGE}}$	95.3	97.7	307.7M	218.95G
LayoutLLM	95.3	98.6	6914.38M	8654.62G

Experimental Comparison Results with Document LLM

ACL 2024 Bangkok, Thailand