Graph-based Document Structure Analysis

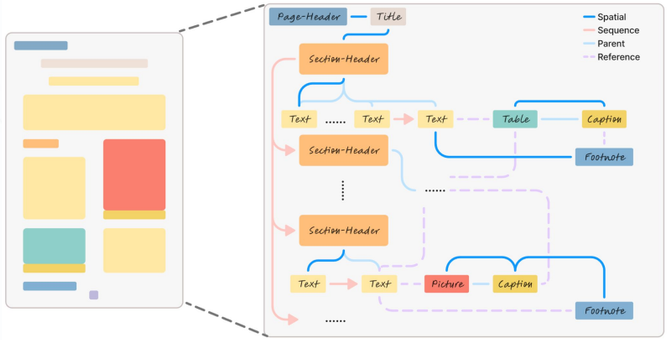
Yufan Chen, Ruiping Liu, Junwei Zheng, Di Wen, Kunyu Peng, Jiaming Zhang*, Rainer Stiefelhagen

Karlsruhe Institute of Technology

cv:hci @KIT Computer Vision for Human-Computer Interaction

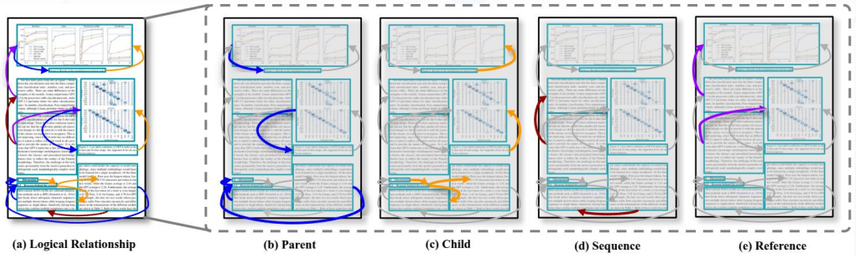
Towards Structured Understanding of Documents

- Traditional DLA focuses only on document layout element detection.
- Not allowed for relation analysis: reading order, hierarchy, and cross-references.
- We models document elements as nodes and spatial/logical relations as graph edges.
- This enables a more human-like and structured understanding of documents.



Graph-Based Document Structure Analysis

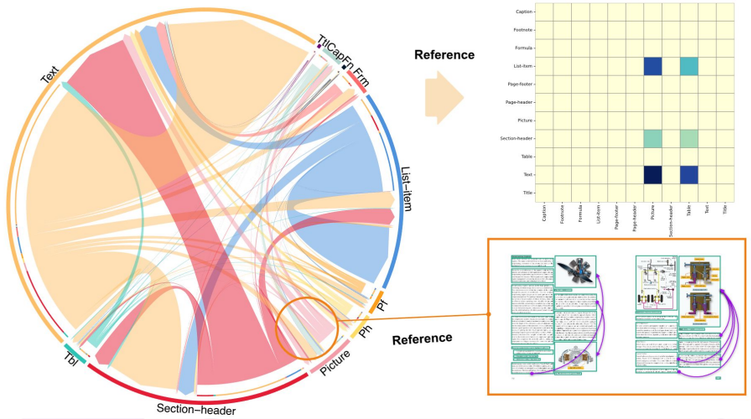
- Required **document layout detection** and **relation extraction**.



- Relation types:
  - **Spatial:** Up, Down, Left, Right
  - **Logical:** Parent, Child, Sequence, Reference
- Multiple relations can exist between layouts.
- Enables structure analysis tasks, e.g., hierarchical parsing, reference linking.

GraphDoc: Graph-based Document Dataset

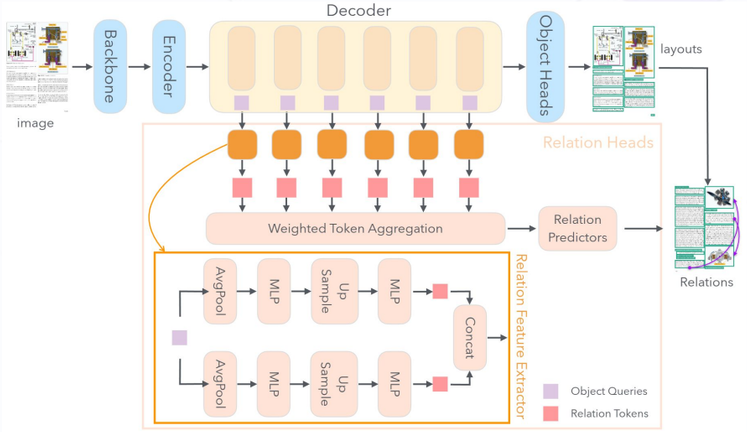
- **80K** documents, **1.1M** layout instances, and **4.13M** relation annotations.



- Relation distribution: **64%** Spatial, **36%** Logical
- Long-tail in distribution: **Reference** are **rare** but **critical**.

DRGG: Document Relation Graph Generator

- **Plug-and-play** Relation Generator
- **Parallel** Layout and Relation Extraction



Empirical Evaluation

Results with DRGG on GraphDoc Dataset:

Backbone	Detector	Relation Head	DLA	gDSA			
			mAP@50:5:95	mR_g@0.5	mAP_g@0.5	mAP_g@0.75	mAP_g@0.95
InternImage	RoDLA	-	80.5	-	-	-	-
InternImage	DETR	DRGG (Ours)	68.2	7.1	19.8	13.5	7.5
	Deformable DETR		73.4	11.5	25.4	11.8	8.5
	DINO		79.5	19.2	25.2	18.7	14.5
	RoDLA		**81.5**	**30.7**	**57.6**	**56.3**	**46.5**
ResNet	RoDLA	DRGG (Ours)	71.0	13.8	45.8	17.6	13.3
ResNeXt			77.9	16.9	40.3	18.4	13.6
Swin			73.7	11.4	26.1	13.5	7.9
InternImage			**81.5**	**30.7**	**57.6**	**56.3**	**46.5**

Per-category relation detection results:

Backbone	Detector	Relation Head	Up	Down	Left	Right	Parent	Child	Sequence	Reference
InternImage	DETR	DRGG (Ours)	32.4	29.7	8.9	8.9	22.8	18.8	27.7	8.9
	Deformable DETR		16.8	19.8	**99.0**	11.9	12.9	12.9	20.8	8.9
	DINO		37.1	38.3	18.8	18.8	11.9	15.8	53.5	7.6
	RoDLA		**49.0**	**49.0**	**99.0**	**99.0**	**45.5**	**45.5**	**56.4**	**16.8**
ResNet	RoDLA	DRGG (Ours)	15.1	17.2	27.7	27.7	6.9	4.0	17.8	16.8
ResNeXt			23.6	24.6	**99.1**	**99.1**	11.9	11.9	33.7	**18.8**
Swin			18.8	19.8	33.7	99.0	3.9	3.8	23.5	5.6
InternImage			**49.0**	**49.0**	99.0	99.0	**45.5**	**45.5**	**56.4**	16.8

Qualitative Results: