omni:us

CVC

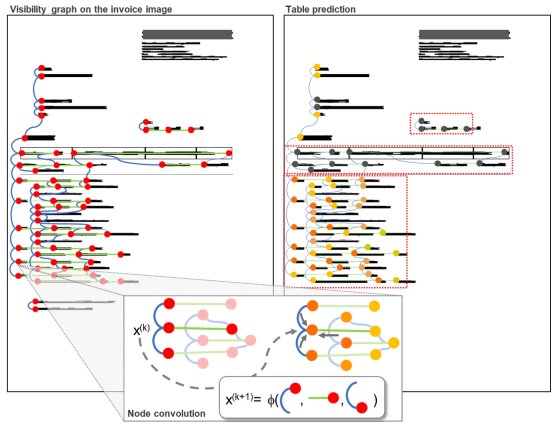
Centre de Visió per Computador

Table Detection in Invoice Documents by Graph Neural Networks

Pau Riba, Anjan Dutta, Lutz Goldmann, Alicia Fornés, Oriol Ramos, Josep Lladós

Computer Vision Center, omni:us

ICDAR, Sydney, Australia, 23rd September, 2019



UAB Universitat Autònoma de Barcelona

Generalitat de Catalunya

Parc de Recerca UAB

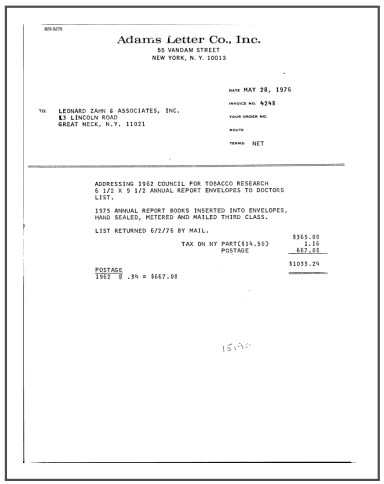
Outline

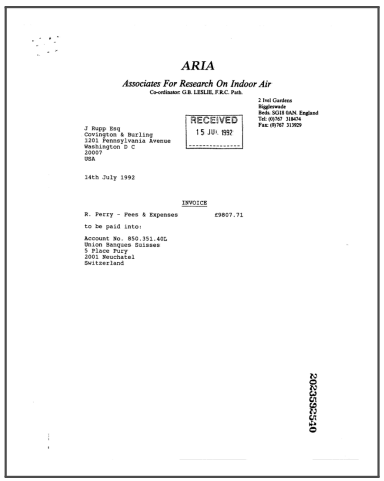
Introduction

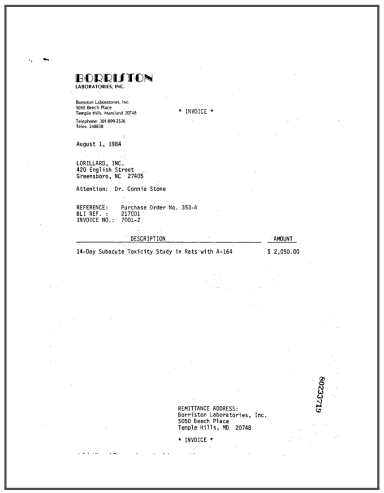
Introduction

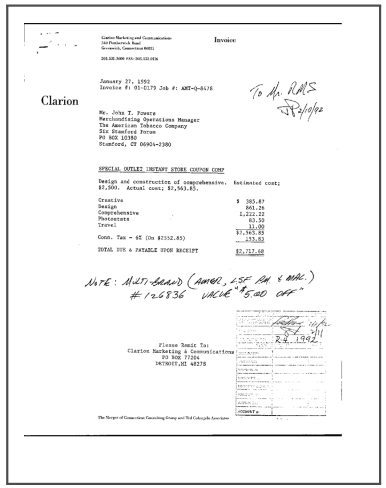
Business Documents

- **Information extraction**: Finance, insurance, manufacturing...
- **Manual extraction**: Tedious and time consuming.
- **Automatic extraction**: Reduced time and improved quality.





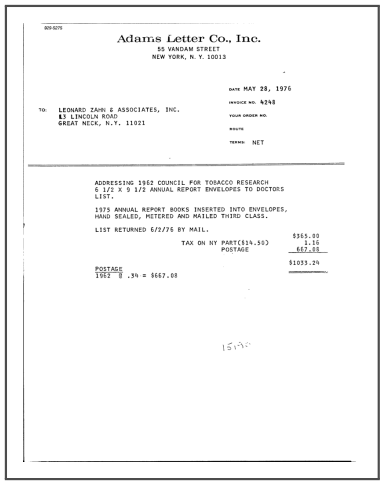


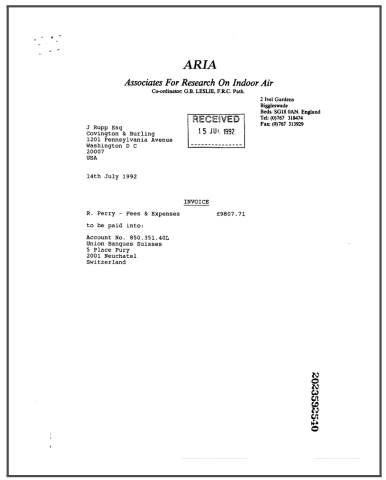


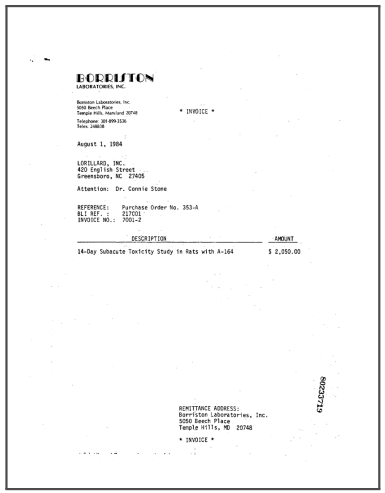
Introduction

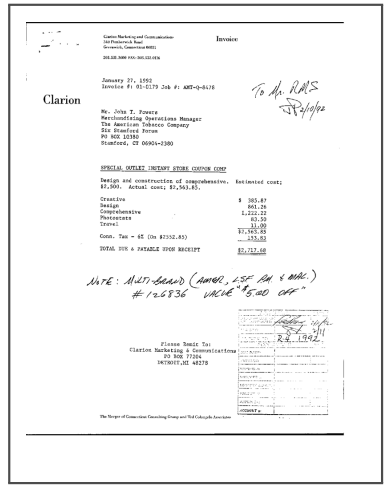
Semi-Structured Documents

- **Structured Documents**: Existing methods, high accuracy.
- **Unstructured Documents**: Human assistance and validation.
- **Semi-structured Documents**:
  - Without a fixed spatial layout.
  - Sharing a common set of components.





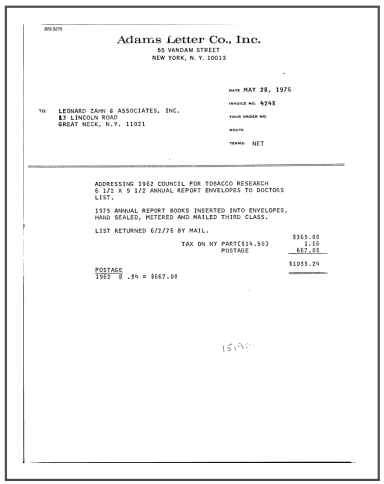


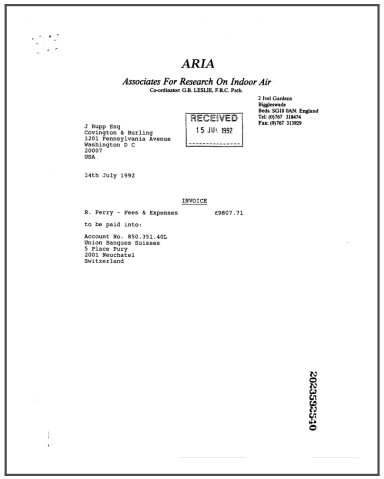


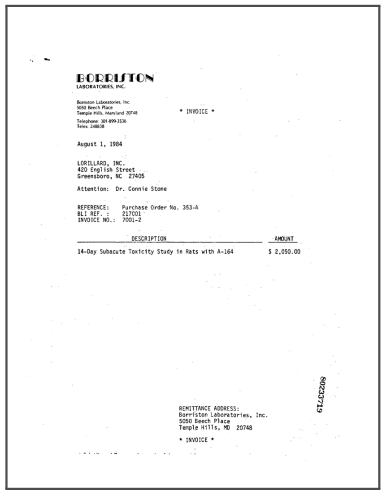
Introduction

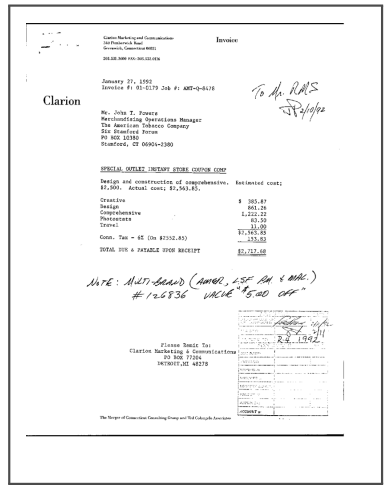
Invoice Documents

- **Semi-structured** documents with flexible layouts
- **Spatial arrangement** roughly perceived as a tabular layout
- **Tables** are commonly used to condense information





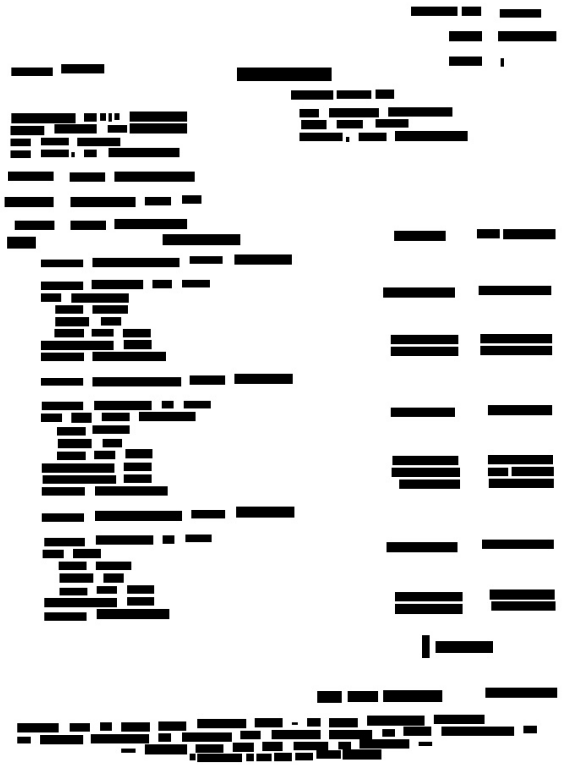




Introduction

Constraint

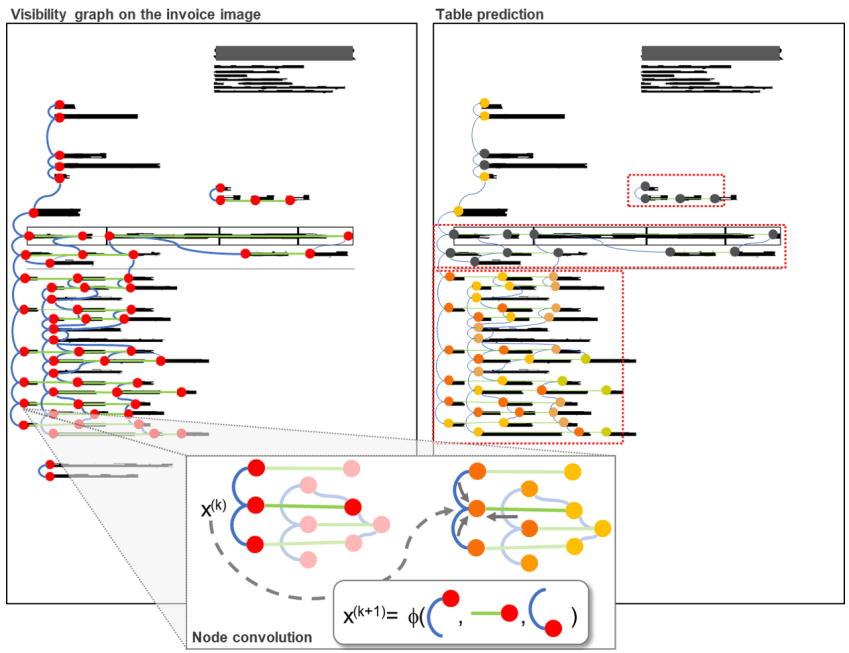
Industrial collaboration - Anonymized data



Introduction

Objective

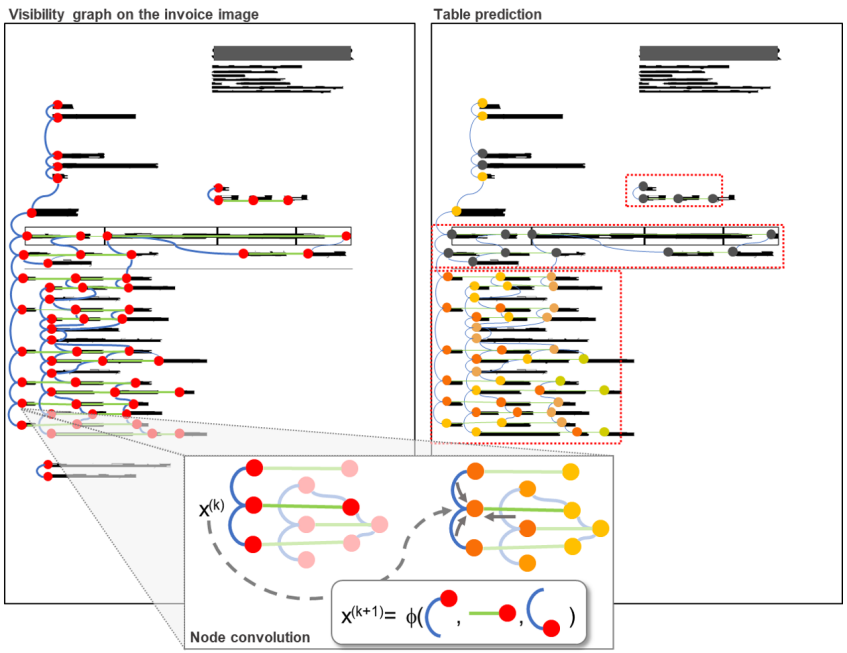
- **Graph based representation**: Exploit repetitive patterns



Introduction

Objective

- **Graph based representation**: Exploit repetitive patterns
- **Classification**: GNN classification for nodes and edges



Introduction

Objective

- **Graph based representation**: Exploit repetitive patterns
- **Classification**: GNN classification for nodes and edges
- **Table detection**: Group nodes into table regions

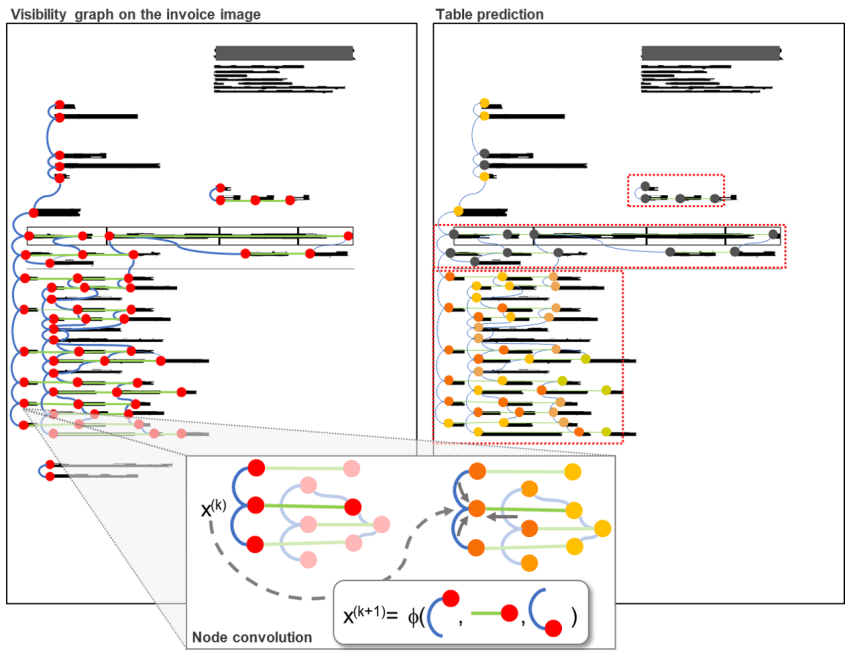
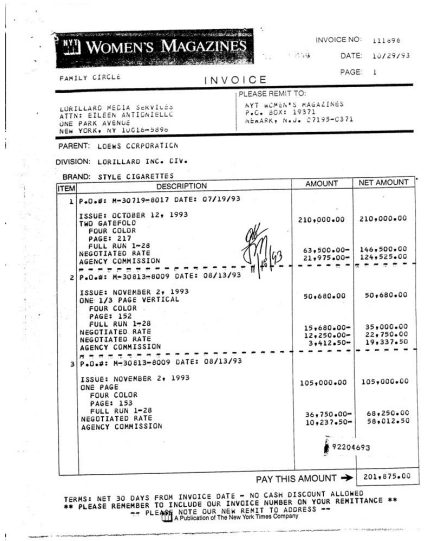
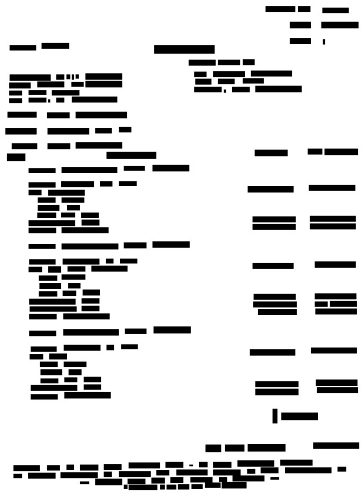


Table Detection Framework

Graph Representation

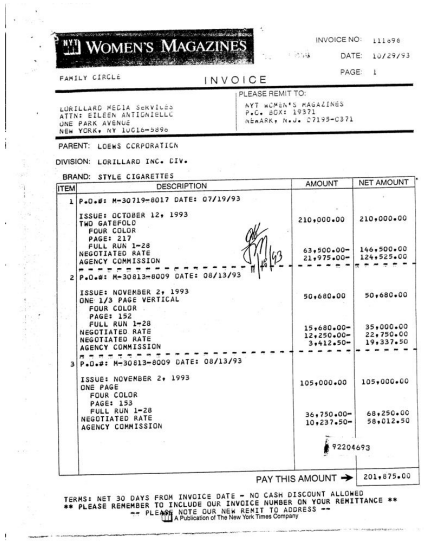
- Commercial OCR (by the industrial partner)
- Textual attributes (numeric, alphabet or symbol)
- Visibility graph:
  - Nodes: Document regions
  - Edges: Visibility relations (vertical and horizontal)





Graph Representation

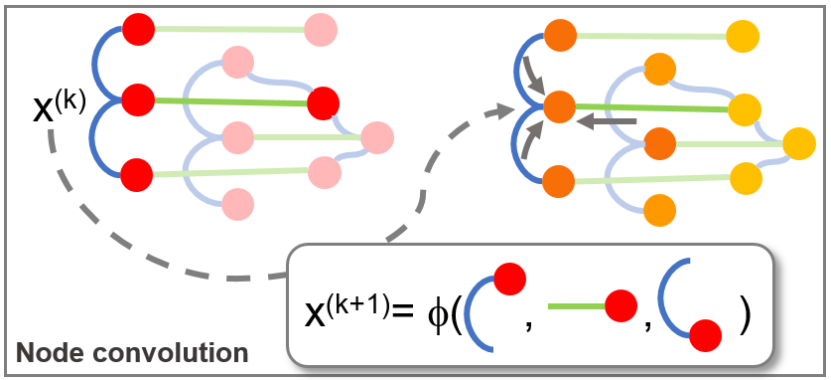
- Commercial OCR (by the industrial partner)
- Textual attributes (numeric, alphabet or symbol)
- Visibility graph:
  - Nodes: Document regions
  - Edges: Visibility relations (vertical and horizontal)



Graph Neural Network

GNN layer

- Notation introduced in [2]
- $\mathcal{A}$ - Graph intrinsic linear operators
- ρ - Activation function (ReLU)
- θ - Learnable parameters



$$x^{(k+1)} = G_C(x^{(k)}) = \rho\left(\sum_{B\in\mathcal{A}^{(k)}} Bx^{(k)}\theta_B^{(k)}\right)$$

[2] V. Garcia *et. al.*, *Few-shot learning with graph neural networks*, in ICLR, 2018.

Graph Neural Network

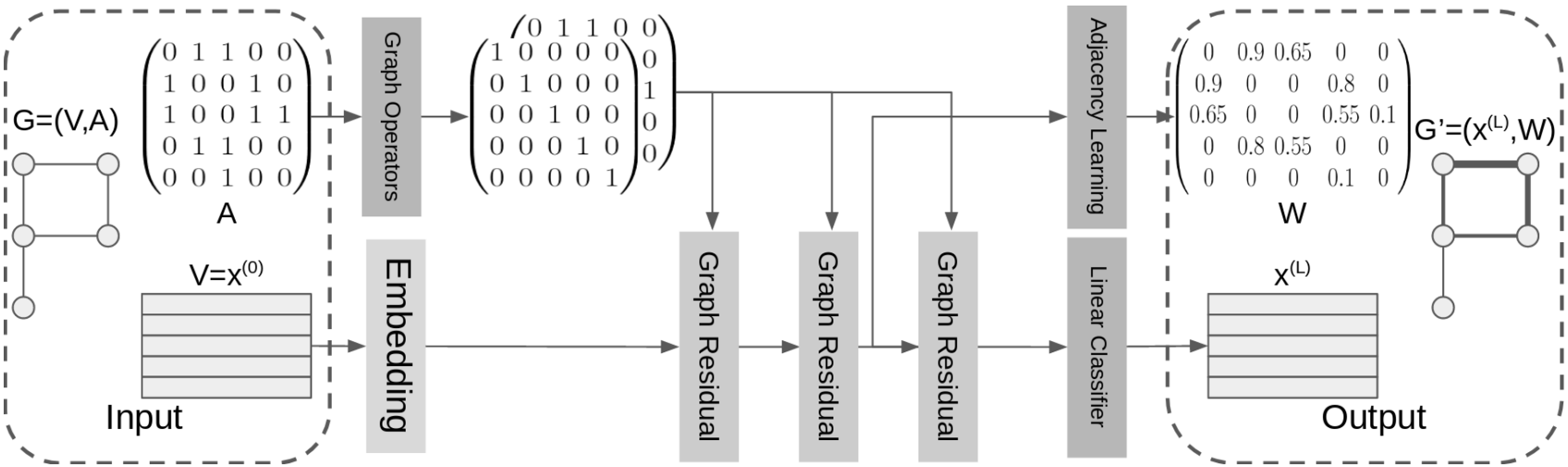
Graph Adjacency Layer

- Importance of the neighbourhood connection
- MLP - MultiLayer Perceptron
- σ - Activation function (Sigmoid)
- Absolute difference provides the symmetry property

$$\phi_k(B)_{i,j} = \begin{cases} 0 & \text{if } B_{i,j} = 0 \\ \sigma\left(\text{MLP}_{\tilde{\theta}}\left(\left|x_i^{(k)} - x_j^{(k)}\right|\right)\right) & \text{otherwise} \end{cases}$$

GNN Architecture

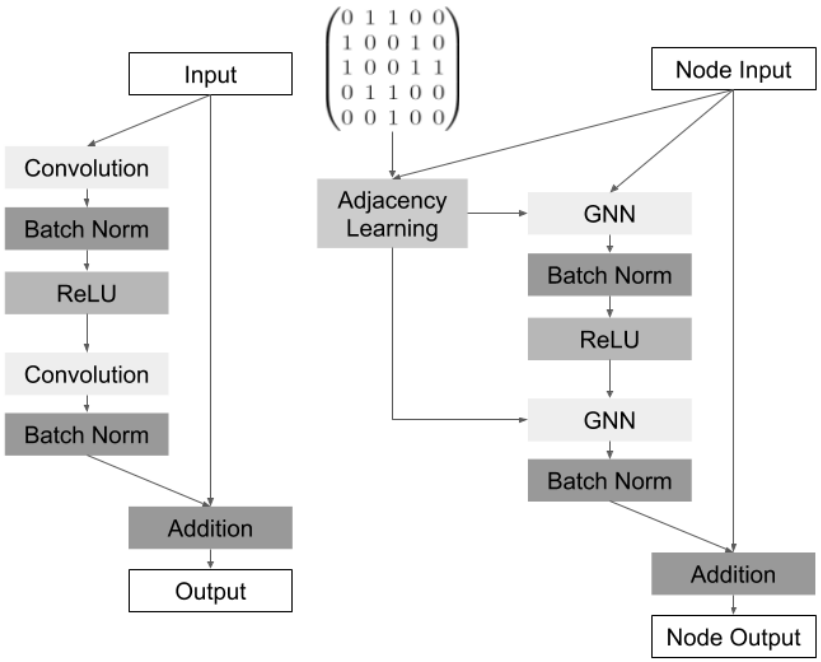
Pipeline



GNN Architecture

Graph Residual Block

- Idea of ResNet [1]
- GNN layers with a skip connection
- Edge weights are learned at the beginning of the block



[1] K. He *et. al.*, *Deep residual learning for image recognition*, in CVPR, 2016.

GNN Architecture

Objective functions

- **Node classifier:** Linear classifier with Softmax operation
- **Edge classifier:** Binary Cross entropy
  - 0 - Edge connects two different regions
  - 1 - Edge connects elements in the same region

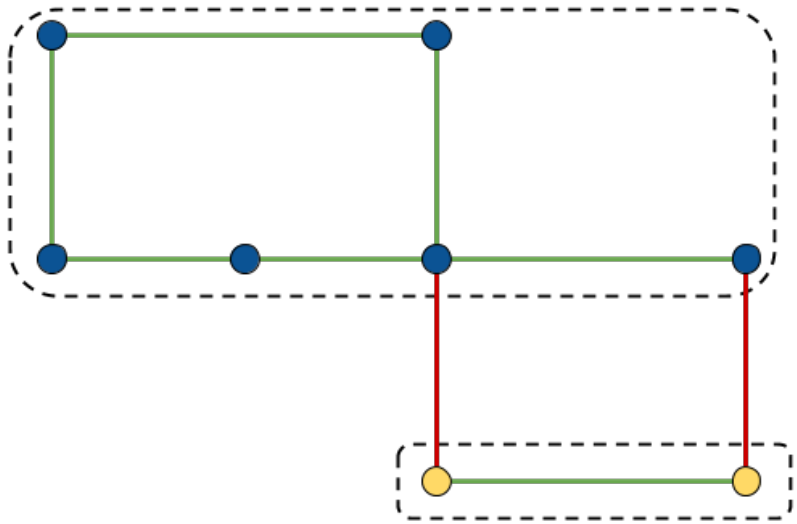
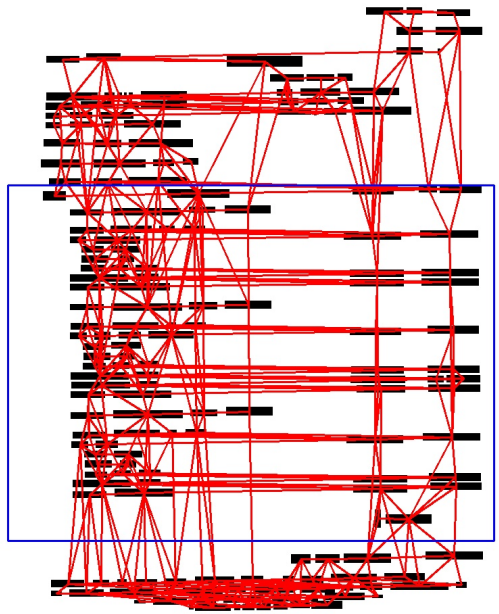


Table Detection

- Discard 0'ed edges
- Subgraphs with nodes classified as Table are considered
- The confidence score of these subgraphs are thresholded for the final decision



Experimental Validation

Datasets

CON-ANONYM

- 960 documents
- 8 region annotation
- Common car invoices
- Not publicly available

RVL-CDIP

- Overall 25,000 images
- 5 region annotation
- Selected 518 invoice class
- Publicly available [1]

[1] https://zenodo.org/record/3257319

[3] A. W. Harley *et. al.*, *Evaluation of deep convolutional nets for document image classification and retrieval,* in ICDAR, 2015.

Datasets

	CON-ANONYM	RVL-CDIP
Total # documents (tr, va, te)	950 (665, 95, 195)	518 (362, 52, 104)
Total # pages	1252	518
Total # tables	1202	485
Total # classes	8	6
Avg. # nodes/page	245.50	124.03
Avg. # edges/page	1354.81	619.55

Node/Edge Classification

Task	CON-ANONYM			RVL-CDIP		
	All	Table	Edge	All	Table	Edge
Pow 2	82.8	96.4	–	57.8	80.9	–
+ Edge	84.2	97.0	93.4	58.2	79.1	84.1
Pow 5	82.7	96.2	–	56.5	82.3	–
+ Edge	84.5	97.2	93.4	62.3	83.9	84.0

Table Detection

- Intersection over Union

Task	CON-ANONYM			RVL-CDIP		
	F1-Score	Precision	Recall	F1-Score	Precision	Recall
Pow 2	69.4	65.8	73.4	28.6	23.9	35.4
+ Edge	70.8	65.2	77.6	30.8	26.7	36.5
Pow 5	68.4	65.3	71.8	22.6	20.0	26.0
+ Edge	73.7	78.4	69.5	30.8	25.2	39.6

Table Detection

- Proper detection

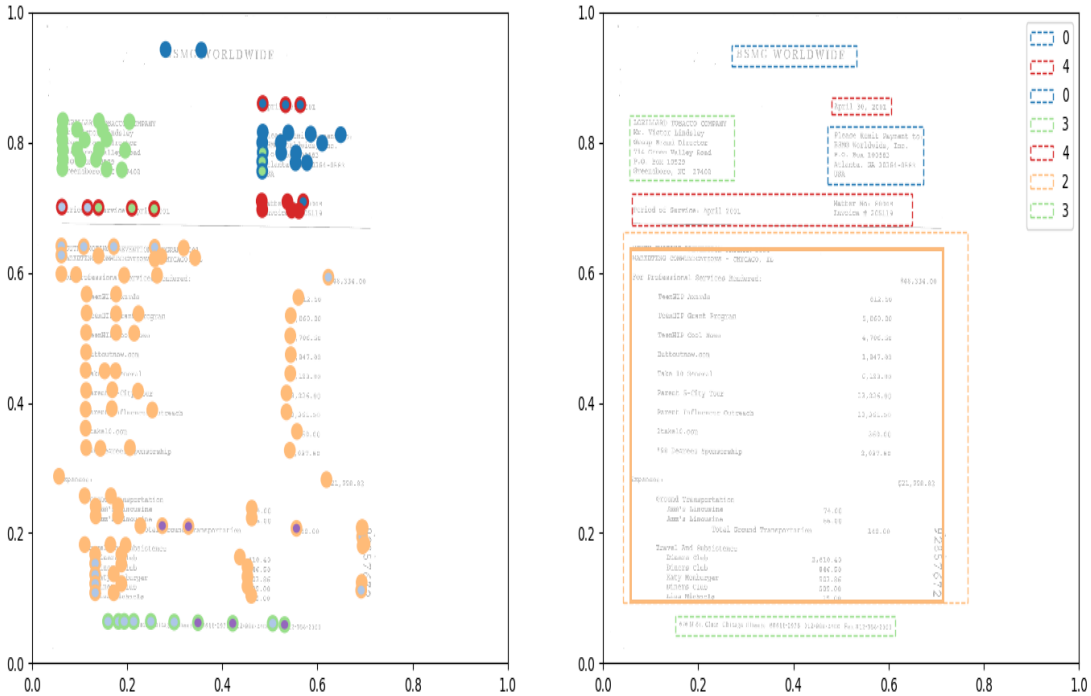


Table Detection

- Proper detection

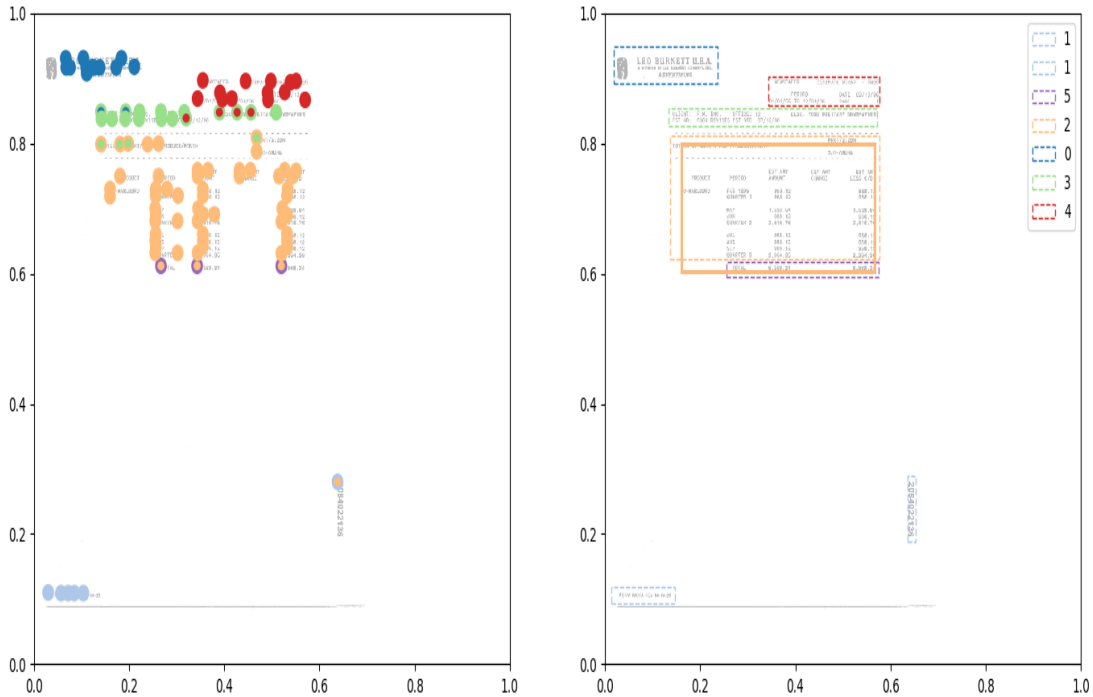


Table Detection

- Proper detection
- Preprocessing problems

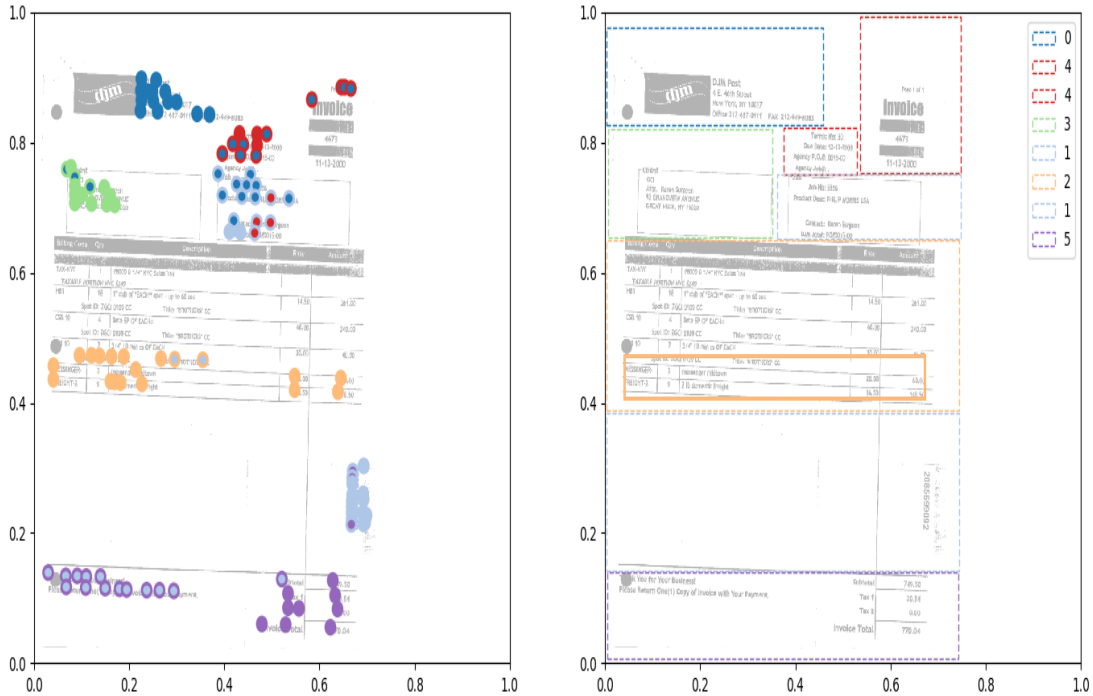
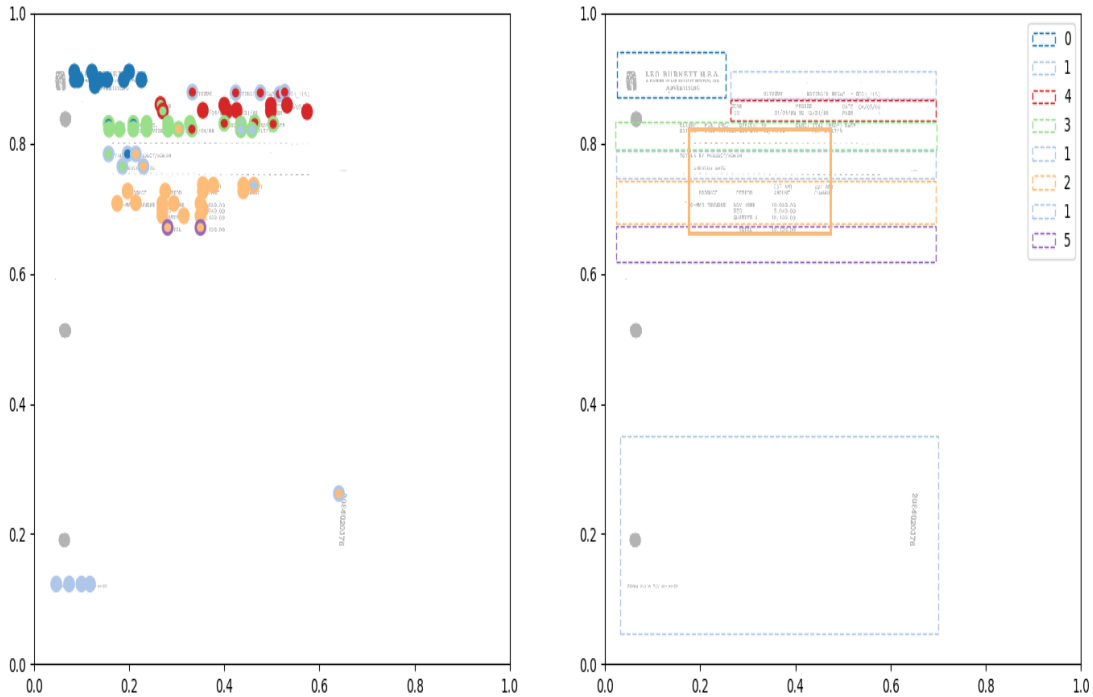


Table Detection

- Proper detection
- Preprocessing problems
- Tabular layout



Conclusions and Future Work

Conclusions and Future Work

- First *Table Detection* based on structural information.
- A Graph models the underlying structure of the document.
- Publicly available RVL-CDIP invoice dataset.
- Deal with anonymized data.

- Generalize to unconstrained tabular layout.

Thank you for your attention!

Pau Riba
Computer Vision Center
priba@cvc.uab.cat
`http://www.cvc.uab.cat/people/priba/`