

Key Information Extraction from Vietnamese Invoices by Combining Layout and Context

Capstone Project Report

Tran Manh Cuong
Ngo Tuan Anh

Supervisor MSc. Le Dinh Huynh

Bachelor of Computer Science
Hoa Lac Campus – FPT University
April 27, 2021



Content

1. Introduction
2. Related Work
3. Proposal Methodology
4. Implementation and Analysis

1. Introduction - Problem

- In business, an VAT invoice that lists information such as the seller, the seller's address, the tax code, the buyer's name, the address of the buyer or products purchased, when and how they were purchased. However, all information in the invoice cannot be extracted or imported into the data system by software. Works like this are all done by humans.

1. Introduction - Problem and motivation

- In business, an VAT invoice that lists information such as the seller, the seller's address, the tax code, the buyer's name, the address of the buyer or products purchased, when and how they were purchased. However, all information in the invoice cannot be extracted or imported into the data system by software. Works like this are all done by humans.
- Key information extraction(KIE) is extracting key information from textual sources to enable finding entities and classifying.

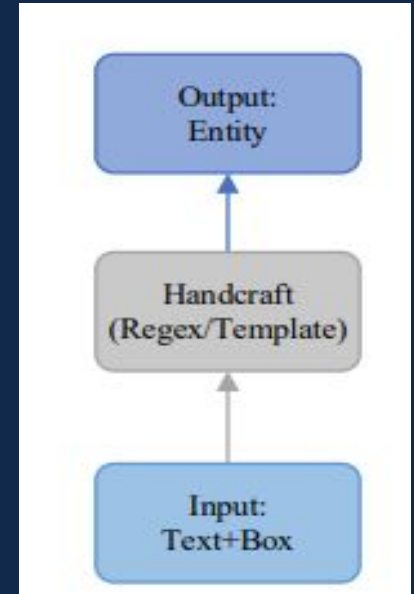
1. Introduction - Main objective

- *Learn literature review.*
- **Implementing the model**
- **Contribute**

1. Introduction - Main objective

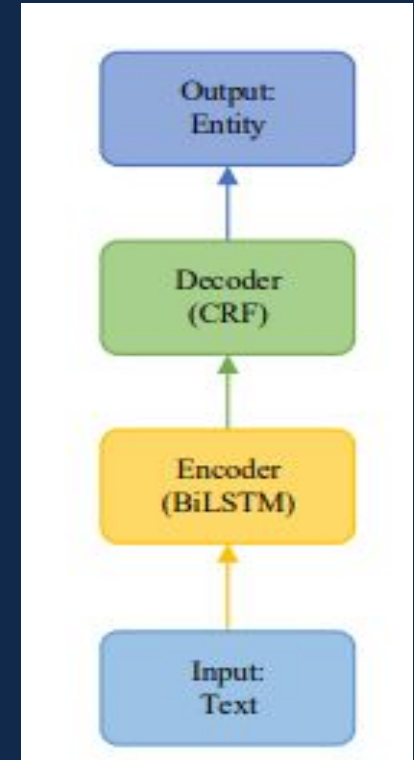
Intelligently-trained information extraction for document archiving

- This approaches use hand-craft features (e.g., regex and template matching) to extract key information.
- This solution only uses text and position information to extract entity and need a large amount of task-specific knowledge and human-designed rules, which does not extend to other types of documents



1. Introduction - Main objective

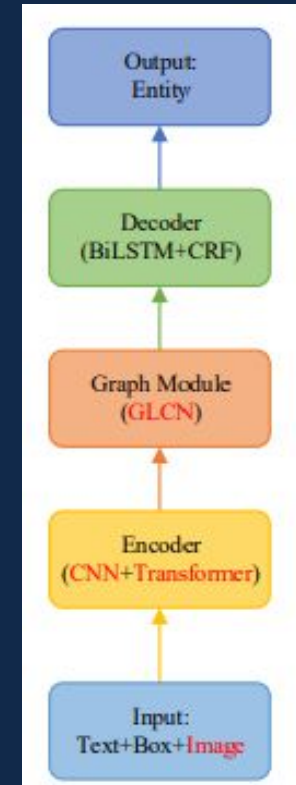
- Have many methods considered KIE as a sequence taggers problem
- It is much more challenging to distinguish entity without ambiguity from complicated documents for a machine.




1. Introduction - Main objective

Implementing the model

- *Method:* Processing Key Information Extraction from Documents using Improved Graph Learning-Convolutional Networks.
- *Data:* 2 weeks preparing the data.
- *Environment requirements:* python = 3.6 and framework: $\geq 1.5.1$



1. Introduction - Problem



Contribute to the community how to generate data from a template for training model

1. Introduction - Motivation

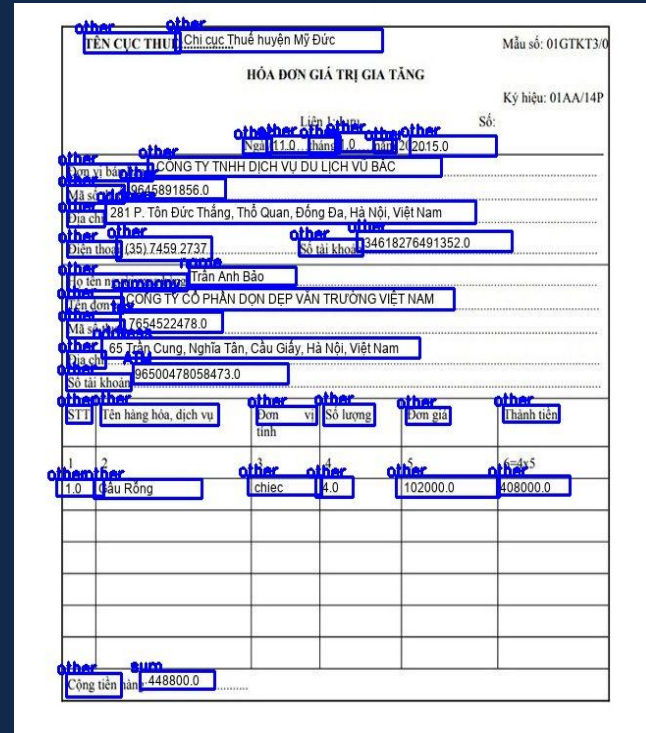
Base on a task of challenge ICDAR 2019 Robust Reading Challenge on Scanned Receipts OCR and Information Extraction

- Task 1 - Scanned Receipt Text Localisation
- Task 2 - Scanned Receipt OCR
- Task 3 - Key Information Extraction from Scanned Receipts

1. Introduction - Motivation

Task 1 - Scanned Receipt Text Localisation

- The aim of this task is to accurately localize texts with 4 vertices.



HÓA ĐƠN GIÁ TRỊ GIA TĂNG

Mẫu số: 01GTK/3.0
Kỳ hiệu: 01AA/14P

Liên hệ: Số: [other]
Ngày [11.0] tháng [0] năm [2015.0]

Đơn vị bán: [other] CÔNG TY TNHH DỊCH VỤ DU LỊCH VỤ BẠC

Mã số thuế: [other] 96745891856.0

Địa chỉ: [other] 281 P. Tôn Đức Thắng, Thủ Quan, Đống Đa, Hà Nội, Việt Nam

Điện thoại: [other] (35) 7459.2737 Số tài khoản: [other] 34618276491352.0

Họ tên người bán: [other] Trần Anh Bảo

Đơn vị: [other] CÔNG TY CỔ PHẦN ĐÓN ĐEP VÀN TRƯỜNG VIỆT NAM

Mã số thuế: [other] 7654522478.0

Địa chỉ: [other] 65 Trần Cung, Nghĩa Tân, Cầu Giấy, Hà Nội, Việt Nam

Số tài khoản: [other] 86500478058473.0

STT	Tên hàng hóa, dịch vụ	Đơn vị tính	Số lượng	Đơn giá	Thành tiền
1	[other] Gấu Rồng	[other] chiếc	[other] 4.0	[other] 102000.0	[other] 408000.0

[other] **sum**
Tổng tiền hàng: [other] 448800.0

1. Introduction - Motivation

Task 2 - Scanned Receipt OCR

- The aim of this task is to accurately recognize the text in a receipt image
- No localisation information is provided, or is required.

1. Introduction - Motivation

Task 3 - Key Information Extraction from Scanned Receipts

Đơn vị bán hàng: CÔNG TY TNHH TƯ VẤN DU HỌC VÀ ĐÀO TẠO HỌC BỔNG
 Mã số thuế: 8568984039
 Địa chỉ: BT5-5 Khu đoàn ngoại giao, Phố Đỗ Nhuận, Phường Xuân Tảo.
 Điện thoại: (81) 0067 6185 Email:
 Tài khoản số: 3269524998726

HÓA ĐƠN GIÁ TRỊ GIA TĂNG Mẫu số: 01GTKT3/001
 Liên 1: Lưu Kỳ hiệu: DT/13P
 Ngày 30...tháng 6.....năm 202007. Số: 41497

Họ tên người mua hàng: Vũ Thành Luân
 Tên đơn vị: CÔNG TY TNHH VS AGRO
 Mã số thuế: 3069473773
 Địa chỉ: Số 1, Ngõ 118 Đường Đồng Ốc, thôn 3 xã Lai Yên, Xã Lai Yên, Huyện Hoài Đức, Th
 Hình thức thanh toán: CK Số tài khoản: 13418914614371

STT	Tên hàng hóa, dịch vụ	Đơn vị tính	Số lượng	Đơn giá	Thành tiền
(1)	(2)	(3)	(4)	(5)	(6) = (4) x (5)
1	Giày Thể Thao Nữ Adidas EG3113	chiec	1	146000	146000

Cộng tiền hàng :
 Thuế suất GTGT : 10 % Tiền thuế GTGT : 14600
 Tổng cộng tiền thanh toán : 160600

Số tiền viết bằng chữ : nan

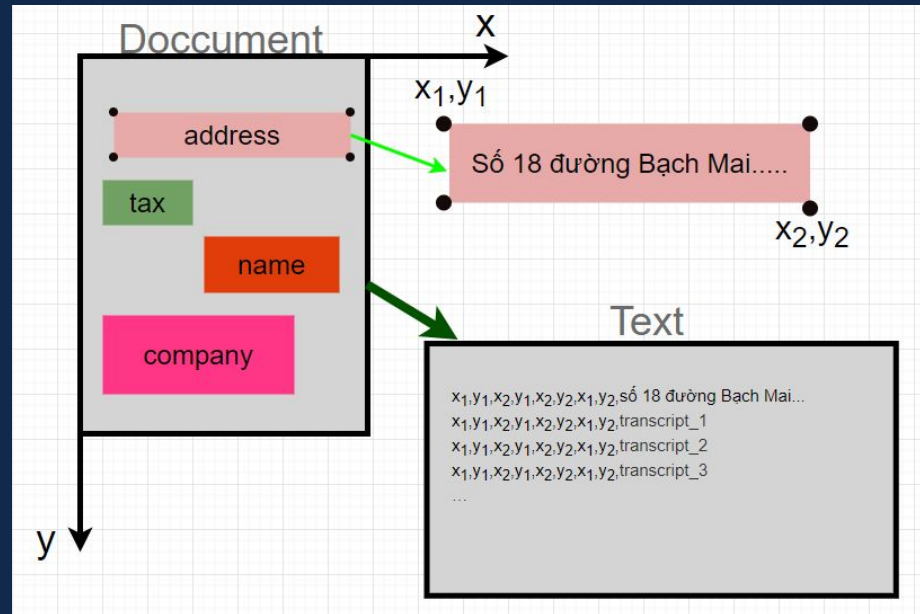
Người mua hàng (Ký, ghi rõ họ, tên)
 Người bán hàng (Ký, đóng dấu, ghi rõ họ, tên)

(Cần kiểm tra số chiều khi lập, giao, nhận hóa đơn)

Scanned invoice image

1. Introduction - Motivation

Task 3 - Key Information Extraction from Scanned Receipts



1. Introduction - Motivation

Task 3 - Key Information Extraction from Scanned Receipts

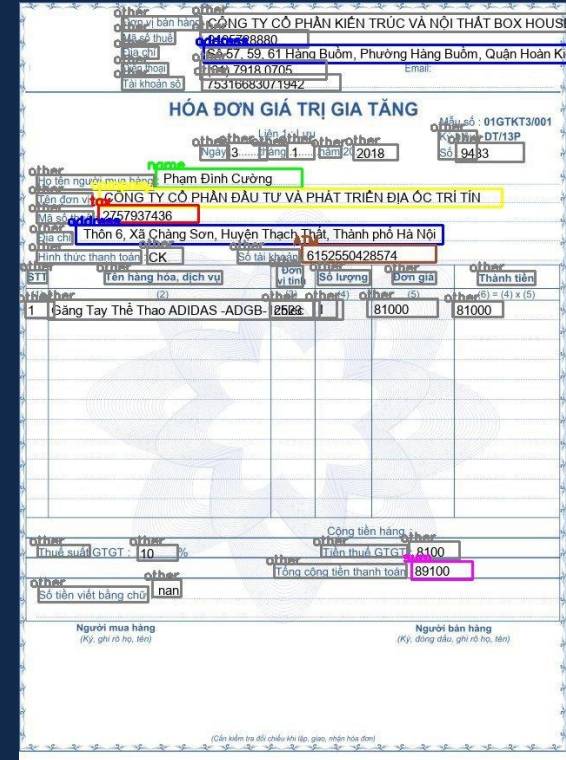


Image + Box + Key of Invoice

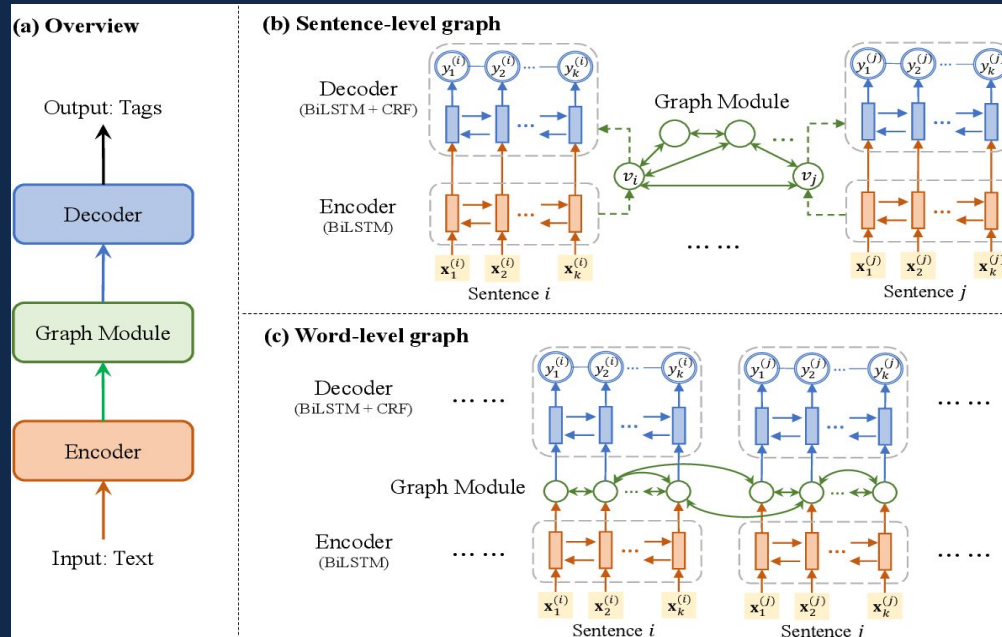
2. Related Work

- GraphIE: A Graph-Based Framework for Information Extraction
- Graph convolution for multimodal information extraction from visually rich documents

2. Related Work - GraphIE: A Graph-Based Framework for Information Extraction

- A framework that improves predictions by automatically learning the interactions between local and non-local dependencies in the input space.
- Integrates a graph module with the encoder-decoder architecture for sequence tagging
- The algorithm operates over a graph, where nodes correspond to textual units (i.e. words or sentences) and edges describe their relations.

2. Related Work - GraphIE: A Graph-Based Framework for Information Extraction

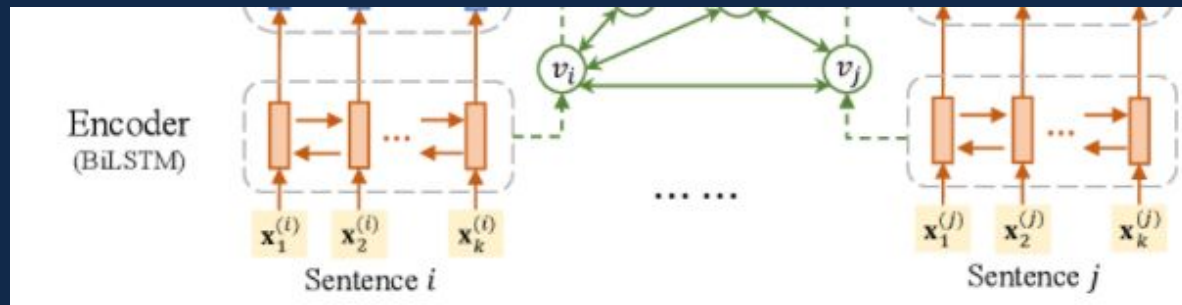


2. Related Work - GraphIE: A Graph-Based Framework for Information Extraction

Encoder

- Given a sentence $s_i = (w^{(i)}_1, w^{(i)}_2, \dots, w^{(i)}_k)$ of length k , each word $w^{(i)}_t$ is represented by a vector $x^{(i)}_t$, which is the concatenation of its word embedding and a feature vector learned with a character-level convolutional neural network

Encode the sentence with a recurrent neural network (RNN)



2. Related Work - GraphIE: A Graph-Based Framework for Information Extraction

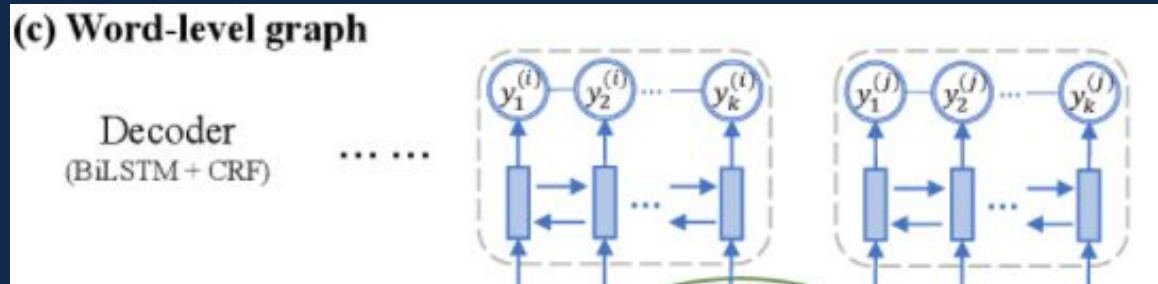
Graph Module

- The graph module is designed to learn the nonlocal and non-sequential information from the graph.
- Adapt the graph convolutional network (GCN) to model the graph context for information extraction.

2. Related Work - GraphIE: A Graph-Based Framework for Information Extraction

Decoder

- The decoder is instantiated as a BiLSTM+CRF tagger.
- The output representation of the graph module, $GCN(s_i)$, is split into two vectors of the same length, which are used as the initial hidden states for the forward and backward LSTMs, respectively.
- In this way, the graph contextual information is propagated to each word through the LSTM.



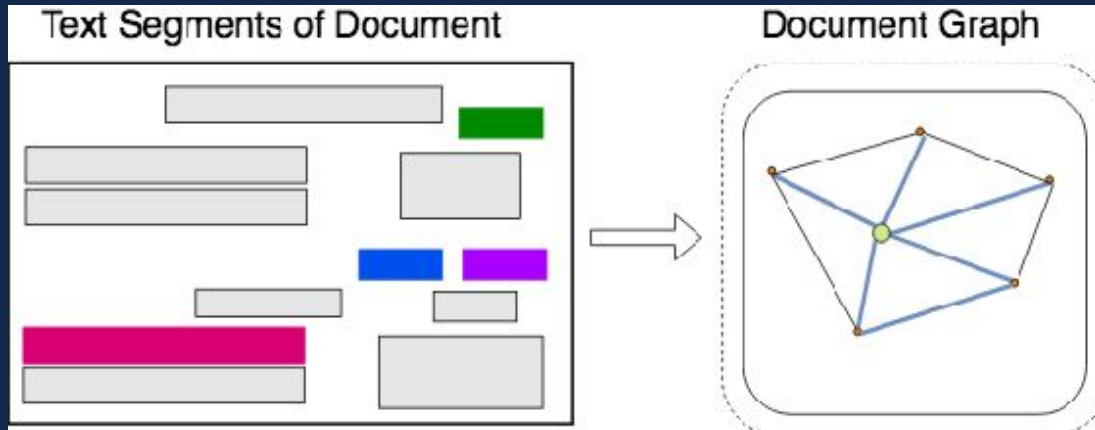
2. Related Work - Graph convolution for multimodal information extraction from visually rich documents

The algorithm introduce a graph convolution based model to combine textual and visual information presented in Visually rich documents (VRDs)

2. Related Work - Graph convolution for multimodal information extraction from visually rich documents

Document graph:

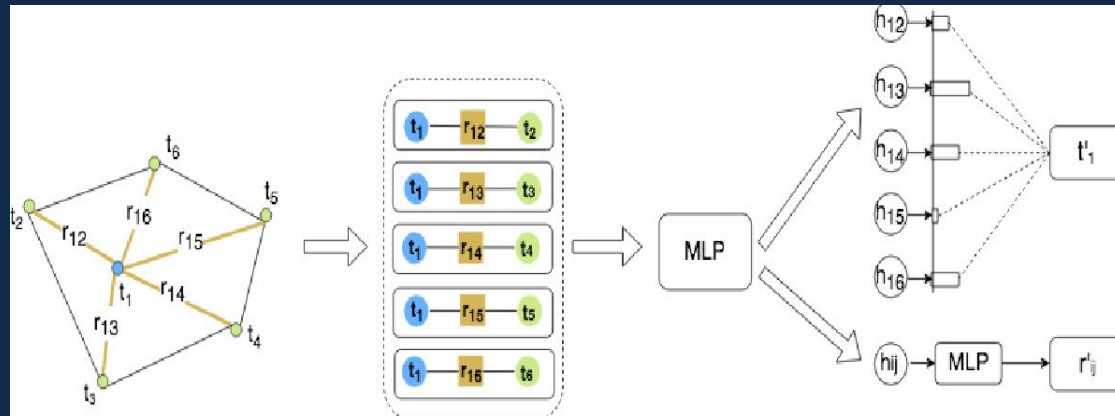
- Every node in the graph is fully connected to each other.



2. Related Work - Graph convolution for multimodal information extraction from visually rich documents

Graph convolution of document graph

- Convolution is defined on node-edge-node triplets (t_i, r_{ij}, t_j) .
- Each layer produces new embeddings for both nodes and edges.

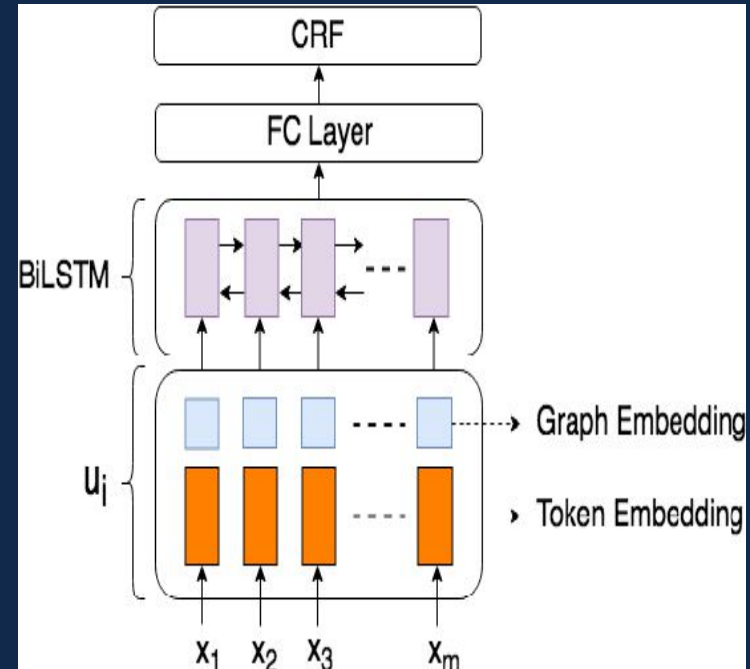


2. Related Work - Graph convolution for multimodal information extraction from visually rich documents

- Graph convolution is applied to compute visual text embeddings of text segments in the graph. Define convolution on the node-edge node triplets (t_i, r_{ij}, t_j) instead of on the node alone.
- For node t_i , we extract features h_{ij} for each neighbour t_j using a multi-layer perceptron (MLP) network.

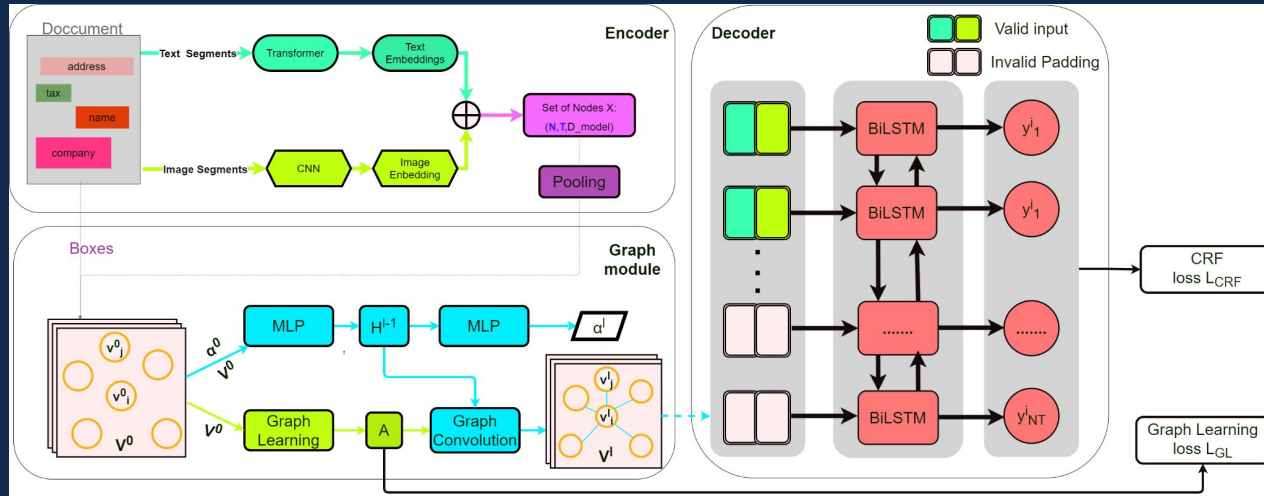
2. Related Work - Graph convolution for multimodal information extraction from visually rich documents

- Combine graph embeddings with token embeddings and feed them into standard BiLSTM-CRF for entity extraction.
- Intuitively, graph embedding adds contextual information to the input sequence.
- Then the input embeddings are fed into a BiLSTM network to be encoded, and the output is further passed to a fully connected network and then a CRF layer.



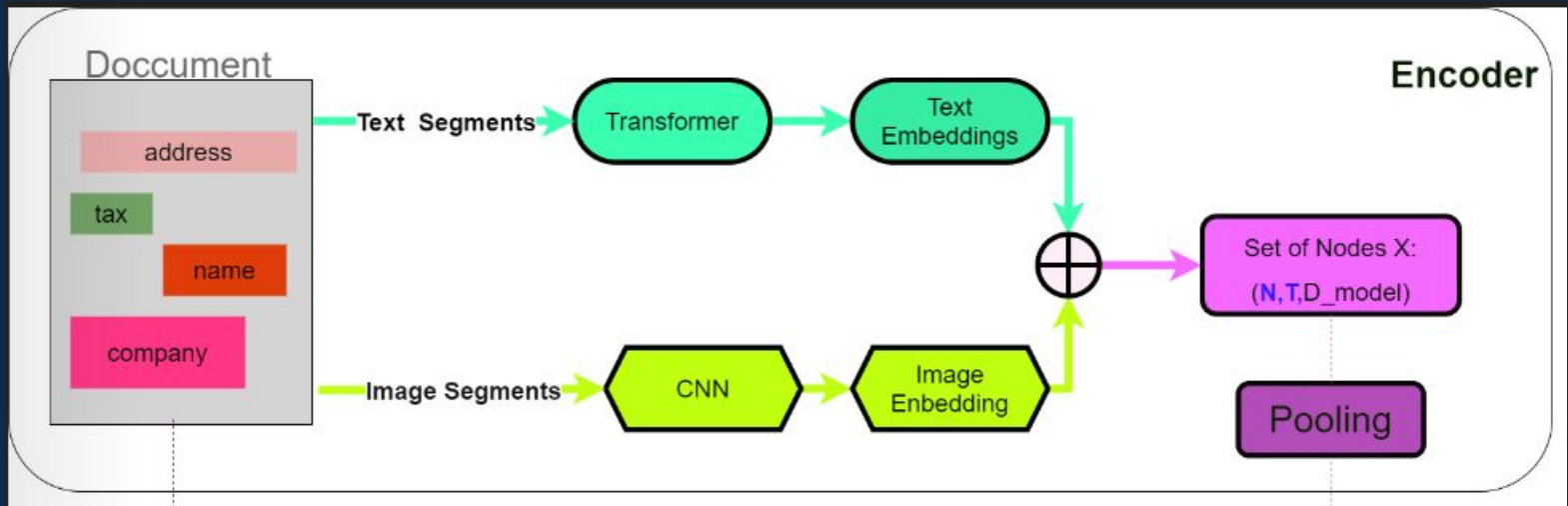
3. Proposal Methodology

- Encoder
- Graph Module
- Decoder



3. Proposal Methodology - Encoder

Extracting features from regular data

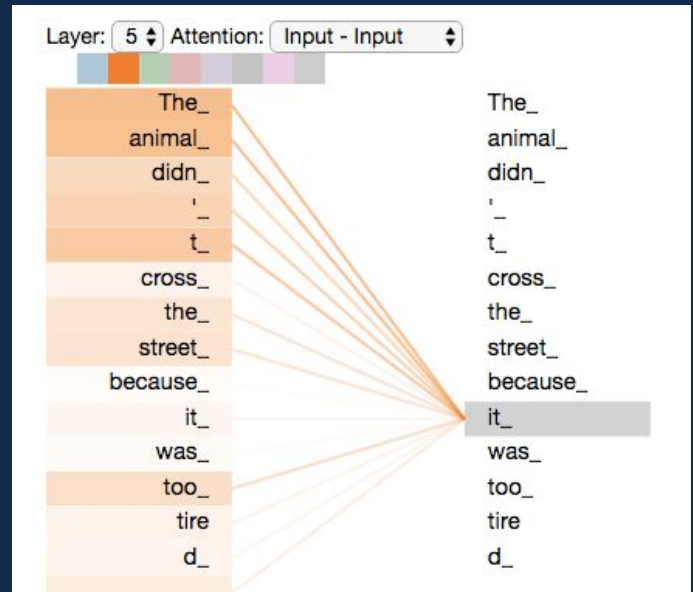


3. Proposal Methodology - Encoder

Transformer Encoder - Self-attention
extract features from text

”The animal didn't cross the street because it was too tired”

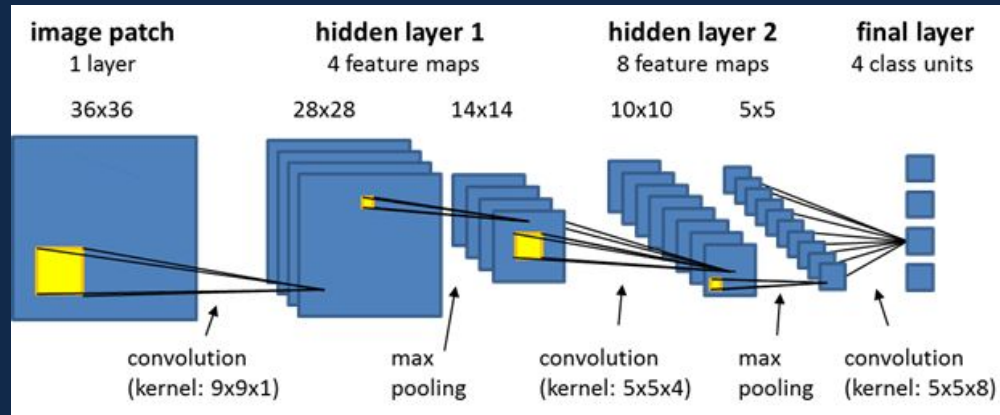
- What does “it” in this sentence refer to?
- Is it referring to the street or to the animal?



3. Proposal Methodology - Encoder

Convolutional Neural Networks

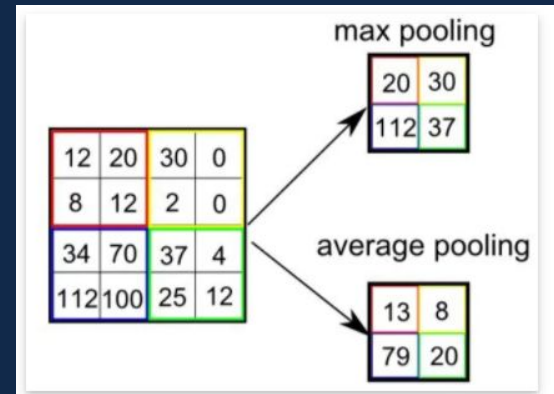
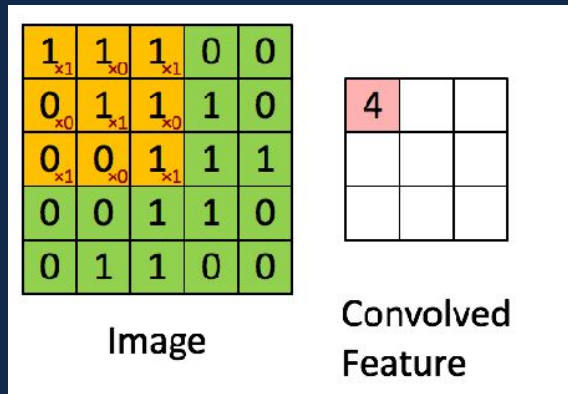
- Extract features from image



3. Proposal Methodology - Encoder

Convolutional Neural Networks

- Extract features from image



3. Proposal Methodology - Encoder

Convolutional Neural Networks

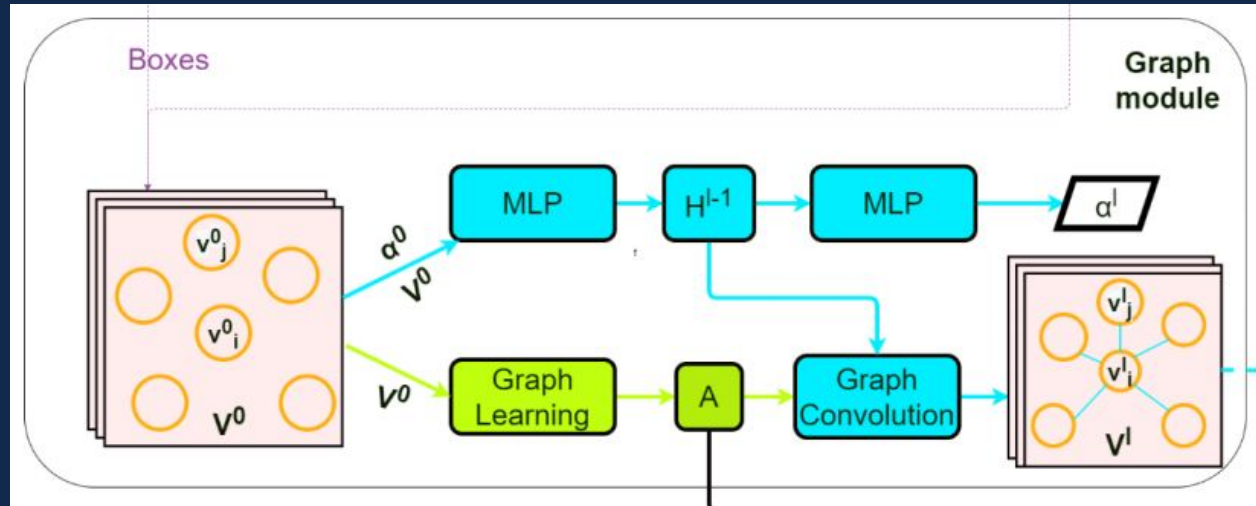
- Extract feature from Resnet-50

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
		3×3 max pool, stride 2				
conv2_x	56×56	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

model	top-1 err.	top-5 err.
VGG-16 [41]	28.07	9.33
GoogLeNet [44]	-	9.15
PReLU-net [13]	24.27	7.38
plain-34	28.54	10.02
ResNet-34 A	25.03	7.76
ResNet-34 B	24.52	7.46
ResNet-34 C	24.19	7.40
ResNet-50	22.85	6.71
ResNet-101	21.75	6.05
ResNet-152	21.43	5.71

3. Proposal Methodology - Graph Module

Learning graph structure by integrating both Graph Learning and Graph Convolution.



3. Proposal Methodology - Graph Module

Learning graph structure

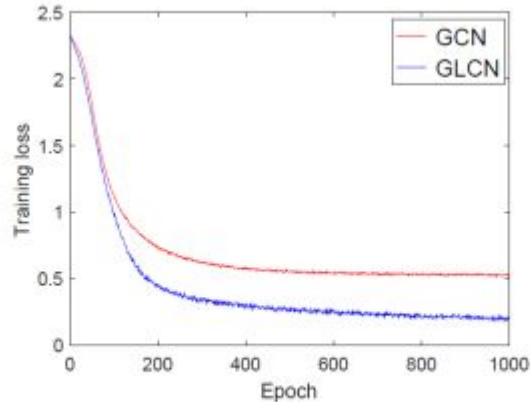


Figure 3: Demonstration of cross-entropy loss values across different epochs on MNIST dataset.

Graph Learning-Convolutional Networks

3. Proposal Methodology - Graph Module

Graph learning

- Soft adjacent matrix A
- Relationship between node v_i and v_j denote by e_{ij}

$$\begin{cases} A_{ij} = \text{softmax}(e_{ij}), i = 1, \dots, N, j = 1, \dots, N, \\ e_{ij} = \text{LeakRelu}(w_i^T |v_i - v_j|) \end{cases}$$

- Loss Function of Graph Learning

$$L_{GL} = \frac{1}{N^2} \sum_{i,j=1}^N \exp(A_{ij} + \eta \|v_i - v_j\|_2^2) + \gamma \|A\|_F^2$$

3. Proposal Methodology - Graph Module

Graph Convolutional

- Relation embedding α

$$\alpha_{ij}^0 = \mathbf{W}_\alpha^0 \left[x_{ij}, y_{ij}, \frac{w_i}{h_i}, \frac{h_j}{h_i}, \frac{w_j}{h_i}, \frac{T_j}{T_i} \right]^T$$

- Extract hidden features h between node v_i and v_j

$$h_{ij}^l = \sigma \left(W_{v_i h}^l v_i^l + W_{v_j h}^l v_j^l + \alpha_{ij}^l + b^l \right),$$

3. Proposal Methodology - Graph Module

Graph Convolutional

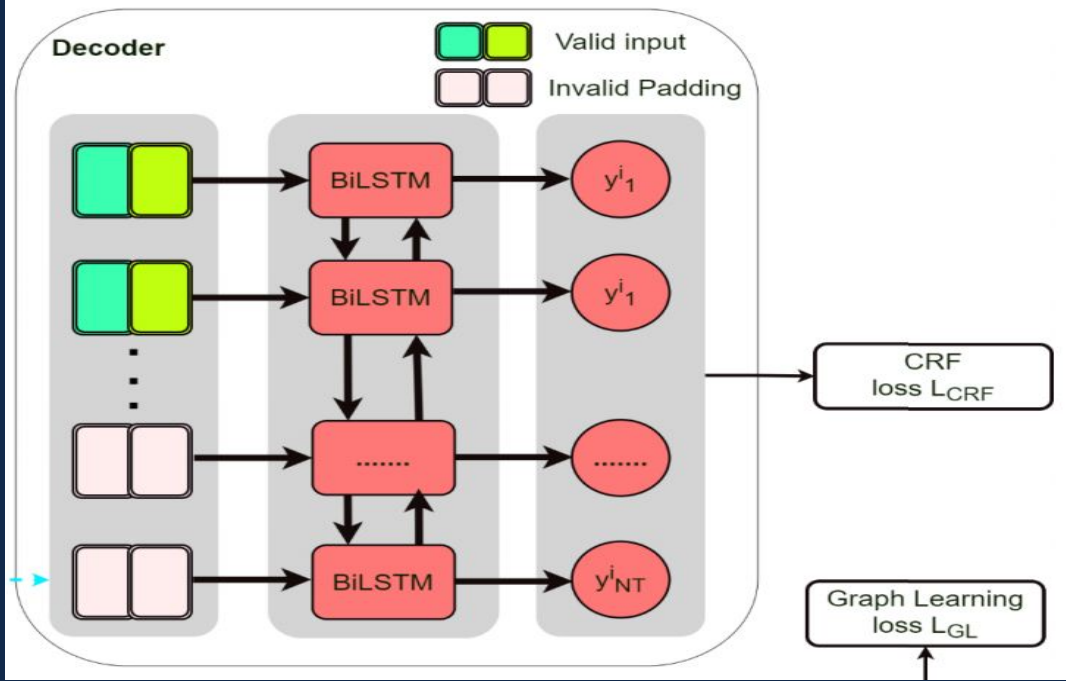
- Updating node embedding v for $l+1$ -th

$$v_i^{(l+1)} = \sigma(A_i h_i^l W^l)$$

- Updating relation embedding α on $l+1$ -th

$$\alpha_{ij}^{l+1} = \sigma(W_\alpha^l h_{ij}^l)$$

3. Proposal Methodology - Decoder



3. Proposal Methodology - Decoder

- Loss Function of Graph Learning

$$L_{GL} = \frac{1}{N^2} \sum_{i,j=1}^N \exp(A_{ij} + \eta \|v_i - v_j\|_2^2) + \gamma \|A\|_F^2$$

- Loss Function of CRF

$$\begin{cases} L_{crf} = -\log(p(y|X_{hat})) = -s(X_{hat}, y) + Z \\ Z = \log(\sum_{y_{hat} \in y(X_{hat})} e^{s(X_{hat}, y_{hat})}) = \text{logadd}_{y_{hat} \in y(X_{hat})} s(X_{hat}, y_{hat}) \end{cases}$$

- Loss Function of total

$$L_{total} = L_{crf} + \lambda L_{GL}$$



4. Implementation and Result-Dataset

Đơn vị bán hàng: []
Mã số thuế: []
Địa chỉ: []
Điện thoại: []
Tài khoản số: []

HÓA ĐƠN GIÁ TRỊ GIA TĂNG

Liên 1: Lưu Mẫu số: 01GTKT3/001
Ngày: 30 tháng 06 năm 2007 Ký hiệu: DT/13P
Số: []

Họ tên người mua hàng: []
Tên đơn vị: []
Mã số thuế: []
Địa chỉ: []
Hình thức thanh toán: [] Số tài khoản: []

STT	Tên hàng hóa, dịch vụ	Đơn vị tính	Số lượng	Đơn giá	Thành tiền
(1)	(2)	(3)	(4)	(5)	(6) = (4) x (5)
[]	[]	[]	[]	[]	[]

Cộng tiền hàng: []
Thuế suất GTGT: [] % Tiền thuế GTGT: []
Tổng cộng tiền thanh toán: []

Số tiền viết bằng chữ: []

Người mua hàng (Ký, ghi rõ họ, tên): []
Người bán hàng (Ký, đóng dấu, ghi rõ họ, tên): []

(Cần kiểm tra đối chiếu khi lập, giao, nhận hóa đơn)

Đơn vị bán hàng: CÔNG TY TNHH TƯ VẤN DU HỌC VÀ ĐÀO TẠO HỌC BỒN
Mã số thuế: 8568984039
Địa chỉ: BT5.5 Khu đoàn ngoại giao, Phố Đỗ Nhuận, Phường Xuân Tảo
Điện thoại: (81) 0067 6185 Email: []
Tài khoản số: 3269524998726

HÓA ĐƠN GIÁ TRỊ GIA TĂNG

Liên 1: Lưu Mẫu số: 01GTKT3/001
Ngày 30 tháng 6 năm 2007 Ký hiệu: DT/13P
Số: 41497

Họ tên người mua hàng: Vũ Thành Luân
Tên đơn vị: CÔNG TY TNHH VS AGRO
Mã số thuế: 3069473773
Địa chỉ: Số 1, Ngõ 118 Đường Đồng Óc, thôn 3 xã Lai Yên, Xã Lai Yên, Huyện Hoài Đức, Hà Nội
Hình thức thanh toán: CK Số tài khoản: 13418914614371

STT	Tên hàng hóa, dịch vụ	Đơn vị tính	Số lượng	Đơn giá	Thành tiền
(1)	(2)	(3)	(4)	(5)	(6) = (4) x (5)
1	Giày Thể Thao Nữ Adidas EG3113	chiec	1	146000	146000

Cộng tiền hàng: []
Thuế suất GTGT: 10 % Tiền thuế GTGT: 14600
Tổng cộng tiền thanh toán: 160600

Số tiền viết bằng chữ: nan

Người mua hàng (Ký, ghi rõ họ, tên): []
Người bán hàng (Ký, đóng dấu, ghi rõ họ, tên): []

(Cần kiểm tra đối chiếu khi lập, giao, nhận hóa đơn)

4. Implementation and Result-Result

TÊN CỤC THUẾ Chi cục Thuế quận Hoàng Mai		Mẫu số: 01GTKT3/0			
HÓA ĐƠN GIÁ TRỊ GIA TĂNG					
Đơn vị bán hàng: CÔNG TY TNHH XUẤT NHẬP KHẨU THIỆP NGỌC DŨNG		Liên hệ: Số: Ký hiệu: 01AA/14P			
Mã số thuế: 4566652854	Năm: 29/12/2008				
Địa chỉ: Số 10 Hẻm 143/45/39 Đường Xuân Phương, Phường Phương Canh, Quận Nam Từ Liêm, Thành phố Hà Nội					
Điện thoại: (67) 1615 0127	Số tài khoản: 85513546187839				
Ho tên người bán hàng: Chu Văn Dũng					
Tên đơn hàng: CÔNG TY TNHH SẢN XUẤT THƯƠNG MẠI VÀ DỊCH VỤ TỔNG HỢP HNT					
Mã số đơn hàng: 7438212600					
Địa chỉ: Số 3 ngõ 129/4 Trần Phú, Phường Văn Quán, Quận Hà Đông, Thành phố Hà Nội					
Số tài khoản: 89740288157505					
STT	Tên hàng hóa, dịch vụ	Đơn vị tính	Số lượng	Đơn giá	Thành tiền
1	Điện Thoại iPhone XS 256GB Hàng Nhập Khẩu			138000	276000
Công tiền bán: 303600					

address Số 10 Hẻm 143/45/39 Đường Xuân Phương\, Phường Phú
 name Chu Văn Dũng,name
 company CÔNG TY TNHH SẢN XUẤT THƯƠNG MẠI VÀ DỊCH VỤ TỔNG H
 tax 7438212600,tax
 address Số 3 ngõ 129/4 Trần Phú\, Phường Văn Quán\, Quận H
 ATM 89740288157505,ATM
 sum 303600,sum

4. Implementation and Result-Result

HÓA ĐƠN GIÁ TRỊ GIA TĂNG

Ngày 3 tháng 1 năm 2018

Địa chỉ: Thôn 6, Xã Chàng Sơn, Huyện Thạch Thất, Thành phố Hà Nội

Số tài khoản: 6152550428574

STT	Tên hàng hóa, dịch vụ	Đơn vị tính	Số lượng	Đơn giá	Thành tiền
1	Bảng Tay Thể Thao ADIDAS -ADGB-1822		1	81000	81000

Thuế suất GTGT: 10% Tổng tiền thuế GTGT: 8100

Tổng cộng tiền thanh toán: 89100

Số tiền viết bằng chữ: Nan

```

address số 57\, 59\, 61 Hàng Buồm\, Phường Hàng Buồm\, Quận
name Phạm Đình Cường,name
company CÔNG TY CỔ PHẦN ĐẦU TƯ VÀ PHÁT TRIỂN ĐỊA ỐC TRÍ TÍ
tax 2757937436,tax
address Thôn 6\, Xã Chàng Sơn\, Huyện Thạch Thất\, Thành p
ATM 6152550428574,ATM
sum 89100,sum
  
```



4. Implementation and Result-Result

- Mean entity prediction **mEP**, mean entity recall **mER**

$$mEP = \sum_{i=0}^{I_p-1} \mathbb{I}(y^i == g^i) / I_p$$

$$mER = \sum_{i=0}^{I_g-1} \mathbb{I}(y^i == g^i) / I_g$$

- Mean entity F-measure **mEF** is the harmonic average of mEP and mER

$$\frac{2}{mEF} = \frac{1}{mEP} + \frac{1}{mER}$$

4. Implementation and Result-Result

Result in invoices by mEP(mean entity prediction), mER(mean entity recall), mEF(Mean entity F-measure)

Entities	mEP	mER	mEF
ATM	0.917506	0.899287	0.908305
tax	0.915836	0.912856	0.914343
address	0.957516	0.912541	0.934652
name	0.934924	0.915429	0.924093
sum	0.906892	0.927898	0.9173727
company	0.901946	0.926472	0.914044
Overall	0.922436	0.915747	0.919079

Thanks for your attention

Any question please contact our via email

- cuongtmhe130625@fpt.edu.vn
- anhenthe130104@fpt.edu.vn