Unified Text Detection And Layout Analysis On Hierarchical Vietnamese Signboard Dataset

Trần Như Cẩm Nguyên [1] *and* **Trần Thị Cẩm Giang** [1]

[1] University of Information Technology, Vietnam National University Ho Chi Minh City

What ?

We unify two separate tasks, Scene text detection and Layout analysis, on Vietnamese signboards, specifically:

- Proposing a method for Unified Scene text detection and Layout analysis on Vietnamese signboards.
- Built the first hierarchical Vietnamese signboard dataset, annotated at three levels (word, line, paragraph), with multi-angle captures: each signboard photographed from three angles (left, frontal, right).
- Evaluating several methods on this dataset.

Why ?

- Scene text detection and document layout analysis have been treated as separate tasks. During our research, we observed the interrelation between these tasks and the unique of text layouts on Vietnamese signboards. Therefore, we decided to unify these two tasks and apply them specifically to Vietnamese signboards.
- Currently, there is limited research focused on layout analysis in Vietnamese scene text, and there is no high-quality dataset available to support this unified task.

Overview

Unified Text detection and Layout analysis on Vietnamese signboard

A Vietnamese signboard image

Figure 1. An image of a signboard.

Scene text detection

Each bounding box contains a detected word

Layout analysis

Words in the same paragraph will have the same bounding boxes color

Description

1. Hierarchical Vietnamese Signboard Dataset

- This dataset contains images of Vietnamese signboards annotated hierarchically (word, line, paragraph) for scene text detection and layout analysis purposes.
- Moreover, it includes GPS coordinates and each signboard is captured from three angles (left, frontal, right), aiming to create a high-quality dataset with diverse perspectives that could be beneficial for various tasks.

```
"image_id": "demo", # Image name
"image_width": 988, # Image width (pixels)
"image_height": 767, # Image height (pixels)
"latitude": 10.803527, # Latitude of the signboard
"longitude": 106.814262, # Longitude of the signboard
"paragraphs": [ # Detected paragraphs on the signboard
 {
  # First paragraph
  "label": "store_type", # Label of this paragraph is "Store Type"
  "lines": [ # Lines of text within this paragraph
   {
    # First line of text
    "words": [ # Words within this line of text
     {
      # First word
      "vertices": [[257,61], [386,70], [387,149], [253,150]], #Bounding box vertices for this word
      "text": "CỬA", # Content of this word
     },
     {
      # Second word
      "vertices": [[407,71], [571,74], [573,150], [404,151]], #Bounding box vertices for this word
      "text": "HÀNG", # Content of this word
     },
     {
      # Third word
      "vertices": [[586,94], [742,102], [738,166], [587,159]], #Bounding box vertices for this word
      "text": "VLXD", # Content of this word
     }
    ]
   }
  ]
 },
 {
  # Second paragraph
  ...
```

Figure 2. Annotation of the signboard in Figure 1.

2. Unified Text detection and Layout analysis on Vietnamese Signboard

- We propose a method that combines the **Transformer** architecture with the **Interactive Attention** module.
- Transformer architecture excels in capturing intricate spatial relationships within text components in images, facilitating comprehensive feature extraction.
- Interactive Attention module optimizes deep learning procedures by improving the model's capacity to understand intricate relationships between text parts.
- This integration aims to greatly increase the precision of layout analysis and text recognition on Vietnamese signboards.

Figure 5. Research content

Figure 3. Input: An image containing a signboard.

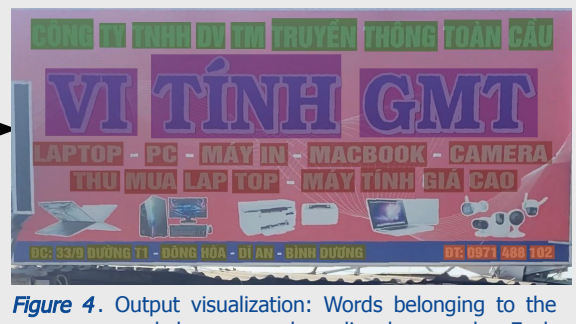

Figure 4. Output visualization: Words belonging to the same paragraph have same bounding boxes color. Each paragraph have a label to classify components on the signboard (store name, address, phone number, etc.). For example: Label of paragraph ["VI", "TÍNH", "GMT"] is "store_name".

NII

Trần Thị Cẩm Giang – University of Information Technology, Vietnam National University Ho Chi Minh City

TEL : 0395728107 Email : 22520361@gm.uit.edu.vn