Unified Line and Paragraph Detection by Graph Neural Networks

Shuang Liu, Renshen Wang, Michalis Raptis, Yasuhisa Fujii

s3liu@eng.ucsd.edu, {rewang, mraptis, yasuhisaf}@google.com

Google Research

Agenda

- Problem Statement

- Related Work

- Main Challenge

- Proposed Method

- Experiments

Problem Statement

Given Properly connected word-level bounding boxes for a document, output line-level and paragraph-level clusters

Left: word-level bounding boxes which do not necessarily correspond to words or characters
Right: word-level bounding boxes with proper connections (2-hop β-skeleton Graph)

Related Work — Layout Analysis

- Using heuristic algorithms to get line-level clusters (e.g. Text Flow [1])
- Clustering line-level bounding boxes into clusters [2]
- Image-based detection models [3]
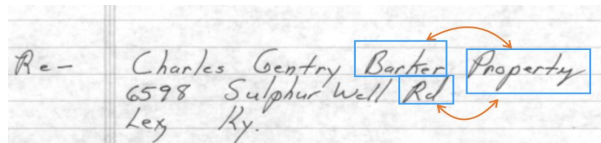
A difficult handwriting example for Text Flow

[1] Shangxuan Tian et al. Text Flow: A Unified Text Detection System in Natural Scene Images. In Proceedings of the IEEE/CVF International Conference on Computer Vision, 2015

[2] Renshen Wang et al. Post-OCR Paragraph Recognition by Graph Convolutional Networks. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022.

[3] Xu Zhong et al. Publaynet: Largest dataset ever for document layout analysis. International Conference on Document Analysis and Recognition, 2019

Related Work — Graph Convolutional Networks

Input

Input

Input

Input

Input

Input

Input

Input
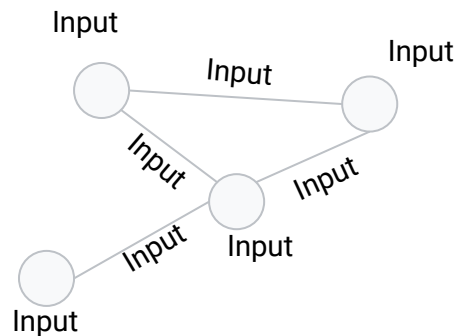
Iterative
Message
Passing

Output

Output

Output

Output

Output

Output

Output

Output

Google
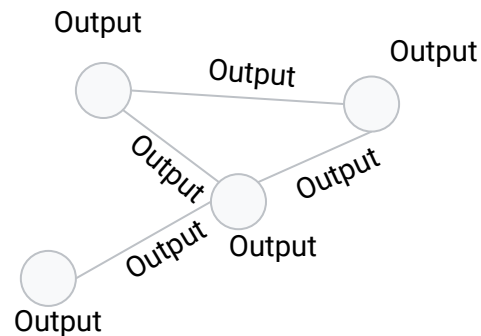
GCN prediction – box_left

- **box_left**(**b1, b2**): **1** if **b1** and **b2** are in the same line and **b1** is adjacent to **b2** to the left, **0** otherwise

GCN boxes and **box_left** labels

GCN prediction – box_above

- **box_above(b1, b2)**: conditioned on **b1** being the first box in a line, **1** if **b1** and **b2** are in the same paragraph and **b1** is in a line that is directly above the line **b2** is in.
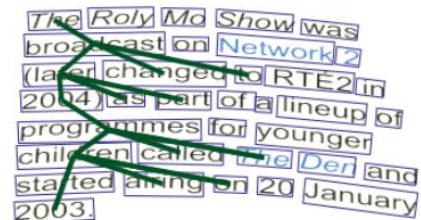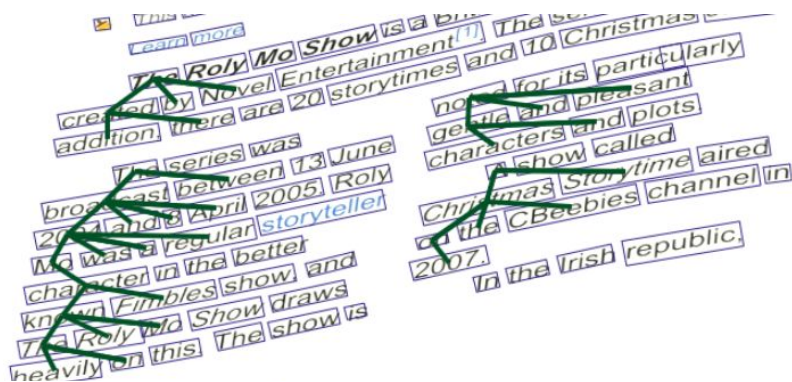
GCN boxes and **box_above** labels

GCN prediction – box_below
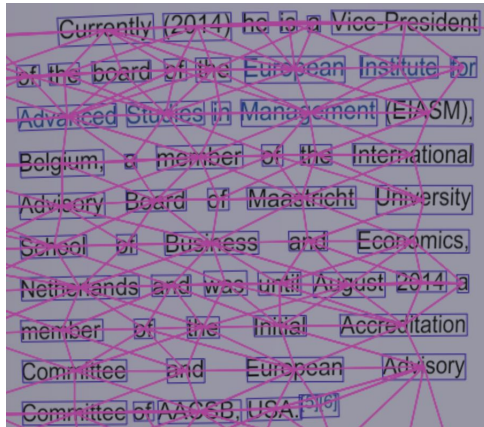
- **box_below(b1, b2)**: conditioned on **b1** being the first box in a line, **1** if **b1** and **b2** are in the same paragraph and **b1** is in a line that is directly below the line **b2** is in.

GCN boxes and **box_below** labels

Do we really need both box_above and box_below?

- **box_below** fails but **box_above** succeeds

The left image shows the graph edges. **box_below** could miss the connection between the first two lines (middle image), but this could be complemented by **box_above** (right image).

Do we really need both box_above and box_below?

- **box_above** fails but **box_below** succeeds
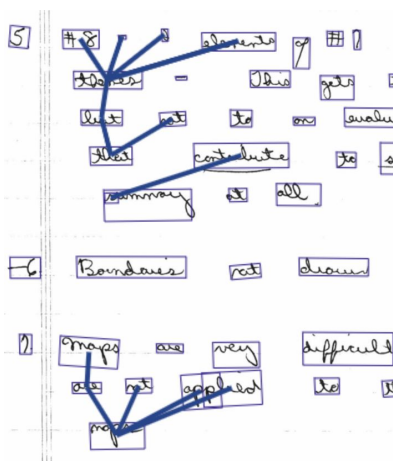
The left image shows the graph edges. **box_above** could miss the connection between the first two lines of paragraphs (right image), but this could be complemented by **box_below** (middle image).

GCN prediction visualization

- Blue: box_left
- Magenta: box_above/box_below

Google

Forming lines and paragraphs

Forming Lines

1. Initially all GCN boxes belong to a different line.
2. Iteratively, if **box_left**(**b1**, **b2**) = 1, then merge **b1** and **b2** into the same line.
3. A GCN box **b** is identified as the first box of a line if there is no **b'** such that **box_left**(**b'**, **b**) = 1.

Forming Paragraphs

1. Initially all lines belong to a different paragraph.
2. Iteratively, if **box_above**(**b1**, **b2**) = 1 or **box_below**(**b1**, **b2**) = 1, then merge **b1**'s line and **b2**'s line into the same paragraph.

Experiments —PubLayNet

Experiments —Successful Real-World Examples

Google

Experiments —Failed Real-World Examples

Thank You

Google Research