LGPMA: Complicated Table Structure Recognition with Local and Global Pyramid Mask Alignment

Liang Qiao*, Zaisheng Li*, Zhanzhan Cheng, Peng Zhang, Shiliang Pu, Yi Niu, Wenqi Ren, Wenming Tan, and Fei Wu

Hikvision Research Institute, Hangzhou, China

Zhejiang University, Hangzhou, China
1. Background
**Table Recognition**

| Dog  | Cat<sup>a</sup> |
|------|-----------------|
| Woof | Arf             | Meow           |

**Targeted HTML code (output)**

```html
<html>
<body>
<table>
<thead>
<tr>
    <td colspan="2">Dog</b></td>
    <td><b>Cat</b><sup>a</sup><sup>a</sup></td>
</tr>
</thead>
<tbody>
<tr>
    <td>Woof</td>
    <td>Arf</td>
    <td>Meow</td>
</tr>
</tbody>
</table>
</body>
</html>
```

**Table Structure Recognition**

- Global-object-based methods
- Local-object-based methods
Previous Method——Global-object-based

Problem of global-object-based methods:
- Lack of grid boundaries
- Cells spanning multiple rows/columns
Problem of local-object-based methods:
- Rules have limitations
- Empty cell ambiguity
2. Method
**Insight**

| Treatment phase | Adverse event | No. of patients |
|----------------|--------------|-----------------|
| T1             | Headache    | 1               |
|                | Burning     | 2               |
|                | Fever       | 1               |
|                | Chest Infection | 1   |
|                | Injury      | 1               |
| T2             | Headache    | 1               |
|                | Headache    | 1               |
|                | Headache    | 1               |

- **Lack of Annotations**: No visible boundaries in the table structure.
- **Position of Aligned Cell**: LGPMA algorithm is used for detection.
Workflow of LGPMA
**Aligned Bounding Box Detection**

- Ground Truth of aligned bounding boxes for non-empty cell

Method to approximate real cell regions
**Local Pyramid Mask Alignment**

- Binary mask of text region
- Pyramid mask in horizontal
- Pyramid mask in vertical

For pixel \((h, w)\), pyramid masks formed as

\[
\begin{align*}
t_h^{(w, h)} &= \begin{cases} 
\frac{w-x_1}{W-w} & w \leq x_{mid} \\
\frac{w-x_2}{W-w} & w > x_{mid}
\end{cases}, &
\begin{cases} 
\frac{h-y_1}{H-h} & h \leq y_{mid} \\
\frac{H-h}{H-y_2} & h > y_{mid}
\end{cases},
\end{align*}
\]

\(1\)

\(\text{Global Pyramid Mask Alignment}\)

- Binary mask of all aligned cells (including empty cell)
- Pyramid mask of all non-empty cells in horizontal
- Pyramid mask of all non-empty cells in vertical

PMFD: Pyramid Mask Text Detector, 2019
Aligned Bounding Box Refine

- Re-scoring strategy to compromise local and global pyramid mask

Final pyramid mask of \((x, y)\) can be re-scored as:

\[
F(x) = \begin{cases} 
\frac{x-x_1}{x_{mid}-x_1} F^{(L)}_{hor}(x, y) + \frac{x_{mid}-x}{x_{mid}-x_1} F^{(G)}_{hor}(x, y) & x_1 \leq x \leq x_{mid} \\
\frac{x-x_2}{x_{mid}-x_2} F^{(L)}_{hor}(x, y) + \frac{x_{mid}-x}{x_{mid}-x_2} F^{(G)}_{hor}(x, y) & x_{mid} < x \leq x_2 
\end{cases}
\]

\[
F(y) = \begin{cases} 
\frac{y-y_1}{y_{mid}-y_1} F^{(L)}_{ver}(x, y) + \frac{y_{mid}-y}{y_{mid}-y_1} F^{(G)}_{ver}(x, y) & y_1 \leq y \leq y_{mid} \\
\frac{y-y_2}{y_{mid}-y_2} F^{(L)}_{ver}(x, y) + \frac{y_{mid}-y}{y_{mid}-y_2} F^{(G)}_{ver}(x, y) & y_{mid} < y \leq y_2 
\end{cases}
\]

- Final horizontal and vertical pyramid mask can fit two planes respectively.

The four planes' intersection lines with the zero plane are the refined boundaries.
Table Structure Recovery Pipeline

Refined according to LPMA and GPMA results
3. Experiment
Implementation Details:

- Backbone: ResNet-50 + FPN
- 4 x feature map
- Pre-trained model of ImageNet
- Anchor ratios: 1/20, 1/10, 1/5, 1/2, 1, 2
- Pytorch, 8 32GB-Tesla-V100 GPUs
- Data augmentations: multi-scale training
- Single scale testing
**Visualization Results:**

| Threshold for EF (L/min/m²) | Lower limit | Middle limit | Upper limit | Mean value |
|-----------------------------|-------------|--------------|-------------|------------|
| 1.1-1.4 | 20 | 30 | 40 | 30 |
| 1.4-1.6 | 30 | 40 | 50 | 40 |
| 1.6-1.8 | 40 | 50 | 60 | 50 |

| Semantic labels | Cityscapes image | PSNR (dB) | SSIM | EQE | VIF |
|-----------------|------------------|-----------|------|-----|-----|
| ps2pcGAN | 15.74 | 0.707 | 0.272 | 0.206 |
| PAN | 16.06 | 0.828 | 0.111 | 0.0658 |

| Edges | Networks |
|-------|----------|
| IDGAN | 20.07 | 0.756 | 0.272 | 0.206 |
| PAN | 19.31 | 0.781 | 0.342 | 0.256 |

| Cityscapes images | Semantic labels | PSNR (dB) | SSIM | EQE | VIF |
|-------------------|----------------|-----------|------|-----|-----|
| IDGAN | 16.50 | 0.657 | 0.357 | 0.172 |
| PAN | 15.90 | 0.657 | 0.357 | 0.172 |

| Annal pattern | Maps |
|---------------|------|
| IDGAN | 19.85 | 0.757 | 0.357 | 0.172 |
| PAN | 20.62 | 0.772 | 0.172 | 0.1638 |

| Annotations | Maps |
|-------------|------|
| IDGAN | 26.10 | 0.645 | 0.012 | 0.0213 |
| PAN | 28.34 | 0.762 | 0.337 | 0.1017 |
Visualization Results:
**Performance Evaluation:**

| Methods                  | Training Dataset | ICDAR 2013 | SciTSR | SciTSR-COMP |
|--------------------------|------------------|------------|--------|-------------|
|                          |                  | P         | R      | F1          | P         | R      | F1          | P         | R      | F1          |
| DeepDeSRT [30]           | -                | 0.959     | 0.874 | 0.914       | 0.906     | 0.887 | 0.890       | 0.863     | 0.831 | 0.846       |
| Split [33]               | Private          | 0.869     | 0.866 | 0.868       | -         | -     | -           | -         | -     | -           |
| DeepTabStR [31]          | ICDAR 2013       | 0.931     | 0.930 | 0.930       | -         | -     | -           | -         | -     | -           |
| Siddiqui et al. [32]     | Synthetic 500k   | 0.934     | 0.934 | 0.934       | -         | -     | -           | -         | -     | -           |
| ReS2TIM [36]             | ICDAR 2013†      | 0.734     | 0.747 | 0.740       | -         | -     | -           | -         | -     | -           |
| GTE [38]                 | ICDAR 2013†      | 0.944     | 0.927 | 0.935       | -         | -     | -           | -         | -     | -           |
| GraphTSR [2]             | SciTSR           | 0.885     | 0.860 | 0.872       | 0.950     | 0.948 | 0.953       | 0.964     | 0.945 | 0.955       |
| TabStruct-Net [26]       | SciTSR           | 0.915     | 0.897 | 0.906       | 0.927     | 0.913 | 0.920       | 0.909     | 0.882 | 0.895       |
| LGPMA                    | SciTSR           | 0.930     | 0.977 | 0.953       | 0.982     | 0.993 | 0.988       | 0.973     | 0.987 | 0.980       |
| LGPMA                    | ICDAR 2013†      | 0.967     | 0.991 | 0.979       | -         | -     | -           | -         | -     | -           |

Table 1: Results on ICDAR 2013, SciTSR, SciTSR-COMP datasets

| Methods       | Training Dataset | Testing Dataset | TEDS (All) | TEDS-Struct. (All) |
|---------------|------------------|-----------------|------------|--------------------|
| EDD [39]      | PTN-train        | PTN-val         | 88.3       | -                  |
| TabStruct-Net [26] | SciTSR   | PTN-val         | 90.1       | -                  |
| GTE [38]      | PTN-train        | PTN-val         | 93.0       | -                  |
| LGPMA (ours)  | PTN-train        | PTN-val         | 94.6       | 96.7               |

Table 2: Results on PubTabNet
**Ablation Summary:**

Table 3: Ablation experiments on modules effect the aligned bounding box detection

| Models       | Modules | Det of text regions | Det of non-empty aligned bounding boxes | TEDS-Struc. |
|--------------|---------|---------------------|-----------------------------------------|-------------|
|              |         | Precision Recall Hmean | Precision Recall Hmean |              |
|              | LPMA   | GPMA   | AL [26] | -     | -    | -    | - | 81.32 | 81.31 | 81.31 | 94.63 |
| Faster R-CNN | ✔       | ✔      | ✔      | ✔     | ✔    | ✔    | ✔ | 91.71 | 91.53 | 91.62 | 81.83 | 81.82 | 81.83 | 94.65 |
| Mask R-CNN   | ✔       | ✔      | ✔      | ✔     | ✔    | ✔    | ✔ | 91.92 | 91.66 | 91.79 | 84.29 | 84.10 | 84.20 | 95.22 |
|              | ✔       | ✔      | ✔      | ✔     | ✔    | ✔    | ✔ | 91.98 | 91.50 | 91.74 | 83.48 | 83.18 | 83.33 | 95.04 |
| Mask R-CNN   | ✔       | ✔      | ✔      | ✔     | ✔    | ✔    | ✔ | 92.27 | 91.86 | 92.06 | 85.14 | 84.77 | 84.95 | 95.53 |
|              | ✔       | ✔      | ✔      | ✔     | ✔    | ✔    | ✔ | 92.11 | 91.85 | 91.98 | 81.91 | 81.79 | 81.85 | 94.94 |
|              | ✔       | ✔      | ✔      | ✔     | ✔    | ✔    | ✔ | 92.05 | 91.65 | 91.85 | 84.87 | 84.50 | 84.68 | 95.31 |
Contact Email:
qiaoliang6@hikvision.com
lizaisheng@hikvision.com

More works from Davar-Lab:
https://davar-lab.github.io/
Thank you