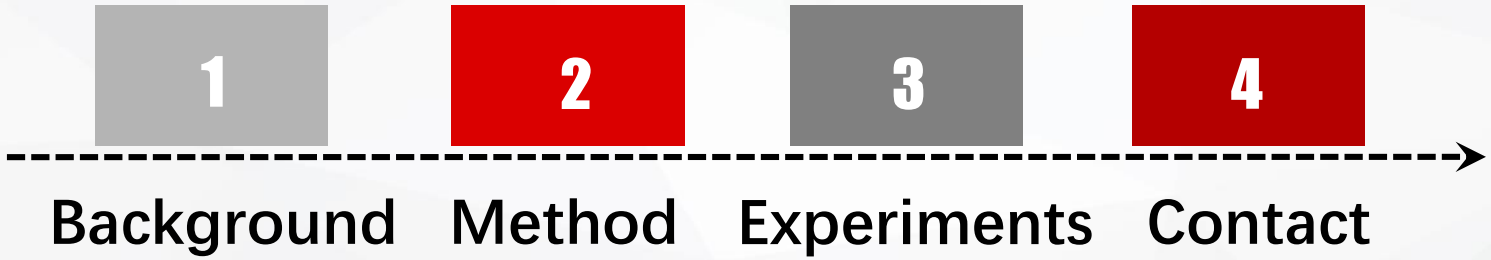
LGPMA: Complicated Table Structure Recognition with Local and Global Pyramid Mask Alignment

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

CONTENT



1. Background

Table Recognition

Table image (input)

Dog		Cat[a]
Woof	Arf	Meow

Targeted HTML code (output)

```
<html>
  <body>
    <table>
      <thead>
        <tr>
          <td colspan="2"><b>Dog</b></td>
          <td><b>Cat</b><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

Powders	Degradation Efficiency (%)					
	DB			MO		
	303 K	318 K	333 K	303 K	318 K	333 K
FE powders	18.9	35.5	64.1	14.7	40.9	30.5
AM powders	38.8	75.9	91.2	76.0	81.9	84.8
NP powders	97.6	98.1	99.0	96.0	97.8	98.2

Column Detection

Powders		Degradation Efficiency (%)					
		DB			MO		
	303 K	318 K	333 K	303 K	318 K	333 K
FE powders	18.9	35.5	64.1	14.7	40.9	30.5
AM powders	38.8	75.9	91.2	76.0	81.9	84.8
NP powders	97.6	98.1	99.0	96.0	97.8	98.2

Row Detection

Powders	Degradation Efficiency (%)					
	DB			MO		
	303 K	318 K	333 K	303 K	318 K	333 K
FE powders	18.9	35.5	64.1	14.7	40.9	30.5
AM powders	38.8	75.9	91.2	76.0	81.9	84.8
NP powders	97.6	98.1	99.0	96.0	97.8	98.2

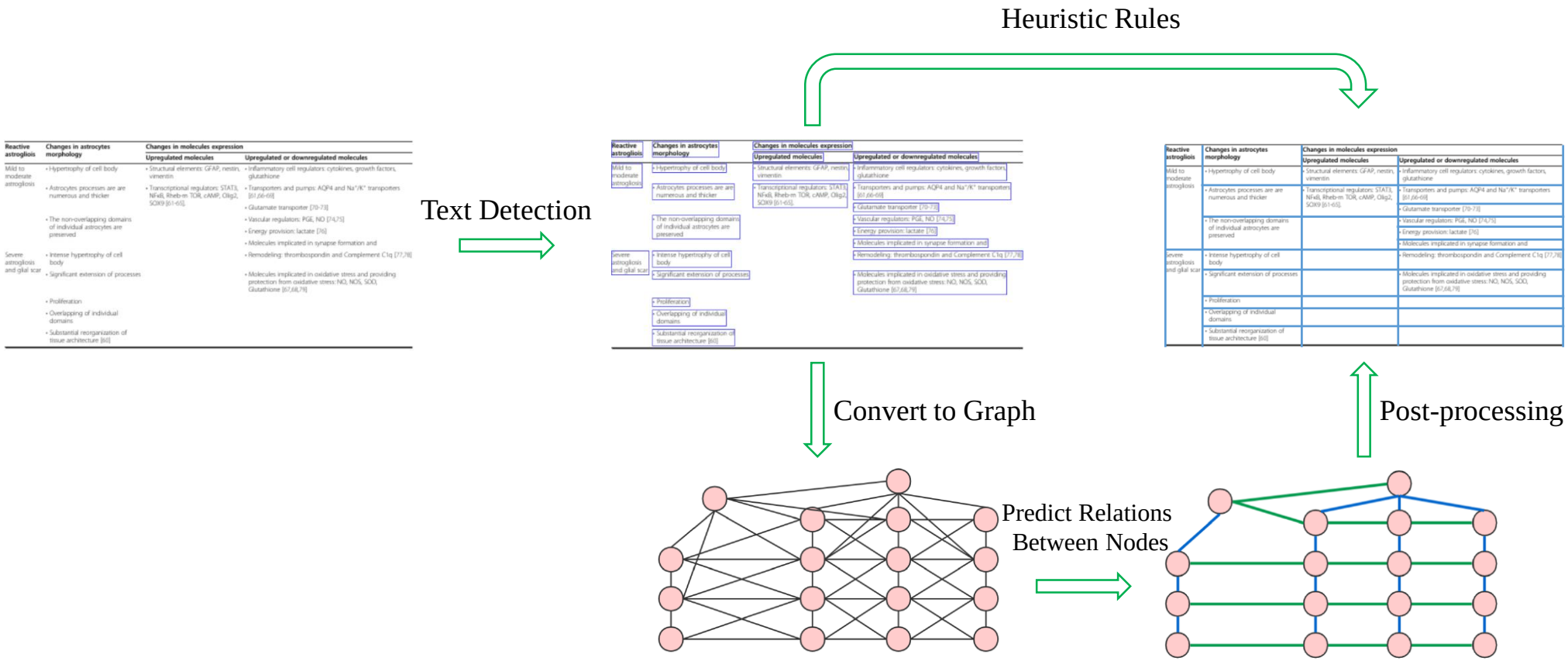
Heuristic Rules

Powders	Degradation Efficiency (%)					
	DB			MO		
	303 K	318 K	333 K	303 K	318 K	333 K
FE powders	18.9	35.5	64.1	14.7	40.9	30.5
AM powders	38.8	75.9	91.2	76.0	81.9	84.8
NP powders	97.6	98.1	99.0	96.0	97.8	98.2

Problem of global-object-based methods:

- Lack of grid boundaries
- Cells spanning multiple rows/columns

✵ Previous Method——Local-object-based

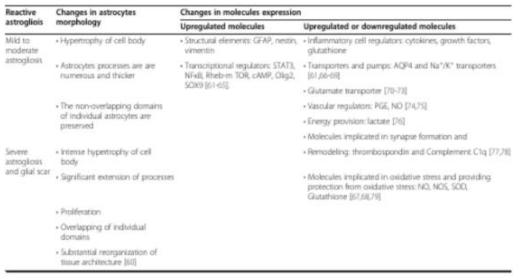


Problem of local-object-based methods:

- Rules have limitations
- Empty cell ambiguity

2. Method

✵ Insight



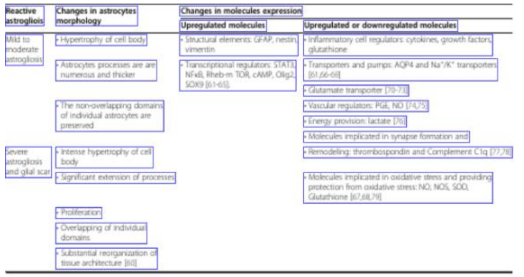
Reactive astrogliosis	Changes in astrocytes morphology	Changes in molecules expression	
		Upregulated molecules	Upregulated or downregulated molecules
Mild to moderate astrogliosis	• Hypertrophy of cell body	• Structural elements: GFAP, nestin, vimentin	• Inflammatory cell regulators: cytokines, growth factors, glutathione
	• Astrocytes processes are are numerous and thicker	• Transcriptional regulators: STAT3, NFκB, Rheb-m TOR, cAMP, Olig2, SOX9 [61-65].	• Transporters and pumps: AQP4 and Na^+/K^+ transporters [61,66-69]
			• Glutamate transporter [70-73]
	• The non-overlapping domains of individual astrocytes are preserved		• Vascular regulators: PGE, NO [74,75]
			• Energy provision: lactate [76]
			• Molecules implicated in synapse formation and
Severe astrogliosis and glial scar	• Intense hypertrophy of cell body		• Remodeling: thrombospondin and Complement C1q [77,78]
	• Significant extension of processes		• Molecules implicated in oxidative stress and providing protection from oxidative stress: NO, NOS, SOD, Glutathione [67,68,79]
	• Proliferation		
	• Overlapping of individual domains		
	• Substantial reorganization of tissue architecture [80]		

Image of Table

→

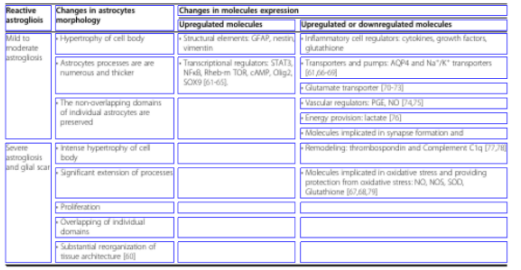
Detection Algorithm

↗



Position of Text

↘



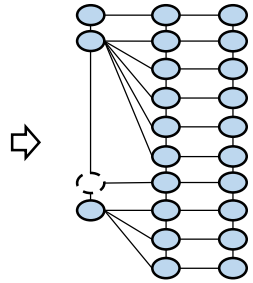
Position of Aligned Cell

↗

Table Structure

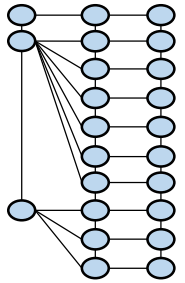
Insight

Treatment phase	Adverse event	No. of patients
T1	Swelling	1
	Itching	1
	Fever	4
	Throat infection	1
	Chest Congestion	2
	Total	9
T2	Diarrhea	1
	Body Pain	1
	Total	2



(a)

Treatment phase	Adverse event	No. of patients
T1	Swelling	1
	Itching	1
	Fever	4
	Throat infection	1
	Chest Congestion	2
	Total	9
T2	Diarrhea	1
	Body Pain	1
	Total	2



(b)

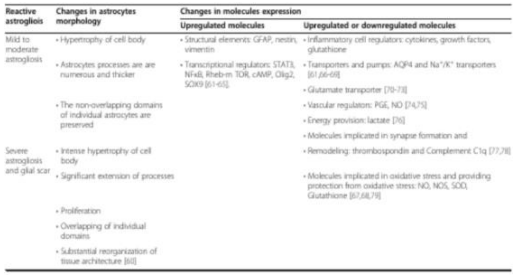
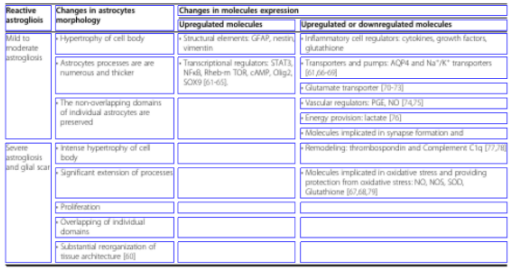


Image of Table

LGPMA

No Visible Boundaries

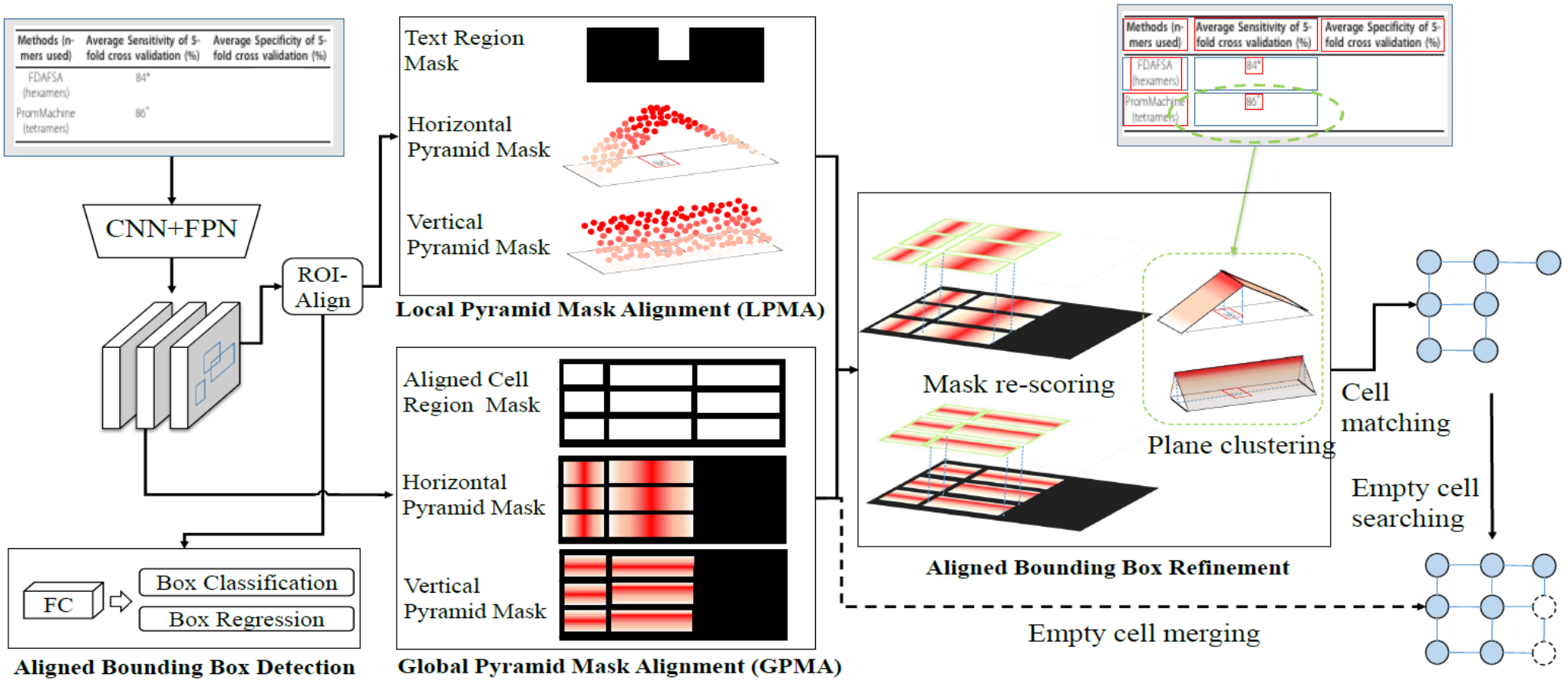
Table Structure



Position of Aligned Cell

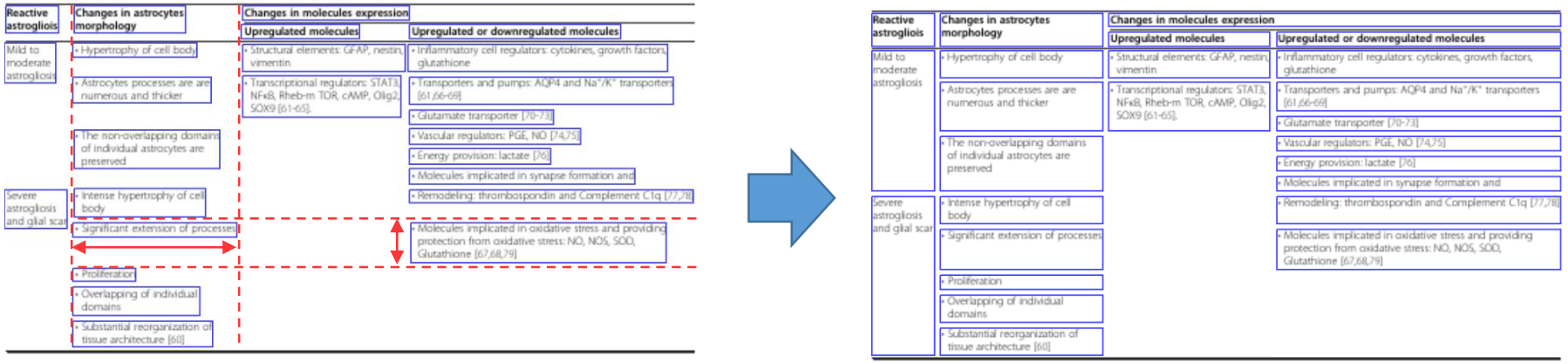
Lacking of Annotations

✵ Workflow of LGPMA



✵ Aligned Bounding Box Detection

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



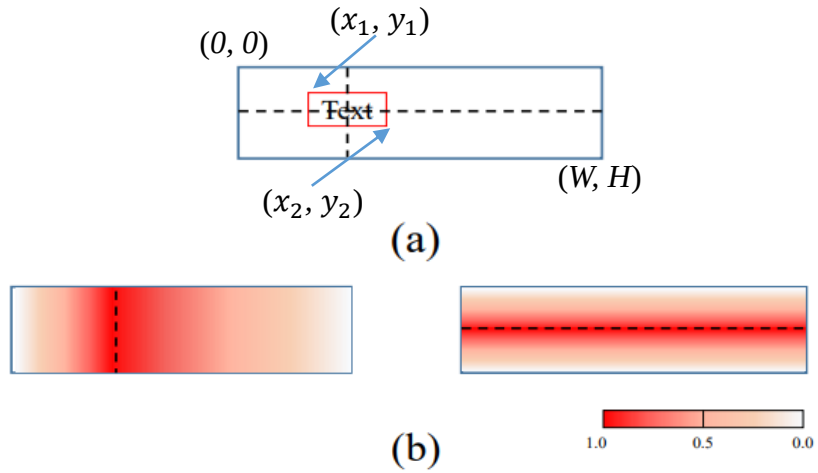
Method to approximate real cell regions

✵ 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

$$t_h^{(w,h)} = \begin{cases} w/x_1 & w \le x_{mid} \\ \frac{W-w}{W-x_2} & w > x_{mid} \end{cases}, \quad t_v^{(w,h)} = \begin{cases} h/y_1 & h \le y_{mid} \\ \frac{H-h}{H-y_2} & h > y_{mid} \end{cases}, \qquad (1)$$

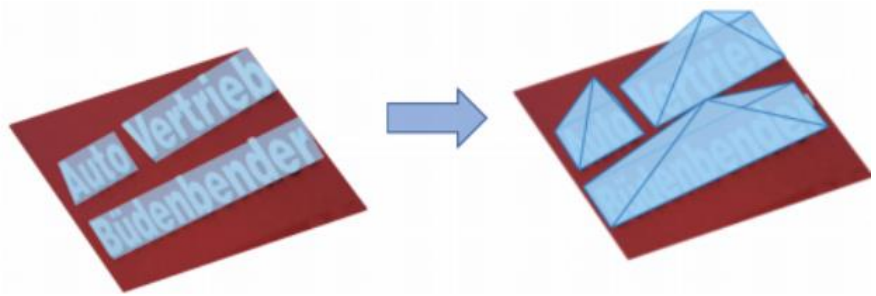


(a)aligned cell box (blue) and text region box (red).

(b) pyramid mask labels in horizontal and vertical

✵ 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



PMTD: Pyramid Mask Text Detector, 2019

✵ Aligned Bounding Box Refine

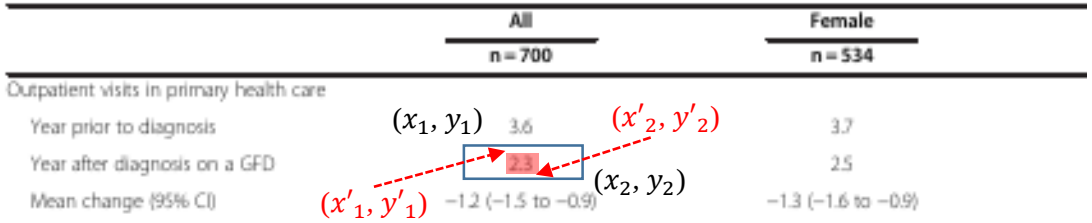
- Re-scoring strategy to compromise local and global pyramid mask

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

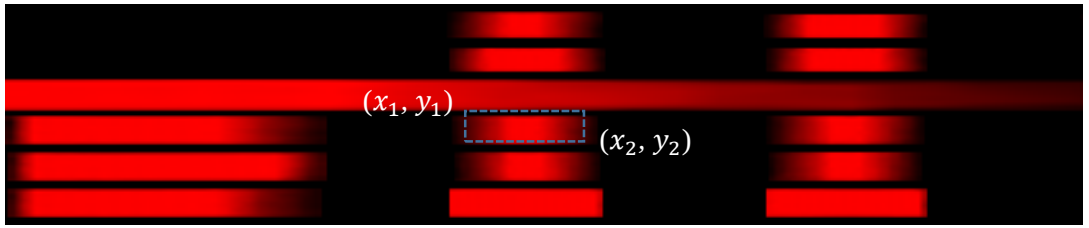
$$F(x) = \begin{cases} \frac{x-x_1}{x_{mid}-x_1} F^{(L)}_{hor}(x,y) + \frac{x_{mid}-x}{x_{mid}-x_1} F^{(G)}_{hor}(x,y) & x_1 \le x \le x_{mid} \\ \frac{x-x_2}{x_{mid}-x_2} F^{(L)}_{hor}(x,y) + \frac{x_{mid}-x}{x_{mid}-x_2} F^{(G)}_{hor}(x,y) & x_{mid} < x \le x_2 \end{cases},$$

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

Local pyramid masks



Original bounding box of cell and mask of text

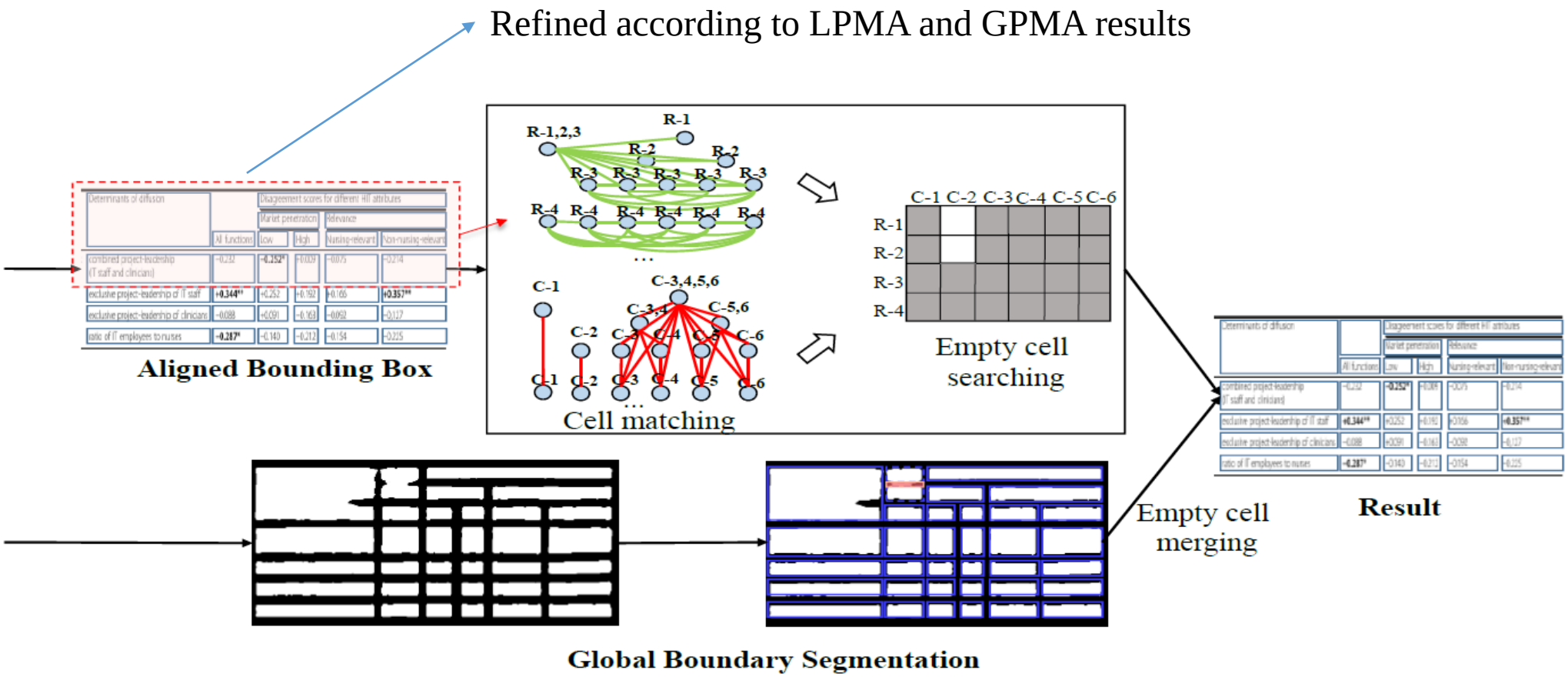


Global pyramid masks

- 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



3. Experiment

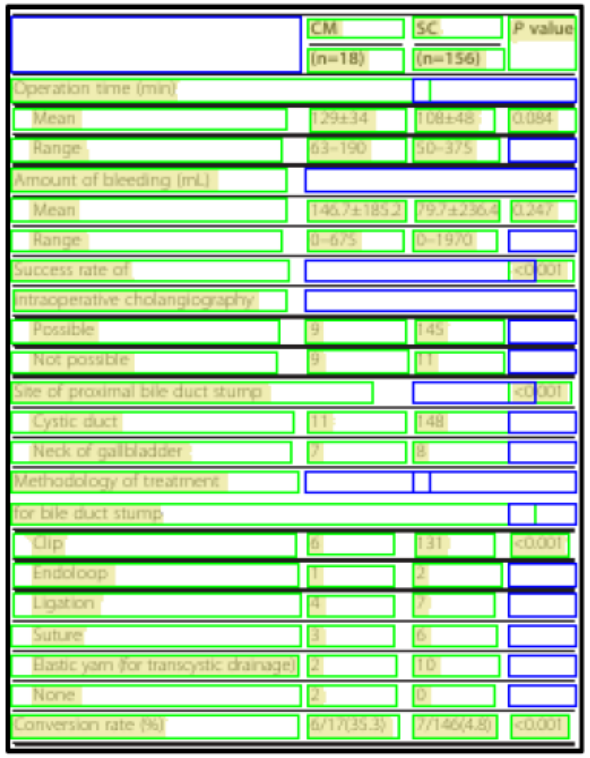
✵ Implement Details:

- 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: mutli-scale training
- Single scale testing

✵ Visualization Results:

	THRESHOLD FOR RELEASES		
	to air kg/year	to water kg/year	to land kg/year
1,2,3,4,5,6- hexachlorocyclohexane (HCH)	10	1	1
Alachlor	-	1	1
Aldrin	1	1	1
Atrazine	-	1	1
Chlordane	1	1	1
Chlordecone	1	1	1
Chlorfenvinphos	-	1	1
Chlorpyrifos	-	1	1
DDT	1	1	1
Diuron	-	1	1
Endosulphan	-	1	1
Endrin	1	1	1
Heptachlor	1	1	1
Isodrin	-	1	-
Isoproturon	-	1	1
Lindane	1	1	1
Mirex	1	1	1
Simazine	-	1	1
Toxaphene	1	1	1
Tributyltin and compounds	-	1	1
Trifluralin	-	1	1
Triphenyltin and compounds	-	1	1

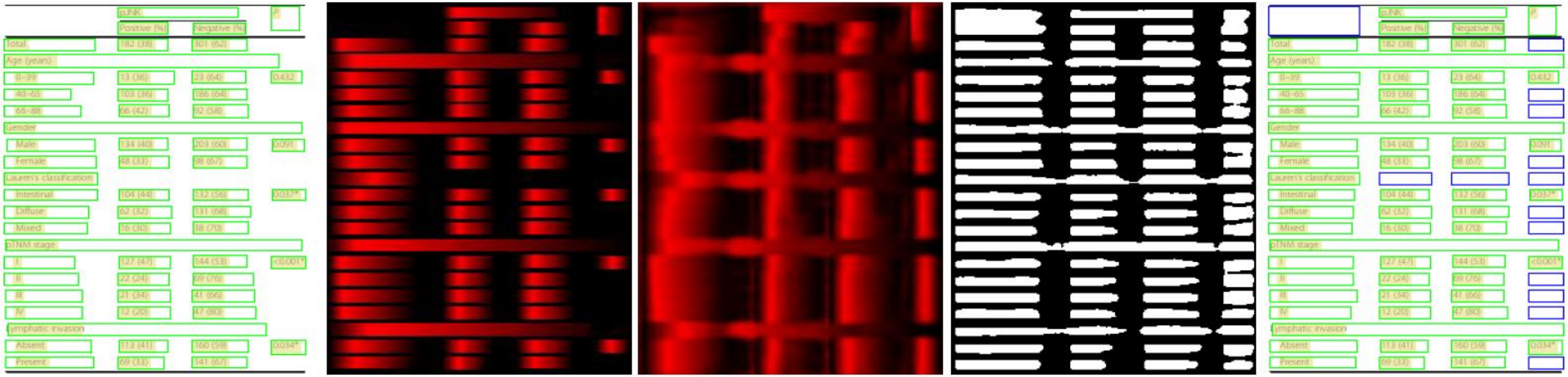
Semmatic labels → Cityscapes imags				
	PSNR(dB)	SSIM	UQI	VIF
pix2pix-cGAN	15.74	0.4275	0.07315	0.05208
PAN	**16.06**	**0.4820**	**0.1116**	**0.06581**
Edges → Shoes				
	PSNR(dB)	SSIM	UQI	VIF
ID-cGAN	**20.07**	0.7504	0.2724	0.2268
PAN	19.51	**0.7816**	**0.3442**	**0.2393**
Edges → Handbags				
	PSNR(dB)	SSIM	UQI	VIF
ID-cGAN	**16.50**	0.6307	0.3978	0.1723
PAN	15.90	**0.6570**	**0.4042**	**0.1841**
Cityscapes images → Semantic labels				
	PSNR(dB)	SSIM	UQI	VIF
ID-cGAN	19.46	0.7270	0.1555	0.1180
PAN	**20.67**	**0.7725**	**0.1732**	**0.1638**
Aerial photos → Maps				
	PSNR(dB)	SSIM	UQI	VIF
ID-cGAN	26.10	0.6465	0.09125	0.02913
PAN	**28.32**	**0.7520**	**0.3372**	**0.1617**



	CM (n=18)	SC (n=156)	P value
Operation time (min)			
Mean	129±34	108±48	0.084
Range	83–190	50–375	
Amount of bleeding (mL)			
Mean	146.7±185.2	79.7±236.4	0.247
Range	0–675	0–1970	
Success rate of			<0.001
Intraoperative cholangiography			
Possible	9	145	
Not possible	9	11	
Site of proximal bile duct stump			<0.001
Cystic duct	11	148	
Neck of gallbladder	7	8	
Methodology of treatment			
for bile duct stump			
Clip	8	131	<0.001
Endoloop	1	2	
Ligation	4	7	
Suture	3	6	
Elastic yarn (for transcystic drainage)	2	10	
None	2	0	
Conversion rate (%)	0/17(35.3)	7/146(4.8)	<0.001

	LRS+ Mean±SD/ (N,%)	LRS− Mean±SD/ (N,%)	P-value
	3,079 (53.9)	2,634 (46.1)	
Gender			
Male (N, %)	3,346 (43.72%)	3,126 (42.75%)	0.462
Female (N, %)	3,733 (56.28%)	3,508 (57.25%)	
Age (Mean±SD)	67.79 (±10.68)	67.63 (±11.15)	0.599
Age (N,%)			
<65 years	3,092 (35.47%)	952 (36.14%)	0.491
65-74 years	3,150 (37.35%)	944 (35.84%)	
>74 years	837 (27.18%)	738 (28.02%)	
ECI score (N,%)			
0	272 (8.83%)	341 (12.95%)	<0.001*
1	217 (7.05%)	245 (9.30%)	
≧2	2,590 (84.12%)	2,048 (77.75%)	
Comorbidity			
Hypertension			
No	3,443 (46.87%)	3,300 (49.35%)	0.061
Yes	3,636 (53.13%)	3,334 (50.65%)	
Hyperlipidemia			
No	2,371 (77.01%)	2,138 (81.17%)	<0.001*
Yes	708 (22.99%)	496 (18.83%)	
Diabetes			
No	2,345 (76.16%)	2,054 (77.98%)	0.103
Yes	734 (23.84%)	580 (22.02%)	
ESRD			
No	2,490 (80.87%)	2,072 (78.66%)	0.038*
Yes	589 (19.13%)	562 (21.34%)	

✵ 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	0.959	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	Tesing Dataset	TEDS (All)	TEDS-Struc. (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.
	LPMA	GPMA	AL[26]	Precision	Recall	Hmean	Precision	Rrecall	Hmean	
Faster R-CNN				-	-	-	81.32	81.31	81.31	94.63
Mask R-CNN				91.71	91.53	91.62	81.83	81.82	81.83	94.65
	✓			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
	✓	✓		**92.27**	**91.86**	**92.06**	**85.14**	**84.77**	**84.95**	**95.53**
Mask R-CNN			✓	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