

Towards Robust Tampered Text Detection in Document Image: New dataset and New Solution (Supplementary Material)

Chenfan Qu¹, Chongyu Liu¹, Yuliang Liu², Xinhong Chen¹, Dezhi Peng¹, Fengjun Guo³,
Lianwen Jin^{1,*}

¹South China University of Technology, ²Huazhong University of Science and Technology,
³IntSig Information Co., Ltd

202221012612@mail.scut.edu.cn, eelwj@scut.edu.cn

Abstract

In this supplementary material, we first show some extensive ablation experiments on the proposed Frequency Perception Head (FPH), Multi-view Iterative Decoder (MID) and Curriculum Learning for Tampering Detection (CLTD). Then we show image-level authentication experiments on the DocTamper dataset, the T-SROIE dataset and the T-IC13 dataset. Moreover, we show more details about the DocTamper dataset and results about three kinds of tampered text respectively. Afterwards, we show the models' cross domain performance on the T-SROIE dataset and DTD's performance on handmade tampered data. In addition, we show scene text tampering detection experiment on the T-IC13 dataset. Finally, we show some representative source images of the DocTamper dataset and some visualization results of the experiments.

1. Extensive ablation experiments

1.1. Ablation study on FPH

Previous works on image manipulation detection utilized information from various noise domains to help model locate tampered regions [3, 5, 8, 11, 12, 14]. In this section, we replace the input of the FPH with image filtered by some commonly used noise domain filters to explore their performance on the DocTamper dataset.

The original FPH computes the absolute value of DCT coefficients and truncates them to [0, 20], each pixel on DCT feature map will be a one-hot vector after being embedded by orthonormal basis. We replace the structure before the down-sampling layer in FPH with different noise domain filters. The structure after the down-sampling layer of FPH keeps the same in all experiments.

Results are shown in Table 1. 'DCT coef' denotes that using raw DCT coefficients as input as in Wang et al. [12]; 'LoG' denotes that using Laplacian of Gaussian with a

residual connection structure filtered image as input as in Wang et al. [11]; 'Bayar' denotes that using constrained convolutional layer [2] filtered image as input as in MVSS-Net [3]; 'High-pass' denotes that using high-pass filtered image as input as in ObjectFormer [8]. 'SRM' denotes that using Steganography Rich Model (SRM) [4] filtered image as input as in RGB-N [14]; 'JALM' denotes that using JPEG Artifacts Learning Module's output as input as in CAT-Net [5]. Results with IoU metric in different image compression settings are also shown in Table 2. Obviously the input design of FPH is better in helping our model locate tampered text in all the experiments.

1.2. Ablation study on MID

In this section, we evaluate model's performance with different numbers of iterations in MID. As shown in Table 6, adding iteration can help improve model's performance efficiently, especially when MID has less than three iterations. we can also get the same conclusion from experiments with different image compression settings, as shown in Table 7.

1.3. Ablation study on CLTD

In this section, we conduct experiments with different temperature factors T in CLTD. Larger T means longer transition from easier samples to harder samples. Results are shown in Table 8, we can find that DTD is relatively robust to the value of T and a moderate value will be the best. Experiments with different image compression settings are as shown in Table 9.

2. Image-level authentication experiments

Image-level authentication denotes identifying whether an input image contains tampered region or not. We conduct binary classification experiments on the DocTamper testing sets and their authentic images. Results are shown in Table 10. We can find that DTD achieves satisfactory performance on image-level authentication task on the Doc-

Table 1. Ablation study of FPH on the DocTammer dataset. All images are compressed randomly one to three times with random quality factors choiced from 75 to 100 and the same random seed. "P" denotes precision, "R" denotes recall and "F" denotes F-score.

Method	Testing set				DocTammer-FCD				DocTammer-SCD			
	IoU	P	R	F	IoU	P	R	F	IoU	P	R	F
DCT coef [12]	0.371	0.637	0.581	0.608	0.247	0.537	0.271	0.360	0.526	0.591	0.591	0.591
LoG [11]	0.640	0.637	0.555	0.593	0.403	0.614	0.435	0.509	0.538	0.589	0.575	0.582
Bayar [3]	0.704	0.673	0.605	0.637	0.523	0.665	0.568	0.613	0.560	0.617	0.621	0.619
High-pass [8]	0.723	0.693	0.619	0.654	0.512	0.647	0.556	0.598	0.582	0.639	0.637	0.638
SRM [14]	0.759	0.736	0.672	0.702	0.549	0.693	0.589	0.637	0.601	0.667	0.681	0.674
JALM [5]	0.812	0.797	0.743	0.769	0.669	0.837	0.697	0.761	0.668	0.723	0.745	0.734
FPH (Ours)	0.828	0.814	0.771	0.792	0.749	0.849	0.786	0.816	0.691	0.745	0.762	0.754

Table 2. Ablation study of FPH on the DocTammer dataset with different image compression settings. IoU metric is used in all the experiments. "Q" denotes the lowest compression quality factor in a series image compression.

Method	Testing set				DocTammer-FCD				DocTammer-SCD			
	Q 75	Q 80	Q 85	Q 90	Q 75	Q 80	Q 85	Q 90	Q 75	Q 80	Q 85	Q 90
DCT coef [12]	0.371	0.373	0.369	0.369	0.247	0.248	0.264	0.301	0.526	0.538	0.554	0.593
LoG [11]	0.640	0.659	0.676	0.712	0.403	0.411	0.430	0.492	0.538	0.551	0.568	0.607
Bayar [3]	0.704	0.720	0.733	0.769	0.523	0.538	0.538	0.612	0.560	0.574	0.588	0.626
High-pass [8]	0.723	0.740	0.756	0.788	0.512	0.527	0.534	0.623	0.582	0.595	0.610	0.642
SRM [14]	0.759	0.775	0.795	0.825	0.549	0.570	0.576	0.675	0.601	0.614	0.633	0.671
JALM [5]	0.812	0.834	0.856	0.883	0.669	0.699	0.734	0.799	0.668	0.693	0.726	0.763
FPH (Ours)	0.828	0.848	0.870	0.893	0.746	0.785	0.804	0.827	0.691	0.716	0.747	0.780

Tamper dataset, the T-SROIE dataset [12] and the T-IC13 dataset [11].

Table 3. F-score of three kinds of tampered text respectively on the DocTammer dataset with Q75 setting.

Method	Copy-Move	Splicing	Generation
Mantra-Net [13]	0.1020	0.1171	0.2490
MVSS-Net [3]	0.2532	0.4215	0.7036
PSCC-Net [6]	0.2184	0.3935	0.6251
BEIT-Uper [1]	0.2891	0.4662	0.8247
Swin-Uper [7]	0.4746	0.6408	0.8789
CAT-Net [5]	0.5705	0.6897	0.8969
DTD (Ours)	0.7018	0.8086	0.9205

3. More details about the DocTammer dataset

In this section, we show more details about the DocTammer dataset. The DocTammer dataset has a total of 582549 tampered text instances. The tampered texts in the DocTammer dataset have various heights, widths and angles. Most of the tampered texts in the DocTammer dataset have an area ranging from 0 to 5000 pixels. The area distribution of the tampered texts is shown in Fig. 1. The height of most tam-

Table 4. Models' performance on the T-SROIE dataset when trained with the DocTammer dataset only. "P" denotes precision, "R" denotes recall and "F" denotes F-score.

Method	P	R	F
MVSS-Net [3]	0.3521	0.7032	0.4692
BEIT-Uper [1]	0.3321	0.6293	0.4348
Swin-Uper [7]	0.5943	0.5146	0.5516
CAT-Net [5]	0.6933	0.7565	0.7235
DTD (Ours)	0.8072	0.7958	0.8014

Table 5. Comparison on the scene text tampering detection T-IC13 dataset. "P" denotes precision, "R" denotes recall and "F" denotes F-score.

Method	P	R	F
EAST [15]	0.7321	0.7515	0.7417
PSENet [9]	0.8495	0.8391	0.8443
ATRR [10]	0.8610	0.9084	0.8840
Wang et al. [11]	0.8843	0.9185	0.9011
DTD (Ours)	0.9217	0.8934	0.9073

pered texts ranges from 0 to 80 pixels, as shown in Fig.2.

Table 6. Ablation study of MID on the DocTamper dataset. All images are compressed randomly one to three times with random quality factors choiced from 75 to 100 and the same random seed. "P" denotes precision, "R" denotes recall and "F" denotes F-score.

Num. of iteration	Testing set				DocTamper-FCD				DocTamper-SCD			
	IoU	P	R	F	IoU	P	R	F	IoU	P	R	F
One iteration	0.706	0.715	0.591	0.647	0.594	0.838	0.603	0.701	0.570	0.671	0.578	0.621
Two iterations	0.753	0.777	0.739	0.758	0.715	0.836	0.769	0.801	0.651	0.712	0.727	0.719
Three iterations	0.793	0.795	0.746	0.770	0.733	0.853	0.770	0.809	0.668	0.726	0.735	0.730
Four iterations	0.828	0.814	0.771	0.792	0.749	0.849	0.786	0.816	0.691	0.745	0.762	0.754

Table 7. Ablation study of MID on the DocTamper dataset with different image compression settings. IoU metric are used in all the experiments. "Q" denotes the lowest compression quality factor in a series image compression.

Num. of iteration	Testing set				DocTamper-FCD				DocTamper-SCD			
	Q 75	Q 80	Q 85	Q 90	Q 75	Q 80	Q 85	Q 90	Q 75	Q 80	Q 85	Q 90
One iteration	0.706	0.737	0.776	0.827	0.594	0.633	0.682	0.749	0.570	0.600	0.643	0.701
Two iterations	0.753	0.778	0.800	0.824	0.715	0.750	0.777	0.798	0.651	0.677	0.714	0.752
Three iterations	0.793	0.814	0.838	0.866	0.733	0.767	0.789	0.823	0.668	0.694	0.727	0.767
Four iterations	0.828	0.848	0.870	0.893	0.746	0.785	0.804	0.827	0.691	0.716	0.747	0.780

Table 8. Ablation study of CLTD on the DocTamper dataset. All images are compressed randomly one to three times with random quality factors choiced from 75 to 100 and the same random seed. "P" denotes precision, "R" denotes recall and "F" denotes F-score.

Value of T	Testing set				DocTamper-FCD				DocTamper-SCD			
	IoU	P	R	F	IoU	P	R	F	IoU	P	R	F
T=2048	0.781	0.776	0.723	0.749	0.634	0.830	0.660	0.735	0.651	0.706	0.715	0.710
T=4096	0.793	0.785	0.730	0.757	0.702	0.835	0.725	0.776	0.664	0.727	0.715	0.721
T=16384	0.806	0.790	0.744	0.766	0.710	0.851	0.738	0.790	0.657	0.720	0.737	0.728
T=8192	0.828	0.814	0.771	0.792	0.749	0.849	0.786	0.816	0.691	0.745	0.762	0.754

Table 9. Ablation study of CLTD on the DocTamper dataset with different image compression settings. IoU metric are used in all the experiments. "Q" denotes the lowest compression quality factor in a series image compression.

Value of T	Testing set				DocTamper-FCD				DocTamper-SCD			
	Q 75	Q 80	Q 85	Q 90	Q 75	Q 80	Q 85	Q 90	Q 75	Q 80	Q 85	Q 90
T=2048	0.781	0.803	0.830	0.862	0.634	0.678	0.720	0.792	0.651	0.678	0.712	0.754
T=4096	0.793	0.816	0.843	0.871	0.702	0.742	0.777	0.817	0.664	0.690	0.721	0.759
T=16384	0.806	0.829	0.857	0.886	0.710	0.749	0.778	0.813	0.657	0.688	0.730	0.770
T=8192	0.828	0.848	0.870	0.893	0.746	0.785	0.804	0.827	0.691	0.716	0.747	0.780

Table 10. Image-level authentication experiments. "Q" denotes the lowest compression quality factor in a series image compression. "R-T", "P-T", "F-T" denotes recall, precision and F-score for tampered image, respectively. "R-A", "P-A", "F-A" denotes recall, precision and F-score for authentic image, respectively. "mF" denotes mean F-score.

Dataset	R-T	P-T	F-T	R-A	P-A	F-A	mF
DocTamper Testing set (Q 75)	0.9769	0.9941	0.9854	0.9942	0.9773	0.9857	0.9855
DocTamper Testing set (Q 80)	0.9816	0.9949	0.9882	0.9949	0.9818	0.9883	0.9882
DocTamper Testing set (Q 85)	0.9847	0.9958	0.9902	0.9958	0.9848	0.9903	0.9902
DocTamper Testing set (Q 90)	0.9866	0.9973	0.9919	0.9973	0.9868	0.9920	0.9920
DocTamper-FCD (Q 75)	0.9840	0.9875	0.9857	0.9875	0.9841	0.9858	0.9857
DocTamper-FCD (Q 80)	0.9845	0.9890	0.9867	0.9890	0.9846	0.9868	0.9867
DocTamper-FCD (Q 85)	0.9865	0.9910	0.9887	0.9910	0.9866	0.9888	0.9887
DocTamper-FCD (Q 90)	0.9875	0.9960	0.9917	0.9960	0.9876	0.9918	0.9917
DocTamper-SCD (Q 75)	0.9681	0.9999	0.9837	0.9999	0.9691	0.9843	0.9840
DocTamper-SCD (Q 80)	0.9697	0.9999	0.9846	0.9999	0.9706	0.9851	0.9848
DocTamper-SCD (Q 85)	0.9728	1.0	0.9862	1.0	0.9736	0.9866	0.9864
DocTamper-SCD (Q 90)	0.9737	1.0	0.9867	1.0	0.9743	0.9870	0.9868
T-SROIE [12]	0.9916	1.0	0.9958	-	-	-	-
T-IC13 [11]	0.9831	0.9943	0.9887	0.9818	0.9473	0.9642	0.9765

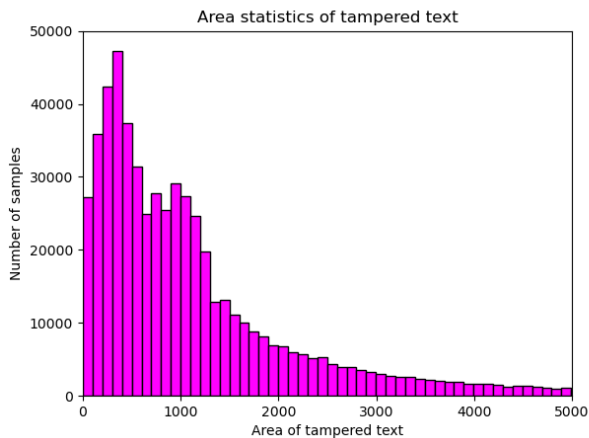


Figure 1. Area Statistics of tampered texts in the DocTamper dataset.

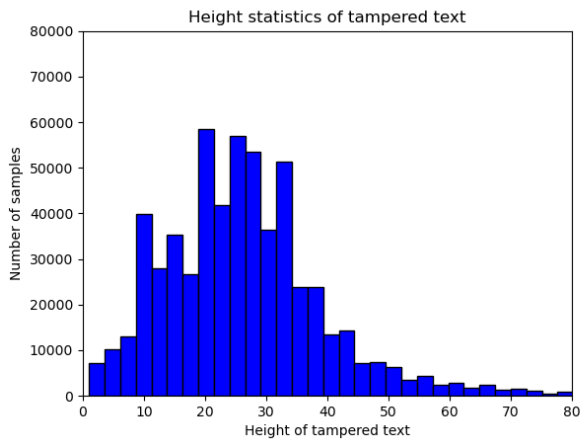


Figure 2. Height Statistics of tampered texts in the DocTamper dataset.



Figure 3. Predictions on handmade tampered text.

4. Results about the three types of tampered text respectively

In this section, we show the results about the three types of tampered text respectively on the DocTamper testing set, results are shown in Table 3. Copy-move is the hardest due to the consistency between font and background. Splicing may leave more manipulation clues thus is relatively easier. Generation is the easiest as the tampered texts have the most differences from the authentic ones.

5. Cross domain performance on the T-SROIE dataset

In this section, we show the cross domain performance of the models on the T-SROIE dataset. The pixel-level precision, recall and F-measure of the models trained with DocTamper dataset only are shown in Table 4, the results show that the proposed DTD generalizes relatively well.

6. Performance on handmade tampered data

Although the tampered samples in the DocTamper dataset are generated automatically, it’s worth noting that the handmade tampered image also requires synthetic technique in digital image processing software, such as PhotoShop and GIMP. Our carefully designed synthesis pipeline follows the same way to mimic the real-world tampering. We further conduct an experiment on an in-house handmade tampered dataset. DTD trained with the DocTamper dataset achieves 96.5% accuracy. Some examples are shown in Fig. 3.

7. Scene Text tampering detection

In this section, we train and evaluate the proposed Document Tampering Detector (DTD) on the scene text tampering detection dataset, T-IC13 [11]. The prediction mask of the model are binarized and clustered with a dilation kernel K , the kernel K has a height $1/200$ of the input image’s height and a width $1/30$ of the input image’s width. Then we get the maximum circumscribed boxes of every connected components as prediction boxes and evaluate the model’s performance with the official evaluation tool provided by the authors of the T-IC13 dataset. Results are shown in Table 5. Although DTD is designed for detecting subtle tampered regions that have few visual tampering clue on document image, instead of detecting tampered scene text of various shapes with a very small-scale training set, it also get comparable results to previous SOTA method on the T-IC13 dataset in F-measure due to its powerful tampering feature extraction ability.

8. Visualization

In this section, we first show some of the representative source images of the DocTamper dataset. Then we show some of the visualization results of the experiments.

The representative source images of the DocTamper dataset are shown in Figure 4, 5, 6. The visualization results of ablation study are shown in Figure 7, 8. The visualization results on the T-SROIE dataset [12] are shown in Figure 9, 10, 11. The visualization results on the T-IC13 dataset [11] are shown in Figure 12, 13.

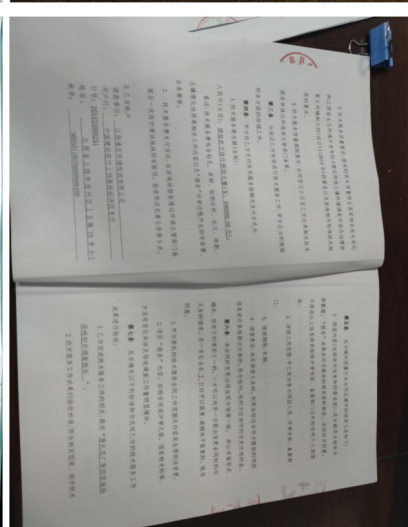
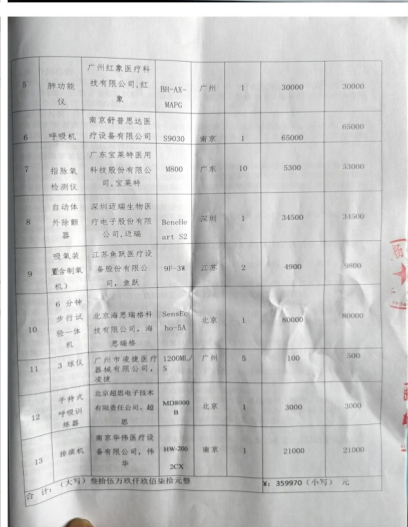
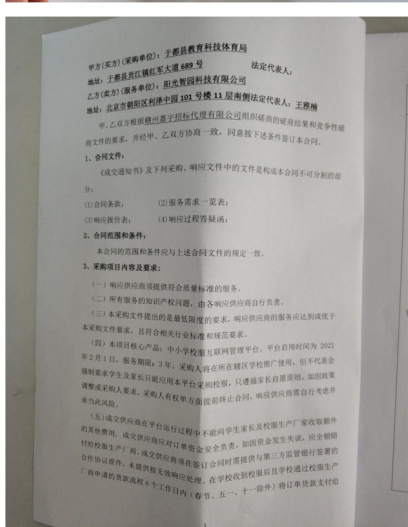
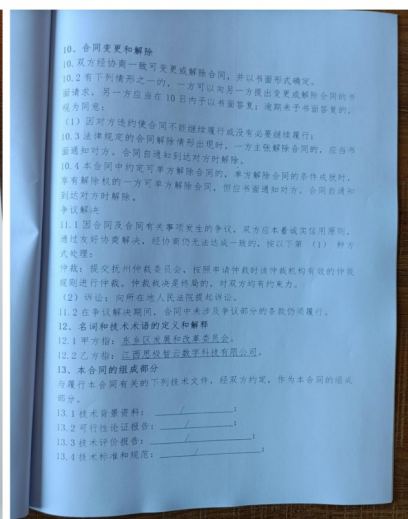
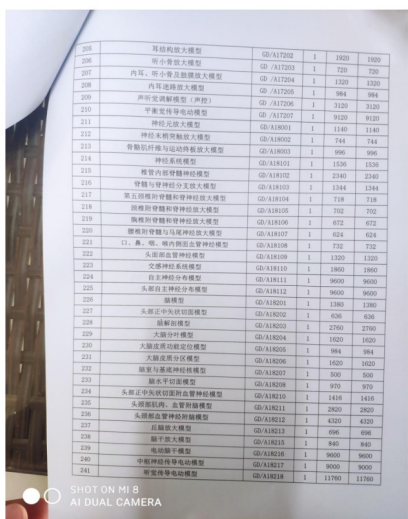
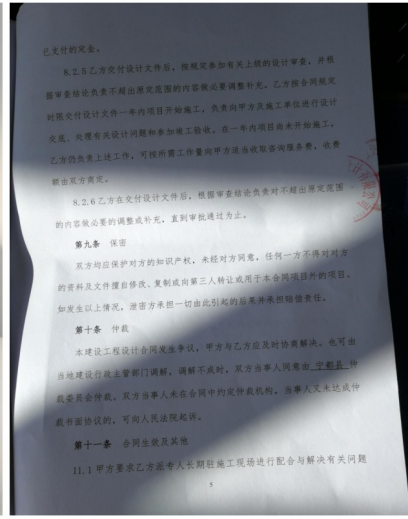
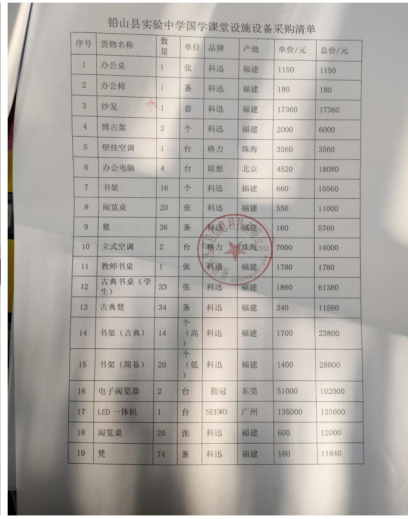
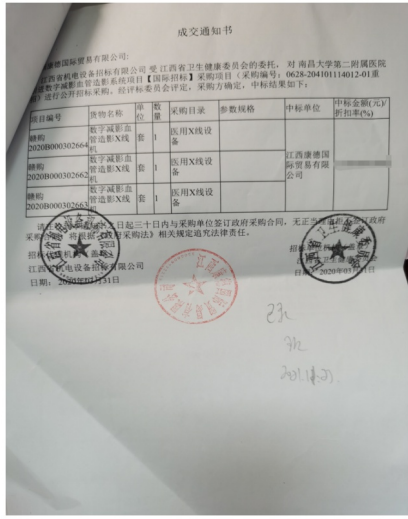


Figure 4. Contract images included in the source images of the DocTammer dataset.

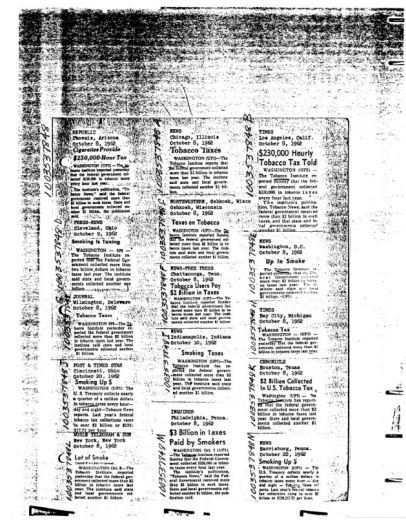
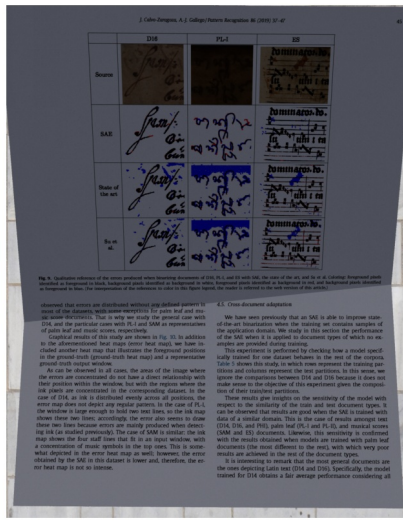
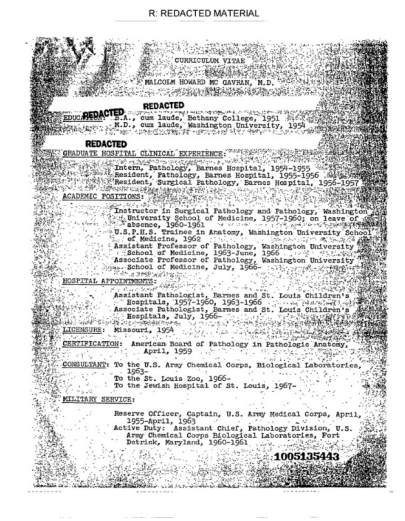
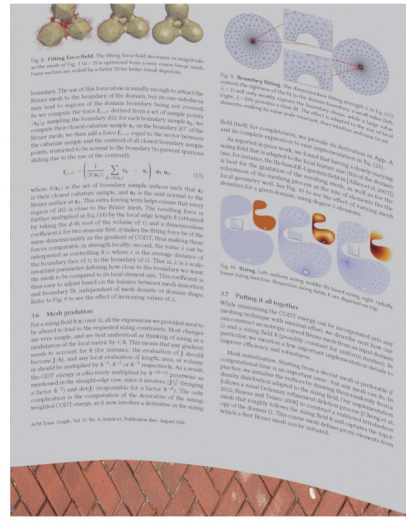
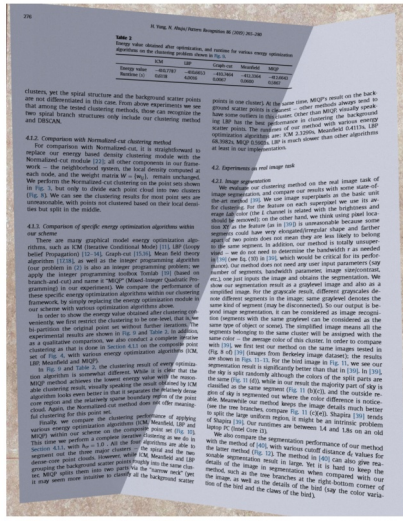


Figure 5. Invoices and normal pages included in the source images of the DocTamer dataset.

A new offline handwritten database for the Spanish language, which contains full Spanish sentences, has recently been developed: the Spartacus database (which stands for Spanish Restricted-domain Task of Cursive Script). There were two main reasons for creating this corpus. First of all, most databases do not contain Spanish sentences, even though Spanish is a widespread major language. Another important reason was to create a corpus from semantic-restricted tasks. These tasks are commonly used in practice and allow the use of linguistic knowledge beyond the lexicon level in the recognition process.

As the Spartacus database consisted mainly of short sentences and did not contain long paragraphs, the writers were asked to copy a set of sentences in fixed places: dedicated one-line fields in the forms. Next figure shows one of the forms used in the acquisition process. These forms also contain a brief set of instructions given to the writer.

A new offline handwritten database for the Spanish language, which contains full Spanish sentences, has recently been developed: the Spartacus database (which stands for Spanish Restricted-domain Task of Cursive Script). There were two main reasons for creating this corpus. First of all, most databases do not contain Spanish sentences, even though Spanish is a widespread major language. Another important reason was to create a corpus from semantic-restricted tasks. These tasks are commonly used in practice and allow the use of linguistic knowledge beyond the lexicon level in the recognition process.

As the Spartacus database consisted mainly of short sentences and did not contain long paragraphs, the writers were asked to copy a set of sentences in fixed places: dedicated one-line fields in the forms. Next figure shows one of the forms used in the acquisition process. These forms also contain a brief set of instructions given to the writer.

WHOLE FOODS MARKET

West Orange WQP
235 Prospect St
West Orange
New Jersey, 07052
973-669-3196

Food/Beverage	
GNJI RAMEN MINI VEGGIE	\$8.00 T
0G BARTLETT PEARS	
0.82 lb @ \$1.99 / lb	\$1.63 F
Tare Weight 0.011b	
365 COCONUT WATER	\$3.29 F
LNDB WLD BLND RICE	\$3.99 F
BBY BRSSL SPRT 500GR	\$4.99 F

Subtotal: \$21.90
Total Savings: \$0.00
Net Sales: \$21.90
Tax/Fee: \$0.55
Total: \$22.45

Sold Items: 5
Paid: \$22.45
Cash

Seafood snack bar

The third avenue.
DATA: 07-06-2021 at 8:10:11 PM
Cash Counter No. 2 Bill No. 101

Mains	Rate	Amount
1 x Linguini	\$ 10.00	\$ 10.00
2 x Lasagna	\$ 7.50	\$ 15.00
Desserts		
1 x Moules-frite	\$ 3.50	\$ 3.50
1 x Waffle	\$ 2.20	\$ 2.20

SUBTOTAL \$ 31.00
TAX \$ 3.11
GRATUITY \$ 0.00
TOTAL \$ 34.10

THANK YOU FOR DINING WITH US!
PLEASE COME AGAIN

CARE RITE PHARMACY

SUNRISE, FL
Lomita, CA
COPY RECEIPT
CLERK

Description	Qty	Rate	Amount
salbutamol	1	7.7	
phenacetin	1	3.3	
ephedrine	1	6.5	6.5

TAX: \$ 16.50

TOTAL \$ 16.50
05/25/2018 14:32

新子荷超市
NO. 12202203100095
2022.03.10 09:03:11
收银机:12 收银员:1016

货号/品名	单价	数量	小计
2203903004702/花蛤(小)	19.800.24	4.70	
2204521001500/水豆腐	5.000.30	1.50	

数量:0.54 件数:2
原价合计: 6.20
折扣:0.00
应付合计: 6.20

付款:思迅Pay-微信 6.20
备注信息:80640959011220220310009
5005/2022-03-10 09:03:31
谢谢惠顾! 欢迎下次光临!
欢迎下次光临

Jack's Food

DAYTONA BEACH, FL

Purchase
DATE:06/04/2018 TIME:09:43 AM

Description	Qty	Rate	Amount
Butter Tart	1	\$4.0	\$4.0
Egg-n-bag	1	\$7.5	\$7.5
Covidax	2	\$13.0	\$26.0

TAX \$ 0.00
TOTAL \$ 37.5

THANK YOU

家家樂超市 (南灣店)

684738002002	綠中骨 (02225)	
684738002001	綠中骨 (00019)	
684738002002	綠中骨 (00019)	

商品名稱 數量 單位 金額
884304022863 康康子雞腿包 6 1 5.90

220043	金針菇	2.00	2.00
684738002002	綠中骨 (02225)	3.80	
220040	干菇蟹肉卷	12.00	12.00
220154	豬頭	2.00	2.00
220154	豬頭	2.00	2.00
220055	蒜末	1.40	1.40
220050	肉醬	5.20	5.20
220004	綠豆糕	2.50	2.50
3 肉醬捲-大		18.50	6.00
220187	脆皮雞腿	0.40	0.40
220029	1罐裝	2.00	2.00
220038	雞蛋	2.50	2.50
220011	康康子	0.225	6.56
220011	康康子	0.225	6.56
684738002001	綠中骨 (00019)	8.76	3.80
684738002001	綠中骨 (00019)	8.76	3.80

原收: 82.00
未收: 82.00
收帳方式: 現金
總計: 82.00

立即送达

【如遇缺货】缺货时电话与我联系

商品

1. 原味薯条	约20g/袋	9.90	29.70
x 3			
2. 【果切】凤梨	约250g	16.90	16.90
x 1			
3. 【果切】乌梅小番茄	约220g	19.90	19.90
x 1			

其他费用

包装费	2.00
配送费	5.00

商品合计 66.50
其他费用合计 7.00
顾客优惠合计 -41.80

总件数: 5 在线支付: 31.70

为保护隐私, 顾客地址已隐藏。您可登陆饿了么零售商家端或骑手端查看

See back of receipt for our choice to win \$1000

Walmart

Save money. Live better.

6 114 3 789 1025
MEMBER PRICE TOPICS
MEMBER CHECKOUT
WELLS FARGO BANK

Store # 072216
7/22/21 20:53:11

CHARGE DIE 0.00

ITEMS SOLD 1

100 7003 0722 1089 1649 4155

Low Prices You Can Trust. Every Day.

Barcode and QR code

Figure 6. Receipts and notes included in the source images of the DocTamper dataset.

Image	GT	Baseline	w/o FPH	w/o MID	w/o CLTD	DTD(Ours)
文件另有规定的技术要求外,本次 试、安装检验应不低于中华人民共和国 GB-2011《电梯安装验收规范》; T-17006《电梯监督检验和定期检验 规则》;T10058-2009《电梯技术条件》; GB-2009《电梯试验方法》; GB-2016《电梯工程施工质量验收规范 》; 《备安全监察条例》。 GB7588《电梯制造与安装安全规范》						
6.平衡输入电压: 200V/100V 7.信噪比: >95dB 8.尖度: <0.035 9.额定输入功率: <= 100W 10.综合效率: >70% 11.阻尼系数: <100s 12.转换效率: >95% 13.冷却系统: Two stage fan 14.电源: AC 220-230V 50Hz 15.箱体重量: <1.5kg 16.箱体尺寸: 400x300x100						
限、地点、方式 限: 签订合同后7天内完成所有工作 点: 业主指定地点 式: 甲方为乙方式成工作提供必要件 乙。 定合同规定的付合进度向乙方支付 务,质保期内提供 24 小时应急响应						
按程序的规定, 按施工进度 按选本项目的定组组工验收, 竣工工 —— 接收全部或部分工程 其他人承接工程的期限, 竣工工 合同的人接收全部或部分工程的, 该 竣工工程的, 违约金计算方法为)						
第1种方式: 采用综合 关于各可选因子, 实 约采: —— 第2种方式: 采用选价 (2) 关于基准价格的约 专用合同条款①承有 单价低于基准价格的, 单 价格为基准超过 —— 或中载明材料单价方						
二、材料采购及 本合同履行期应为一 三、合同金额、履约 上, 本合同总金额为 合同明确约定的费用外, 不 展期(包括但不限于材料 2、本合同签订之日起 本合同总金额的 90%, 即人 余 10%即人民币或按付力向 方向乙方一次性付清, 不 3、乙方还须按本合同 元) 缴纳履约保证金, 履约 方, 甲方在货到履约保证金 4、乙方指定账户);						
2.1 乙方工作范围: 按甲方规定的 5. 免费为甲方提供食品检测 止, 处置食品安全突发事件和基 提供技术支持。 9. 履行其他服务承诺, 需应急 2 小时内到达甲方指定位置, 否则 九、违约责任, 违约事宜, 双方签订合同时按合同 十、其他的约定; 1. 本合同未尽事宜, 按《合同						
十二、税费 在甲国境外发生 十三、合同生效 本合同在甲乙 十四、其它 1. 所有经双方 函), 谈判的件和程 分割的有效组成部 签字盖章确认之日 2. 如一方地 否则, 应承担相应						

Figure 7. Ablation study for DTD on the DocTamer testing set. "GT" denotes ground-truth annotation.

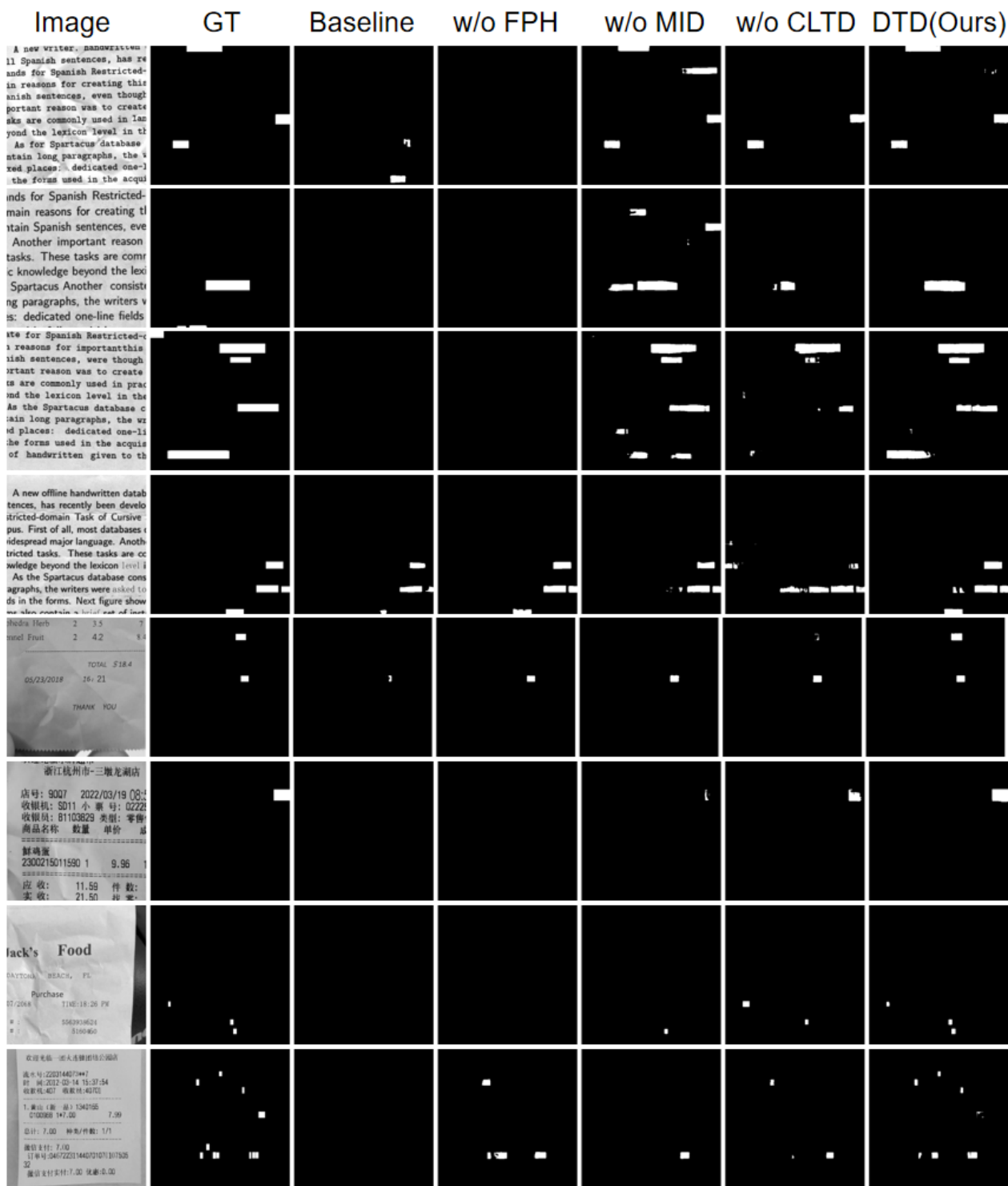


Figure 8. Ablation study for DTD on the DocTamer-FCD and the DocTamer-SCD. "GT" denotes ground-truth annotation.

Image

Prediction

GroundTruth

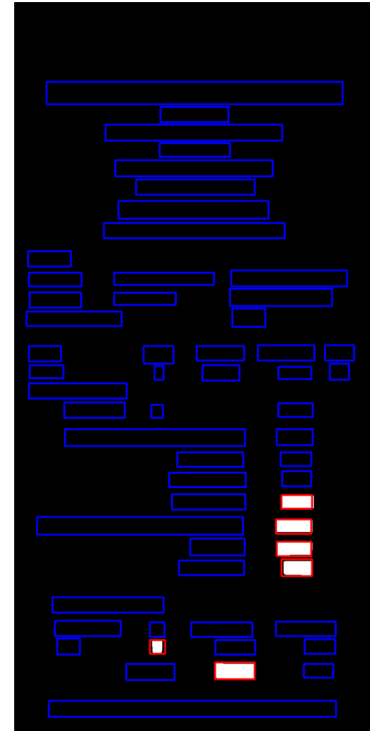
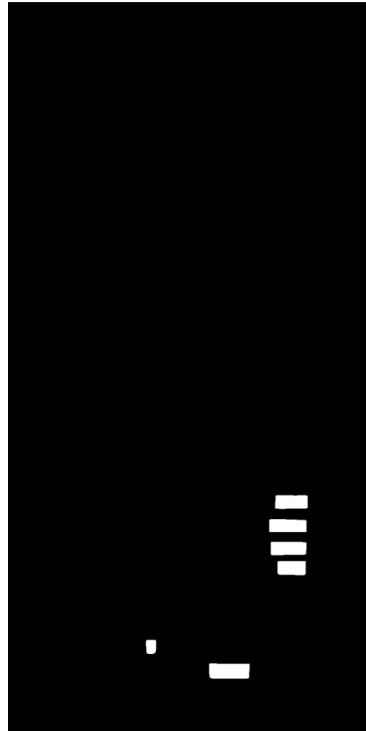
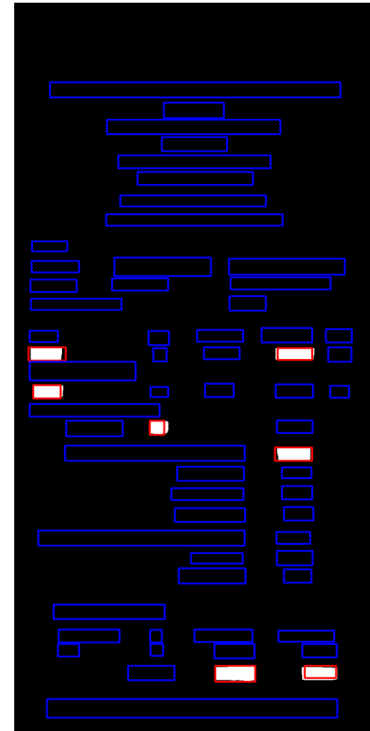
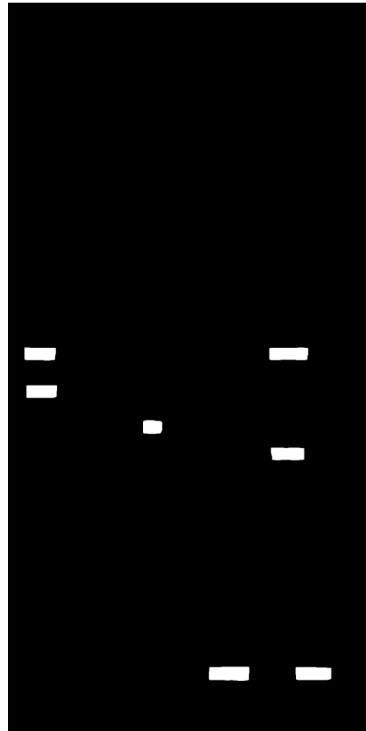


Figure 9. Predictions on the T-SROIE dataset. Blue boxes denote authentic text boxes, red boxes denote tampered text boxes.

Image

Prediction

GroundTruth

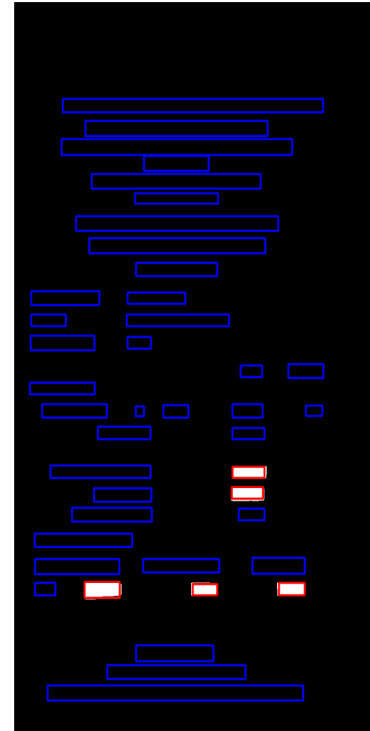
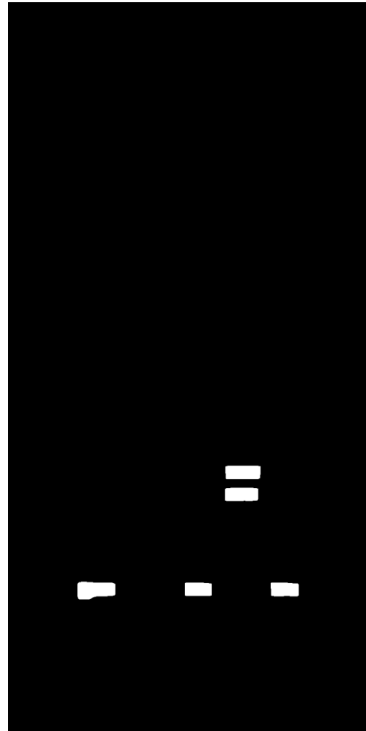
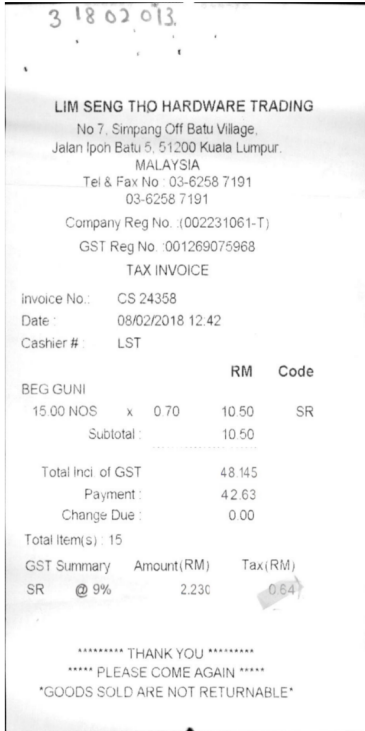
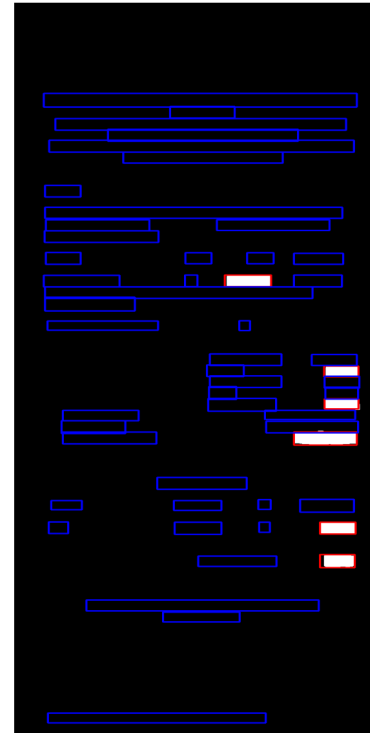


Figure 10. Predictions on the T-SROIE dataset. Blue boxes denote authentic text boxes, red boxes denote tampered text boxes.

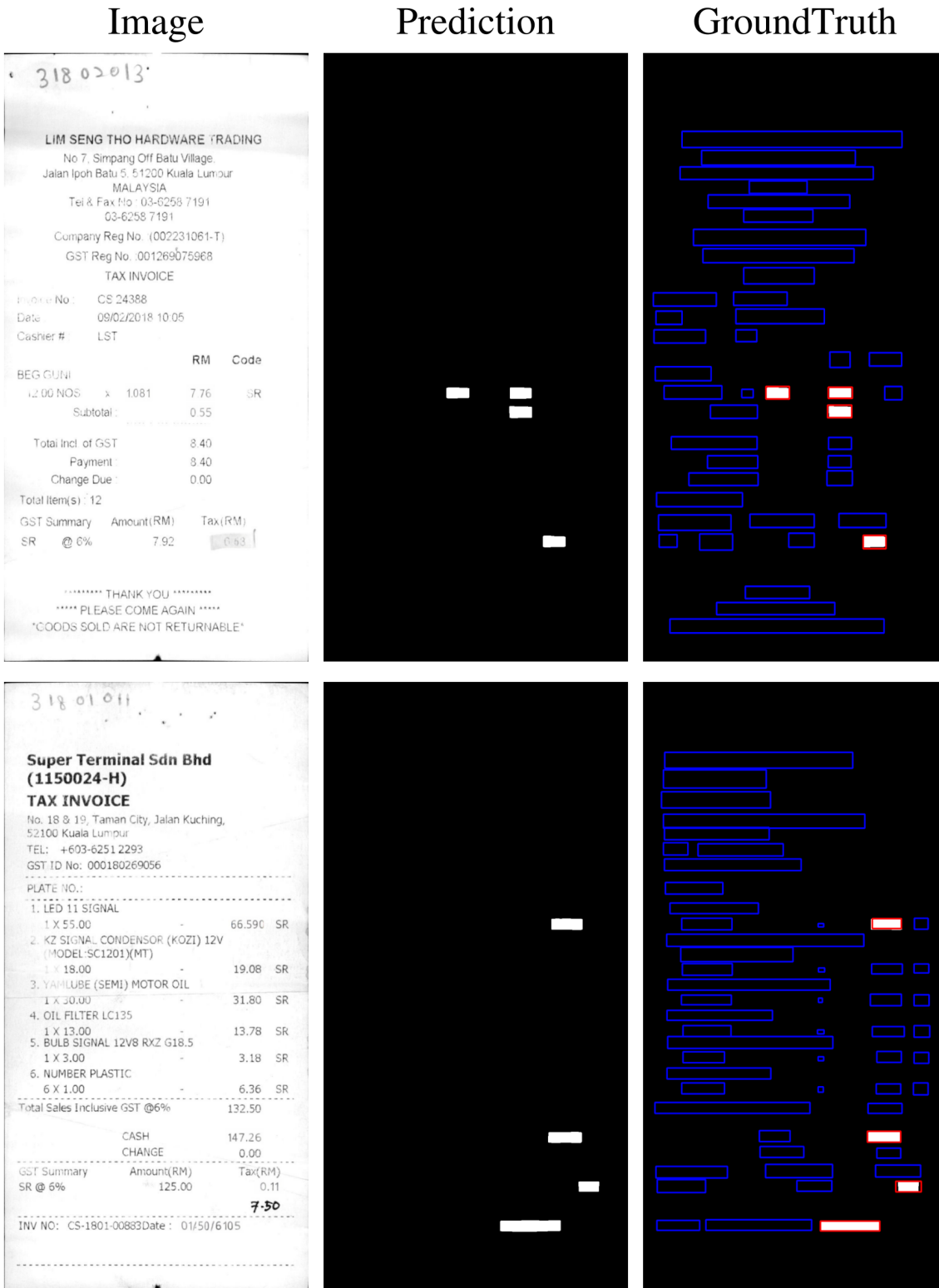


Figure 11. Predictions on the T-SROIE dataset. Blue boxes denote authentic text boxes, red boxes denote tampered text boxes.

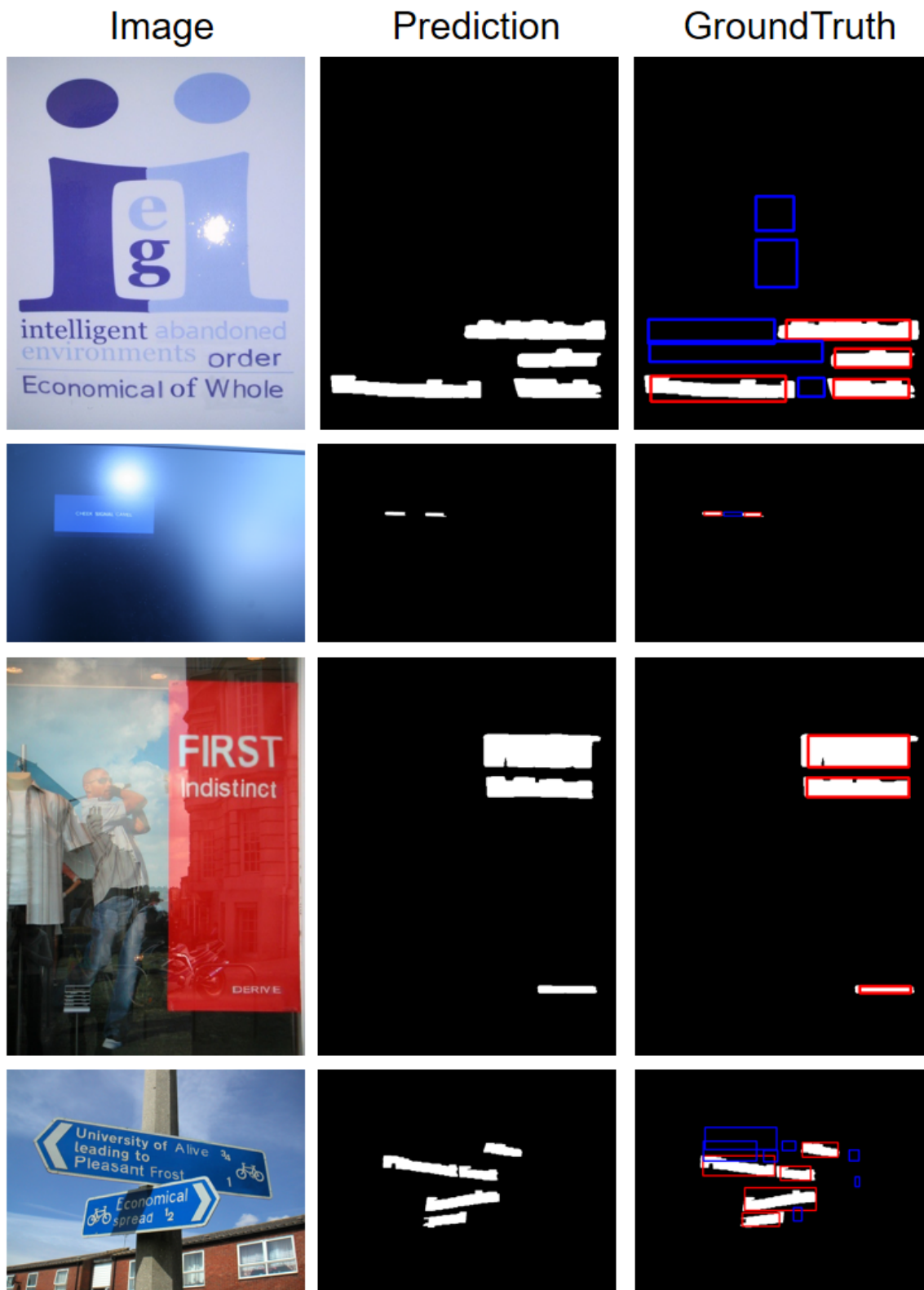


Figure 12. Predictions on the T-IC13 dataset. Blue boxes denote authentic text boxes, red boxes denote tampered text boxes.

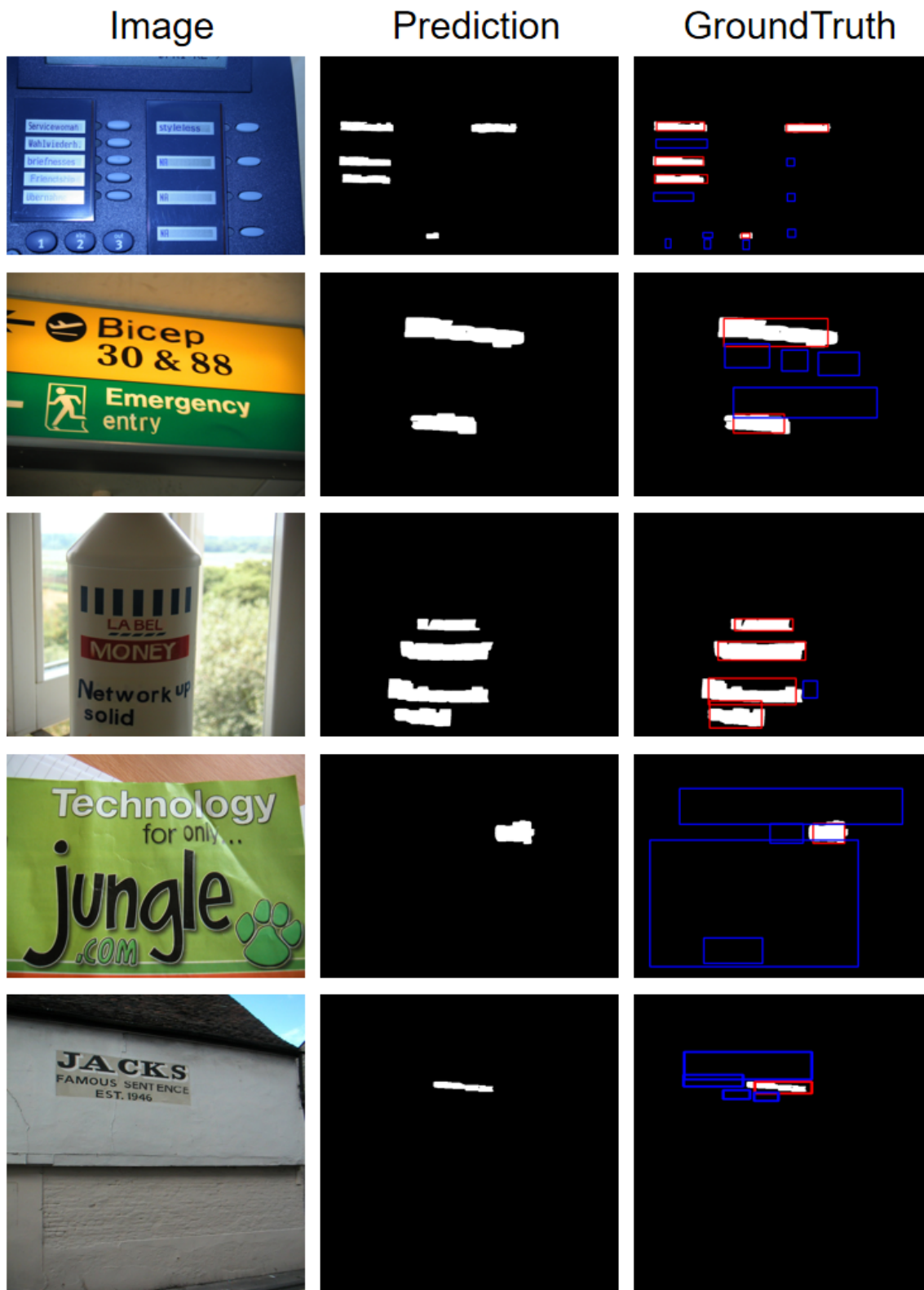


Figure 13. Predictions on the T-IC13 dataset. Blue boxes denote authentic text boxes, red boxes denote tampered text boxes.

References

- [1] Hangbo Bao, Li Dong, Songhao Piao, and Furu Wei. BEit: BERT pre-training of image transformers. In *International Conference on Learning Representations*, 2022.
- [2] Belhassen Bayar and Matthew C Stamm. Constrained convolutional neural networks: A new approach towards general purpose image manipulation detection. *IEEE Transactions on Information Forensics and Security*, 13(11):2691–2706, 2018.
- [3] Chengbo Dong, Xinru Chen, Ruohan Hu, Juan Cao, and Xirong Li. Mvss-net: Multi-view multi-scale supervised networks for image manipulation detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.
- [4] Jessica Fridrich and Jan Kodovsky. Rich models for steganalysis of digital images. *IEEE Transactions on Information Forensics and Security*, 7(3):868–882, 2012.
- [5] Myung-Joon Kwon, Seung-Hun Nam, In-Jae Yu, Heung-Kyu Lee, and Changick Kim. Learning jpeg compression artifacts for image manipulation detection and localization. *International Journal of Computer Vision*, pages 1875–1895, 2022.
- [6] Xiaohong Liu, Yaojie Liu, Jun Chen, and Xiaoming Liu. Pscn-net: Progressive spatio-channel correlation network for image manipulation detection and localization. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022.
- [7] Ze Liu, Han Hu, Yutong Lin, Zhuliang Yao, Zhenda Xie, Yixuan Wei, Jia Ning, Yue Cao, Zheng Zhang, Li Dong, et al. Swin transformer v2: Scaling up capacity and resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12009–12019, 2022.
- [8] Junke Wang, Zuxuan Wu, Jingjing Chen, Xintong Han, Abhinav Shrivastava, Ser-Nam Lim, and Yu-Gang Jiang. Objectformer for image manipulation detection and localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2364–2373, 2022.
- [9] Wenhai Wang, Enze Xie, Xiang Li, Wenbo Hou, Tong Lu, Gang Yu, and Shuai Shao. Shape robust text detection with progressive scale expansion network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [10] Xiaobing Wang, Yingying Jiang, Zhenbo Luo, Cheng-Lin Liu, Hyunsoo Choi, and Sungjin Kim. Arbitrary shape scene text detection with adaptive text region representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [11] Yuxin Wang, Hongtao Xie, Mengting Xing, Jing Wang, Shenggao Zhu, and Yongdong Zhang. Detecting tampered scene text in the wild. In *European Conference on Computer Vision*, pages 215–232. Springer, 2022.
- [12] Yuxin Wang, Boqiang Zhang, Hongtao Xie, and Yongdong Zhang. Tampered text detection via rgb and frequency relationship modeling. *Chinese Journal of Network and Information Security*, 8(3):29–40.
- [13] Yue Wu, Wael AbdAlmageed, and Premkumar Natarajan. Mantra-net: Manipulation tracing network for detection and localization of image forgeries with anomalous features. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9543–9552, 2019.
- [14] Peng Zhou, Xintong Han, Vlad I Morariu, and Larry S Davis. Learning rich features for image manipulation detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1053–1061, 2018.
- [15] Xinyu Zhou, Cong Yao, He Wen, Yuzhi Wang, Shuchang Zhou, Weiran He, and Jiajun Liang. East: An efficient and accurate scene text detector. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.