Document Tampering Detection (Find-it Challenge Analysis)

500 -	CRF-0 33 RU 17000 Te1	DITY LA R DE DE LA D LA R 0 5.46.2	CHELLE SCIERIE OCHELLE 7.02.12	1
	DESCRIPTION		QTE	MONTANT
.500 -	*500G PAIN DE MIE *4X100 ALLUHETTES *200G GALEPEAUTRE *BAGUETTE 250G			2.85€ 1.78€ 1.20€ 1.34€
2000 -	4 ARTICLE(S)	TOTAL A	PAYER	<mark>7.17€</mark>
	CB EMV SANS CO	INTACT E	UR	5.81€
	0006 002	000346	26/06/2017	19:50:32
		MERCI DE A	VOTRE VISI BIENTOT	TE
	L. 0 200 400	600 800	1000 1200	0 1400 160

Fig1. Tampered image with the tampered region highlighted

Overview

- 1. Challenge (Classification & Detection)
- 2. Methods for detection
 - a. Yashas Method
 - b. Fusion(Jpeg Artifacts using DCT)
 - c. CMFD(Copy Move Forgery Detection)
 - d. Splicebuster(Steganalysis)
 - e. Noiseprint(Deep learning based steganalysis)
- 3. Results
 - a. Confusion Matrix
 - b. F1 score
- 4. PaySlip Dataset

Challenge Description

For classification we have around 1240 images out of which 240 images are forged

The dataset was divided into 2 parts. Training and Testing.

	Training	Testing
Forged	130	110
Not Forged	500	500

For training & testing distribution of dataset

Challenge Description

A forger can create various types of forgery. They are described as follows:

- CPI (copy and paste inside the document)
- CPO (copy and paste from another document)
- IMI (creation of a text box imitating the font)
- CUT (deletion of one or more characters/words)
- Other: drawing, copy and paste from web...

TABLE I NUMBER OF DOCUMENTS PER TYPES OF ALTERATION IN EACH CORPUS

(700)	T1Train	T2Train	T1Test	T2Test	Total
CPI	9	34	13	27	83
CPI CPO	1	5		4	10
CPI CPO CUT	1	1		1	3
CPI CPO IMI		1		1	2
CPI CUT	9	18	6	17	50
CPI CUT IMI		4		6	10
CPI IMI	2	6	1	4	13
CPO		7	3	2	12
CPO CUT		1	1	2	4
CUT	3	9	2	8	22
CUT CPI CPO IMI	1				1
IMI	3	11	3	7	24
IMI CUT		1	1		2
Other	1	2		1	4
Total	30	100	30	80	240

The distribution of forgeries across the dataset

Methods proposed

Different techniques provide solutions to different types of tampering.

Yashas' Method[1]: Augment and Adapt, Yashas Anandani & C.V. Jawahar, ICPR 2018

Fusion[2]: Improved dct coefficient analysis for forgery localization in jpeg images, A. Piva, ICASSP, 2011

CMFD[<u>3</u>] : Copy Move Forgery detection Cozzolino, Davide & Verdoliva, Luisa, ICIP 2014

Splicebuster[4]: D. Cozzolino & L. Verdoliva, WIFS 2015

Noiseprint[<u>5</u>]: a CNN-based camera model fingerprint. Cozzolino, Davide & Verdoliva, Luisa. Submitted, Uploaded: Aug 2018.

Augment and Adapt: A Simple Approach to Image Tampering Detection Yashas Annadani, C.V. Jawahar, ICPR [1]

Divide an image into 64x64

If the more than image is 10% of the patch is tampered. We label the patch as tampered.

If k patches are tampered in an image. The image is classified as tampered

K was brought to reduce false positives. Here K can can take values from 1 to 8.

Using only 7 layer CNN doesn't learn a feasible representation of tampering



The distribution of tampered patch size.

Augment and Adapt: A Simple Approach to Image Tampering Detection Yashas Annadani, C.V. Jawahar, ICPR [1]

CDE_CITY IA		
33 RUE DE L	A SCIERIE	
17000 LA	ROCHELLE	
101 : 03.40	.27.02.12	
DESCRIPTION	QTE	NONTAN
CROTTIN CHEVRE X2	₹CROTTIN CH	HEVRE X2 2.73
*GUISNETTE VERT 75CL	3x 7	.95 23.85
*PINEAU DES CHARENTES	CROTTIN CH	HEURE X2 8.654
	A DAVED	20 100
O HRITCLE(3) TOTAL	H FHTEN	
ESPECES	FUR	40 18€
Votre Monnaie	LON	-1.00€
	5 12/01/201	7 10:37:00
0006 002 00005	12/01/201	
0006 002 00005	12/01/201	
0006 002 00005 MERCI	DE VOTRE VI A BIENTOT	SITE

acolio Road CC		
2 *	2,00	4,00 €
*Blang.vola11FE	2,00	4,00 €
*Parm.poulet FE	2 50	E 00.e
*Tartiflette FE	2,30	2,00 €
*Poulet pestoFE	2.00	4.00.0
*Poulet basquai	2,00	4,00 €
2 *	2,00	4,00 €
2 *	2,00	4,00 €
*Chili Carne FE	2.00	4.00 e
Filet poulet 1k	2,00	10,35 €
Escalopes dinde		12,40 €
TOTAL (17	PSOR	53 75 F
IUIAL LII.		JJ,15 C
C		53,75 €
9 miline Mt Hi	T Mt TVA	Nt TTC
V1 5,5 % 50,9	95 2,80	53,75
Calsse Date	Heure T	icket Empl
2 29.03.17	18:59 4	03614 GFIL
PROGRAMME DE FIDE	ELITE - Ins	crivez-vous
et recevez regul	lerement :	reductions,
Rendez-vous sur	WWW.PICARD	ETMOI.FR

111	-	TIC		
	HL	111	IN.	 ·
4154 Pui	i 1boreau			
Rue du 1	18 Juin, ZAI	C Sarah		
Vous ete	es servi pa	r: Beaulie	E4101-1	0144424
APTTCI FS	14 10:09:5.	41	34101-1	FLIR
2531102	bar-le-du	c eau min.	non	0.32
	gazeuse 5	00m1		
	TOTAL			0,32
MODE DE	PAIEMENT			
	Espèces			1,00
	Monnaie (Espèces)		-0,68
DÉTAILS	TVA	TVA	Excl.	Incl.
5,5%		0,02	0,30	0,32
TOTAL		0,02	0,30	0,32
				b
4	154101170	1418093	707455	0
	Merci e	all rei	50	
Écha	inge sous 80	iours a		t.
LOTIC	inge obue of		0	

BOUCHERIE CHARCUTERIE VINCENT 1 C LA ROCHELLE
BALANCE 0 08/04/2017 09:20 VEND 2 CLIENT 5
DIVERS 1,105 kg 18,90 €rkg 20,88 €
1 ENREGISTREMENT TOTAL = 20,98 € MERCI A BIENTOT

White region Composted

Copy-paste(Below TOTAL)

Inpainting

Inpainting

Augmented Images

Augment and Adapt: A Simple Approach to Image Tampering Detection Yashas Annadani, C.V. Jawahar, ICPR [1]



Good training accuracy but poor test results. Distribution of patches selected for training not appropriate Very high false positive rate, most likely model overfitted on training patches. It learned how to detect edges instead of tampering

Improved dct coefficient analysis for forgery localization in jpeg images T. Bianchi, A. D. Rosa, and A. Piva, IEEE International Conference on Acoustics, Speech and Signal Processing, 2011[1]





Findit dataset has Jpeg images, hence this method is used to find CPO(copy paste from other document) and imitation based on forgery.

Results

Acc: 0.959 Precision: 0.804 Recall: 0.994 F1: 0.889

Forgeries missed were CPI(copy paste from Same document)

	Pred True	Pred False
Gr True	193	47
Gr False	1	938

Confusion Matrix

How it works

Let Q_1 and Q_2 quantization steps used in the first and second compression.

To estimate Q_1 , they minimize the difference between 2 histogram of DCT coefficients, one of the image h(x) and one predicted by using Q_1 and Q_2 , p(x).

They have discovered that double jpeg artifacts cause a periodic shift in the DCT coefficient of the doubled quantized region with a period $Q_1 / gcd(Q_1, Q_2)$. Let that function be n(x).

They estimate the histogram as a mixture model of single compressed region (H_1) and double compressed regions (H_0). Let it be

 $p(x; Q_1, \alpha) = \alpha \cdot n(x; Q_1) \cdot \tilde{n}(x \mid H_1) + (1 - \alpha) \cdot \tilde{n}(x \mid H_1)$

 $\tilde{n}(x|H_1)$ is histogram calculated by using only Q_2 , $\tilde{n}(x|H_0)$ is calculated as $\tilde{n}(x|H_0) = n(x, Q_1) \cdot \tilde{n}(x|H_1)$

Estimation of Q_1 using L_2 loss : Q_1 = argmin sum([h(x) - p(x; Q_1, \alpha)])²

After estimating Q_1 the probability of a 8x8 block

$$p = p(\mathcal{H}_0|x_0, \dots, x_{63}) = \frac{\prod_i p(x_i|\mathcal{H}_0)}{\prod_i p(x_i|\mathcal{H}_0) + \prod_i p(x_i|\mathcal{H}_1)}$$

being double compressed (p) is given by joint distribution over all coefficients

Analysis on Fusion

A very basic but highly effective approach, it is able to beat many of the current methods. But works only on JPEG



Correctly predicted

Original Image 1	railuit	- Case		- QZJ
(B) citi			UCT JI	•
CRF-CITY LA ROCHELLE 33 RUE DE LA SCIERIE 17000 LA ROCHELLE Tel : 05.46.27.02.12	2			
DESCRIPTION QTE	NONTANT			
200GR BISC NAPPE CHO	3.25€			
*KRI CREME X 8	3.500			
*SCE PESTO CRF CDM	6.19C			
ARTICLE(S) TOTAL A PAYER	7.93C			
CARTE BANCAIRE ENV USD	4.95€			
0001 004 000504 08/03/2017	17:16:14			
MERCI DE VOTRE VIS A BIENTOT	ITE			
A BIENTOT				

 $\mathsf{Foilure} \mathsf{Coop} (\mathsf{O1} = -\mathsf{O2})$

CMFD(Copy Move Forgery detection) Cozzolino, Davide & Poggi, Giovanni & Verdoliva, Luisa, ICIP 2014[4]

The paper proposes using <u>Patch Match</u> algorithm for detecting tampering.



Original Image	CMFD results
0 SAISONS	
TEL: 09, 53, 76, 56, 30	
www.osaisons.com	
balance: 1 Client 634	
Vendeur 1 Vendeur 1	
fich licket : 76	
kg €/kg €	
DIGNON BID	
FOMME DE TERRE BID	
0,500 2,30 1,15	
1 LITRE	
RADIS	
1 Act x 1,10 1,10	
1014	
IUTAL: 34,63	
book 128 5.50 x: 0.76 s	
total TVA : 0.24 c total UT : 4.39 c	
4 transactica(s)	
Faye 5.65 €	
icado: 1,02 C	
osaisons@gmail.com	
Merci de votre visite	
A & & & & & & & & & & & & & & & & & & &	

Example of CMFD. "3" is copied from 1 location to another

Results

Acc: 0.84 Precision: 0.220 Recall: 0.981 F1: 0.360

Forgeries CPO & Imitation Misclassified

	Pred True	Pred False
Gr True	53	187
Gr False	1	938

CMFD(Copy Move Forgery detection) Cozzolino, Davide & Poggi, Giovanni & Verdoliva, Luisa, ICIP 2014[4]

Used to detect copy-paste from same document

Patch Match algorithm quickly finds correspondences between small square regions (or patches) of an image. It is extremely robust to rotation. Also it is faster than other methods (due to the random nature of the algorithm).

Patch Match works by defining a NNF(nearest neighbour field) f: $R^2 \rightarrow R^2$ for pixel to an offset. Two regions in image with high correspondence can be concluded to be originated from the same patch. (one patch is copied to another location)

Analysis on CMFD

CMFD is very useful in the challenge. Large number of tampered cased were CPI.



Top region of the image was pasted just below it. CMFD was able to detect the forgery and also give the original location of the copied patch

Multiple forgies were attempted. Some at character level, while others were used to hide information. CMFD is able to detect these

Analysis on CMFD

False positive Large smooth regions such as the black background. Lead to false positives.

Original Image N	CMFD results	
Mille Sabu	ords	
Laisse : 1 Date : 17/05/2017 & Venceur :	Ticket : 105 17:41	
PARA-BS ≫ Nombre d'articies :	1x 35,00	
<u>Total à payer</u>	35,00 E.R	
17/4 208 Total +7	5,85 ELR 25,17 ELR	
Ròglé en Carte	32.00 ELR	
20/22 rue du calais - tel 05.45 41 72. httl:/scbordsiaru.	17000 la Rochelle 2 - Egoestork HelleSprail.com et à biant21	

False Negatives Different camera: Patch Match will not work on CPO. Hence we move to splice buster and noiseprint to detect these

Small forgeries: While experimentation we noticed CMFD is unable to catch very small patches. We have look further into patch match to get better results





CMFD was able to detect CPI but not CPO

CMFD was unable to detect small patches

Splicebuster & Steganalysis

D. Cozzolino, G. Poggi and L. Verdoliva, 2015 IEEE International Workshop on Information Forensics and Security (WIFS), 2015 [2]

Feature-based algorithm to detect image splicings without any prior information.

Splicing and host images are assumed to be characterized by different parameters

For splice detection, the high level noise is more important than the image content.

Linear high pass filters is used to create "residue".

Co-occurrence matrix is calculated for residue. This matrix is used as features for the gaussian mixture model/SVM.

This co-occurrence matrix is then used as a feature for GMM.

Using expectation-maximization we can cluster pixels into 2 classes, forged or background.

Analysis of Splice Buster

Splicebuster is good at finding CPO.



Noiseprint: a CNN-based camera model fingerprint Cozzolino, Davide & Verdoliva, Luisa. (2018).[3]

This approach tries to find the tampered regions by extracting the 'noiseprint'(fingerprint of the camera model used).

One of these high frequency noise is PRNU(Photo Response Non Uniformity) caused by the output of the sensors. This noise is not only camera dependant but also depends on the location of the sensor in the camera. A siamese network is trained using different camera models.

The paper proposed a 17 layer FCNN. Given an image, the output is the noiseprint.

Results

Acc: 0.932 Precision: 0.675 Recall: 0.993 F1: 0.803

Forgeries CPI & Imitation Misclassified

	Pred True	Pred False
Gr True	162	78
Gr False	1	938

Results(Classification)

Different methods specialize in different types of methods

We apply all above methods on a tampered image. If any of the above images detect forgery. We classify the image as forged. T1 Train Acc: 0.993 Precision: 0.933 Recall: 0.933 F1: 0.933

	Pred True	Pred False
Gr True	28	2
Gr False	2	468

T1+T2 Train Acc: 0.991 Precision: 0.984 Recall: 0.969 F1: 0.976

	Pred True	Pred False
Gr True	126	4
Gr False	2	468

T1 Test Acc: 0.983 Precision: 0.966 Recall: 0.805 F1: 0.878

	Pred True	Pred False
Gr True	29	1
Gr False	7	462

T1+T2 Test Acc: 0.982 Precision:0.972 Recall: 0.938 F1: 0.954

	Pred True	Pred False
Gr True	107	3
Gr False	7	462

Failure Cases(False Positive)

Splice Buster Heatmap







Splice Buster Heatmap



DCT IPEG





Original Image 0

CMFD results Splice Buster Heatmap



Splicebuster Triggered by torn region

Splicebuster Triggered by torn region

CMFD predicted white space as copied

Failure Cases(False Negative)

Original Image 1



Splice Buster Heatmap





CMFD results





Splice Buster Heatmap





CMFD results



DCT JPEG





Conclusion

They have achieved reasonable accuracy on find-it challenge from ICPR 2017.

Payslip Dataset

A Dataset for Forgery Detection and Spotting in Document Images

Nicolas Sidere, Francisco Cruz, Mickal Coustaty and Jean-Marc OgierL3i Laboratory, University of La Rochelle

2017 Seventh International Conference on Emerging Security Technologies (EST)

Payslip Dataset

Dataset of Genuine Documents : Synthetic real-like Payslips

- 200 documents
- 5 fonts, 4 text sizes
- 477 Forged Documents
- 241 Genuine Documents

		# documents	# forgeries
Imitation	Case 1	22	298
	Case 2	37	493
C/P intra		224	1627
C/P Inter	Case 1	50	931
	Case 2	44	811
	Case 3	100	1798
Total		477	5958

Types of forgeries

Payslip Dataset(Examples)

	BULLET	'IN DE PAIE			
	EMP	LOYEUR			
Nom :	PLASTIVALOIRE				
Adresse :		LD LES	S VALLEES - ZI N	ORD	
CP et Ville :		3	7130 LANGEAIS		
Numéro APE :			2229A		
Numéro SIRET :		e	4480016100015		
	SA	LARIE			
Nom et Prénom :		(GUERIN Frederic		
Adresse :		28 Avenu	e de l'Amiral Gan	teaume	
CP et Ville :		854	58 LA BOURBOU	E	
Numéro SS :		1	59083084331962		
Date Entrée :			22/04/01		
Emploi :		Ouvriers of	qualifiés de type ir	ndustriel	
Salaire de base	151,67	15,44 €	2 341,78 €		
HS à 25%	11	19,30 €	212,30 €		
SALAIRE BRUT		L	2 554,08 €		
COTICATIONS		PART SALARIALE		PART PATRONALE	
COTISATIONS	Base	Taux	Montant	Taux	Montant
CSG non déductible	2 477,46 €	2,40%	59,46 €		
CRDS non déductible	2 477,46 €	0,50%	12,39 €		
Csg déductible	2 477,46 €	5,10%	126,35 €		
Sécurité sociale					
Assurance maladie	2 554,08 €	0,75%	19,16 €	12,80%	326,92
Assurance veuvage	2 554,08 €	0,10%	2,55 €		
Assurance vieillesse					
AV déplafonée	2 554,08 €	6,55%	167,29 €	1,60%	40,87
AV plafonnée	2 554,08 €			8,20%	209,43
Accidents du travail	2 554,08 €			7,30%	186,45
Allocation familiales	2 554,08 €			5,40%	137,92
Aide au logement					
AL déplafonée	2 554,08 €			0,40%	10,22
AL plafonnée	2 554,08 €			0,10%	2,55
ASSEDIC					
Ass. chômage tranche A	2 554,08 €	2,40%	61,30 €	4,00%	102.16
Ass. chômage tranche B	0,00 €	2,40%	0,00 €	4,00%	0,00
TOTAL des cotisations			448,50 €		1 016,53
Payé par chèque de banque			Net à payer	2 105	,59 €

Imitation

Salahan kana kana kana kana kana kana kana	BULLI	ETIN DE PAIE	ile en sen de la sen Internet de la sen de	and Galler	
Nom :		F	LASTIVALOIRE		
Adresse :	ID LES VALLEES - ZI NORD				
CR at Villa :		20 220		0110	
OF BUTTING .		0	100 EANGEAIG		
Numero APE :			2229A		
Numero SIRE1 :		6	4480016100015		
New of Deferrer		ALARIE			
Nom et Prenom :		00.4	OERIN Prederic		
Adresse :		20 Avenue	e de l'Amirai Gant	eaume	
CP et ville :		3/1	TO VILLEDOMER		
Numero SS :		10	190630643319 15		
Date Entree :		Our stars of	22/04/01	decount of	
Emploi :		Odwhers d	ualities de type in	dustriel	
Salaire de base	151,67	15,44 €	2 341,78 €		
-IS à 25%	11	19,30 €	212,30 €		
SALAIRE BRUT		L	2 554,08 €		
007/017/01/0		PART SALARIALE		PART PATRONALE	
COTISATIONS	Base	Taux	Montant	Taux	Montant
CSG non déductible	2 477,46 €	2,40%	59,46 €		
CRDS non déductible	2 477,46 €	0,50%	12,39 €		
Csg déductible	2 477,46 €	5,10%	126,35 €		
Sécurité sociale					
Assurance maladie	2 554,08 €	0,75%	19,16 €	12,80%	326,92 €
Assurance veuvage	2 554,08 €	0,10%	2,55 €		
Assurance vieillesse					
AV déplafonée	2 554,08 €	6,55%	167,29 €	1,60%	40,87€
AV plafonnée	2 254,08 €			8,20%	209,43 €
Accidents du travail	2 554,08 €			7,30%	186,45€
Vlocation familiales	2 554,08 €			5,40%	137,92 €
Aide au logement					
AL déplafonée	2 554,08 €			0,40%	10,22€
AL plafonnée	2 554,08 €			0,10%	2,55€
ASSEDIC					
Ass. chômage tranche A	2 554,08 €	2,40%	61,30 €	4,00%	102,16€
Ass. chômage tranche B	0,00 €	2,40%	0,00 €	4,00%	0,00 €
OTAL des cotisations			448,50 €		1 016,53 €
Pavé par virement bancaire	Г		Net à paver	2 554	08€
and a second sec					

Copy Paste same document

	BULLE	TIN DE PAIE			
	EMF	PLOYEUR			
Nom :		PLASTIVALOIRE			
Adresse :		LD LE	S VALLEES - ZI N	ORD	
CP et Ville :			37130 LANGEAIS		
Numéro APE :			2229A		
Numéro SIBET			64480016100015		
	s				
Nom et Prénom :	0.	10000	GUERIN Clemer	il.	
Adresse :		28 Aven	ue de l'Amiral Gan	teaume	
CP et Ville :		3	110 VILLEDOMER	2	
Numéro SS :		2	59083084331962		
Date Entrée :			22/04/01		
Emploi :		Ouvriers	qualifiés de type ir	ndustriel	
Calaire da hana	454.07	15.11.6	0.044.70.6		
Salaire de base	151,67	15,44 €	2 341,78 €		
SALAIRE BRUT		15,30 €	2 554 08 6		
SAEAIRE BRUT		l	2 334,08 €		
COTISATIONS		PART S	ALARIALE	PART PAT	RONALE
	Base	Taux	Montant	Taux	Montant
CSG non déductible	2 477,46 €	2,40%	59,46 €		
CRDS non deductible	2 477,46 €	0,50%	12,39 €		
Csg deductible	2477,46€	5,10%	126,35 €		
Sécurité sociale					
Assurance maladie	2 554,08 €	0,75%	19,16 €	12,80%	326,92 €
Assurance veuvage	2 554,08 €	0,10%	2,55 €		
Assurance vieillesse					
AV déplafonée	2 554,08 €	6,55%	167,29 €	1,60%	40,87 €
AV plafonnée	2 554,08 €			8,20%	209,43 €
Accidents du travail	2 554,08 €			7,30%	186,45 €
Allocation familiales	2 554,08 €			5,40%	137,92 €
Aida au lananan					
Alde au logement	2 554 09 6			0.40%	10 22 6
AL deplatonee	2 554 08 6			0,40%	255 6
AL platonites	2 004,00 0			0,1070	2,00 0
ASSEDIC					
Ass. chômage tranche A	2 554,08 €	2,40%	61,30 €	4,00%	102,16 €
Ass. chômage tranche B	0,00 €	2,40%	0,00 €	4,00%	0,00 €
TOTAL des cotisations			448,50 €		1 016,53 €
Payé par virement bancaire	Г	-	Net à payer	2 105	5,59 €
le: 25/06/13	ŀ		Net imposable	2 165	.05€
	L				

Copy Paste another document

References

1. Augment & Adapt

Yashas Anandani, C.V. Jawahar ICPR

2. <u>CMFD(Copy-Move Forgery Detection)</u>

Cozzolino, Davide & Poggi, Giovanni & Verdoliva, Luisa. (2015). Copy-move forgery detection based on PatchMatch. 2014 IEEE International Conference on Image Processing, ICIP 2014. 5312-5316. 10.1109/ICIP.2014.7026075.

3. Fusion

T. Bianchi, A. D. Rosa, and A. Piva, "Improved dct coefficient analysis for forgery localization in jpeg images," in IEEE International Conference on Acoustics, Speech and Signal Processing, 2011, pp. 2444–2447.

4. <u>Splicebuster</u>

D. Cozzolino, G. Poggi and L. Verdoliva, "Splicebuster: A new blind image splicing detector," 2015 IEEE International Workshop on Information Forensics and Security (WIFS), Rome, 2015, pp. 1-6. doi: 10.1109/WIFS.2015.7368565

5. Noiseprint(not published)

Cozzolino, Davide & Verdoliva, Luisa. (2018). Noiseprint: a CNN-based camera model fingerprint. Url: https://arxiv.org/abs/1808.08396