Supplementary Material BioFors: A Large Biomedical Image Forensics Dataset

Ekraam Sabir, Soumyaroop Nandi, Wael AbdAlmageed, Prem Natarajan USC Information Sciences Institute, Marina del Rey, CA, USA

{esabir, soumyarn, wamageed, pnataraj}@isi.edu

1. Image Collection and Statistics

We described the image collection procedure from compound biomedical figures in the main paper. The overall procedure involved automatic extraction of compound figures from documents followed by manual cropping of images. Figure 1 shows a sample compound figure and its decomposition. Furthermore, there is a significant variation in the number of images extracted from each document. Figure 2 shows the frequency of images extracted per document. Finally, images in BioFors have a wide range of dimensions. Figure 3 shows a scatter-plot of BioFors image dimensions as compared to two other natural-image forensic datasets, Columbia [7] and COVERAGE [10].



Figure 1. We crop compound biomedical figures in two stages: 1) crop sub-figures and 2) crop images from sub-figures. Synthetic images such as charts and plots are filtered.

2. Orientation

Duplicated regions in BioFors may have an orientation difference. These differences may occur between duplicated regions across two images (external duplication detection (EDD) task) or within an image (internal duplication detection (IDD) task). We found five major categories of differing orientation: 0° , 90° , 180° , horizontal and vertical flip. Figure 4 shows the frequency of each orientation between duplicated regions.

3. Baseline for CSTD

Existing forgery detection methods are not specifically designed to detect cuts/sharp transitions in images. The ab-



Figure 2. Distribution of images extracted from documents. The distribution peaks at 25 images from most documents. The right-most entry has 219 images from one document.



Figure 3. BioFors images have much higher variation in dimension compared to two popular image forensic datasets.

sence of diverse detection methods prompted us to train a simple convolutional neural network (CNN) baseline trained on synthetic manipulations in pristine blot/gel images from BioFors. Figure 5 shows the CNN architecture.



Figure 4. Frequency of differing orientations between duplicated regions in EDD and IDD tasks.



Figure 5. Our baseline CNN architecture. Padding ensures the same height and width for input image and output mask.

4. Synthetic Data Generation

Image forensic datasets usually do not have sufficient samples to train deep learning models. Previous works [11, 12] created suitable synthetic manipulations in natural images for model pre-training. The synthetic manipulations were created by extracting objects from images and pasting them in the target image with limited data augmentation such as rotation and scale change. Similar to previous works, we created suitable synthetic manipulations in biomedical images corresponding to each task for training and validation. However, biomedical images do not have well defined objects and boundaries and manipulated regions are created with rectangular patches. Manipulation process for each task is discussed ahead.

External Duplication Detection (EDD): Corresponding to the two possible sources of external duplication, we create manipulations by 1) Cropping two images with overlapping regions from a source image and 2) Splicing i.e. copy-pasting rectangular patches from source to target image. Manipulations of both types are shown in Figure 6. We generate pristine, spliced and overlapped samples in a 1:1:1 ratio. Images extracted with overlapping regions are resized to 256x256 image dimensions.

Internal Duplication Detection (IDD): Internal duplications are created with a copy-move operation within an image. Rectangular patches of random dimensions are copypasted within the same image. Figure 7 shows samples.



Figure 6. Synthetic manipulations in image pairs created using (a) overlapping image regions and (b) spliced image patches.



Figure 7. Internal duplications created with copy-move operations.



Figure 8. Synthetic cut/sharp transitions created in blot/gel images.

Cut/Sharp-Transition Detection (CSTD): We simulated synthetic manipulations by randomly splitting an image along two horizontal or vertical lines in the image and rejoining the image. The line of rejoining represents a synthetic cut or sharp transition and is used for training. Figure 8 shows synthetic CSTD manipulations.

5. Experiment Details

We list the hyper-parameter and finetuning details of baselines corresponding to each task.

Keypoint-Descriptor: We implemented classic image matching algorithm using keypoint-descriptor based methods such as SIFT, ORB and BRIEF. Keypoints are matched using kd-tree and a consistent homography is found using RANSAC to remove outlier matches. A rectangular bounding box is created around the furthest matched keypoints. We keep a threshold of minimum 10 matched keypoints to consider an image pair to be manipulated.

DenseField: We evaluated DenseField [4] on IDD task with three reported transforms - zernike moments (ZM), polar cosine transform (PCT) and fourier-mellin transform (FMT). ZM and PCT are evaluated with polar sampling grid. Feature length for ZM, PCT and FMT are 12, 10 and 25. Since DenseField is a copy-move detection algorithm,

	Microscopy		Blot/Gel		Macroscopy		FACS		Combined	
Method	Image	Pixel	Image	Pixel	Image	Pixel	Image	Pixel	Image	Pixel
SIFT [6]	8.48%	5.82%	9.37%	11.74%	6.98%	10.52%	6.09%	2.32%	8.18%	5.83%
ORB [8]	30.48%	28.56%	5.97%	12.45%	9.87%	19.34%	22.53%	8.86%	20.66%	21.47%
BRIEF [2]	27.42%	25.59%	3.74%	9.74%	13.07%	16.22%	20.09%	9.33%	18.22%	18.58%
DF - ZM [4] DMVN [11]	42.00% 16.40%	42.08% 31.54%	15.42% 18.61%	19.17% 42.07%	27.48% 10.29%	25.82% 21.78%	54.17% 8.94%	50.24% 18.85%	27.06% 16.31%	32.46% 27.55%

Table 1. F_1 scores for external duplication detection (EDD) task. The scores correspond to the experiments in main document.

	Microscopy		Blot/Gel		Macroscopy		Combined	
Method	Image	Pixel	Image	Pixel	Image	Pixel	Image	Pixel
DF - ZM [4]	74.07%	12.04%	52.46%	43.99%	53.33%	38.95%	56.10%	30.48%
DF - PCT [4]	74.07%	12.45%	51.97%	46.02%	70.59%	40.03%	57.31%	31.90%
DF - FMT [4]	58.33%	9.94%	50.00%	38.53%	49.18%	39.84%	50.62%	27.74%
DCT [5]	16.33%	3.50%	28.57%	17.06%	23.53%	15.53%	23.39%	10.35%
DWT [1]	25.97%	7.53%	40.00%	25.73%	63.16%	22.16%	37.04%	16.23%
Zernike [9]	17.95%	3.94%	34.78%	18.45%	42.86%	13.49%	28.99%	11.13%
BusterNet [12]	13.99%	13.38%	27.33%	4.57%	24.14%	14.94%	30.25%	6.04%

Table 2. $\overline{F_1}$ scores for external duplication detection (EDD) task. The scores correspond to the experiments in main document.

it expects a single image input. For evaluation on EDD task, we concatenated image pairs along the column axis to form a single input and used the best reported transform (ZM).

DMVN: The model is finetuned on synthetic data using adam optimizer with a learning rate of 1e-5, batch size 16 and binary crossentropy loss. The model has two outputs: 1) binary mask prediction and 2) image level forgery classification. We found fine-tuning to be unstable for joint training of both outputs. We set image classification loss weight to zero, tuning only the pixel loss. For image level classification we used the protocol similar to BusterNet [12]. Postprocessing by removing stray pixels with less than 10% of image area improved image classification performance.

BusterNet: We finetune BusterNet [12] on synthetic data using adam optimizer with a learning rate of 1e-5, batch size of 32 and categorical-crossentropy loss. BusterNet predicts a 3-channel mask to identify source, target and pristine pixels. Since we do not need to discriminate between source and target pixels, we consider both classes as manipulated.

Block Feature Matching: Discrete cosine transform (DCT), discrete wavelet transform (DWT) and Zernike features are matched with a block size of 16 pixels and minimum euclidean distance of 50 pixels between two matched blocks using the CMFD algorithm reported in [3].

ManTraNet: We finetuned the model using adam optimizer with learning rate of 1e-3, batch size of 32 with gradient accumulation and binary-crossentropy loss. Since, cuts and transitions have thin pixel slices which can be distorted by resizing, we use images with original dimension.

Baseline CNN: We trained the CNN using adam optimizer with learning rate of 1e-3, mean squared error loss and batch size of 10.

6. Alternate Metric: F_1 score

Table 1 and 2 report F_1 scores for EDD and IDD tasks respectively. The experiments are identical to those reported in the main document, but with F_1 scores.

7. Sample Predictions

Prediction samples for EDD, IDD and CSTD tasks respectively in Figures 9,10 and 11. For EDD we show predictions from ORB [8] and DMVN [11]. Samples for IDD include DCT [5], DenseField [4], DWT [1], Zernike [9] and BusterNet [12] baselines. Similarly, CSTD predictions are from ManTraNet [13] and our cnn baseline.

8. Ethical Considerations

We have used documents from PLOS to curate BioFors, since it is open access and can be used for further research including modification and distribution. However, the purpose of BioFors is to foster the development of algorithms to flag potential manipulations in scientific images. Bio-Fors is explicitly not intended to malign or allege scientific misconduct against authors whose documents have been used. To this end, there are two precautions (1) We have anonymized images by withholding information about the source publications. Since scientific images have abstract patterns, matching documents from the web with BioFors images is a significant hindrance to the identification of source documents. (2) Use of pristine documents and documents with extenuating circumstances such as citation for duplication and justification. As a result, inclusion of a document in BioFors does not assure scientific misconduct.



Figure 9. Rows of image pairs and corresponding predicted masks. The text in sample (d) misleads the prediction from both models.



Figure 10. Rows of images and forgery detection predictions. There is significant variation in prediction across models. Rotated predictions in sample (a) are not identified by any model.



Figure 11. Predictions from ManTraNet and baseline CNN. It is evident that current forensic models are not suitable for the CSTD task.

References

- M. Bashar, K. Noda, N. Ohnishi, and K. Mori. Exploring duplicated regions in natural images. *IEEE Transactions on Image Processing*, pages 1–1, 2010. 3
- Michael Calonder, Vincent Lepetit, Christoph Strecha, and Pascal Fua. Brief: Binary robust independent elementary features. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 778–792. Springer, 2010.
 3
- [3] Vincent Christlein, Christian Riess, Johannes Jordan, Corinna Riess, and Elli Angelopoulou. An evaluation of popular copy-move forgery detection approaches. *IEEE Transactions on Information Forensics and Security*, 7(6):1841– 1854, 2012. 3
- [4] Davide Cozzolino, Giovanni Poggi, and Luisa Verdoliva. Efficient dense-field copy-move forgery detection. *IEEE Transactions on Information Forensics and Security*, 10(11):2284–2297, 2015. 2, 3
- [5] A Jessica Fridrich, B David Soukal, and A Jan Lukáš. Detection of copy-move forgery in digital images. In *in Proceedings of Digital Forensic Research Workshop*. Citeseer, 2003.
- [6] David G Lowe. Distinctive image features from scaleinvariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004. 3
- [7] Tian-Tsong Ng, Jessie Hsu, and Shih-Fu Chang. Columbia image splicing detection evaluation dataset. DVMM lab. Columbia Univ CalPhotos Digit Libr, 2009. 1

- [8] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. Orb: An efficient alternative to sift or surf. In 2011 International Conference on Computer Vision, pages 2564– 2571. Ieee, 2011. 3
- [9] Seung-Jin Ryu, Min-Jeong Lee, and Heung-Kyu Lee. Detection of copy-rotate-move forgery using zernike moments. In *Proceedings of the 12th International Conference on Information Hiding*, pages 51–65, 2010. 3
- [10] Bihan Wen, Ye Zhu, Ramanathan Subramanian, Tian-Tsong Ng, Xuanjing Shen, and Stefan Winkler. Coverage—a novel database for copy-move forgery detection. In 2016 IEEE International Conference on Image Processing (ICIP), pages 161–165. IEEE, 2016. 1
- [11] Yue Wu, Wael Abd-Almageed, and Prem Natarajan. Deep matching and validation network: An end-to-end solution to constrained image splicing localization and detection. In *Proceedings of the 25th ACM international conference on Multimedia*, pages 1480–1502, 2017. 2, 3
- [12] Yue Wu, Wael Abd-Almageed, and Prem Natarajan. Busternet: Detecting copy-move image forgery with source/target localization. In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018. 2, 3
- [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 (CVPR), June 2019. 3