

# GreatSplicing: A Semantically Rich Splicing Dataset

\*Xiuli Bi, Jiaming Liang

Chongqing University of Posts and Telecommunications  
bixl@cqupt.edu.cn, chinaliangjm@gmail.com

October 23, 2023

## Abstract

In existing splicing forgery datasets, the insufficient semantic varieties of spliced regions cause a problem that trained detection models overfit semantic features rather than splicing traces. Meanwhile, because of the absence of a reasonable dataset, different detection methods proposed cannot reach a consensus on experimental settings. To address these urgent issues, GreatSplicing, a manually created splicing dataset with a considerable amount and high quality, is proposed in this paper. GreatSplicing comprises 5,000 spliced images and covers spliced regions with 335 distinct semantic categories, allowing neural networks to grasp splicing traces better. Extensive experiments demonstrate that models trained on GreatSplicing exhibit minimal misidentification rates and superior cross-dataset detection capabilities compared to existing datasets. Furthermore, GreatSplicing is available for all research purposes and can be downloaded from [www.greatsplicing.net](http://www.greatsplicing.net).

## Introduction

In recent years, image authentication has become more and more critical. Besides Deepfake and other generated images, traditional image forgery, altering the local semantic information of an image while maintaining its overall authenticity and integrity, is still always exploited by malicious users. Splicing forgery is the most common local forgery (Bi et al. 2023), which is defined (Ng and Chang 2004) as a cut-and-paste of image regions from one image onto another image to create a spliced image. Splicing forgery detection (Ahmad and Khurshed 2021) is to localize the spliced regions and disclose them via a binary map.

Thanks to the powerful capabilities of neural networks, masses of validated network architectures for splicing detection have been proposed. However, subject to the limitations of existing datasets, these methods cannot be fully exploited and fail to perform well in cross-dataset cases. Meanwhile, because of the absence of a dataset that can reflect the performance of these methods faithfully, experimental settings in different papers often are based upon independently synthesized datasets for training. During the training-testing process, specific techniques like image augmentation (Shorten and Khoshgoufar 2019) and fine-tuning are employed. It raises concerns about fairness in this research field.

Therefore, to address these problems caused by the absence of well-structured splicing datasets, GreatSplicing is introduced in this paper. Based on BossBase, GreatSplicing comprises 5,000 spliced images manually created using Adobe Photoshop, without introducing any automated

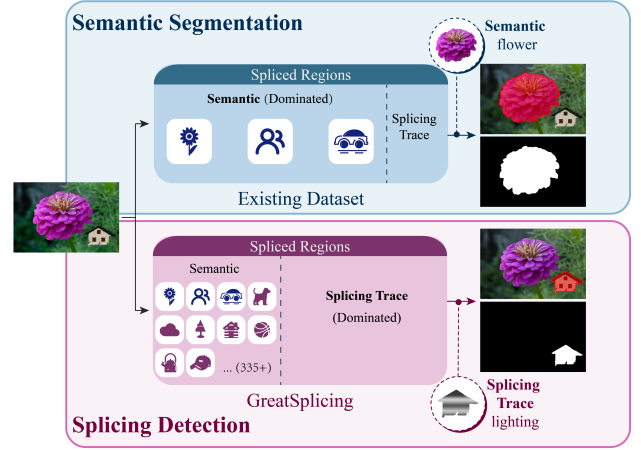


Figure 1: The distinctions between GreatSplicing and existing splicing datasets. The spliced regions in GreatSplicing exhibit a wider variety of semantic categories, which enhances the network’s learning of splicing traces.

mechanisms. The spliced regions in GreatSplicing exhibit diverse semantic categories, allowing neural networks to learn splicing traces more effectively. This dataset features high resolution and follows a standardized production process, ensuring the faithful preservation of splicing traces. The distinction between GreatSplicing and existing datasets is illustrated in Figure 1.

The contributions of this paper can be summarized as follows:

- A significant defect of insufficient semantic diversity in existing splicing forgery datasets is revealed in this paper. Extensive experiments demonstrate that this limitation can result in severe misidentification.
- A semantically rich splicing dataset, GreatSplicing, is proposed to address the challenge. Its high resolution and verisimilitude make neural networks get splicing traces more effectively.
- Based on GreatSplicing, five recommended experimental settings are proposed to promote fairness in comparison.

The rest of the paper is organized as follows: In Section 2, a detailed introduction to existing splicing forgery datasets will be provided. Additionally, a concise overview of BossBase will be presented. In Section 3, we will present the production process of GreatSplicing and offer five recom-

mended experimental settings based on GreatSplicing. Subsequently, Section 4 will present extensive experimental results to verify the superiority of GreatSplicing. Finally, Section 5 will summarize the whole paper.

## Related Works

This section will briefly introduce existing splicing forgery datasets and their limitations. Besides, a concise description of the authentic image dataset BossBase is also stated.

### Existing Splicing Datasets

**CASIA(Dong, Wang, and Tan 2013)** consists of V1.0 and V2.0, comprising 8,000 authentic color images and 2,300 spliced images. The authentic images used in creating CASIA encompass nine different scenes.

**DEFACTO(Mahfoudi et al. 2019)** is a large-scale image tampering dataset comprising 105,000 spliced images. However, the spliced regions cover only nine semantic categories: airplane, bird, clock, frisbee, ball, traffic light, and stop sign.

**NIST16(Guan et al. 2019)** comprises 288 spliced images, with over half of the images featuring combinations of basic shapes, such as spheres, triangles, and cubes.

**FantasticReality(Kniaz, Knyaz, and Remondino 2019)** consists of 16,592 authentic images and 14,546 spliced images. These authentic images were collected undisclosed and later employed to create the spliced images. Besides, the spliced regions in FantasticReality encompass only ten semantic categories: person, car, truck, van, bus, building, cat, dog, tram, and boat.

**Columbia(Hsu and Chang 2006)** only contains 180 spliced images, and all spliced regions in these images are of random-shapes.

**IMD2020(Novozamsky, Mahdian, and Saic 2020)** offers 2,010 spliced images and their corresponding authentic images. However, these spliced images are derived from the Internet and created by random individuals. Additionally, these spliced images have undergone post-processing, like JPEG compression. Thus, there are substantial concerns regarding the rigorousness of production standards.

### BossBase

The GreatSplicing dataset proposed in this paper is based on BossBase(Bas, Filler, and Pevný 2011) dataset. BossBase is a widely used dataset in image steganalysis, comprising 10,000 images taken with digital cameras and stored in the original camera image format. This dataset encompasses images from 7 distinct camera models, namely Canon EOS 400D, Canon EOS 40D, Canon EOS 7D, Canon EOS DIGITAL REBEL XSi, PENTAX K20D, NIKON D70, and M9 Digital Camera, covering diverse scenes from daily life. The images are saved in multiple formats, including CR2, PEF, DNG, and NEF, and exhibit high resolution and considerable size.

BossBase was selected for the following reasons:

1. This dataset ensures image authenticity because it originates from cameras;

Order	Datasets	Capacity	Semantics	O	R
1	CASIA V2.0	1, 849	<30	•	
2	FantasticReality	14, 546	= 10	•	
3	DEFACTO	105, 000	= 7	•	
4	NIST16	288	<20	•	
5	Columbia	180	-		•
6	IMD2020	2, 010	Various	•	
7	GreatSplicing	5, 000	=335	•	•

(a) Summary of existing public splicing datasets. Semantics indicates the number of semantic categories covered by the spliced regions in the dataset. O and R denote the inclusion of object-aware and shape-random spliced images, respectively.

Time	Paper	Dataset	Syn	Fine	Aug	Auth
2022	MVSS-Net	1, 3, 4, 5, 6				•
2022	Objectformer	1, 4, 5, 6	•	•	•	•
2022	PSCC-Net	1, 4, 5, 6	•		•	
2021	RTAG	1, 2, 4, 5				
2020	SPAN	1, 4, 5	•	•		
2019	RRU-Net	1, 5			•	
2019	Mantra-Net	1, 5, 6	•			•
2019	MAGritte	1, 2, 5				•
2019	HLED	4	•	•	•	
2018	RGB-Noise	1, 4, 5	•	•		
2017	ESS-Net	4			•	

(b) Summary of different experimental settings in the past six years. Dataset indicates the datasets used in the paper, with the numbers corresponding to the dataset indices in Table (a). Syn denotes the use of synthetic datasets for training. Fine refers to fine-tuning. Aug and Auth represent the inclusion of augmented and authentic images during training.

Table 1: Summary of datasets and experimental settings in the current splicing detection field.

2. BossBase encompasses diverse real-life scenarios, facilitating the creation of spliced images that align better with practical applications;
3. The abundant semantic information present in these real-life scenario images provides an advantage for generating a splicing detection dataset with rich semantic content;
4. The high resolution and large size of the images in BossBase could effectively preserve traces of splicing manipulations.

## GreatSplicing

### Limitations of Existing Splicing Forgery Datasets

While various splicing forgery datasets are currently available, less rigor in their creation processes has introduced several significant deficiencies. Consequently, existing neural networks for splicing detection cannot realize their potential fully.

CASIA includes numerous identical spliced regions, as shown in Figure 2(a). Spliced regions in the first row are the same people, and in the second row are the same sky. Similar

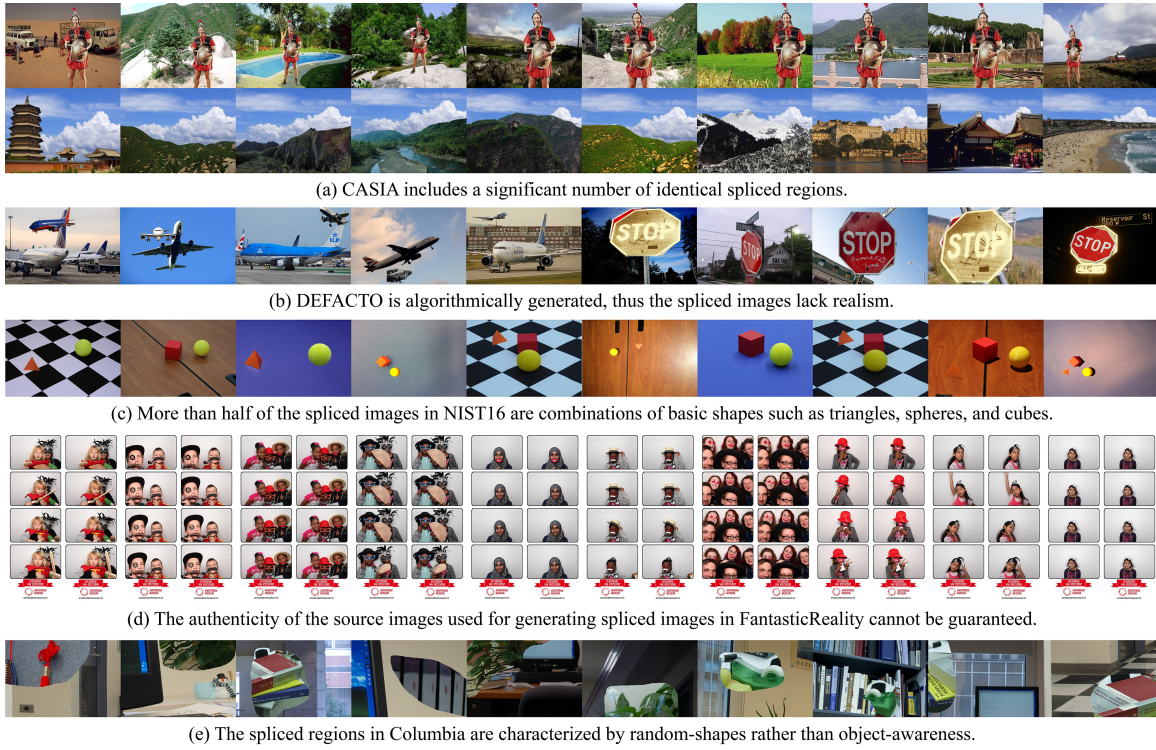


Figure 2: Examples of limitations in existing splicing datasets.

instances are considerable in CASIA, leading networks to overfit these identical semantic information, thus impeding the learning of splicing traces.

DEFACTO is algorithmically generated based on MSCOCO(Lin et al. 2014), as illustrated in Figure 2(b). These spliced images lack verisimilitude and do not align with our real world, and thus, networks trained on DEFACTO cannot attain applicability.

More than half of the spliced regions in NIST16 consist of basic geometric shapes, as depicted in Figure 2(c). Networks trained on NIST16 tend to overfit these simple shapes rather than splicing traces.

The authenticity of the authentic images employed for generating spliced images in FantasticReality cannot be guaranteed, as illustrated in Figure 2(d). Each of these 10 images is spliced of smaller images. However, they are annotated as authentic within FantasticReality. This error could make the network confused regarding splicing traces.

Spliced regions within Columbia exhibit random shapes, as illustrated in Figure 2(e). These images show noticeable edge artifacts and differ from our practical spliced images. While these spliced images lead the network to overfit edge artifacts, they cannot facilitate learning of comprehensive splicing traces.

In addition to the aforementioned concerns, a significant prevailing issue across existing splicing datasets is the limited semantic diversity within spliced regions. A summary of the number of semantic categories within the spliced regions of each dataset is provided in Table 1(a). As a result,

the network becomes overly sensitive and overfits to specific semantics, hindering its progress in learning effective splicing traces, and it leads to severe misidentification and diminished splicing detection capability.

In short, there are many challenges in the field of splicing detection. To address these issues, we introduce the high-quality splicing dataset GreatSplicing as a solution.

## GreatSplicing

For ease of describing GreatSplicing, the following definitions are stated briefly.

A spliced image is defined as

$$s := x_f \odot M + x_b \odot (1 - M). \quad (1)$$

Here,  $x_f$  and  $x_b$  refer to two different authentic image sources, while  $M$  represents a binary image mask where each element takes value from  $\{0, 1\}$ , with the symbol  $\odot$  indicating element-wise multiplication. It should be pointed out that

$$\sum_{(i,j)} M(i,j) < \sum_{(i,j)} [1 - M(i,j)]. \quad (2)$$

This inequation suggests that the area of the spliced region expressed by  $x_f \odot M$  should be smaller than the area of the background region expressed by  $x_b \odot (1 - M)$ .

The process is meticulously controlled while creating GreatSplicing to avoid shortcomings present in existing splicing datasets. It can be summarized as

$$M \odot (\Pi_{i=1}^{k_f} \varphi_i)(x_f) + (1 - M) \odot (\Pi_{j=1}^{k_b} \phi_j)(x_b) \rightarrow s. \quad (3)$$

Specifically, the process involves these four steps:





Figure 3: Selected spliced images from GreatSplicing. GreatSplicing exhibits significant advantages, such as semantic richness, high verisimilitude, high resolution, etc.

Category	Detail
<b>Cut-Out Tools</b>	1) marquee tool, 2) lasso tool,
	3) magic wand tool, 4) pen tool,
	5) quick selection tool
<b>Image Manipulations</b>	1) scale, 2) rotate, 3) skew,
	4) distort, 5) brightness/contrast,
	6) flip, 7) levels, 8) hue/saturation,
	9) curves, 10) color balance

Table 2: Summary of the tools and image manipulations used in creating GreatSplicing.

1. Selecting two images,  $x_f$  and  $x_b$ , from BossBase, to be used for the spliced region source and the background region source respectively. Each image in BossBase is used only once, creating 5,000 spliced images from 10,000 authentic images.
2. Applying one or more of the image manipulations shown in Table 2 consecutively to  $x_f$  and  $x_b$  (i.e.,  $\prod_{i=1}^{k_f} \varphi_i$  and  $\prod_{j=1}^{k_b} \phi_j$  in formula 3).
3. Using the cut-out tools shown in Table 2, we extract a region from  $x_f$  as the spliced region, paste it onto  $x_b$  as a new layer, and save it in PNG format. If the spliced region is a recognizable semantic object, the spliced image is called an object-awareness spliced image. Conversely, if the spliced region is a random area without explicit content, the spliced image is called a shape-random spliced image. GreatSplicing comprises 2,887 object-awareness spliced images and 2,113 shape-random spliced images.
4. Utilizing the Color Overlay in blending options, we set spliced regions to white and background regions to black, generating groundtruths.

In addition to spliced images and their corresponding groundtruths, the following information for each spliced im-

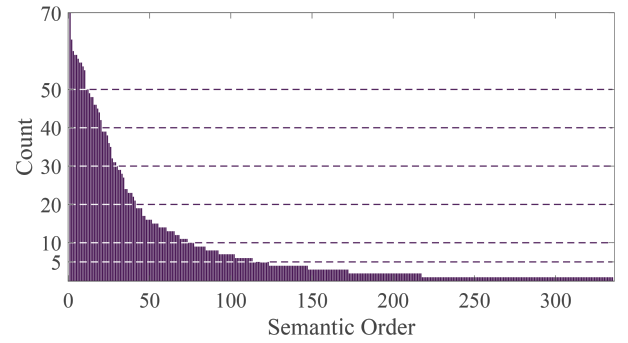


Figure 4: GreatSplicing comprises 335 distinct semantic categories for its spliced regions. In the distribution diagram, the horizontal axis represents the index of each semantic category, while the vertical axis represents the amount of spliced images corresponding to each semantic category, arranged in descending order.

age is also recorded and compiled:

1. Corresponding image index of  $x_f$  and  $x_b$  in BossBase;
2. Semantic content of the spliced regions;
3. Image manipulations undergone by  $x_f$  and  $x_b$ ;
4. Number of the connected spliced regions.
5. Annotation of the spliced region in  $x_f$ .

GreatSplicing includes 335 distinct semantic categories for spliced regions, surpassing all existing splicing datasets. Figure 4 presents a distribution diagram displaying the amount of spliced images for each semantic category.

Additionally, the manual creation process of GreatSplicing ensures a high level of realism in the object-awareness spliced images. These images are comparable to those in real-world applications. Selected object-awareness spliced images from GreatSplicing are illustrated in Figure 3.



## Recommended Experiment Settings

Based on GreatSplicing, five fair and reasonable recommended experimental settings for comparison experiments are proposed.

- **Self-Sufficiency Mode.** GreatSplicing is divided into training and testing sets. The neural network is trained on the training set without any additional tricks like data augmentation and is subsequently tested on the testing set.
- **Cross-Dataset Validation Mode.** All spliced images from GreatSplicing compose the training set, while other existing splicing detection datasets form the testing set, without any other tricks.
- **Synthetic-Finetuning Mode.** Pre-training the neural network on a synthetic dataset is permitted. GreatSplicing is divided into fine-tuning training and testing sets. The pre-trained model is fine-tuned on the fine-tuning training set and subsequently tested on the testing set.
- **Large-Sample Mode.** GreatSplicing is divided into training and testing sets. The training set is augmented to generate a mass of training samples.
- **Authentic-Image Introduction Mode.** Authentic images from BossBase are included in the training set. This mode can be used in combination with the above modes. (Important note: Authentic images cannot be present as spliced or background regions in the testing set.)

*Self-Sufficient Mode* and *Cross-Dataset Validation Mode* are basic patterns as experimental settings. The rationality will be verified in the following Train-Test Validation and Cross-Dataset Splicing Detection experiments separately. The other three modes could be adopted according to concrete situations.

## Experiment

This section conducted extensive comparison experiments and evaluations among GreatSplicing and existing splicing datasets, with state-of-the-art network designs for splicing detection. The results demonstrate that rich semantic categories in spliced regions make the detection model learn splicing traces better with a superior detection rate.

### Experiment Settings

**Experimental environment.** All experiments were conducted on a single NVIDIA Tesla T4 GPU, with Ubuntu 18.04, Python 3.8, and PyTorch 1.9.1.

**Datasets.** The existing splicing datasets used in the experiments include CASIA V2.0, DEFACTO, FantasticReality, NIST16, and Columbia. All images were resized to  $448 \times 608$  and saved in PNG format. Specific processing methods for each dataset are detailed below.

- CASIA-SP-1320 and CASIA-AU-700 were created by randomly selecting 1,320 spliced images and 700 authentic images from CASIA V2.0;
- DEFACTO-SP-3500 was formed by randomly selecting 500 images from each of 7 semantic classes of spliced images;

- FR-SP-3500 consists of 3,500 randomly selected spliced images from FantasticReality, while FR-AU-700 contains 700 authentic images randomly chosen from authentic images;
- Columbia-SP-180 comprises all 180 spliced images from Columbia;
- NIST16-SP-280 includes 280 spliced images from NIST16;
- We randomly selected 2,100 object-awareness spliced images and 1,400 shape-random spliced images from GreatSplicing to form GS-SP-3500. Additionally, we randomly selected 300 object-awareness spliced images and 400 shape-random spliced images from the remaining GreatSplicing dataset to create GS-SP-700. Furthermore, 700 images were randomly selected from BossBase to create GS-AU-700.

The selected authentic images have not been used as the source for spliced regions or background regions in the chosen spliced images.

**Neural networks.** Three basic visual networks and two state-of-the-art splicing detection networks were selected for experimental evaluation. The three basic networks are VGG-16(Simonyan and Zisserman 2014), ResNet-50(He et al. 2016), U-Net(Ronneberger, Fischer, and Brox 2015). We made minor adjustments to adapt them for end-to-end splicing detection. The two splicing detection networks are RRU-Net(Bi et al. 2019) and MVSS-Net(Dong et al. 2022). They provided an objective and comprehensive reflection of each dataset.

### Train-Test Validation

To validate the effectiveness of GreatSplicing, the train-test evaluation was conducted. GS-SP-3500 was used as the training set, and GS-SP-700 as the testing set. These five network architectures were trained and tested on GreatSplicing, with F1-score and IoU as the metric. The experimental results are depicted in Table 3.

The table reveals that U-Net and RRU-Net with segmentation capabilities excel in learning splicing traces from GreatSplicing, demonstrating superior splicing detection. Moreover, standard visual convolutional networks like VGG-16 and ResNet-50, lacking dedicated modules for splicing detection, exhibit subpar performance in train-test validation.

It is worth noting that the relatively lower F1-score and IoU metrics in our experiments, compared to other research findings, can be attributed to the modest sample size of the 3,500 training samples we employed. Other splicing detection models are typically trained on substantially larger datasets achieved through synthesis or augmentation techniques. Furthermore, the comparison experiments in existing literature are often conducted on datasets with inherent semantic cues, enabling the networks to identify the target better. In contrast, GreatSplicing eliminates semantic interference, compelling the network to learn genuine splicing traces, thus increasing the learning complexity.

Model	VGG-16	U-Net	ResNet-50	MVSS-Net	RRU-Net
F1	.286	.373	.282	.326	.400
IoU	.216	.305	.221	.267	.340

Table 3: Results of Train-Test Validation. Training set is GS-SP-3500 and testing set is GS-SP-700.

Dataset	Model				
	VGG-16	ResNet-50	U-Net	RRU-Net	MVSS-Net
CASIA V2.0	.046	.070	.023	<b>.015</b>	.012
FantasticReality	.067	.112	.082	.083	.063
GreatSplicing	<b>.007</b>	<b>.008</b>	<b>.005</b>	.028	<b>.008</b>

Table 4: Results of authentic image misidentification. The criterion used in the table is MIR.

### Authentic Image Misidentification

Due to the limited semantic diversity in spliced regions, existing splicing datasets hinder networks from effectively learning splicing traces. This leads to the extraction of irrelevant features and severe misidentification. In contrast, GreatSplicing offers diverse semantic categories in spliced regions, facilitating better splicing trace learning, reducing the extraction of irrelevant features, and lowering the misidentification of authentic images.

To quantify misidentification, the misidentification rate (MIR) is defined as the ratio of the misidentification region’s area to the authentic image’s area, expressed as

$$MIR = \frac{FP}{P + N}. \quad (4)$$

In this experiment, CASIA and FantasticReality datasets were chosen as they offer authentic images, avoiding cross-domain confusion. Five networks were trained on CASIA-SP-1320, FR-SP-3500, and GS-SP-3500, and subsequently tested on CASIA-AU-700, FR-AU-700, and GS-AU-700 respectively, with MIR calculated. The results are shown in Table 4.

The table shows that models trained on CASIA and FantasticReality have a high misidentification rate for authentic images. In contrast, models trained on GreatSplicing consistently perform best, with misidentification rates approaching zero in most cases.

In addition, we display misidentification results in Figure 5. The color images in the figure correspond to authentic images from CASIA, FantasticReality, and BossBase. Notably, CASIA-trained models show sensitivity to sky regions, while FantasticReality-trained models exhibit sensitivity to buses. Conversely, models trained on GreatSplicing display significantly reduced misidentification.

This fact shows that datasets that lack diverse semantic categories in spliced regions, cause the network to learn irrelevant features, leading to severe misidentification. GreatSplicing effectively helps the network learn splicing traces and reduces the misidentification rate.

It is important to emphasize that GreatSplicing does not trade between misidentification and splicing detection rates.

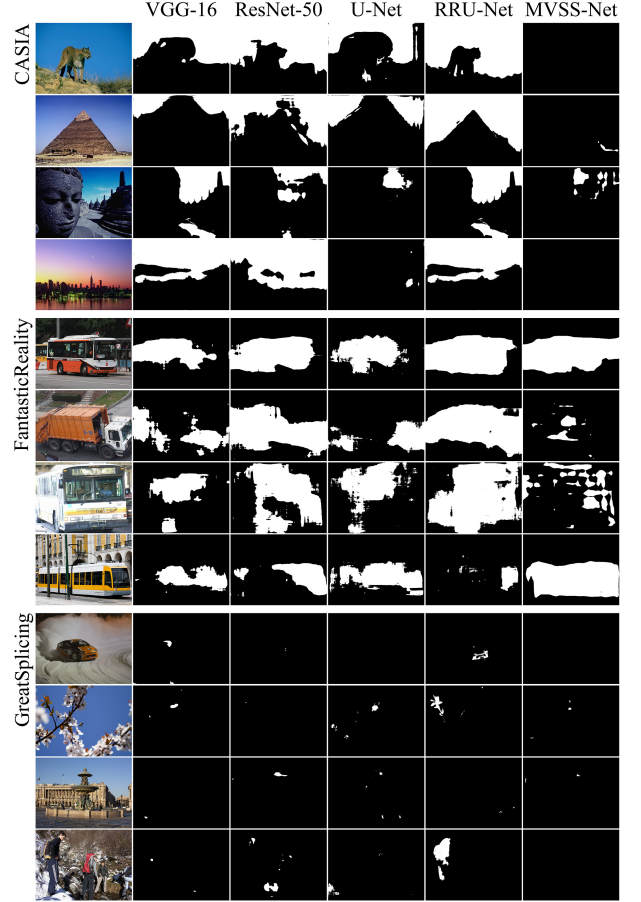


Figure 5: Results of authentic image misidentification. Twelve authentic color images were sourced from CASIA, FantasticReality, and BossBase. These images were tested with models trained on CASIA, FantasticReality, and GreatSplicing datasets respectively, resulting in misidentification outcomes.

Instead, it achieves a simultaneous improvement in both aspects. It will be evident in the following cross-dataset splicing detection experiments.

### Cross-Dataset Splicing Detection

This experiment aims to validate whether models trained on GreatSplicing exhibit superior generalization capabilities for splicing detection compared to existing datasets.

We conducted cross-dataset experiments using CASIA-SP-1320, FantasticReality-SP-3500, DEFACTO-SP-3500, GreatSplicing-SP-1320, and GreatSplicing-SP-3500 as the training sets, and Columbia-SP-180 and NIST16-SP-280 as the test sets. The results are presented in Table 5, with F1-score and IoU as the evaluation metric. Some visual results are presented in Figure 6.

The models trained on GreatSplicing exhibit consistently superior performance in cross-dataset splicing detection across almost all scenarios. This is attributed to the semantic richness of GreatSplicing in the spliced regions, which en-

Test	Train	Nums	VGG-16		ResNet-50		Model U-Net		RRU-Net		MVSS-Net	
			F1	IoU	F1	IoU	F1	IoU	F1	IoU	F1	IoU
Columbia	(Criterion)											
	CASIA V2.0	1,320	.191	.118	.064	.036	.258	.168	.234	.159	.210	.139
	GreatSplicing	1,320	<b>.226</b>	<b>.138</b>	<b>.204</b>	<b>.128</b>	<b>.292</b>	<b>.194</b>	<b>.390</b>	<b>.298</b>	<b>.301</b>	<b>.209</b>
	FantasticReality	3,500	.351	.242	.085	.242	.160	.278	.194	.090	.090	.057
	DEFACTO	3,500	<b>.394</b>	<b>.290</b>	.006	.003	.240	.157	.062	.037	.019	.011
NIST16	GreatSplicing	3,500	.282	.185	<b>.316</b>	<b>.211</b>	<b>.362</b>	<b>.245</b>	<b>.536</b>	<b>.443</b>	<b>.337</b>	<b>.246</b>
	CASIA V2.0	1,320	.176	.116	.131	.088	.222	.154	.337	.261	.277	.198
	GreatSplicing	1,320	<b>.280</b>	<b>.198</b>	<b>.314</b>	<b>.228</b>	<b>.331</b>	<b>.248</b>	<b>.411</b>	<b>.318</b>	<b>.336</b>	<b>.254</b>
	FantasticReality	3,500	.028	.018	.031	.020	.052	.033	.280	.214	.046	.030
	DEFACTO	3,500	.302	.204	.064	.048	.341	.238	.261	.199	.070	.049
	GreatSplicing	3,500	<b>.305</b>	<b>.217</b>	<b>.305</b>	<b>.225</b>	<b>.356</b>	<b>.272</b>	<b>.448</b>	<b>.358</b>	<b>.337</b>	<b>.253</b>

Table 5: Results of Cross-Dataset Splicing Detection. GreatSplicing exhibits optimal performance in nearly all test scenarios.

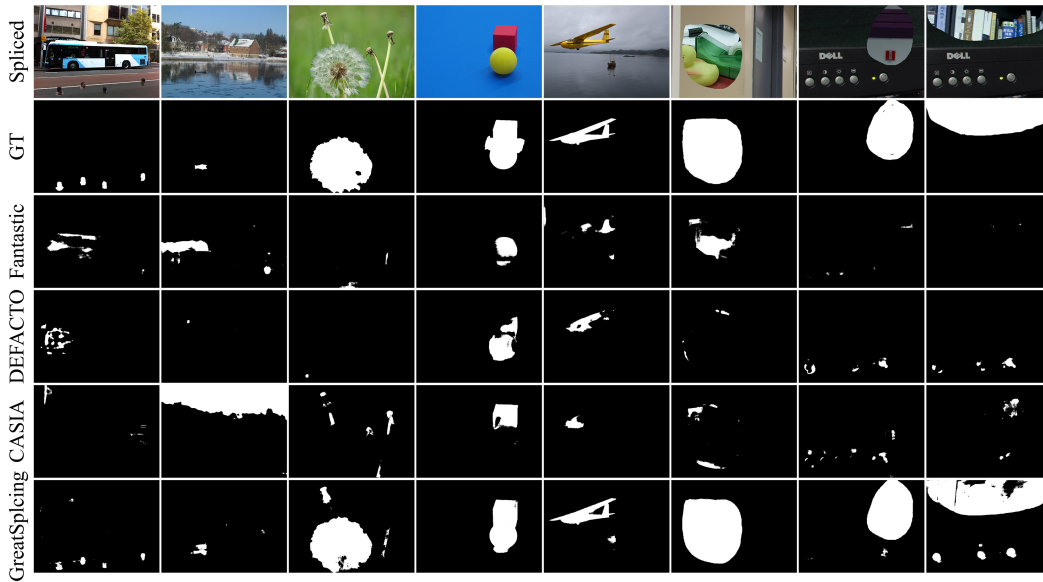


Figure 6: Cross-Dataset Splicing Detection. We trained FantasticReality-SP-3500, DEFACTO-SP-3500, CASIA-SP-1320, and GreatSplicing-SP-3500 with RRU-Net and tested on NIST16 and Columbia.

ables the network to learn more general splicing traces, reducing the reliance on irrelevant features. Besides, the cross-dataset experiment confirms that GreatSplicing achieves a dual benefit, reducing the misidentification rate while simultaneously enhancing the splicing detection rate, without any trade-off between the two.

## Conclusion

The existing splicing datasets suffer from limited semantic diversity in the spliced regions, hindering networks from learning general splicing traces effectively, leading to high misidentification rates on authentic images and decreased splicing detection capabilities. In addition, existing datasets exhibit a variety of flaws stemming from improper construction. To address these issues, our proposed GreatSplicing dataset offers rich semantic diversity in the spliced regions,

encompassing 335 semantic categories. Furthermore, GreatSplicing follows standardized production procedures, overcoming the deficiencies present in existing splicing datasets. Extensive experiments confirm that models trained on GreatSplicing significantly reduce misidentification rates while enhancing splicing detection performance, simultaneously improving both aspects. Based on GreatSplicing, five recommended experimental settings are proposed for splicing detection in comparison to promote fairness in evaluations.

## References

Ahmad, M.; and Khursheed, F. 2021. Digital image forgery detection approaches: a review. In *Applications of Artificial Intelligence in Engineering: Proceedings of First Global Conference on Artificial Intelligence and Applications (GCAIA 2020)*, 863–882. Springer.



- Bas, P.; Filler, T.; and Pevný, T. 2011. "Break our steganographic system": the ins and outs of organizing BOSS. In *International workshop on information hiding*, 59–70. Springer.
- Bi, X.; Wei, Y.; Xiao, B.; and Li, W. 2019. RRU-Net: The ringed residual U-Net for image splicing forgery detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 0–0.
- Bi, X.; Yan, W.; Liu, B.; Xiao, B.; Li, W.; and Gao, X. 2023. Self-Supervised Image Local Forgery Detection by JPEG Compression Trace. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 232–240.
- Dong, C.; Chen, X.; Hu, R.; Cao, J.; and Li, X. 2022. Mvss-net: Multi-view multi-scale supervised networks for image manipulation detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3): 3539–3553.
- Dong, J.; Wang, W.; and Tan, T. 2013. Casia image tampering detection evaluation database. In *2013 IEEE China summit and international conference on signal and information processing*, 422–426. IEEE.
- Guan, H.; Kozak, M.; Robertson, E.; Lee, Y.; Yates, A. N.; Delgado, A.; Zhou, D.; Kheyrkhah, T.; Smith, J.; and Fiscus, J. 2019. MFC datasets: Large-scale benchmark datasets for media forensic challenge evaluation. In *2019 IEEE Winter Applications of Computer Vision Workshops (WACVW)*, 63–72. IEEE.
- He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.
- Hsu, J.; and Chang, S. 2006. Columbia uncompressed image splicing detection evaluation dataset. *Columbia DVMM Research Lab*, 6.
- Kniaz, V. V.; Knyaz, V.; and Remondino, F. 2019. The point where reality meets fantasy: Mixed adversarial generators for image splice detection. *Advances in neural information processing systems*, 32.
- Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; and Zitnick, C. L. 2014. Microsoft coco: Common objects in context. In *Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part V 13*, 740–755. Springer.
- Mahfoudi, G.; Tajini, B.; Retraint, F.; Morain-Nicolier, F.; Dugelay, J. L.; and Marc, P. 2019. DEFACTO: Image and face manipulation dataset. In *2019 27th european signal processing conference (EUSIPCO)*, 1–5. IEEE.
- Ng, T.-T.; and Chang, S.-F. 2004. A model for image splicing. In *2004 International Conference on Image Processing, 2004. ICIP'04.*, volume 2, 1169–1172. IEEE.
- Novozamsky, A.; Mahdian, B.; and Saic, S. 2020. IMD2020: A large-scale annotated dataset tailored for detecting manipulated images. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops*, 71–80.
- Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18*, 234–241. Springer.
- Shorten, C.; and Khoshgoftaar, T. M. 2019. A survey on image data augmentation for deep learning. *Journal of big data*, 6(1): 1–48.
- Simonyan, K.; and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.