LOTA: Bit-Planes Guided AI-Generated Image Detection

Supplementary Material
Project Website: https://github.com/hongsong-wang/LOTA

A. New Dataset of AI-generated Image Detection: HardGVD

As detection accuracy on the GenImage dataset [3] has approached saturation (i.e., nearly 100%), we introduce a more challenging dataset, HardGVD, to better evaluate AI-generated image detection methods.

We sample frames from AI-generated and real videos to construct the challenging AI-generated image detection eval set. Details about this constructed eval set are shown in Table 1, and results are shown in Table 2. While other SOTA methods completely fail on this dataset, our approach achieves an accuracy of approximately 70%.

Table 1. The challenging AI-generated image detection eval set.

Subset	Generator	Label	Sampled Videos	Sampled frames
MuseV	MuseV	Fake	5,000	10,000
SVD	Diffusion	Fake	5,000	10,000
CogV	CogVideo	Fake	5,000	10,000
Mora	Mora	Fake	5,000	10,000
HD-VG	—	Real	10,000	40,000
COG	CogVideo	Fake	500	2,500
T2VZ	Text2Video-Zero	Fake	500	2,500
TAV	Tune-A-Video	Fake	500	2,500
VC	VideoCrafter	Fake	500	2,500
HD-VG	—	Real	2,000	10,000

Table 2. Results of AI-generated image detection on the HardGVD.

Subset	MuseV	SVD	Mora	CogV	HD-VG	Avg.
LaRE^2 [2]	7.1	6.8	23.6	37.5	63.8	41.3
ESSP [1]	33.4	38.4	32.6	39.8	61.5	48.8
LOTA	**75.1**	**83.0**	**81.3**	**80.9**	**73.0**	**76.5**
Subset	COG	T2VZ	TAV	VC	YT-BI	Avg.
LaRE^2 [2]	13.1	15.8	32.1	38.0	55.8	40.3
ESSP [1]	28.8	18.8	15.4	19.7	47.4	34.0
LOTA	**71.8**	**72.3**	**89.6**	**78.2**	**48.3**	**63.1**

B. Generalization Across Datasets

We also construct new datasets by using FLUX.1-dev and Google Imagen respectively to generate 1,000 images each. In addition, we construct two partial manipulation subsets: ForgeryNet and DF40 (with 4 different local forgery techniques), each containing 1,000 images. To evaluate cross-dataset generalization, we directly apply the model trained on GenImage to four different subsets: FLUX.1-dev, Google Imagen, ForgeryNet and DF40.

As shown in Table 3, without additional finetuning, our model consistently achieves accuracy exceeding 90% on all four subsets. Our proposed AI-generated image detection framework LOTA works well on partial manipulations such as inpainting or editing.

Table 3. Comparison with SOTA on new datasets with more advanced image generation models and partial manipulations.

Method	FLUX.1-dev	Google-Imagen	ForNet	DF-40	Avg.
LaRE^2 [2]	23.4	1.3	6.7	4.7	9.03
ESSP [1]	28.7	54.6	3.7	0.20	21.8
LOTA	**100**	**93.0**	**91.3**	**99.2**	**95.9**

C. Results on High Resolution Images

We provide an evaluation of AI-generated image detection on high-resolution images in Table 4, and analyze the impact of patch size on detection performance. We find that the best patch size is approximately 1/8 of the input image resolution, and our approach is highly robust against image resolutions.

Table 4. Multi-resolution analysis with varying patch strategies.

Resolution	16×16	32×32	64×64	96×96
128×128	**87.3**	78.4	19.8	10.4
256×256	7.5	**67.1**	0.1	0
512×512	0.19	53.5	**98.9**	89.7
1024×1024	75.4	93.0	98.9	**100**

References

[1] Jiaxuan Chen, Jieteng Yao, and Li Niu. A single simple patch is all you need for ai-generated image detection. *arXiv preprint arXiv:2402.01123*, 2024. 1

[2] Yunpeng Luo, Junlong Du, Ke Yan, and Shouhong Ding. Lareˆ 2: Latent reconstruction error based method for diffusion-generated image detection. In *CVPR*, pages 17006–17015, 2024. 1

[3] Mingjian Zhu, Hanting Chen, Qiangyu Yan, Xudong Huang, Guanyu Lin, Wei Li, Zhijun Tu, Hailin Hu, Jie Hu, and Yunhe Wang. Genimage: A million-scale benchmark for detecting ai-generated image. *NeurIPS*, 2023. 1