

TERRAFly-FORENSICS: A DATASET FOR FORENSIC DETECTION OF GENERATED MAP IMAGES WITH QUALITY ASSESSMENT OF GENERATIVE MODELS

Arpan Mahara, Liangdong Deng, Naphtali D. Rishe

Florida International University
Knight Foundation School of Computing and Information Sciences
11200 SW 8th St. Miami, Florida, United States

ABSTRACT

The rapid advancement of generative models has significantly influenced various domains, including Remote Sensing (RS), through applications in map generation, navigation, and environmental monitoring. However, alongside these positive impacts, the misuse of generative models has raised critical concerns, such as adversarial attacks. Existing forensic detection studies rely on publicly available datasets but often lack a focus on map images, which are essential for navigation and critical decision-making tasks. To bridge this gap, we introduce TerraFly-Forensics, a comprehensive dataset comprising real and synthesized map images generated using advanced generative models. The real maps were downloaded using the TerraFly mapping system, and state-of-the-art generative models were applied to produce high-quality synthesized map images, accompanied by a detailed evaluation of their generative quality. We have conducted an experimental evaluation of several recent forensic detection methods on the TerraFly-Forensics dataset, providing insights into and contributing a new foundation for forensic detection in Remote Sensing.

Index Terms— Generative AI, geospatial analysis, map images, adversarial attacks, forensic detection, Generative Adversarial Networks (GANs), diffusion models

1. INTRODUCTION

Vector tiles [2] are vector-based representations of geographical features derived from satellite or aerial imagery. These tiles can be rendered into visual forms, commonly referred to as map images. Such representations facilitate the development of user-friendly, customizable, and interactive mapping systems. Map images have a wide range of applications, including geospatial analysis, navigation, urban planning, smart city development, environmental monitoring of wildlife and vegetation, emergency rescue operations, and national defense through terrain analysis. This versatility arises from their capacity for interactive exploration of objects of interest within satellite imagery.

The construction of map images traditionally involves labor-intensive manual processes, including rendering and

styling using tools such as Mapnik [3], often guided by style sheets and spectral band processing. The emergence of Generative Artificial Intelligence (AI) presents a promising alternative for automating map image construction. Generative approaches leverage learned knowledge of map features to produce styled maps through models trained on representative datasets. For example, [4] introduced the dataset Aerial Photograph \leftrightarrow Maps, which facilitated generative models to translate aerial images into map images. Subsequently, several works [5, 6, 7] based on Generative Adversarial Networks (GANs) extended the exploration of generative performance on this dataset. Despite these advancements, the application of advanced unconditional GANs, such as ProGAN [8] and StyleGAN2-ADA [9], as well as diffusion models like Latent Diffusion [10] and DiT [11], remains underexplored due to constraints in dataset availability.

The proliferation of advanced generative methods also raises concerns about adversarial misuse, similar to the challenges posed by deepfakes. Systems involving map images may face critical issues from adversarial attacks. To address these risks, the detection of synthesized maps is a crucial first step. While several forensic detection methods [12, 13, 14, 15, 16] have been developed to identify synthesized images in popular datasets like ImageNet [17] and MSCOCO [18], there is a notable lack of similar efforts dedicated to map images.

The primary reasons for these gaps in the literature are the scarcity of map image datasets required to train advanced generative models and the absence of dedicated forensic tools for map images. To address these gaps, the contributions of this study are as follows:

- Development of a curated dataset of map images, obtained using the TerraFly mapping system, with extensive preprocessing and data curation.
- Application of advanced generative models from both the diffusion and GAN families to generate high-quality map images.
- Forensic detection analysis of state-of-the-art methods on the generated map images, marking the first attempt

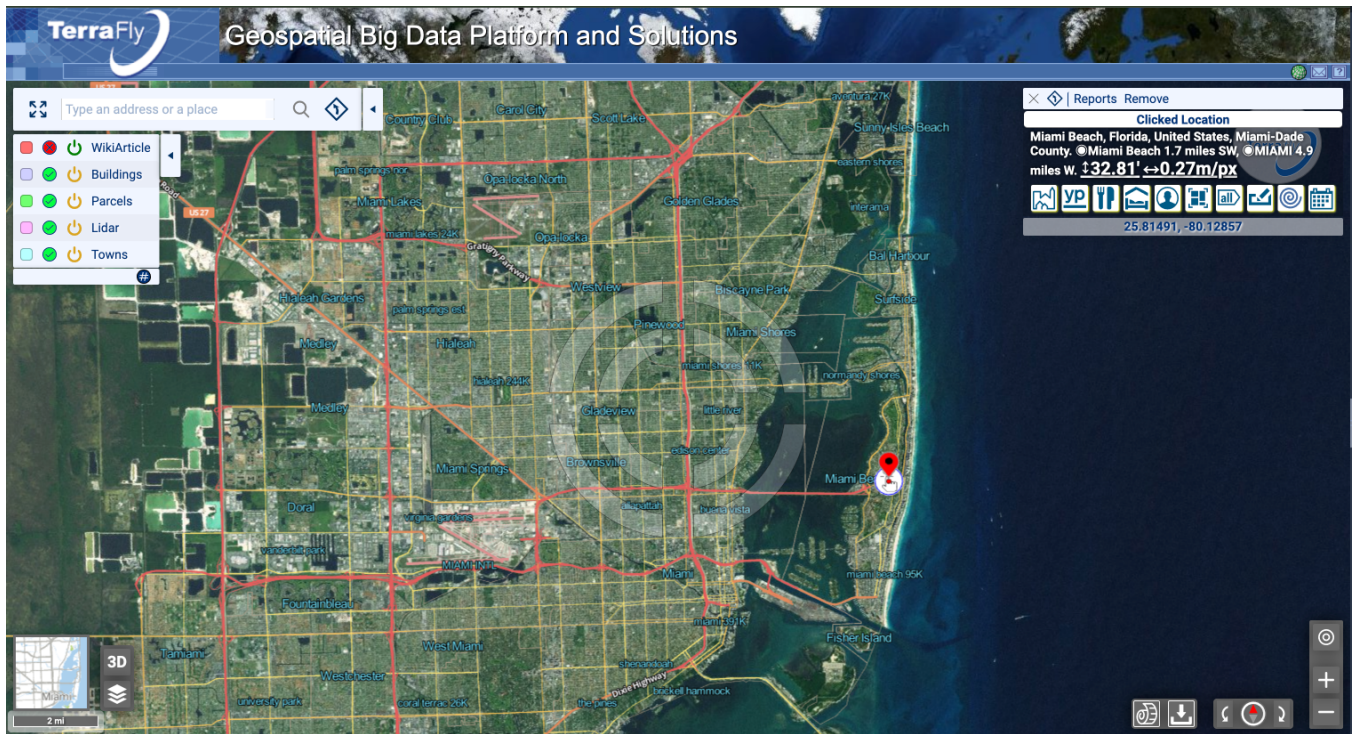


Fig. 1. TerraFly Mapping System [1]

at addressing forensic detection and adversarial attack prevention in the context of map image generation.

- Introduction of TerraFly-Forensics, a new dataset containing 47,467 real and synthesized map images.

2. RELATED WORK

The research community has been actively developing forensic detection methods for generated images to address concerns about unethical practices, societal harm, privacy breaches, and adversarial attacks. To facilitate the experimental evaluation of these methods and to assess their effectiveness in mitigating adversarial attacks across various domains, several datasets have been made publicly available.

[12] introduces a dataset generated by training on the LSUN dataset [19], which includes 20 object categories. Similarly, the CIFAKE dataset, proposed in [20], consists of generated images created using models trained on CIFAR-10 [21], consisting classes such as airplanes, automobiles, and others. In [22], the ImagiNet dataset is presented, containing generated images of diverse content, including paintings and faces. While these datasets are large-scale and diverse, they do not include datasets focused on critical infrastructure, such as those used for geospatial analysis.

Although some studies, as mentioned in [23], address the generation and detection of satellite imagery, the study of forensic detection of vector tiles, particularly map images,

remains unexplored. More importantly, no existing dataset focuses on forensic detection of map images within the context of geospatial analysis. To address this gap, we present the TerraFly-Forensics dataset, which includes both real and generated map images. This dataset is complemented by a generative quality study of map generation and a forensic detection analysis for map images, paving the way for the development and maintenance of robust geospatial systems.

3. METHOD

Our study presents TerraFly-Forensics dataset that contains the map images downloaded using TerraFly [1] mapping system and the generated images from advanced generative models by applying on the downloaded dataset. Primarily, the TerraFly-Forensics dataset is presented so that the research community can use it for measuring the forensic detection robustness beyond what the current detection dataset consists. The following subsection details the process on the data download, followed by the selection of different generative models and finally the selection of different synthetic image detection.

3.1. Data Download

The TerraFly mapping system (illustrated in Fig. 1), which implements the Microsoft Bing projection, was utilized to obtain the map images. Specifically, TerraFly's vector API

was used to download RGB map images with a resolution of 256×256 pixels. Three U.S. states, Florida, Georgia, and Texas, were selected as the regions of interest. Using the TerraFly user interface, X and Y tile coordinates were identified to select areas that captured meaningful geographic features within these states. The zoom level, Z , representing the map scale, was uniformly set to 15 to maintain consistent resolution across the dataset. A Python-based web scraping script was integrated to automate the download process, ensuring that each selected tile represented distinct and meaningful features such as roads, buildings, or natural elements.

To enhance the dataset's diversity and applicability, specific map layers were extracted: *osm_land*, *osm_landuse*, *osm_water*, *osm_buildings*, and *osm_roads*. These layers cover critical geographic and infrastructural features such as land coverage, land use types, water bodies, buildings, and road networks, making the dataset suitable for a wide range of geospatial analysis applications.

3.2. Generative AI Model Selection and Training

Generative AI has achieved significant advancements in the Computer Vision domain, particularly in the generation of imagery data. In this study, both GANs and diffusion models were selected for generating map images.

GANs, first proposed in [24], consist of two competing neural networks: a generator G and a discriminator D , trained adversarially by optimizing an adversarial loss function. For unconditional generation, two widely recognized GAN architectures were selected: ProGAN [8] and StyleGAN2-ADA [9]. ProGAN introduced a progressive training approach, incrementally increasing the resolution of G and D during training, resulting in high-quality and high-resolution images. ProGAN was implemented using its official TensorFlow framework and trained for 12,000 Kaimings. StyleGAN2-ADA, on the other hand, incorporated Adaptive Discriminator Augmentation (ADA), which dynamically adjusts stochastic augmentations applied to the discriminator to prevent overfitting. We trained StyleGAN2-ADA using its official PyTorch implementation for 12,000 Kaimings, consistent with ProGAN.

For conditional map generation, we utilized CycleGAN [25] and AttentionGAN [26], which are designed for unpaired image-to-image translation tasks. CycleGAN operates on the principle of cycle consistency and AttentionGAN incorporates an attention mechanism to identify and focus on relevant features. Both models were implemented using their official PyTorch frameworks and trained for 400 epochs. Diffusion models, first introduced by Sohl-Dickstein et al. [27], have gained significant attention for their effectiveness in image generation tasks. These models generate data by progressively transforming a Gaussian noise distribution into a target data distribution through a forward and reverse diffusion process. In this study, we implemented Latent Diffusion [10]

and DiT [11] for map generation. Latent Diffusion operates within a latent representation space, utilizing a Kullback-Leibler (KL)-regularized latent space to achieve efficient and high-quality image generation. Similarly, DiT replaces the conventional U-Net backbone with a Transformer-based architecture, facilitating class-conditional image generation. For this purpose, map images from three states were treated as distinct classes. Both models were trained for 800 epochs to ensure effective learning of the underlying map distributions.

3.3. Forensic Detection Techniques

As highlighted in Section 1, map images play a critical role in fields such as navigation, geospatial analysis, and other related domains. However, generative AI technologies may be exploited by malicious actors or third parties for unethical purposes or adversarial attacks. To address these threats, this study introduced the TerraFly-Forensics dataset, conducted a comprehensive review of current forensic synthetic image detection methods, and selected four state-of-the-art models for evaluation on our dataset. Specifically, the following models were considered: CNNSpot [12], DIRE [14], AIDE [15], and RINE [16].

The CNNSpot [12] method demonstrated that a carefully configured ResNet-50 architecture, combined with effective image preprocessing, can achieve generalizable detection performance for synthesized images generated by CNN-based GAN models. While earlier detection methods achieved robust performance on GAN-generated images, they faced challenges when applied to diffusion-generated images. To address this limitation, Diffusion Reconstruction Error (DIRE) [14] was proposed. DIRE measures the reconstruction error between an input image and its regenerated counterpart produced by a pre-trained diffusion model, such as DDIM [28], and has shown effectiveness in detecting diffusion-generated images.

Although existing methods succeeded in detecting images generated by various families of generative models, AIDE [15] and RINE [16] identified edge cases where previous approaches failed. To address these limitations, both methods leverage the capabilities of the large vision-language model, CLIP [29]. AIDE uses CLIP's pretrained visual encoder to compute patchwise features, capturing both high-level semantics and low-level artifacts for detecting AI-generated images. Similarly, RINE utilizes intermediate representations from CLIP's Transformer blocks and processes them through a projection network to extract forgery-aware features, thereby enhancing the detection of synthetic images.

All methods were implemented in the PyTorch framework using their official implementations. Training was conducted on both generated and real map images from the TerraFly-Forensics dataset. The experimental results are presented and discussed in the Results section below.

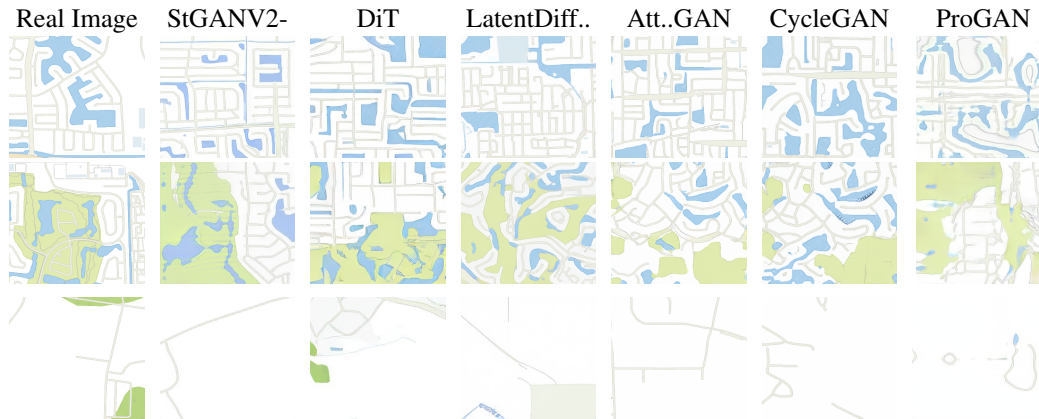


Fig. 2. Visualization of images generated by each generative model, along with example real images from the TerraFly-Forensics dataset. The images are not exact matches, as most of the models perform unconditional generation.

Table 1. Quantitative comparison of FID and KID scores across various models for map generation. Lower scores indicate better performance.

Model	FID ↓	KID (Mean ± Std) ↓
ProGAN [8]	104.3720	0.0709 ± 0.0018
StyleGANV2-ADA [9]	38.0042	0.0215 ± 0.0011
LatentDiffusion [10]	73.6138	0.0573 ± 0.0025
DiT [11]	62.4027	0.0362 ± 0.0018
CycleGAN [25]	41.6488	0.0251 ± 0.0013
AttentionGAN [26]	39.5313	0.0237 ± 0.0013

4. EXPERIMENTAL SETUP AND RESULTS

The training and evaluation of the generative models and forensic detection methods were conducted using NVIDIA GPUs on three different nodes. The first node contains A100 GPUs with 80 GB PCIe memory and 8 GPUs, the second node contains the same A100 GPUs but with 4 GPUs, and the final node contains A100 GPUs with 40 GB PCIe memory and 8 GPUs. All models were implemented in Python, with TensorFlow or PyTorch frameworks as specified in Subsections 3.2 and 3.3.

In total, we downloaded approximately 6,781 real map images. Following the training of the generative models described in Subsection 3.2, we synthesized an additional 6,781 map images for each of the six generative models, resulting in a total of 40,686 generated images. Consequently, the TerraFly-Forensics dataset comprises approximately 47,467 images, including real and generated map images.

4.1. Analysis and Results

To evaluate the quality of the images generated by the selected generative models, Fréchet Inception Distance (FID) and Kernel Inception Distance (KID) metrics were selected. The results are summarized in Table 1, where StyleGAN2-

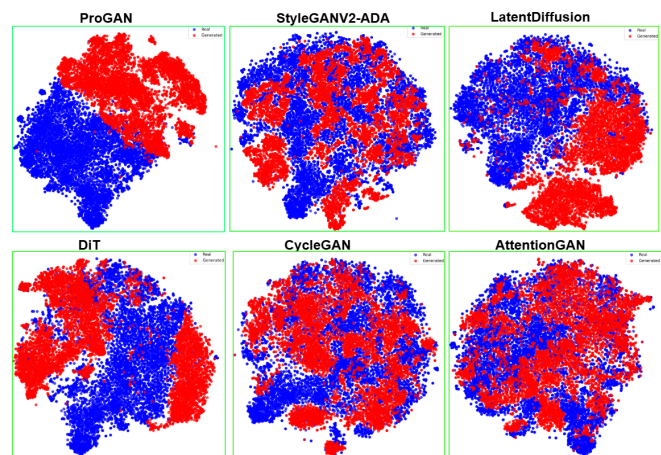


Fig. 3. t-SNE plot illustrating the distribution of real and generated images for generative models. Blue points represent real images, while red points correspond to generated images.

ADA achieved the best performance among all models, with an FID score of 38.0042 and a KID score of 0.0215 ± 0.0011 . Additionally, a t-distributed Stochastic Neighbor Embedding (t-SNE) plot is provided in Fig. 3, illustrating the distribution of generated map images from each generative model in comparison to real map images. Fig. 2 presents example images generated by each model, chosen randomly, along with real images from the TerraFly-Forensics dataset.

Similarly, our experimental setup involved training the detection methods on images generated by one type of generative model, combined with real images, and then testing the models on images generated by different types of models. Table 2 presents the classification accuracy for detecting fake images among 6,780 synthesized images per model. As seen in the table, RINE [16], the most recent method, demonstrated better performance than other methods. Missing values in the table correspond to cases where the detection model was trained on the full set of 6,780 images generated by the re-

Table 2. Cross-Generator Detection Accuracy on TerraFly-Forensics. The table shows the classification accuracy for detecting generated images as fake, evaluating the generalization capability of each detection method when trained on one type of generative model and tested on others.

Model	Training	Testing on Generative Models Individually					
		ProGAN [8]	StGANv2-ADA [8]	LatentDiff. [10]	DiT [11]	CycleGAN [25]	Att.GAN [26]
CNNSpot [12]	DiT	88.75	63.24	99.09	–	99.63	72.44
	StGANv2-ADA	60.48	–	67.36	4.78	93.88	84.74
	ProGAN	–	49.48	73.68	17.77	94.44	77.38
	Att.GAN	17.36	0.80	2.95	17.77	100.0	–
	CycleGAN	3.69	0.40	61.05	3.75	–	99.63
	LatentDiff.	40.66	14.28	–	1.81	87.04	60.80
DIRE [14]	DiT	83.11	38.97	99.37	–	95.73	62.51
	StGANv2-ADA	77.26	–	74.79	8.36	98.01	92.62
	ProGAN	–	39.62	63.17	21.87	93.12	84.29
	Att.GAN	9.70	0.10	61.60	17.77	99.92	–
	CycleGAN	25.05	5.22	73.72	4.04	–	99.88
	LatentDiff.	36.47	12.75	–	1.65	88.77	71.03
AIDE [15]	DiT	66.05	74.03	92.74	–	72.94	37.84
	StGANv2-ADA	60.26	–	96.29	77.92	96.58	54.55
	ProGAN	–	35.92	94.45	31.63	42.81	29.80
	Att.GAN	34.24	54.52	64.14	46.85	99.54	–
	CycleGAN	12.21	20.59	52.73	42.72	–	18.26
	LatentDiff.	22.71	14.57	–	11.16	47.44	2.09
RINE [16]	DiT	98.20	99.90	99.80	–	78.20	98.1
	StGANv2-ADA	99.50	–	99.90	84.30	99.70	99.00
	ProGAN	–	89.10	96.80	82.10	74.30	95.40
	Att.GAN	93.30	66.70	92.8	20.0	100.0	–
	CycleGAN	91.00	31.80	89.5	4.70	–	100.0
	LatentDiff.	85.30	36.10	–	24.70	67.70	82.80

spective model, divided into training and validation sets, and therefore had no additional data for testing.

5. CONCLUSION AND FUTURE WORK

While generative AI has brought numerous benefits to critical sectors in the geospatial domain, it also presents significant challenges, specifically through adversarial attacks. Existing studies on forensic detection to mitigate adversarial threats often focus on publicly available datasets, leaving a gap in addressing vector tiles or map images due to the lack of specialized datasets. This study addresses this gap by collecting a comprehensive map dataset from the TerraFly mapping system, training advanced generative models, and presenting the TerraFly-Forensics dataset comprising 47,467 map images. In addition, this study provides a comparative analysis of the generative capabilities of various state-of-the-art generative models and introduces the first forensic detection of synthesized map images using advanced detection methods. The present work contributes a resource for the research community to evaluate the performance of forensic detection methods and offers a direction toward developing adversari-

ally robust geospatial technologies. The dataset also serves as a valuable resource for evaluating the performance of image generation methods and can contribute to their improvement. This dataset, along with the experimental results, is available at <https://github.com/amaha7984/TerraFly-Forensics>.

While this study provides benchmarks for generative models, there is potential for improvement by exploring adjustments such as model training configurations, batch sizes, and alternative backbones, for instance, integrating different architectures in the DiT model [11]. As part of future work, we plan to expand the TerraFly-Forensics dataset by collecting additional map images and generating more synthetic images using advanced generative models.

6. ACKNOWLEDGEMENT

This material is based in part upon work supported by the National Science Foundation under Grant Nos. MRI20 CNS-2018611 and MRI CNS-1920182, and FDEP Grant C-2104.

7. REFERENCES

- [1] N. Rishe, Y. Sun, M. Chekmasov, A. Selivonenko, and S. Graham, "System architecture for 3d terrafly online gis," in *IEEE Sixth International Symposium on Multimedia Software Engineering*. IEEE, 2004, pp. 273–276.
- [2] L. Martinelli and M. Roth, *Vector Tiles from OpenStreetMap*, Ph.D. thesis, HSR Hochschule für Technik Rapperswil, 2015.
- [3] B. Bourdrez, "Improving Mapnik with Label Placement Algorithms," 2011.
- [4] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1125–1134.
- [5] J. Song, J. Li, H. Chen, and J. Wu, "Mapgen-gan: A fast translator for remote sensing image to map via unsupervised adversarial learning," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 2341–2357, 2021.
- [6] Y. Liu, W. Wang, F. Fang, L. Zhou, C. Sun, Y. Zheng, and Z. Chen, "Cscgan: Conditional scale-consistent generation network for multi-level remote sensing image to map translation," *Remote Sensing*, vol. 13, no. 10, pp. 1936, 2021.
- [7] A. Mahara and N. D. Rishe, "Generative adversarial model equipped with contrastive learning in map synthesis," in *Proceedings of the 2024 6th International Conference on Image Processing and Machine Vision*, 2024, pp. 107–114.
- [8] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," in *International Conference on Learning Representations*, 2018.
- [9] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, and T. Aila, "Training generative adversarial networks with limited data," *Advances in Neural Information Processing Systems*, vol. 33, pp. 12104–12114, 2020.
- [10] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 10684–10695.
- [11] W. Peebles and S. Xie, "Scalable diffusion models with transformers," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 4195–4205.
- [12] S.-Y. Wang, O. Wang, R. Zhang, A. Owens, and A. A. Efros, "Cnn-generated images are surprisingly easy to spot... for now," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 8695–8704.
- [13] H. Wu, J. Zhou, and S. Zhang, "Generalizable synthetic image detection via language-guided contrastive learning," *arXiv preprint arXiv:2305.13800*, 2023.
- [14] Z. Wang, J. Bao, W. Zhou, W. Wang, H. Hu, H. Chen, and H. Li, "Dire for diffusion-generated image detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 22445–22455.
- [15] Shilin Yan, Ouxiang Li, Jiayin Cai, Yanbin Hao, Xiaolong Jiang, Yao Hu, and Weidi Xie, "A sanity check for ai-generated image detection," *CoRR*, vol. abs/2406.19435, 2024.
- [16] C. Koutlis and S. Papadopoulos, "Leveraging representations from intermediate encoder-blocks for synthetic image detection," in *European Conference on Computer Vision*. Springer, 2025, pp. 394–411.
- [17] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al., "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, pp. 211–252, 2015.
- [18] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part V 13*. Springer, 2014, pp. 740–755.
- [19] F. Yu, A. Seff, Y. Zhang, S. Song, T. Funkhouser, and J. Xiao, "Lsun: Construction of a large-scale image dataset using deep learning with humans in the loop," *arXiv preprint arXiv:1506.03365*, 2015.
- [20] J. J. Bird and A. Lotfi, "Cifake: Image classification and explainable identification of ai-generated synthetic images," *IEEE Access*, 2024.
- [21] A. Krizhevsky, G. Hinton, et al., "Learning multiple layers of features from tiny images," 2009.
- [22] D. Boychev and R. Cholakov, "Imagenet: A multi-content dataset for generalizable synthetic image detection via contrastive learning," *arXiv preprint arXiv:2407.20020*, 2024.
- [23] X. Ding, Y. Nie, J. Yao, J. Tang, and Y. Lang, "Forensic research of satellite images forgery: a comprehensive survey," *Artificial Intelligence Review*, vol. 57, no. 9, pp. 253, 2024.
- [24] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in Neural Information Processing Systems*, vol. 27, 2014.
- [25] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2223–2232.
- [26] H. Tang, H. Liu, D. Xu, P. H. S. Torr, and N. Sebe, "Attentiongan: Unpaired image-to-image translation using attention-guided generative adversarial networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 4, pp. 1972–1987, 2021.
- [27] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli, "Deep unsupervised learning using nonequilibrium thermodynamics," in *International Conference on Machine Learning*. PMLR, 2015, pp. 2256–2265.
- [28] Jiaming Song, Chenlin Meng, and Stefano Ermon, "Denoising diffusion implicit models," in *International Conference on Learning Representations*, 2021.
- [29] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, et al., "Learning transferable visual models from natural language supervision," in *International Conference on Machine Learning*. PMLR, 2021, pp. 8748–8763.