

This paper is an extended version of a contribution presented
at the [Graphicon 2025 conference](#).

FakeSegment: Deep Feature Maps Analysis and Visualization for Enhancing of Synthetic Image Realism

V. Aleksandrov^{1,A}, V.V. Kniaz^{2,A,B}, M. Samodurov^{3,A}

^A State Research Institute of Aviation Systems (GosNIIAS), Moscow, Russia

^B Moscow Institute of Physics and Technology (MIPT), Moscow, Russia

¹ ORCID: 0009-0005-6084-9510, vsaleksandrov@gosniias.ru

² ORCID: 0000-0003-2912-9986, vl.kniaz@gosniias.ru

³ ORCID: 0009-0005-2575-5308, manifest@gosniias.ru

Abstract

The generation of synthetic datasets for training neural models in object detection and recognition tasks has become a prevalent approach due to the cost and time savings compared to collecting real world data. However, synthetic images often lack critical details, which can degrade the performance of trained models. To address this issue, we propose the FakeSegment neural model, designed to annotate unrealistic parts of synthetic images. Our method utilizes two Single Shot Multibox Detector (SSD) networks with shared weights. By analyzing the differences in corresponding feature maps from real and images. By comparing these feature maps, we can pinpoint areas where synthetic images diverge from expected patterns observed in real-world data. FakeSegment automatically detects unnatural areas within the synthetic data. We evaluate our model on two datasets: the FantasticReality dataset and a newly introduced UnrealLanding dataset focused on aircraft safety during landing. Our contributions include (1) the development of the FakeSegment model, (2) the creation of the UnrealLanding dataset with paired synthetic and real images, and (3) a comprehensive evaluation demonstrating that Fake Segment outperforms baseline methods by 15% in Intersection over Union (IoU) for segmenting unreal parts.

Keywords: generative modeling, image synthesis, image enhancement, object detection.

1. Introduction

The advent of deep learning has revolutionized the field of computer vision, particularly in tasks such as object detection and recognition. Central to the success of these models is the availability of large-scale annotated datasets. However, acquiring such datasets is often labor-intensive and costly, prompting researchers to explore synthetic data generation as a viable alternative. Synthetic datasets are created using advanced 3D modeling and rendering techniques, offering a controlled environment where diverse scenarios can be simulated efficiently. Despite their potential, even the most sophisticated 3D modeling suites struggle to produce synthetic images that match the quality and realism required for training models with performance comparable to those trained on real-world data.

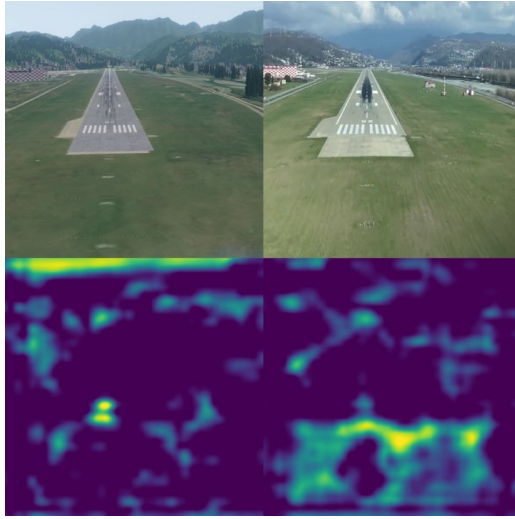


Figure 1: Example visualization of unrealistic details in synthetic images.

While there is a substantial body of literature focused on enhancing the visual fidelity of synthetic images through techniques like style transfer and domain adaptation, there remains a critical gap in understanding which specific features render synthetic images distinct from their real counterparts. Identifying these features is crucial for improving the utility of synthetic datasets in training robust neural models. This paper seeks to address this gap by visualizing unrealistic details in synthetic images using feature maps derived from neural networks, thereby providing insights into the discrepancies that affect model performance.

The pursuit of enhancing the perceptual realism of synthetic images has seen significant advancements in recent years. A seminal contribution in this domain was made by the CycleGAN model [1], which demonstrated the potential for unpaired image-to-image translation, thereby improving the visual fidelity of synthetic images. Building upon this foundation, subsequent research introduced methods such as Enhancing Photorealism Enhancement, which utilized the HRNetV2 model [2] alongside intermediate graphical buffers like normal and depth maps to further refine image realism [3]. More recently, diffusion models such as Flux Kontext have emerged, offering superior capabilities in incorporating realistic features into synthetic datasets [4]. Despite these advancements, a critical challenge remains: these models do not provide insights into which specific features make learning on synthetic images less effective compared to real-world data. Addressing this gap is essential for improving the utility of synthetic datasets in training robust neural networks. In this paper, we propose a novel neural framework FakeSegment aimed at the segmentation of unrealistic areas within synthetic images. Our objective is to leverage feature maps of a neural network to identify and delineate regions that deviate from realism, thereby facilitating targeted enhancements. This approach not only aids in understanding the specific features that distinguish synthetic images from real ones but also provides a practical tool for improving the quality of synthetic datasets. By employing linear combinations of texture map stacks, our method allows for the direct refinement of unrealistic elements within synthetic images, potentially bridging the gap between synthetic and real-world data.

In this study, we introduce the FakeSegment model, a novel approach for the segmentation of unnatural regions in synthetic images. To facilitate the training and evaluation of our model, we developed the UnrealLanding dataset, which comprises paired real and synthetic images depicting landings at Sochi airport. This dataset serves as a benchmark for assessing the performance of models in identifying unrealistic features in synthetic imagery. Our comprehensive evaluation includes comparisons with three contemporary baseline models, demonstrating that FakeSegment not only achieves but also competes with state-of-the-art performance in unnatural region segmentation. These results underscore the efficacy of our approach in enhancing the realism of synthetic datasets, thereby contributing to more robust

machine learning applications. Examples of visualizations predicted by our FakeSegment model are presented in Figure 1.

2. Related Work

The field of synthetic image generation and its application in training neural networks has garnered significant attention in recent years. Researchers have explored various methodologies to bridge the gap between synthetic and real-world data, aiming to improve model performance and generalization. This section reviews the key advancements in enhancing synthetic images and the use of such images for network training.

2.1. Enhancement of Synthetic Images

The enhancement of perceptual realism in synthetic images has been a significant area of research, with early successes marked by the development of the CycleGAN model [1]. This model pioneered the use of generative adversarial networks (GANs) to translate images from one domain to another, thereby improving their realism. Subsequent advancements include the work presented in "Enhancing Photorealism Enhancement," [3] which leveraged the HRNetV2 model [2] alongside intermediate graphical buffers, such as normal and depth maps, to achieve enhanced photorealistic effects. In the realm of human face synthesis, methods like "Beyond Reconstruction" have introduced a physics-based Neural Deferred Shader for photo-realistic rendering, significantly advancing the field [5]. Additionally, approaches such as "Generation of Synthetic Images for Pedestrian Detection Using a Sequence of GANs" have demonstrated effective strategies for generating realistic pedestrian images through sequential GAN architectures [6]. More recently, novel diffusion models like Flux Kontext have emerged, offering superior performance in incorporating realistic features into synthetic images by modeling complex distributions with greater fidelity [4]. These developments collectively highlight the diverse methodologies employed in enhancing the realism of synthetic images across various domains.

2.2. Artifact Explanation in Fake and Synthetic Images

The segmentation of unrealistic features in synthetic images is intricately linked to the broader challenge of detecting fake or manipulated images. This problem domain has seen significant advancements through frameworks like the Mixed Adversarial Generators [7]. In this work, a novel framework is proposed for training a discriminative segmentation model via an adversarial process. The framework involves simultaneous training of four models: a generative retouching model (GR) that translates manipulated images to the real image domain, a generative annotation model (GA) that estimates the pixel-wise probability of an image patch being either real or fake. Another significant contribution is the ManTraNet model, which offers a unified deep neural architecture for both detection and localization of manipulated regions without requiring extra preprocessing or postprocessing steps [8]. As a fully convolutional network, ManTraNet handles images of arbitrary sizes and various known forgery types such as splicing, copy-move, removal, enhancement, and even unknown types. In addition to these methods, the ContRail framework utilizes ControlNet [9] for synthesizing realistic railway images, showing casing advancements in domain-specific synthetic image generation [10]. Finally, the FakeVLM model provides capabilities to describe which features appear unnatural in fake images, offering insights into feature-level discrepancies that contribute to perceived unrealism [11].

3. Method

We aim developing a neural model capable of unsupervised detection and segmentation of unrealistic features in synthetic images generated using 3D visualization frameworks such as Unreal Engine and Blender. We use the SSD model [12] as the starting point for our research. Leveraging the deep feature maps produced by the SSD model, we aim indicating such details

of a synthetic image that decrease the performance of the object detection model that was trained using a synthetic data.

3.1. FakeSegment Framework Overview

In this work, we present the FakeSegment framework, designed to visualize and segment unrealistic details in synthetic images utilizing neural network feature maps. Our framework takes inspiration from the Single Shot Multibox Detector (SSD) model, renowned for its efficient object detection capabilities. Specifically, we employ two instances of the SSD model with shared weights, enabling the extraction of consistent feature representations from both real and synthetic images. The shared-weight architecture facilitates a robust comparison between genuine and artificial image features, ensuring that subtle discrepancies are effectively captured. Central to our framework is the use of an additional U-Net ResNet-based segmentation neural model, denoted as S . This model is responsible for converting the intermediate feature maps generated by the SSD into a heat map P . The heat map P quantitatively describes the probability of each pixel being real or fake, offering an intuitive visualization of potentially manipulated regions.

To train our model effectively, we employ a paired dataset consisting of real and synthetic images. This pairing allows for precise calculation of feature map differences between actual images, F_{real} , and fabricated ones, F_{fake} . We compute the difference between these feature maps and determine the average value across each pixel, which serves as an approximation of the desired heat map P . This heat map is crucial as it directs the segmentation neural model S to learn and predict fake regions from F_{fake} effectively. By minimizing the prediction error of these discrepancies, S learns to map the complex feature space to a probability distribution over each pixel, highlighting areas that are likely fake. The integration of these components forms our FakeSegment framework, visually summarized in Figure 2, illustrating its capacity to discern synthetic elements in images with a high degree of accuracy. This method reinforces the utility of combining powerful object detection models with nuanced segmentation approaches to tackle the challenges presented by synthetic image manipulation. The overview of the framework is presented in Figure 2.

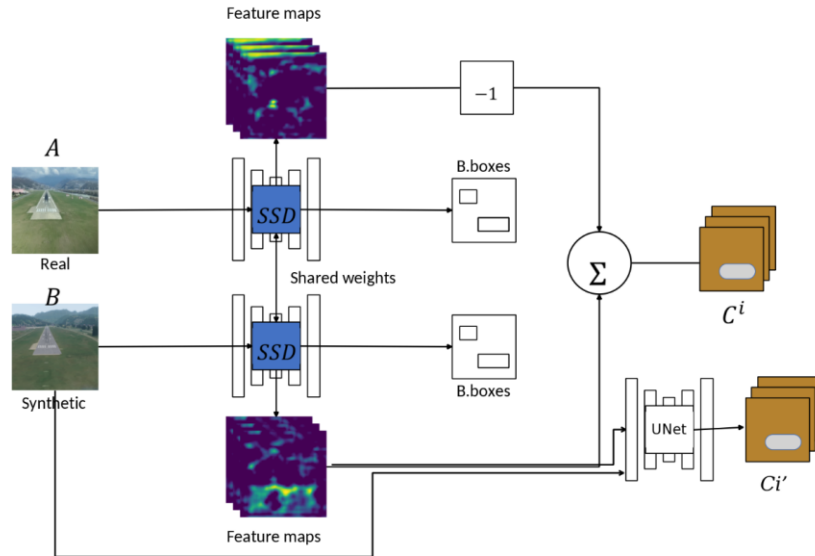


Figure 2: FakeSegment Framework Overview.

3.2. Network Architecture

In our network architecture, we address the challenge of discerning unrealistic details in synthetic images by operating within three distinct domains: the real image domain $A \in \mathbb{R}^{w \times h \times 3}$, the synthetic image domain $B \in \mathbb{R}^{w \times h \times 3}$, and the probability heatmap domain $S \in [0,1]^{w \times h}$. Our objective is to train a mapping function G that translates an input image $A \in A$ into a corresponding fake probability heatmap $S \in S$, effectively formulated as $G: A \rightarrow S$. Given the inherent ill-defined nature of this problem, compounded by the subjective perception of image realism, we adopt a two-stage approach to predict the probability heatmap S .

In the first stage, we utilize an off-the-shelf Single Shot Multibox Detector (SSD) to predict an intermediate probability map F . This map serves as a rich repository of feature vectors that can distinguish real image parts from fake regions. In the subsequent stage, the feature maps F are fed into a U-Net architecture, which has been specifically trained to translate these feature vectors into the desired probability map S . During training, we generate an approximation of S as the absolute difference between the real and fake feature maps, expressed by the equation [1]:

$$S = |F_{\text{real}} - F_{\text{fake}}|. \quad (1)$$

To enhance the dataset's diversity, we draw inspiration from the training methodology of the OASIS generative network [13]. We create composite images by randomly mixing real and synthetic images across various regions. For these mixed images, the annotation assigns a zero value to the real regions, effectively marking those as genuine parts of the frame. This approach not only enriches the dataset with a broader variety of mixed-content images but also reinforces the model's ability to discriminate between real and manipulated areas, thus improving the generalization capability of our framework.

The process of synthetic image enhancement at training and inference phases is shown in Figure 3.

So, the tree-like diagram can be supplemented by visualization of some integral characteristics:

- on the graph of a tree-like function corresponding to a certain time moment, can be plotted a chart of the maximum (or minimum) values of the corresponding parameter at a given point; for pipelines, this is usually the dependence on the co- ordinate of the maximum pressures reached in the corresponding time;
- on the graph of a tree-like function corresponding to a certain point in time can be supplemented by a plot of the maximum permissible values of the corresponding value; for pipelines, such a value is usually set to the maximum allowable pressure values above which the operation of the pipeline can cause its destruction;
- on the graph of a tree-like function corresponding to a certain time moment, dangerous spatial intervals can be demonstrated; at these intervals the maximum permissible values have already been exceeded by this time moment; in case of consideration of the pipeline system, the places of exceeding the maximum allowable pressure at a given time moment can be highlighted, for example by color, directly on the line of the function.

In conclusion, it should be particularly noted that the presented approach to visualization using tree-like graphs is applicable not only to displaying at specified moments of time, but also in the form of animated films. In this case, in our opinion, the visibility of the proposed approach in the data representation increases due to the continuous perception of the entire spacetime flow pattern.

In addition, animation visualization can be more visual in case of its implementation in real-time systems, when all the changes are reflected in real or advanced time, for example, in the control centers, from which the real pipeline systems are controlled. Animated visualization is indispensable in analyzing the appearances and development of emergencies.

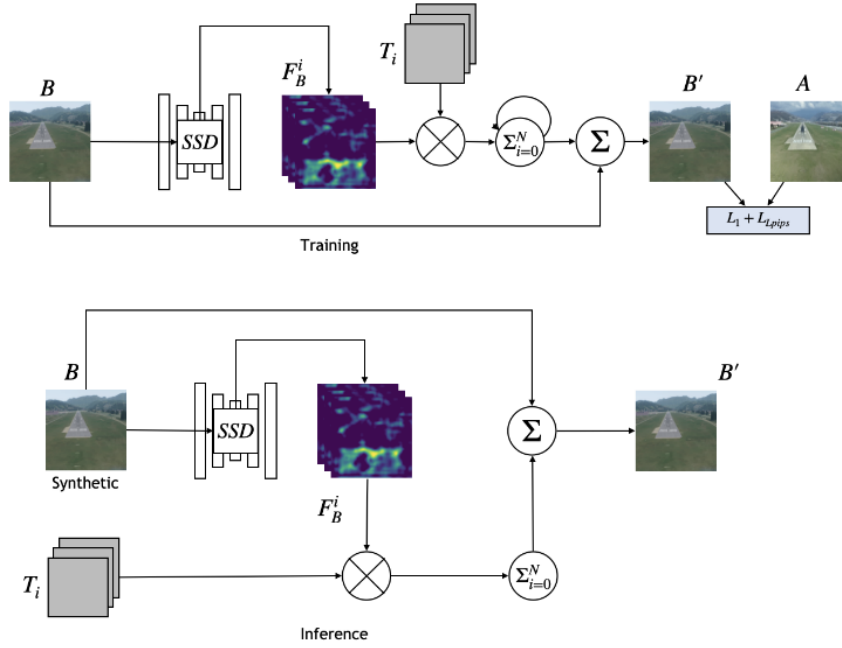


Figure 3: Synthetic image enhancement. Top: training phase; bottom: inference phase.

3.3. Loss Function

To effectively train the FakeSegment model, we employ a dual loss function strategy designed to robustly capture unrealistic details while enhancing the model’s discriminative capabilities. The two primary loss functions guiding our training process are the negative log-likelihood loss (NLL) \mathcal{L}_{NLL} and an adversarial loss \mathcal{L}_{adv} . The negative log-likelihood loss is critical for penalizing the omission of fake regions in our predicted probability map. Specifically, it evaluates the fidelity of the predicted probability map against ground-truth annotations, focusing on ensuring comprehensive detection of synthetic areas. The NLL loss is formulated as follows [2]:

$$\mathcal{L}_{\text{NLL}} = -\sum_{i,j} [y_{i,j} \log(p_{i,j}) + (1 - y_{i,j}) \log(1 - p_{i,j})], \quad (2)$$

where $y_{i,j} \in \{0,1\}$ denotes the ground-truth label of pixel (i,j) , indicating real (0) or fake (1), and $p_{i,j}$ represents the predicted probability of pixel (i,j) being fake. This formulation effectively punishes high confidence predictions of fake regions where none exist, as well as low confidence predictions where fake regions are present.

In addition to the NLL loss, we incorporate an adversarial loss \mathcal{L}_{adv} to imbue the model with enhanced perceptual realism capabilities. The adversarial component is inspired by generative adversarial networks (GANs) and seeks to improve the model’s ability to indistinguishably blend synthetic textures with real elements. This is achieved through an adversarial setup where the FakeSegment model competes against a discriminator network that attempts to distinguish between the generated probability maps and real annotations. Mathematically, the adversarial loss can be expressed as [3]:

$$\mathcal{L}_{\text{adv}} = \mathbb{E} A \sim \mathcal{A} [\log(D(A))] + \mathbb{E} F \sim G(A) [\log(1 - D(F))], \quad (3)$$

where D is the discriminator network, A represents the input data from the real image domain \mathcal{A} , and F is the fake probability map generated by mapping through G . Together, these loss functions ensure that our FakeSegment model remains not only accurate in detecting synthetic components but also sophisticated in rendering the intricacies of real versus fake distinctions, thereby achieving high levels of perceptual authenticity in visual outputs.

3.4. Dataset Generation

For our experiments, we generate a diverse set of synthetic images using state-of-the-art rendering software. These images encompass various object categories and environmental

conditions to simulate realistic scenarios encountered in practical applications. Additionally, we curate a corresponding set of real-world images for comparative analysis within our framework.

To effectively train our FakeSegment model, it was imperative to create a paired dataset of real and synthetic images, tailored to capture the intricacies of cockpit views during the critical landing phase of a flight. We leveraged a sophisticated 3D model of an airport, crafted using the robust Unity 3D framework, to simulate this complex environment. This virtual model allowed us to generate images that realistically mimic the visual dynamics experienced during landing, focusing particularly on elements such as runway and taxiway layouts. This approach ensured that our synthetic images maintained a high degree of authenticity in terms of both aesthetic appeal and spatial accuracy, providing a reliable basis for the development and refinement of our segmentation model.

To complement these synthetic images, we sourced real images from videos recorded by pilots during actual landings. These videos were collected from the internet. The critical challenge was determining the camera pose relative to the runway, a task we approached systematically. Initially, we employed the MLZ+ algorithm [14] to derive a preliminary trajectory based on the detected boundaries of the runway. This automated estimation served as a foundational step, which was then meticulously refined through manual calibration processes to enhance the precision of both camera pose and rotational parameters. This dual approach enabled us to (a) Original Image (b) Model Image Figure 4: Illustrative examples from the dataset. (c) Label Image obtain accurate camera poses for 2000 images. Correspondingly, for each determined pose, we rendered a synthetic counterpart alongside a semantic segmentation delineating two primary classes: 'runway' and 'taxiway'. Illustrative examples from this dataset are depicted in Figure 4, showcasing the alignment between real and synthetic representations.

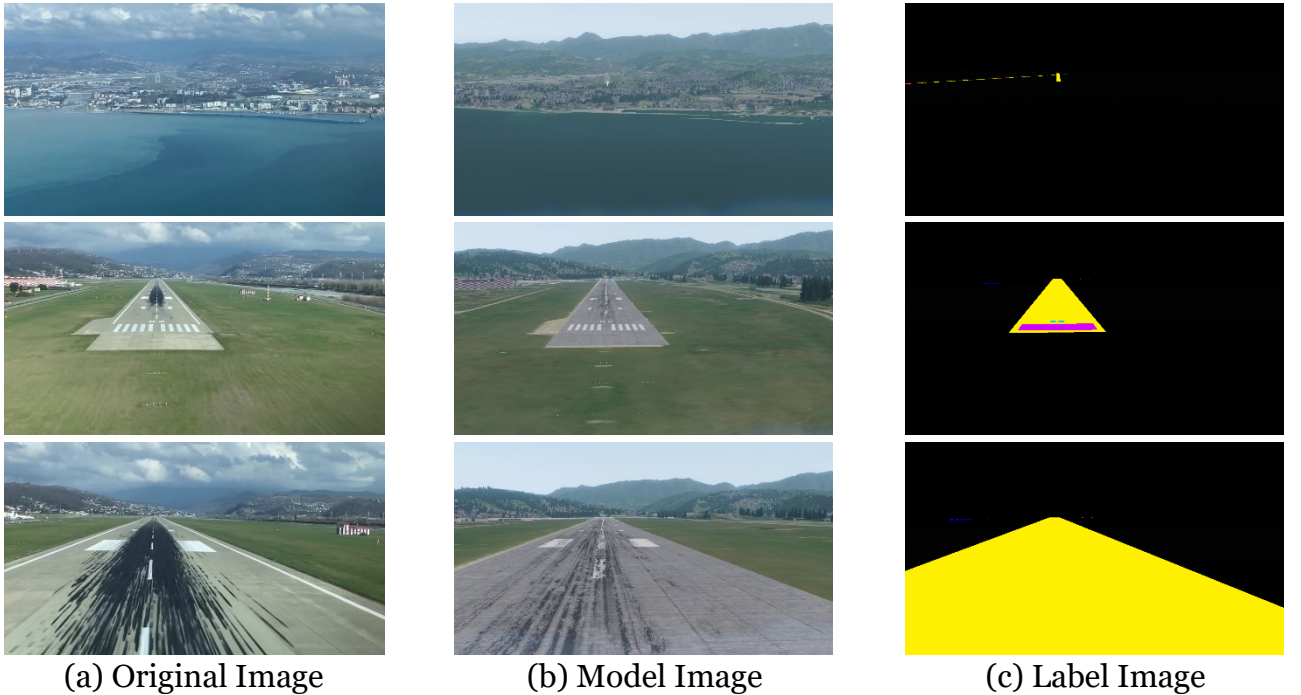


Figure 4: Illustrative examples from the dataset.

3.5. Evaluation

Our evaluation strategy encompasses both qualitative and quantitative assessments to validate the effectiveness of our method in identifying unrealistic details in synthetic images.

3.6. Evaluation Protocol

Our evaluation strategy encompasses both qualitative and quantitative assessments to validate the effectiveness of our method in identifying unrealistic details in synthetic images. We establish a rigorous evaluation protocol that involves comparing feature map activations across multiple CNN architectures trained on both synthetic and real datasets. This protocol ensures that our findings are consistent and generalizable across different network configurations.

In the evaluation phase of our study, we critically assess the performance of our FakeSegment models in comparison with three state-of-the-art baseline models designed for the segmentation of unrealistic or artificially manipulated image regions: ManTraNet [8], MAGritte [7], and CAT-Net [15]. The evaluation protocol employs two core metrics to quantify the accuracy in identifying manipulated regions: the Intersection over Union (IoU) metric and the Dice coefficient. These metrics are widely recognized for their robustness in measuring the overlap and similarity between predicted and ground truth segmentations, thereby providing a comprehensive evaluation of the model’s segmentation capabilities.

Given that the base dataset does not include traditional segmentation-related labels—with real images labeled as entirely ones and synthetic images as entirely zeros—our evaluation methodology addresses this limitation through a dynamic mixing of real and synthetic images in the test dataset preparation. By creating composite images that include both genuine and synthetic regions, we can better evaluate the model’s ability to accurately segment and identify these regions. The labeled real regions are utilized as ground truth for validation purposes, allowing for a precise assessment of how well each model discriminates between real and manipulated content. This dynamic dataset preparation enhances the relevance and applicability of our evaluation protocol, ensuring that it faithfully represents varied scenarios encountered in real-world applications.



Figure 5: Synthetic image enhancement. Top: training phase; bottom: inference phase.

3.7. Qualitative Evaluation

Qualitative evaluation involves visual inspection of highlighted regions within synthetic images where significant activation discrepancies occur. These visualizations provide intuitive insights into specific areas requiring enhancement for improved realism.

In complement to our quantitative assessment, we conduct a qualitative evaluation of the FakeSegment framework against established baselines by examining visual outputs generated during the segmentation process. This evaluation involves generating predictive masks for synthetic images selected from the test split of our dataset and subsequently comparing these masks with the respective ground truth labels, designed to reflect precise contours and delineations of unrealistically rendered regions. Figure 5 illustrates the comparative effectiveness of the various models, providing visual insight into the degree of concordance between predicted segmentation boundaries and their ground-truth counterparts.

The qualitative results presented underscore the superior performance of the FakeSegment model, particularly in terms of its ability to generate labels that closely align with ground-truth expectations. The model exhibits enhanced proficiency in accurately tracing the contours of manipulated regions, such as the unrealistic appearances of these sea and adjacent buildings, which are often challenging for segmentation models due to their complex textures

and reflections. The qualitative analysis reveals that our model not only excels in defining clear boundaries but also displays an impressive capability to identify and isolate areas with unrealistic features, which are more frequently misclassified by the baseline models. This alignment with ground-truth labels reflects the robustness and reliability of FakeSegment in practical applications where precise segmentation is critical.

3.8. Quantitative Evaluation

For quantitative assessment, we measure the degree of alignment between feature map activations from synthetic versus real datasets using statistical metrics such as mean squared error (MSE) and structural similarity index (SSIM). These metrics offer objective measures of how closely aligned the representations are across domains.

The quantitative evaluation of the FakeSegment framework, alongside the evaluated baselines, is centered around its performance on the test split of our SyntheticLanding dataset using two pivotal metrics: the Dice coefficient and Intersection over Union (IoU). As presented in Table 1, the results unequivocally indicate the superior performance of the FakeSegment framework, marking a distinct improvement in both metrics over the competing baseline models. This enhancement in segmentation accuracy is indicative of the model’s advanced capability to align its predicted labels with the ground truth, highlighting its efficacy in detecting and demarcating synthetic or manipulated regions in images. Of the baseline models, ManTraNet emerges as the closest competitor, yet it trails behind FakeSegment in both evaluation metrics. The results are shown in the Table 1.

Table 1. Quantitative evaluation in terms of Dice coefficient.

Class	ManTraNet	MAG	FakeSegment
Runway (R/W)	0.38	0.25	0.51
Background	0.41	0.28	0.49
Average	0.40	0.27	0.50

The comparative analysis reveals that while ManTraNet is adept at identifying small, localized patches indicative of image manipulation—such as those resulting from subtle alterations—it lacks the broader sensitivity required to address the complexities found in synthetic images. This is especially evident when juxtaposing the qualitative results, which show ManTraNet’s propensity to focus on discrete unrealistic patches. In contrast, the FakeSegment framework demonstrates an enhanced sensitivity to both unrealistic textures and 3D graphics artifacts. These include common challenges such as aliasing and the absence of shadows, which are often indicative of artificial image rendering. The ability of our framework to comprehensively capture these subtle yet significant imperfections across entire regions rather than isolated spots underscores its robustness in synthetic segmentation tasks and suggests superior adaptability to diverse forms of image manipulation and generation.

4. Conclusion

In this paper, we introduced the FakeSegment framework, a novel approach designed to enhance the detection of unrealistic details in synthetic images. This framework is constructed by integrating two Single Shot Multibox Detector (SSD) models with shared weights, which extract and refine feature maps, in conjunction with a U-Net architecture that translates these feature maps into precise segmentation of artificially generated regions. Through this method, FakeSegment effectively delineates unrealistically rendered areas in synthetic imagery, illustrating its potential to significantly improve image manipulation detection in both academic and applied settings.

Furthermore, we have developed the SyntheticLanding dataset, leveraging a sophisticated environment simulator to produce a comprehensive collection of 16,000 samples that capture the complexities of the landing stage of flight scenarios. This dataset was instrumental in

training the FakeSegment framework as well as several baseline models. Our subsequent evaluations, which focused on critical metrics such as the Dice coefficient and Intersection over Union (IoU) loss, reveal that the FakeSegment model significantly surpasses baseline models, achieving a remarkable 17% improvement in IoU over ManTraNet, the next best-performing model. These results underscore the efficacy of our approach in detecting synthetic anomalies and pave the way for future advancements in automated image analysis and integrity verification.

Acknowledgments

The research was carried out at the expense of a grant from the Russian Science Foundation No. 24-21-00314, <https://rscf.ru/project/24-21-00314/>

References

1. J. Zhu, T. Park, P. Isola, A. A. Efros, Unpaired image-to-image translation using cycle consistent adversarial networks, in: IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017, IEEE Computer Society, 2017, pp. 2242–2251. URL: <https://doi.org/10.1109/ICCV.2017.244>. doi:10.1109/ICCV.2017.244.
2. J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, D. Liu, Y. Mu, M. Tan, X. Wang, W. Liu, B. Xiao, Deep high-resolution representation learning for visual recognition, IEEE Trans. Pattern Anal. Mach. Intell. 43 (2021) 3349–3364. URL: <https://doi.org/10.1109/TPAMI.2020.2983686>. doi:10.1109/TPAMI.2020.2983686.
3. S. R. Richter, H. A. Alhaija, V. Koltun, Enhancing photorealism enhancement, IEEE Trans. Pattern Anal. Mach. Intell. 45 (2023) 1700–1715. URL: <https://doi.org/10.1109/TPAMI.2022.3166687>. doi:10.1109/TPAMI.2022.3166687.
4. B. F. Labs, S. Batifol, A. Blattmann, F. Boesel, S. Consul, C. Diagne, T. Dockhorn, J. English, Z. English, P. Esser, S. Kulal, K. Lacey, Y. Levi, C. Li, D. Lorenz, J. M'uller, D. Podell, R. Rombach, H. Saini, A. Sauer, L. Smith, FLUX.1 kontext: Flow matching for in-context image generation and editing in latent space, CoRR abs/2506.15742 (2025). URL: <https://doi.org/10.48550/arXiv.2506.15742>. doi:10.48550/ARXIV.2506.15742. arXiv:2506.15742.
5. Z. He, P. Henderson, N. Pugeault, Beyond reconstruction: A physics based neural deferred shader for photo-realistic rendering, ArXiv abs/2504.12273 (2025). URL: <https://api.semanticscholar.org/CorpusID:277824169>.
6. V. Seib, M. Roosen, I. Germann, S. Wirtz, D. Paulus, Generation of synthetic images for pedestrian detection using a sequence of gans, ArXiv abs/2401.07370 (2024). URL: <https://api.semanticscholar.org/CorpusID:266999343>.
7. V. V. Kniaz, V. A. Knyaz, F. Remondino, The point where reality meets fantasy: Mixed adversarial generators for image splice detection, in: H. M. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alch'e-Buc, E. B. Fox, R. Garnett (Eds.), Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada, 2019, pp. 215–226. URL: <https://proceedings.neurips.cc/paper/2019/hash/98dce83da57b0395e163467c9dae521b-Abstract.html>.
8. Y. Wu, W. AbdAlmageed, P. Natarajan, Mantra-net: Manipulation tracing network for detection and localization of image forgeries with anomalous features, in: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019, Computer Vision Foundation / IEEE, 2019, pp. 9543–9552. URL: http://openaccess.thecvf.com/content_CVPR_2019/html/Wu_ManTra-Net_Manipulation_Tracing_Network_for_Detection_and_Localization_of_Image_CVPR_2019_paper.html. doi:10.1109/CVPR.2019.00977.
9. L. Zhang, A. Rao, M. Agrawala, Adding conditional control to text-to-image diffusion models, in: IEEE/CVF International Conference on Computer Vision, ICCV 2023, Paris,

France, October 1-6, 2023, IEEE, 2023, pp. 3813–3824. URL: <https://doi.org/10.1109/ICCV51070.2023.00355>. doi:10.1109/ICCV51070.2023.00355.

10. A. Alexandrescu, R. Petec, A. Manole, L. Diosan, Contrail: A framework for realistic railway image synthesis using controlnet, CoRR abs/2412.06742 (2024). URL: <https://doi.org/10.48550/arXiv.2412.06742>. doi:10.48550/ARXIV.2412.06742. arXiv:2412.06742.

11. S. Wen, J. Ye, P. Feng, H. Kang, Z. Wen, Y. Chen, J. Wu, W. Wu, C. He, W. Li, Spot the fake: Large multimodal model-based synthetic image detection with artifact explanation, CoRRabs/2503.14905(2025). URL: <https://doi.org/10.48550/arXiv.2503.14905>. doi:10.48550/ARXIV.2503.14905. arXiv:2503.14905.

12. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. E. Reed, C. Fu, A. C. Berg, SSD: single shot multibox detector, in: B. Leibe, J. Matas, N. Sebe, M. Welling (Eds.), Computer Vision ECCV2016-14thEuropeanConference, Amsterdam, TheNetherlands, October11-14, 2016, Proceedings, PartI, volume9905ofLectureNotesinComputerScience, Springer, 2016, pp.21–37. URL: https://doi.org/10.1007/978-3-319-46448-0_2. doi:10.1007/978-3-319-46448-0\ _2.

13. E. Schonfeld, V. Sushko, D. Zhang, J. Gall, B. Schiele, A. Khoreva, You only need adversarial supervision for semantic image synthesis, in: 9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021, Open-Review.net, 2021. URL: <https://openreview.net/forum?id=yvQKLqNE6M>.

14. V. V. Kniaz, I. I. Greshnikov, D. E. Tonkikh, A. N. Bordodymov, V. S. Aleksan drov, S. Y. Zheltov, Improving camera exterior orientation estimation using van ishing point detection, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLVIII-2/W9-2025 (2025) 143–150. URL: <https://isprs-archives.copernicus.org/articles/XLVIII-2-W9-2025/143/2025/>. doi:10.5194/ isprs-archives-XLVIII-2-W9-2025-143-2025.

15. M.-J. Kwon, I.-J. Yu, S.-H. Nam, H.-K. Lee, Cat-net: Compression artifact tracing network for detection and localization of image splicing, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2021, pp. 375–384.