УДК 528.727
DOI: 10.25686/978-5-8158-2474-4-2025-514-516

Visualization of unrealistic details in synthetic images using feature maps of a neural network

Victor Aleksandrov[1], Vladimir V. Kniaz[1,2], Pavel Gomozov[1] and Daniil Novikov[1]
[1]State Research Institute of Aviation Systems (GosNIIAS), Moscow, Russia
[2]Moscow Institute of Physics and Technology (MPT), Moscow, Russia

Abstract. The generation of synthetic datasets has become a pivotal tool in training neural models for object detection and recognition tasks. These datasets offer an efficient and scalable solution to the challenges of acquiring large volumes of labeled data. However, synthetic images often suffer from a lack of realistic details, which can significantly degrade the performance of models trained on them. The absence of nuanced features such as texture, lighting variations, and complex backgrounds leads to a gap between synthetic and real-world data, hindering model generalization.

In this paper, we propose a novel method for visualizing and identifying unrealistic details in synthetic images by leveraging feature maps extracted from neural networks. Our approach involves analyzing the activations within convolutional layers to highlight discrepancies between synthetic and real images. By comparing these feature maps, we can pinpoint areas where synthetic images diverge from expected patterns observed in real-world data. This visualization technique provides valuable insights into the specific elements that require enhancement, guiding improvements in synthetic image generation processes.

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

Introduction

The advent of deep learning has revolutionized the field of computer vision, particularly in tasks such as object detection and recognition. A critical component in the success of these models is the availability of large, diverse, and well-annotated datasets. However, collecting and annotating real-world images is a labor-intensive and costly process. To address this challenge, synthetic datasets have emerged as a viable alternative, offering an efficient means to generate vast amounts of labeled data. These datasets are created using computer graphics techniques that simulate various objects and scenes. Despite their advantages, synthetic images often lack the intricate details present in real-world images, which can adversely affect model performance when deployed in practical scenarios.

One of the primary issues with synthetic images is their inability to accurately replicate the complex textures, lighting conditions, and environmental variations found in natural settings. These shortcomings result in a domain gap between synthetic and real data, which neural networks struggle to bridge during training. This gap manifests as a degradation in model accuracy and robustness when applied to real-world tasks. Consequently, there is a pressing need for methods that can enhance the realism of synthetic images or at least identify where these images fall short. Example of visualization of unrealistic details in synthetic images in presented in Figure 1.

In recent years, feature maps within neural networks have been utilized to gain insights into how models perceive input data. These feature maps capture hierarchical representations of images at different levels of abstraction, from simple edges to complex shapes and textures. By examining these activations, researchers can infer which aspects of an image are emphasized by the network during processing. This capability presents an opportunity to analyze synthetic images critically and identify unrealistic details that may contribute to the observed domain gap.

In this paper, we introduce a method for visualizing unrealistic details in synthetic images using feature maps extracted from convolutional neural networks (CNNs). Our approach involves comparing feature map activations between synthetic and real images across various layers of the network. By highlighting discrepancies in these activations, we can pinpoint specific areas where synthetic images deviate from expected patterns seen in real-world data. This visualization technique not only aids in diagnosing deficiencies within synthetic datasets but also provides actionable insights for improving image generation processes. Ultimately, our method aims to facilitate the creation of more realistic synthetic datasets that enhance model performance on object detection and recognition tasks.

Figure 1. Example visualization of unrealistic details in synthetic images

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.

Enhancement of Synthetic Images

Enhancing the realism of synthetic images is crucial for reducing the domain gap between synthetic and real-world datasets. Techniques such as style transfer, domain adaptation, and generative adversarial networks (GANs) have been employed to imbue synthetic images with more naturalistic features. Style transfer methods adjust the texture and color distribution of synthetic images to match those found in real images, while domain adaptation techniques focus on aligning feature distributions across domains. GANs, particularly CycleGANs, have shown promise by learning mappings from synthetic to real image domains without paired examples, thereby enhancing visual fidelity.

Network Training using Synthetic Images

Training deep neural networks with synthetic data offers a cost-effective alternative to collecting large-scale annotated datasets. Several studies have demonstrated that models pre-trained on synthetic datasets can achieve competitive performance when fine-tuned with a smaller set of real images. Transfer learning techniques are often employed to adapt models trained on synthetic data to real-world tasks. However, challenges remain in ensuring that models do not overfit to unrealistic artifacts present in synthetic images. Recent research has focused on developing robust training strategies that mitigate these issues by incorporating domain adaptation layers or adversarial training schemes.

Method

In this section, we introduce our proposed method for visualizing unrealistic details in synthetic images using feature maps from convolutional neural networks (CNNs). Our approach leverages a novel framework named FakeSegment, which systematically analyzes discrepancies between feature map activations of synthetic and real images.

FakeSegment Framework Overview

The FakeSegment framework is designed to identify unrealistic details in synthetic images by comparing their feature map activations against those derived from real images. The framework consists ofseveral stages: preprocessing, feature extraction, activation comparison, and visualization. By focusing on areas where activation patterns diverge significantly, FakeSegment highlights regions within synthetic images that may contribute to the domain gap.

Network Architecture

Our method employs a SSD object detection architecture known for its effectiveness in capturing hierarchical image representations. The network comprises multiple convolutional layers followed by pooling layers, culminating in fully connected layers for classification tasks. We utilize intermediate feature maps from various convolutional layers to perform our analysis, as these maps provide insights into different levels of abstraction within the image.

Loss Function

To optimize our framework's performance, we define a custom loss function that penalizes discrepancies between feature map activations of synthetic and real images. This loss function incorporates both pixel-wise differences and structural similarity metrics to ensure comprehensive evaluation across diferent abstraction levels.

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.

Evaluation

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

Evaluation Protocol

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.

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.

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.

Conclusion

In this paper, we presented a novel approach for visualizing unrealistic details in synthetic images using feature maps from neural networks. Our FakeSegment framework effectively identifies areas where synthetically generated content deviates from realistic patterns observed in natural imagery. By providing actionable insights into these discrepancies, our method facilitates improvements in image generation processes aimed at producing more realistic datasets. Future work will explore extending this approach to other modalities such as video sequences or 3D models while integrating additional machine learning techniques for enhanced analysis capabilities.

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/

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.