Fixing the Synthetic Data for Representational Learning

Aayush Prakash Parmar | Pankaj |Vatsal Trivedi
Advisor: Prof. Shanmuganathan Raman, PhD Student Prajwal Singh
IIT Gandhinagar

Problem Statement

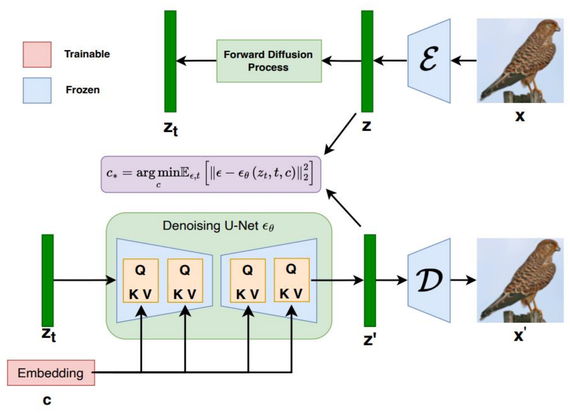
- Deep learning models in computer vision rely heavily on large-scale, high-quality labeled datasets.
- Real-world data collection is expensive, time-consuming, and limited by privacy concerns.
- Synthetic data from GANs and diffusion models offers a scalable alternative but often leads to lower performance in downstream tasks.
- This project investigates why synthetic data underperforms by identifying missing features critical for learning and explores methods to enhance its effectiveness for downstream tasks.

Literature Review

- **Geometric Discrepancies**: Generative models often fail to preserve key geometry, causing artifacts like misaligned vanishing points and inconsistent shadows, which contribute to the domain gap between real and synthetic data .
- **Representation Learning**: While synthetic data provides scalability, it lacks fine details and context, making it less effective for training robust models that require nuanced data .
- **Improvement Techniques**: Methods like Diffusion Inversion condition synthetic data on real image distributions, reducing domain discrepancies and improving classification performance .
- **Average Representation**: Synthetic data tends to focus on average representations, neglecting edge cases, which limits its ability to capture the full diversity of real-world data.

Diffusion Inversion

- Adapts a frozen, pretrained diffusion model using a small set of trainable embedding vectors.
- Embeddings are injected into the U-Net at each timestep during the reverse diffusion process.
- Only the embeddings are optimized using gradient descent; model weights remain unchanged.



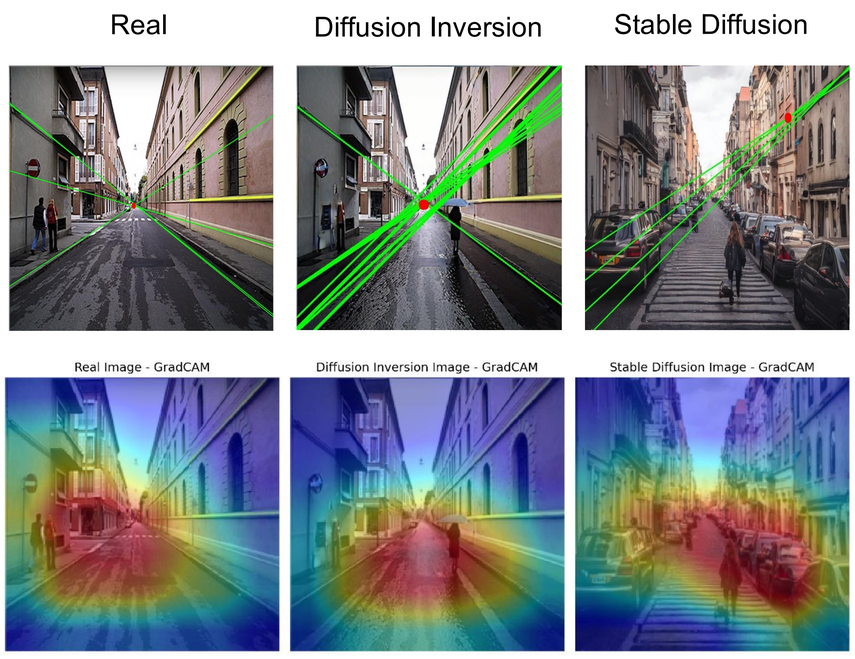
- Minimizes denoising loss between predicted and actual noise to align generations with real data.
- Enables the model to generate more realistic images while preserving its original architecture.

Methodology

- Explored synthetic datasets downstream performance and investigated reasons behind accuracy differences.
- We used CIFAR-10 dataset which consists of 10 classes.
- Used Stable Diffusion and Diffusion Inversion to generate synthetic counterparts of the real dataset.
- Trained a ResNet-18 classifier separately on real, Stable Diffusion, and Diffusion Inversion datasets.
- Compared test accuracies across datasets and used Grad-CAM and projective geometry (vanishing points) to study differences in attention and visual realism.

Visualizing the Realism Gap

Analyzing Projective Geometry



- Vanishing points are key elements in projective geometry, where parallel lines in 3D space appear to converge in a 2D image, providing information about the spatial structure, depth, and orientation of scenes.

Dataset	Focus Score	Attention Spread	Intensity
Real Images	0.439	0.4177	0.439
Fake Inverted Images	0.4361	0.4068	0.4361

- Grad-CAM is a technique used to visualize which areas of an input image highlights important regions for a model's prediction, by generating heat-maps. Warmer colors (red/yellow) in the heat-maps indicate stronger influence.

- Focus Score and Intensity are almost identical, suggesting that the model is equally confident for both image types.
Attention Spread is slightly lower for fake images, indicating that attention is slightly more concentrated in certain regions for fake images. This suggests that fake images relies more on localized features.

Results

Images	Test accuracies (in %)
Real	81.80
Diffusion Inversion	78.63
Stable Diffusion	52.57
Real + Diffusion Inversion	86.14
Real + Stable Diffusion	83.94

- Diffusion Inversion performed better than Stable Diffusion but worse than the real dataset, indicating that while it generates more realistic samples than Stable Diffusion, it still doesn't match the quality of real data.
- The lower performance of Stable Diffusion is likely due to its inability to accurately capture the true data distribution, leading to less useful synthetic samples for training.
- Combining real data with synthetic data (data augmentation) led to higher accuracies, showing that well-crafted synthetic data can enhance model performance when used alongside real examples.

Future Prospects and Applications

- Fine-tune only the later layers of synthetic-trend models to close the accuracy gap with minimal real data.
- Use projective geometry and Grad-CAM feedback to guide better synthetic generation.
- Apply refined synthetic data for domain adaptation in low data areas like medical or autonomous driving.
- Train in privacy sensitive domains using synthetic data to avoid exposing real personal information.

References

[1] 'Shadows Don't Lie and Lines Can't Bend! Generative Models don't know Projective Geometry...for now' https://projective-geometry.github.io/
[2] 'AI-Generated Images as Data Sources: The Dawn of Synthetic Era',https://arxiv.org/pdf/2310.01830
[3] 'Training on Thin Air: Improve Image Classification with Generated Data', https://sites.google.com/view/diffusion-inversion
[4] ' Mind the Gap Between Synthetic and Real: Utilizing Transfer Learning to Probe the Boundaries of Stable Diffusion Generated Data ', https://arxiv.org/pdf/2405.03243