Class-Consistent and Diverse Image Generation Through StyleGAN

Harsh Rangwani[1]* Lavish Bansal[1,3]* Kartik Sharma[1,4] Tejan Karmali[2] Varun Jampani[2] R. Venkatesh Babu[1]

What is the problem?

The paper "NoisyTwins: Class-Consistent and Diverse Image Generation Through StyleGANs" addresses the issue of performance degradation in StyleGANs when trained on large-scale, long-tailed datasets with class conditioning. It introduces "NoisyTwins," an augmentation strategy for class embeddings that prevents collapse of latents, ensuring diverse and consistent image generation across classes.

What has been done earlier?

Use of StyleGANs :

- StyleGANs have been successfully used to generate high-quality, diverse images from well-curated datasets. They produce a semantically disentangled latent space, useful for image editing and manipulation tasks.

Challenges with Class-Conditioning:

- Training StyleGANs with class-conditioning on diverse and in-the-wild datasets has shown to be problematic, especially with long-tailed data distributions. The performance tends to degrade due to issues like mode collapse and poor recall.

Existing Conditioning Techniques:

- arious conditioning and regularization techniques have been implemented, but they either fail to prevent mode collapse or lead to class confusion. This includes techniques from recent StyleGAN variants and other regularized methods.

Ashutosh Panda,B421012

Enhancing Class-Consistent and Diverse Image Generation in StyleGANs on Long-Tailed Datasets

What are the remaining challenges? What novel solution proposed by the authors to solve the problem?

- **Remaining Challenges**:
    - **Mode Collapse**: Particularly severe in minority or tail classes where the generator produces limited varieties of outputs, often leading to repeated, non-diverse images.
    - **Class Confusion**: The generator confuses classes, particularly when classes are numerous and visually or semantically similar, which complicates training on diverse datasets like ImageNet-LT.
- **Novel Solutions Proposed**:
    - **Class Embedding Augmentation**: Introduces noise to the class embeddings to create augmented, but consistent, versions of class information. This helps in preventing the collapse of the generator's latent space by diversifying the input conditions.
    - **Self-Supervised Decorrelation**: Uses a self-supervised technique to decorrelate the latent variables within the generator's architecture. This approach ensures that changes in the noise-augmented class embeddings lead to meaningful and diverse changes in the generated images, enhancing intra-class diversity while maintaining class consistency.

Ashutosh Panda, B421012