Overcoming Few-Shot Challenges in Generative Modelling with Adaptive IMLE

(Theory)

Jaichiddharth Rangasamy Karthigeyan - A20527281
Sreeved Krishna P S - A20525827

Professor Dr. Yan Yan

INTRODUCTION

- In the realm of image synthesis, remarkable strides have been made in the last decade, largely attributed to advancements in deep learning. Traditional GANs struggle due to mode collapse, so Implicit Maximum Likelihood Estimation (IMLE) was proposed. However, IMLE has limitations, leading to the introduction of Adaptive IMLE. This new approach adapts to varying difficulties, improving model quality and establishing a new standard in few-shot image synthesis.

We compare quantitative results with FastGAN and qualitative results with vanilla IMLE

PROBLEM DESCRIPTION

- Our work introduces a generalized Implicit Maximum Likelihood Estimation (IMLE) to address mode collapse in Generative Adversarial Networks (GANs) in few-shot settings.
- This formulation, termed Adaptive IMLE, adapts to varying difficulty levels of training examples and outperforms existing methods in few-shot image synthesis tasks.
- The theoretical guarantees of this approach are valid under broader conditions than vanilla IMLE.

METHODOLOGIES

- **FastGAN (Generative Adversarial Networks)**: A deep learning model that consists of two parts, a generator and a discriminator, which are trained simultaneously. The generator creates images that the discriminator evaluates. The goal is for the generator to produce images so realistic that the discriminator cannot distinguish them from actual images.
- **Vanilla IMLE (Implicit Maximum Likelihood Estimation)**: A technique used in few-shot image synthesis to ensure that generated images resemble all training examples. It addresses the issue of mode collapse in GANs by encouraging diversity and coverage of all modes in the training data.

PROPOSED METHODOLOGY

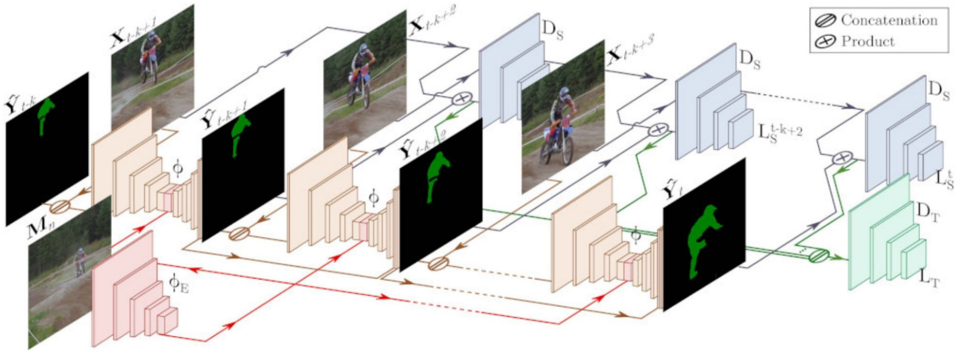
- **Generalized Formulation**: Adaptive IMLE presents a more generalized formulation of Implicit Maximum Likelihood Estimation (IMLE), which holds theoretical guarantees under less restrictive conditions than the original IMLE.
- **Adaptation to Training Examples**: The method adapts to the varying difficulty of different training examples by individualizing the pace at which generated samples converge to each training example.
- **Curriculum Learning Strategy**: It employs a curriculum learning strategy, progressively decreasing the neighborhood radius around data points to maximize the likelihood of the immediate neighborhood.
- **Practical Algorithm**: The Adaptive IMLE algorithm is designed for practical efficiency, utilizing a shared pool of generated samples among all data examples to speed up convergence and reduce computational expense.

DATASET DESCRIPTION

1. **100-shot-obama:**
   - The purpose of this dataset is to classify images associated with Barack Obama. Each class within this dataset has 100 training examples.
2. **100-shot-panda:**
   - It contains images of pandas. It serves as a training dataset for panda image recognition. Each class has 100 training examples.
3. **ffhq (Flickr-Faces-HQ):**
   - The Flickr-Faces-HQ (FFHQ) dataset consists of high-quality images of human faces sourced from Flickr.
   - We use FFHQ for tasks like face generation, style transfer, and facial attribute analysis.
4. **AnimalFace (with cat and dog variations):**
   - This dataset contains animal face images, specifically focusing on cats and dogs.
   - It’s to differentiate between different animal species.

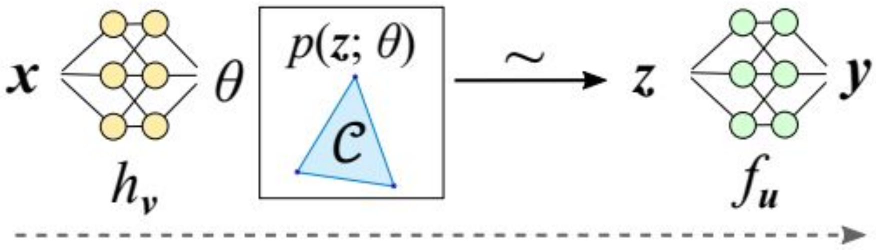
FastGAN Architecture

- **Input**: The process begins with an input image, likely of a person or object that needs to be segmented.
- **Silhouette Extraction**: The architecture extracts the silhouette of the person or object from the original image, focusing on the shape and outline.
- **Layered Refinement**: Through several layers, each denoted as Ds in the image, the architecture refines the details and quality of the generated image. These layers progressively enhance the features and fidelity of the silhouette.
- **Operations**: The image shows various operations such as concatenation and product, which combine features extracted at different stages to improve the final output.
- **Final Output**: The end result is a high-quality generated image that retains essential characteristics of the original input while enhancing certain features, such as the clarity of the silhouette.



IMLE Architecture

- **Input Layer (x)**: The diagram starts with an input layer labeled "x," which represents the initial data that the model receives.
- **Hidden Layer (hv)**: The input data is processed through a hidden layer "hv," which can be thought of as an intermediate processing step that captures the underlying features of the input.
- **Latent Variable Model (p(z; θ))**: At the center of the architecture is the latent variable model, denoted by "p(z; θ)." This model generates latent variables "z" based on the processed data from the hidden layer and the parameters theta (θ).
- **Constraint ©**: Within the latent variable model, there's a constraint "C" that ensures the generated data adheres to certain conditions or characteristics, which is crucial for the model to produce meaningful outputs.
- **Transformation to Output (fu)**: The latent variables "z" are then transformed into the final output data "y" through a transformation function "fu."
- **Output Layer (y)**: The final output layer is labeled "y," where the processed data is presented after being transformed from the latent variables.



Adaptive IMLE

Core Concept

Adaptive IMLE is built upon the foundation of Implicit Maximum Likelihood Estimation (IMLE), which is a method for training generative models.

The key idea behind IMLE is to generate samples that are <u>close to real data points</u>, thus covering all modes of the data distribution.

Challenges in Few-Shot Learning

- **Mode Collapse**: In few-shot settings, where only a limited number of training examples are available, GANs tend to suffer from mode collapse, ignoring some training examples and overfitting to a subset of the dataset.
- **Uniform Optimal Likelihood**: Traditional IMLE assumes that the optimal likelihood at <u>all data points is the same</u>, which is a restrictive condition.

Adaptive IMLE

Adaptive IMLE Solution

- **Few-Shot Image Generation**: It is particularly effective for few-shot unconditional image generation, even with as few as 100 data examples.
- **Pre Training-Free**: Adaptive IMLE can learn from a few samples from scratch without any auxiliary datasets or pretraining.
- **Performance**: Adaptive IMLE significantly outperforms existing methods in terms of image quality and mode coverage.
- **Adaptation to Training Examples**: The algorithm adapts to the varying difficulty of different training examples, pulling generated samples towards training examples at an individualized pace.
- **Mode Coverage**: It ensures that all modes of the data distribution are covered, which is crucial for generating diverse and high-quality results.

EVALUATION METRICS

- The Fréchet Inception Distance **(FID)** to measure the perceptual quality of the generated images, where we randomly generate 5000 images and compute FID between the generated samples and real images in each dataset.
- To evaluate the mode modelling accuracy (**precision)** and coverage **(recall)**, we use the precision metric to measure the former, and use the recall metric and LPIPS backtracking score to measure the latter.
- For **LPIPS** backtracking, we use 90% of the full dataset for training and evaluate the metric using the remaining 10% of the dataset

QUANTITATIVE RESULTS

- Comparison of image quality, as measured by FID between real data and 5000 randomly generated samples from each method.

- Lower FID value is better.
(We choose an input latent dimension of 1024, m = 10000 and a tightening coefficient $\delta = 0.98$. We train our model for less than 200k iterations with a mini-batch size of 4 using the Adam optimizer with a learning rate of $2 \times 10{-}6$ on a single NVIDIA V100 GPU)

	Obama	Panda	Cat	Dog	FHHQ subset
FastGAN	41.1	10.0	35.1	50.7	54.2
Adaptive IMLE	25.0	7.6	24.9	43.0	33.2

QUANTITATIVE RESULTS

- Precision and recall are computed for 1000 randomly generated samples and the target dataset.
- Higher precision shows better fitting to the target dataset, while higher recall corresponds to better mode coverage.

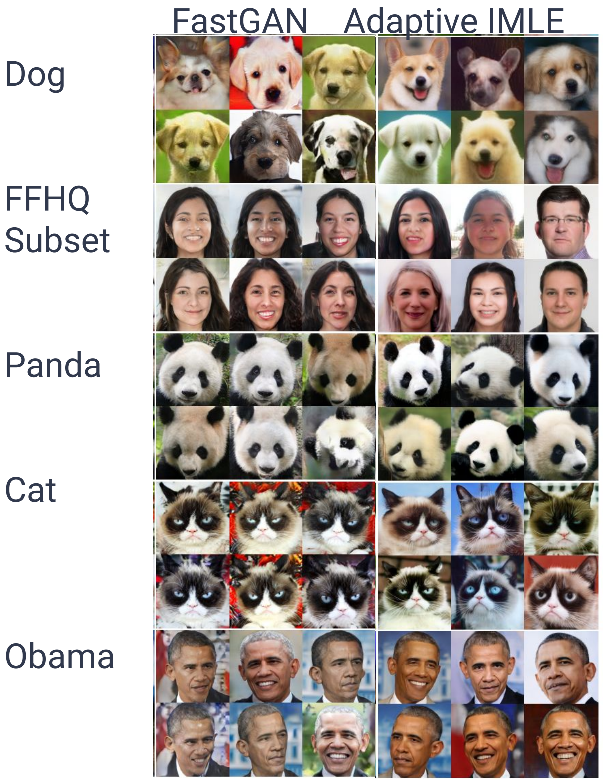
	Obama		Panda		Cat		Dog		FHHQ subset	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
FastGAN	0.92	0.09	0.96	0.16	0.97	0.08	0.96	0.19	0.91	0.13
Adaptive IMLE	0.99	0.68	0.98	0.63	0.98	0.86	0.97	0.61	0.99	0.77

QUANTITATIVE RESULTS

- LPIPS below represents LPIPS backtracking score (Liu et al., 2021). For this metric, each model is trained on 90% of the dataset.
- The resulting model is used to backtrack in the latent space and reconstruct the remaining 10%.
- Lower LPIPS backtracking score shows better mode coverage of the training data.

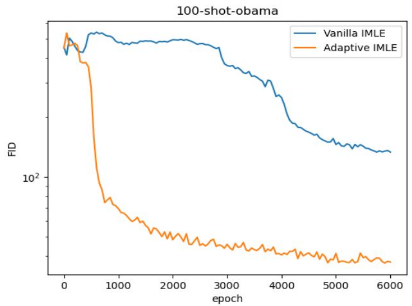
	Obama	Panda	Cat	Dog	FHHQ subset
FastGAN	0.370	0.339	0.467	0.430	0.357
Adaptive IMLE	0.036	0.039	0.074	0.072	0.014

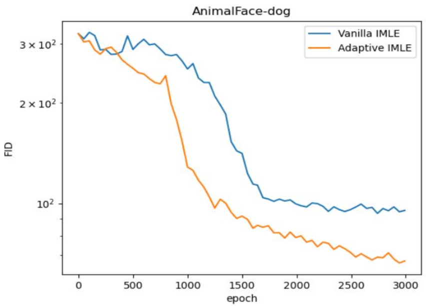
QUANTITATIVE RESULTS



- Quantitative comparison of images generated by our method and the baseline, FastGAN.
- Our method shows higher sample quality and diversity.
- In contrast, the samples generated by the baseline exhibit distortions and limited diversity, as supported by the results of precision and recall.
- These results indicate that the baselines suffer from spurious modes or mode collapse.

QUALITATIVE RESULTS





FID comparison of Adaptive IMLE and Vanilla IMLE during training on the **Obama and Dog** datasets. Adaptive IMLE achieves better image quality (lower FID) and faster convergence rate compared to vanilla IMLE.

- FID comparison between vanilla IMLE and the proposed method, Adaptive IMLE.
- Adaptive IMLE consistently outperforms vanilla IMLE, demonstrating better generated image quality.

	Obama	Panda	Cat	Dog	FHHQ subset
Vanilla IMLE	37.4	8.2	34.4	61.9	54.1
Adaptive IMLE	25.0	7.6	24.9	43.0	33.2

CONCLUSION

- **Superior Performance**: Adaptive IMLE Demonstrated significant improvements over FastGAN and IMLE in image quality and mode coverage across six few-shot benchmark datasets, establishing a new state-of-the-art for few-shot image synthesis without the need for pre-training on auxiliary datasets.
- **Generalized IMLE**: Adaptive IMLE is a more generalized formulation of Implicit Maximum Likelihood Estimation (IMLE) that holds under **less restrictive** conditions, enabling **adaptation to training examples** with varying difficulty levels.

REFERENCES

[1] Brock, A., Donahue, J., and Simonyan, K. Large scale GAN training for high fidelity natural image synthesis. arXiv preprint arXiv:1809.11096, 2019.

[2] Child, R. Very deep VAEs generalize autoregressive models and can outperform them on images. arXiv preprint arXiv:2011.10650, 2021.

[3] Dhariwal, P., and Nichol, A. Diffusion models beat GANs on image synthesis. arXiv preprint arXiv:2105.05233, 2021.

[4] Dinh, L., Sohl-Dickstein, J., and Bengio, S. Density estimation using Real NVP. arXiv preprint arXiv:1605.08803, 2017.

[5] Finn, C., Abbeel, P., and Levine, S. Model-agnostic meta-learning for fast adaptation of deep networks. In Proc. of the 34th International Conference on Machine Learning-Volume 70, pages 1126–1135. JMLR. org, 2017

THANK YOU