TRAINING DIFFUSION MODELS WITH FEDERATED LEARNING

Matthijs de Goede M.B.J.degoede@student.tudelft.nl

Responsible Professor: Jérémie Decouchant
Supervisor: Bart Cox

TU Delft

1. BACKGROUND

Fig 1: Diffusion Models
Gradually add noise to images and learn the reverse process.

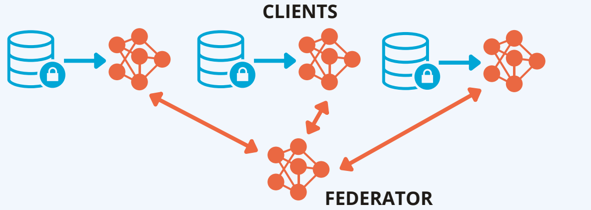

Fig 2: Federated Learning
Collaborative training of ML models without sharing raw data.

2. MOTIVATIONS

- Strong dependency on Big Tech for image generation models
- Privacy, data authority, and transparency issues
- Data variety among clients exploitable to build better models
- High communication overhead in collaborative training

3. OBJECTIVES

- Train Diffusion Models using Federated Learning
- Maximize image quality
- Minimize communication overhead
- Robustness against Statistical Heterogeneity

4. METHODOLOGY

Communication Efficient Training

Exploit UNet Architecture to reduce number of communicated parameters

FULL $\mathcal{O}(R \cdot K \cdot 2|\theta|)$
Send full set of parameters back and forth between federator and K clients

USPLIT $\mathcal{O}(R \cdot K \cdot \frac{3}{2}|\theta|)$
Federator sends full set of parameters, K clients send complementary updates

UDEC $\mathcal{O}(R \cdot K \cdot |\theta_{\mathrm{dec}}|)$
Train only the decoder in federated way, encoder and bottleneck locally

ULATDEC $\mathcal{O}(R \cdot K \cdot |\theta_{bot} \frown \theta_{dec}|)$
Train the decoder and bottleneck in federated way, encoder locally

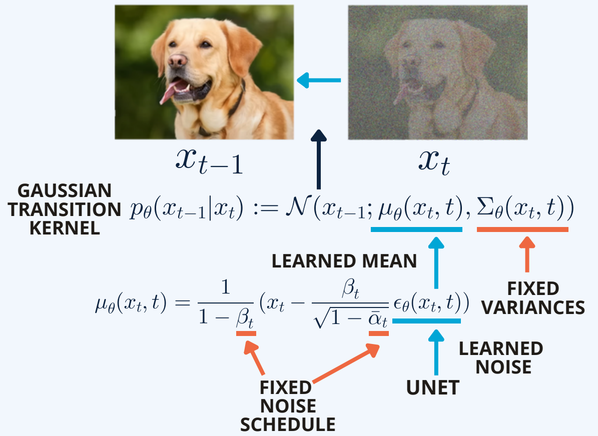

Fig 3: DDPM
Denoising Diffusion Probabilistic Model [1] that models the transition kernels of the reverse diffusion process as Gaussians.

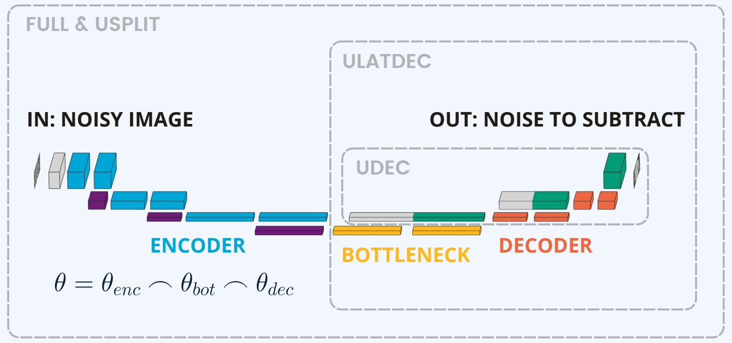

Fig 4: UNet
Convolutional Neural Network with Encoder, Decoder and Bottleneck to predict the noise added to a sample at a specific timestep [2].

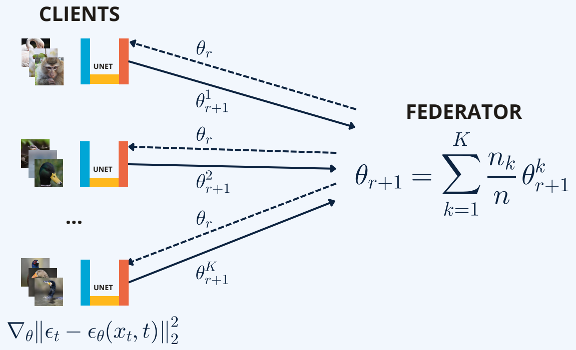

Fig 5: FedDiff
Federated Averaging [3] of UNet parameters based on client dataset size. Clients perform batch SGD on local dataset to update their parameters.

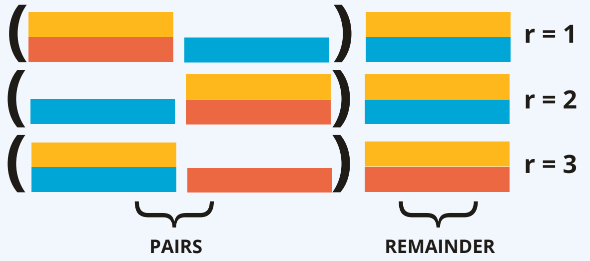

Fig 6: USplit Task Assignment
Every round, the federator splits the tasks for reporting about network parts complementarily within random client pairs.

5. EXPERIMENTS AND RESULTS

Experimental Setup

- Fashion-MNIST Dataset (28x28 grayscale)
- Communication efficiency measured by Cumelative Number of Communicated Parameters
- Image quality measured by Fréchet inception distance (FID) [4]
- Dirichlet Distribution (β = 0.5) to introduce label distribution skew (l-skew) and quantity skew (q-skew) [5]

Centralized Baseline

- Sufficient quality after 10 rounds FID <= 72
- Minor improvement after 15 rounds FID = 43

Full Federated Learning

- Considered scenarios with 2, 5 and 10 clients
- Local epochs increased to 5 to achieve comparable results

Comparing Training Methods

- With 2, 5 and 10 clients on IID, l-skewed and q-skewed data (Tab 1)
- With 5 clients and IID data, all produced good looking images (Fig 7)

METHOD	#CLIENTS	#PARAMS (M)		FID SCORE		
				IID	L-SKEW	Q-SKEW
BASELINE	1	0		43 ± 1	n/a	n/a
FULL	2	179.78		39 ± 2	33 ± 1	33 ± 3
	5	449.45		39 ± 4	43 ± 4	23 ± 5
	10	898.89		61 ± 2	64 ± 3	76 ± 11
USPLIT	2	134.81		37 ± 3	38 ± 4	55 ± 4
	5	243.73	−25%	41 ± 5	61 ± 5	39 ± 9
	10	674.17		62 ± 3	70 ± 8	87 ± 19
ULATDEC*	2	105.50		45 ± 13	49 ± 4	54 ± 24
	5	263.75	−41%	53 ± 15	72 ± 30	122 ± 138
	10	527.51		70 ± 14	101 ± 83	137 ± 125
UDEC*	2	47.54		49 ± 16	49 ± 5	78 ± 48
	5	118.85	−74%	51 ± 15	75 ± 31	139 ± 135
	10	237.69		72 ± 20	94 ± 67	147 ± 119

Tab 1: Communication efficiency and image quality
Communication efficiency and image quality in different federated settings, using 15 communication rounds and 5 loal epochs. The * denotes that the FID scores have been averaged over all local client models. FID scores that exceed 72 within one standard deviation are marked in orange.

Fig 7: Fashion-MNIST samples
Fashion-MNIST samples generated with each of the training methods using 5 clients, IID data, 15 communication rounds and 5 local epochs.

6. CONCLUSIONS

- Diffusion models can successfully be trained using FedDiff with multiple local epochs
- FULL federated training robust to q-skew and l-skew with a limited number of clients
- USPLIT reduces communication by 25% but limits the robustness against non-IID data
- UDEC reduces communication by 74% but leads to variations in image quality among local client models. Works only in conjunction with a limited number of clients and IID data
- Comparable image quality for all methods with IID data and a limited number of clients

7. FUTURE WORK

- Design federated solutions for diffusion models based on Stochastic Differential Equations (SDEs) and Stochastic Score-Based Generative Models (SGMs)
- Improve robustness against non-IID data with alternative aggregation methods
- Establish theoretical bounds for the convergence of our methods
- Integrate Latent Diffusion Models (LDMs) to work with higher resolution datasets

REFERENCES

[1] Ho et al., "Denoising Diffusion Probabilistic Models," NeurIPS 2020.
[2] McMahan et al., "Communication-Efficient Learning of Deep Networks from Decentralized Data," AISTATS 2017.
[3] Ronneberger et al., "U-net: Convolutional networks for biomedical image segmentation," MICCAI 2015.
[4] Heusel et al., "Gans trained by a two time-scale update rule converge to a local nash equilibrium," NeurIPS 2017.
[5] Yurochkin et al., "Bayesian nonparametric federated learning of neural networks," ICML 2019.