

DP Training & Synthetic Data via DP Clustering

Felix Zhou, Samson Zhou, Vahab Mirrokni, Alessandro Epasto, Vincent Cohen-Addad

Yale



Differential Privacy

A (randomized) algorithm $\mathcal{A}$ is (ϵ, δ) -**differentially private** if for all neighboring datasets D and D' and all sets of possible outputs Y :

$$e^{-\epsilon} - \delta \leq \frac{\Pr[\mathcal{A}(D) \in Y]}{\Pr[\mathcal{A}(D') \in Y]} \leq e^{\epsilon} + \delta$$

How to DP-fy ML Pipeline?

A1: DP-SGD

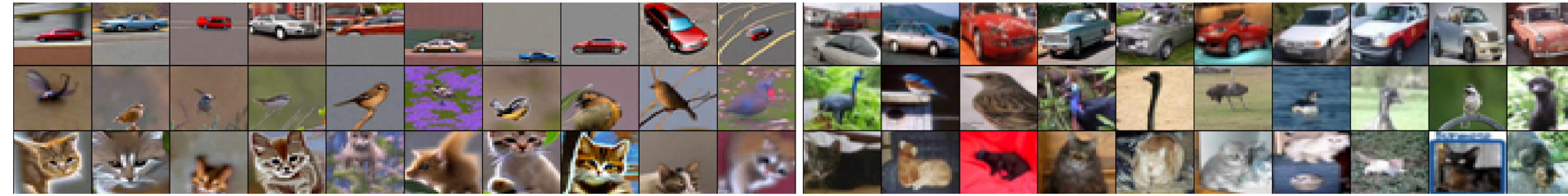
- Access data through noisy gradients

A2: DP Synthetic Data

- Use “approximate” synthetic dataset
- Any (non-private) training is allowed!

Benefits of Synthetic Data

- Exploratory data analysis without privacy loss
- Avoid changes to (complex) existing codebases
- Train many models for more epochs



Example: CIFAR-10 (Above+Right)

- Left: Synthetic CIFAR-10 data
 - $\epsilon = 8, \delta = 1e - 5$
- Right: CIFAR-10 training data
- Each row is a single class

Results: Private Training (Below)

Task: Train DP image classifier

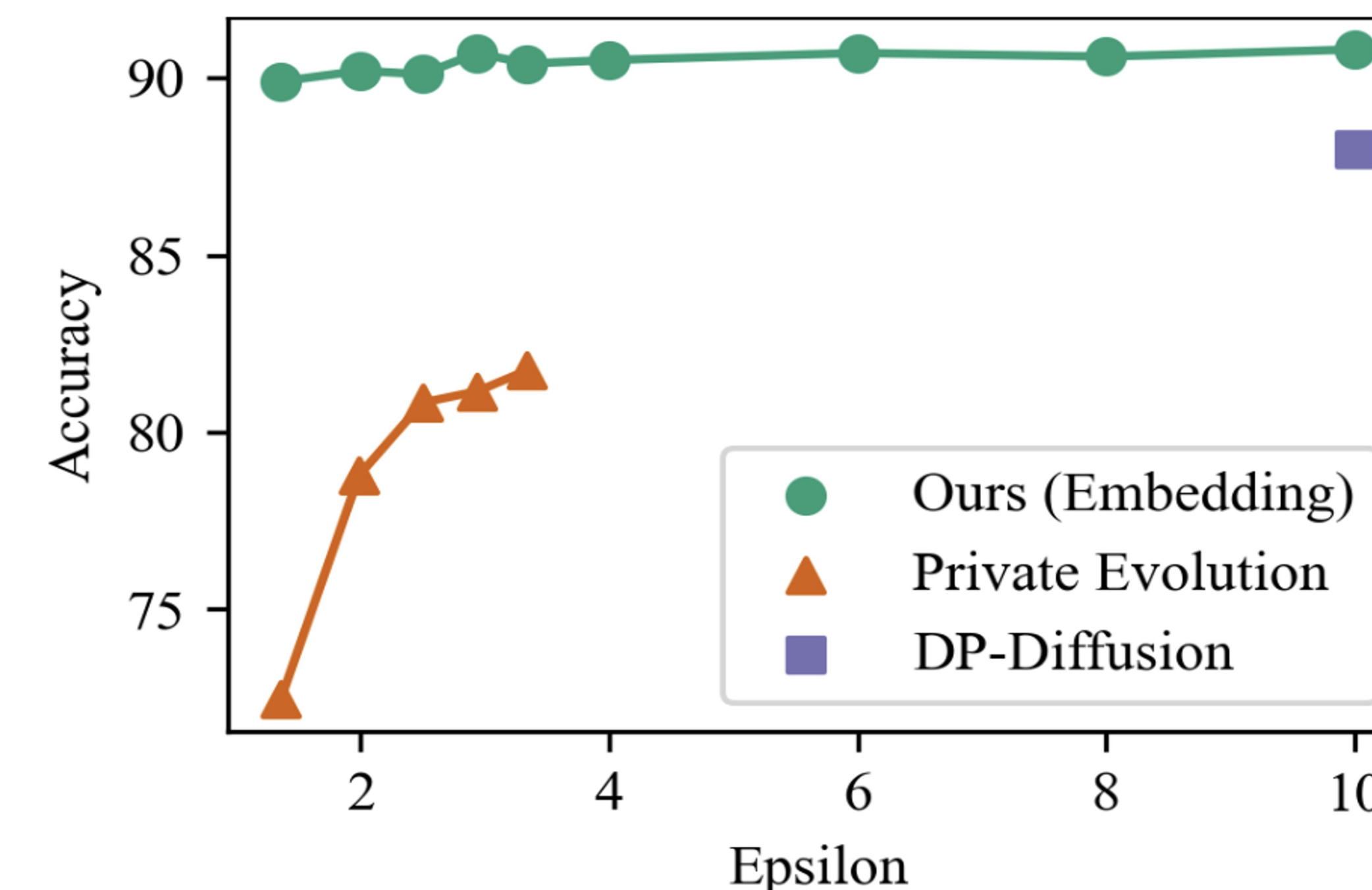
- Generate DP synthetic **embeddings**
- Train 2-layer neural network on embeddings
- Test on embeddings of original test set
- **Better** results than DP-finetuning baseline!
 - [DBH+22]

DATASET	ϵ	SOTA	OURS
CIFAR-10	8	96.6	97.0 ± 0.01
CIFAR-100	8	81.8	80.5 ± 0.177
CAMELYON17	10	91.1	93.1 ± 0.067

Results: Synthetic Data (Below)

Task: Generate DP synthetic images

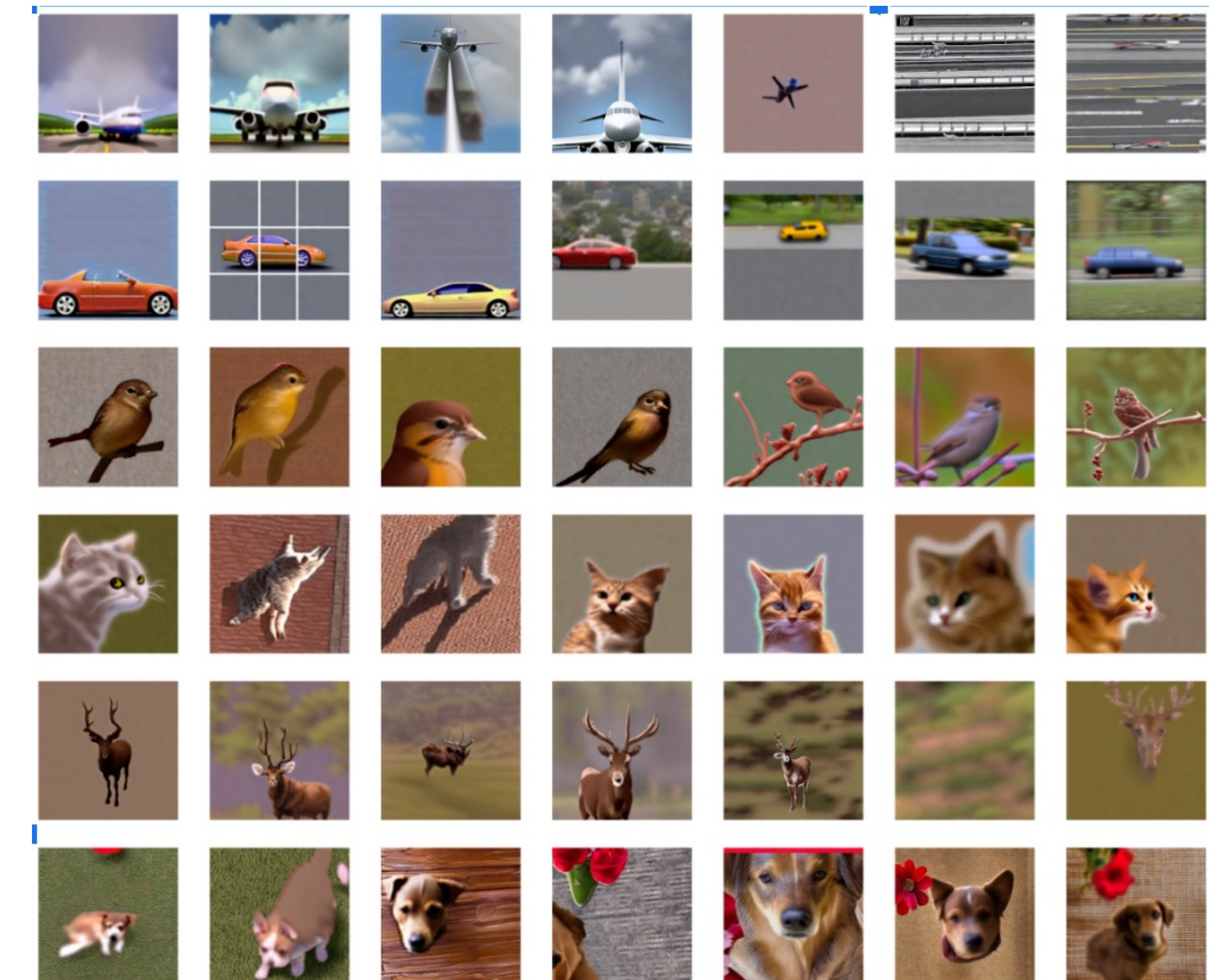
- Generate DP synthetic **images**
- Train ResNet50 on images
- Test directly on original test set
- **Better** results than DP synthetic data baselines!
 - Private evolution [LGK+24]
 - DP-Diffusion [GBG+23]



Methodology

Novel Approach: DP Gaussian Mixture Model on embeddings

1. **Encode** the images to embedding space
2. Run **DP Clustering** on embeddings to obtain centers (**means**)
3. Privately estimate covariance per cluster
4. **Sample** (unlimited) new “embeddings” from Gaussian mixture model
5. **Decode** “embeddings” into images



Work done while Felix Zhou was a student researcher at Google