



Speakers: Elvi Mihai, Fabian Gobet,
Pietro Miotto, Yassine Oueslati



Course: Advanced Topics in Machine Learning,
MSc AI

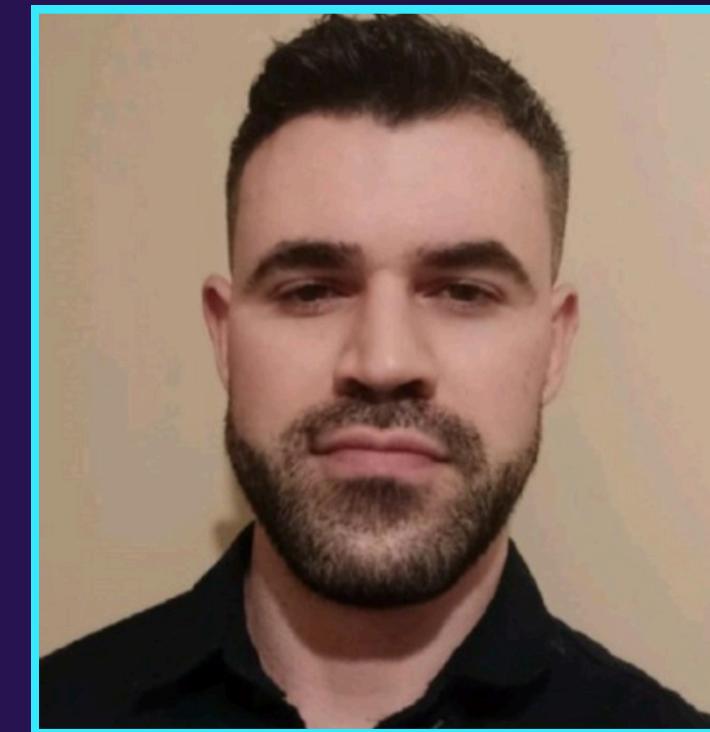
Deep Generative Clustering with Multimodal Diffusion Variational Autoencoders

ATML 01 - Reproducibility Challenge

The Team



Pietro Miotto



Fabian Gobet



Elvi Mihai Sabau

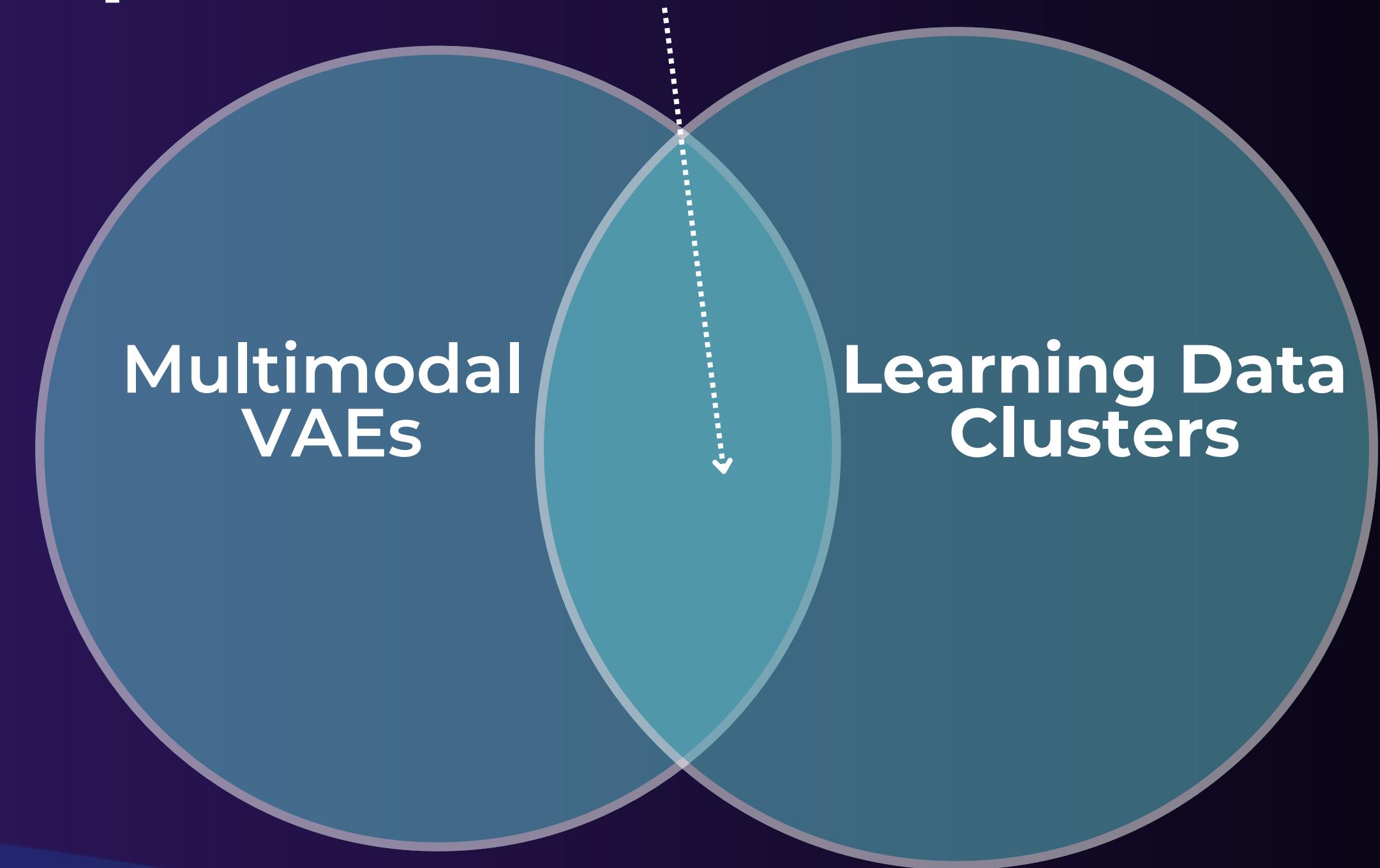


Yassine Oueslati



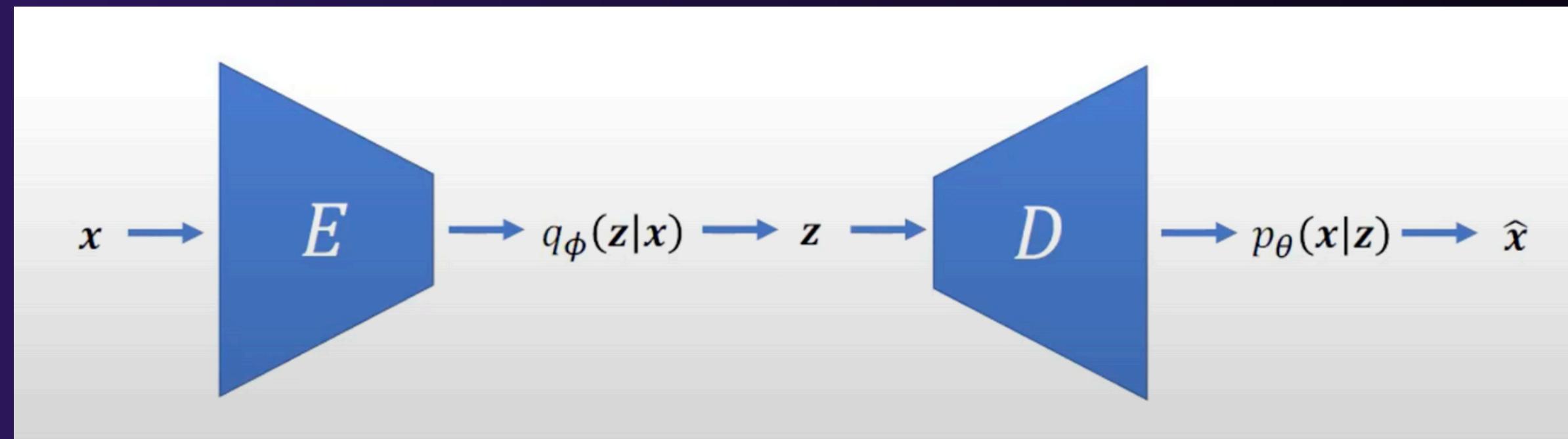
The Objective

Deep Generative Clustering



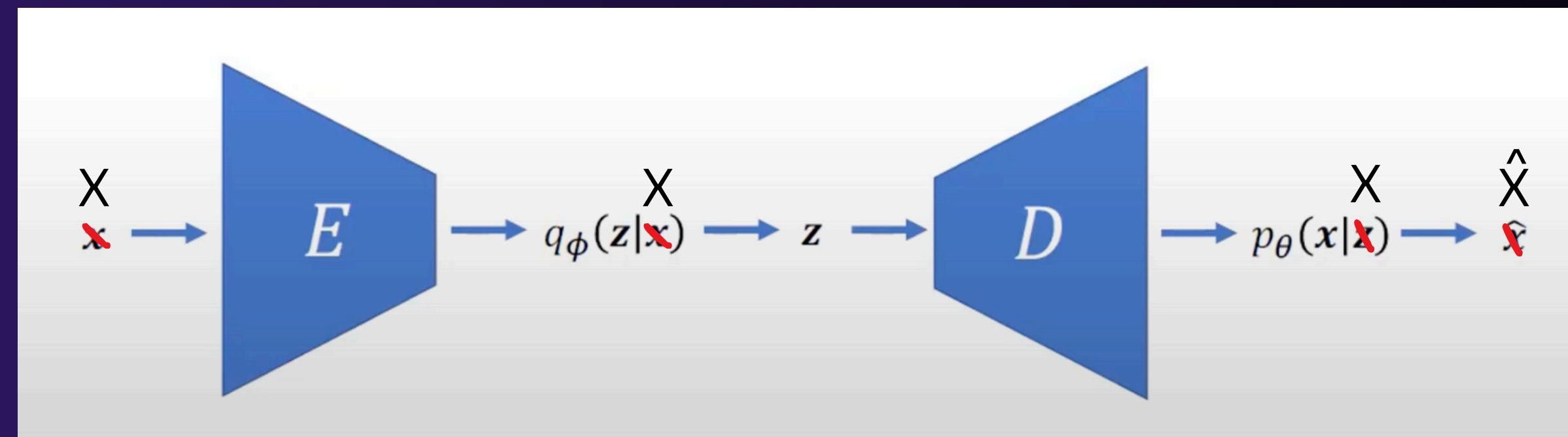
Multimodal VAEs

Unimodal VAE



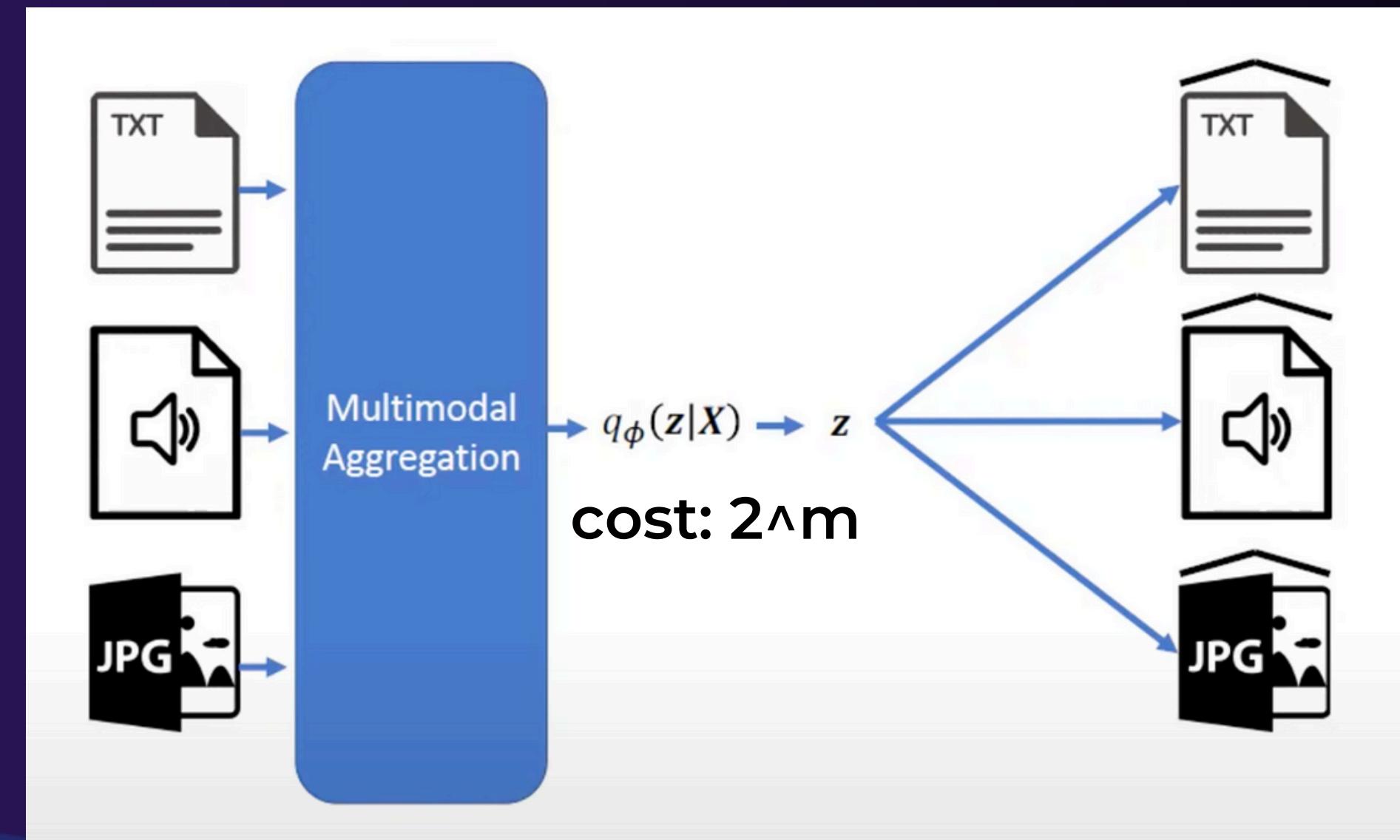
Multimodal VAEs

MultiModal VAE



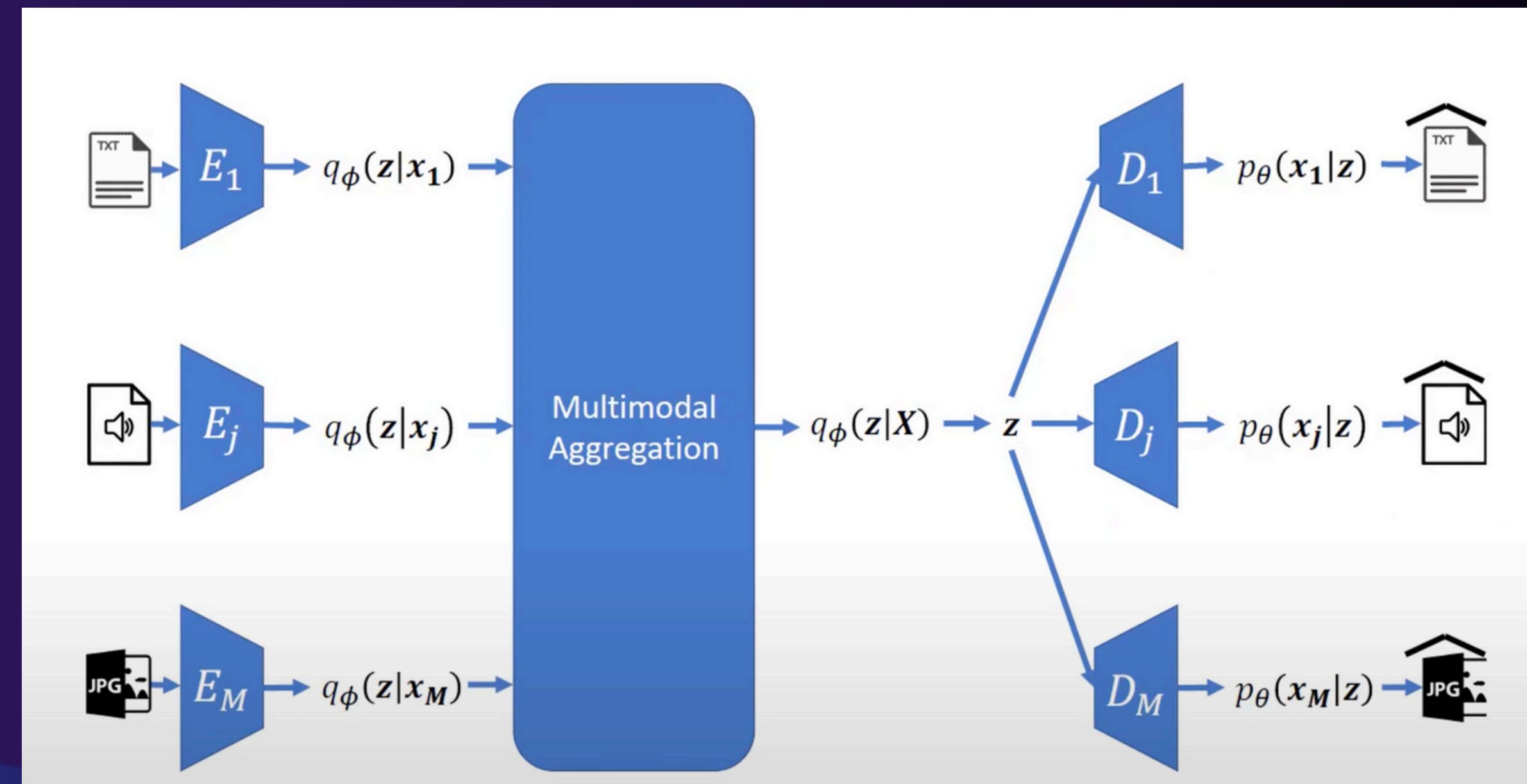
Multimodal VAEs

Scalability



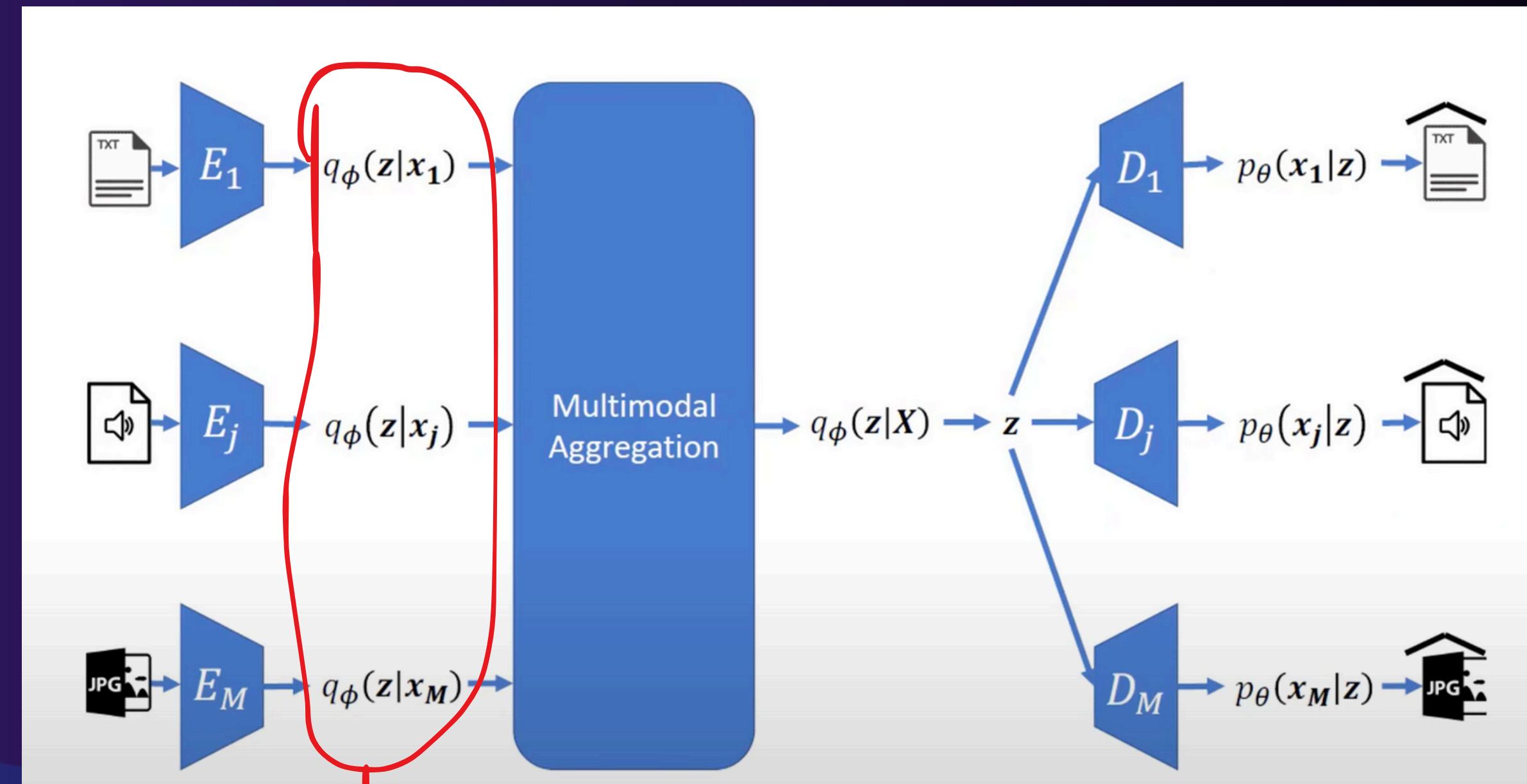
Multimodal VAEs

Merging modality specific encoders outputs



Multimodal VAEs

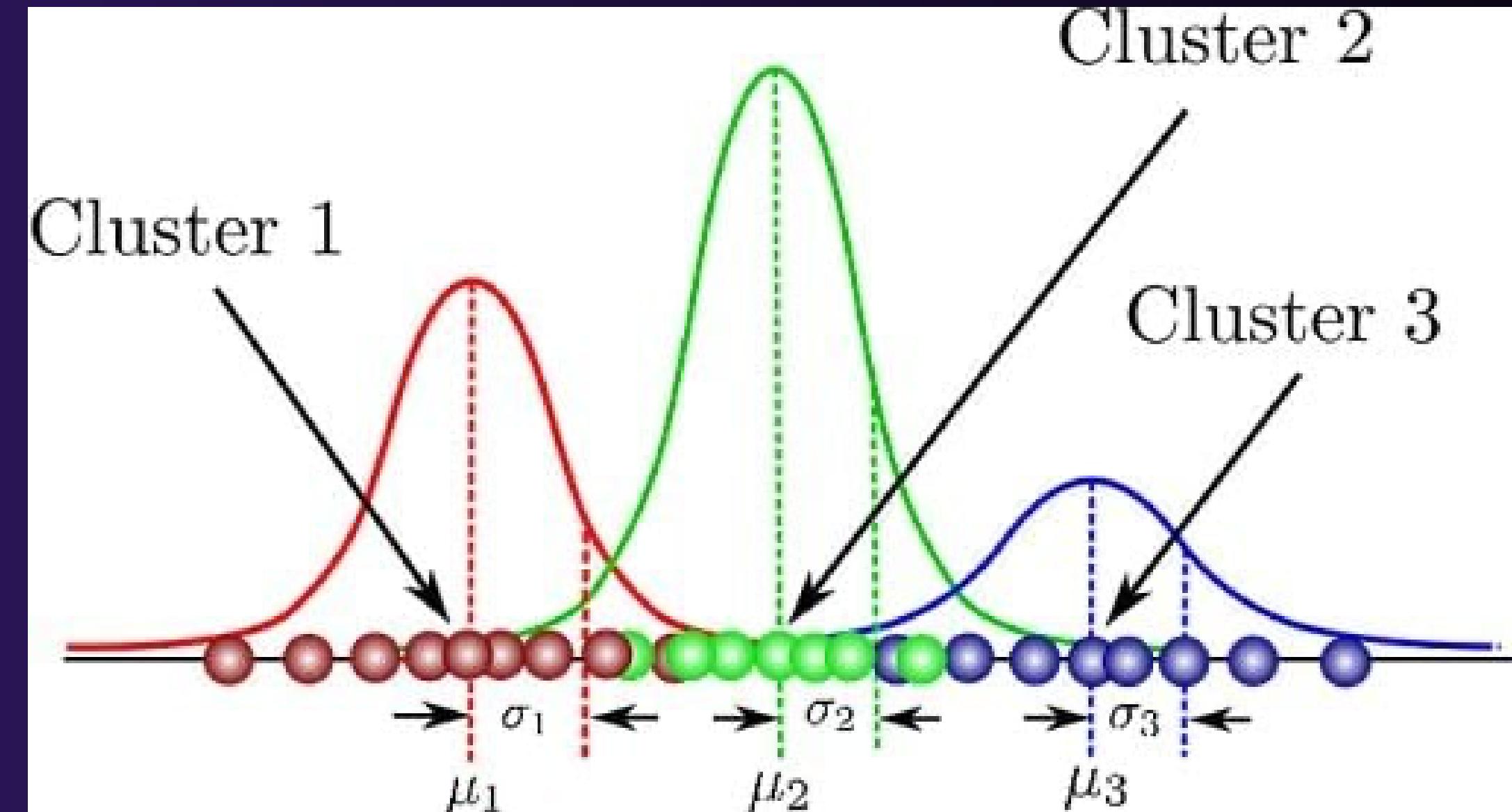
Importance of modality specific subspaces



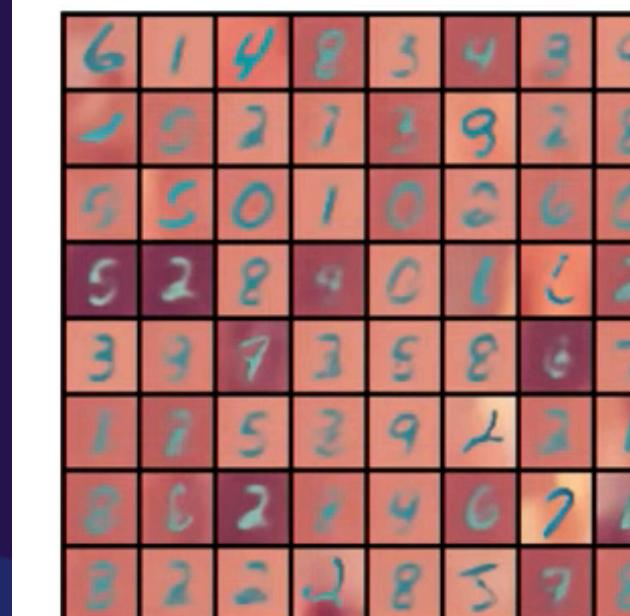
We integrate this modality-specific subspaces

Learning Data Clusters

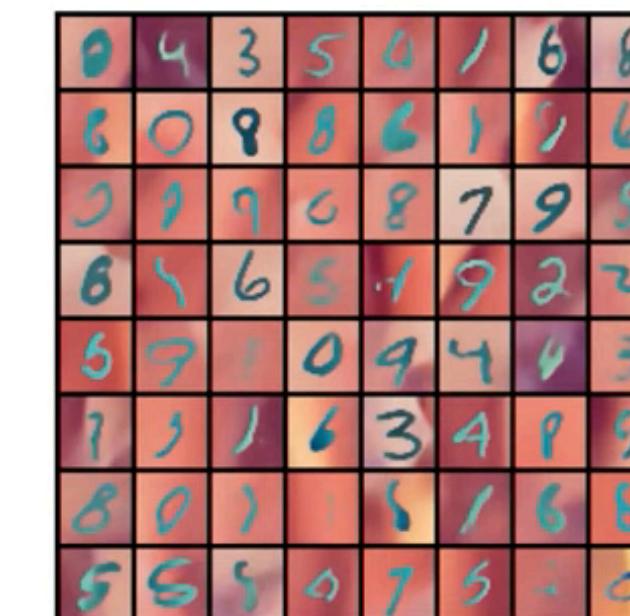
What does it mean to learn clusters in latent space?



Quick Look to the Datasets: PolyMNIST

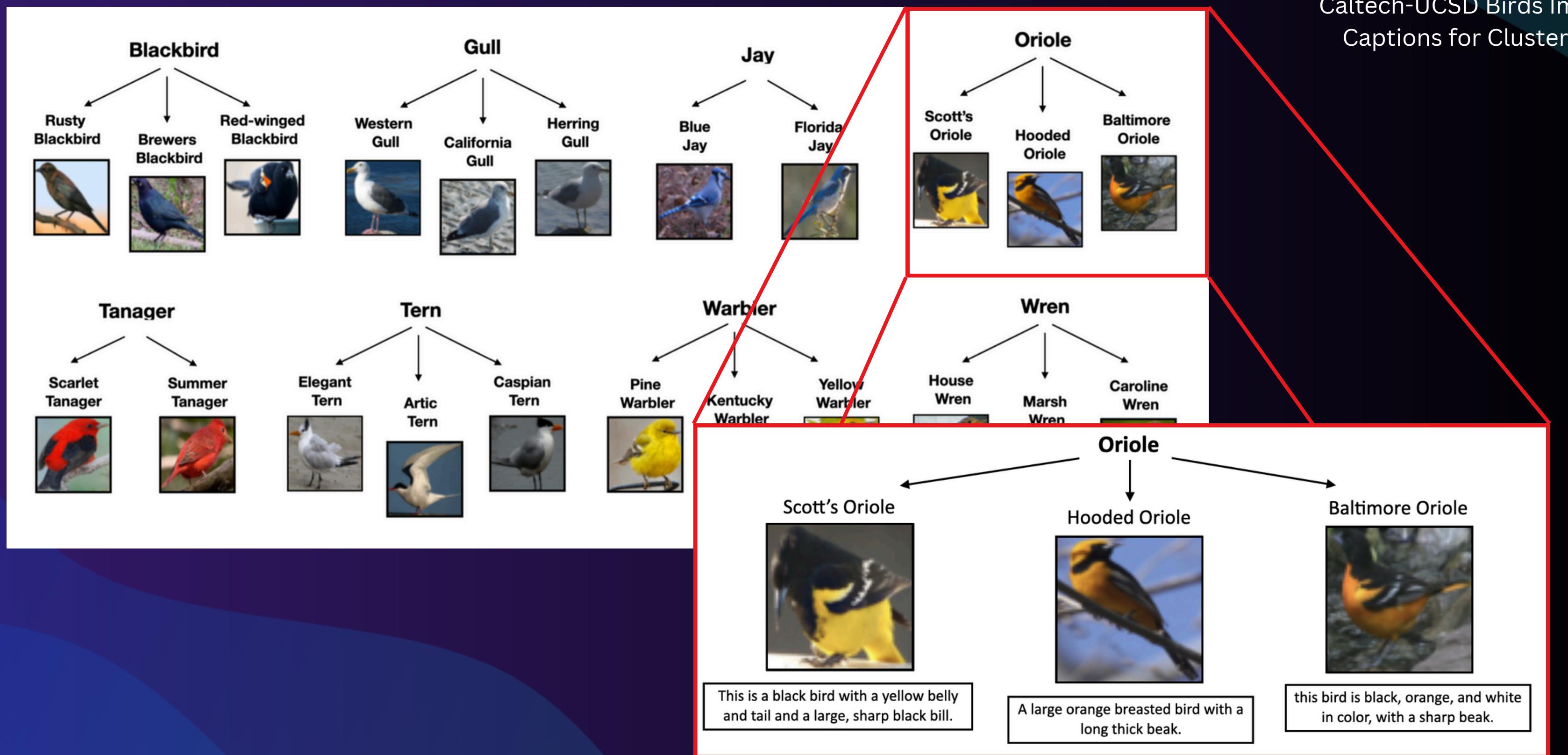


(a) MMVAE with a single joint latent space



(b) MMVAE with additional modality-specific subspaces

Quick Look to the Datasets: CUBICC

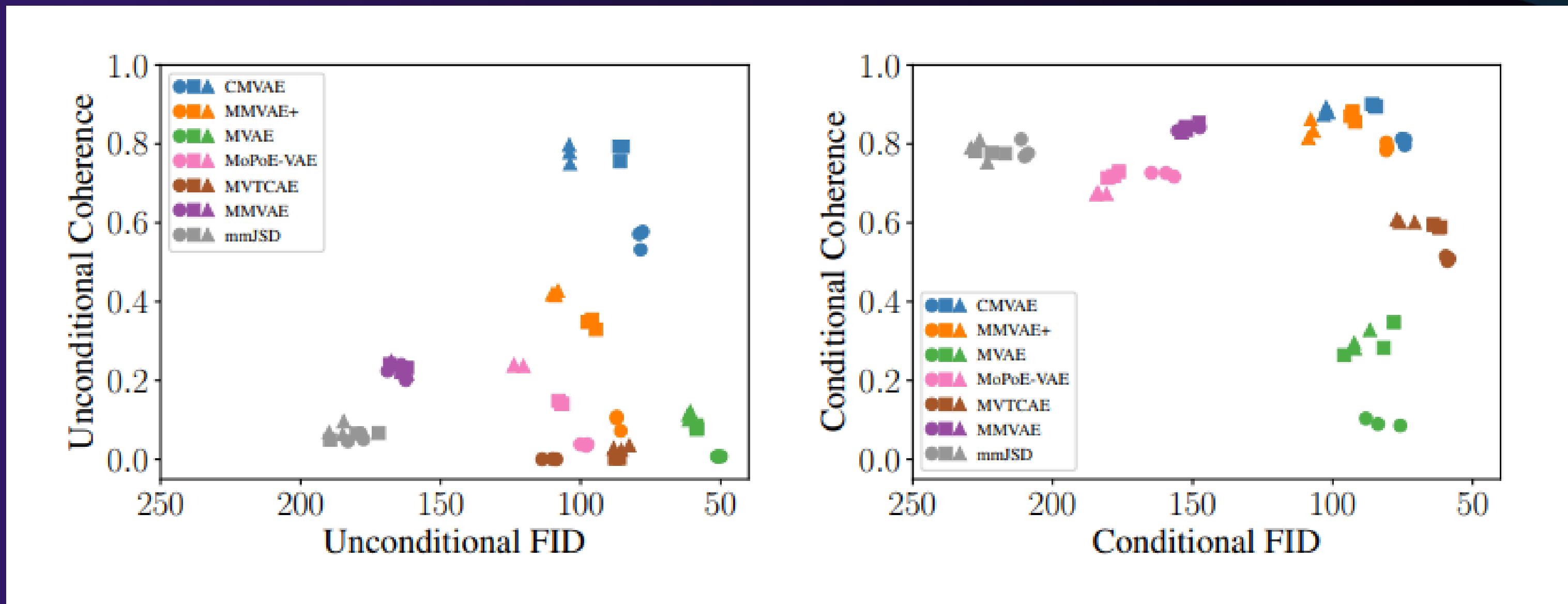


Caltech-UCSD Birds Image-Captions for Clustering

Claims



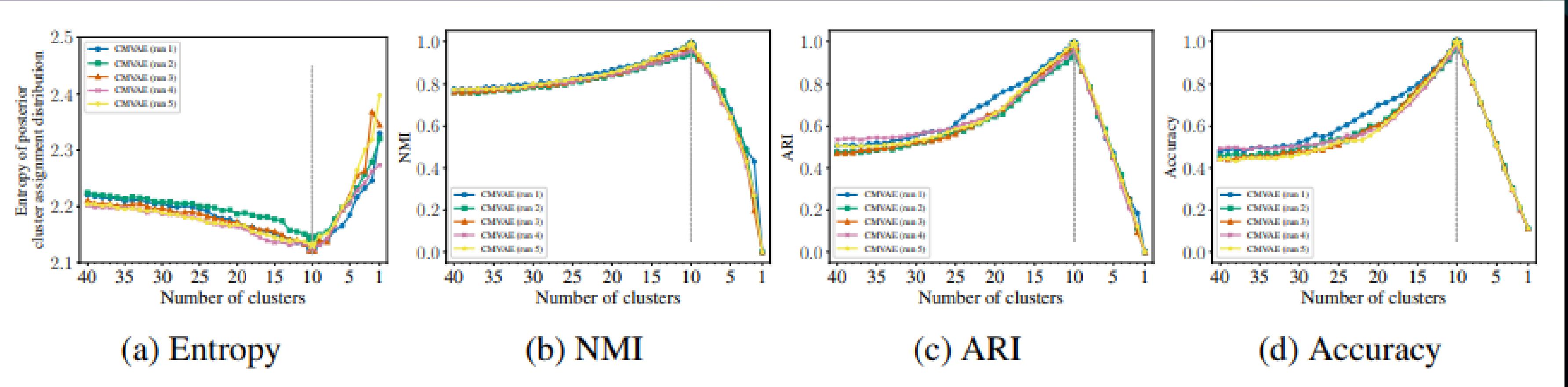
Claims



Claim 1. Novel Model for Multimodal Clustering:

The paper introduces the Clustering Multimodal VAE (CMVAE), which enhances the clustering capability by enforcing a mixture distribution in the shared latent space. It demonstrates superior performance from other models in unconditional generation and weakly-supervised clustering compared to existing methods.

Claims



(a) Entropy

(b) NMI

(c) ARI

(d) Accuracy

Claim 2. Post-Hoc Clustering Optimization:

CMVAE employs an innovative post-hoc procedure to infer the optimal number of clusters at test time, overcoming the necessity to define the number of clusters during training. This approach reduces entropy in cluster assignments and ensures effective modeling of latent clusters.

Claims

Claim 3. Integration of Diffusion Models:

The paper describes how incorporation of Denoising Diffusion Probabilistic Models (DDPMs) could improve the generative performance of real-world multimodal data.

Overview of Current Training Procedures and Results



Key Concepts going forward...

- **Conditional generation:** Data generated given a pre-condition (input)
- **Unconditional generation:** Data generated without conditions (no input)
- **Fréchet inception distance (FID):** Measures image quality similarity
- **Coherence:** learned latent space meaning/interpretability vs. real-world concepts
- *Clustering metrics:*
 - **Adjusted Rand Index (ARI):** Measures clustering agreement accuracy
 - **Normalized Mutual Information (NMI):** Measures clustering mutual dependence
 - **Accuracy (ACC):** Measures clustering match accuracy
- *Parameters:*
 - **K:** Number of samples for k-reparametrization trick
 - **β:** Balances generation vs coherence by weighting D_{KL} term

Overview of Current Training Procedures

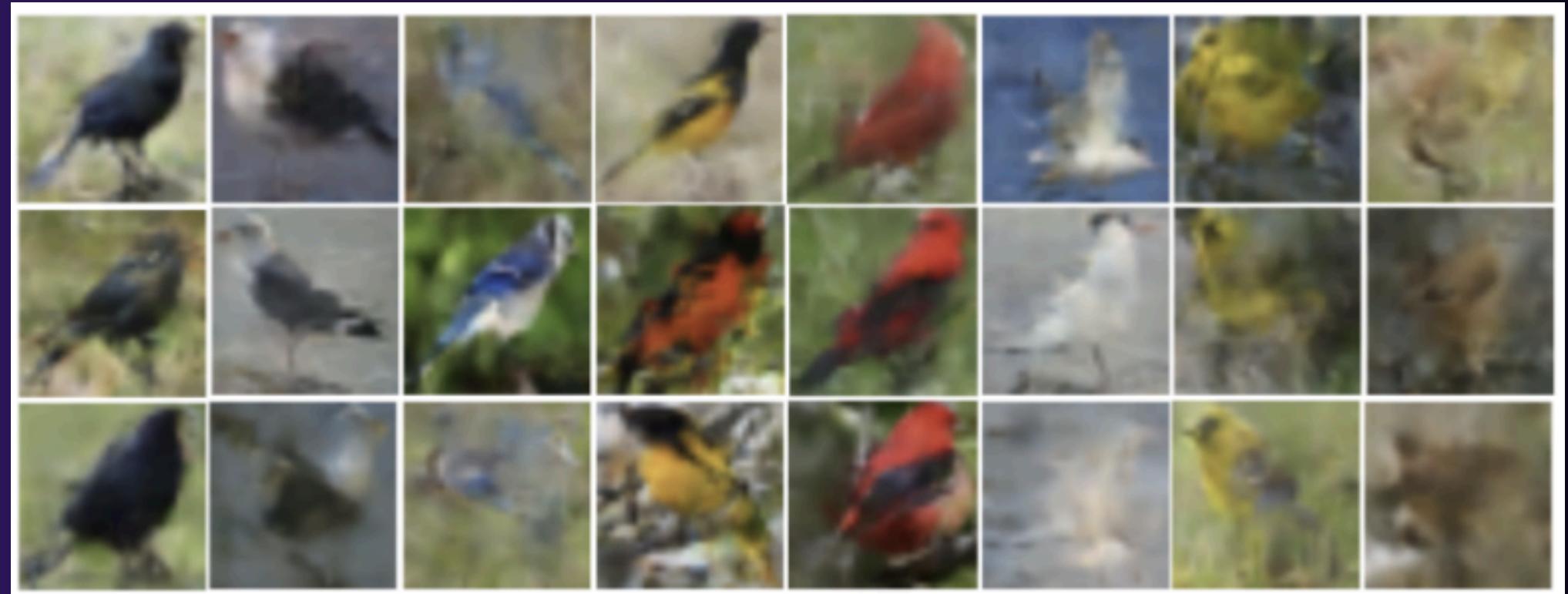
Table 1: Overview of models trained in the paper.

Dataset	Betas	Dim z	Dim w	K
CUBICC	<u>1.0, 2.5, 5.0</u>	64	32	<u>10, 40</u> (with beta 5.0)
PolyMNIST	<u>1.0, 2.5, 5.0</u>	32	32	<u>10, 40</u> (with beta 5.0)

- CUBICC Dataset:
 - Training includes first three models only ($K \neq 40$).
 - Initial number of clusters 35
- PolyMNIST Dataset:
 - Training involves only the first model with $\beta = 1.0$.
 - Initial number of clusters 40

Observations: Validation loss consistently increases (Figures 4, 17).
 Challenges: Limited training instances compared to original study.

CUBICC Qualitative Results - Unconditional Generation

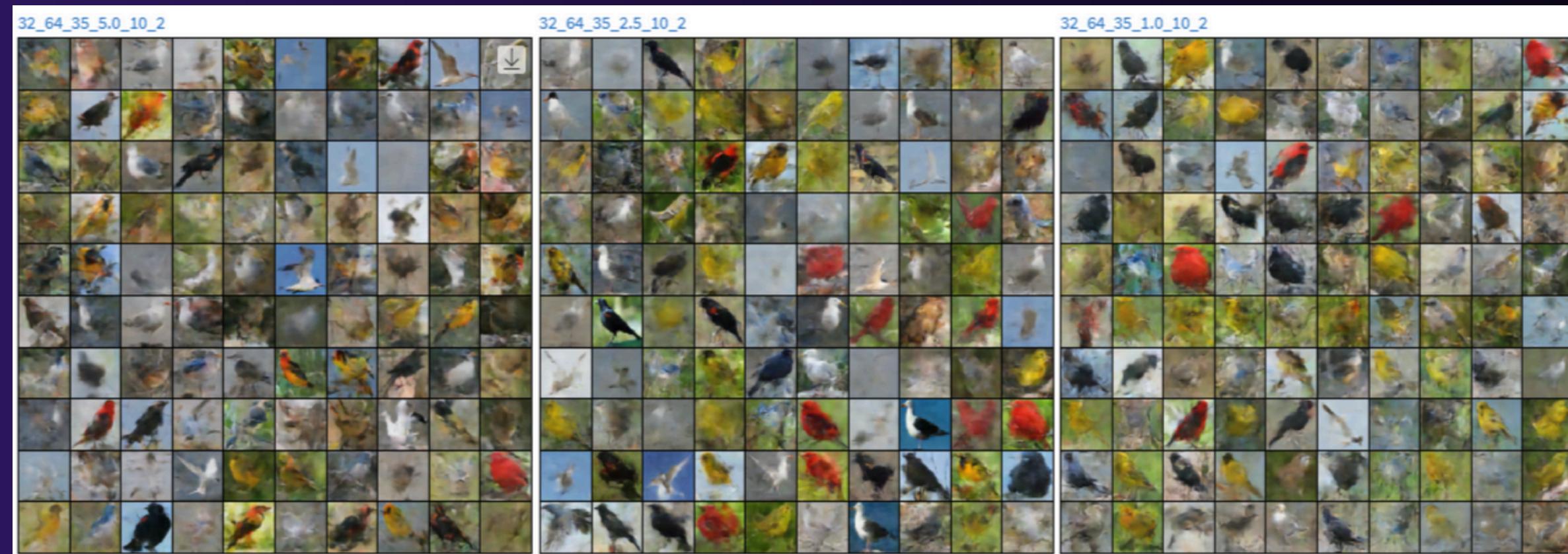


Research paper

Unconditional Generation (1 model, m0, K=10):

- Observations: Low-quality samples, undefined textures, blended colors.
- Alignment: Matches study's findings, where DDPM is suggested for improvement.
- Claim Validated: Generative quality impacted by diverse modality types.

CUBICC Qualitative Results - Unconditional Generation

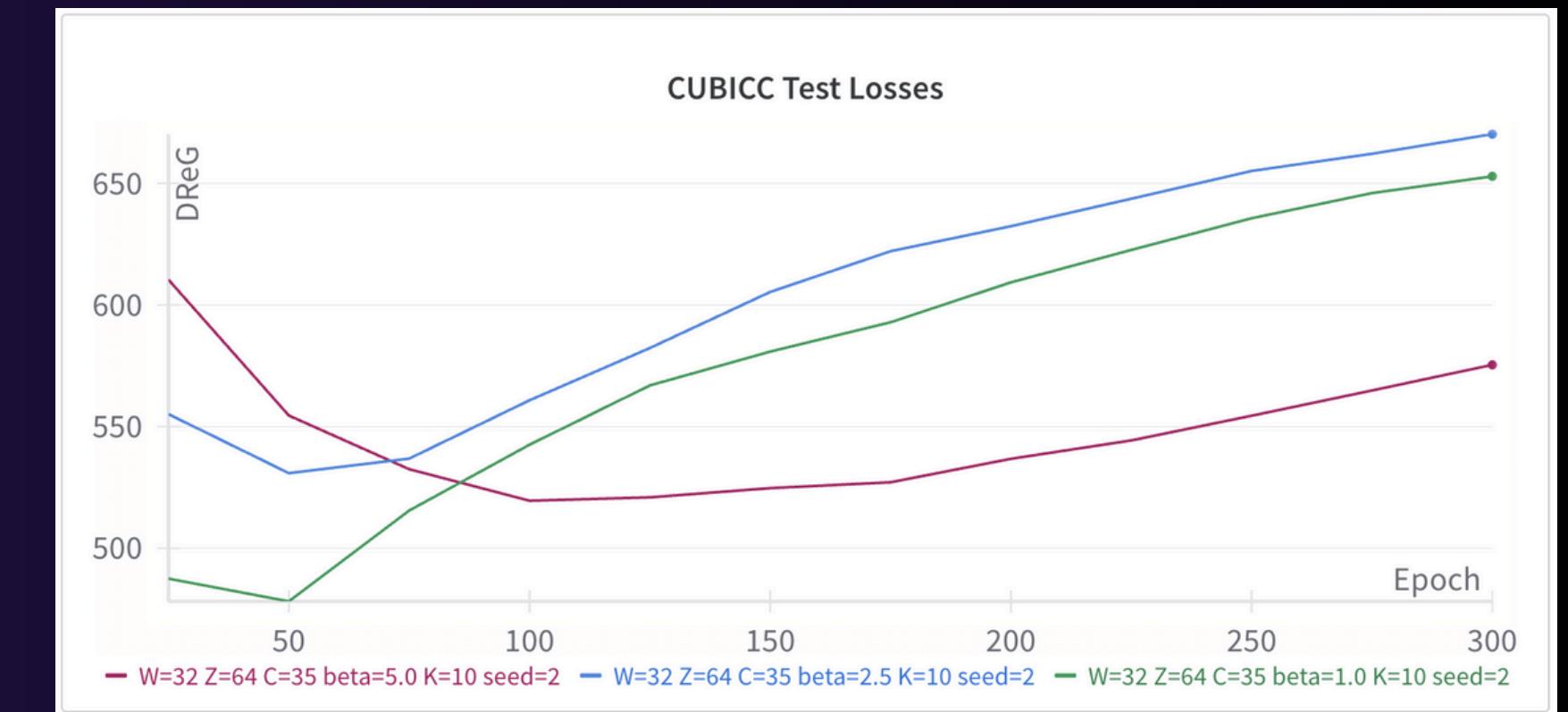
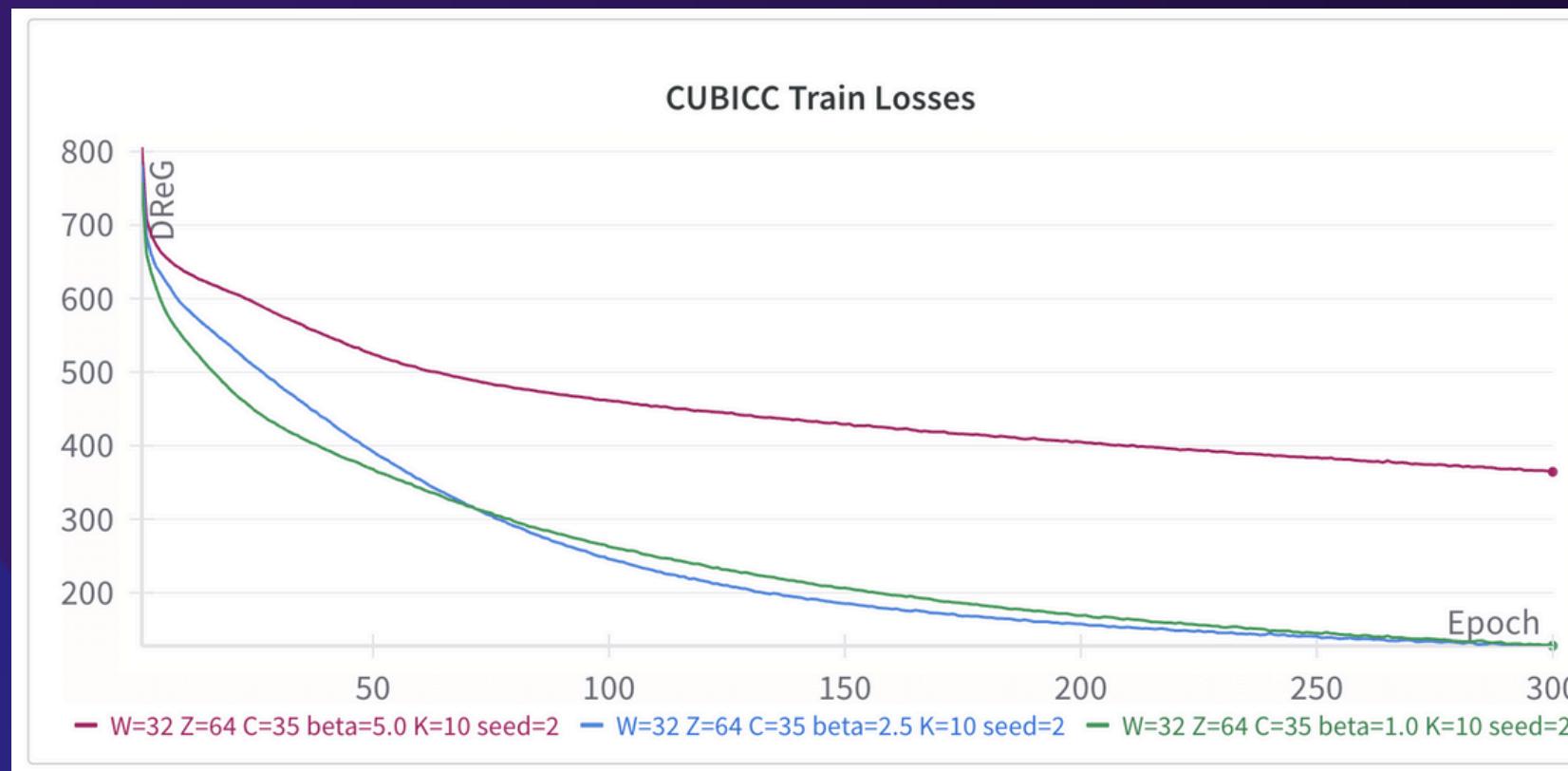


Our project

Unconditional Generation (3 models, m0, K=10):

- Observations: Low-quality samples, undefined textures, blended colors.
- Alignment: Matches study's findings, where DDPM is suggested for improvement.
- Claim Validated: Generative quality impacted by diverse modality types.

CUBICC Quantitative Results - Training Losses



- Not reported in the paper
- Suggests model is learning the data and not generalizing
- May create bias in future results

CUBICC Quantitative Results - Generation

Conditional generation performance:

- Metrics: FID
- Results for β values:
 - $\beta = 1.0 \rightarrow 74.53$
 - $\beta = 2.5: \rightarrow 85.12$
 - $\beta = 5.0 \rightarrow 102.95$

	$\beta = 1.0$	Unconditional		Conditional
	FID	Coherence	FID	Coherence
MVAE	50.65 (0.72)	0.007 (0.001)	82.59 (6.22)	0.093 (0.009)
MVTCAE	110.85 (2.61)	0.000 (0.000)	58.98 (0.62)	0.509 (0.006)
mmJSD	179.76 (2.97)	0.054 (0.011)	209.98 (1.26)	0.785 (0.023)
MoPoE-VAE	98.56 (1.32)	0.037 (0.002)	160.29 (4.12)	0.723 (0.006)
MMVAE	165.17 (3.40)	0.222 (0.019)	152.11 (4.11)	0.837 (0.004)
MMVAE+	86.64 (1.04)	0.095 (0.020)	80.75 (0.18)	0.796 (0.010)
CMVAE	78.52 (0.63)	0.560 (0.025)	74.53 (0.64)	0.806 (0.009)
	$\beta = 2.5$	Unconditional		Conditional
	FID	Coherence	FID	Coherence
MVAE	58.53 (0.12)	0.080 (0.006)	85.23 (9.37)	0.298 (0.044)
MVTCAE	87.07 (0.89)	0.003 (0.000)	62.55 (1.30)	0.591 (0.004)
mmJSD	180.55 (8.67)	0.060 (0.010)	222.09 (5.34)	0.778 (0.003)
MoPoE-VAE	107.11 (0.780)	0.141 (0.005)	178.27 (2.01)	0.720 (0.008)
MMVAE	164.71 (3.17)	0.232 (0.010)	150.83 (2.69)	0.844 (0.010)
MMVAE+	96.01 (2.10)	0.344 (0.013)	92.81 (0.78)	0.869 (0.013)
CMVAE	85.68 (0.66)	0.781 (0.021)	85.12 (0.75)	0.897 (0.003)
	$\beta = 5.0$	Unconditional		Conditional
	FID	Coherence	FID	Coherence
MVAE	61.25 (0.40)	0.112 (0.010)	90.37 (3.20)	0.301 (0.024)
MVTCAE	85.43 (2.80)	0.029 (0.001)	74.61 (3.41)	0.604 (0.004)
mmJSD	186.49 (2.89)	0.076 (0.018)	226.20 (2.91)	0.784 (0.029)
MoPoE-VAE	122.68 (1.96)	0.238 (0.001)	182.99 (1.96)	0.673 (0.002)
MMVAE	164.29 (2.97)	0.229 (0.017)	152.11 (3.18)	0.839 (0.010)
MMVAE+	109.08 (1.41)	0.421 (0.006)	107.78 (0.88)	0.836 (0.023)
CMVAE	103.95 (0.16)	0.775 (0.024)	102.36 (0.83)	0.882 (0.010)

CUBICC Quantitative Results - Generation

Our results

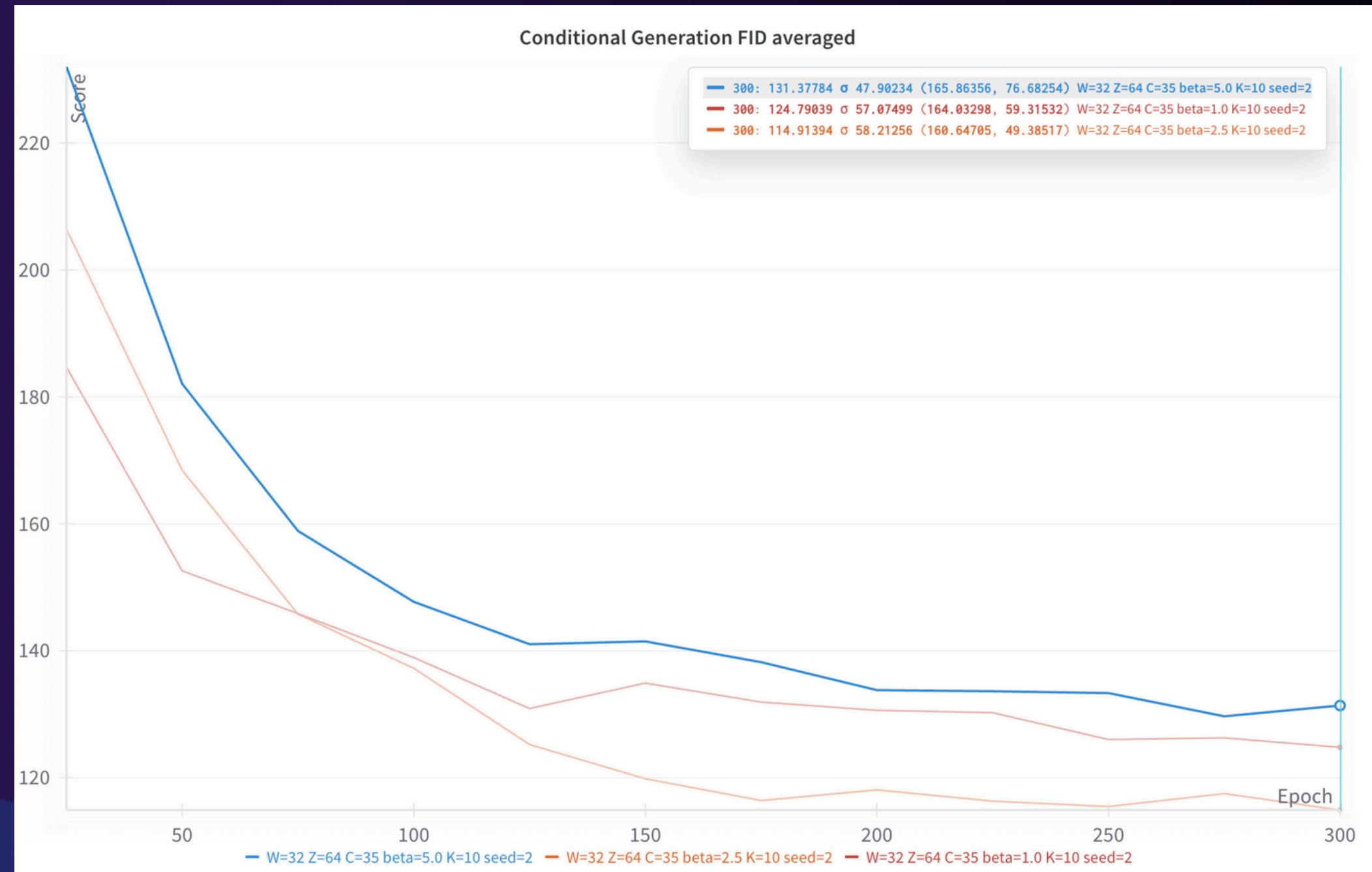
Conditional Generation:

- Metrics: FID
- Results for β values:
 - $\beta = 1.0 \rightarrow$ avg 124.8, std 57.07
 - $\beta = 2.5: \rightarrow$ avg 114.91, std 58.22
 - $\beta = 5.0 \rightarrow$ avg 131.38, std 47.09

Research Paper

Generation performance:

- Metrics: FID
- Results for β values:
 - $\beta = 1.0 \rightarrow 78.52$
 - $\beta = 2.5: \rightarrow 85.68$
 - $\beta = 5.0 \rightarrow 103.95$



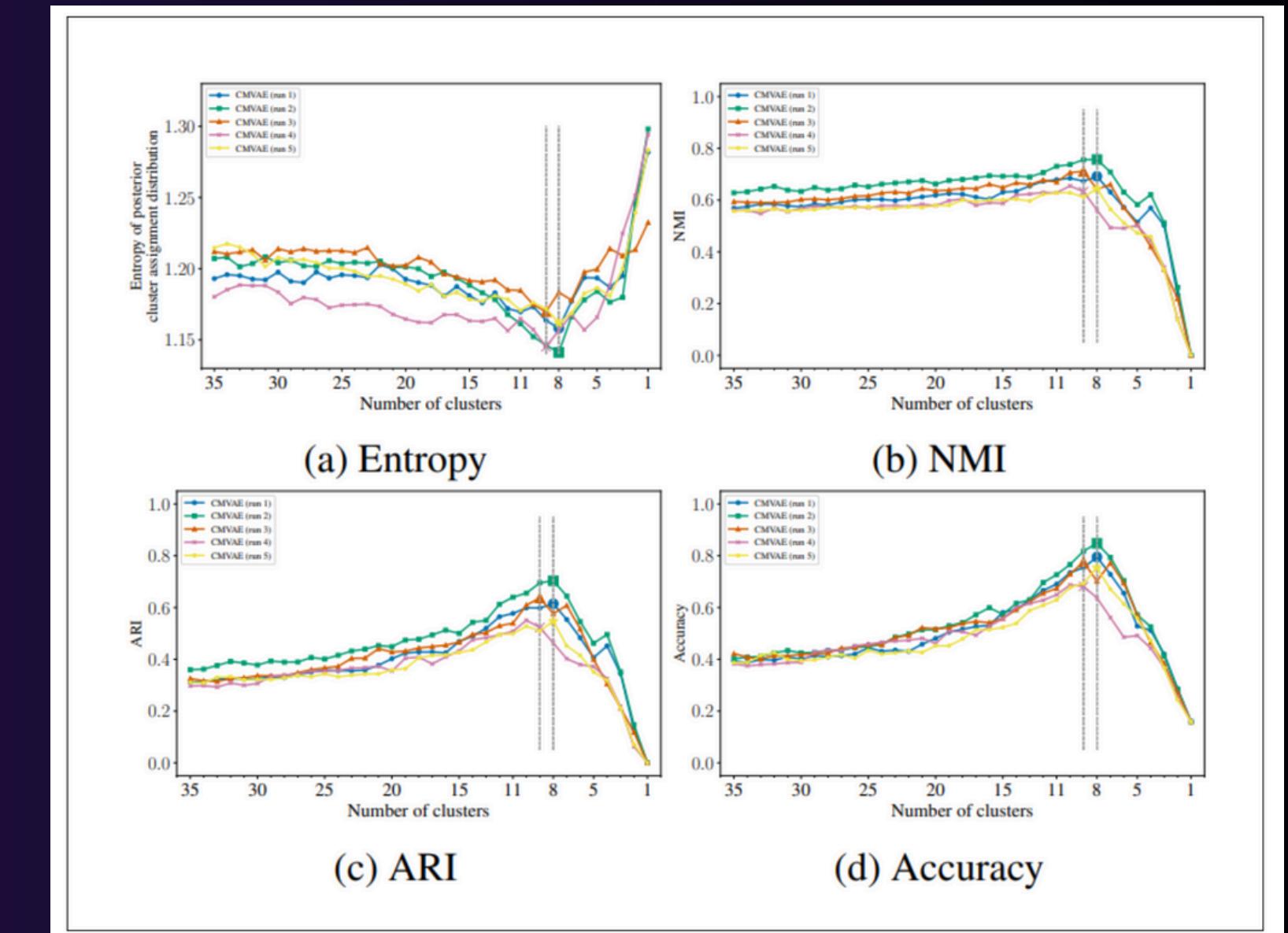
High standard deviation across all betas is insufficient to derive a conclusion.

CUBICC Quantitative Results - Clustering

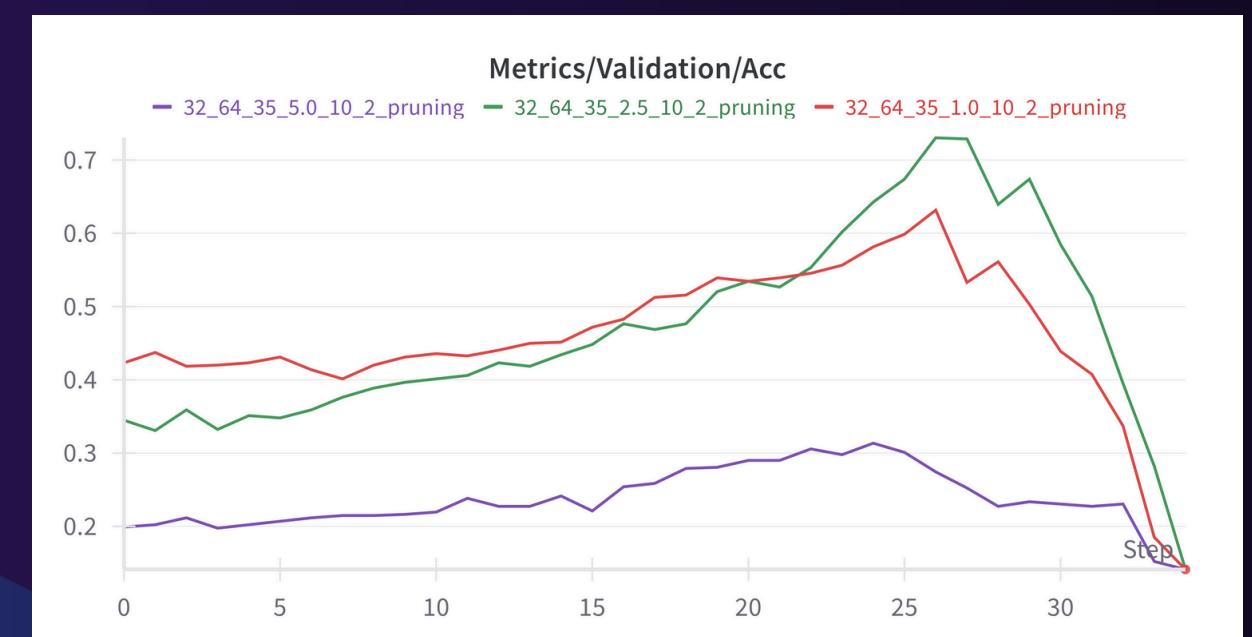
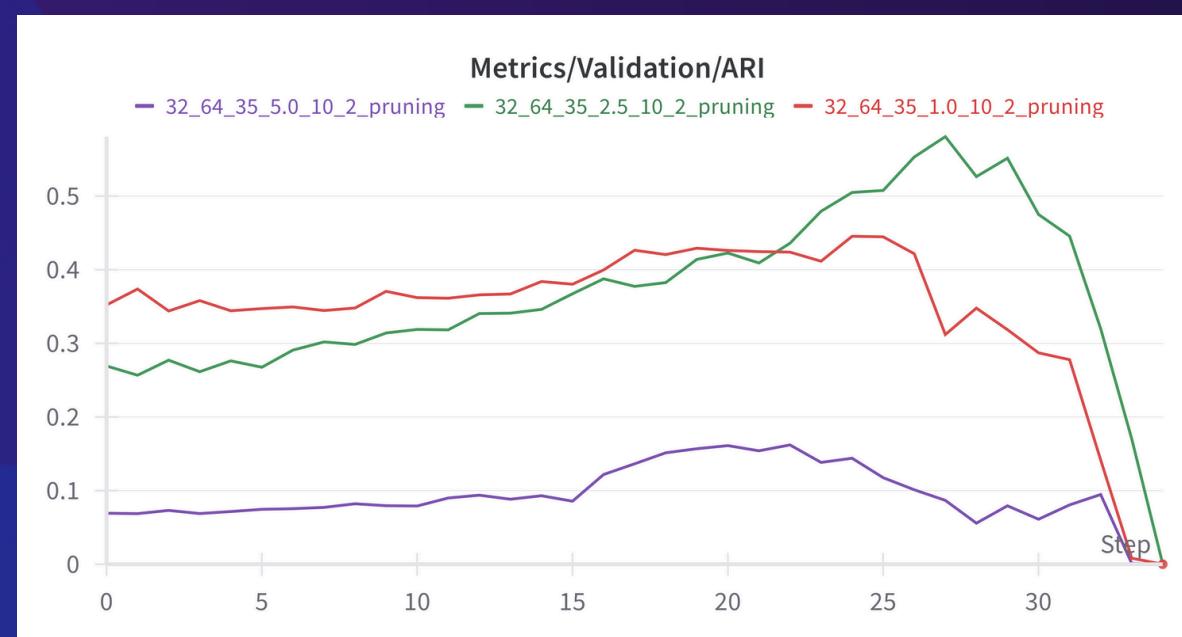
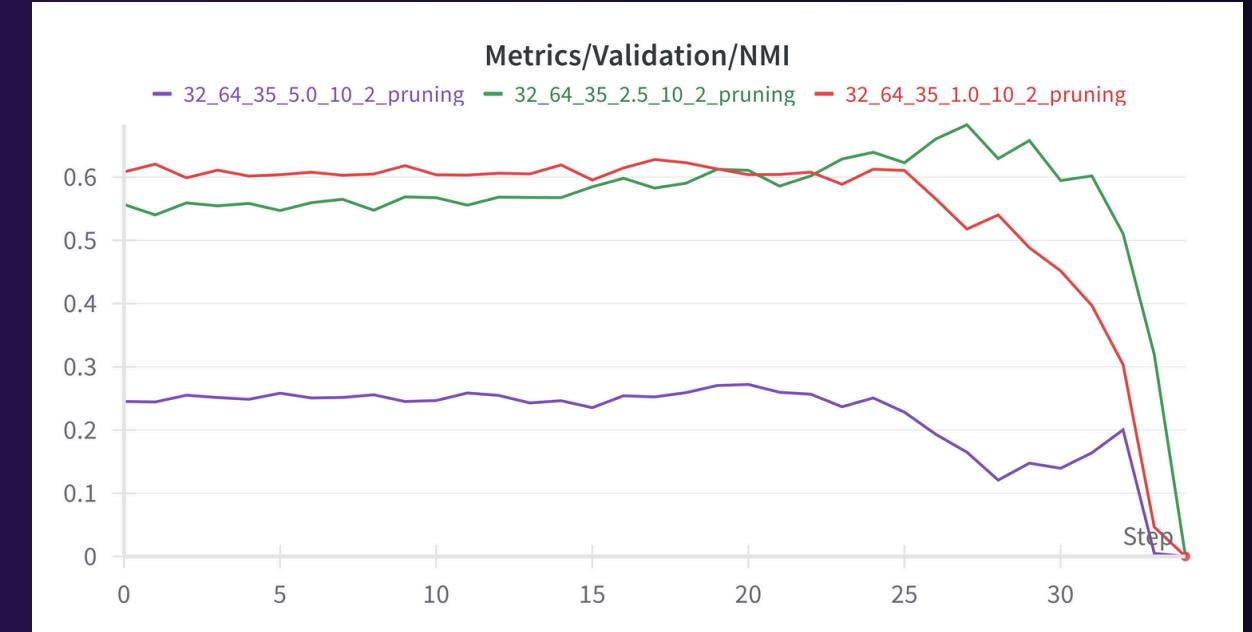
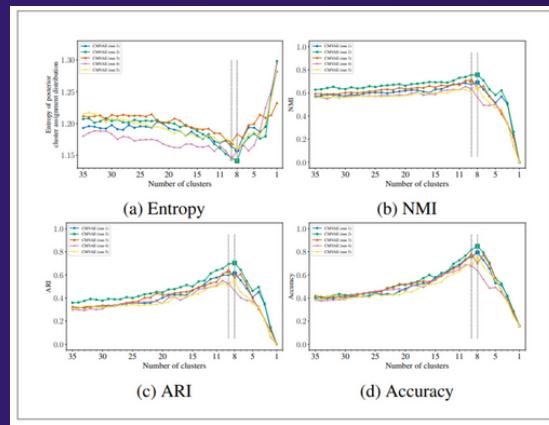
Research Paper

- 5 different seed runs on beta=2.5
- Optimal number of clusters found is 8
- NMI: ~0.67
- ARI: ~ 0.59
- ACC: ~ 0.76

	PolyMNIST			CUBICC		
	NMI	ARI	ACC	NMI	ARI	ACC
VaDE	0.43 (0.04)	0.36 (0.04)	0.54 (0.05)	0.15 (0.01)	0.08 (0.01)	0.27 (0.01)
DeepCluster	0.12 (0.02)	0.08 (0.02)	0.26 (0.04)	0.19 (0.01)	0.10 (0.01)	0.29 (0.01)
CMC	0.97 (0.01)	0.97 (0.01)	0.99 (0.01)	0.37 (0.05)	0.10 (0.03)	0.31 (0.04)
CMVAE	0.97 (0.02)	0.97 (0.02)	0.99 (0.01)	0.67 (0.07)	0.59 (0.09)	0.76 (0.07)



CUBICC Quantitative Results - Clustering



Research Paper

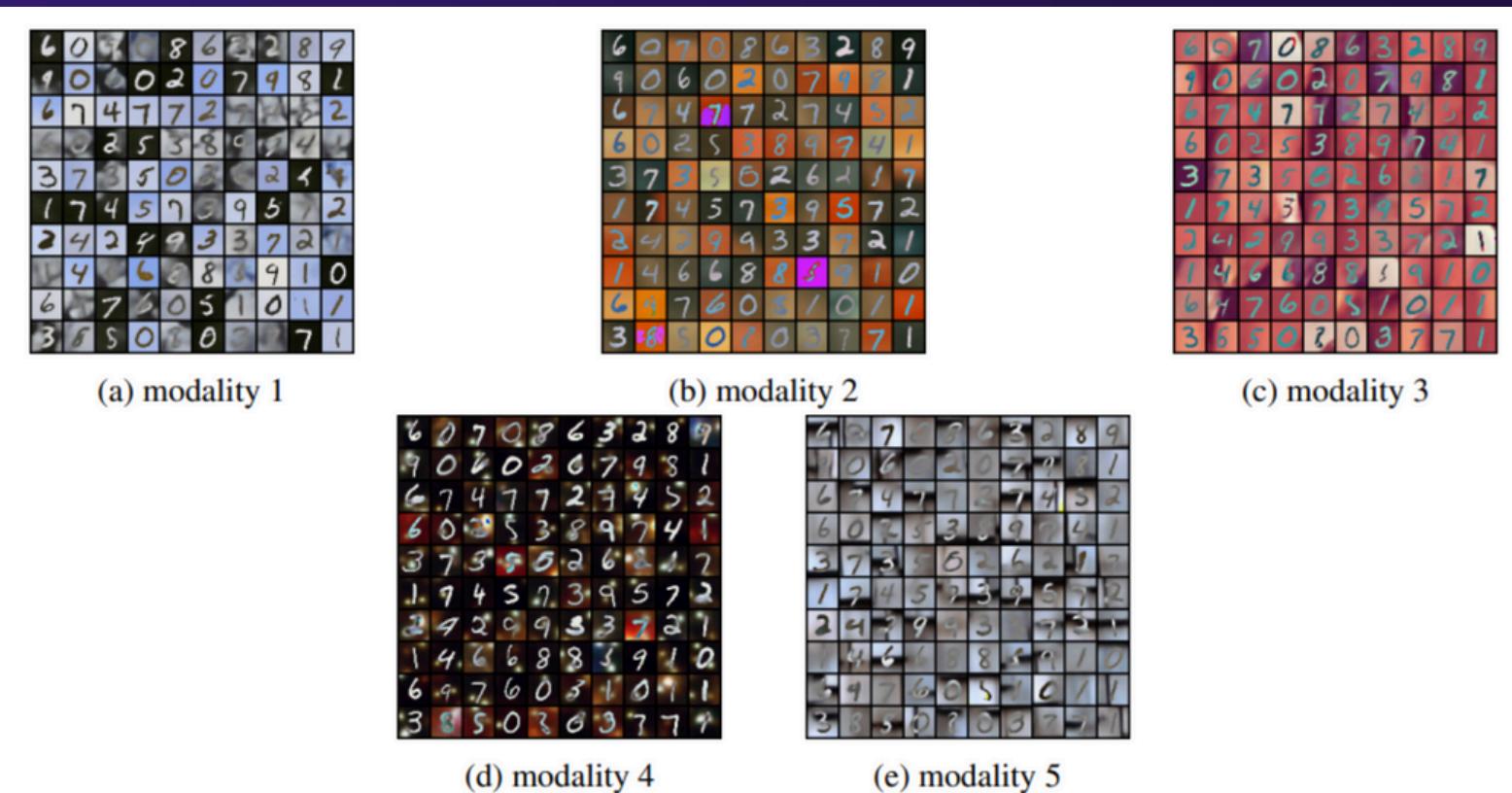
- 5 different seed runs on beta=2.5
- Optimal number of clusters found is 8
- NMI: ~ 0.67 (-0.03)
- ARI: ~ 0.59 (-0.08)
- ACC: ~ 0.76 (+0.01)

Our project

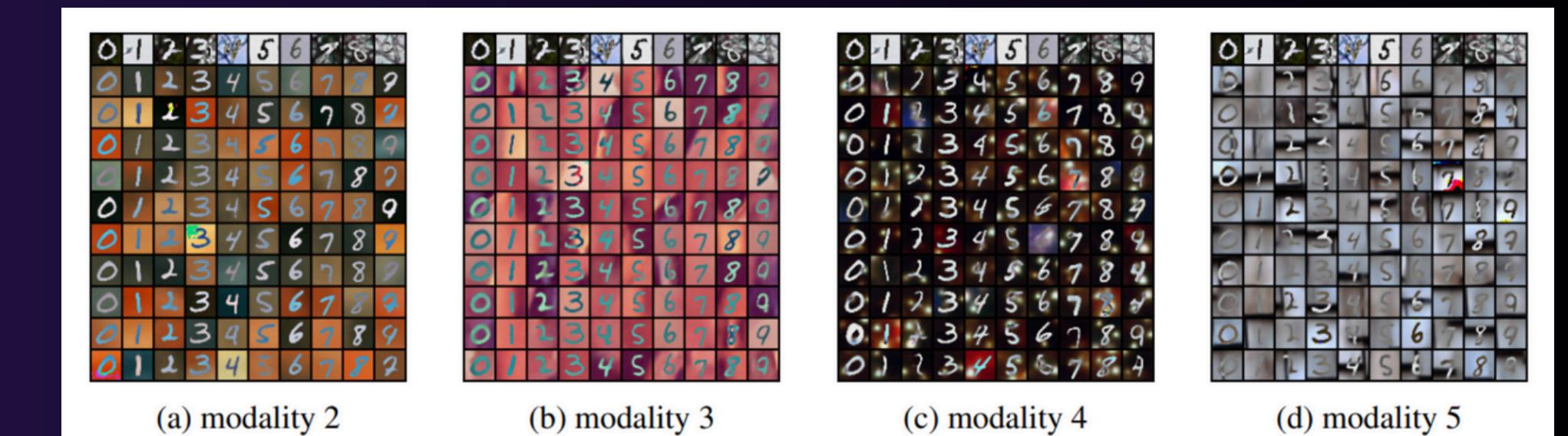
- 1 seed run, best results on beta=2.5
- For beta=2.5:
 - Optimal number of clusters found is 8
 - NMI: ~ 0.7
 - ARI: ~ 0.67
 - ACC: ~ 0.75

Results suggest congruency
with claims on paper on
CUBICC dataset

PolyMNIST Qualitative Results - Generation



Unconditional

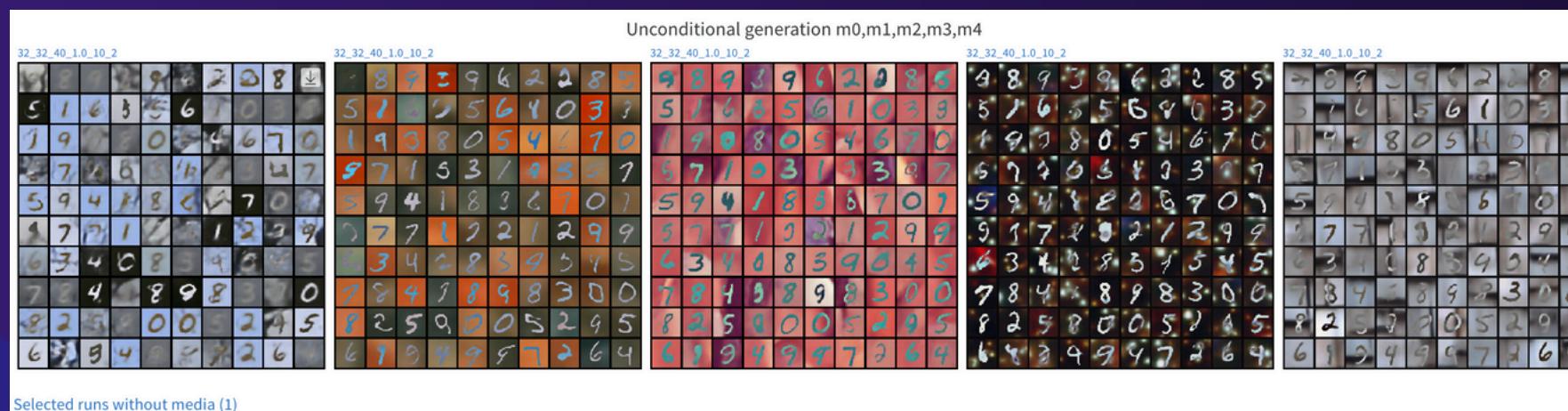


Conditional from m0 to others

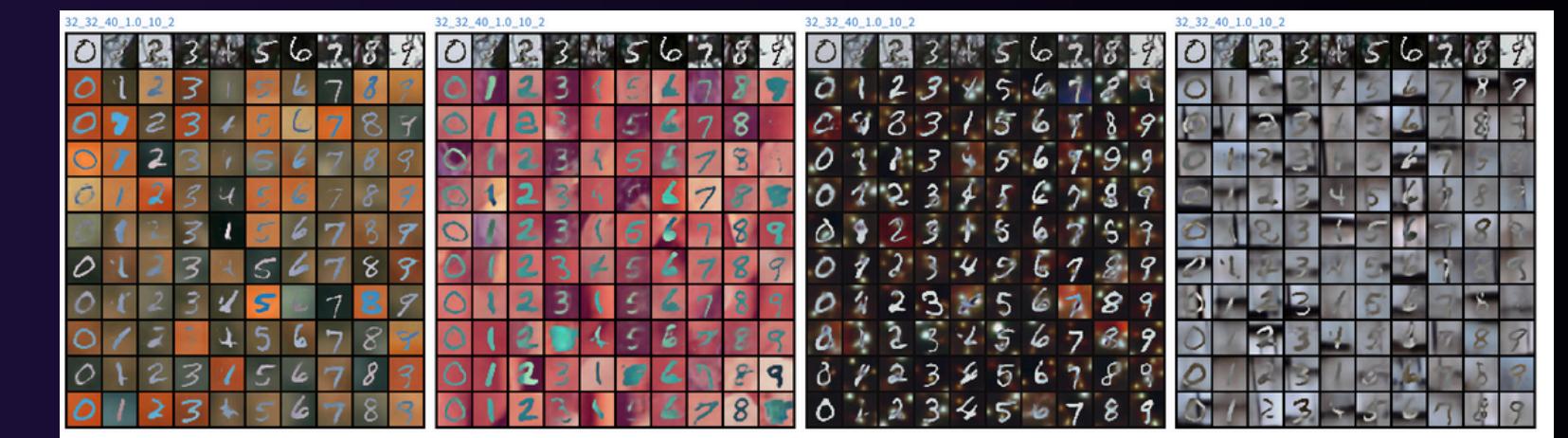
Research paper

- Observations: Strong cross-modality consistency.
- Features: Preserves modality-specific traits like color, texture, and style.
- Validation: High suitability for multimodal generative tasks.

PolyMNIST Qualitative Results - Generation



Unconditional

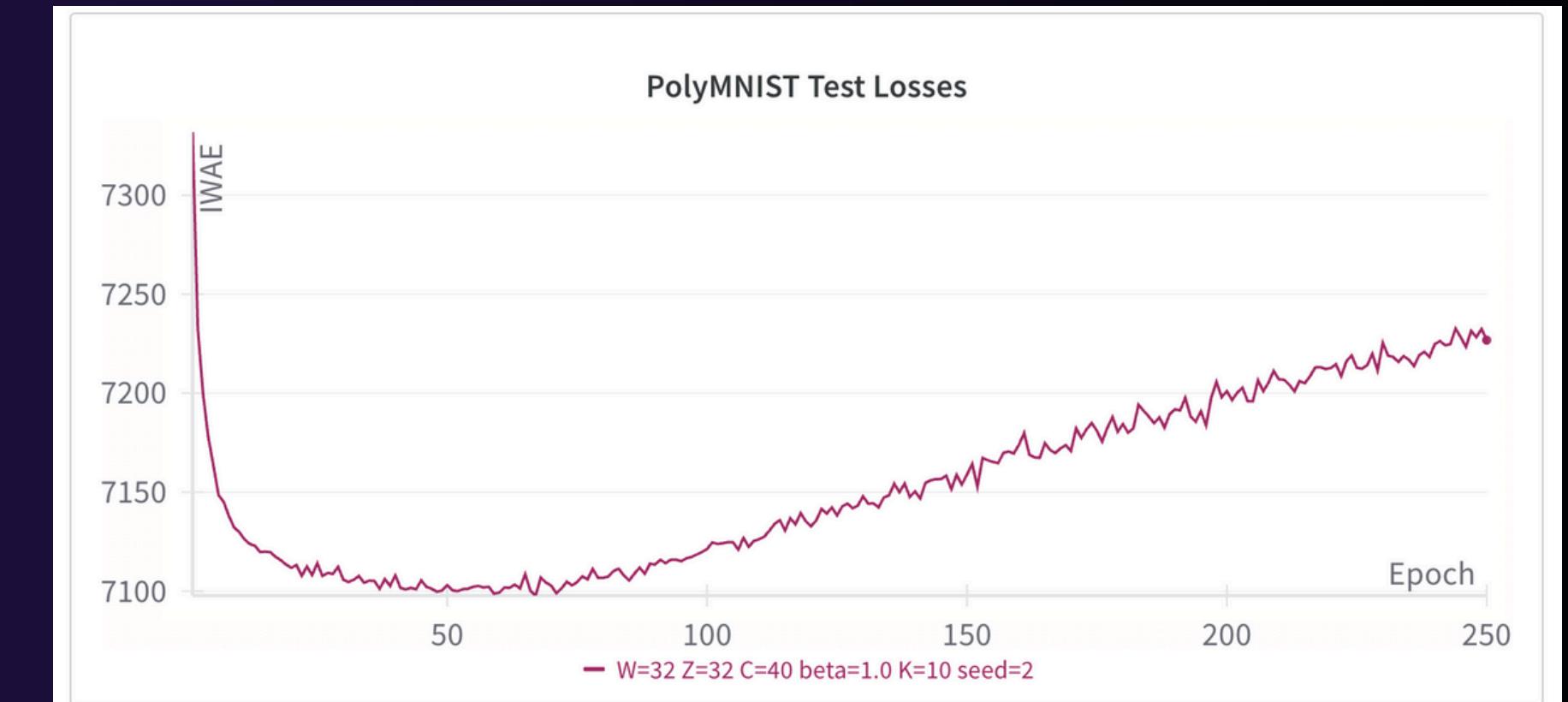


Conditional from m0 others

Our project

- Observations: Strong cross-modality consistency.
- Features: Preserves modality-specific traits like color, texture, and style.
- Validation: High suitability for multimodal generative tasks.
- Qualitative results are congruent with research paper

PolyMNIST Quantitative Results - Training Losses



- Not reported in the paper
- Suggests model is learning the data and not generalizing
- May create bias in future results

PolyMNIST Quantitative Results - Generation

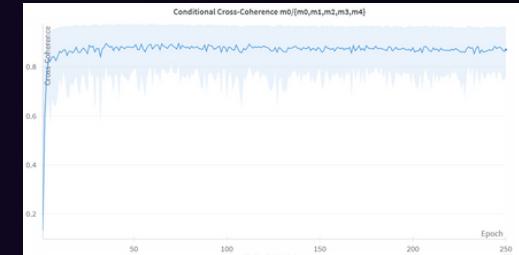
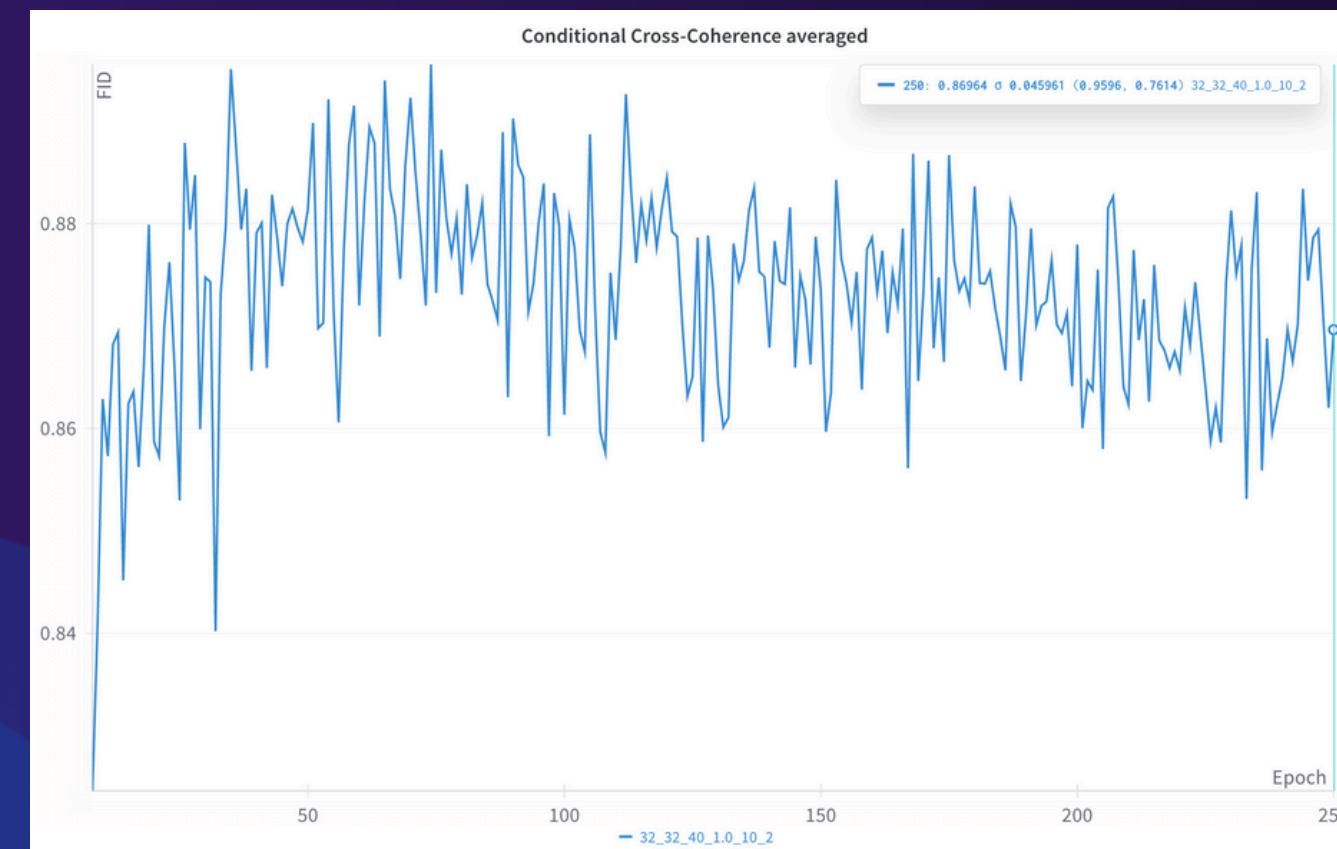
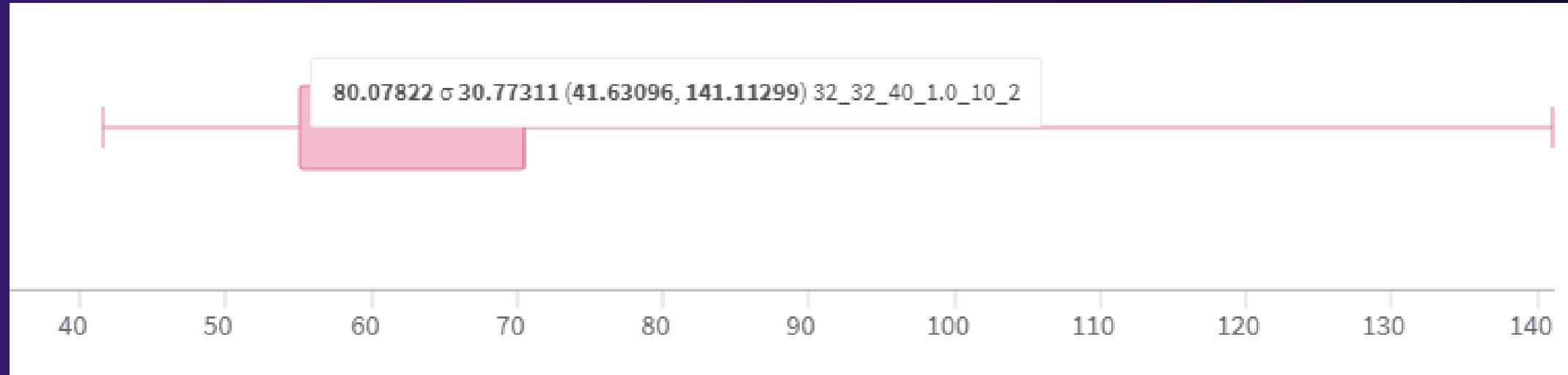
Research Paper

Conditional Generation performance:

- Metrics: FID, Coherence
- Results for $\beta = 1.0$:
 - FID 74.53
 - Coher. 0.806

	$\beta = 1.0$	Unconditional		Conditional
	FID	Coherence	FID	Coherence
MVAE	50.65 (0.72)	0.007 (0.001)	82.59 (6.22)	0.093 (0.009)
MVTCAE	110.85 (2.61)	0.000 (0.000)	58.98 (0.62)	0.509 (0.006)
mmJSD	179.76 (2.97)	0.054 (0.011)	209.98 (1.26)	0.785 (0.023)
MoPoE-VAE	98.56 (1.32)	0.037 (0.002)	160.29 (4.12)	0.723 (0.006)
MMVAE	165.17 (3.40)	0.222 (0.019)	152.11 (4.11)	0.837 (0.004)
MMVAE+	86.64 (1.04)	0.095 (0.020)	80.75 (0.18)	0.796 (0.010)
CMVAE	78.52 (0.63)	0.560 (0.025)	74.53 (0.64)	0.806 (0.009)
	$\beta = 2.5$	Unconditional		Conditional
	FID	Coherence	FID	Coherence
MVAE	58.53 (0.12)	0.080 (0.006)	85.23 (9.37)	0.298 (0.044)
MVTCAE	87.07 (0.89)	0.003 (0.000)	62.55 (1.30)	0.591 (0.004)
mmJSD	180.55 (8.67)	0.060 (0.010)	222.09 (5.34)	0.778 (0.003)
MoPoE-VAE	107.11 (0.780)	0.141 (0.005)	178.27 (2.01)	0.720 (0.008)
MMVAE	164.71 (3.17)	0.232 (0.010)	150.83 (2.69)	0.844 (0.010)
MMVAE+	96.01 (2.10)	0.344 (0.013)	92.81 (0.78)	0.869 (0.013)
CMVAE	85.68 (0.66)	0.781 (0.021)	85.12 (0.75)	0.897 (0.003)
	$\beta = 5.0$	Unconditional		Conditional
	FID	Coherence	FID	Coherence
MVAE	61.25 (0.40)	0.112 (0.010)	90.37 (3.20)	0.301 (0.024)
MVTCAE	85.43 (2.80)	0.029 (0.001)	74.61 (3.41)	0.604 (0.004)
mmJSD	186.49 (2.89)	0.076 (0.018)	226.20 (2.91)	0.784 (0.029)
MoPoE-VAE	122.68 (1.96)	0.238 (0.001)	182.99 (1.96)	0.673 (0.002)
MMVAE	164.29 (2.97)	0.229 (0.017)	152.11 (3.18)	0.839 (0.010)
MMVAE+	109.08 (1.41)	0.421 (0.006)	107.78 (0.88)	0.836 (0.023)
CMVAE	103.95 (0.16)	0.775 (0.024)	102.36 (0.83)	0.882 (0.010)

PolyMNIST Quantitative Results - Generation



Our project

Conditional Generation performance:

- Metrics: FID, Coherence
- Results for $\beta=1.0$:
 - FID: avg. 80.07, std. 30, med. 41.63
 - Coher.: avg. 0.86, std. 0.05

Research Paper

Conditional Generation performance:

- Metrics: FID, Coherence
- Results for $\beta = 1.0$:
 - FID 74.53
 - Coher. 0.806

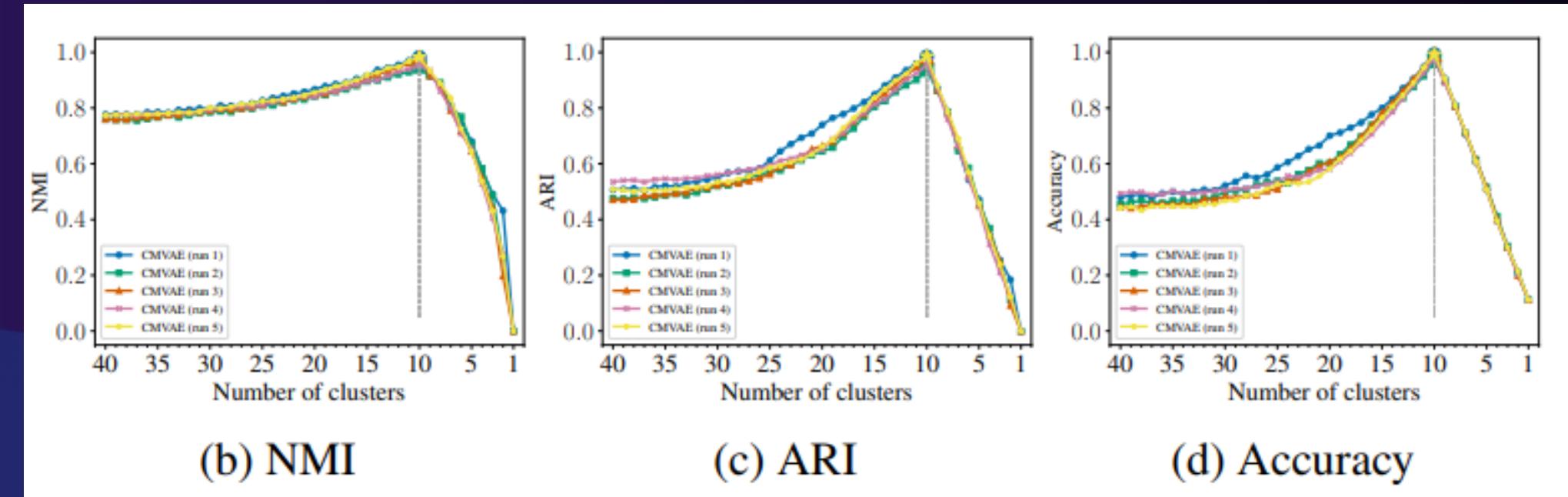
Coherence results suggest congruity with claims on paper on PolyMNIST dataset
 FID research needs to be refined and re-evaluated.

PolyMNIST Quantitative Results - Clustering

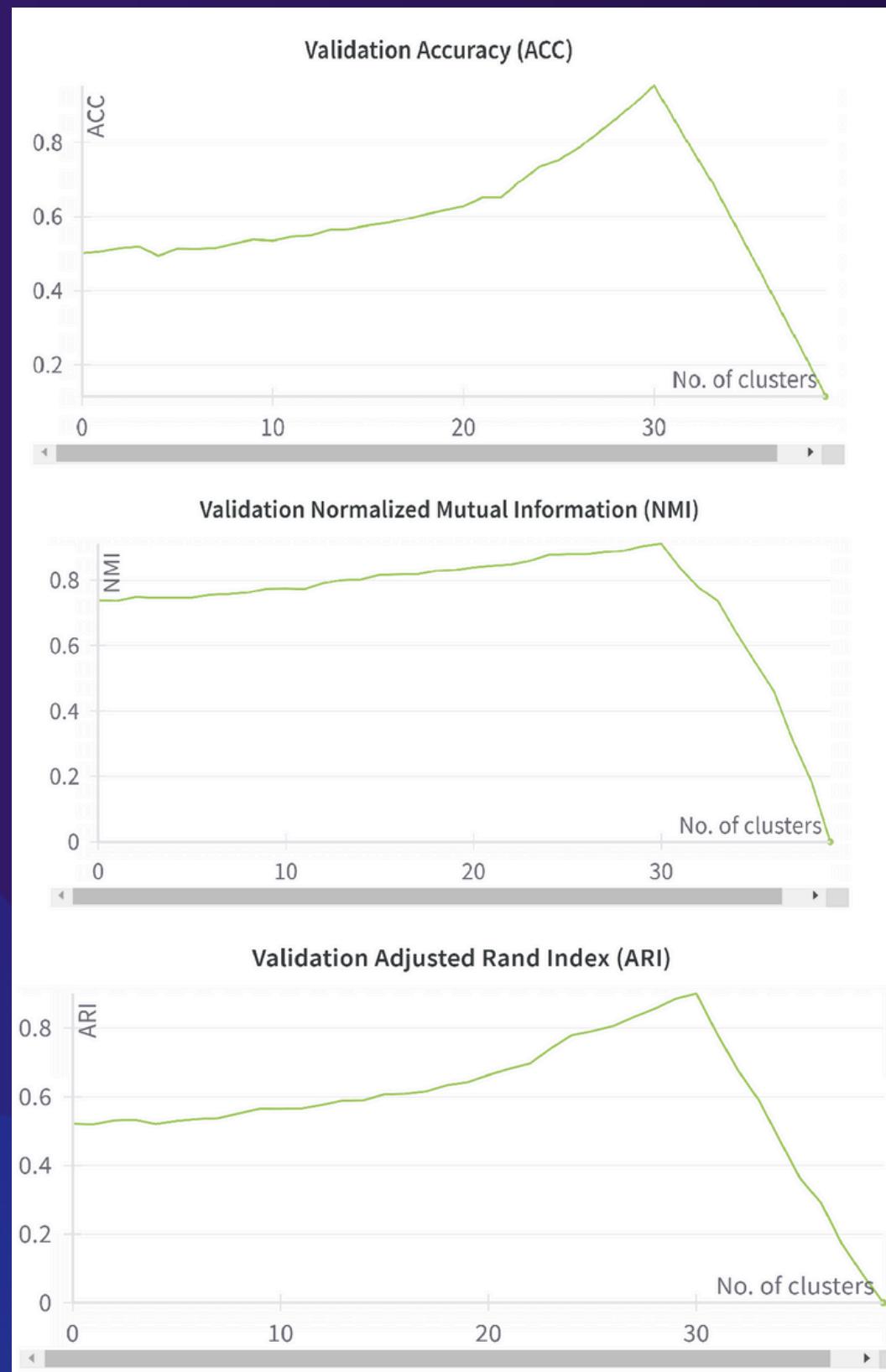
Research Paper

- 5 different seed runs on beta=1.0
- Optimal number of clusters found is 10
- NMI: ~0.97
- ARI: ~ 0.97
- ACC: ~ 0.99

	PolyMNIST			CUBICC		
	NMI	ARI	ACC	NMI	ARI	ACC
VaDE	0.43 (0.04)	0.36 (0.04)	0.54 (0.05)	0.15 (0.01)	0.08 (0.01)	0.27 (0.01)
DeepCluster	0.12 (0.02)	0.08 (0.02)	0.26 (0.04)	0.19 (0.01)	0.10 (0.01)	0.29 (0.01)
CMC	0.97 (0.01)	0.97 (0.01)	0.99 (0.01)	0.37 (0.05)	0.10 (0.03)	0.31 (0.04)
CMVAE	0.97 (0.02)	0.97 (0.02)	0.99 (0.01)	0.67 (0.07)	0.59 (0.09)	0.76 (0.07)



PolyMNIST Quantitative Results - Clustering



Our project

- 1 seed runs on beta=1.0
- Optimal number of clusters found is 10
- NMI: ~0.91 (-0.06)
- ARI: ~ 0.91 (-0.06)
- ACC: ~ 0.96 (-0.03)

Research Paper

- 5 different seed runs on beta=1.0
- Optimal number of clusters found is 10
- NMI: ~0.97
- ARI: ~ 0.97
- ACC: ~ 0.99

Results suggest congruency with claims on paper on PolyMNIST dataset

Consolidating Findings

Key Findings

- CUBICC:
 - Generative quality impacted by diverse modalities. Needs DDPM.
 - Train/Test losses pattern diverge continuously
 - Effective clustering at $\beta = 2.5$ (8 clusters)
 - NMI, ARI, ACC scores congruent with paper across all β
- PolyMNIST:
 - High fidelity in conditional and unconditional generation
 - Train/Test losses pattern diverge continuously
 - Coherence scores are congruent with paper for $\beta = 1.0$
 - Effective clustering at $\beta = 1.0$ (10 clusters)
 - NMI, ARI, ACC scores congruent with paper for $\beta=1.0$

Claims Validated:

- Implementation seems to improves over existing methods for clustering; generation needs finer assessment
- Post-hoc procedure effectively reduces entropy finding optimal number of clusters

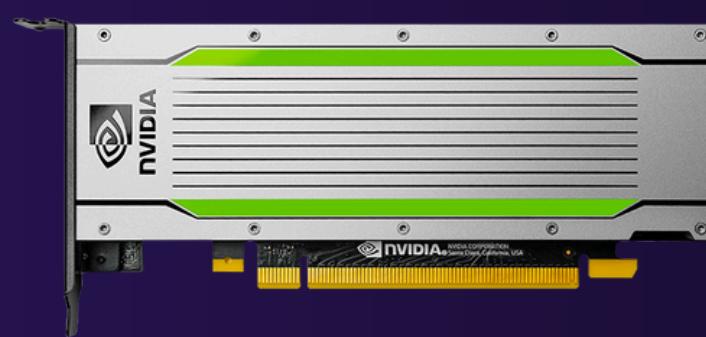
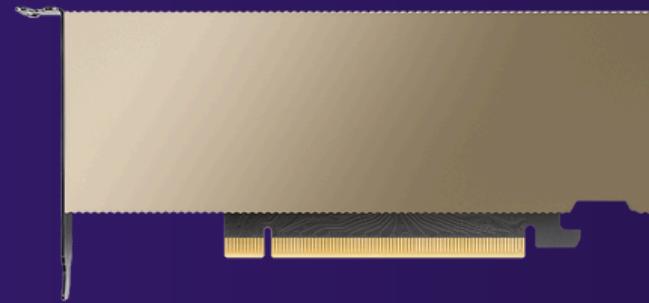
Challenges:

- Single seed-instance training limits conclusions
- Lack of resources for validating all claims
- No DDPM and hyperparameters are mentioned
- Could not validate FID for PolyMNIST
- Could not validate FID for CUBICC
- No way to perform statistical tests due to lack of data from authors

Hardware and Training Infrastructure



Hardware and Training Infrastructure



CPU:

- 4 Core, 16GB RAM, 3.1 Ghz

GPU:

- T4, 24GB VRAM, 30.3 TFLOPs
- L4, 16GB VRAM, 8.1 TFLOPs
- RTX 3070, 8GB VRAM, 20.2 TFLOPS
- M2, 16GB VRAM, 3.6 TFLOPs

From:

- Lightning.ai
- Own Hardware
- USI Mac Books

Times:

- 300 hours total compute
- PolyMNIST: 170 hours
- CUBICC: 35 hours

Issues Found



Issues Found

```

  OUTPUTDIR='..outputs'
  EXPERIMENT="PolyMNIST_1"
  DATADIR='..data'
  EPOCHS=250
  SEED=2
  SHARED_LAT_DIM=16
  MS_LAT_DIM=8
  # train_CMVAE
  python train_CMVAE_polyMNIST.py M
  --latent-dim 10 --latent-dim-z $SHARED_LAT_DIM --experiment $EXPERIMENT --obj "iwaе" --K 6 --batch-size 32 --seed $SEED \
  --datadir $DATADIR --outputdir $OUTPUTDIR \
  --inception_path "${DATADIR}/pt_inception-2015-12-05-6726825d.pth" \
  --pretrained-clfs-dir-path "${DATADIR}/trained_clfs_polyMNIST" \
  --priorposterior 'Laplace'
  # Entropy-based latent cluster selection
  python prune_polyMNIST.py --save-dir "${OUTPUTDIR}/${EXPERIMENT}/checkpoints/${MS_LAT_DIM}_${SHARED_LAT_DIM}" --{BETA} ${SEED} \
  --epoch $EPOCHS --seed $SEED \
  # calculate frechet_distance
  offset = np.eye(sigma1.shape[0]) * eps
  covmean = linalg.sqrtm((sigma1 + offset).dot(sigma2 + offset))
  # Numerical error might give slight imaginary component
  if np.iscomplexobj(covmean):
    if not np.allclose(np.diagonal(covmean).imag, 0, atol=1e-3):
      m = np.max(np.abs(covmean.imag))
      covmean = np.nan
      print('Imaginary component {}.'.format(m))
    else:
      covmean = covmean.real
  # Ensure covmean is at least 2D
  covmean = np.atleast_2d(covmean)
  tr_covmean = np.trace(covmean)
  return (diff.dot(diff) + np.trace(sigma1 +
    np.trace(sigma2) - 2 * tr_covmean)

```

```

  train_CMVAE_polyMNIST.py M
  utils.py > plot_text_as_image_tensor
  plt.axis('off')
  # Draw the canvas and retrieve the image as a Numpy array
  fig.canvas.draw()
  image_np = np.frombuffer(fig.canvas.tostring_rgb(), dtype=np.uint8)
  image_np = image_np.reshape([fig.canvas.get_width_height()[:-1] + (3,)])
  # Convert the Numpy array to a PyTorch tensor
  image_np = np.copy(image_np)
  image_tensor = torch.from_numpy(image_np).permute(2, 0, 1).float() / 255
  imgs.append(image_tensor)
  # Clean up the figure
  plt.close(fig)
  # Utils for clustering
  def cluster_acc(y_true, y_pred, return_index=False):
    """
    Calculate clustering accuracy
    fabia@Fabian-PC MINGW64 ~/Desktop/CMVAE/src (main)
    $ bash commands/run_polyMNIST_experiment_prune.sh
    ./outputs/POLYMNIST_FINAL/checkpoints/32_32_1.0_2
    Path does not exist
    ./outputs/POLYMNIST_FINAL/checkpoints/32_32_1.0_2
    False
    C:/Users/fabia/Desktop/CMVAE
    ['CUBICC_1', 'PolyMNIST_1', 'POLYMNIST_FINAL']
    FileNotFoundErr... [Errno 2] No such file or directory: '../data/pt_inception-2015-12-05-6726825d.pth'
  
```

- Dataset Installation Process Challenges
- Confusing Tables in the Paper
- Inconsistent Parameter Specifications in the Paper
- CUDA Dependency and Scalability Constraints

Issues Found

- Anomalies in Code Implementation
 - Checkpoints saved without opt. parameters
 - Missing implementation for Stable Diffusion
 - Complexity in Modeling Conditional Dependencies
 - Communication with Original Authors

```
OUTPUTDIR='..outputs'
EXPERIMENT="PolyMNIST_1"
DATADIR='..data'
EPOCHS=250
SEED=2
SHARED_LAT_DIM=16
MS_LAT_DIM=8

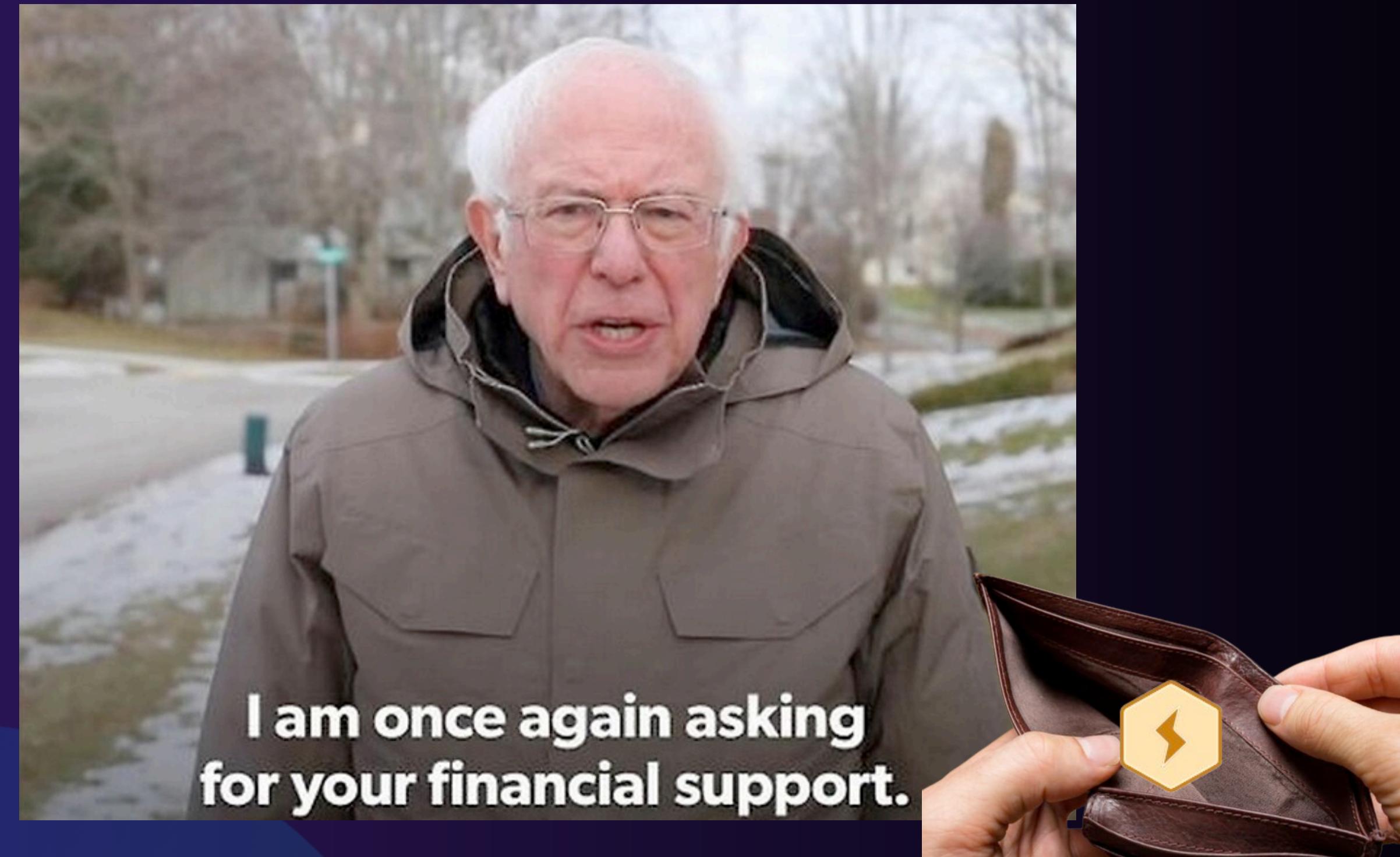
# Train CMVAE
python train_CMVAE_polyMNIST.py --experiment $EXPERIMENT --obj "iwae" --K 6 --ba
--latent-dim-c 10 --latent-dim-z $SHARED_LAT_DIM --latent-dim-w $MS_LAT_DIM
--datadir $DATADIR --outputdir $OUTPUTDIR \
--inception_path "${DATADIR}/pt_inception-2015-12-05-6726825d.pth" \
--pretrained-clfs-dir-path "${DATADIR}/trained_clfs_polyMNIST" \
--priorposterior 'Laplace'

# Entropy-based latent cluster selection
python prune_polyMNIST.py --save-dir "${OUTPUTDIR}/${EXPERIMENT}/checkpoint"
--epoch $EPOCHS --seed $SEED \
--fid_score.py 9+, M X uti
~/CMVAE/src/fid/fid_s

train_CMVAE_polyMNIST.py M train_CMVAE_CUBICC.py 7, M
CMVAE > src > fid > fid_score.py > calculate_frechet_distance
137 def calculate_frechet_distance(mu1, sigma1, mu2, sigma2, eps=1e-6):
138     offset = np.eye(sigma1.shape[0]) * eps
139     covmean = linalg.sqrtm((sigma1 + offset).dot(sigma2 + offset))
140
141     # Numerical error might give slight imaginary component
142     if np.iscomplexobj(covmean):
143         if not np.allclose(np.diagonal(covmean).imag, 0, atol=1e-3):
144             m = np.max(np.abs(covmean.imag))
145             covmean = np.nan
146             print('Imaginary component {}'.format(m))
147             # raise ValueError('Imaginary component {}'.format(m))
148
149     else:
150         covmean = covmean.real
151
152     # Ensure covmean is at least 2D
153     covmean = np.atleast_2d(covmean)
154
155     tr_covmean = np.trace(covmean)
156
157     return (diff.dot(diff) + np.trace(sigma1) +
158            np.trace(sigma2) - 2 * tr_covmean)

• Anoma
• Checkp
• Missin
```

Conclusion



Thank you!

