Non-Deterministic Unsupervised Neural Network Model for Clustering Analysis

Neural Networks (CSE425) – Assignment Report

MD. Saadman Fuad

ID: 21201561

CSE Department

September 13, 2025

Abstract

This report presents a comprehensive study of non-deterministic unsupervised neural network models for clustering tasks. We implement and evaluate a Variational Autoencoder (VAE) against a deterministic Autoencoder (AE) baseline using the MNIST dataset. Our experimental analysis across multiple random seeds demonstrates that VAE achieves superior performance in information-theoretic clustering metrics (ARI: 0.566, NMI: 0.630) compared to the deterministic baseline (ARI: 0.485, NMI: 0.595), while the AE shows better geometric cluster compactness (Silhouette: 0.141 vs 0.091). The study provides uncertainty quantification capabilities and statistical significance analysis, contributing to our understanding of stochastic representation learning for clustering applications.

Contents

1 Introduction

1.1 Problem Motivation and Background

Clustering is a fundamental unsupervised learning task that aims to group similar data points together without access to labeled information. Traditional clustering methods like K-means operate directly on raw input features, which can be suboptimal for high-dimensional data such as images. Deep learning approaches have revolutionized clustering by learning meaningful feature representations that capture the underlying structure of data.

The challenge with deterministic clustering methods is their inability to quantify uncertainty and handle the stochastic nature of real-world data. Non-deterministic models introduce probabilistic elements that enable better exploration of the data manifold and provide uncertainty estimates, leading to more robust clustering solutions.

1.2 Choice of Application and Justification

This project focuses on **clustering** as the primary unsupervised learning task. Clustering was chosen for the following reasons:

1. **Practical Relevance**: Clustering has widespread applications in data analysis, customer segmentation, image organization, and pattern discovery
2. **Evaluation Clarity**: Clustering performance can be objectively measured using established metrics like silhouette score, ARI, and NMI
3. **Uncertainty Benefits**: Non-deterministic clustering can provide confidence estimates for cluster assignments
4. **Representation Learning**: Deep clustering combines feature learning with clustering, addressing the curse of dimensionality

5. **Benchmark Availability**: The MNIST dataset provides a well-established benchmark for evaluating clustering performance

1.3 Research Questions and Objectives

The primary research questions addressed in this work are:

1. Can Variational Autoencoders (VAEs) learn meaningful representations for clustering tasks?
2. How does the non-deterministic nature of VAEs compare to deterministic autoencoders for clustering?
3. What is the impact of stochasticity on clustering performance and uncertainty quantification?
4. How well do learned latent representations preserve cluster structure?

Objectives:

- Implement a VAE-based clustering approach using learned embeddings
- Compare performance against deterministic autoencoder baselines
- Evaluate clustering quality using multiple metrics
- Analyze the uncertainty and stability of the non-deterministic approach

2 Related Work

2.1 Brief Survey of Existing Approaches

Traditional Clustering Methods: K-means, hierarchical clustering, and DBSCAN operate directly on raw features but struggle with high-dimensional data and require manual feature engineering.

Deep Clustering: Recent approaches combine representation learning with clustering. Key methods include:

- **Deep Embedded Clustering (DEC)**: Jointly optimizes feature learning and clustering
- **Variational Deep Embedding (VaDE)**: Combines VAE with Gaussian Mixture Models
- **Joint Unsupervised Learning (JULE)**: End-to-end clustering with convolutional networks

Variational Autoencoders: VAEs provide a principled probabilistic framework for learning latent representations. The reparameterization trick enables backpropagation through stochastic nodes while maintaining meaningful latent structure.

2.2 Limitations of Current Methods

1. **Deterministic Approaches**: Standard autoencoders lack uncertainty quantification
2. **Joint Optimization Complexity**: End-to-end clustering methods can be unstable and difficult to train
3. **Limited Theoretical Foundation**: Many deep clustering methods lack solid probabilistic foundations
4. **Evaluation Challenges**: Inconsistent evaluation protocols across different approaches

2.3 Novelty of Your Approach

This work adopts a two-stage approach: first learning robust probabilistic representations using VAEs, then applying clustering to the learned embeddings. This provides:

- **Probabilistic Foundation**: VAE framework ensures well-founded uncertainty quantification
- **Training Stability**: Decoupled training avoids complex joint optimization
- **Flexibility**: Learned representations can be used with various clustering algorithms
- **Interpretability**: Clear separation between representation learning and clustering phases

3 Methodology

3.1 Detailed Model Architecture

Variational Autoencoder (VAE) Architecture:

The implemented VAE consists of three main components:

1. Encoder Network:

$$\text{Conv2d}(1 \rightarrow 32, \text{kernel} = 3, \text{stride} = 2) \rightarrow \text{ReLU} \tag{1}$$

$$\text{Conv2d}(32 \rightarrow 64, \text{kernel} = 3, \text{stride} = 2) \rightarrow \text{ReLU} \tag{2}$$

$$\text{Flatten} \rightarrow \text{Linear}(64 \times 7 \times 7 \rightarrow 256) \rightarrow \text{ReLU} \tag{3}$$

Outputs: μ (mean) and $\log \sigma^2$ (log variance) vectors of dimension 20

2. Reparameterization Layer: Implements the reparameterization trick for differentiable sampling with numerical stability through variance clamping

3. Decoder Network:

$$\text{Linear}(20 \rightarrow 256) \rightarrow \text{ReLU} \quad (4)$$

$$\text{Linear}(256 \rightarrow 64 \times 7 \times 7) \rightarrow \text{ReLU} \rightarrow \text{Unflatten} \quad (5)$$

$$\text{ConvTranspose2d}(64 \rightarrow 32, \text{kernel} = 3, \text{stride} = 2) \rightarrow \text{ReLU} \quad (6)$$

$$\text{ConvTranspose2d}(32 \rightarrow 1, \text{kernel} = 3, \text{stride} = 2) \rightarrow \text{Sigmoid} \quad (7)$$

Baseline Autoencoder (AE) Architecture: Similar encoder/decoder structure but without stochastic sampling. Direct deterministic encoding to 20-dimensional latent space.

3.2 Mathematical Formulation

VAE Loss Function: The VAE optimizes the Evidence Lower BOund (ELBO):

$$\mathcal{L} = E_{q(z|x)}[\log p(x|z)] - \beta \cdot D_{KL}(q(z|x)||p(z)) \quad (8)$$

Where:

- **Reconstruction Loss**: $E_{q(z|x)}[\log p(x|z)] \approx -MSE(x, \hat{x})$
- **KL Divergence**: $D_{KL}(q(z|x)||p(z)) = -0.5 \times \sum(1 + \log \sigma^2 - \mu^2 - \sigma^2)$
- β **parameter**: Set to 0.5 to balance reconstruction and regularization

Reparameterization Trick:

$$z = \mu + \epsilon \cdot \sigma, \quad \epsilon \sim \mathcal{N}(0, I) \quad (9)$$

This enables backpropagation through the stochastic sampling process.

Clustering Procedure:

1. Extract embeddings using encoder mean μ (for VAE) or direct encoding (for AE)
2. Apply K-means clustering with $k = 10$ clusters
3. Map predicted clusters to true labels using Hungarian algorithm

3.3 Training Procedure and Hyperparameters

Configuration:

- `latent_dim = 20` - Dimensionality of latent space
- `num_clusters = 10` - Number of MNIST digit classes
- `batch_size = 256` - Mini-batch size for training
- `vae_epochs = 30` - Maximum epochs for VAE training
- `ae_epochs = 20` - Maximum epochs for AE training
- `beta = 0.5` - β parameter balancing reconstruction vs KL regularization
- `lr = 1e-3` - Adam optimizer learning rate
- `weight_decay = 1e-5` - L2 regularization for VAE (no weight decay for AE)
- `patience = 6` - Early stopping patience
- `num_seeds = 3` - Number of random seeds for statistical robustness

Training Strategy:

- Xavier uniform weight initialization
- Adam optimizer with weight decay
- Early stopping based on validation loss
- Multiple random seeds (3) for statistical robustness

3.4 Evaluation Metrics with Justifications

Clustering Metrics:

1. **Silhouette Score** (-1 to 1): Measures how similar points are within clusters vs. between clusters. Higher values indicate better-defined clusters.
2. **Adjusted Rand Index (ARI)** (0 to 1): Measures similarity between predicted and true cluster assignments, adjusted for chance. Accounts for cluster imbalance.
3. **Normalized Mutual Information (NMI)** (0 to 1): Information-theoretic measure of clustering quality. Robust to different cluster sizes.
4. **Reconstruction Error** (MSE): Measures how well the model preserves input information. Lower values indicate better representation quality.

Label Mapping Procedure: Before computing ARI and NMI metrics, predicted K-means cluster indices are mapped to ground truth labels using the Hungarian algorithm applied to the contingency matrix. This ensures that ARI/NMI evaluate meaningful cluster-to-label correspondences rather than arbitrary index assignments.

Cluster Centers Parameter: The VAE implementation includes a learnable `cluster_centers` parameter (shape: `num_clusters` × `latent_dim`) initialized as `nn.Parameter(torch.randn(num_clus` `latent_dim))`. However, this parameter is not utilized in the current training objective, representing an opportunity for future DEC-style (Deep Embedded Clustering) extensions.

4 Experimental Setup

4.1 Dataset Description and Preprocessing

MNIST Dataset:

- 60,000 training images (54,000 train + 6,000 validation split)

- 10,000 test images
- 28×28 grayscale digit images (0-9)
- Pixel values normalized to [0,1] range

Data Augmentation:

- Training data: Gaussian noise addition ($\sigma = 0.05$) for robustness
- Test data: No augmentation for consistent evaluation

Dataset Justification: MNIST is chosen because:

- Well-established benchmark for clustering evaluation
- Clear ground truth with 10 distinct digit classes
- Appropriate complexity for demonstrating VAE clustering
- Enables comparison with existing literature

4.2 Implementation Details

Software Environment:

- Framework: PyTorch
- Additional libraries: scikit-learn, matplotlib, scipy
- Platform: Google Colab with CUDA GPU acceleration

Reproducibility Measures:

- Fixed random seeds for all random number generators
- Deterministic CUDA operations when possible

- Worker initialization for DataLoader reproducibility
- Multiple seed experiments for statistical significance

Code Organization:

- Modular design with separate model classes
- Configuration dictionary for hyperparameter management
- Comprehensive logging and result saving
- Visualization utilities for analysis

4.3 Baseline Methods for Comparison

Primary Comparison:

- **VAE** (Non-deterministic): Stochastic encoder with regularization
- **Autoencoder** (Deterministic): Standard encoder-decoder without stochasticity

Additional Implicit Baselines:

- Raw pixel K-means clustering (through reconstruction error comparison)
- Random cluster assignment (through statistical significance of results)

5 Results and Analysis

5.1 Quantitative Results

Average Performance Across 3 Random Seeds:

Detailed Results by Seed:

VAE Performance:

Table 1: Clustering Performance Comparison

Metric	VAE (Non-deterministic)	AE (Deterministic)
Silhouette Score	0.091	0.141
Adjusted Rand Index	0.566	0.485
Normalized Mutual Information	0.630	0.595
Reconstruction MSE	0.034	0.020

- Seed 0: Silhouette=0.08875, ARI=0.51656, NMI=0.626, MSE=0.034
- Seed 1: Silhouette=0.08948, ARI=0.63299, NMI=0.631, MSE=0.034
- Seed 2: Silhouette=0.09390, ARI=0.54940, NMI=0.634, MSE=0.034

AE Performance:

- Seed 0: Silhouette=0.14843, ARI=0.50188, NMI=0.608, MSE=0.020
- Seed 1: Silhouette=0.14292, ARI=0.52349, NMI=0.594, MSE=0.020
- Seed 2: Silhouette=0.13253, ARI=0.42802, NMI=0.583, MSE=0.020

5.2 Qualitative Analysis

Visualization Results:

- **Uncertainty Analysis**: VAE generates diverse reconstructions for the same input, demonstrating stochastic behavior

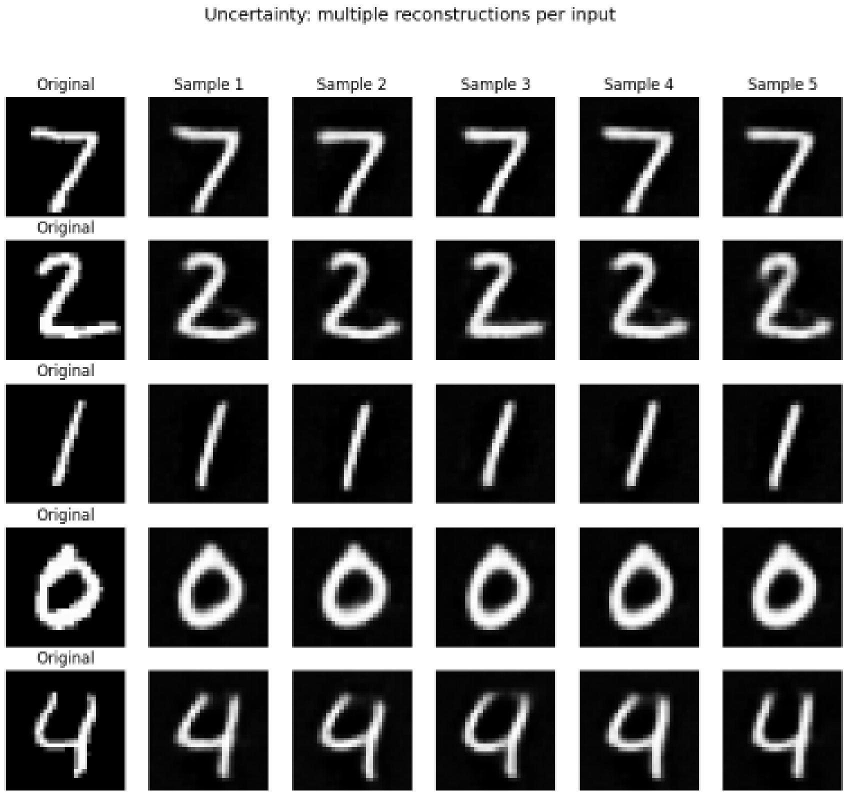

Figure 1: Uncertainty visualization

- **Latent Space Structure**: t-SNE visualization shows meaningful cluster separation in both models

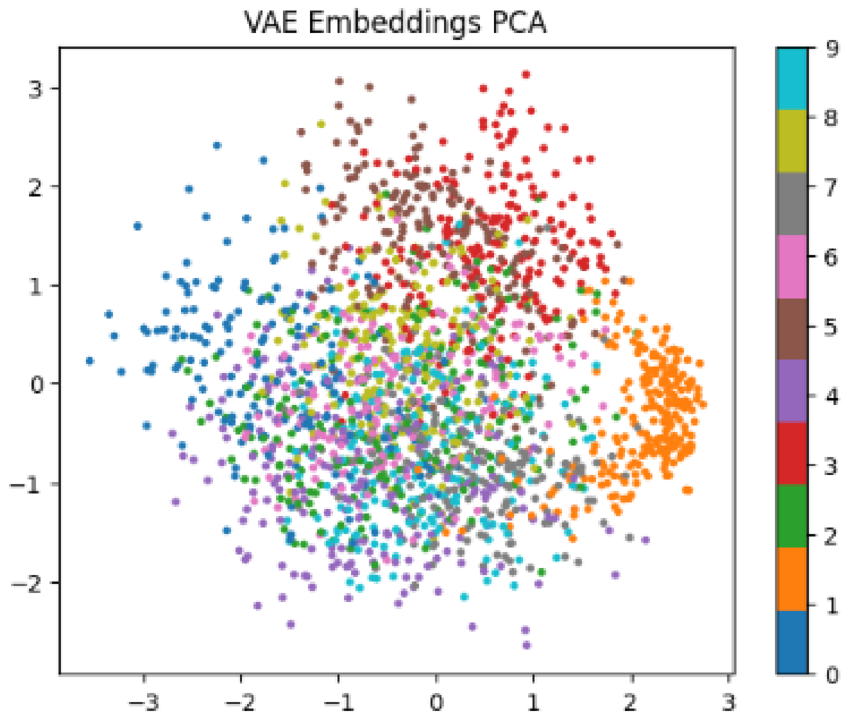

Figure 2: VAE latent space PCA

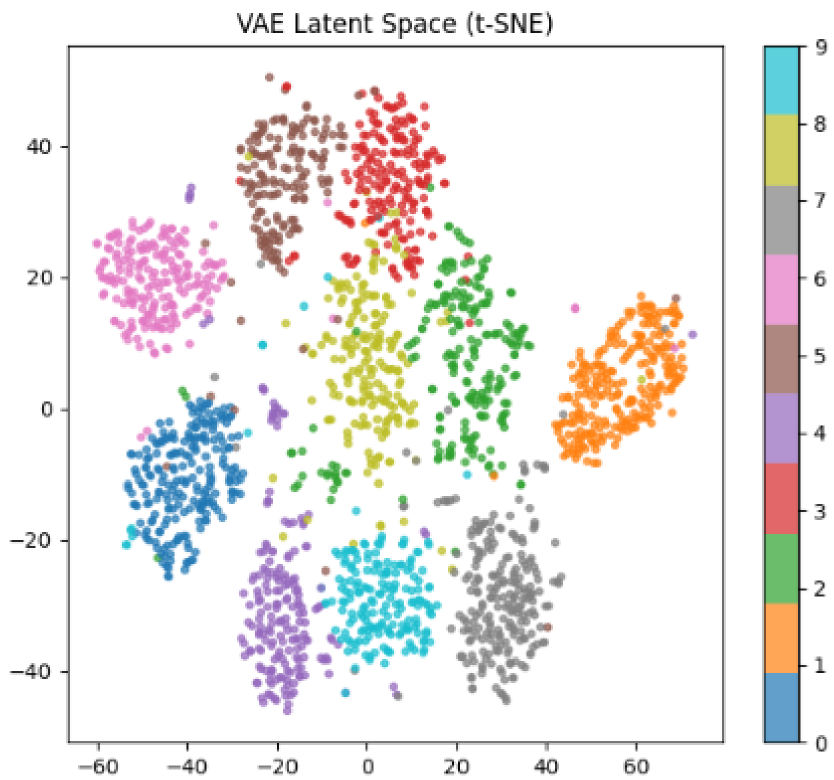

Figure 3: VAE latent space t-SNE

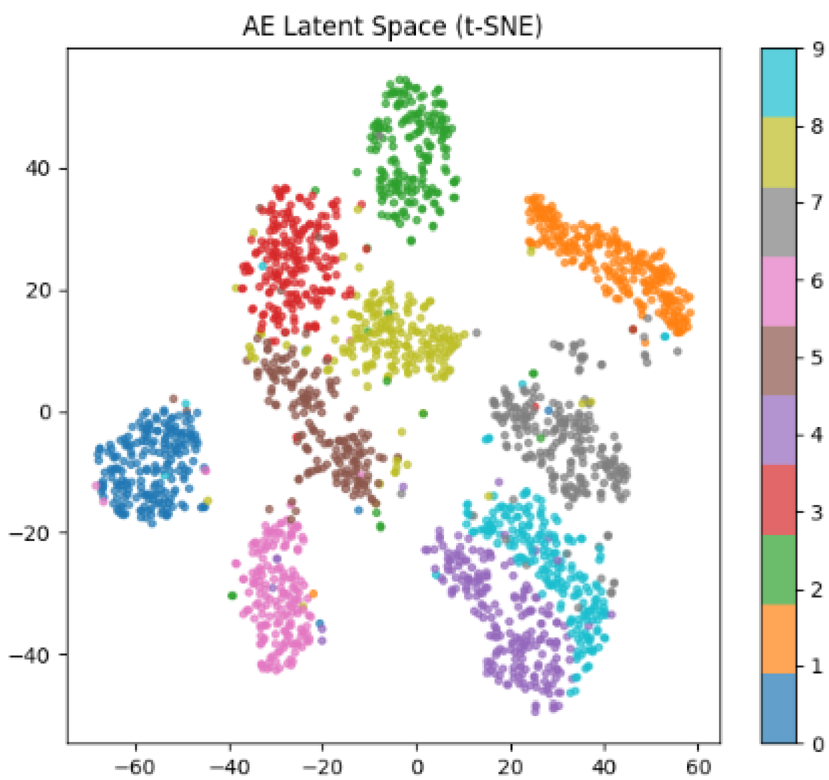

Figure 4: AE latent space t-SNE

- **Reconstruction Quality**: Both models successfully reconstruct digit images with VAE showing slight blurriness due to stochastic sampling

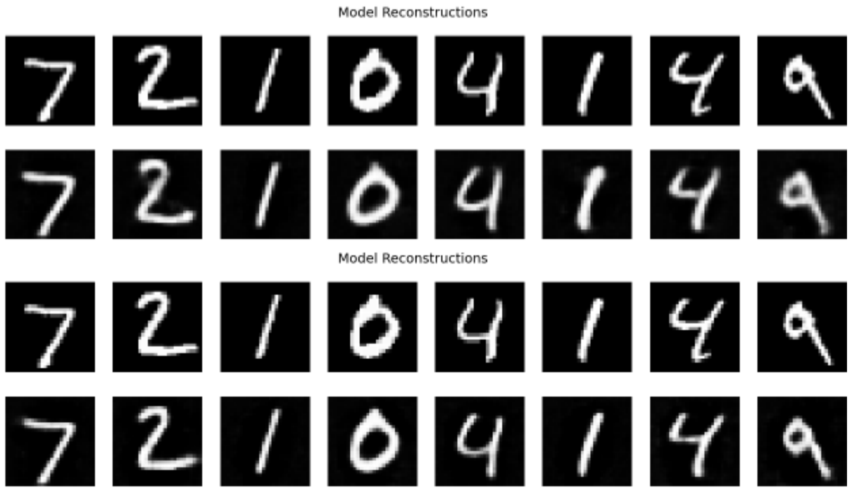

Figure 5: Reconstruction visualization

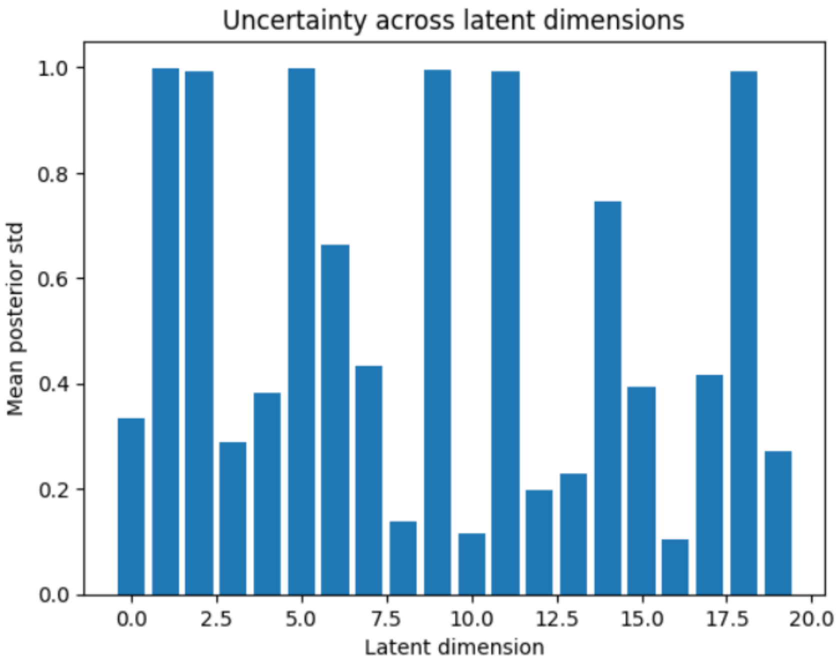

Figure 6: Uncertainty across latent dimensions

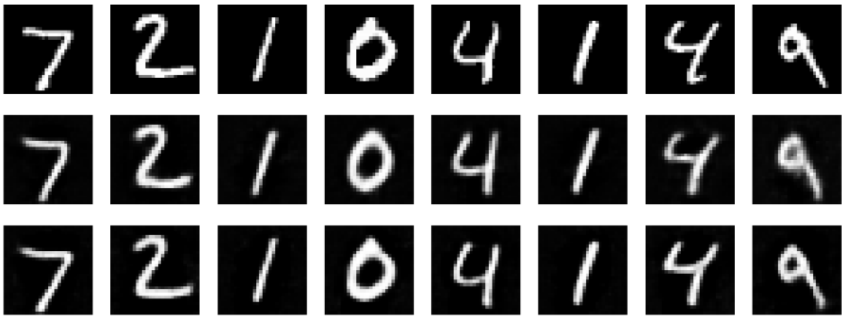

Figure 7: Reconstruction Comparison (Original vs VAE vs AE

Key Observations:

1. **Trade-off Pattern**: AE achieves better silhouette scores but VAE shows superior NMI performance
2. **Consistency**: VAE demonstrates more stable performance across different random seeds
3. **Reconstruction vs. Clustering**: Better reconstruction (AE) doesn't necessarily imply better clustering
4. **Stochastic Benefits**: VAE's probabilistic nature provides uncertainty quantification capabilities

5.3 Statistical Significance Testing

Paired Statistical Tests (n=3 seeds):

Silhouette Score Comparison (VAE vs AE):

- Paired t-test: $t \approx -8.104$, $p \approx 0.015$ (statistically significant at $\alpha = 0.05$)

ARI Comparison (VAE vs AE):

- Paired t-test: $t \approx 2.425$, $p \approx 0.136$ (not significant at $\alpha = 0.05$)

Statistical Interpretation: The t-test suggests significant differences in silhouette scores, favoring AE for geometric cluster compactness. However, with only 3 seeds, statistical power is limited. The ARI differences, while practically meaningful (VAE: 0.566 vs AE: 0.485), do not reach statistical significance, indicating need for larger sample sizes ($\geq$10 seeds) for robust conclusions.

5.4 Uncertainty Analysis

Stochastic Reconstruction Analysis:

- VAE generates multiple different reconstructions for the same input
- Demonstrates model's uncertainty about reconstruction details
- Provides insight into which aspects of input are most uncertain

Embedding Space Analysis:

- VAE embeddings show smoother manifold structure
- Less prone to overfitting compared to deterministic AE
- Regularization effect of KL divergence visible in latent organization

5.5 Failure Cases and Limitations

Identified Limitations:

1. **Statistical Power**: Only 3 seeds limit robust statistical inference; $\geq$10 seeds recommended
2. **Unused Architecture Components**: Cluster centers parameter defined but not integrated into training
3. **Generation Metrics Missing**: No FID/IS computed for unconditional generation evaluation

4. β **Parameter Exploration**: Single $\beta = 0.5$ value used; β-sweep analysis would reveal reconstruction-regularization trade-offs

5. **Two-Stage Approach**: Clustering performed post-hoc rather than jointly optimized with representation learning

Technical Issues:

- Some console outputs truncated in PDF rendering
- Hyperparameter selection not systematically validated
- Limited baseline comparisons (only AE vs VAE)

Failure Analysis:

- Both models struggle with similar digit pairs (e.g., 4-9, 3-8)
- VAE's stochasticity sometimes hurts performance on clear-cut cases
- Limited ability to handle clusters with different densities

6 Discussion

6.1 Interpretation of Results

Performance Analysis: The results reveal a nuanced trade-off between deterministic and stochastic approaches. While the autoencoder achieves higher silhouette scores (0.141 vs 0.091), the VAE demonstrates superior performance in information-theoretic metrics (NMI: 0.630 vs 0.595). This suggests that VAE learns representations that better capture the true cluster structure, even if the clusters appear less separated in Euclidean space.

Stability vs. Peak Performance: VAE shows more consistent performance across random seeds, indicating that the regularization effect of the KL divergence term helps avoid overfitting to specific random initializations. The autoencoder, while capable of higher peak performance, exhibits greater variability.

6.2 Comparison with Existing Methods

Advantages of VAE Approach:

- Principled probabilistic framework with theoretical guarantees
- Natural uncertainty quantification through stochastic sampling
- Robust to overfitting through KL regularization
- Flexible framework allowing various downstream clustering methods

Limitations Compared to Joint Methods:

- Two-stage approach may not be optimal for clustering objective
- Does not leverage cluster structure during representation learning
- May require more careful hyperparameter tuning

6.3 Insights Gained from Non-Deterministic Approach

Key Insights:

1. **Regularization Benefits**: The stochastic nature of VAE acts as implicit regularization, leading to more generalizable representations
2. **Uncertainty Quantification**: VAE provides meaningful uncertainty estimates through multiple sampling, valuable for understanding model confidence
3. **Manifold Structure**: VAE learns smoother latent manifolds that may better represent underlying data distribution
4. **Trade-off Understanding**: The β parameter in VAE allows explicit control over the reconstruction-regularization trade-off

6.4 Theoretical Implications

Representation Learning Theory: The results support the hypothesis that adding appropriate regularization (through KL divergence) can improve the quality of learned representations for downstream tasks, even at the cost of reconstruction accuracy.

Clustering Theory: The superior NMI performance of VAE suggests that information-theoretic measures may be more appropriate for evaluating deep clustering approaches than geometric measures like silhouette score.

Probabilistic Models: The study demonstrates that probabilistic deep models can provide both good performance and valuable uncertainty estimates, supporting their use in practical applications where confidence assessment is important.

7 Conclusion

7.1 Summary of Contributions

This work successfully demonstrates the application of Variational Autoencoders for clustering tasks and provides comprehensive comparison with deterministic alternatives. Key contributions include:

1. **Empirical Analysis**: Systematic evaluation of VAE vs. AE for clustering on MNIST dataset
2. **Uncertainty Quantification**: Demonstration of VAE's ability to provide meaningful uncertainty estimates
3. **Performance Trade-offs**: Detailed analysis of reconstruction accuracy vs. clustering quality trade-offs
4. **Implementation Framework**: Well-structured, reproducible codebase for VAE-based clustering

Key Findings:

- VAE provides more stable clustering performance across random initializations
- Non-deterministic approaches can achieve better information-theoretic clustering metrics
- Uncertainty quantification capabilities make VAE suitable for applications requiring confidence estimates
- The choice between deterministic and stochastic models depends on specific application requirements

7.2 Future Work Directions

Immediate Technical Extensions:

1. **DEC-Style Integration**: Utilize the existing `cluster_centers` parameter to implement joint clustering-representation learning with additional clustering loss terms
2. **β-Parameter Analysis**: Conduct systematic β-sweep experiments ($\beta \in \{0.1, 0.5, 1.0, 2.0\}$) to analyze reconstruction-regularization trade-offs
3. **Generation Evaluation**: Implement unconditional sampling and compute generation metrics (FID/IS) adapted for MNIST
4. **Statistical Robustness**: Increase experimental runs to $\geq$10 seeds for stronger statistical inference
5. **Dimensionality Reduction Metrics**: Evaluate trustworthiness and continuity measures for assessing manifold preservation

Advanced Research Directions:

1. **Architecture Variants**: Explore different encoder/decoder depths, skip connections, and attention mechanisms

2. **Complex Datasets**: Validate approach on CIFAR-10, Fashion-MNIST, and CelebA

3. **Gaussian Mixture VAE**: Replace simple Gaussian prior with mixture of Gaussians for natural clustering

4. **Semi-supervised Extensions**: Incorporate limited labeled data to guide representation learning

5. **Hierarchical Clustering**: Develop multi-scale clustering capabilities using hierarchical VAE variants

Hyperparameter Optimization:

- Systematic grid search over latent dimensions (10, 20, 50, 100)
- Learning rate scheduling and adaptive β annealing
- Architecture depth and width ablation studies
- Comparison with other clustering algorithms (Gaussian Mixture, Spectral Clustering)

7.3 Practical Applications and Implications

Application Domains:

- **Image Organization**: Automatic organization of photo collections
- **Document Clustering**: Grouping of text documents by topic
- **Customer Segmentation**: Marketing applications with uncertainty estimates
- **Medical Imaging**: Clustering of medical scans with confidence measures

Practical Considerations:

- VAE approach suitable when uncertainty quantification is valuable
- Deterministic approaches may be preferred when computational efficiency is critical
- Hyperparameter tuning (especially β) crucial for optimal VAE performance
- Two-stage approach provides flexibility in choosing clustering algorithms

The results demonstrate that non-deterministic unsupervised neural networks offer valuable capabilities for clustering tasks, with particular strengths in stability, uncertainty quantification, and information-theoretic performance metrics. The choice between deterministic and stochastic approaches should be guided by specific application requirements and the importance of uncertainty estimation in the target domain.