Few-Shot Image Generation Using Cross-Aligned and Vector Quantized VAE with Meta-Learning

A Novel Approach for Limited Data Scenarios

Contents

Problem Statement

- **Challenge**: Image generation with limited data
- **Current Limitations**:
  - Traditional models require large datasets and more training time
  - Poor performance in few-shot scenarios
  - Difficulty maintaining image quality
- **Goal**: Generate quality and diverse images from 10-30 examples with less training time

Research Objectives

Primary Goals

- Few-shot framework development
- Cross-modal learning
- High-quality generation
- Text-guided synthesis

Technical Goals

- Modal alignment
- Stable training
- Efficient meta-learning
- Comprehensive loss design

Core Architecture Overview

- **Hybrid Design**:
  - VAE component for continuous latent space
  - VQ-VAE for discrete semantic representation
  - Cross-modal alignment mechanisms
- **Key Components**:
  - Image encoder/decoder
  - Semantic processor
  - Vector quantizer
  - Distribution and cross alignment module

Architectural Components

```
class CADA_VAE(nn.Module):
    def __init__(self, latent_size, num_embeddings, embedding_dim):
        self.image_encoder = ImageEncoder(latent_size)
        self.image_decoder = ImageDecoder(latent_size)
        self.semantic_encoder = SemanticEncoderVQVAE(
            latent_size, num_embeddings, embedding_dim)
        self.semantic_decoder = SemanticDecoder(latent_size)
```

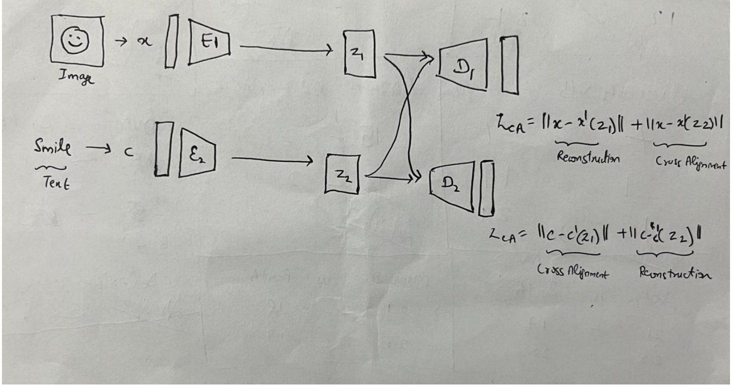


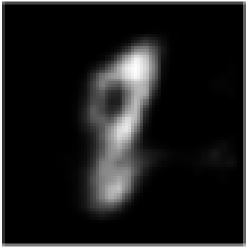
Figure: Model Architecture

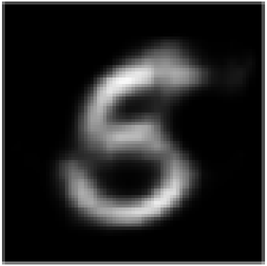
Loss Function Components

- **Standard VAE Loss**:
  - Reconstruction term
  - KL divergence term ($= 2.0$)
- **Cross-Alignment Loss**:
  - Bidirectional mapping
  - Semantic consistency
- **Distribution Alignment Loss**:
  - Distribution alignment
  - Feature matching
- **Contrastive Loss**:
  - Feature discrimination
  - Class separation

MNIST Dataset

- Initial validation dataset
- Simple binary pixel representation
- Clear class separation
- Generated Results:





CIFAR-10 Implementation

- More complex natural images
- 32x32 RGB format
- 10 diverse object classes
- Challenges:
    - Color complexity
    - Object variation
    - Textural details
    - High quality generation

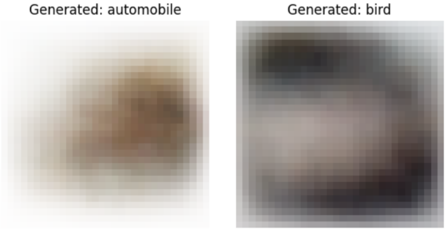


Figure: CIFAR-10 generated samples

Fashion-MNIST: Standard Training

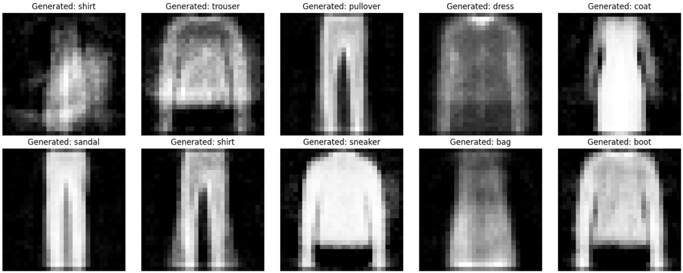


Figure: Full dataset training without meta-learning

- Baseline performance with traditional training
- Used complete dataset
- Shows potential but lacks consistency

Fashion-MNIST: Meta-Training

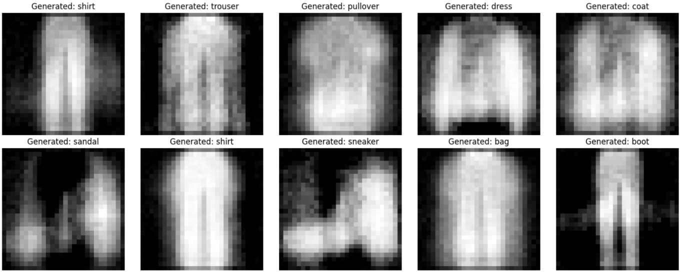


Figure: 30 images per class with meta-learning

- Significant improvement with meta-learning
- Better quality despite limited data
- More consistent style and structure

Standard Training Dynamics

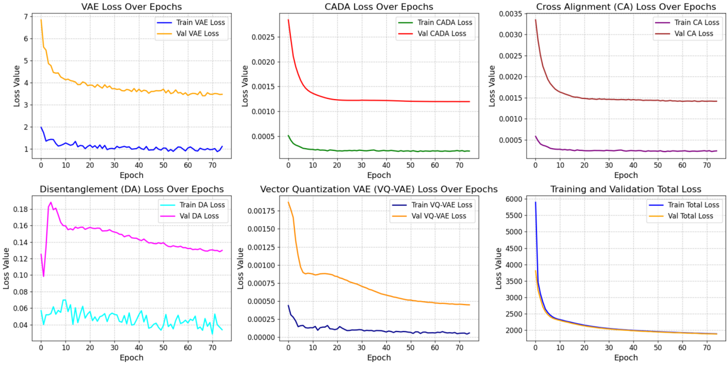


Figure: Loss progression without meta-training

- Higher overall loss values
- Less stable convergence
- Slower learning rate

Meta-Training Dynamics

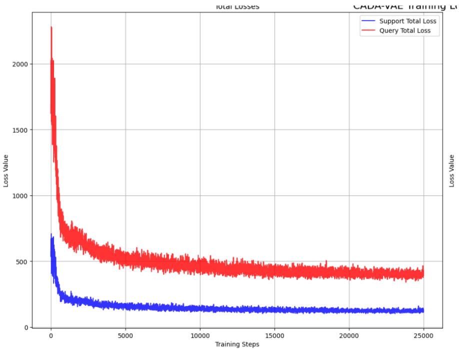


Figure: Loss progression with meta learning

- More stable convergence
- Better loss optimization
- Efficient learning from limited data

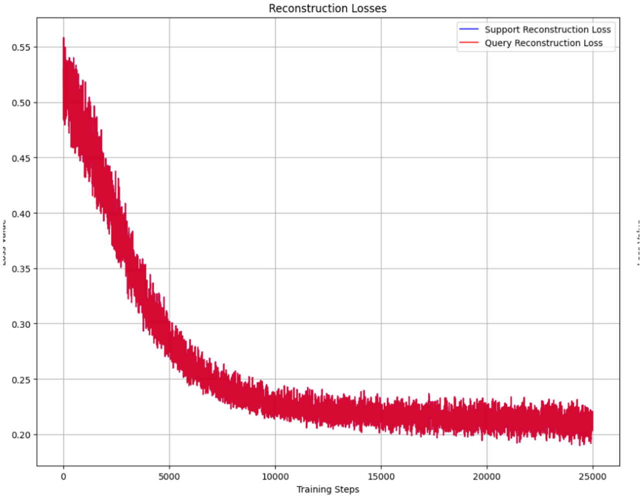


Figure: Reconstruction loss

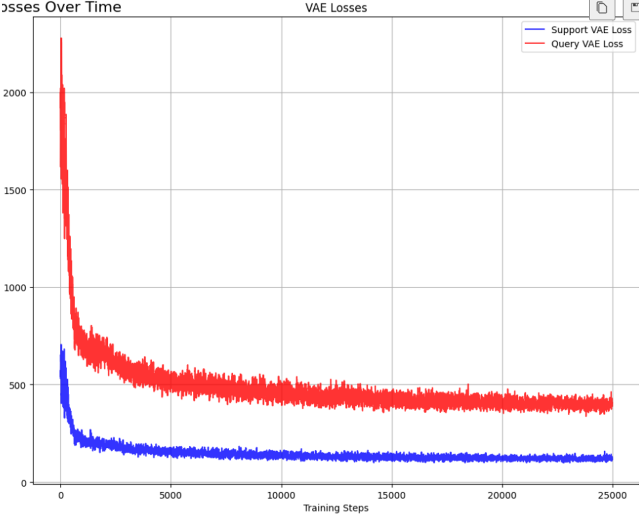


Figure: VAE loss

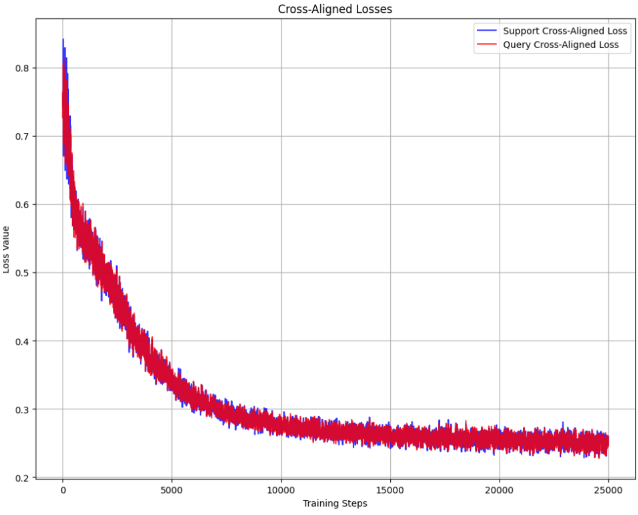


Figure: Cross Alignment loss

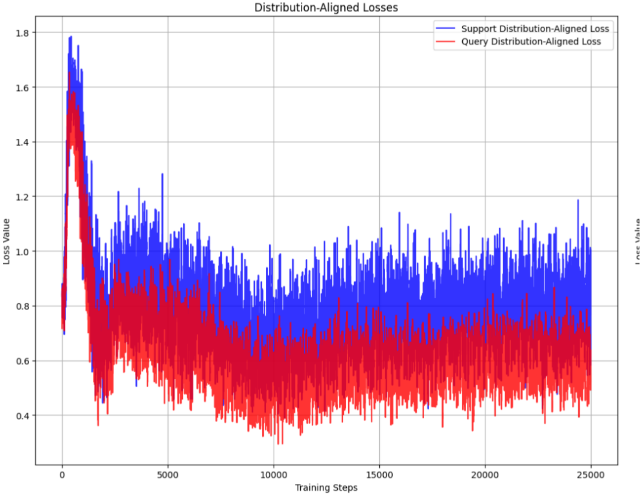


Figure: Distribution Alignment Loss

Standard Training Latent Space

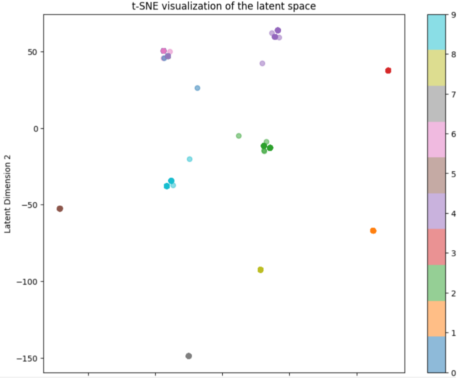


Figure: 100 epochs

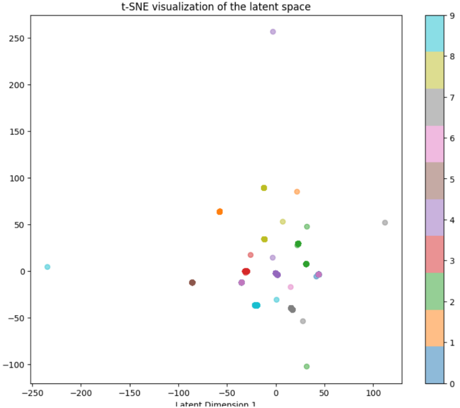


Figure: 75 epochs

- Descent distinct class separation
- Overlapping clusters
- Inconsistent distribution

Meta-Training Latent Space

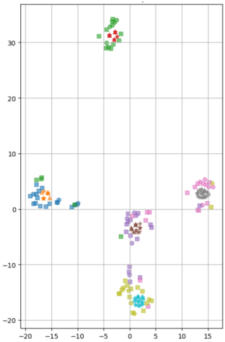


Figure: 30 images per class with meta-learning and incorporation of contrastive loss

- Clear class separation
- Well-defined clusters
- Better semantic organization

Class Reference

- **Fashion-MNIST Classes**:
  - T-shirt/top
  - Trouser
  - Pullover
  - Dress
  - Coat
  - Sandal
  - Shirt
  - Sneaker
  - Bag
  - Ankle boot

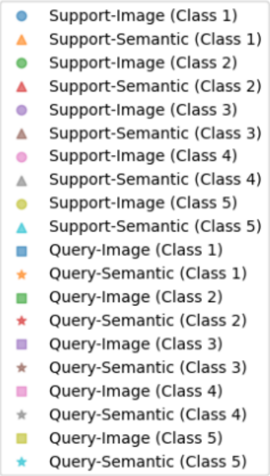


Figure: Class indices

Advantages of Meta-Learning

- **Data Efficiency**:
  - Learns from limited examples
  - Better generalization than memoization
  - Reduced data requirements
- **Training Stability**:
  - More consistent convergence
  - Better loss optimization
  - Reduced overfitting
- **Generation Quality**:
  - Improved image fidelity
  - Better class distinction and latent space separation
  - More consistent style

Technical Improvements

- **Distribution Alignment**:
  - MMD loss effectiveness
  - Better modal matching
  - Improved feature distribution
- **Feature Learning**:
  - Enhanced discrimination
  - Better semantic capture
  - Improved class separation
- **Cross-Modal Integration**:
  - Text-image alignment
  - Semantic consistency
  - Better generation control

Key Achievements

- **Technical Success**:
  - Descent few-shot generation for simple datasets
  - Effective meta-learning integration
  - Stable training framework Can work with less data
- **Performance Gains**:
  - Better latent organization
  - Improved efficiency
- **Future Potential**:
  - Scalability to complex datasets
  - Real-world applications
  - Further architectural improvements

Final Remarks

- Successfully developed novel few-shot framework
- Demonstrated viability on multiple datasets
- Achieved quality results with limited data
- Established foundation for future research