Assignment: Variational Autoencoders for Color Image Synthesis

Foundational Models and Generative AI 2026

Dataset: CIFAR-10 (32x32 RGB images)

1 Objective

The goal of this assignment is to build an AI model capable of generating its own color images from scratch. You will develop a **Variational Autoencoder (VAE)**, explore how it organizes high-dimensional information within its "hidden memory" (latent space), and experiment with a modified version known as a **β-VAE** to understand the trade-offs in generative modeling.

2 Implementation Tasks

Task 1: Architectural Design

You are required to build two neural networks that work in tandem:

- **The Encoder:** A network that processes a color image and compresses it into two sets of numbers representing the "mean" and "variance" of that image's distribution.
- **The Decoder:** A network that takes a sampled "hidden code" (latent vector) and transforms it back into a full 32×32 color image.
- **The Bridge:** Implement the "reparameterization trick." This is the mathematical step that allows the model to sample hidden codes while remaining trainable via backpropagation.

Task 2: Training and Performance

Train your model using the CIFAR-10 dataset.

- You must design the architecture, deciding the appropriate number of layers and filters to handle RGB color channels and image details effectively.
- Sufficient training is required so that the generated images resemble recognizable objects (e.g., cars, birds, airplanes) rather than unstructured noise.

Task 3: Latent Space Interpolation

Test the model's "imagination" by blending two distinct images together.

- **The Procedure:** Select two random points (z_1 and z_2) from the model's hidden memory. Generate a sequence of 10 images by moving in linear increments from the first point to the second.

- **Deliverable:** A plot displaying this 10-step "morphing" process, showing the smooth transition between classes.

Task 4: [Advanced] β-VAE Modification

Extend your standard VAE by modifying the loss function. Introduce a multiplier, β, to the term that handles the hidden memory (the KL Divergence).

- **The Experiment:** Compare a baseline version where $\beta = 1$ to a version where β is significantly higher (e.g., $\beta = 5$).
- **Analysis:** Provide an explanation of how this change impacts the visual quality of the images versus the model's ability to organize independent features (such as shape, color, or orientation).

3 Submission Requirements

Please submit your completed Google Colab notebook containing:

1. Full source code for the **Encoder**, **Decoder**, and the modified **Loss function**.
2. Training curves (graphs) showing the model's improvement over time.
3. A generated grid of 16 images "dreamed up" by the model from random noise.
4. The 10-step **interpolation visualization**.
5. A brief written summary (max 300 words) discussing the observations made when adjusting the β value.