InfoGAN: Implementation and Results – Final Project Report

Nasykhova Anastasia Slovyagina Anna Tarasova Sofia

December 29, 2025

1 Individual Contributions

The project was developed through a well-defined division of responsibilities, followed by collaborative integration and joint presentation preparation. The team believes that everyone has made an equal contribution to the project.

- **Anastasia Nasykhova**
  - Optimized the majority of the codebase for GPU execution (CUDA), significantly accelerating training.
  - Designed and implemented a unified pipeline for Configuration (class Config), data processing section, and feature extraction module (class SharedFeatureExtractor) — ensuring consistent behavior across all datasets.
  - Conducted training experiments on **MNIST**, **Fashion-MNIST**, and **CelebA**.
- **Anna Slovyagina**
  - Implemented the core InfoGAN architecture in PyTorch, including the **Generator**, **Discriminator**, **Auxiliary Network** and utils section.
  - Performed training and evaluation on the **Chairs** and **SVHN** datasets.
- **Sofia Tarasova**
  - Developed the end-to-end training loop with model checkpointing (weight saving/loading).
  - Implemented visualization tools.
  - Generated the custom **Faces** dataset following the procedure described in the original InfoGAN paper and ran all associated experiments.

After individual development, all team members jointly integrated the components into a cohesive system. To manage computational constraints—some dataset runs required up to 5 hours—each member independently re-executed experiments on their assigned datasets for final tuning and validation.

Additionally, all three team members contributed equally to the preparation of all project presentations, including slides and speaker notes for the *Project Proposal*, *Intermediate Status Report*, and *Final Defense*.

2 Main Insights from the Project

Through rigorous experimentation across six datasets (MNIST, Fashion-MNIST, SVHN, CelebA, Chairs, and Faces), we gained several key insights:

- **Reproducibility is achievable with careful implementation.** We successfully replicated the qualitative results reported in the original InfoGAN paper (Chen et al., 2016). On MNIST, the model consistently learned disentangled representations—for example, one continuous latent variable controlled digit rotation, while categorical variables captured digit identity.

- **Dataset complexity directly impacts disentanglement quality.** Simple, grayscale datasets with clear semantic factors (e.g., MNIST) yielded excellent disentanglement. In contrast, more complex datasets like CelebA and Faces showed partial or moderate disentanglement, often limited to coarse attributes (e.g., presence of glasses or hair color) rather than fine-grained variations.

- **Architecture and hyperparameters are critical.** We confirmed that DCGAN-style design choices—such as batch normalization in the generator and LeakyReLU in the discriminator—stabilize training. However, we also observed that learning rates, batch size, and mutual information weight (λ) required dataset-specific tuning. For instance, Chairs required a smaller batch size (32) due to memory constraints, which slightly affected convergence.

- **Fashion-MNIST served as a successful out-of-domain validation.** Although not included in the original paper, we chose to experiment with Fashion-MNIST as a completely new dataset to test the generalizability of the InfoGAN framework. The model demonstrated clear disentanglement of semantic attributes (e.g., distinguishing between types of clothing such as sneakers, coats, or dresses), confirming that the approach adapts well to unseen data distributions with similar structural properties.

Overall, the project validated that mutual information maximization enables unsupervised learning of interpretable latent codes, but its effectiveness is modulated by data complexity, model capacity, and available compute.

3 Future Plans

Building on our current results, we plan to pursue the following directions:

1. **Quantitative evaluation of generation quality and disentanglement.** Currently, our assessment is largely qualitative. We intend to compute established metrics such as Fréchet

Inception Distance (FID) for image quality and Mutual Information Gap (MIG) or SAP score for disentanglement, enabling objective comparison across datasets and architectures.

2. **Architectural enhancements.** We aim to replace the current DCGAN backbone with more modern generators (e.g., StyleGAN2) to improve sample fidelity, especially on high-variance datasets like CelebA. Integrating progressive growing or attention mechanisms could further boost performance.

3. **Interactive latent space exploration.** We will develop a user interface to manipulate latent codes in real time, demonstrating the practical utility of disentangled representations for tasks like image editing or data augmentation.

These steps will help bridge the gap between theoretical disentanglement and real-world applications in controllable generative modeling.