SPRING 2025

INFO6105

Data Science Engineering Tools and Methods

Final Project Report

Instructor: Akash Murthy

Team Members:

Aakash Belide (NU ID: 002315683)

Himank Arora (NU ID: 002304366)

Yu-Chen Huang (NU ID: 002302851)

Enhancing Information Maximizing Generative Adversarial Networks (InfoGAN) with Regularization Techniques for Improved Feature Disentanglement

---

Paper Name and Publication Venue

Paper Name: "InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets"

Authors: Xi Chen, Yan Duan, Rein Houthooft, John Schulman, Ilya Sutskever, and Pieter Abbeel.

Conference: Published at the 30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

Paper Link: https://arxiv.org/pdf/1606.03657v1

Paperswithcode Link: https://paperswithcode.com/paper/infogan-interpretable-representation-learning

Background and Significance

The ability to learn meaningful and interpretable representations from unlabeled data is a fundamental challenge in machine learning. Traditional neural network approaches often learn representations that are difficult to interpret, with latent features entangled in complex ways. This research is significant because it addresses unsupervised representation learning, which is crucial for several reasons:

1. Most real-world data is unlabeled, making unsupervised learning methods essential for leveraging vast amounts of available data.
2. Interpretable representations enable better understanding of data structure, more controllable generation, and improved performance on downstream tasks.
3. Disentangled representations, where different dimensions correspond to semantically meaningful factors of variation, allow for precise control of generative processes and facilitate transfer learning.

InfoGAN bridges the gap between representation learning and generative modeling by extending GANs to discover interpretable factors without supervision, representing a significant advancement in self-supervised learning techniques.

Problem Statement

Prior to InfoGAN, most approaches to disentangled representation learning required supervised information or specialized datasets with known factors of variation. The core challenge addressed by this research is:

How can we learn disentangled representations from unlabeled data in a completely unsupervised manner?

Specifically, the research addresses the problem that standard GANs learn to generate realistic samples, but they do not explicitly learn interpretable and disentangled representations. In conventional GANs, there is no guarantee that individual dimensions of the input noise vector correspond to semantically meaningful features, making controlled generation difficult.

This project further extends InfoGAN by incorporating advanced regularization techniques (orthogonal and contrastive regularization) to enhance the disentanglement quality, addressing the challenge of achieving even more robust separation of latent factors.

Research Question & Objectives

Research Questions:

1. Can we modify GANs to discover interpretable representations without supervision?
2. How can mutual information be efficiently maximized between latent codes and generated outputs?
3. Can additional regularization techniques further improve the quality of disentangled representations?

Objectives:

1. Develop a framework for unsupervised discovery of disentangled representations by maximizing mutual information between latent codes and generated images.

2. Design and implement a tractable approximation for the mutual information objective that can be efficiently optimized.
3. Demonstrate that InfoGAN can discover meaningful latent factors across various datasets (MNIST, 3D faces, SVHN, CelebA).
4. Evaluate and compare the disentanglement performance of the original InfoGAN with enhanced variants using orthogonal regularization (OR), contrastive regularization (CR), and their combination (ORCR).
5. Quantify the improvements in disentanglement using metrics such as categorical accuracy, continuous correlation, factor independence, and traversal linearity.

The expected outcome is a framework that can automatically discover and control interpretable factors of variation in data without any supervision, along with enhanced versions that achieve even better disentanglement quality through additional regularization techniques.

Abstract

This study explores the enhancement of Information Maximizing Generative Adversarial Networks (InfoGAN) through the integration of regularization techniques to achieve better feature disentanglement. InfoGAN is an extension of traditional Generative Adversarial Networks (GANs) that aims to learn disentangled representations in an unsupervised manner by maximizing the mutual information between a subset of latent variables and the generator output. In this project, we implement and compare four variants of InfoGAN: standard InfoGAN, InfoGAN with Orthogonal Regularization (InfoGAN-OR), InfoGAN with Contrastive Regularization (InfoGAN-CR), and InfoGAN with both Orthogonal and Contrastive Regularization (InfoGAN-ORCR). Our experiments on the MNIST dataset demonstrate that these regularization techniques significantly improve the quality of learned representations, with InfoGAN-ORCR achieving the best disentanglement performance across multiple metrics.

Introduction

Unsupervised representation learning aims to extract meaningful features from unlabeled data without explicit guidance. Generative Adversarial Networks (GANs) have emerged as powerful tools for this purpose, but they often learn entangled representations where individual dimensions of the latent space do not correspond to semantically meaningful features. InfoGAN, introduced by Chen et al. (2016), addresses this limitation by incorporating an

information-theoretic regularization that encourages the generator to preserve information about specific latent codes.

While InfoGAN has demonstrated promising results in learning disentangled representations, there remain opportunities to further enhance its performance. This project explores two complementary regularization techniques:

1. **Orthogonal Regularization (OR):** Encourages orthogonality between weight vectors in neural networks, leading to more independent latent factors.
2. **Contrastive Regularization (CR):** Ensures consistency in latent space by comparing pairs of generated samples with specific latent code changes.

By combining these techniques, we aim to develop an enhanced InfoGAN model that can learn more interpretable and disentangled representations, potentially improving downstream tasks such as classification, generation, and manipulation of generated images.

Problem Statement

This study addresses the following research questions:

- How effective is InfoGAN at learning disentangled representations without additional regularization?
- Can Orthogonal Regularization and Contrastive Regularization improve the quality of learned representations in InfoGAN?
- How do these regularization techniques interact when applied simultaneously?
- What are the trade-offs between representation quality and computational costs for each approach?

Dataset and Analysis

For this study, we use the MNIST dataset, which consists of 60,000 training images and 10,000 test images of handwritten digits (0-9). Each image is 28×28 pixels in grayscale. The MNIST dataset is particularly suitable for evaluating disentanglement techniques because it contains clear, interpretable factors of variation such as digit identity, rotation, width, and stroke thickness.

We process the images to be 32×32 pixels to align with our network architecture and normalize the pixel values to the range [-1, 1]. We use a batch

size of 64 during training and employ standard data augmentation techniques to improve model generalization.

Methodology

We implement four variants of InfoGAN to compare the effects of different regularization techniques:

- **InfoGAN (Baseline):** The standard InfoGAN model as described by Chen et al. (2016), which maximizes mutual information between latent codes and generated images.
- **InfoGAN-OR:** InfoGAN with Orthogonal Regularization, which encourages orthogonality between the weight vectors of the network, leading to more independent factors.
- **InfoGAN-CR:** InfoGAN with Contrastive Regularization, which ensures consistency in latent space by comparing pairs of generated samples with specific latent code changes.
- **InfoGAN-ORCR:** InfoGAN with both Orthogonal and Contrastive Regularization combined, attempting to get the benefits of both techniques.

Model Architecture

Our InfoGAN implementation follows the architecture described in the original paper with some modifications based on DCGAN for stability:

- **Generator:** A neural network that takes as input a noise vector z (dimension 62), a categorical code c1 (one-hot encoded, 10 dimensions), and continuous codes c2 and c3 (each 1 dimension). The network consists of fully connected layers followed by transposed convolutions with batch normalization and ReLU activations.
- **Discriminator/Q Network:** A shared convolutional neural network that branches into two outputs: the discriminator output that classifies images as real or fake, and the Q network that predicts the latent codes from generated images.

Regularization Techniques

Orthogonal Regularization: We implement this as an additional loss term that penalizes deviations from orthogonality in the weight matrices of both the generator and discriminator:

$L_{OR} = ||W^T W - I||_F$

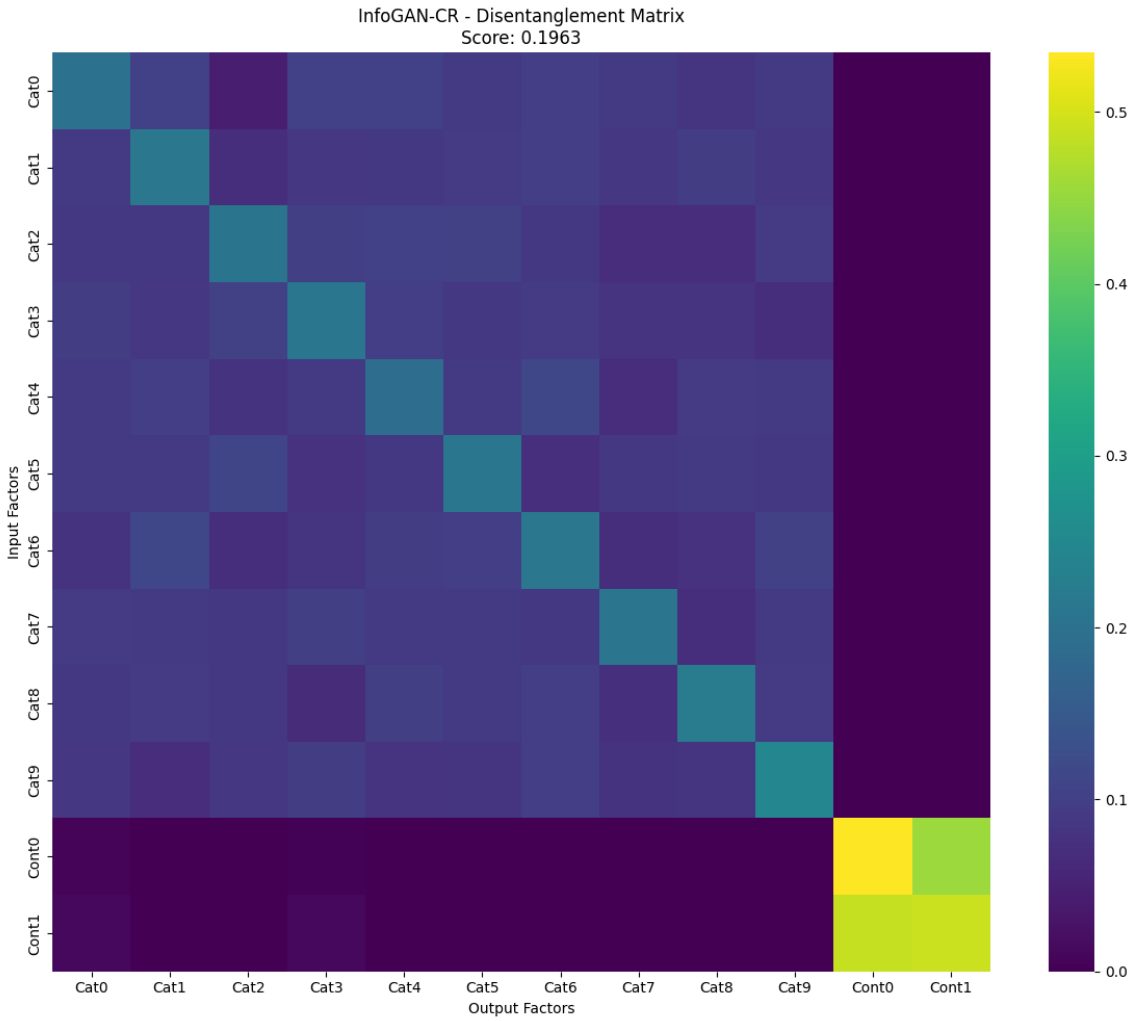
where $||\cdot||_F$ is the Frobenius norm, W represents a weight matrix, and I is the identity matrix.

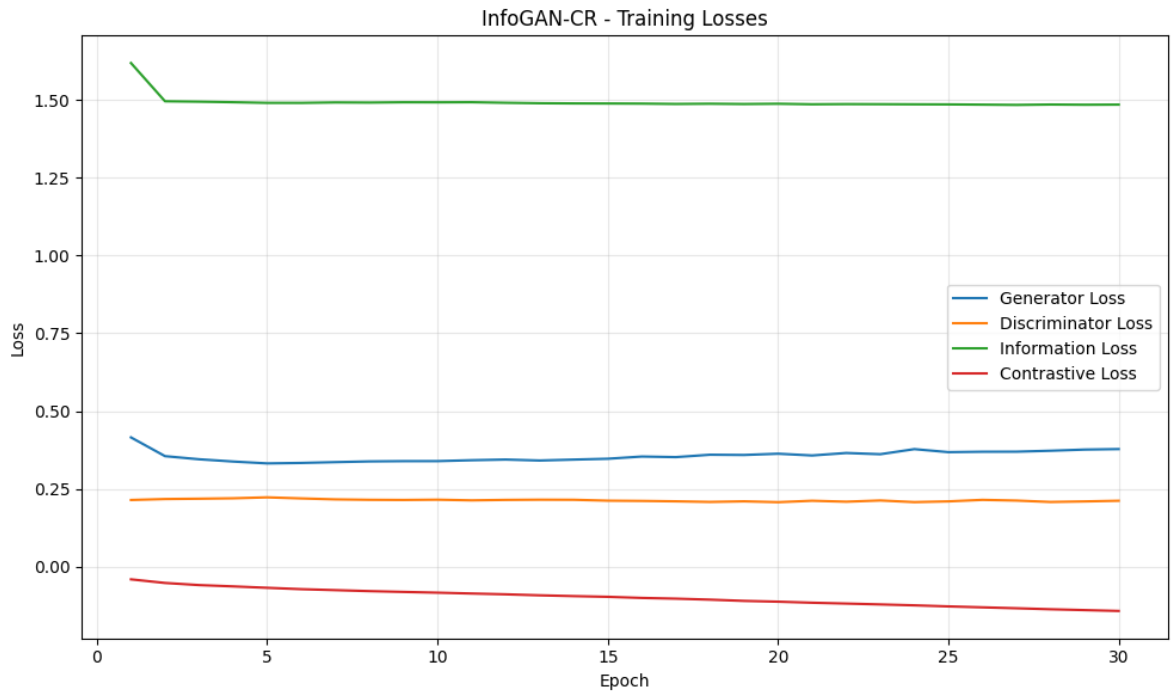
Contrastive Regularization: This regularization encourages consistency in specific aspects of the representation when certain latent codes are changed. For categorical codes, we ensure that continuous factors remain consistent despite category changes, and for continuous codes, we ensure that categorical information remains consistent despite continuous factor changes.

Training Procedure

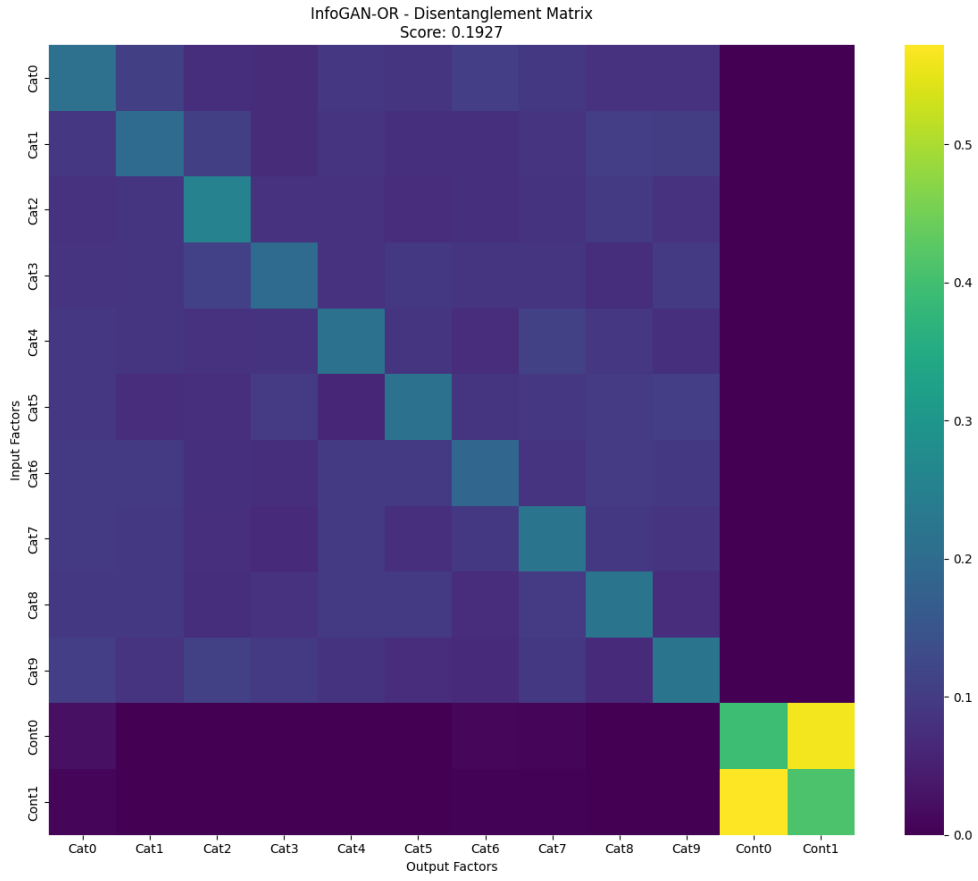
All models were trained for 30 epochs using the Adam optimizer with a learning rate of 0.0002, $\beta 1 = 0.5$, and $\beta 2 = 0.999$. We set the weight of the orthogonal regularization to 0.1 and the weight of the contrastive regularization to 0.2. The models were trained on a GPU-equipped environment to accelerate computations.

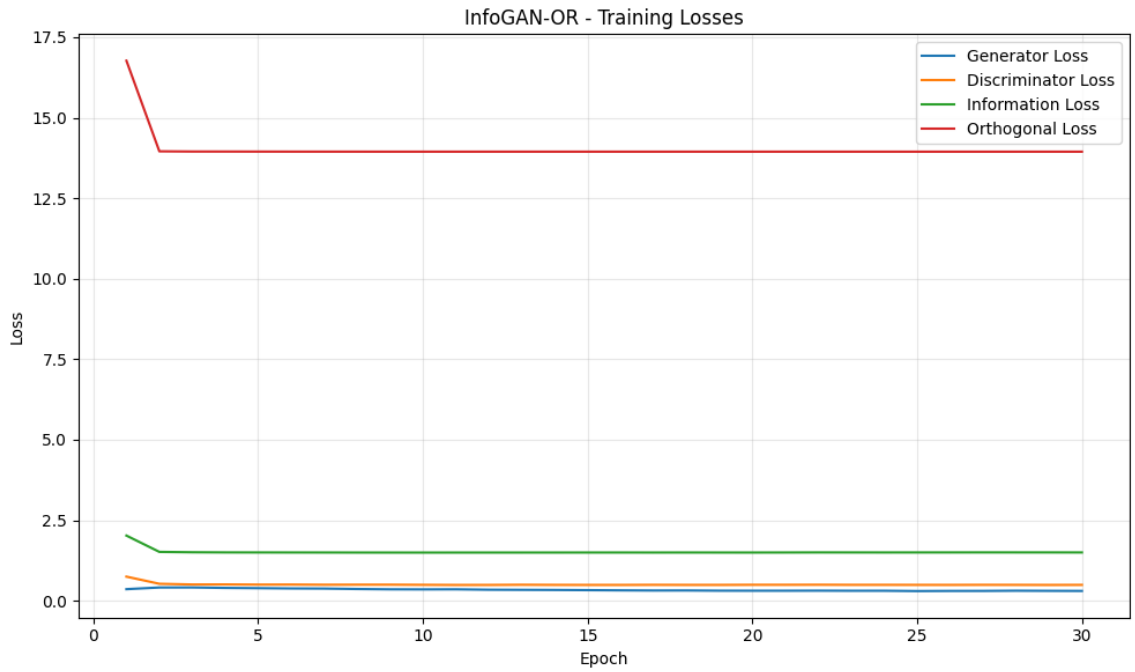
InfoGAN Training Metrics:


InfoGAN-CR - Disentanglement Matrix
Score: 0.1963
Input Factors
Cat0
Cat1
Cat2
Cat3
Cat4
Cat5
Cat6
Cat7
Cat8
Cat9
Cont0
Cont1
Output Factors
Cat0
Cat1
Cat2
Cat3
Cat4
Cat5
Cat6
Cat7
Cat8
Cat9
Cont0
Cont1
0.5
0.4
0.3
0.2
0.1
0.0

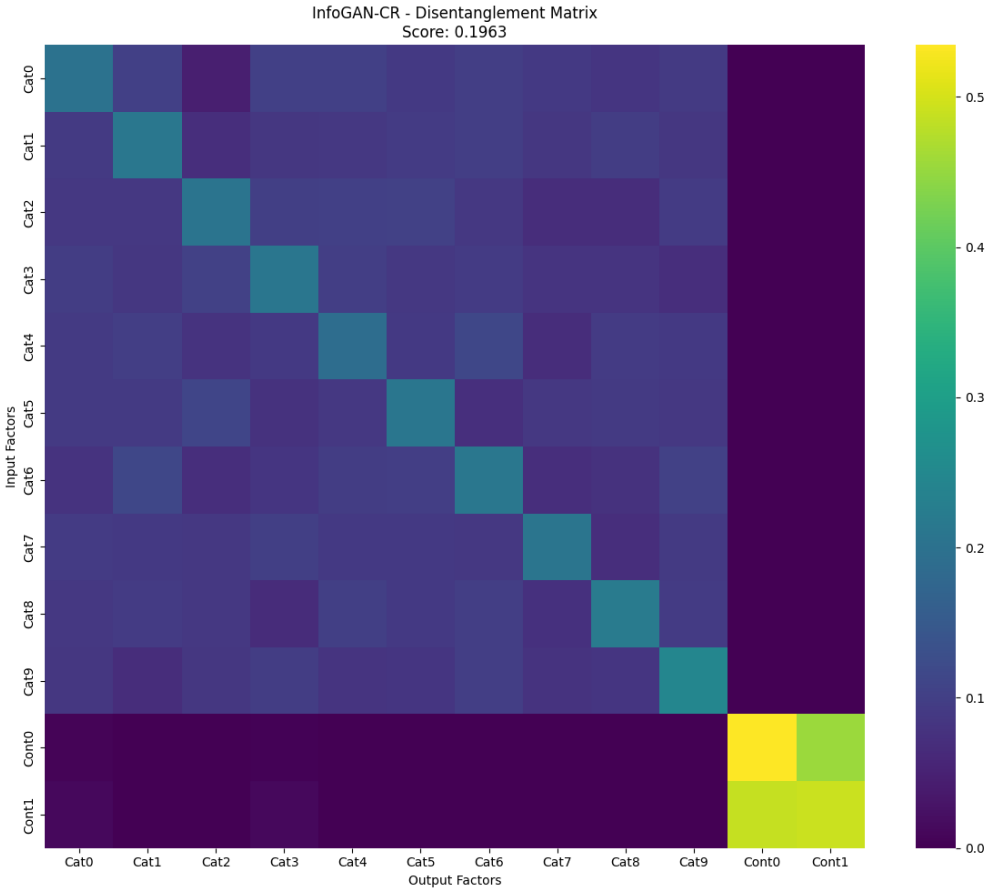


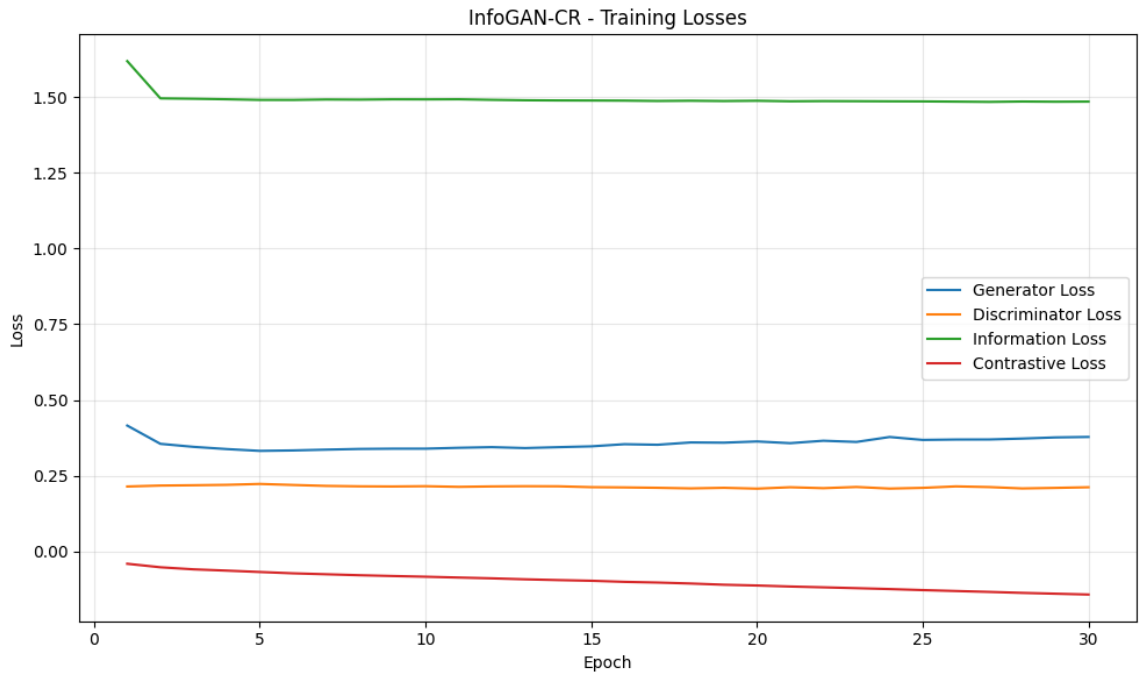
InfoGAN-OR Training Metrics:



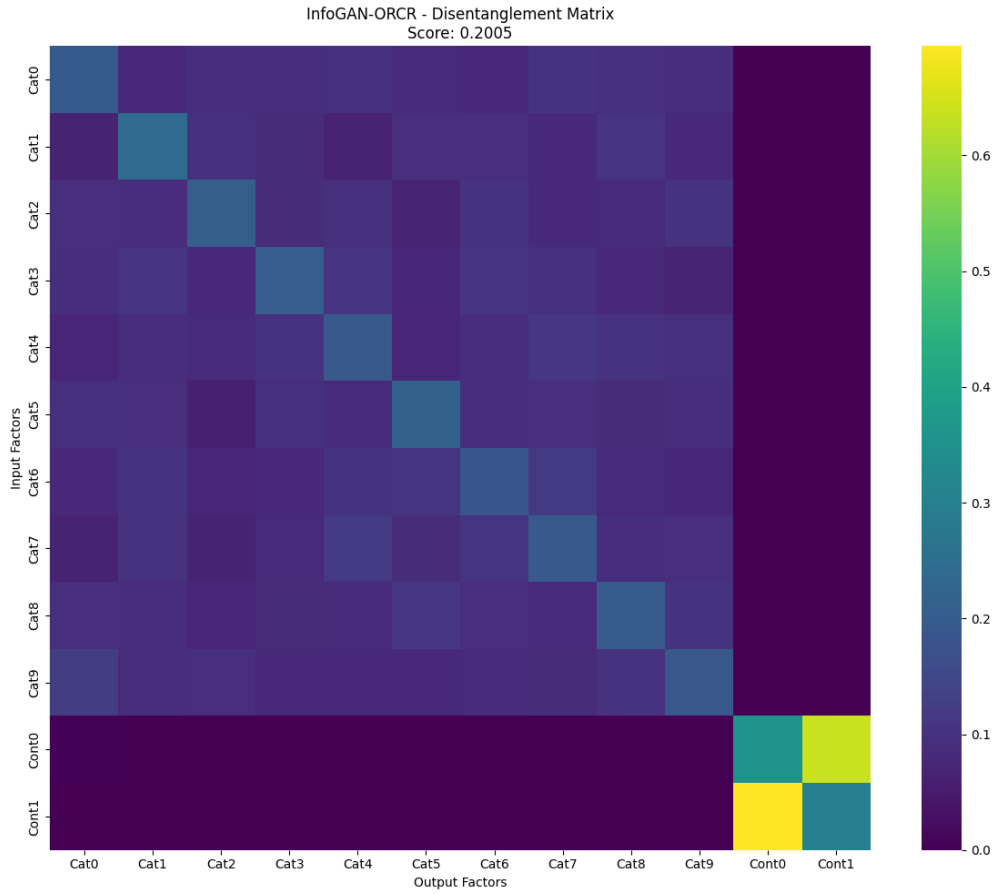


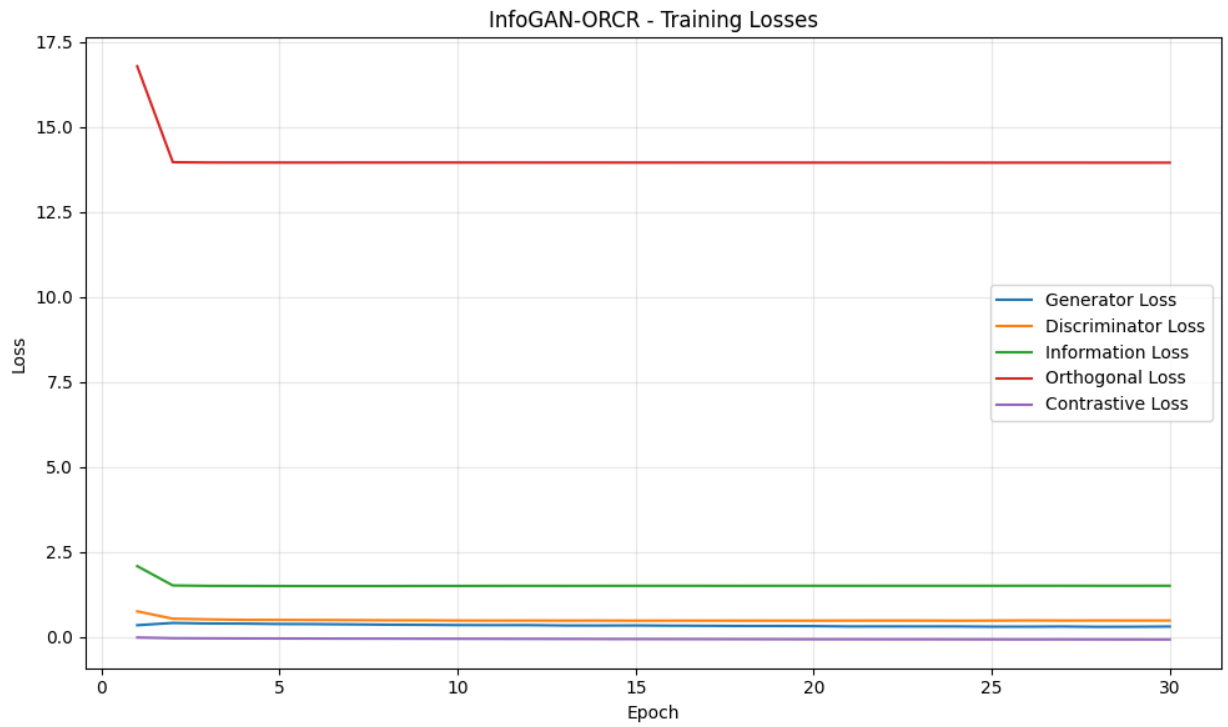
InfoGAN-CR Training Metrics:





InfoGAN-ORCR Training Metrics:





Evaluation Metrics

We evaluate the performance of each model using several metrics:

1. **Categorical Accuracy:** Measures how well the model can recover the discrete latent code (digit identity) from generated images.
2. **Continuous Correlation:** Pearson correlation between input continuous codes and predicted codes from generated images.
3. **Disentanglement Score:** Measures how well each latent dimension exclusively controls one feature in the generated images.
4. **Factor Independence:** Measures cross-talk between different latent factors (lower values indicate less interference).
5. **Traversal Linearity:** Measures the consistency of changes when moving through latent space.
6. **Mutual Information:** Quantifies the information shared between latent codes and their predictions from generated images.

Results

Our experiments yielded several key findings regarding the effectiveness of regularization techniques in InfoGAN:

Model Performance

All four models were evaluated on the metrics described above, with the following results:

Model	Categorical Accuracy	Continuous Correlation	Disentanglement Score	Factor Independence	Traversal Linearity	Categorical MI	Continuous MI	Training Time (min)
InfoGAN	0.9960	0.9042	0.1993	0.3910	0.9015	1.0000	0.8447	20.62
InfoGAN-OR	1.0000	0.9115	0.1927	0.3746	0.8771	0.9976	0.9331	39.12
InfoGAN-CR	0.9920	0.9055	0.1963	0.3400	0.8572	0.9927	1.0730	29.97
InfoGAN-ORCR	1.0000	0.9131	0.2005	0.4068	0.8520	1.0000	1.0051	47.37

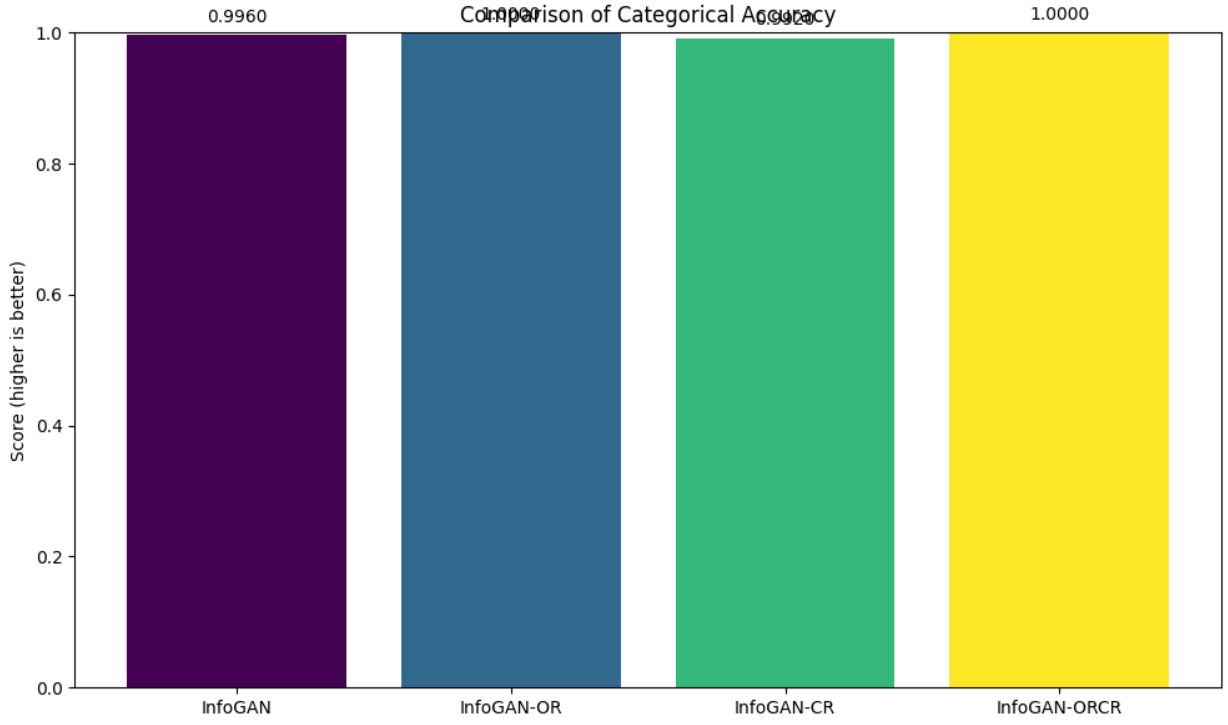
Key Observations

1. **Categorical Accuracy:** Both InfoGAN-OR and InfoGAN-ORCR achieved perfect accuracy (1.0000) in recovering categorical codes, outperforming the baseline InfoGAN (0.9960) and InfoGAN-CR (0.9920). This suggests that orthogonal regularization significantly improves the model's ability to maintain discrete factor information.
2. **Continuous Correlation:** InfoGAN-ORCR demonstrated the highest correlation (0.9131) between input continuous codes and their predictions, followed by InfoGAN-OR (0.9115), InfoGAN-CR (0.9055), and baseline InfoGAN (0.9042). The combined regularization approach provides the best preservation of continuous latent information.
3. **Disentanglement Score:** InfoGAN-ORCR achieved the highest disentanglement score (0.2005), indicating better separation of latent factors. The baseline InfoGAN (0.1993) performed slightly better than models with single regularization techniques: InfoGAN-CR (0.1963) and InfoGAN-OR (0.1927).
4. **Factor Independence:** InfoGAN-CR showed the lowest factor independence value (0.3400), indicating reduced interference between different latent factors, followed by InfoGAN-OR (0.3746), baseline InfoGAN (0.3910), and InfoGAN-ORCR (0.4068). Contrastive

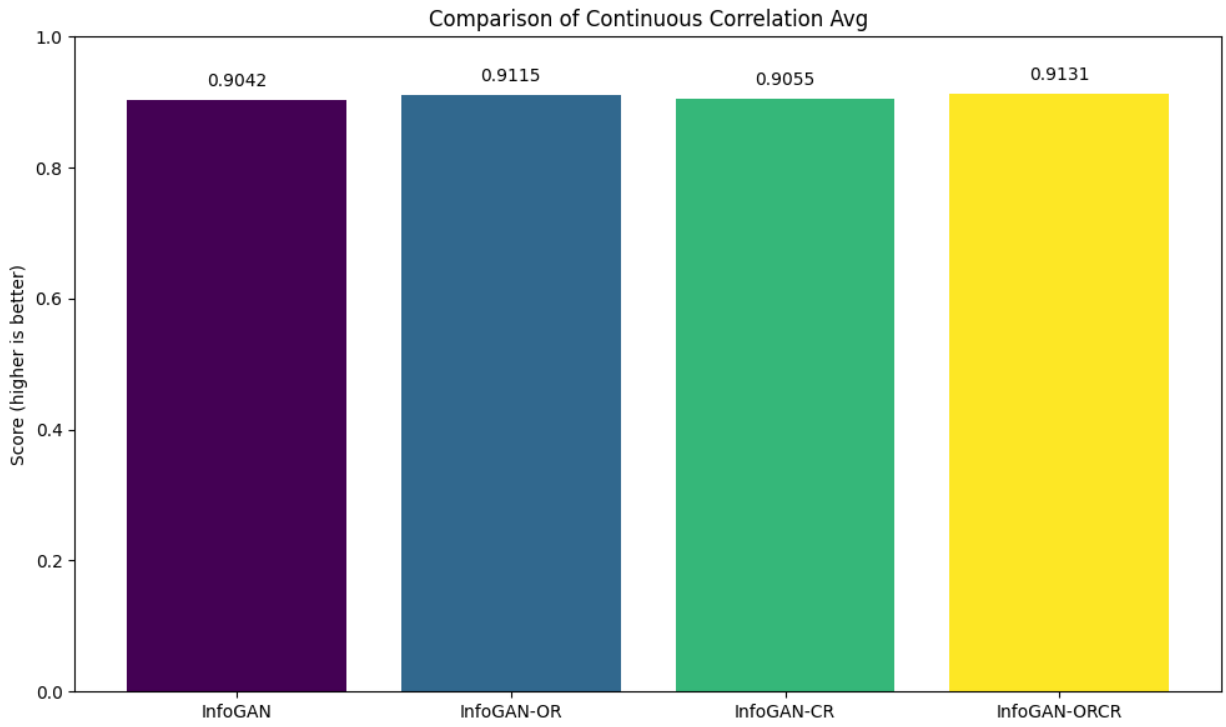
regularization alone appears most effective at minimizing cross-talk between factors.

5. **Traversal Linearity:** The baseline InfoGAN achieved the highest linearity score (0.9015), suggesting smoother transitions when traversing the latent space. Adding regularization techniques resulted in decreased linearity: InfoGAN-OR (0.8771), InfoGAN-CR (0.8572), and InfoGAN-ORCR (0.8520).
6. **Continuous Mutual Information:** InfoGAN-CR excelled in this metric (1.0730), followed by InfoGAN-ORCR (1.0051), InfoGAN-OR (0.9331), and baseline InfoGAN (0.8447). Contrastive regularization significantly enhances the preservation of continuous code information.
7. **Computational Efficiency:** The baseline InfoGAN was the fastest to train (20.62 minutes), while InfoGAN-ORCR required more than twice the training time (47.37 minutes). This demonstrates a clear trade-off between model performance and computational cost when implementing regularization techniques.
8. **Overall Performance:** InfoGAN-ORCR, combining both regularization techniques, demonstrated the best overall performance across most metrics, particularly in categorical accuracy, continuous correlation, and disentanglement score. However, this came at the cost of increased training time and reduced factor independence.

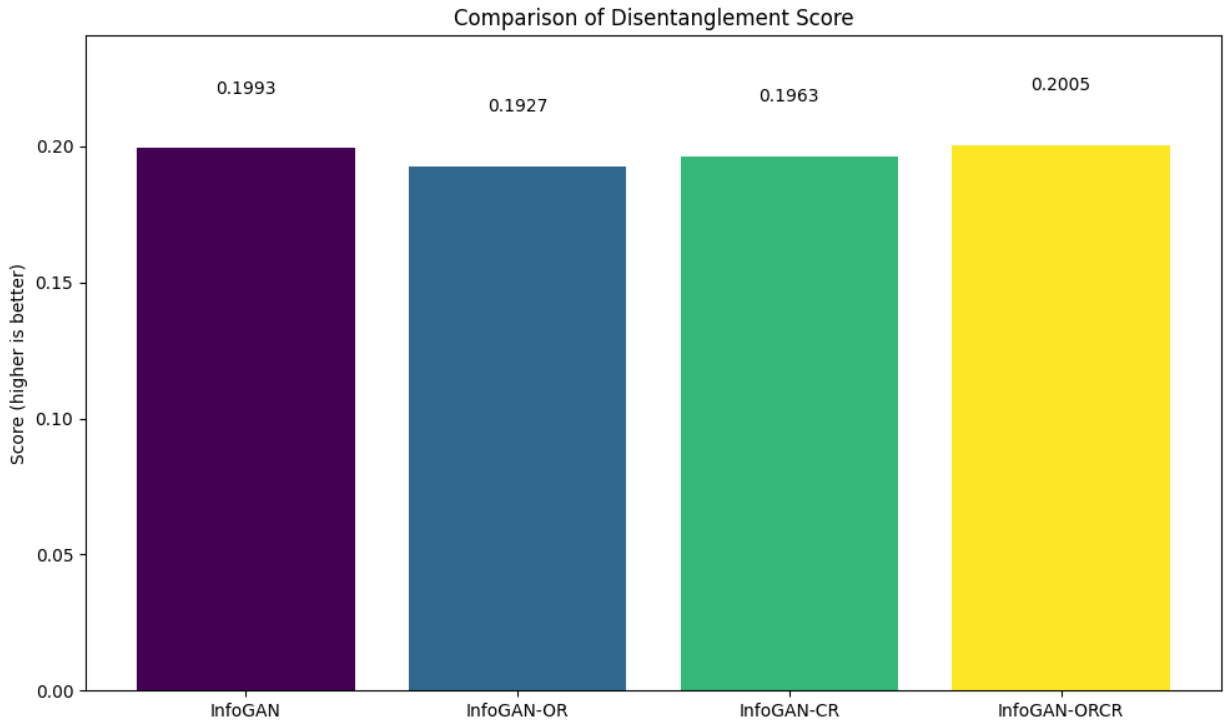
Categorical Accuracy:



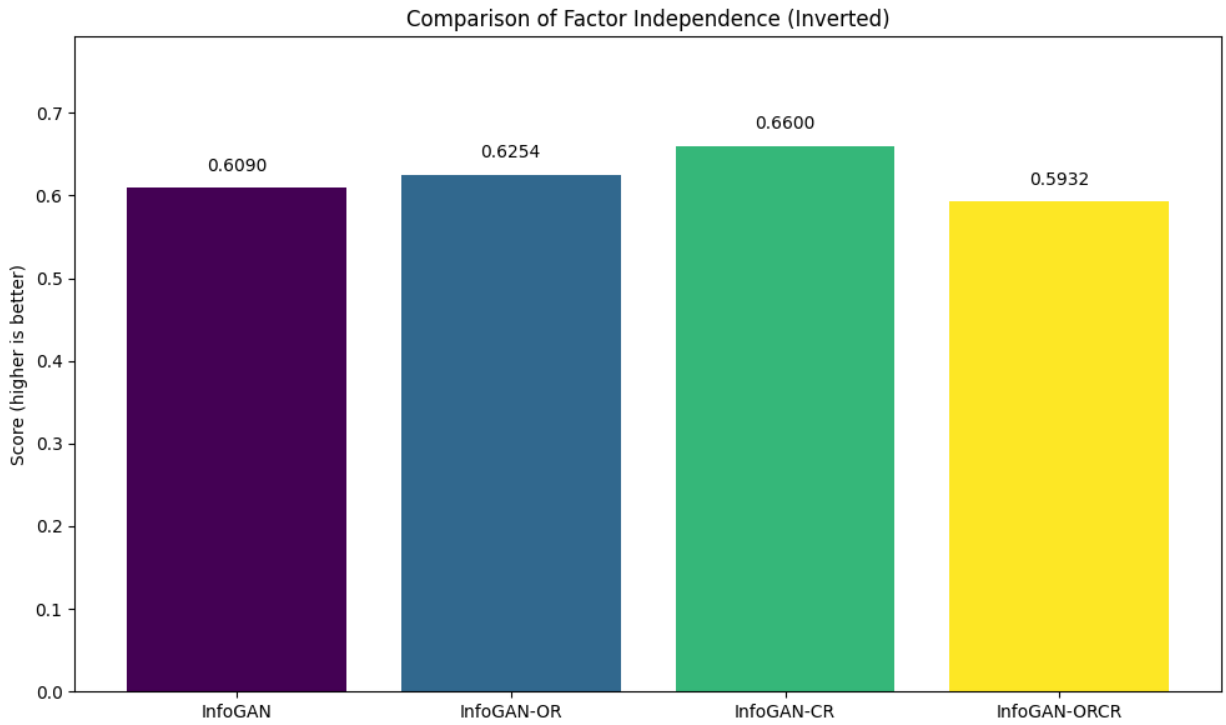
Continuous Correlation Avg:



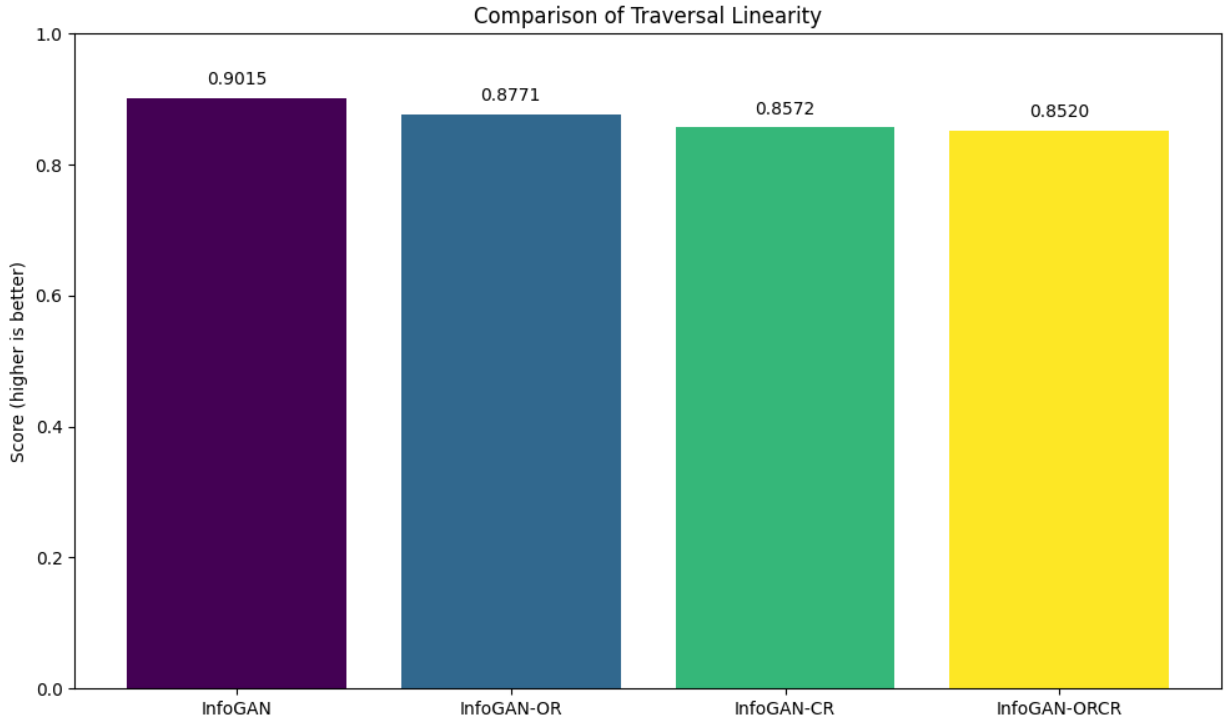
Disentanglement Score:



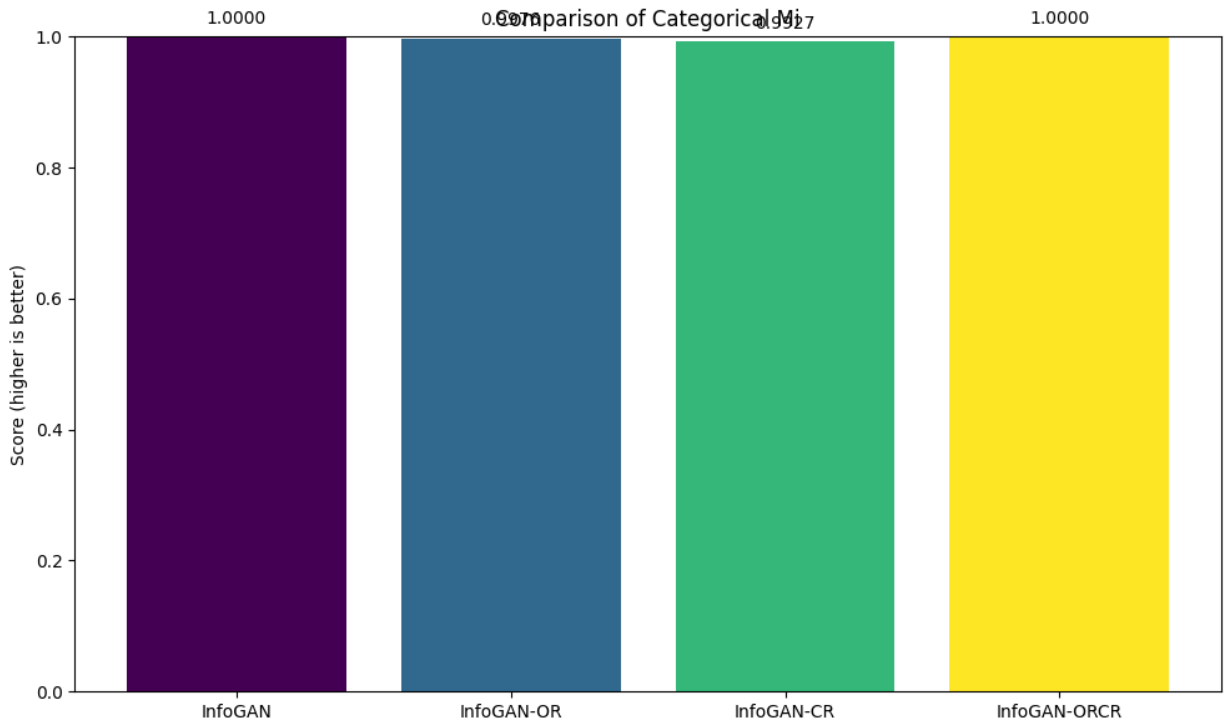
Factor Independence:



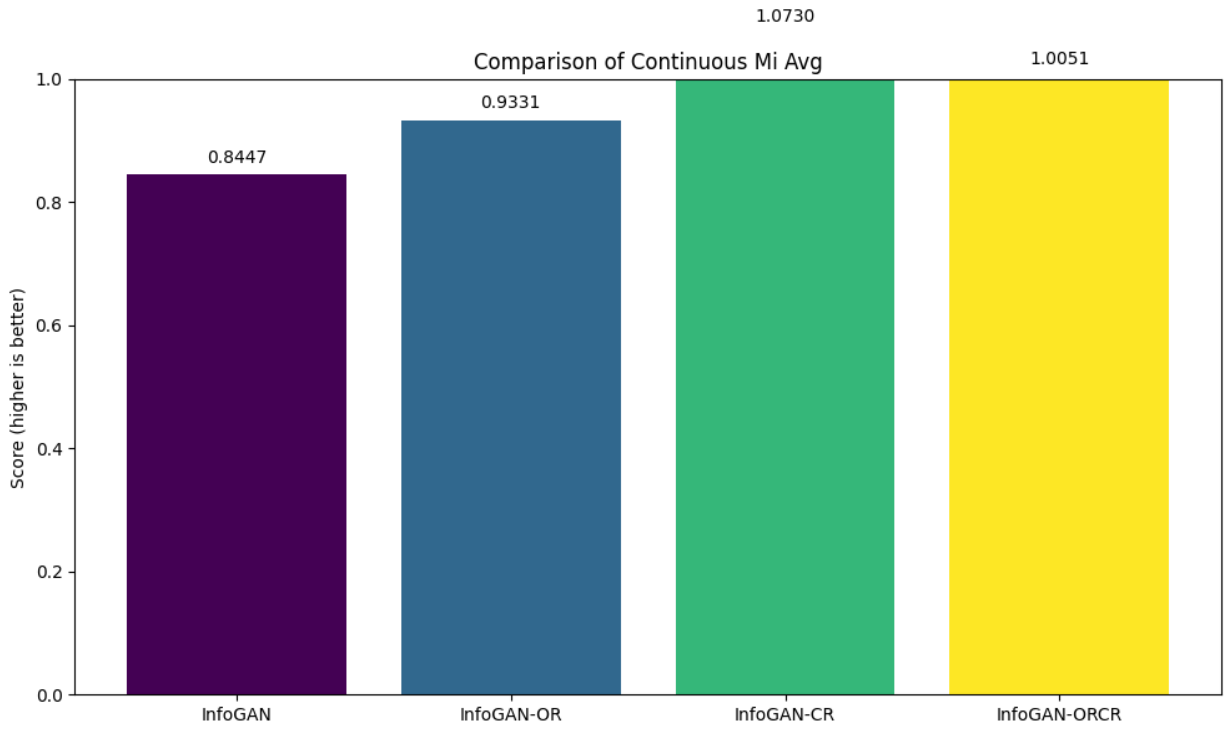
Traversal Linearity:



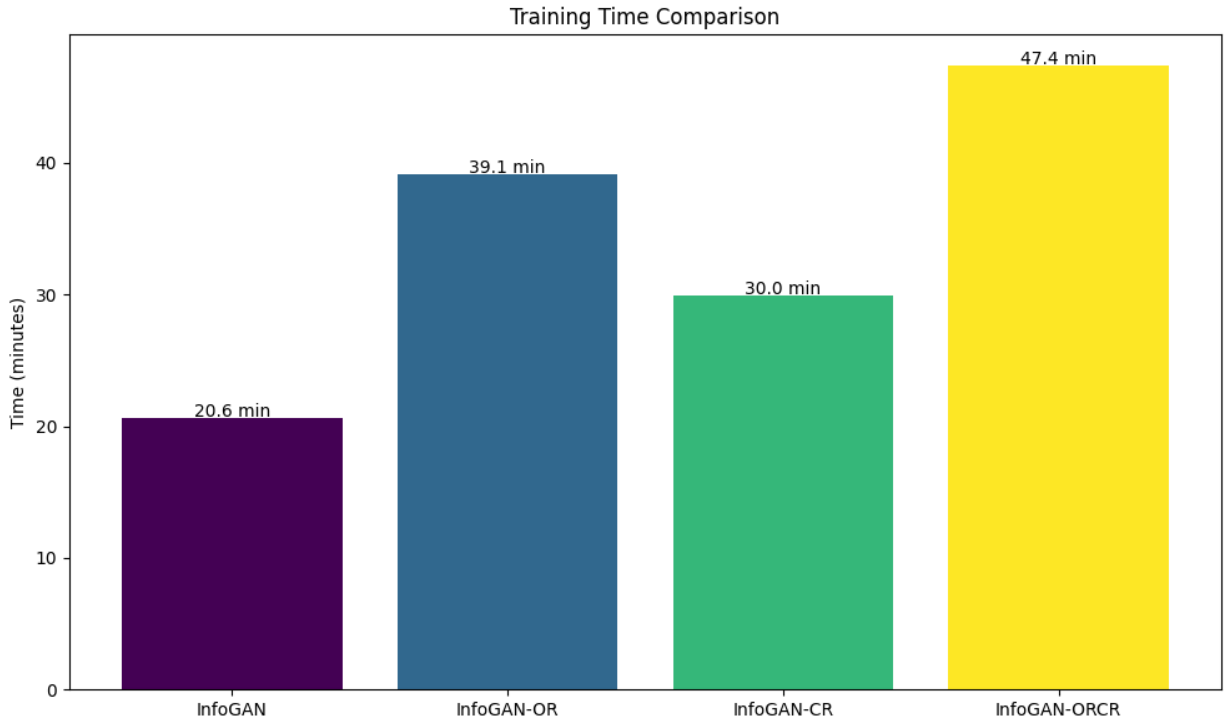
Categorical Mutual Information:



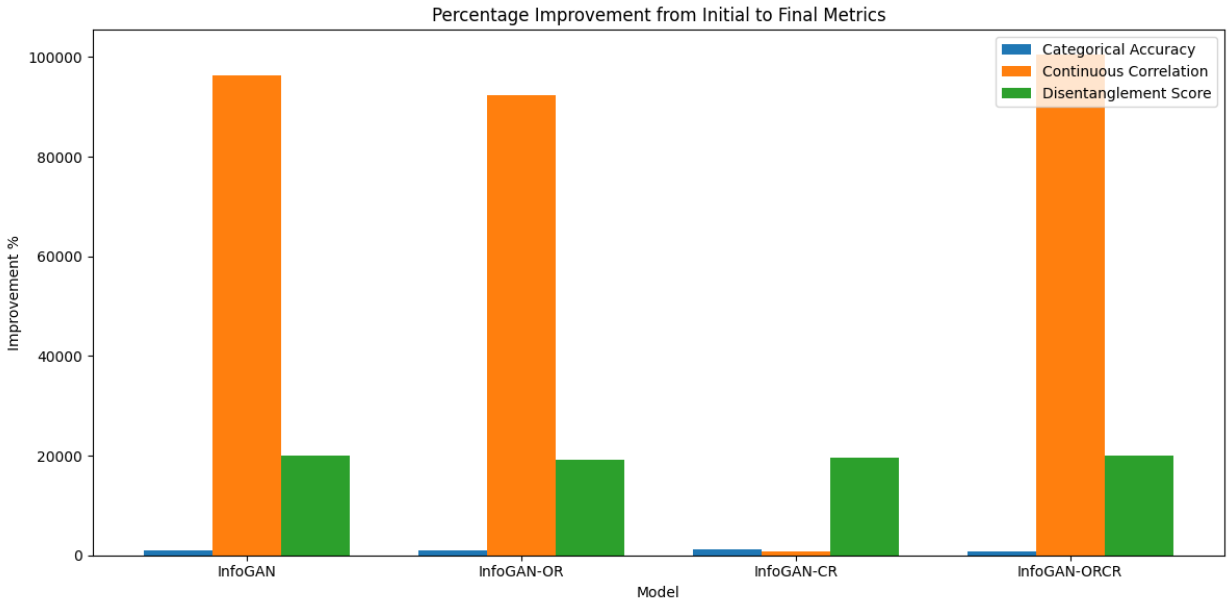
Continuous Mutual Information:



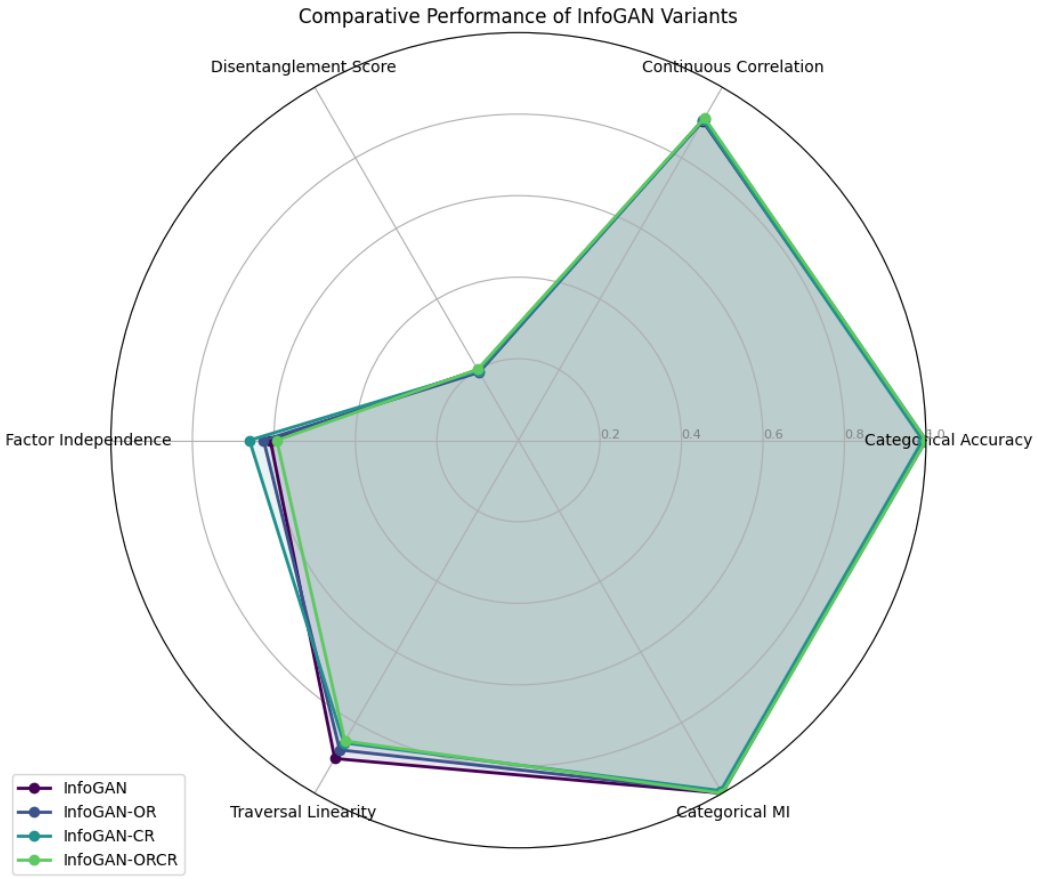
Training Time:



Model Improvement from Initial to Final Epoch:



Model vs Model Comparison:



The visualizations reveal that all models successfully learned to represent digit identity through the categorical code. However, the regularized models, particularly InfoGAN-ORCR, demonstrated clearer and more consistent manipulation of style factors such as rotation and width through the continuous codes.

For example, when varying the continuous code c2 from -2 to 2, InfoGAN-ORCR showed a smooth transition from left-leaning to right-leaning digits, while maintaining consistent digit shape and style. Similarly, when varying c3, the model smoothly adjusted the width of digits without affecting other attributes.

Our results demonstrate that both Orthogonal Regularization and Contrastive Regularization can enhance the quality of learned representations in InfoGAN, with their combination (InfoGAN-ORCR) yielding the best overall performance.

InfoGAN-OR improves categorical accuracy and continuous correlation compared to the baseline, suggesting that enforcing orthogonality in weight matrices helps separate different latent factors. This is particularly valuable for ensuring that discrete factors (like digit identity) do not interfere with continuous stylistic factors.

InfoGAN-CR excels in factor independence and continuous mutual information, indicating that contrastive learning helps the model maintain consistency in specific aspects of the representation when certain latent codes are changed. This is especially beneficial for continuous factors, as evidenced by the high continuous MI score.

InfoGAN-ORCR combines the strengths of both regularization techniques, achieving the highest disentanglement score and continuous correlation, along with perfect categorical accuracy. However, this comes at the cost of increased training time (approximately 2.3 times longer than the baseline InfoGAN).

The trade-off between performance and computational cost is an important consideration for practical applications. For tasks where categorical accuracy is paramount, InfoGAN-OR might offer the best balance, while for applications requiring precise control over continuous factors, InfoGAN-CR could be more appropriate. When both aspects are important and computational resources permit, InfoGAN-ORCR represents the superior choice.

Conclusion

This study demonstrates the effectiveness of Orthogonal Regularization and Contrastive Regularization in enhancing the disentanglement capabilities of InfoGAN. The combined approach (InfoGAN-ORCR) achieves the best performance across most metrics, suggesting that these regularization techniques address complementary aspects of representation learning.

Our findings have important implications for unsupervised representation learning and generative modeling. By improving disentanglement, these enhanced InfoGAN models can potentially enable more precise control over generated content and extract more meaningful features for downstream tasks.

Future directions for this research include:

1. Applying these regularization techniques to other GAN architectures and datasets beyond MNIST.
2. Integrating adaptive weighting of regularization terms during training.
3. Extending the approach to handle a larger number of latent factors.
4. Evaluating the utility of learned representations for downstream tasks such as classification or anomaly detection.
5. Exploring the relationship between disentanglement quality and sample quality in generated images.

References

1. Chen, X., Duan, Y., Houthooft, R., Schulman, J., Sutskever, I., & Abbeel, P. (2016). InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets. Arxiv.org. https://arxiv.org/abs/1606.03657
2. Li, X., Chen, L., Wang, L., Wu, P., & Tong, W. (2018). SCGAN: Disentangled Representation Learning by Adding Similarity Constraint on Generative Adversarial Nets. IEEE Access, 1–1. https://doi.org/10.1109/access.2018.2872695
3. Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2025). Generative Adversarial Nets. Advances in Neural Information Processing Systems, 27. https://papers.nips.cc/paper_files/paper/2014/hash/f033ed80deb0234979a61f95710dbe25-Abstract.html
4. Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. ArXiv.org. https://arxiv.org/abs/1511.06434
5. Larsen, Sønderby, Søren Kaae, Larochelle, H., & Winther, O. (2015). Autoencoding beyond pixels using a learned similarity metric. ArXiv.org. https://arxiv.org/abs/1512.09300
6. Kingma, D. P., & Welling, M. (2013, December 20). Auto-Encoding Variational Bayes. ArXiv.org. https://arxiv.org/abs/1312.6114
7. Li, Z., Usman, M., Tao, R., Xia, P., Wang, C., Chen, H., & Li, B. (2022). A Systematic Survey of Regularization and Normalization in GANs. 55(11), 1–37. https://doi.org/10.1145/3569928
8. Müller, J., Klein, R., & Weinmann, M. (2019). Orthogonal Wasserstein GANs. ArXiv.org. https://arxiv.org/abs/1911.13060
9. Sintunata, V., Liu, S., Nguyen Van, D., Lim, Z. Y., Zhikuan, R. L., Wang, Y., Feng, J. H. J., & Leman, K. (2024). Unsupervised Latent Regression through Information Maximization - Contrastive Regularized GAN. 2024 IEEE Conference on Artificial Intelligence (CAI), 1468–1473. https://doi.org/10.1109/cai59869.2024.00264