Improving Image Reconstruction using Incremental PCA-Embedded Convolutional Variational Auto-Encoder

Amir Azizi
CYENS CoE
Nicosia
CYPRUS
a.azizi@cyens.org.cy

Panayiotis Charalambous
CYENS CoE
Nicosia
CYPRUS
p.charalambous@cyens.org.cy

Yiorgos Chrysanthou
CYENS CoE
Nicosia
CYPRUS
y.chrysanthou@cyens.org.cy

ABSTRACT

Traditional image reconstruction methods often face challenges like noise, artifacts, and blurriness, requiring handcrafted algorithms for effective resolution. In contrast, deep learning techniques, notably Convolutional Neural Networks (CNNs) and Variational Autoencoders (VAEs), present more robust alternatives. This paper presents a novel and efficient approach for image reconstruction employing Convolutional Variational Autoencoders (CVAEs). We use Incremental Principal Component Analysis (IPCA) to enhance efficiency by discerning and capturing significant features within the latent space. This model is integrated into both the encoder and sampling stages of CVAEs, refining their capability to generate high-fidelity images. Our incremental strategy mitigates scalability issues associated with traditional PCA while preserving the model’s aptitude for identifying crucial image features. Experimental validation utilizing the MNIST dataset showcases noteworthy reductions in processing time and enhancements in image quality, underscoring the efficacy and potential applicability of our model for large-scale image generation tasks.

Keywords
Image Processing, Image Reconstruction, Principal Component Analysis, Convolutional Variational Auto-Encoders

1 INTRODUCTION

Generative models like Variational Autoencoder (VAE) are a significant advancement in deep learning, using probabilistic approaches to generate new data from existing datasets. VAEs, consisting of an encoder and decoder pair, capture key properties [CGD+20, BTLLW21]. VAEs have a wide range of applications, from image [LSM+21, YSYL16, WX21], video [YZAS21, WRO21, ZLS+21, DLW+22], text [LPL21, YDML21, SLW+21, ZDYC21] and music generation [JY23, WT22, WY21] to anomaly detection [NYW20, LCB+20], and they are easier to train than their competitors, such as Generative Adversarial Networks (GANs) [BTLLW21]. They ensure training stability by offering a greater range of realistic and varied facts [DB21]. CVAEs integrates the functionalities of VAEs with those of a CNNs, constituting a specialized form of deep generative model [YKK21]. CVAEs possess an architecture that leverages CNNs as both encoders and decoders, harnessing the spatial relation-capturing capabilities of CNNs and the fidelity in data generation characteristic of VAEs [SDRM21]. This method demonstrates exceptional proficiency in processing grid-based data, especially images, leveraging the spatial understanding capabilities inherent in CNNs [BLD22]. CVAEs stand as foundational components across a diverse range of applications [KSZ+21, WML21, LPC22, CQWZ21], spanning image synthesis, reconstruction, and anomaly detection. Their robustness consistently delivers stable, realistic, and semantically coherent outcomes [JJ21]. In Figure 1, the structure of CVAEs and the interaction between their components are illustrated. This paper is dedicated to enhancing both the efficiency and quality of image reconstruction through the utilization of CVAEs.

Our contributions are summarized as follows:

1. Introducing a novel method that combines IPCA with CVAEs, enhancing the sampling and encoding stages.

2. considerably reduces processing time, thus boosting the algorithm’s overall efficiency.

3. Demonstrating substantial enhancements in image quality by implementing our proposed algorithm on MNIST datasets.

The paper is structured as follows: Section 2 delves into the related works; in Section 3, we outline our novel approach and methodology. Section 4 is our Experimental results. Finally Section 5, discusses the research challenges, limitations, and potential future directions.
Figure 1: CVAEs utilize convolutional layers to encode input images, compress information into latent space, and decode samples, enabling efficient data representation and generation of new samples.

2 RELATED WORKS

The VAE, introduced by Kingma et al. [KW13], they applied variational Bayesian inference principles for image generation. It consists of two neural networks. Figure 2 shows an inference network for generating latent variable distributions, and a generation network for approximation. Deep learning and Artificial Intelligence have revolutionized image reconstruction and image generation from diverse data sources, with Variational Auto-encoders emerging as key methodologies for high-quality image generation [EEAMT22, LSC20, IB23, WCQ23].

Figure 2: The model employs variational Bayesian inference principles for image generation, utilizing two neural networks for inference and generation to capture data structure and variability.

Conditional Variational Autoencoder (CVAE), which is shown in Figure 3, improves the unsupervised model by incorporating category information labels, transforming it into a semi-supervised mode within the CVAE framework [SLY15, HNW21]. The Very Deep Variational Autoencoder (VDVAE) is an extension of the standard Variational Autoencoder (VAE) with increased depth in its neural network architecture [Chi20]. This depth allows for more intricate hierarchical representation of latent variables, enabling high-fidelity and nuanced reconstructions while navigating large-scale dataset complexities. VDVAE-SR [CHW+23] uses transfer learning on pre-trained VDVAEs to improve image super-resolution. Fusion-VAE [DVZN22] is a novel deep hierarchical variational autoencoder for generative image fusion, outperforming traditional methods. VAE, a neuro-symbolic generative model, combines VAEs and probabilistic Logic Programming strengths, enabling the generation of new data points satisfying logical constraints while capturing complex relationships VAE architecture illustrated in Figure 4 VDVAEs, while powerful for image generation, face computational complexity due to high-dimensional data, but ongoing advancements aim to improve efficiency and scalability [ASY+22, VK20].

Figure 3: CVAE, similar to VAE, adds category information to input data, maximizing logarithmic marginal likelihood and lower bound function of variation, but not solving image blur or high synthetic data accuracy.

Figure 4: The VAE model consists of three components: an encoder, a Prob-Log program, and a decoder, which compute latent variables, parameterize programs, and reconstruct images.

3 PROPOSED METHOD

The study uses IPCA and CVAEs to enhance image reconstruction. CVAEs extract hierarchical features from images using convolutional layers, preserving spatial information. They encode input images into a latent space, capturing meaningful representations of digits in a lower-dimensional space. Our novel approach lies in integrating IPCA within the training paradigm of CVAEs for image reconstruction. IPCA is used after encoding to reduce dimensionality in data obtained from the CVAE. It refines the latent space representation before decoding, enhancing efficiency and retaining meaningful information for the decoding step.

https://www.doi.org/10.24132/CSRN.3401.12
thereby optimizing the handling of latent representations. The workflow of our proposed method, as illustrated in Figure 5 is structured as follows:

1. CVAEs are used to improve image reconstruction by extracting hierarchical features from images using convolutional layers. These layers preserve spatial information, ensuring image structural integrity. By encoding input images into a latent space, CVAEs capture meaningful representations of digits, allowing the model to learn essential features while reducing data dimensionality.

2. IPCA is a technique used in the training paradigm of CVAEs to optimize image reconstruction by refining the latent space representation obtained from the encoding step, thereby enhancing efficiency and retaining crucial information for the decoding step.

3. The decoder uses deconvolutional layers, also known as transposed convolutional layers or upsampling layers, to transform the refined latent representation from IPCA into a high-dimensional feature map. This process optimizes the handling of latent representations, enhancing the efficiency and accuracy of image reconstruction tasks. The integrated approach ensures meaningful information is retained and utilized throughout the decoding process.

The integration of IPCA enhances the CVAE model’s efficiency by refining latent space representation before decoding, reducing data dimensionality while retaining crucial information for accurate image reconstruction. This approach enhances overall image reconstruction task performance and provides a novel solution for efficient dimensionality reduction and preservation of meaningful information. It is applicable across different domains and datasets, offering versatility and effectiveness in handling latent representations. Integrating IPCA into CVAEs training provides a robust and efficient solution for image reconstruction, contributing to image processing and deep learning techniques.

4 RESULTS

The study utilizes a PC with a Core i7-8700K CPU, 16GB RAM, and Google Colab for computational resources. The results of this work are presented in two parts:

i) Analyzing the impact of the IPCA on processing time.
ii) Exploring the relationship between reconstructed image quality and the proposed method.

We use two main datasets: MNIST [LC10] and Fashion-MNIST [XRV17]. MNIST has 70,000 grayscale images of ten handwritten digit classes. Fashion-MNIST is similar but covers ten categories of clothing, including items like t-shirts, trousers, and pullovers. To enhance compression, we utilized three models: CVAE, PCA-CVAE, and IPCA-CVAE. The architecture of the encoder and decoder involved a consistent structure. The encoder started with an input layer for gray-scale images, followed by two Conv2D layers implementing 32 and 64 filters. A flattening layer condensed the output, leading to a dense layer with 16 neurons activated by ReLU. Two dense layers generated parameters for shaping the variational distribution within the latent space. The decoder played a crucial role in the reconstruction process, starting with an input layer for 20-dimensional latent space representation. The output was rearranged into a (7, 7, 64) structure, enabling two Conv2DTranspose layers to progressively upscale the encoded latent space. The output layer, a Conv2DTranspose, reconstructed the original image dimensions, concluding the decoding phase.

Figure 6: Training Time Comparison for CVAE Models on MNIST digits dataset.

In Figure 6, a comparative analysis of various methodologies is presented, focusing on the training duration. The visual representation of PCA-CVAE and CVAE...
methodologies reveals significant differences. PCA-CVAE offers sustained functionality beyond the fourth epoch and reduces processing time, while IPCA-CVAE demonstrates accelerated training process completion, which is beneficial for large image databases. The proposed IPCA-CVAE method shows heightened efficiency, especially in scenarios with voluminous image collections, demonstrating its superior performance.

Figure 8: Reconstructed Images based on IPCA-CVAE on MNIST digits dataset.

Figure 9: Reconstructed Images based on IPCA-CVAE on MNIST fashion dataset.

Figure 7 provides a comparison of Training and Validation Loss across various CVAE methods. Figure 8 and Figure 9 display reconstructed images generated using the IPCA-CVAE method applied to the MNIST datasets. Furthermore, Figure 10 highlights a noticeable discrepancy in training duration on the MNIST dataset, emphasizing the tangible advantages of employing the proposed algorithm. Significantly, the discernible reduction in processing time attests to the algorithm’s efficacy in optimizing training efficiency. Additionally, Figure 11 presents a comparative analysis between CVAE and IPCA-CVAE methods. The left image corresponds to the CVAE method, the middle image represents PCA-CVAE, and the right image pertains to IPCA-CVAE. Notably, the Figure distinctly reveals a more pronounced reduction in loss during the training process in the proposed method compared to alternative approaches.

4.1 Metrics Evaluation
In this section, we quantitatively assess the performance of our proposed method, CVAE-IPCA, on the
We employ three widely used image quality metrics: Peak Signal-to-Noise Ratio (PSNR), Mean Squared Error (MSE), and Structural Similarity Index (SSIM). PSNR measures the ratio between the maximum possible power of a signal and the power of corrupting noise, defined as:

\[
PSNR = 10 \times \log_{10}(\text{peakval}^2)/MSE
\]

where peakval is the maximum possible pixel value of the image. MSE calculates the average squared difference between the original and reconstructed images, which is defined for two images such as \( \hat{g}(n,m) \) and \( g(n,m) \) as below:

\[
MSE = 1/MN \times (\sum_{n=0}^{M} \sum_{m=1}^{N} |\hat{g}(n,m) - g(n,m)|^2)
\]

Lastly, SSIM evaluates the similarity between two images based on luminance, contrast, and structure, and is defined as

\[
SSIM(x,y) = \left[l(x,y)\right]^\alpha \times \left[c(x,y)\right]^\beta \times \left[s(x,y)\right]^\gamma
\]

Here, \( L \) denotes luminance, \( C \) represents contrast, and \( S \) signifies structure. These parameters help gauge brightness, intensity range differences, and local pattern similarities between images, respectively. \( \alpha, \beta, \) and \( \gamma \) are constants used for computation stability. Results of metrics evaluation illustrated in Figure 12.

### 4.2 Ablation Study

In our ablation study, we systematically varied three key parameters: learning rate, batch size, and latent space dimensionality, to understand their impact on the performance of our proposed method.

1. **Learning Rate Variation**: We tested learning rates of 0.001, 0.0001, and 0.01 to analyze their effect on model convergence and performance metrics such as PSNR, SSIM, and MSE. This experiment provided insights into how different learning rates influence training dynamics and model effectiveness.

2. **Batch Size Variation**: We explored batch sizes of 64, 128, and 256 to assess their impact on training stability and computational efficiency. By observing training speed and model accuracy under different batch size settings, we gained an understanding of their trade-offs and implications.

3. **Latent Space Dimensionality Variation**: We varied latent space dimensions between 20, 50, and 100 to examine how they affect the model’s ability to capture and represent input data features. Analyzing reconstruction quality across different latent space sizes provided insights into the dimensionality’s role in feature representation.

The default parameter values for our method are learning rate = 0.001, batch size = 128, and latent space dimension = 20. These values serve as reference points for comparison against the variations tested in our ablation study.

#### 4.2.1 Findings

1. Our investigation found that employing a learning rate of 0.0001 failed to yield any notable improvements. Although a rate of 0.01 showed some enhancements compared to 0.0001, the most consistent and optimal outcomes were achieved with a learning rate of 0.001.

2. Changing the batch size did not lead to any improvement in the results. However, the alterations in batch size yielded better outcomes compared to those from modifying the learning rate. Notably, the optimal results were observed with a batch size of 128. As anticipated, increasing the batch size to 64 resulted in longer processing times, while decreasing it to 256 reduced processing times.

3. Increasing the latent space dimensions to 100 yielded improved results. However, this enhancement came at the cost of significantly increased processing time. The results of the ablation study are shown in Table 1.

Figure 13 also presents the outcomes of PSNR. Figure 14 shows the reconstructed images with different parameters and finally, Figure 15 illustrates the results of SSIM, MSE, and training curves.

### 5 CONCLUSION AND FUTURE WORKS

This paper introduces a novel approach to image reconstruction using CVAEs. The efficiency of the model is optimized through IPCA, a technique used to identify and capture significant features in the latent space. This
Table 1: Summary of the Ablation Study (lr=Learning Rate , ld=Latent Dimension , bs=Batch Size , TT=Training Time(s))

<table>
<thead>
<tr>
<th>Parameter</th>
<th>PSNR</th>
<th>SSIM</th>
<th>MSE</th>
<th>TT</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>lr=0.01</td>
<td>20.439</td>
<td>0.8907</td>
<td>0.0167</td>
<td>1031</td>
<td>ld=20, bs=128</td>
</tr>
<tr>
<td>lr=0.001</td>
<td>21.0432</td>
<td>0.9097</td>
<td>0.0133</td>
<td>974</td>
<td>ld=20, bs=128</td>
</tr>
<tr>
<td>lr=0.001</td>
<td>19.3043</td>
<td>0.8898</td>
<td>0.0193</td>
<td>1021</td>
<td>ld=20, bs=128</td>
</tr>
<tr>
<td>ld=20</td>
<td>21.0432</td>
<td>0.9097</td>
<td>0.01505</td>
<td>974</td>
<td>lr=0.001, bs=128</td>
</tr>
<tr>
<td>ld=50</td>
<td>20.8135</td>
<td>0.8992</td>
<td>0.0193</td>
<td>1255</td>
<td>lr=0.001, bs=128</td>
</tr>
<tr>
<td>ld=100</td>
<td>22.002</td>
<td>0.9101</td>
<td>0.0123</td>
<td>1245</td>
<td>lr=0.001, bs=128</td>
</tr>
<tr>
<td>bs=64</td>
<td>20.754</td>
<td>0.8998</td>
<td>0.0163</td>
<td>1182</td>
<td>lr=0.001, ld=20</td>
</tr>
<tr>
<td>bs=128</td>
<td>21.0432</td>
<td>0.9097</td>
<td>0.0133</td>
<td>1021</td>
<td>lr=0.001, ld=20</td>
</tr>
<tr>
<td>bs=256</td>
<td>20.9097</td>
<td>0.9001</td>
<td>0.0140</td>
<td>1084</td>
<td>lr=0.001, ld=20</td>
</tr>
</tbody>
</table>

Figure 13: PSNR results (a) learning rate=0.0001, (b) learning rate=0.01 (c) latent dimension=50 (d) latent dimension=100 (e) batch size=64 (f) batch size=256

Figure 14: Reconstructed Images (a) learning rate=0.0001, (b) learning rate=0.01 (c) latent dimension=50 (d) latent dimension=100 (e) batch size=64 (f) batch size=256

incremental strategy addresses scalability concerns associated with traditional PCA while preserving the model’s proficiency in identifying essential image features. Experimental results on the MNIST dataset show:

1. Reduces processing time and improves image quality.
2. Applicable in large-scale image generation tasks.

3. CVAEs’ sensitivity to input data distribution variations.

Our proposed method demonstrates significant potential across various image-related tasks, extending beyond the datasets examined thus far. In object detection, denoising, and other applications, such as image classification using extensive datasets like CIFAR-10 and ImageNet, integrating IPCA-CVAE could enhance feature extraction and dimensionality reduction. This enhancement may lead to improved classification accuracy and resilience against variations in image content. Moreover, in tasks like style transfer and image synthesis, generating realistic and diverse images is crucial. IPCA-CVAE’s refined latent representations could offer finer control over visual attributes, enabling the creation of more compelling images. Additionally, in medical imaging tasks like tumor detection and segmentation, where handling high-dimensional images is common, IPCA-CVAE integration could help capture meaningful anatomical structures while reducing computational complexity and memory usage. Furthermore, in satellite imagery analysis for environmental monitoring and disaster management, IPCA-CVAE could facilitate efficient feature extraction and anomaly detection, enabling timely identification of critical changes in land cover and environmental conditions.

Moving forward, future research directions should focus on adapting the model to diverse and more complex datasets, exploring real-time optimization strategies, and improving the robustness of the CVAE framework. Additionally, incorporating domain-specific knowledge or constraints into the model may further refine its applicability for specific image reconstruction tasks.

6 ACKNOWLEDGEMENTS

This project has received funding from the European Union’s Horizon 2020 Research and Innovation Programme under Grant Agreement No 739578 and the Government of the Republic of Cyprus through the Deputy Ministry of Research, Innovation and Digital

https://www.doi.org/10.24132/CSRN.3401.12

112
Policy. This work has received funding from the European Union’s research and innovation program Horizon Europe under grant agreement No 10192889. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

7 REFERENCES


Yann LeCun and Corinna Cortes. MNIST handwritten digit database. 2010.


