Learning Representations for High-Fidelity Image Compression

Leila Abdelrahman

***Abstract*—In this investigation, we explore how autoencoders (AE)s and Variational Autoencoders (VAE)s serve a powerful role in image compression. By investigating the reconstructed images' mean squared error, and the space needed to encode an image, we demonstrate how the AE is a strong alternative to traditional JPEG; VAEs need more fine-tuning and enhancement to achieve our AE's performance. We further present qualitative analysis, and difference maps to discuss where our approaches excelled and struggled, as compared to the traditional JPEG compression algorithm. Overall, our methods show that efficient compression needs further investigation, and has great potential to significantly minimize large amounts of disk storage space at scale.**

***Index Terms*—OCR, Digital Image Processing, Feature Selection.**

I. Introduction

IMAGES add richness and dimension to human life. In Medicine, radiologists often refer to patients' mammogram or chest x-ray. As the world's population increases and ages, radiologists are storing and analyzing petabytes of medical images at a time. Image quality and precision are critical for radiologists to make accurate diagnostic and prognostic decisions. Yet, these factors generally point to costly data storage techniques. Traditional image compression techniques, like JPEG [1], JPEG2000 [2], and WebP suffice in domains with low-fidelity standards (e.g. websites, image thumbnails, etc.). However, for healthcare domains, high-fidelity imaging is necessary.

Techniques like learned image representations and generative networks [3] are alternative compression algorithms that can accurately reconstruct images from tiny, encoded representations.

II. Related Works

A. Traditional Compression Techniques

Traditional image compression techniques include compression algorithms like JPEG [1] and JPEG2000 [2]. These techniques work in two stages: (1) The lossy stage removes information undetectable to the human eye. The lossless stage uses fewer bits to code the salient symbols in the remaining data. The Royal College of Radiologists [4] suggests different lossy compression ratio recommendations for JPEG compression on modalities ranging from mammography (20:1) to radiotherapy CT scans (no compression recommended). While JPEG compression is the classic standard for coding images, recent studies into neural networks have shown how autoencoders can outperform these classical techniques.

L. Abdelrahman was with the Department of Electrical and Computer Engineering at the University of Miami

B. Autoencoders for Learned Representations

Autoencoders (AE)s [5] are unsupervised neural networks that minimize the reconstruction loss of an image by learning a latent representation. Cheng et al. [6] recommend a deep convolutional autoencoder-based lossy image compression algorithm. The authors use a rate-distortion loss function to minimize the reconstruction loss and the number of bits used to encode the compressed data. After training the model, the authors show how their methods are comparable to JPEG2000 compression. Choi et al. [7] recommend a variable rate autoencoder that users can use to compress images based on user-defined bit-rate preferences. Considering image context, Toderici et al. [8] motivate a variable rate convolutional autoencoder that uses a recurrent neural network for spatial context. The authors also allow for user-defined bitrates and show how their compression methods outperform JPEG, JPEG2000, and WebP benchmarks. Mentzer et al.[9] show how adding a context module to the autoencoder's latent representation learns the conditional probability of these high-level representations. The context model is updated to learn a better latent representation, allowing for optimal bit allocation in the compressed latent coding. Learning optimal latent space configurations also underlies Variational Autoencoders, which we discuss in the next section.

C. Variational Autoencoders

Variational Autoencoders (VAE)s [10] build on traditional autoencoders but optimize compressed latent codings by sampling points from the encoded input's latent space distribution. As the model learns, it optimizes the distribution, minimizing the KL Divergence, which measures the entropy between two probability distributions.

Zhou et al. [11] use VAEs with a combined mean square error and perceptional loss to compress images using models with different compression bit rates. Dumas et al. suggest one model that automatically learns the quantization rate, eliminating the need for multiple models.

Context is critical when detecting and encoding features in an image. Thus, Wen et al. [12] propose a multi-scale pyramidal encoder for generating the VAE's latent encoding. Although the multi-resolution method improves encoding performance, the authors train eight models to compress each image, with each model having a different bit rate. Despite this shortcoming,

D. Encodings for Compression in Medical Imaging

Researchers are now using Deep Learning-based compression methods like Autoencoders and VAEs to show their

viability in generating high-fidelity medical image reconstructions. Steudel et al. [13] were the first to recommend neural networks for medical image compression in 1995. Tan et al. [14] demonstrate autoencoder viability for compressing DDSM [15] mammograms. Tellez et al. [16] use VAEs to compress histopathology images prior to classification, showing how light-weight representations suffice to extrapolate critical diagnostic interpretation. Shen et al. [17] recommend a stacked autoencoder for malaria blood cell images before infection diagnosis, illustrating how compressed image representations have clinical utility. Finally, Sushmit et al. [18] show how context through an autoencoder's recurrent layers is critical for accurately compressing X-ray images.

Motivated by the current literature in the field, we aim to compare traditional JPEG, JPEG2000, Convolutional AEs, and Convolutional VAEs based on their compression efficiency and image reconstruction fidelity.

III. Approach

A. Model Architectures

We begin by constructing our AE and VAE neural networks. Both networks contain bottleneck encoding layers that we use for compression. The main difference between the AE and VAE is that the VAE learns to also minimize the bit-size of the encoding by sampling from a distribution, as the bottleneck in the lower part of Figure 1 B) shows. Both networks have an identical decoder sub-network.

In both networks, we rely on internal batch normalization [19] to improve learning and numerical stability throughout the network. Batch normalization recenters and scales the values inside a layer to help reduce the number of training epochs required to achieve good results.

We use the LeakyReLU activation function [20] instead of the traditional ReLU function, as LeakyReLU can handle negative inputs values to the activation far better than ReLU. The encoder has two convolutional layers that rescale the image from 1024×1024 to $256 \times 256 \times 64$ tensors. In the AE, the critical compression point comes during the flattening and dense layer steps, which learn a low-dimension representation of the data. In the VAE, there are also flattening and dense layers. However, the compression step also includes the mean, variance, and sampling layers, which learn an even smaller representation by optimizing the encoding's distribution. Both the AE and VAE eventually learn a compression encoding vector that is 16×1 in dimension.

For the decoding sub-network, the network takes the learned dense encoding as input and then uses deconvolution layers to

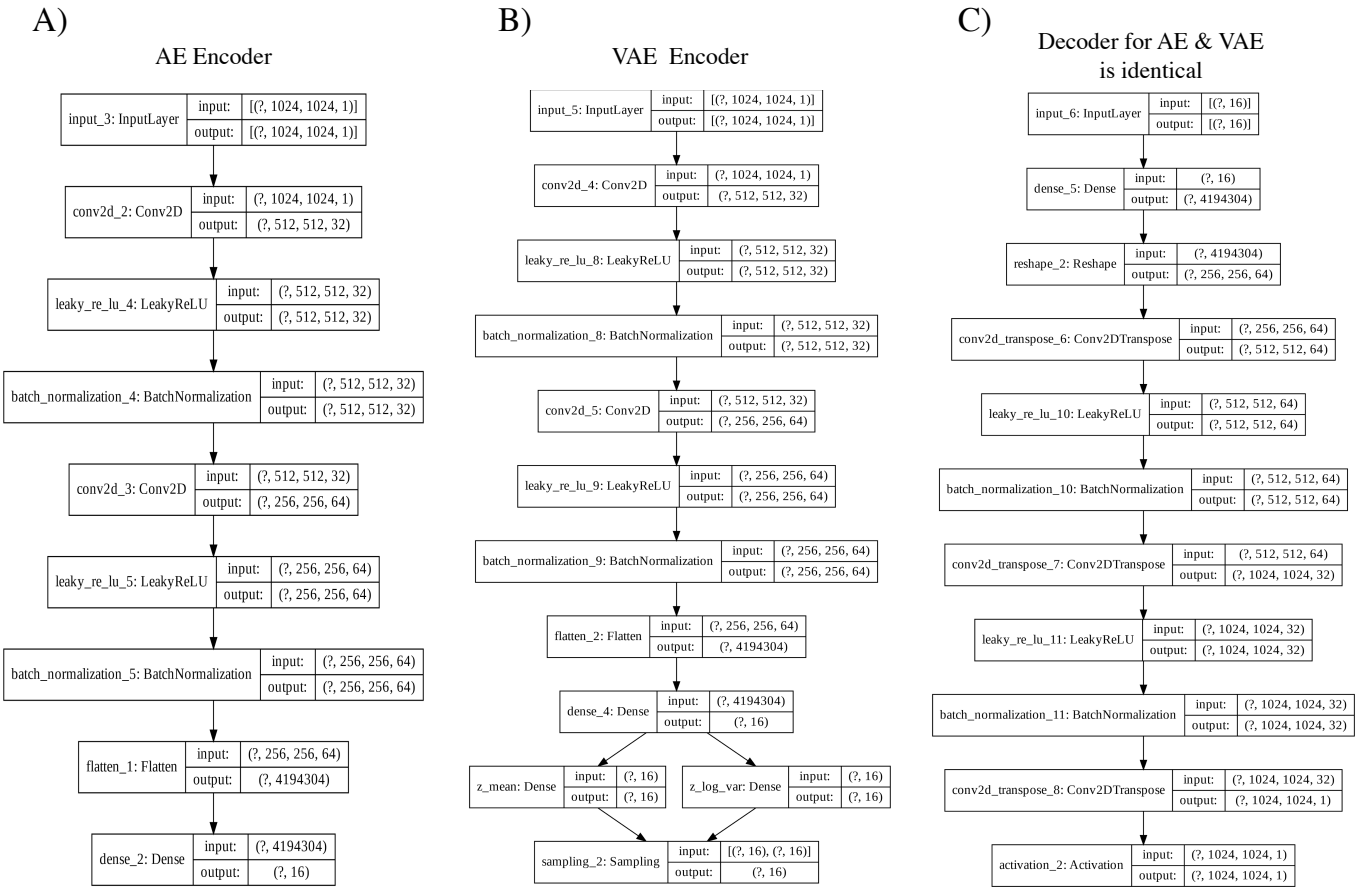

Fig. 1. A) The encoder network of the entire AE neural net. The network learns a dense, bottleneck representation of the image through convolutions and minimizing the reconstruction loss. B) The VAE encoder samples from a distribution and learns the mean and covariance to minimize differences between the learned and the image's ground truth distribution. C) Both the AE and VAE have an identical decoder that expands the original image through a series of Conv2DTranspose deconvolution layers.

expand the representation back to the original image resolution. To optimize the network's weights, we use backpropagation to minimize model loss functions, Equation 1 and Equation 2, for the AE and VAE, respectively.

$$L_{AE} = \frac{\sum_{i=1}^{N}(x_i - \hat{x_i})^2}{N} \quad (1)$$

Equation 1 shows the mean squared error, which is related to the unsigned difference between the predicted ($\hat{x_i}$) and the ground truth (x_i) images, summed over all images for $i \in N$.

$$L_{VAE} = BCL + -\frac{1}{2} \times (1 + \log(\mathbf{\Sigma}) - \mu^2 - e^{\Sigma}) \quad (2)$$

The first term, BCL, represents the binary cross-entropy loss, which is another way of measuring the reconstruction error. The remaining terms in the equation are related to the KL Divergence loss, which measures how well the model's sampling distribution approximates the actual image distribution. During training and backpropagation, the model optimizes its μ and Σ parameters to minimize this loss. Thus, the loss function for the VAE, given in Eqution 2 is the sum of the reconstruction loss and the KL divergence loss.

B. The Data

To test our model architectures, we retrieve and preprocess x-ray mammograms from the CBIS-DDSM public dataset [15]. The images are initially in DICOM format and contain pictures of the right and left breast, in the craniocaudal (CC) and mediolateral oblique (MLO) anatomical views. The dataset also labels the mammograms as containing masses or calcifications. For this investigation, we select images in the MLO view, which include masses. To ensure our models can generalize, we choose to keep both the left and right breast images in our filtered dataset. We also convert the DICOM images to .png and resize the images using bilinear interpolation to 1024×1024 pixels. Thus, the final dataset we use for this study contains $n = 637$ images. We split the data into a training and test set, using 80% of the data for training and 20% for validation.

C. Implementation

When training the networks, we use an NVIDIA $V - 100$ GPU. We train both models for 1000 epochs, using checkpoints to save model weights associated with the lowest training loss functions.

When working with the images, we store both the encodings and the reconstructed, decoded image for analysis. We work with compressing the pictures using the traditional JPEG algorithm at 85% quality for our baseline comparison. We build and train our models using TensorFlow and generate figures and visual data using Python 3.6.

IV. Results

A. Quantitative Results

1) Mean Squared Error: The MSE is often used to measure the image reconstruction quality, as it uses a pixel-wise difference calculation to determine how similar two images are. We plot the results for this metric in Figure 2.

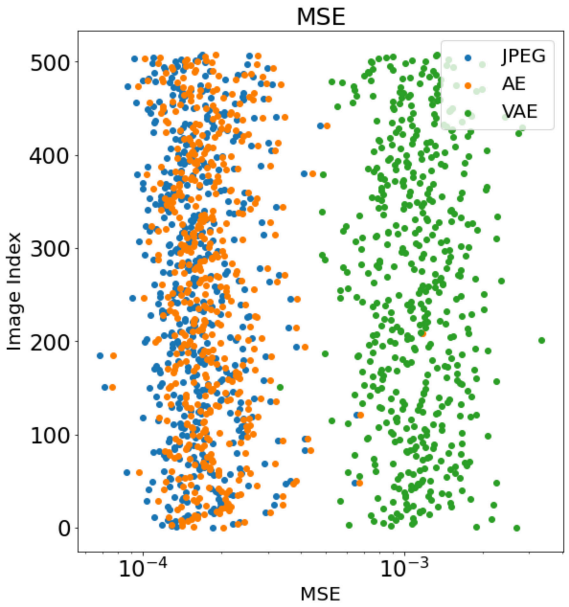

Fig. 2. The MSE measured for all the training images, where each scatterpoint represents the MSE for one image, whose index in the dataset is on the y axis. While the VAE has significantly high error rates, the AE achieves a low MSE, as compared with the JPEG baseline.

As Figure 2 shows, images that are difficult for the AE to reconstruct are also difficult for JPEG reconstruction too. Likewise, there is clear correlation between points with lower MSE values. This emphasizes how AE performs comparably to the JPEG compression algorithm.

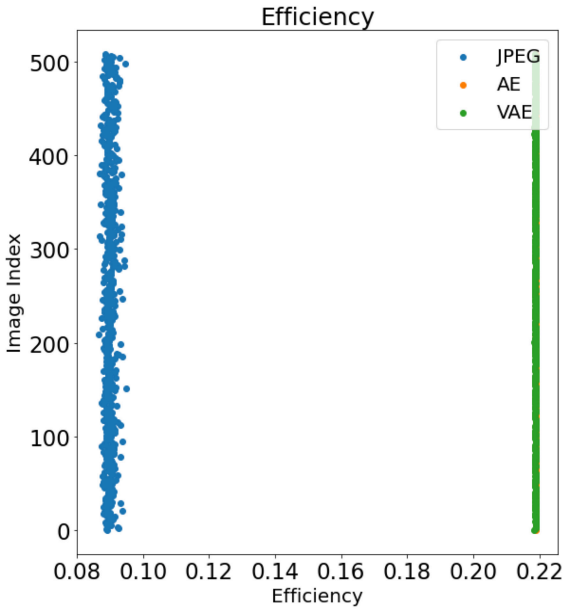

Fig. 3. The compression efficiency results for the three different approaches. As compared with the baseline JPEG algorithm, the AE and VAE, achieve significantly higher efficiency, as their encoding bit-size is generally on the order of e1, wheres for JPEG, is its on the order of e3.

2) Compression Efficiency: In assessing our compression algorithm performance, we use an efficiency performance metric, given by Equation 3. It represents a ratio of the mean squared error to the learned encoding bit-size for each image.

$$Eff_i = \frac{1 - (x_i - \hat{x_i})^2}{\log(S_i)} \tag{3}$$

Here, the first term is essentially the reconstruction accuracy, which we obtain by subtracting the mean-squared error from 1. Then, we divide by the log of the number of bits required to encode the image, given by S_i.

To measure the image reconstruction performance, we also plot the mean square error alone. In making our comparison, we record the results for the AE, VAE, and JPEG. Figure 3 highlights the outcome when using the compression efficiency as our performance metric. By far, the AE and VAE can reconstruct images with high efficiency, measured as the ratio of reconstruction accuracy to encoding size. The main contributor to their efficiency is the coding size. JPEG encoding generates codings that are on the order of 10^3 bits in size, while AE and VAE encoding generate encodings that are less than 100 bits in size. When working with massive datasets, this two order-of-magnitude difference scales, and can lead to terabytes of saved disk space.

B. Qualitative Results

1) Displaying Reconstructions: We selected sample five sample images representing the three compression approaches and showcase their image reconstructions compared to the original resolution image. Qualitatively, the VAE model struggles to capture vital details present in the original mimeograph. As Figure 4 shows, the model misses striations in the breast and even entire mass regions. These regions are critical in a mammograph, as they signal tumors and potentially metastatic cancers.

Because the VAE learns to represent the image from an optimized sampling distribution, it indirectly optimizes its encodings for the reconstruction. Moreover, the VAE has an additional loss term (the KL divergence), which it uses to minimize its encoding bit-size. Because of this, the VAE bottleneck may have too many constraints, as it attempts to generate a representation that is too simple to recreate the original image accurately.

On the other hand, the autoencoder does a good job overall at capturing and reconstructing the image. It can recreate the mass regions accurately and still captures the striations present in the original image. Of all three techniques compared (JPEG, AE, and VAE), the AE seems to be the optimal reconstruction

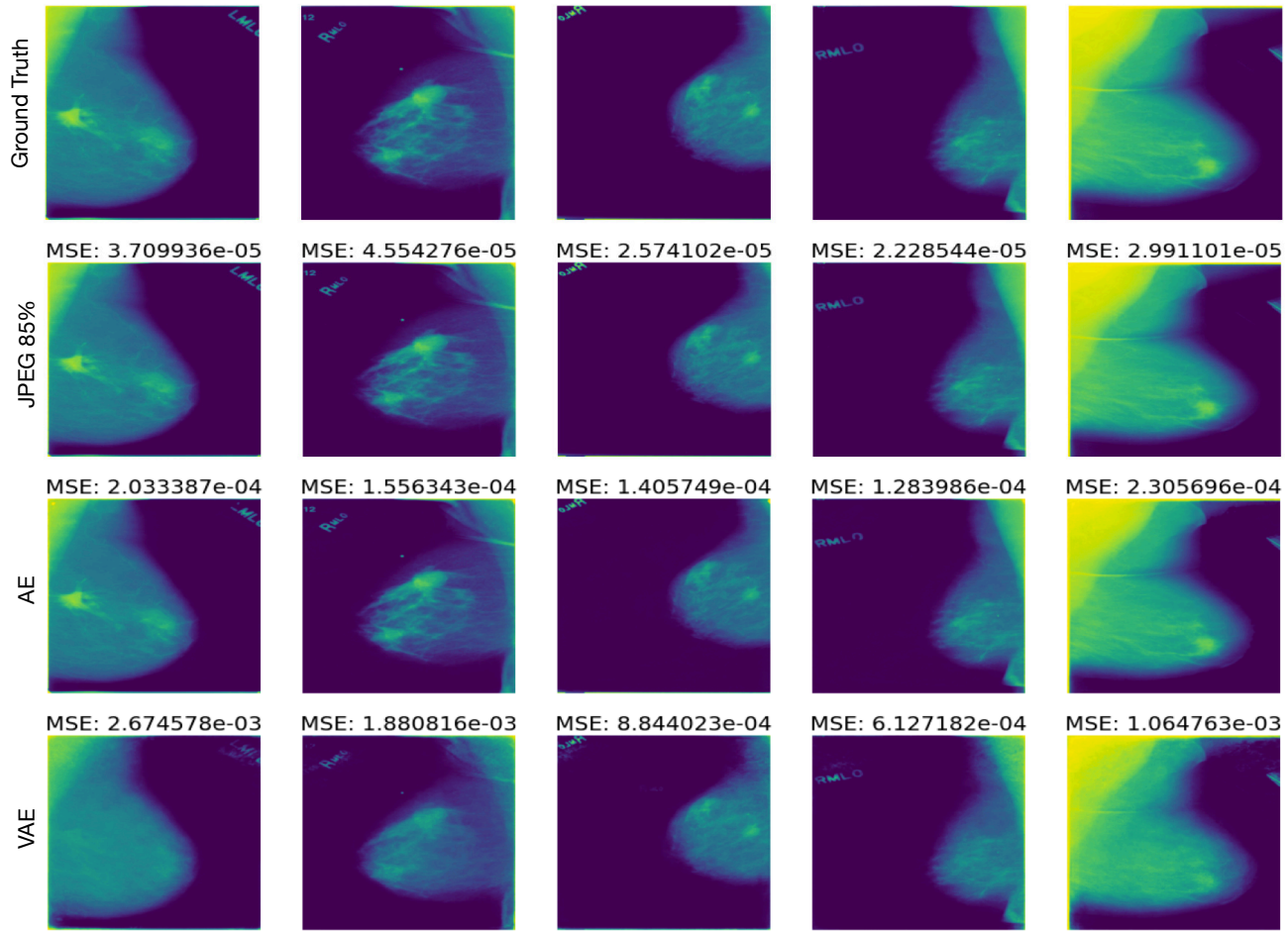

Fig. 4. Select image reconstructions based on the different methods. While the VAE model struggles to restore finer details of striation in the breast, the AE and JPEG methods can accurately reconstruct these.

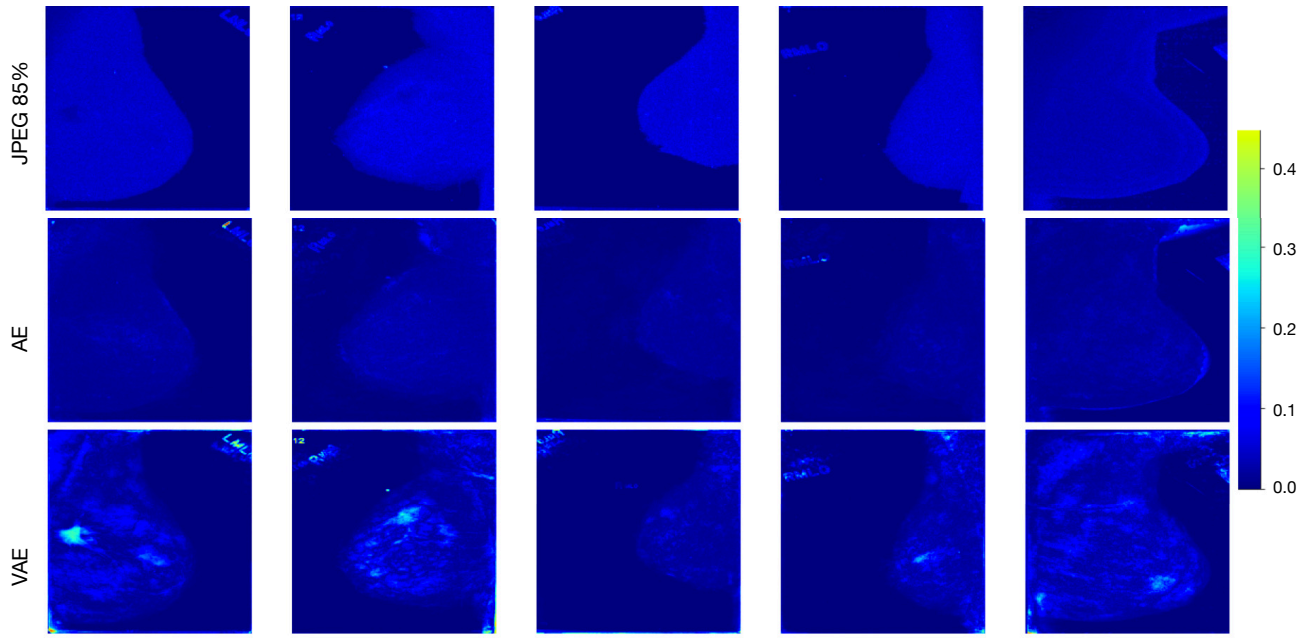

Fig. 5. Absolute difference maps between the image reconstructions and the original image. Brighter regions correlate with higher differences, while darker regions correlate with well-preserved reconstructions.

choice. The vector it uses to represent the original image is just 16×1 in dimension, requiring roughly 80 bits to encode. Its high fidelity reconstructions and small encoding size mean that it can create light-weight encodings that preserve crucial details in the original image. Although its MSE is on the order of 10^{-4}, its small encoding size means that AEs merit further investigation and implementation for efficient image compression.

2) Analyzing Reconstruction Error: In interpreting the qualitative results further, we also investigate absolute value difference maps to identify regions where the algorithms accurately reconstruct the images and areas where they fail. As Figure 5 shows, brighter colors in the difference map correlate with higher absolute value difference. The compression algorithm often preserves details about masses (these are usually darker in the difference maps) for the JPEG baseline. Simultaneously, there is a slight overall error in the background (perhaps due to the incorrect color intensity). On the other hand, the AE algorithm sometimes fails to capture striation, but the largest error sources are at the breast contours. This is most apparent in Figure 5's right-most column. Other sources of error include the writing at the images' corners. Because these details are small and fine-grained, the letters are often the first sources of error during image reconstruction. The letters as a source of error persist in all three compression approaches yet is most prevalent in the VAE.

By far, the weakest compression algorithm for this task is the VAE. Specifically, as highlighted by the intense regions inside the breast tissue, the algorithm misses critical striation patterns in the tissue, and more importantly, entire masses. The model needs more work and refinement. GAN algorithms, predicated upon VAEs, show promise to produce better encodings, faithful to the original image.

V. Discussion

This study examines alternative image compression algorithms to classic benchmark methods, like the JPEG standard. We show that neural networks like Autoencoders can perform well compared to JPEG compression and significantly outperform JPEG in compression efficiency. On the other hand, VAEs, variational methods based on sampling from distributions, need more fine-tuning and investigation to reach higher performance levels.

This investigation highlights how the VAE misses necessary characteristics in the mammogram, such as apparent masses and striation indicative of breast density and cancers. On the other hand, the AE has high fidelity when reconstructing these features.

We also emphasize that all three algorithms struggled to capture the written text in the image text accurately, as Figure 5 highlights. While images like mammograms may not always require preserving written text, situations that require keeping the text in images (like street signs, photos of license plates, and other applications) need robust image compression algorithms. The AE did a much better job reconstructing the breast tissue than the JPEG compression; the AE's primary source of error came when it reconstructed the letters. This may be because not every image in the training set contained letters.

Despite the shortcomings of the neural network approach, it shows promise: its ability to generate reasonable image reconstructions, with a more than two-order-of-magnitude encoding size, highlights its potential for efficient compression—investigations into sophisticated approaches founded on the autoencoder merit further work.

VI. Conclusion

Throughout this investigation, we explore autoencoder-based methods for image compression as alternatives to the classical JPEG algorithm. We show that while JPEG does slightly outperform our strategies in terms of reconstruction accuracy, autoencoders' strength lies in their efficient encodings. Small bit-size representations suffice to reconstruct an image reasonably. Depending on the context, autoencoders may help optimize data storage while preserving essential features in a snap.

Acknowledgment

The author would like to thank Dr. Negahdaripour Shahriar for providing the dataset and inspiring the representation learning for this problem.

References

[1] G. K. Wallace, "The jpeg still picture compression standard," *IEEE transactions on consumer electronics*, vol. 38, no. 1, pp. xviii–xxxiv, 1992.

[2] M. Rabbani, "Jpeg2000: Image compression fundamentals, standards and practice," *Journal of Electronic Imaging*, vol. 11, no. 2, p. 286, 2002.

[3] E. Agustsson, M. Tschannen, F. Mentzer, R. Timofte, and L. Van Gool, "Extreme learned image compression with gans." in *CVPR Workshops*, vol. 1, 2018, p. 2.

[4] F. Liu, M. Hernandez-Cabronero, V. Sanchez, M. W. Marcellin, and A. Bilgin, "The current role of image compression standards in medical imaging," *Information*, vol. 8, no. 4, p. 131, 2017.

[5] M. A. Kramer, "Nonlinear principal component analysis using autoassociative neural networks," *AIChE journal*, vol. 37, no. 2, pp. 233–243, 1991.

[6] Z. Cheng, H. Sun, M. Takeuchi, and J. Katto, "Deep convolutional autoencoder-based lossy image compression," in *2018 Picture Coding Symposium (PCS)*. IEEE, 2018, pp. 253–257.

[7] Y. Choi, M. El-Khamy, and J. Lee, "Variable rate deep image compression with a conditional autoencoder," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 3146–3154.

[8] G. Toderici, S. M. O'Malley, S. J. Hwang, D. Vincent, D. Minnen, S. Baluja, M. Covell, and R. Sukthankar, "Variable rate image compression with recurrent neural networks," *arXiv preprint arXiv:1511.06085*, 2015.

[9] F. Mentzer, E. Agustsson, M. Tschannen, R. Timofte, and L. Van Gool, "Conditional probability models for deep image compression," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4394–4402.

[10] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[11] L. Zhou, C. Cai, Y. Gao, S. Su, and J. Wu, "Variational autoencoder for low bit-rate image compression." in *CVPR Workshops*, 2018, pp. 2617–2620.

[12] S. Wen, J. Zhou, A. Nakagawa, K. Kazui, and Z. Tan, "Variational autoencoder based image compression with pyramidal features and context entropy model." in *CVPR Workshops*, 2019, p. 0.

[13] A. Steudel, S. Ortmann, and M. Glesner, "Medical image compression with neural nets," in *Proceedings of 3rd International Symposium on Uncertainty Modeling and Analysis and Annual Conference of the North American Fuzzy Information Processing Society*. IEEE, 1995, pp. 571–576.

[14] C. C. Tan and C. Eswaran, "Using autoencoders for mammogram compression," *Journal of medical systems*, vol. 35, no. 1, pp. 49–58, 2011.

[15] R. S. Lee, F. Gimenez, A. Hoogi, K. K. Miyake, M. Gorovoy, and D. L. Rubin, "A curated mammography data set for use in computer-aided detection and diagnosis research," *Scientific data*, vol. 4, p. 170177, 2017.

[16] D. Tellez, G. Litjens, J. van der Laak, and F. Ciompi, "Neural image compression for gigapixel histopathology image analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[17] H. Shen, W. D. Pan, Y. Dong, and M. Alim, "Lossless compression of curated erythrocyte images using deep autoencoders for malaria infection diagnosis," in *2016 Picture Coding Symposium (PCS)*. IEEE, 2016, pp. 1–5.

[18] A. S. Sushmit, S. U. Zaman, A. I. Humayun, T. Hasan, and M. I. H. Bhuiyan, "X-ray image compression using convolutional recurrent neural networks," in *2019 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*, 2019, pp. 1–4.

[19] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026–1034.