Autoencoder Image Compression Algorithm for Reduction of Resource Requirements

Young Joon (Fred) Kwon, MSE
Sinai BioDesign
Icahn School of Medicine at Mount Sinai
fred.kwon@icahn.mssm.edu

Danielle Toussie, MD
Department of Radiology
Icahn School of Medicine at Mount Sinai
danielle.toussie@mountsinai.org

G Anthony Reina, MD
Intel Corporation
g.anthony.reina@intel.com

Ping Tak Peter Tang, PhD
Facebook Research
pingtaktang@gmail.com

Amish H Doshi, MD
Department of Radiology
Icahn School of Medicine at Mount Sinai
amish.doshi@mountsinai.org

Eric K Oermann, MD
Departments of Neurosurgery and Radiology
New York University
Grossman School of Medicine
eric.oermann@nyulangone.org

Anthony B Costa, PhD
Sinai BioDesign
Icahn School of Medicine at Mount Sinai
anthony.costa@mssm.edu

1 Introduction

Staggering compute resources are used in the largest artificial intelligence (AI) models, a trend that has been increasing exponentially (doubling time of 3.4 months since 2012) [1]. Meanwhile, the expected doubling time of computational power (Moore's Law) has remained steady at 2 years, and compute and network infrastructure in the developing world improves inconsistently, if at all [1, 3]. Commercial compression schemes provide excellent compression rates but result in image quality losses that are unacceptable in a diagnostic radiology setting [10]. The development of a lightweight medical image compression machine learning (ML) algorithm that preserves diagnostic features can alleviate resource requirements for ML algorithm training and image interpretation.

2 Related Works

Autoencoders reduce the number of features that describe data via a type of dimensionality reduction [9]. Previous attempts to compress high-dimensional data, like radiological images, using autoencoders have failed to retain intricate details with high fidelity and diversity at high compression ratios and thus eliminated clinical utility [2, 16]. Recently, autoencoders have been more prominently incorporated as part of generative models rather than compression algorithms [5, 17, 13]. In particular, a vector quantized variational autoencoder (VQ-VAE) framework has been previously used to generate images with greater fidelity and diversity to compress radiographs while maintaining diagnostically relevant features [13]. In the VQ-VAE framework, the encoder models a categorical distribution to obtain integer values used as indices to a dictionary of embeddings, from which indexed values are passed to the decoder [13]. The dictionary that is learned during training remains fixed after

34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

deployment and is input agnostic. A VQ-VAE can thus encode images to a set of integer indices in a latent space of an even smaller size that can be transferred and subsequently decoded to the original image. To further increase the diversity of reconstructed images from a discrete set of latent values, a multi-level encoder approach (VQ-VAE-2) can extract both local and global features (Figure 1A).

3 Materials and Methods

We used the radiographs from the CheXpert dataset for training and previously unseen MIMIC-CXR dataset for testing [8, 7]. For the classification task, we used original images, its compressed latent vectors, or the reconstructed images of MIMIC-CXR dataset to train the DenseNet-121 classifier as previously described in Rajpurkar et al. [12] (Figure 1B).

Our models are based on the architecture of the two-level VQ-VAE-2 model from Razavi et al. [13]. 2D convolutions were of filter size 4, stride 2, and padding 1. There were 2 convolutions in the first level, 1 convolution in the second level. Each encoding layer had an output of 1 channel depth. We used the DenseNet-121 architecture that was first pre-trained on ImageNet, similar to the CheXNet study [4, 6, 12]. For DenseNet-121 training with compressed latent vectors, there was an additional transpose convolution layer that converted the two-channel input (concatenated decoder outputs from two layers) into three-channel output. We measured the MAC (multiply accumulate), memory consumption, and time per epoch of training on a single NVIDIA® V100 Tensor Core GPU. We compared the per-class AUROC for the classification performance.

4 Results

In all models, uncompressed inputs were represented by 16 bit floating point numbers from the original decoded .jpg files. VQ-VAE-2 resulted in a mean compression ratio (CR) of 24.0, Fréchet inception distance (FID) of 3.81, peak signal to noise ratio (PSNR) of 45.12, and structural similarity index measurement (SSIM) of 0.953. On visual inspection, VQ-VAE-2 preserved anatomical details and even the numbers and letters found on the images (Figure 2). JPEG2000, currently widely used compression algorithm, resulted in a mean CR of 15.4, FID of 2.13, PSNR of 46.05, SSIM of 0.968 when configured to give a similar FID (2.13) as that of VQ-VAE-2 (Figure 2). VQ-VAE-2 was also resistant to various input manipulations (e.g. random noise, blocks, swirls) and even generalized to other modalities (transesophageal echocardiography, TEE, shown in (Figure 3).

On a single NVIDIA V100 GPU, average memory utilization and time per epoch for training with latent vectors were reduced by 93.0% (11960 MB to 840 MB). The amount of computation per iteration decreased by 93.8% (60.22 GMAC (Giga multiply–accumulate operations) to 3.76 GMAC). The time per epoch during training was also reduced by 48.5% (25.55 min to 13.15 min). The DenseNet-121 algorithm gave an average AUROC of 0.8373 with the original image input, 0.9182 on the reconstructed image input, and 0.9100 on the latent vector input (Table 1).

5 Discussion

Most successful ML algorithms are using increasingly larger amounts of compute [1]. Compression of larger medical imaging data, including higher resolution data and 3D or 4D data, can allow researchers

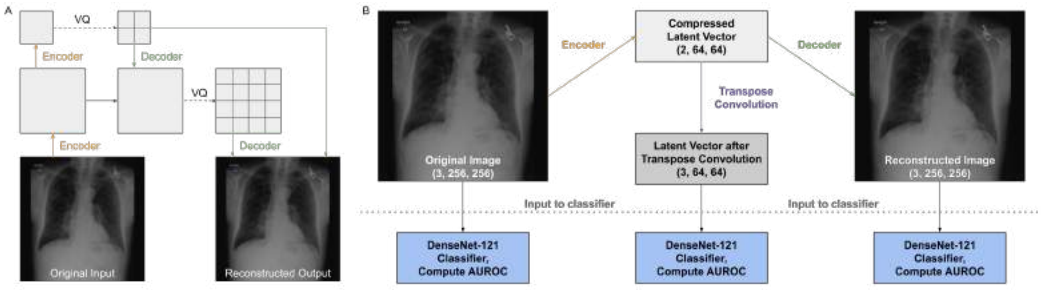

Figure 1: A) Two level VQ-VAE architecture. B) Training scheme with various inputs.

Original	JPEG2000 (Current Standard)	VQ-VAE-2
Compression Ratio	15.4	24.0
Fréchet Inception Distance	2.13	3.81
Peak Signal to Noise Ratio	46.05	45.12
Structural Similarity Index	0.968	0.953

Figure 2: Image reconstructions and associated quantitative metrics from JPEG2000 and VQ-VAE-2.

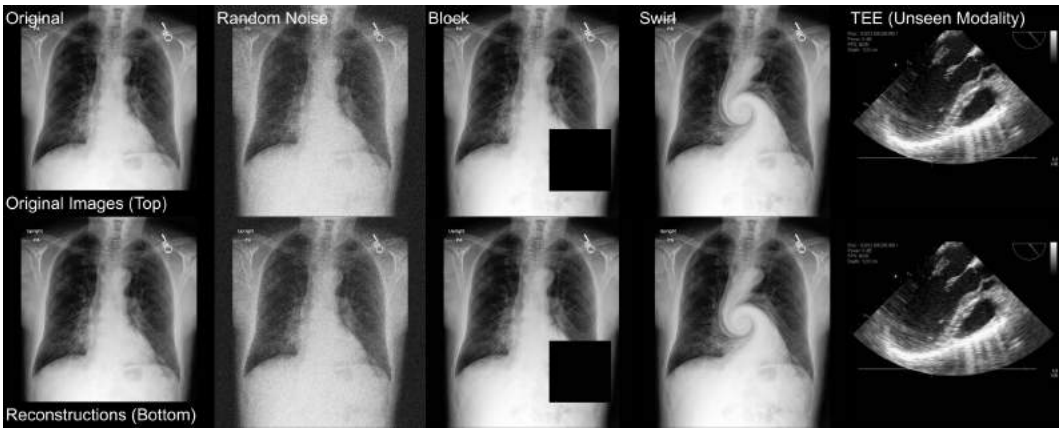

Figure 3: Robustness of VQ-VAE-2 to input manipulations and previously unseen modality transesophageal echocardiography (TEE). Top: original image input. Bottom: VQ-VAE reconstructions.

to train ML algorithms that showed promising results with 2D data but could not accommodate larger inputs due to memory or computational constraints. The advancement of novel compression schemes shown to retain classification performance and, in this case, diagnostic utility, have the potential to significantly democratize medical ML research to those with limited computational capacity, such as single-GPU systems with limited on-device memory. Even for the training of smaller networks, compression of data can lead to faster prototyping and less memory utilization. We have demonstrated similar AUROC performance when training classification methods with inputs from both the original image input and compressed latent vector representations, demonstrating the possible utility of VQ-VAE-2 in reducing the input file size and memory constraints. Furthermore, reducing bandwidth constraints for image transfer can increase access to radiological services in the most remote areas of the world.

Inputs	**Original**	**Reconstructed**	**Latent Vectors**
AUROC	0.8373	0.9182	0.9100
Train Time	25.55 min	25.57 min	13.15 min
Memory	11960 MB	11960 MB	840 MB
# Computes	60.22 GMAC	60.22 GMAC	3.7 GMAC

Table 1: AUROC and training metrics of DenseNet-121 trained with the original images, reconstructed images, or the compressed latent vectors.

Interestingly, the AUROC increased when the DenseNet-121 classifier was trained with the latent vectors and reconstructed images of VQ-VAE-2 as inputs compared to the original image alone tab:metrics. In fact, we obtain greater than the state of the art AUROCs in classifying such pathologies on certain pathologies [7]. This was an unexpected outcome. We interpret these results as a likely denoising effect driven by the VQ-VAE-2 encoder and reconstruction, improving the robustness of the classification task – a phenomenon extensively reported in engineering and medicine alike [18, 11, 15]. We have planned future studies to investigate the robustness of this phenomenon.

VQ-VAE-2 not only addresses the image compression but also does so with a relatively small number of layers (3 convolution layers total) in its network. Training of VQ-VAE-2 requires only images and no additional labels because the loss function is calculated based on how well the reconstructed image (i.e. output) resembles the original image (i.e. input). That is, training here is a self-supervised learning task. VQ-VAE-2 is robust to both small and large scale input manipulations that have previously impacted the performance of neural networks [14]. We have also demonstrated promising reconstruction for a previously unseen modality. Given that training is self-supervised, variational autoencoders may be adapted for other imaging modalities and anatomy without a need for a labeled dataset and alleviate resource requirements for faster ML algorithm training.

6 Broader Impact

Staggering compute resources are used in the largest artificial intelligence (AI) models, a trend that has been increasing exponentially (doubling time of 3.4 months since 2012). Meanwhile, the expected doubling time of computational power (Moore's Law) has remained steady at 2 years, while compute and network infrastructure in the developing world improves inconsistently, if at all. Commercial compression schemes provide excellent compression rates but result in image quality losses that are unacceptable in a diagnostic radiology setting. The development of a lightweight medical image compression machine learning (ML) algorithm that preserves diagnostic features can alleviate resource requirements for ML algorithm training and image interpretation. We hope to inspire an in-depth discussion of various techniques to alleviate technical resources required for machine learning research and lowering the barrier for new researchers to enter the field.

References

[1] D. Amodei. AI and compute. `https://openai.com/blog/ai-and-compute/`, May 2018. Accessed: 2020-1-13.

[2] A. Brock, J. Donahue, and K. Simonyan. Large scale GAN training for high fidelity natural image synthesis. Sept. 2018.

[3] E. Calandro, J. Chavula, and A. Phokeer. Internet development in africa: A content use, hosting and distribution perspective. In *e-Infrastructure and e-Services for Developing Countries*, pages 131–141. Springer International Publishing, 2019.

[4] J. Deng, W. Dong, R. Socher, L. Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, June 2009.

[5] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial networks. June 2014.

[6] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. Aug. 2016.

[7] J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute, H. Marklund, B. Haghgoo, R. Ball, K. Shpanskaya, J. Seekins, D. A. Mong, S. S. Halabi, J. K. Sandberg, R. Jones, D. B. Larson, C. P. Langlotz, B. N. Patel, M. P. Lungren, and A. Y. Ng. CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison. Jan. 2019.

[8] A. E. W. Johnson, T. J. Pollard, S. J. Berkowitz, N. R. Greenbaum, M. P. Lungren, C.-Y. Deng, R. G. Mark, and S. Horng. MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. *Sci Data*, 6(1):317, Dec. 2019.

[9] D. P. Kingma and M. Welling. Auto-Encoding variational bayes. Dec. 2013.

[10] F. Liu, M. Hernandez-Cabronero, V. Sanchez, M. W. Marcellin, and A. Bilgin. The current role of image compression standards in medical imaging. *Information*, 8(4):131, Oct. 2017.

[11] M. M. A. Rahhal, Y. Bazi, H. AlHichri, N. Alajlan, F. Melgani, and R. R. Yager. Deep learning approach for active classification of electrocardiogram signals. *Inf. Sci.*, 345:340–354, June 2016.

[12] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng. CheXNet: Radiologist-Level pneumonia detection on chest X-Rays with deep learning. Nov. 2017.

[13] A. Razavi, A. van den Oord, and O. Vinyals. Generating diverse High-Fidelity images with VQ-VAE-2. June 2019.

[14] J. Su, D. V. Vargas, and S. Kouichi. One pixel attack for fooling deep neural networks. Oct. 2017.

[15] W. Sun, S. Shao, R. Zhao, R. Yan, X. Zhang, and X. Chen. A sparse auto-encoder-based deep neural network approach for induction motor faults classification. *Measurement*, 89:171–178, July 2016.

[16] C. C. Tan and C. Eswaran. Using autoencoders for mammogram compression. *J. Med. Syst.*, 35 (1):49–58, Feb. 2011.

[17] A. van den Oord, O. Vinyals, and K. Kavukcuoglu. Neural discrete representation learning. Nov. 2017.

[18] D. Wang and X. Tan. Label-Denoising auto-encoder for classification with inaccurate supervision information. In *2014 22nd International Conference on Pattern Recognition*, pages 3648–3653, Aug. 2014.