Proceedings Paper:
Keighley, J orcid.org/0000-0002-3647-0683, de Kamps, M orcid.org/0000-0001-7162-4425, Wright, A et al. (1 more author) (2023) Digital pathology whole slide image compression with Vector Quantized Variational Autoencoders. In: Medical Imaging 2023: Digital and Computational Pathology. SPIE Medical Imaging 2023, 19-24 Feb 2023, San Diego, USA. SPIE .

https://doi.org/10.1117/12.2647844

Digital pathology whole slide image compression with Vector Quantized Variational Autoencoders

Jason Keighley[a, e], Marc de Kamps[a, e, f], Alexander Wright[a, b], and Darren Treanor[a, b, c, d]

[a]University of Leeds, Woodhouse Lane, Leeds, LS2 9JT, UK
[b]Leeds Teaching Hospitals NHS Trust, Great George St, Leeds, LS1 3EX, UK
[c]Department of Clinical Pathology, and Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden
[d]Center for Medical Image Science and Visualization (CMIV), Linköping University, Linköping, Sweden
[e]Leeds Institute for Data Analytics (LIDA), University of Leeds, Leeds, UK
[f]The Alan Turing Institute, 96 Euston Rd, London, NW1 2DB, UK

ABSTRACT

Digital pathology Whole Slide Images (WSIs) are large images (∼30 GB/slide uncompressed) of high resolution (0.25 microns per pixel), presenting a significant data storage challenge for hospitals wishing to adopt digital pathology. Lossy compression has been adopted by scanner manufacturers to address this issue - we compare lossy Joint Photographic Experts Group (JPEG) compression for WSIs and investigate the Vector Quantised Variational Autoencoder 2 variant (VQVAE2) as a possible alternative to reduce file size while encoding useful features in the compressed representation. We trained three VQVAE2 models on a Camelyon 2016 subset to the Compression Ratio (CR) of 19.2:1 (CR_1), 9.6:1 (CR_2) and 4.8:1 (CR_3) and tested on a Camelyon 2016 (DS_1) subset; University of California (DS_2) and Internal Validation Set (DS_3). We then compared compression performance to ImageMagick JPEG and JPEG 2000 implementations. Both JPEG and JPEG 2000 compression outperformed the VQVAE2 implementation within the Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM) metrics. The trained VQVAE2 models could visually reproduce WSI tissue structure, but used colours from the original training data within the reconstructions on other datasets.

Keywords: Variational Autoencoder, Quantisation, Compression, Pathology, Whole Slide Images, Digital Pathology

Further author information: (Send correspondence to Jason Keighley)
Jason Keighley: E-mail: ll13jjk@leeds.ac.uk

1. BACKGROUND

1.1 Introduction

Digital pathology: Digital pathology is a technology to digitise entire glass pathology slides, allowing the diagnosis of cancer on Whole Slide Images (WSIs).WSIs allow for remote diagnosis of digital images on computer workstations, and viewing of slides by multiple pathologists concurrently, greatly increasing the flexibility and efficiency of tissue diagnosis in cancer compared to glass slides.[1,2] Due to the high resolution of digital pathology WSIs (0.25 microns/pixel; typically 100,000x80,000 pixels), challenges exist in the storage and transmission of WSIs. To reduce file size, compression methods like Joint Photographic Experts Group (JPEG) 2000 are used by WSI scanners. Lossy compression can lead to image quality reduction for higher compression ratios. This work seeks to investigate the use of alternative compression methods for WSIs using Variational Autoencoders (VAEs), in particular the Vector Quantised Variational Autoencoder 2 variant (VQVAE2).[3]

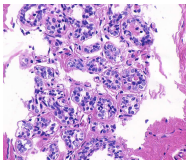
1.2 Related work

Traditional compression techniques: In order to enable the scanning of WSIs, digital pathology scanners use lossy JPEG compression. Although systems attempt to use compression methods that are visually lossless, lossy compression in medical imaging is not without issues, due to possible information loss.[4] This information loss can affect the accuracy or consistency of image analysis. Chen et al.[5] found that up to 85% JPEG compression would maintain a performance of at least 95% for deep learning algorithm detection and segmentation tasks compared to if using uncompressed data from the Camelyon 2017 dataset. Zanjani et al.[4] found similar results when they trained 3 convolutional neural network (CNN) deep learning algorithms on uncompressed images for nucleus segmentation, lymph node metastasis segmentation, and lymphocyte detection tasks. The trained networks were tested on JPEG 2000 compressed images with compression around 4% of their original size. In evaluating the effect of JPEG compression, metrics such as Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM) can be used to estimate the impact on image quality. SSIM is generally accepted as a better method for comparing image similarity versus PSNR, due to its ability to compare images in a similar way to human vision, with the metric being based on luminance, contrast and image structure.

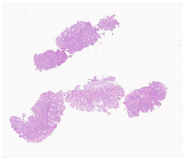
Medical image VAE compression: There is a paucity of evidence to support the use of VAEs for image compression in histopathology, however, more comprehensive compression research of the autoencoder (AE) without stochastic sampling has been previously published.[6–8] VAEs have been used previously for compression and meaning representation tasks within other clinical domains, including within x-ray and CT compression.[9] VAEs have a large advantage over AE models in real world settings. AEs can only generalise latent vectors to previously seen image features. VAEs map image features to distributions represented within the latent vectors. VAEs can traverse the probabilistic latent space generating data with similar latent features. AEs randomly assign mappings for unencoded latent space, therefore unseen data will be random for each trained AE model, reducing model reproducibility and reliability. The probabilistic latent space of VAEs does not experience this issue due to their probabilistic nature.

Proposal of the VAE: AEs have some weaknesses, including the inability to deal with previously unseen data. This means that outside the training datasets, the features identified by the encoder are not stable, and may not represent the actual input data. VAE latent space represents a distribution instead of representing the input space. Each distribution represents an approximate learned latent feature; therefore, each latent vector component can use probability distributions to approximate the latent features of unseen data. This architecture may be able to improve on the shortcomings of AEs, while not increasing complexity by an incomprehensible amount. VAE networks can suffer from issues like posterior collapse. VAE variants like the Vector Quantised Variational Autoencoder (VQVAE) and VQVAE2 aim to combat these issues. If the compression of VAEs could compete with existing methods, the compression would allow for the compressed representation to contain meaning within the latent features, possibly allowing for abstraction of clinically relevant features for faster analysis of the biopsy tissue contained on the slide.

(a) Camelyon 2016 training slide



(b) UCLA slide



(c) Internal Validation slide

Figure 1: Example dataset slides

2. MATERIALS AND METHODS

2.1 Methodology pipeline overview

Pipeline stages: The experiment methodology consisted of four main stages:

- **Data preprocessing**: We extracted tiles (size=(512, 512, 3), overlap=(12, 12, 0)) from Camelyon 2016 (DS_1), and selected tiles containing tissue samples via hysteresis thresholding. Hysteresis thresholding measures the connectiveness of edges between two defined thresholds. Areas that are connected to the high threshold and are above the low threshold are seen to be continuations of high confidence areas. Hysteresis thresholding allows for identification of connected edges. Tissue is made of multiple complex connected edges, whereas slide background noise is not, so is ignored within the thresholding.
- **VQVAE2 training**: We trained 6 VQVAE2 models (cloned from a GitHub repository[10]) on a subset of Camelyon 2016 (DS_1) thresholded tiles (Normal and Cancerous slides 1-10) for 500 epochs. Three VQVAE2s were trained at Compression Ratios (CRs) (∼4.8 [CR_1], ∼9.6 [CR_2] and ∼19.2 [CR_3]) with limited data augmentation, while the second three used a custom data augmentation method to randomise channels to increase model generalise-ability and transfer-ability.
- **VQVAE2 testing:** We tested the trained VQVAE2s on the DS_1 subset, University of California (DS_2) (n=21 benign, 21 malignant) and Internal Validation Set (DS_3) (n=1,935 512x512x3 patches). We then analysed the generated SSIM and PSNR metrics.
- **JPEG compression testing:** We tested the CR (2 to 97 with intervals of 5) for JPEG and JPEG 2000 compression using the ImageMagick implementation. We then analysed the generated SSIM and PSNR metrics.

2.2 Datasets

Camelyon 2016 subset (DS_1): Camelyon was created for cancer segmentation and classification tasks on digital pathology WSIs. Originally presented in the Camelyon challenge, the Camelyon dataset was created in 2016 as the Camelyon 2016 dataset. This was then expanded in 2017 as the Camelyon 2017 dataset. The Camelyon 2016 dataset is from two hospitals, whereas Camelyon 2017 is from five hospitals. For preliminary experiments on the networks, a subset of the Camelyon 2016 dataset was used, containing Normal slides 1-10 (excluding 9) and Cancerous slides (1-10). These WSIs were converted to smaller patches for network training and testing. This dataset was originally created using a 3DHistech P250 scanner with lossy compression, which was set to a compression quality setting of 80%. The public Camelyon dataset pyramidal BigTIFFs that are used within this research were originally compressed, and so may contain artifacts from this previous compression.

UCLA (DS_2): The UCLA dataset (n=21 benign, 21 malignant) is a dataset from the University of California. DS_2 consists of WSI slide scans, mainly of breast cancer tissue.

Internal validation set (DS_3): This was a dataset used for internal validation of the algorithms and testing. The DS_3 consisted of 8,455 512x512x3 patches which, after thresholding was applied, became 1,935 512x512x3 patches containing tissue. The other datasets (DS_1 and DS_2) are from real world data, but DS_3 is from routine histopathology image data.

2.3 Data preprocessing

WSI patch extraction: OpenSlide was used to extract patches of 512x512x3 with an overlap of 12. We then used the scikit-image hysteresis thresholding implementation to create a hysteresis threshold map. The hysteresis threshold map was used to measure the amount of tissue within the patch, with patches containing over 10% tissue area being placed into a thresholded patch set for model training.

Data augmentation: Data augmentation was used on the training image patches (X) in the form of torchvision.transforms.Compose([torchvision.transforms.Resize((512, 512)), torchvision.transforms.toTensor()]).[11] A custom colour randomising algorithm was also used during VQVAE2 training, which used Image:$x_{ijk} \mapsto U \sim (0, 2) \times (x_{ijk}); k \in \{0, 1, 2\}$.

Data augmentation reasoning: Within preliminary testing, the trained VQVAE2s correctly reconstructed cellular and tissue structures within the unseen testing images. DS_1 trained models without data augmentation, overfitted to the DS_1 dataset colour palette of the trained dataset. These models reconstructed the image structure correctly, while applying the learnt colours from the original data. The trained models would not generalise to other scanner datasets, as reconstructions contained the correct tissue structure but colour similar to the DS_1 training images. Initial testing of uniform or normal distribution single colour channel scale factors produced a range of colours within the training images, so trained models learnt from increased colour data, allowing for greater model generalise-ability on datasets with different colour palettes.

2.4 Algorithms

VQVAE2 architecture: The original VQVAE architecture uses an encoder, a decoder and a quantizer. The VQVAE2 modified this with a secondary encoder and decoder layer. The secondary (top) and primary (bottom) encoder and decoder layers encode for high level and low-level features within the image respectively.

Loss function: The loss function within the VQVAE2 has three components: the reconstruction loss (L_1), the embedding loss (L_2), and the commitment loss (L_3). The reconstruction loss encodes for the accuracy of the reconstructed image (X') and can be any existing loss function, including the mean squared error (MSE). The embedding loss moves the code book embeddings ($Z_{embed(X)}$) towards the normal distribution quantised embeddings ($Z_{embed(quant)}$). The commitment loss moves the embedding ($Z_{embed(quant)}$) towards the image encoded embedding ($Z_{embed(X)}$).

VQVAE2 training: The VQVAE2 implementation was cloned from a GitHub repository.[10] The default parameters were used for initial training and preliminary testing. The dimension of embeddings and number of embedding parameters were modified to achieve larger or smaller compression of the 512x512x3 image patches. The VQVAE2 was trained on the DS_1 subset training data, and was then tested on all other datasets.
The default VQVAE2 network parameters (in_channel=3, channel=128, n_res_block=2, n_res_channel=32, embed_dim=64, n_embed=512, decay=0.99, loss_function= LossFunction.mean_squared_error) led to a CR of 0.6. The embedding dimension parameter can be changed to alter the level of compression within the compressed representation. The embedding dimensions for the embed_dim values of 8, 4, 2 are 4.8:1, 9.6:1 and 19.2:1 respectively for the CR between the compressed representation ($X^{'}$) and the original patch image (X). For example, within the training the embedding dimension value of 2 was used to create a 19.2:1 CR with a $R^{((2,128,128))} + R^{((2,64,64))}$ latent space for the $R^{((512,512,3))}$ image data (X).

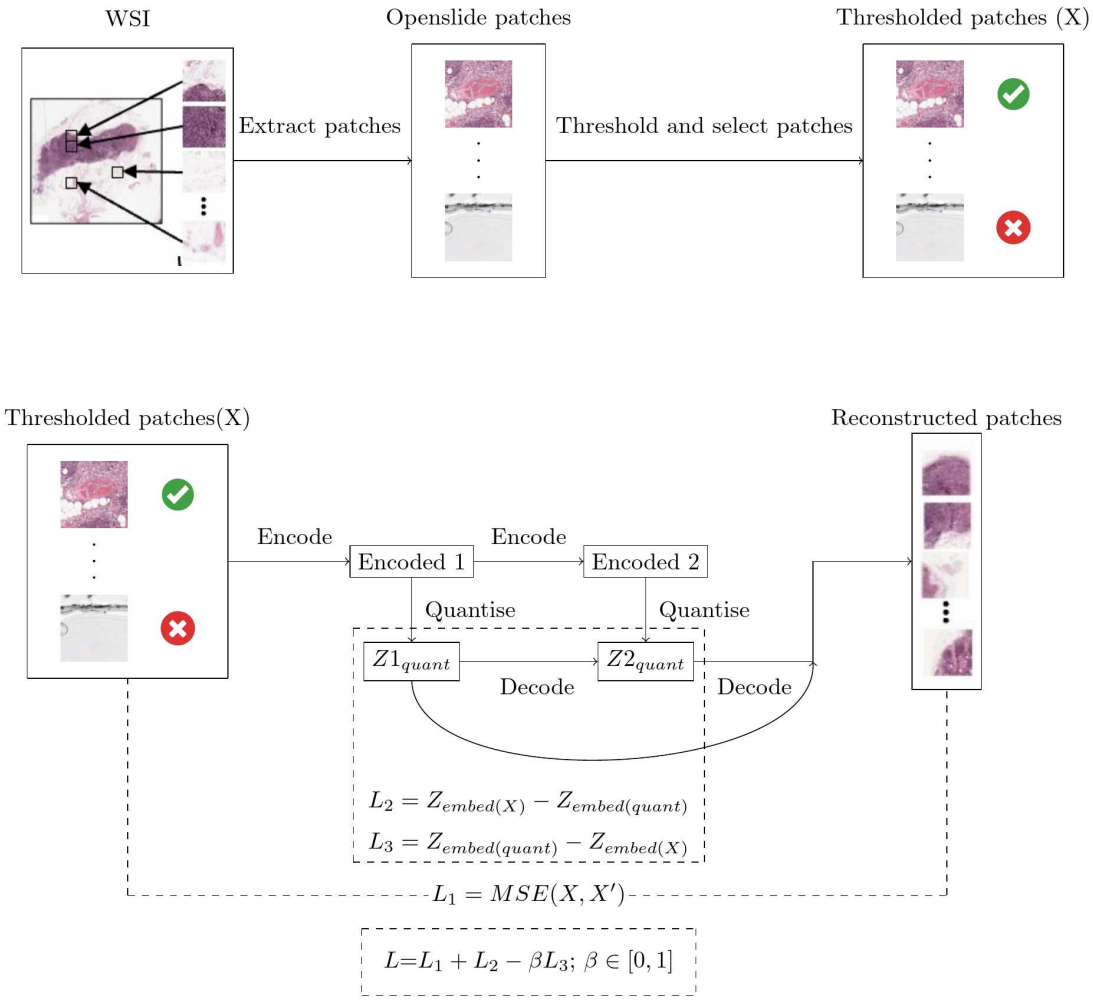
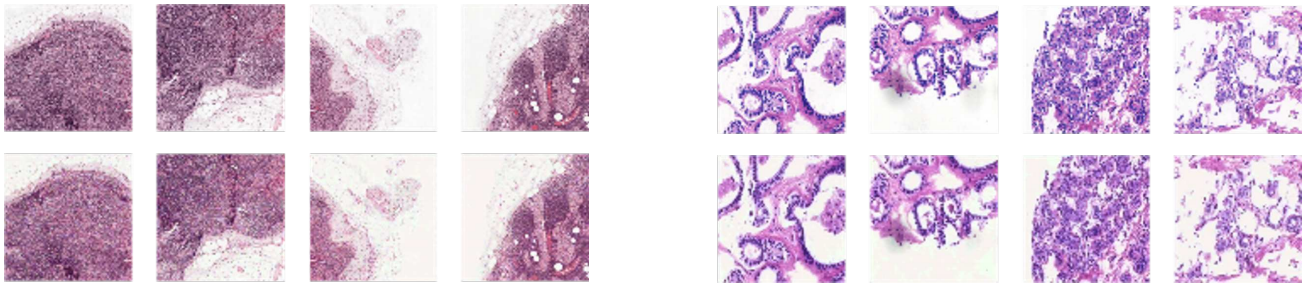


Figure 2: Example path of a whole slide image file. This file is patched using openslide, thresholded using hysteresis thresholding, and then selected or denied for training or testing. After a patch is selected, it is inputted into the VQVAE2 and is reconstructed. This diagram shows an example of 4 patches from the Camelyon 2016 subset dataset and 4 example reconstructed patches.



(a) Camelyon 2016 subset (b) UCLA

Figure 3: Visual comparison of reconstructed (bottom row) and original (top row) patches.

JPEG: Two ImageMagick JPEG implementations were used: the generic JPEG implementation ("magick source_file -strip -quality quality result_file"), and the JPEG 2000 compression ("magick source_file -strip -colorspace RGB -define jp2:quality compression_ratio result_file"). All JPEG compression algorithms were run on the same datasets as the VQVAE2, using the same Pytorch dataloaders, without data augmentation other than "torchvision.transforms.Resize((512, 512))" for all patches within the pathology datasets.
The JPEG compression quality flag was set to a quality from 2% to 99% by intervals of 5%. This was manually mapped to the compression ratio of the compressed representation. The JPEG 2000 used the same numbers from 2 to 99 but within JPEG 2000. This value represented the desired compression ratio. The PSNR and SSIM values were calculated for all reconstructed images and plotted as a function of the compression ratio for both the JPEG and JPEG 2000 algorithms.

3. RESULTS

Visual differences between JPEG and VQVAE2 compression: JPEG and JPEG 2000 are compression techniques used in WSI storage within hospitals. The compression ratio of these stored images is usually around a 15:1 to 20:1 compression ratio. The 19.2:1 compression ratio VQVAE2 reconstructed images were then visually compared to the decompressed images from the JPEG compression testing.
The reconstructions seemed to perform comparably on all datasets compared to the JPEG compression at a similar compression level. Example patch reconstructions for the VQVAE2 can be seen in figure 3.

Custom data augmentation: The VQVAE2 trained without data augmentation outperformed the custom colour augmented datasets in all quantitative tests. The difference in PSNR (mean (μ) [standard deviation (σ)]) between the VQVAE2 trained to the compression ratio of $\sim$19.2 was 1.22 [0.17], -2.2 [0.59], 1.13 [0.42] for the mean and standard deviations of 19.92$\pm$1.83 dB, 20.09$\pm$1.29 dB, 21.95$\pm$1.89 dB (no colour data augmentation) and 18.70$\pm$1.66 dB, 22.29$\pm$0.70 dB, 20.82$\pm$1.47 dB (with randomised colour data augmentation) for the DS_1 subset, DS_2 and DS_3. The difference in the SSIM for the VQVAE2 trained to the compression ratio of $\sim$19.2 was 0.11 [-0.024], 0.03 [-0.001], 0.07 [-0.009] from the mean and standard deviation of 0.72$\pm$0.076, 0.88$\pm$0.018, 0.80$\pm$0.060 (no colour data augmentation) and 0.61$\pm$0.10, 0.85$\pm$0.019, 0.73$\pm$0.069 (with randomised colour data augmentation) for the DS_1 subset, DS_2 and DS_3 respectively.
The boxplots in figure 5 illustrate the differences between the PSNR and SSIM results for JPEG 2000 and the other methods.

JPEG compression quality quantitative results: The peak signal to noise ratio (PSNR) and structural similarity (SSIM) JPEG metrics were calculated for the JPEG compression. The PSNR for JPEG on the tested pathology datasets dropped significantly from 43 to 32 between the 1:1 and 3:1 compression ratios. The PSNR for JPEG maintained stable with a PSNR of around 29 after the 5:1 compression ratio.

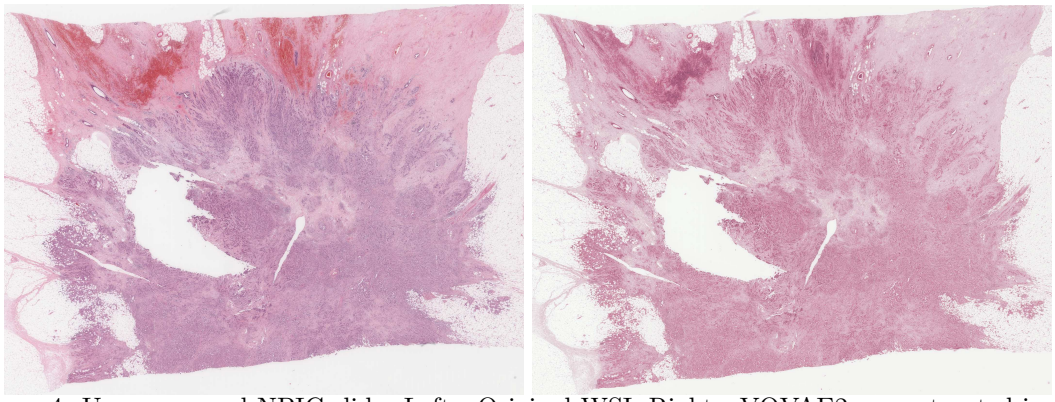


Figure 4: Uncompressed NPIC slide. Left - Original WSI, Right - VQVAE2 reconstructed image

JPEG vs VQVAE2 (mean and standard deviation): The mean (μ) and standard deviation (σ) PSNR ($\mu \pm \sigma$) for DS_2, DS_1 subset and DS_3 were 31.50±0.81 dB, 29.70±1.28 dB, 29.91±1.21 dB for the CR of ∼19.2:1 for JPEG compression. For the VQVAE2 compression the PSNR mean and standard deviation were 20.3±1.09 dB, 19.6±1.77 dB, 21.49±1.83 dB, respectively. SSIM considers image structure, PSNR does not. SSIM ($SSIM \in [0, 1]$) results for the DS_2, DS_1 subset and DS_3 SSIM were 0.84±0.022, 0.83±0.034, 0.80±0.042 for JPEG and 0.85±0.022, 0.72±0.072, 0.79±0.056 for the VQVAE2 implementation.

19:1 JPEG vs 19.2:1 VQVAE2 (medium and interquartile range): The PSNR medium and interquartile range for the UCLA and Camelyon 2016 subset and internal validation set are ∼21, 20, 21 and ∼32, 29, 30 for the VQVAE2 and JPEG respectively. The SSIM medium and interquartile range for the UCLA and Camelyon 2016 subset and internal validation set are ∼0.85, 0.72, 0.79 and ∼0.84, 0.81, 0.8 for the VQVAE2 and JPEG respectively.

4. DISCUSSION

The quantitative metrics for the initial VQVAE2 testing showed that the PSNR and SSIM for the VQVAE2 was lower than that of JPEG and JPEG 2000. The visual comparisons indicate that the VQVAE2 could reconstruct slide image structure, but could not reconstruct the colour correctly.

As the learnt model was able to generalise between different scanners and different datasets from different locations, the latent space was shown to encode features that represented the image features within the image. Future work aims to understand the representations generated by VAE networks, aiming to create clinically useful representations of WSI slides and patches while allowing for compression of the pathology images. Future work will focus on comparing the compression error type to subjective image quality and performance metrics, particularly whether clinician performance would be affected by the VQVAE2 reconstructions, as well as more rigorous testing of the VQVAE2 on uncompressed data.

The Camelyon dataset used for training of the VQVAE model was originally compressed with a lossy compression of 80%, therefore the model may have learnt compression artefacts that may not be present on uncompressed WSIs. The tested VAEs networks may have been mapping too much towards the reconstructed image, rather than ensuring mapping to the normal distribution feature representations, therefore ignoring the normal distribution feature mapping in favour of minimizing the reconstruction image accuracy. This would have lead to a less randomly generalisable model, where the model would aim to move the reconstructions towards its previously seen training data, rather than to a stochastic latent variable. The lower PSNR and SSIM results for the VQVAE2 implementation could be due to the use of suboptimal training hyper parameters, or possibly the use of a loss function that did not focus on the image reconstruction quality, with the latent space embedding representations being favoured or balanced over the reconstruction accuracy metrics of PSNR or SSIM. A full hyperparameter search for the VQVAE2 network was not performed to gain the optimal parameters for the network compression. As the metrics were done on a per patch basis, rather than per WSI, this may not be fully transferable to a WSI compression metric.

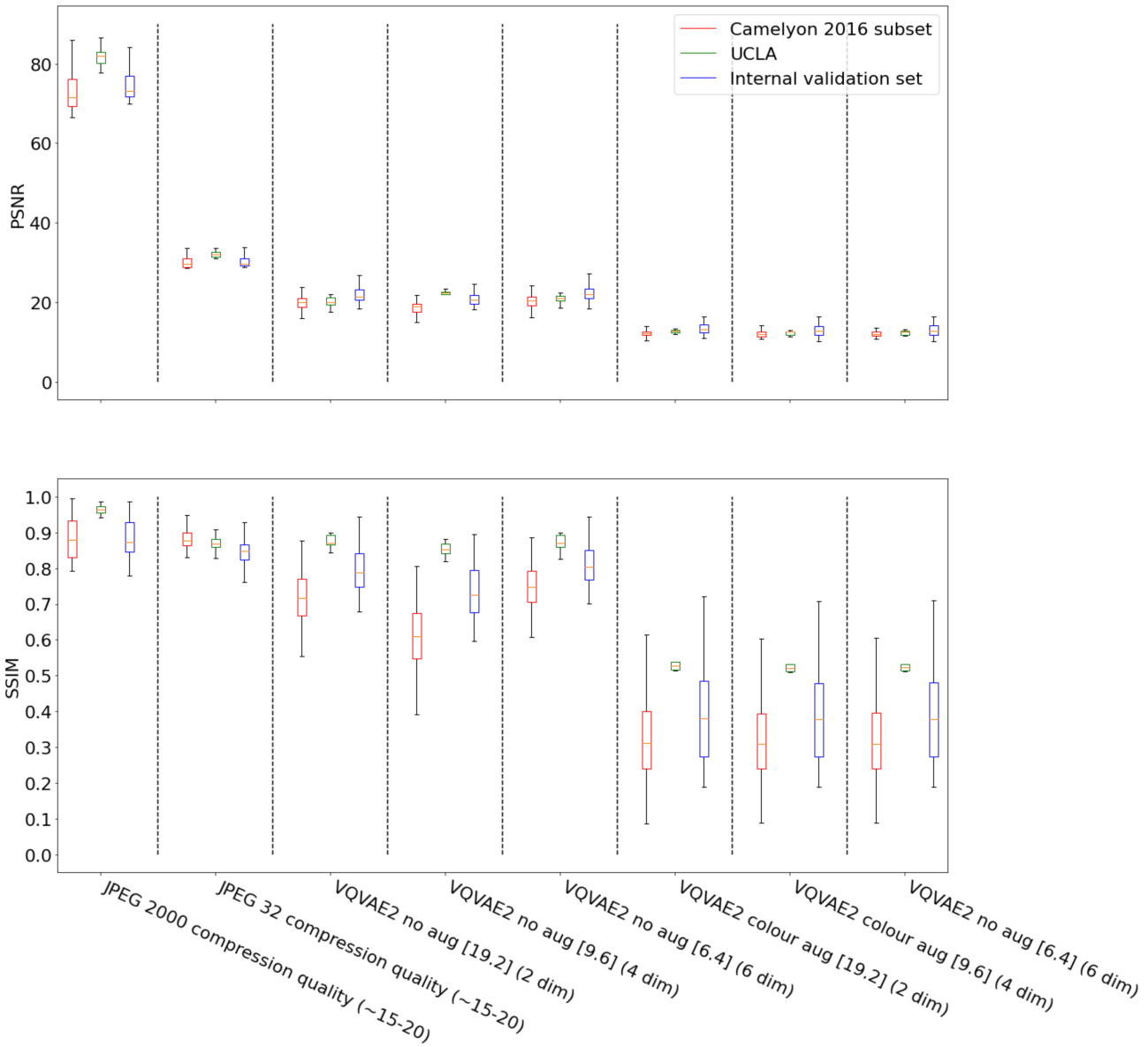


Figure 5: The peak signal to noise ratio and structural similarity metrics for all tested algorithms. The JPEG 2000 and JPEG algorithms are grouped to the left, the DS_1 trained VQVAE2s to the right, with the different data augmented VQVAE2 networks being grouped together. The compression ratio is noted within the square brackets for each algorithm (Camelyon 2016 subset DS_1, UCLA DS_2 and the internal validation set DS_3).

Appendices

Additional data files

Code availability Code can be found at https://github.com/Jjk422/Digipath-WSI-compress-VQVAE.

Data availability Data for the Camelyon 2016 dataset can be found at https://camelyon16.grand-challenge.org/, DS_2 data can be found at https://bioimage.ucsb.edu/research/bio-segmentation. The internal validation set contains patient data and so cannot be publicly shared.

Compliance with Ethical Standards

This work was carried out under regional ethical approval (Leeds West REC, reference 05/Q1205/220).

Compliance with EPSRC open access requirements

In compliance with the UKRI and EPSRC open access requirements, publishable material and code has been made available as part of a Creative Commons Attribution (CC BY) licence. The internal validation dataset contains patient data and so cannot be made publicly available. The code and data used for the experimentation can be found at https://github.com/Jjk422/Digipath-WSI-compress-VQVAE.

Acknowledgements

JK is funded via the Leeds University Centre for Doctoral Training (CDT) for AI in medical diagnosis and care with a grant from EPSRC (Project no. 2271095).
DT and AW are funded by National Pathology Imaging Co-operative, NPIC (Project no. 104687), supported by a £50m investment from the Data to Early Diagnosis and Precision Medicine strand of the government's Industrial Strategy Challenge Fund, managed and delivered by UK Research and Innovation (UKRI).

REFERENCES

[1] Williams, B. J., Bottoms, D., and Treanor, D., "Future-proofing pathology: the case for clinical adoption of digital pathology," *Journal of Clinical Pathology* **70**, 1010–1018 (Dec. 2017).

[2] Williams, B. J., Bottoms, D., Clark, D., and Treanor, D., "Future-proofing pathology part 2: building a business case for digital pathology," *Journal of Clinical Pathology* **72**, 198–205 (Mar. 2019).

[3] Razavi, A., Oord, A. v. d., and Vinyals, O., "Generating Diverse High-Fidelity Images with VQ-VAE-2," (June 2019).

[4] Ghazvinian Zanjani, F., Zinger, S., Piepers, B., Mahmoudpour, S., and Schelkens, P., "Impact of JPEG 2000 compression on deep convolutional neural networks for metastatic cancer detection in histopathological images," *Journal of Medical Imaging* **6**, 1 (Apr. 2019).

[5] Chen, Y., Janowczyk, A., and Madabhushi, A., "Quantitative Assessment of the Effects of Compression on Deep Learning in Digital Pathology Image Analysis," *JCO Clinical Cancer Informatics* , 221–233 (Sept. 2020).

[6] Xu, J., Xiang, L., Liu, Q., Gilmore, H., Wu, J., Tang, J., and Madabhushi, A., "Stacked Sparse Autoencoder (SSAE) for Nuclei Detection on Breast Cancer Histopathology Images," *IEEE Transactions on Medical Imaging* **35**, 119–130 (Jan. 2016).

[7] Roy, M., Kong, J., Kashyap, S., Pastore, V. P., Wang, F., Wong, K. C. L., and Mukherjee, V., "Convolutional autoencoder based model HistoCAE for segmentation of viable tumor regions in liver whole-slide images," *Scientific Reports* **11**, 139 (Dec. 2021).

[8] Fassler, D. J., Abousamra, S., Gupta, R., Chen, C., Zhao, M., Paredes, D., Batool, S. A., Knudsen, B. S., Escobar-Hoyos, L., Shroyer, K. R., Samaras, D., Kurc, T., and Saltz, J., "Deep learning-based image analysis methods for brightfield-acquired multiplex immunohistochemistry images," *Diagnostic Pathology* **15**, 100 (Dec. 2020).

[9] Kwon, Y. J., Toussie, D., Azour, L., Concepcion, J., Eber, C., Reina, G. A., Tang, P. T. P., Doshi, A. H., Oermann, E. K., and Costa, A. B., "Appropriate Evaluation of Diagnostic Utility of Machine Learning Algorithm Generated Images," in [*Proceedings of the Machine Learning for Health NeurIPS Workshop*], 179–193, PMLR (Nov. 2020). ISSN: 2640-3498.

[10] Seonghyeon, K., "vq-vae-2-pytorch." https://github.com/rosinality/vq-vae-2-pytorch (Jan. 2023).

[11] The PyTorch Foundation, "Torchvision main documentation." https://pytorch.org/vision/stable/index.html (2023).