DEEP QUANTIZED REPRESENTATION FOR ENHANCED RECONSTRUCTION

Akash Gupta[†,⋆], *Abhishek Aich*[†], *Kevin Rodriguez*[‡], *G. Venugopala Reddy*[‡], *Amit K. Roy-Chowdhury*[†]

[†]Department of Electrical and Computer Engineering, [‡]Department of Botany and Plant Sciences,
University of California, Riverside

ABSTRACT

While machine learning approaches have shown remarkable performance in biomedical image analysis, most of these methods rely on high-quality and accurate imaging data. However, collecting such data requires intensive and careful manual effort. One of the major challenges in imaging the Shoot Apical Meristem (SAM) of Arabidopsis thaliana, is that the deeper slices in the $z-$stack suffer from different perpetual quality related problems like poor contrast and blurring. These quality related issues often lead to disposal of the painstakingly collected data with little to no control on quality while collecting the data. Therefore, it becomes necessary to employ and design techniques that can enhance the images to make it more suitable for further analysis. In this paper, we propose a data-driven Deep Quantized Latent Representation (DQLR) methodology for high-quality image reconstruction in the Shoot Apical Meristem (SAM) of Arabidopsis thaliana. Our proposed framework utilizes multiple consecutive slices in the z-stack to learn a low dimensional latent space, quantize it and subsequently perform reconstruction using the quantized representation to obtain sharper images. Experiments on a publicly available dataset validate our methodology showing promising results. Our code is available at github.com/agupt013/enhancedRec.git.

Index Terms— Cell reconstruction, quantized representation, shoot apical meristem, arabidopsis thaliana

1. INTRODUCTION

Automated analysis in biomedical research is critical to provide researchers with concrete evidence to prove any proposed hypothesis without any bias. However, automated image analysis requires high-quality imaging data. Image quality related problems are often encountered while imaging deeper layers of the Shoot Apical Meristem (SAM) of arabidopsis thaliana [2]. These quality related problems hinder automated analysis and often lead to disposal of painstakingly collected data. To this end, we propose a data driven Deep Quantized Latent Representation (DQLR) framework for high-quality imaging data reconstruction of the $z-$stack of the SAM. In this work, we propose to project noisy stack in a latent space, quantize the latent representations and utilize the quantized latent representations for reconstruction of enhanced $z-$stack (see Fig. 1 for conceptual overview).

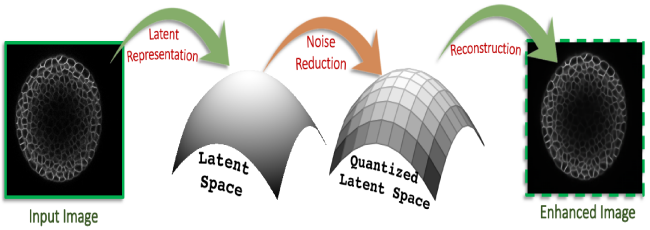

Fig. 1: **Conceptual Overview of DQLR.** The latent representation of the collected image is quantized using $k-$means over the entire dataset [1]. This quantized representation is then used to reconstruct the enhanced image.

Overview. An architectural overview of our approach is illustrated in Fig. 2. During training, the encoder E compresses i^{th} input slice image to a latent representation x_i. The consecutive slices in the z-stack are correlated which implies that they must be correlated in the latent space as well. We employ a recurrent neural network (RNN) R to learn this correlated representation $\{y_i, y_{i+1}, \cdots, y_{i+n}\}$ by passing the latent vector $\{x_1, x_2, \cdots, x_n\}$ through R. The compressed representation x_i is processed through R_i to learn the inter-correlation between this latent representation of the consecutive slices $\{x_i, x_{i+1}, \cdots, x_{i+n}\}$ during training. RNN generated latent codes $\{y_i, y_{i+1}, \cdots, y_{i+n}\}$ are then used as input to quantization module Q_i. Q_i learns a vector dictionary for quantized representation of the network and generates a quantized latent code $\{y_i^q, y_{i+1}^q, \cdots, y_{i+n}^q\}$. In our proposed method, the quantization of the latent code will remove the noisy component of $\{y_i\}$ and the reconstructed/predicted images using the quantized latent codes by generator G should be enhanced. During testing, we pass one slice at a time from the $z-$stack, compress it using the encoder, predict the correlated latent codes using the RNN, and finally quantize it using the quantization dictionary learned during the training stage using Q. This quantized code is then used to reconstruct and predict enhanced consecutive slices from the given $z-$stack.

This work was partially supported by NSF grants 1664172 from the Office of Advanced Cyberinfrastructure and 1762063 from the Division of Mathematical Sciences. ⋆indicates corresponding author. E-mails: {agupt013@, aaich001@, krodr005@, venug@, amitrc@ee.}ucr.edu.

2. RELATED WORK

In this section we describe prior works closely related to the our proposed method. Our method closely relates to reconstruction using auto-encoders [3] and enhancement in the compressed domain [1, 4, 5].

Auto-Encoders. Variations of auto-encoders are extensively used in reconstruction tasks by compressing the input to a latent representation and using the latent representation to retrieve the input as close as possible [1, 3, 6]. However, often the reconstructed images are blurry due to inherent nature of Mean Square Error (MSE) loss to produce blurry results. In the proposed approach we also include Structural Similarity Index (SSIM) [7] loss to enhance the visual results.

Compressed Domain Enhancement. Some works have tried to enhance the images in the compressed domain. In [4] a method based on a contrast measure defined within the discrete cosine transform (DCT) domain is proposed to enhance the image. Authors in [1] propose a vector quantized variations auto-encoder for reconstruction of various media input. We adopt their approach of vector quantization in our framework. However, we exploit the input data correlation using RNN for enhancement task as opposed to reconstruction in [1] where ground truth data was available.

3. METHODOLOGY

We propose a Deep Quantized Latent Representation (DQLR) framework for enhancing $z-$stack imaging in SAM of Arabidopsis thaliana. We apply quantization in the latent space of the noisy $z-$stack for enhanced reconstruction. In this section, we first formulate the problem statement and then explain our proposed approach in details.

3.1. Problem Formulation

Given a $z-$stack $\mathbf{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_n\}$, with $\mathbf{z}_i$ being the i^{th} slice in the stack from the top, we aim to reconstruct $\widehat{\mathbf{Z}} = \{\widehat{\mathbf{z}}_1, \widehat{\mathbf{z}}_2, \cdots, \widehat{\mathbf{z}}_n\}$ such that $\widehat{\mathbf{z}}_i$ is the visually enhanced slice compared to $\mathbf{z}_i, \forall i = 1, 2, \cdots, n$. Let there be a latent representation of input noisy $z-$stack $\mathbf{X_Z} = \{x_1, x_2, \cdots, x_n\}$ where x_i is the latent representation corresponding to the i^{th} slice $\mathbf{z}_i$. Since the slices in $z-$stack are correlated in the pixel space, their latent representations should inherit the same property in the latent space. Therefore, corresponding to each latent representation $\mathbf{X_Z}$ let there be a latent representation $\mathbf{Y_Z} = \{y_1, y_2, \cdots, y_n\}$ such that all $\{y_i\}$ are correlated.

We propose to generate visually enhanced $z-$stack by quantizing the latent representation of the noisy input stack. Our hypothesis is that each correlated latent representation y_i of a slice in the $z-$stack consists of two components; the quantized representation y_i^q and the noise representation y_i^{noise} of y_i, such that $y_i = y_i^q + y_i^{\text{noise}}$. Hence, noise component y_i^{noise} can be removed by applying quantization on the correlated latent codes leaving the representation y_i^q required to generate the enhanced image $\widehat{\mathbf{z}}_i \; \forall \; i = 1, 2, \cdots, n$.

3.2. Proposed Approach

Our proposed framework is shown in Figure 2. It consists of four components: the encoder network E, the recurrent neural network R, the quantization module Q and the generator network G. The encoder network is used to extract latent representation for each slice in the noisy input stack. The recurrent neural network utilizes the latent representations to generate correlated latent representations. These correlated representations are quantized to reduce noise in the latent space by the quantization module. Finally, the quantized representations are used to generate an enhanced $z-$stack.

Input Latent Representation. We employ a convolutional neural network as an encoder E which extracts the latent representation for each slice in a given noisy $z-$stack such that

$$\mathsf{E}(\mathbf{Z}) = \mathsf{E}(\{\, \mathbf{z}_1,\, \mathbf{z}_2, \cdots,\, \mathbf{z}_n\}) = \{\, x_0,\, x_1, \cdots,\, x_n\} \quad (1)$$

where x_i is latent representation corresponding to slice $\mathbf{z}_i$. A set of correlated representations is generated by the recurrent neural network for the latent representations extracted from the encoder E to incorporate the z-resolution dynamics of the $z-$stack in the latent representations.

Recurrent Neural Network (RNN). The consecutive slices in a $z-$stack capture 3D-structure of any cell in the plant. Thus, there must be a correlation between the consecutive slices. The latent representation $\mathbf{X_Z}$ of the noisy input $\mathbf{Z}$ should also be correlated in some space $\mathbf{Y_Z}$. Thus, we employ a recurrent neural network R_i to transform the i^{th} noisy latent representation to the correlated latent representation as RNN can capture dynamics of the sequence given by

$$y_{i+1} = \mathsf{R}_i(y_i, h_i) \quad (2)$$

where h_0 is the hidden state sampled randomly from a Gaussian distribution and $h_i = x_{i-1} \; \forall \; i > 0$. Here, we aim to capture the $z-$resolution dynamics of the stack unlike traditional recurrent neural network where temporal dynamics of the sequence is captured.

Deep Quantized Latent Representation. We propose that a data driven quantization of the latent representation can reduce the average noise in the stack and enhance it visually. In order to quantize the latent representation, we employ vector quantization dictionary learning algorithm as proposed in [1], represented as Q_i in our framework.

Enhanced Stack Generation. We employ a generative model G to transform the quantized representations into

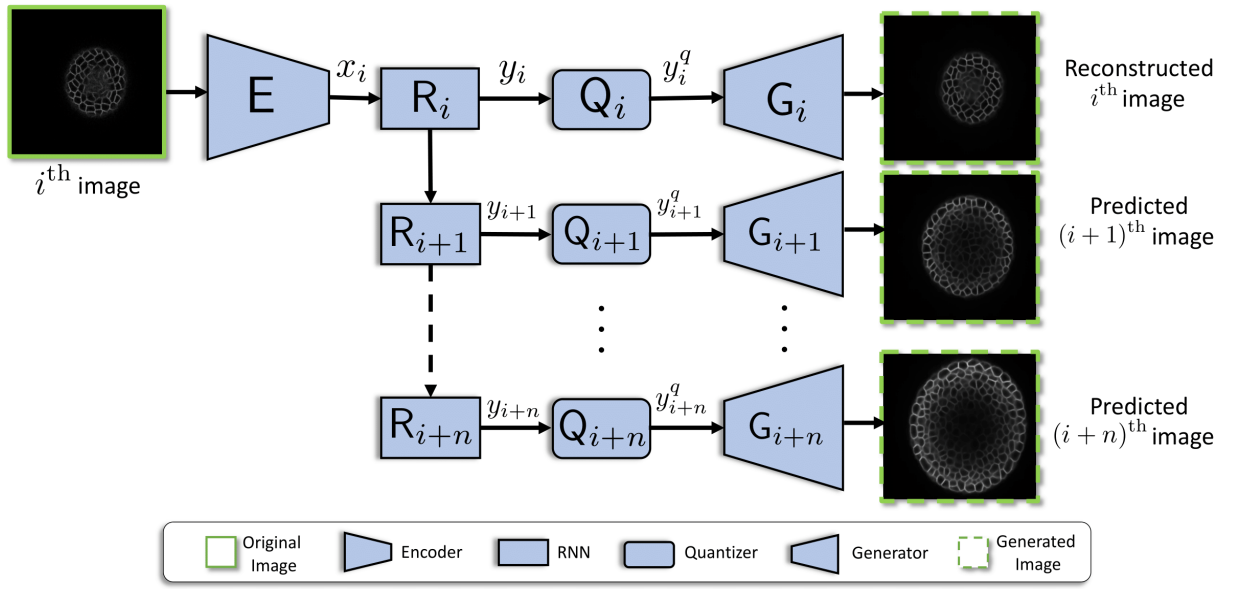

Fig. 2: **Architectural Overview of DQLR (for one slice of the stack).** Encoder E encodes input image to x_i. Recurrent Neural Network (RNN) module generates correlated codes for reconstruction (y_i) and prediction ($\{y_i, y_{i+1}, \cdots, y_{i+n}\}$). Quantizer module Q_i quantizes the latent codes and Generator G reconstructs/predicts the images.

an enhanced stack $\widehat{\mathbf{Z}}$. The quantized representations $\mathbf{Y}_{\mathbf{Z}}^q$ are used by the generator G to synthesize enhanced stack $\widehat{\mathbf{Z}} = \{\widehat{\mathbf{z}}_1, \widehat{\mathbf{z}}_2, \cdots, \widehat{\mathbf{z}}_n\}$ such that $\widehat{\mathbf{z}}_i$ is the visually enhanced image of the slice $\mathbf{z}_i$ in the noisy stack $\mathbf{Z}$.

3.3. Optimization

Our optimization function consists of the Mean Squared Error (MSE) pixel reconstruction loss, the Structural Similarity (SSIM) loss [7] and quantization loss as defined in [1].

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{mse}} + \lambda_{\text{s}}\mathcal{L}_{\text{ssim}} + \lambda_{\text{q}}\mathcal{L}_{\text{quant}} \tag{3}$$

We briefly describe the loss functions below. Define P as the total number of non-overlapping patches in a given image, N as total number of pixels in P, and α and β as the generated and ground truth image, respectively.

$$\mathcal{L}_{\text{mse}}(P) = \frac{1}{N}\sum_{p\in P} \|\alpha(p) - \beta(p)\|_2$$

$$\mathcal{L}_{\text{ssim}}(P) = \frac{1}{N}\sum_{p\in P} 1 - \text{SSIM}(p),$$

$$\text{with, } \text{SSIM}(p) = \left(\frac{2\mu_\alpha\mu_\beta + C_1}{\mu_\alpha^2 + \mu_\beta^2 + C_1}\right)\left(\frac{2\sigma_{\alpha\beta} + C_2}{\sigma_\alpha^2 + \sigma_\beta^2 + C_2}\right)$$

where, $\mu_{(\cdot)}$ and $\sigma_{(\cdot)}$ are computed with a Gaussian filter with standard deviation σ_G, $C_1 < 1$ and $C_2 < 1$ are constants introduced to handle division by zero issue, λ_{s} and λ_{q} weights for SSIM and quantization loss, respectively. For $\mathcal{L}_{\text{quant}}$, we use the loss function as proposed in [1] on the correlated latent space $\mathbf{Y}_{\mathbf{Z}}$ and dictionary $\mathbf{D} = \{d^1, d^2, \cdots, d^k\}$, where $k = 128$ is length of dictionary to learn for quantization.

4. EXPERIMENTATION AND RESULTS

Datasets. We used the publicly available Confocal Membrane dataset [8] consisting of six plants. We train our model using four plant stacks, and use one plant stack each for validation and testing.

Qualitative Results. Fig. 3 shows few examples of the reconstructed slices from the $z-$stack using the our approach along with the input slice. It can be observed that our proposed method is able to generate sharper cell boundaries. Since we learn the quantization dictionary using all the slices in various $z-$stacks, our method is able to generate cleaner images. We also compare our proposed approach with deconvolution operation used to denoise microscopy images using ImageJ [9]. Deconvolution is performed on 2D slices using Gaussian Point Spread Function (PSF) with standard values. It can be seen from Fig. 5 that our proposed approach reconstructs visually enhanced slices compared to deconvolution operation in ImageJ. A key reason that deconvolution doesn't work well is due to the selection of PSF which highly depends on the capturing instrument. This demonstrates the advantage of our approach with respect to existing algorithms. Note that in Fig. 3, Fig. 4, and Fig. 5, input slice is shown inside ☐ and the reconstructed slice using the proposed approach is shown inside ⬚. Results are best viewed when zoomed-in.

Qualitative Ablation. To evaluate the impact of quantization in the latent space, we perform an experiment without applying quantization keeping all other parameters same in the proposed method. Fig. 4 qualitatively shows the contribution on quantization in latent space. The image generated without quantization is less sharp than with quantization. This is due to inherent property of mean square loss to produce blurry results which dominates the reconstruction in absence of latent representation quantization loss.

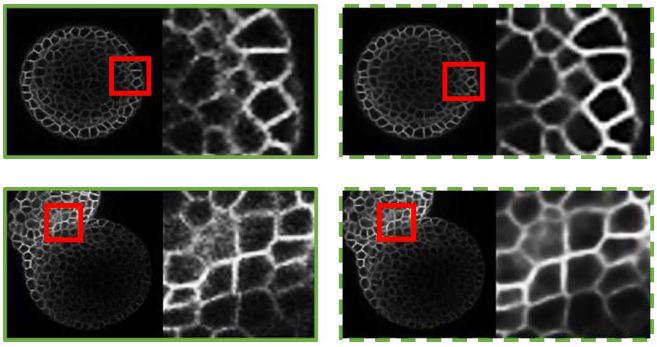

Fig. 3: **Qualitative Results of Proposed Method.** Original image (*left*) and Reconstructed image (*right*) with corresponding zoomed parts are presented here. The proposed method is able to generate sharper images from the given blurry image slices.

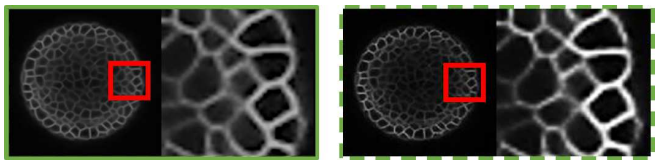

Fig. 4: **Reconstruction Results without Quantization.** Reconstructed image without quantization (*left*) and Reconstructed image (*right*) with quantization with corresponding zoomed parts are presented here. This demonstrates that the quantization module in our proposed approach is effective in deblurring the data.

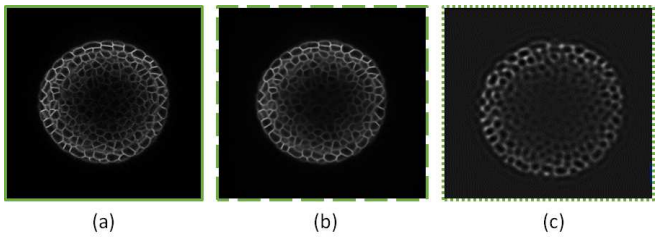

Fig. 5: **Comparison of Reconstructed Results with ImageJ** [9]. **(a)** Original Image, **(b)** Reconstructed using DQLR (**ours**) and **(c)** Reconstructed using deconvolution by ImageJ.

5. CONCLUSION

Micro-imaging data collected for various bio-medical research suffers from inherent blurriness and using this data for further analysis is a challenging task. We present an approach for enhanced reconstruction of microscopic sequential data by leveraging the information from consecutive image slices and using quantization of their latent representation to alleviate blurriness. Our data driven approach demonstrates visually superior results on a publicly available benchmark. The proposed approach would be useful for bio-medical researchers to enhance images where data is scarce and consequently, avoid unwanted laborious efforts for re-imaging the data.

Acknowledgement. We thank Prof. B.S. Manjunath from University of California, Santa Barbara for valuable discussions and helpful suggestions.

References

[1] Aaron van den Oord, Oriol Vinyals, et al., "Neural discrete representation learning," in *Advances in Neural Information Processing Systems*, 2017, pp. 6306–6315.

[2] Min Liu, Anirban Chakraborty, et al., "Adaptive cell segmentation and tracking for volumetric confocal microscopy images of a developing plant meristem," *Molecular Plant*, vol. 4, no. 5, pp. 922–931, 2011.

[3] Jürgen Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85–117, 2015.

[4] Jinshan Tang, Eli Peli, and Scott Acton, "Image enhancement using a contrast measure in the compressed domain," *IEEE Signal Processing Letters*, vol. 10, no. 10, pp. 289–292, 2003.

[5] Yan Wu, Mihaela Rosca, and Timothy Lillicrap, "Deep compressed sensing," in *International Conference on Machine Learning*, 2019, pp. 6850–6860.

[6] Alireza Makhzani, Jonathon Shlens, Navdeep Jaitly, and Ian J. Goodfellow, ," in *Adversarial Autoencoders*, 2015, vol. abs/1511.05644.

[7] Hang Zhao, Orazio Gallo, Iuri Frosio, and Jan Kautz, "Loss functions for image restoration with neural networks," *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 47–57, 2016.

[8] Lisa Willis et al., "Cell size and growth regulation in the arabidopsis thaliana apical stem cell niche," *Proceedings of the National Academy of Sciences*, pp. E8238–E8246, 2016.

[9] Tony J Collins, "ImageJ for microscopy," *Biotechniques*, vol. 43, no. S1, pp. S25–S30, 2007.