Tackling the Nonlinearity Problem in Inverse Modeling: Mixture Density Network-Backed Quantized AutoEncoder

Muhammed Nur Talha Kilic, Yuwei Mao, Vishu Gupta, Alok Choudhary, Wei-keng Liao, Ankit Agrawal
Northwestern University, Evanston, IL, USA
{talha.kilic, yuweimao2019, vishugupta2020}@u.northwestern.edu, {choudhar, wkliao, ankitag}@eecs.northwestern.edu

***Abstract*—Generative models have been widely used in the field of computer vision due to their ability to produce unseen data points. Its application has proven to be useful in various scientific domains such as materials science for generating new microstructure images that require learning nonlinear and one-to-many property-microstructure relationships. However, existing simulation-based solutions for this application are inefficient and time-consuming. Moreover, nonlinearity from the lower to higher dimensions poses considerable challenges. In this work, we propose a novel Mixture Density Network (MDN) based Quantized Autoencoder Network designed to generate microstructure images from only a single data point by establishing one-to-many nonlinear relationships from property to microstructure. Once the autoencoder effectively compresses spatial information into the property domain, we use the combination of produced latent vectors and property values to create a supplementary dataset for MDN. Upon completion of the model training, generative structures are extracted and merged to create a framework that generates images based on target property values (i.e., absorption values in this study). The trained MDN demonstrates proficiency in generating latent vectors within distribution, while the proposed Vector Quantized Variational Autoencoder (VQ-VAE) efficiently maps the embedding table to the latent space, generating images from property values within the range of properties observed during training. We demonstrate that our proposed model consistently outperforms the baselines with respect to generating new microstructure images having target properties and overcoming the above-mentioned challenges.**

***Index Terms*—Generative Networks, Autoencoder, Inverse Modeling, Microstructure Images**

I. Introduction

Artificial intelligence has become widely recognized for its capacity to comprehend nonlinear relations and employ generative models, including Autoencoders, Generative Adversarial Networks (GANs), and Diffusion Models [1]–[3]. These models are engineered to produce new data points by leveraging the learned data distribution during the training phase, allowing for the creation of higher-dimensional outputs.

In this study, we present our proposed model focused on the generation of microstructure images. Our goal is to generate images not from a random value, such as a diffusion models or GANs, but rather from the optical absorption value, which is a property of the microstructure images, as shown in Fig. 1. However, one of the challenges here is the high sensitivity encountered during the creation of a highly accurate model. An illustrative example, as depicted in Fig. 1 using the dataset from this research, reveals that even a small change in the microstructure image may result in significant change in the property value. Despite a 99.73% similarity between images calculated by pixel difference, the measured Residual Error Percentage (REP) value, used as a performance metric in this research, is 3.07%, indicating a notable deviation.

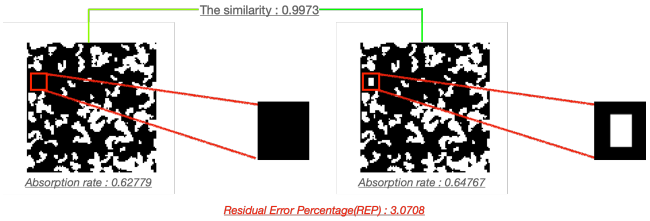

Fig. 1. Comparison of absorption values for microstructure images that appear visually similar, but have significant absorption value differences.

Contributions of this study are as follows:

- We demonstrate that the proposed model can generate realistic microstructure images.
- Our model overcomes challenges such as overfitting, compression loss, blurriness, and overly generalized outputs.
- Unlike simulations, it enables the rapid generation of images with the desired properties in a very short time.

II. Related Work

Autoencoders: Autoencoders stand out among generative models due to their ability to compress input data and extract information from it. Notably, the Variational Autoencoder (VAE) [5] is renowned for its capacity to extract concise information.

Diffusion Models: This popular generative model aims to generate images from reverse Markov transitions and has also started to be used in the field of materials science [6]. However, the linearity inherent of this structure implies that it may struggle to capture the nonlinear relationships present in our study.

GANs: Among the related works, this model structure stands out as the most prominent and is highly popular in materials science [7]–[9] but the training instability and model collapse associated with GAN models have prompted researchers to explore different models.

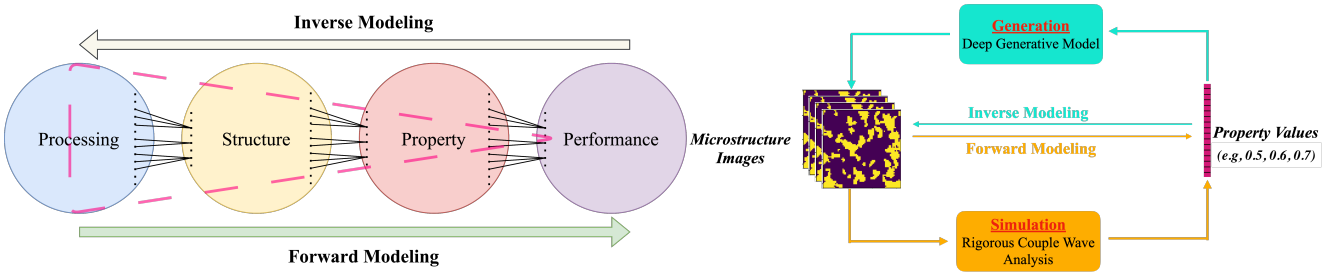

Fig. 2. Processing, Structure, Property, Performance (PSPP) Relationship [4].

PCA-MDN Method: This method utilizes the traditional dimension reduction technique, Principal Component Analysis (PCA), to perform an inverse transform and generate microstructure images [10]. The input data it receives is a reduced set of principal components created by MDN [11].

Optimization-Based Inverse Modeling: The inverse modeling method, based on optimization, involves sampling 250 pairs of design representations [11]. Subsequently, metamodel-based Bayesian optimization is employed to iteratively explore the next potentially optimal design point [12].

III. Background: Nonlinearity and Inverse Modeling

During the course of scientific evolution, four foundational paradigms have historically guided research endeavors: empirical, theoretical, computational, and data-driven [4], [13]. In the current epoch, we observe the significant impact of data-driven science, signaling the rise of the fourth paradigm in scientific history. In this realm, key concepts are built upon four fundamental phenomena known as Processing, Structure, Property, and Performance (PSPP) principles [4].

Fig. 2 visually represents forward modeling, where the sequence from left to right represents a logical flow of cause-effect relationships from processing to structure to property to performance. This emphasizes the predictability of output values in forward modeling, provided the mappings are well-established. However, establishing a connection in the reverse direction—from performance to processing, also known as inverse modeling—poses a significant challenge due to the nonlinearity [14]. Nonlinearity is a key characteristic of AI models. It enables these systems to learn from data in ways that linear models cannot, allowing them to handle a wide range of complex tasks [15]. However, it also introduces challenges that necessitate careful consideration in model design and training. Additionally, using lower-dimensional input data to generate higher-dimensional outputs, one-to-many relationships, can lead to diversity issues because vital values in high-dimensional data are derived from less informative low-dimensional data [16].

IV. Methods

Within this section, we provide an explanation of the deep learning models used in this study.

A. Vector Quantized Variational Autoencoder

Autoencoders comprise three essential components: an encoder, a latent space, and a decoder. The encoder network parameterizes a posterior distribution $q(z|x)$, representing the latent space z given the input data. The latent space includes a prior distribution $p(z)$ and a decoder operates by processing this latent space as $p(x|z)$. Initially, we explored the development of a Variational Autoencoder (VAE) to establish a latent space comprising μ, σ, and e. Due to the posterior collapse [17], we chose to use VQ-VAE as a foundational method [18]. In VQ-VAE, the encoder's output becomes discrete, and the decoder uses these discrete values instead of the original encoded output [5]. The process initiates by generating discrete representations through the transformation of the latent space. The latent space is composed of embedding tables, denoted as e_i, which exist in R^D. Here, k represents the size of the discrete latent space, as defined by the user. This implies that each D-dimensional array ($e_i \in R^D$, where $i \in 1, 2,..., k$) originating from the encoder will be assigned to one of the k classes.

VQ-VAE model incorporates three loss components: the reconstruction loss (notated as R), the VQ loss (notated as V), and the commitment loss (notated as C). Additionally, we propose a new component A, which corresponds to the absorption loss in this study to minimize the discrepancy in properties between the generated output and the desired attributes. The normalization parameters are denoted as α, β, and θ.

$$Total\ Loss = \alpha * R + \beta * A + \theta * V + C \tag{1}$$

The initial loss we consider is the reconstruction loss, measuring the reconstruction error using the L2 norm, representing the disparity between the model's output and the input image.

$$R : reconstruction\ loss = \log p(\hat{z}|z_q(\hat{z})) \tag{2}$$

While updating the embedding table, it is advisable to use exponential moving averages (EMA) [18]. The loss is characterized as the vq loss, computed through the L2 norm. This process guides the adjustment of e towards the encoder output $z_e(\hat{z})$.

$$V : vq\ loss = ||sg[z_e(\hat{z})] - e||_2^2 \tag{3}$$

The commitment loss employs the stop gradient operator (sg) to guarantee the embeddings' commitment to the nearest codebook vector [18].

$$C : commitment\ loss = \beta_c||z_e(\hat{z}) - sg[e]||_2^2, \quad (4)$$

where β_c is a scale factor of the commitment loss.

To enhance the resilience of our proposed framework, we have introduced a novel term known as the *"absorption loss"*, denoted as A. This loss is implemented on a randomly chosen subset of the training data (x) with the generated output ($\hat{x}$), specifically ($n\_batch/32$) samples per epoch.

$$A : \text{absorption loss} = \sum_{k=0}^{n} |\hat{x}_j - x_j|,\ j \sim rand(),\ n = \frac{n_{\text{batch}}}{32} \quad (5)$$

B. Mixture Density Network

The concept of the Mixture Density Network (MDN) was introduced by Bishop in 1994 [19]. The primary objective of this network is to generate not just a singular output but rather a distribution. In this network, each output is characterized by both mean and variance, collectively forming a part of the output distribution. This network produces individual outputs, μ and σ. The accumulation of these values, each multiplied by varying coefficients, π, forms the closest approximation to a Gaussian distribution [20].

$$P(\hat{y}|x) = \sum_{k=1}^{K} \pi_k(x) * \theta(\hat{y}|\mu_k(x), \sigma_k^2(x)) \quad (6)$$

$$J(w) = \sum_{n=1}^{N} -ln \sum_{k=1}^{K} \pi_k(x_n, w) * \theta(\hat{y}|\mu_k(x_n, w), \sigma_k^2(x_n, w)) \quad (7)$$

The symbol θ, in (6) and (7), denotes the conditional density. The loss equation, in (7), employed for model learning during the training phase is provided, where the inputs are represented by x, featuring associated weights w, and K indicates the number of parameters for one of the three partitions (μ, σ, and π) with N samples generated.

C. End-to-End Pipeline

In this section, we explore the integration of both generative models. In the initial phase, the training of the VQ-VAE model focuses exclusively on microstructure images. The primary metric for evaluating the VQ-VAE model revolves around examining the network's capability to reconstruct input images closely. Upon concluding the training phase of the proposed VQ-VAE, the subsequent algorithm incorporates mixture density networks. Dataset images, absorption values, and latent space information from the proposed VQ-VAE is stored for the mixture density models. In the training process of the mixture density model, absorption values are utilized, predicting the latent distribution shown by 30 potential latent vectors. The end-to-end framework outlines the procedures for generating images from a given property value. This involves combining two trained generative models by deactivating the regularizer from the MDN and disabling the encoder component of the VQ-VAE.

TABLE I
THE PERFORMANCE OF THE PROPOSED FRAMEWORK, GAN-MDN METHOD, PCA-MDN METHOD, MDN-BASED DEEP LEARNING INVERSE METHOD, AND OPTIMIZATION BASED INVERSE MODEL

Absorption Value	Residual Error Percentage (REP %)		
(Target)	Min	Average	Standard Deviation
The Proposed Framework			
0.55	**0.07%**	6.17%	4.30%
0.60	**0.36%**	8.40%	6.04%
0.65	**0.10%**	8.21%	5.11%
0.70	**0.03%**	5.45%	4.03%
0.75	**0.32%**	6.84%	3.30%
GAN-MDN Method			
0.55	1.25%	16.19%	8.96%
0.60	0.70%	10.99%	7.93%
0.65	0.18%	7.65%	5.64%
0.70	0.10%	5.00%	4.61%
0.75	0.43%	6.18%	3.51%
PCA-MDN Method			
0.55	4.96%	11.74%	3.05%
0.60	0.07%	2.69%	2.18%
0.65	3.71%	8.79%	2.59%
0.70	0.10%	3.41%	2.44%
0.75	3.17%	6.27%	1.52%
MDN Based Deep Learning Inverse Model			
0.55	2.85%	12.78%	3.89%
0.60	7.87%	14.95%	3.56%
0.65	11.00%	17.33%	2.63%
0.70	3.03%	15.62%	4.09%
0.75	8.44%	12.73%	3.20%
Optimization Based Inverse Model			
0.55	15.51%	-	-
0.60	-	-	-
0.65	1.21%	-	-
0.70	-	-	-
0.75	-	-	-

Note: Bold values indicate the best values of the primary evaluation metric (min REP).

V. EXPERIMENTS

This section provides the dataset information and experimental results.

A. Dataset

The dataset used in this research were synthetically generated using the Gaussian random field (GRF) method [21]. The GRF method relies on three parameters—mean, standard deviation, and volume fraction—and employs a simulation technique. Additionally, the optical absorption property of the microstructure images is simulated using the rigorous coupled wave analysis [22]. The dataset comprises 23,619 microstructure images, each sized at 64x64 pixels. This dataset is accompanied by corresponding optical absorption values spanning from 0.3 to 0.8, with the majority falling between 0.55 and 0.75.

B. Evaluation Metric

We employ the Residual Error Percentage, as indicated by (8). This metric offers a percentage-based measure for comparing the actual outcomes with the target results.

$$\text{Residual Error Percentage (REP \%)} = \frac{|x-\hat{x}|}{x} \times 100 \quad (8)$$

where x represents the target optical absorption value and $\hat{x}$ is the generated microstructure image's optical absorption value.

C. Experimental Results

Initially, the improved VQ-VAE model with absorption loss underwent unsupervised training. Following multiple iterations with our dataset, we identified the most suitable parameters: a batch size of 128, 1500 epochs, a learning rate of 3e-4, α set at 0.1, β at 1.2, and θ at 1.2. Afterward, we trained the Mixture Density Network (MDN) by utilizing optical absorption data as the input and generating a latent space as the output, which was then fed into the autoencoder model. Our MDN model was set up with 40 Gaussian mixture parameters and 30 samples.

Tables I provides a comparative analysis of the proposed framework against existing methods [23]. Our examination involves five targets, each characterized by distinct absorption values within the 0.55 to 0.75 range. These targets are distributed across the dataset, displaying absorption values of 0.55, 0.60, 0.65, 0.70, and 0.75.

VI. CONCLUSION

In this work, we developed a novel generative framework that integrates an improvised MDN and VQ-VAE for inverse modeling. Our results illustrate that our proposed model effectively extracts valuable insights from the provided data, enabling the generation of new images based on specified optical absorption values. Leveraging the MDN approach, our solution produces a distribution of outputs, facilitating the creation of novel images while preserving underlying connections. The experimental results show that the proposed method performs better than the existing solutions. Our study shows that the suggested framework facilitates a smoother shift from intensive mathematical simulations to lighter AI models.

In the future, we plan to explore more informative datasets for generalizability. We also plan to unify the independently trained networks to minimize the connection loss.

ACKNOWLEDGMENT

The Rigorous Couple Wave Analysis simulation is supported by Prof. Cheng Sun's lab at Northwestern University. This work was performed under the following financial assistance award 70NANB19H005 from U.S. Department of Commerce, National Institute of Standards and Technology as part of the Center for Hierarchical Materials Design (CHiMaD). Partial support is also acknowledged from NSF awards CMMI-2053929, OAC-2331329, DOE award DE-SC0021399, and Northwestern Center for Nanocombinatorics.

REFERENCES

[1] D. Bank, N. Koenigstein, and R. Giryes, "Autoencoders," *Machine Learning for Data Science Handbook: Data Mining and Knowledge Discovery Handbook*, pp. 353–374, 2023.
[2] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.
[3] J. G. Wijmans and R. W. Baker, "The solution-diffusion model: a review," *Journal of membrane science*, vol. 107, no. 1-2, pp. 1–21, 1995.
[4] A. Agrawal and A. Choudhary, "Perspective: Materials informatics and big data: Realization of the "fourth paradigm" of science in materials science," *Apl Materials*, vol. 4, no. 5, 2016.
[5] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.
[6] E. Azqadan, H. Jahed, and A. Arami, "Predictive microstructure image generation using denoising diffusion probabilistic models," *Acta Materialia*, vol. 261, p. 119406, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S135964542300736X
[7] A. Henkes and H. Wessels, "Three-dimensional microstructure generation using generative adversarial neural networks in the context of continuum micromechanics," *Computer Methods in Applied Mechanics and Engineering*, vol. 400, p. 115497, Oct. 2022. [Online]. Available: http://dx.doi.org/10.1016/j.cma.2022.115497
[8] Z. Yang, X. Li, L. Catherine Brinson, A. N. Choudhary, W. Chen, and A. Agrawal, "Microstructural materials design via deep adversarial learning methodology," *Journal of Mechanical Design*, vol. 140, no. 11, Oct. 2018. [Online]. Available: http://dx.doi.org/10.1115/1.4041371
[9] A. Iyer, B. Dey, A. Dasgupta, W. Chen, and A. Chakraborty, "A conditional generative model for predicting material microstructures from processing methods," 2019.
[10] A. Maćkiewicz and W. Ratajczak, "Principal components analysis (pca)," *Computers & Geosciences*, vol. 19, no. 3, pp. 303–342, 1993.
[11] Z. Yang, D. Jha, A. Paul, W.-k. Liao, A. Choudhary, and A. Agrawal, "A general framework combining generative adversarial networks and mixture density networks for inverse modeling in microstructural materials design," *arXiv preprint arXiv:2101.10553*, 2021.
[12] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. De Freitas, "Taking the human out of the loop: A review of bayesian optimization," *Proceedings of the IEEE*, vol. 104, no. 1, pp. 148–175, 2015.
[13] A. Agrawal and A. Choudhary, "Deep materials informatics: Applications of deep learning in materials science," *Mrs Communications*, vol. 9, no. 3, pp. 779–792, 2019.
[14] R. A. Ketcham, "Forward and inverse modeling of low-temperature thermochronometry data," *Reviews in mineralogy and geochemistry*, vol. 58, no. 1, pp. 275–314, 2005.
[15] K. Hornik, "Approximation capabilities of multilayer feedforward networks," *Neural networks*, vol. 4, no. 2, pp. 251–257, 1991.
[16] L. Song, Y. Wang, Y. Han, X. Zhao, B. Liu, and X. Li, "C-brain: A deep learning accelerator that tames the diversity of cnns through adaptive data-level parallelization," in *Proceedings of the 53rd Annual Design Automation Conference*, 2016, pp. 1–6.
[17] J. Lucas, G. Tucker, R. Grosse, and M. Norouzi, "Understanding posterior collapse in generative latent variable models," 2019.
[18] A. van den Oord, O. Vinyals, and K. Kavukcuoglu, "Neural discrete representation learning," 2018.
[19] C. M. Bishop, "Mixture density networks," 1994.
[20] C. B. Do, "The multivariate gaussian distribution," *Section Notes, Lecture on Machine Learning, CS*, vol. 229, 2008.
[21] Y. Liu, J. Li, S. Sun, and B. Yu, "Advances in gaussian random field generation: a review," *Computational Geosciences*, vol. 23, pp. 1011–1047, 2019.
[22] S. Yu, Y. Zhang, C. Wang, W.-k. Lee, B. Dong, T. W. Odom, C. Sun, and W. Chen, "Characterization and design of functional quasi-random nanostructured materials using spectral density function," *Journal of Mechanical Design*, vol. 139, no. 7, p. 071401, 2017.
[23] Y. Mao, Z. Yang, D. Jha, A. Paul, W.-k. Liao, A. Choudhary, and A. Agrawal, "Generative adversarial networks and mixture density networks-based inverse modeling for microstructural materials design," *Integrating Materials and Manufacturing Innovation*, vol. 11, no. 4, pp. 637–647, 2022.