Siamese-SR: A Siamese Super-Resolution Model for Boosting Resolution of Digital Rock Images for Improved Petrophysical Property Estimation

Vishal R. Ahuja, Utkarsh Gupta, Shivani R. Rapole, Nishank Saxena, Ronny Hofmann, Ruarri J. Day-Stirrat, Jaya Prakash, and Phaneendra K. Yalavarthy, *Senior Member, IEEE*

***Abstract*— Digital Rock Physics leverages advances in digital image acquisition and analysis techniques to create 3D digital images of rock samples, which are used for computational modeling and simulations to predict petrophysical properties of interest. However, the accuracy of the predictions is crucially dependent on the quality of the digital images, which is currently limited by the resolution of the micro-CT scanning technology. We have proposed a novel Deep Learning based Super-Resolution model called Siamese-SR to digitally boost the resolution of Digital Rock images whilst retaining the texture and providing optimal de-noising. The Siamese-SR model consists of a generator which is adversarially trained with a relativistic and a siamese discriminator utilizing Materials In Context (MINC) loss estimator. This model has been demonstrated to improve the resolution of sandstone rock images acquired using micro-CT scanning by a factor of 2. Another key highlight of our work is that for the evaluation of the super-resolution performance, we propose to move away from image-based metrics such as Structural Similarity (SSIM) and Peak Signal to Noise Ratio (PSNR) because they do not correlate well with expert geological and petrophysical evaluations. Instead, we propose to subject the super-resolved images to the next step in the Digital Rock workflow to calculate a crucial petrophysical property of interest, *viz.* porosity and use it as a metric for evaluation of our proposed Siamese-SR model against several other existing super-resolution methods like SRGAN, ESRGAN, EDSR and SPSR. Furthermore, we also use Local Attribution Maps to show how our proposed Siamese-SR model focuses optimally on edge-semantics, which is what leads to improvement in the image-based porosity prediction, the permeability prediction from Multiple Relaxation Time Lattice Boltzmann Method (MRTLBM) flow simulations as well as the prediction of other petrophysical properties of interest derived from Mercury Injection Capillary Pressure (MICP) simulations.**

***Index Terms*— Image super-resolution, deep learning, generative adversarial networks, siamese networks, digital rock physics, petrophysics, geology, micro computed tomography, local attribution maps.**

Manuscript received September 3, 2021; revised March 24, 2022; accepted April 18, 2022. Date of publication May 9, 2022; date of current version May 18, 2022. This work was supported by Shell. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Lisimachos Paul Kondi. *(Vishal R. Ahuja, Utkarsh Gupta, and Shivani R. Rapole contributed equally to this work.) (Corresponding author: Vishal R. Ahuja.)*

Vishal R. Ahuja is with Shell India Markets Private Ltd., Shell Technology Centre Bangalore, Mahadeva Kodigehalli, Bengaluru, Karnataka 562149, India (e-mail: vishal.ahuja@shell.com).

Utkarsh Gupta, Shivani R. Rapole, and Phaneendra K. Yalavarthy are with the Department of Computational and Data Sciences, Indian Institute of Science, Bengaluru, Karnataka 560012, India (e-mail: utkarshgupta@iisc.ac.in; shivanir@iisc.ac.in; yalavarthy@iisc.ac.in).

Nishank Saxena, Ronny Hofmann, and Ruarri J. Day-Stirrat are with Shell International Exploration and Production Inc., Shell Technology Center Houston, Houston, TX 77082 USA (e-mail: nishank.saxena@shell.com; ronny.hofmann@shell.com; ruarri.day-stirrat@shell.com).

Jaya Prakash is with the Department of Instrumentation and Applied Physics, Indian Institute of Science, Bengaluru, Karnataka 560012, India (e-mail: jayap@iisc.ac.in).

This article has supplementary downloadable material available at https://doi.org/10.1109/TIP.2022.3172211, provided by the authors.

Digital Object Identifier 10.1109/TIP.2022.3172211

I. Introduction

DIGITAL Rock Physics [1] is an emerging framework utilizing advances in imaging technologies and state-of-the-art image processing algorithms to construct digital models of reservoir rocks, which are used to run physics simulations to calculate several petrophysical properties of interest. The accuracy of the digital rock workflow relies crucially on the resolution of the acquired images [2], which is currently limited by the hardware of the micro-CT scanning technology. Super-resolution methods can effectively address this limitation by digitally boosting the resolution of images acquired using micro-CT scanners. Furthermore, for heterogeneous rock samples with a large Representative Elementary Volume (REV), significant acquisition time savings can be achieved by scanning the required large field of view with a lower resolution followed by boosting the resolution digitally using super-resolution [3].

Image super-resolution involves improving the resolution characteristics of a low-resolution (LR) image to obtain an equivalent higher resolution (HR) image by using either classical approaches such as patch-based or dictionary-based learning or deep learning approaches which are based on neural networks. These super-resolution methods tend to perform much better than traditional interpolation methods like bilinear and bicubic interpolation, which compromise on the finer details in the image. Such super-resolution methods have been widely used in a number of different application areas such as optical imaging [4], medical imaging [5], satellite imaging [6], and many such areas. Recently, deep-learning

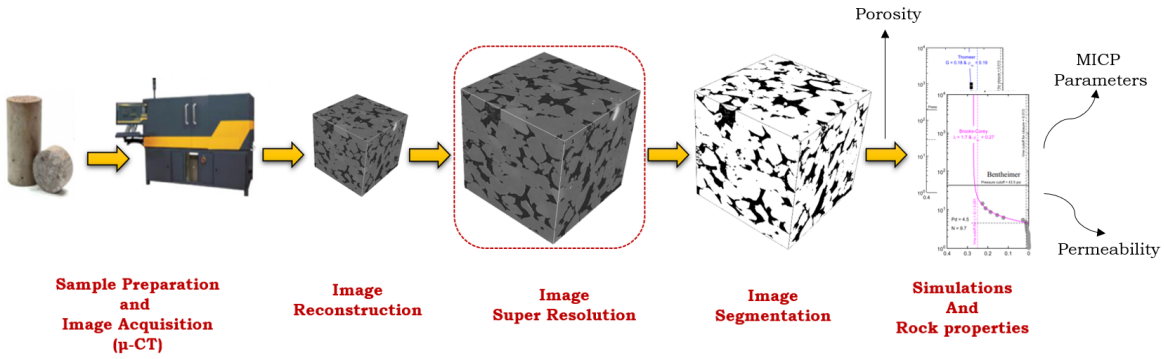

Fig. 1. Key steps involved in a typical digital rock workflow, with the role of super-resolution module highlighted in a red box and the prediction of different petrophysical properties of interest highlighted with black arrows.

based super-resolution techniques have been employed in Digital Rock Physics to improve the quality of the acquired micro-CT images [7], [8], the resolution of which was hitherto limited by the hardware of micro-CT scanning technology.

The key purpose of including the super-resolution module in the Digital Rock workflow shown in Fig. 1 is to have a good image quality with sufficient resolution to resolve the sub-resolution porosity, which remains hidden due to the limited resolution at which the images can be scanned using micro-CT scanning technology. Hence, a digital super-resolution method such as the one we propose (Siamese-SR) that not only enhances the image resolution but also retains and enhances semantic features of interest such as edges between rock and pores leads to an improvement in the petrophysical property prediction accuracy of the consequent steps in the Digital Rock workflow such as Mercury Injection Capillary Pressure (MICP) simulations which predict the capillary pressure curve and the Multiple Relaxation Time Lattice Boltzmann Method (MRTLBM) flow simulations which predict the permeability of the porous rock. Thus, the super-resolution module highlighted in the red box in Fig. 1 ensures high quality images are being supplied to the physics simulators for petrophysical property prediction. Hence, in this study, we have kept this ultimate goal of accurate petrophysical property estimation at the heart of the evaluation criteria for comparing the performance of different super-resolution methods.

Largely speaking, there are two classes of Deep Learning models which are typically used for super-resolution - Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs). The CNN based models started with the traditional Super-Resolution Convolution Neural Network (SR-CNN) [9] but later many improvements were developed such as the Deeply Recursive Convolutional Network (DRCN) [10], the Efficient Sub-Pixel Convolutional Neural Network (ESPCN) [11], the Enhanced Deep Super Resolution (EDSR) network [12] and the Wide Activation for Efficient and Accurate Image Super-Resolution (WDSR) [13]. While these CNN-based models are good at obtaining a high Peak Signal to Noise Ratio (PSNR), they tend to compromise on the high-frequency details like texture since they use generic optimization functions based on Mean Squared Error (MSE) in the image space which are defined based on the pixel-wise differences [14]–[16]. To tackle this issue and to improve the perceptual quality, Generative Adversarial Networks (GANs) have been employed to generate realistic textures. Since our aim is to boost the resolution of the Digital Rock images whilst retaining the original texture and high-frequency components, we have mainly focused our attention on GANs in this study. However, we do include one CNN based network, *viz.* EDSR network in our performance benchmarking for sake of completeness.

In this work, we have proposed a novel Deep Learning based Super-resolution architecture called Siamese-SR and successfully demonstrated that it leads to better performance as compared to some of the other existing super-resolution models described in Section II in terms of more accurate estimation of porosity and other such petrophysical properties of interest, the estimation of which is the ultimate aim of Digital Rock Physics. We have also proposed to move away from image based metrics like Structural Similarity (SSIM) and Peak Signal to Noise Ratio (PSNR) which do not correlate with expert geological and petrophysical evaluations. Instead of using such image based metrics, we have proposed to use porosity, which is an important petrophysical property of interest in the application domain, as a metric for evaluation of the different GAN based super-resolution methods. Besides using porosity as a metric for evaluation, we have also demonstrated that the super-resolution using Siamese-SR model not just yields better porosity but also leads to more accurate estimation of petrophysical properties of interest calculated using Mercury Injection Capillary Pressure (MICP) simulations and Multiple Relaxation Time Lattice Boltzmann Method (MRTLBM) flow simulations.

Thus, the key contributions of this work can be summarized as follows:

- Novel architecture called Siamese-SR (with Siamese Discriminator + MINC Loss Estimator) is proposed specifically for super-resolution of Digital Rock images.
- The network training is performed using pairs of separately acquired LR and HR images using different micro-CT scanning acquisition parameters as against using synthetically under-sampled HR images to generate LR images. Due to this reason, image based metrics like

structural similarity (SSIM) and PSNR are found to be sub-optimal to quantify improvements in super-resolution algorithms.
- Instead of using image based metrics like SSIM and PSNR, we propose to use an important petrophysical property from the application domain *i.e.* porosity as the quantitative metric to evaluate the performance of various SR models.
- We demonstrate the efficacy of our proposed super-resolution model for improved petrophysical property prediction by running physics simulations using the super-resolved images, *viz.* Multiple Relaxation Time Lattice Boltzmann Method (MRTLBM) flow simulations for permeability prediction and Mercury Injection Capillary Pressure (MICP) simulations for prediction of the Capillary Pressure curve from which other petrophysical properties of interest are derived.
- We also show that our proposed architecture optimally focuses on edge-semantics using Local Attribution Maps(LAM) generally used for evaluating the super-resolution network performance.

The paper is organized as follows: We first present a background of existing super-resolution methods in Section II. Then, in Section III, we describe the proposed Siamese-SR model, the image restoration loss functions used, the network architecture, the figures of merit used for evaluation and details about training and testing. We then present the results of applying the above super-resolution models and their evaluation using the quantitative metrics in Section IV. Then, in Section V, we provide a brief discussions on the obtained results. Finally, in section VI, we present the conclusions of this study and discuss the scope for future work in this area.

II. Existing Models

In this section, we mainly focus our attention on some of the widely used existing GAN based models for reasons explained in Section I. The models that are mentioned in this section are essentially used for performance benchmarking of our proposed Siamese-SR model.

Super-resolution Generative Adversarial Network **(SRGAN)** [17] incorporates an adversarial loss and a perceptual loss based on feature space along with the basic MSE loss. This perceptual loss between the generated image and the HR image is based on the Euclidean distance between their corresponding feature maps obtained from the pre-trained VGG19 network. The generator of SRGAN architecture comprises of a set of residual dense blocks stacked between an input convolutional layer and the final set of convolutional and pixel shuffler up-sampling layers. The skip connections among these dense blocks allow the gradients to directly flow across many layers and eliminate issues like vanishing or exploding gradients in deeper networks. The task of the generator is to learn to generate realistic textures. On the other hand, the discriminator is trained to distinguish between the generated super-resolved image and the high-resolution image treated as a pseudo ground truth.

Enhanced Super-Resolution Generative Adversarial Network **(ESRGAN)** [18] is an advanced architecture which enhances the key components of SRGAN to eliminate undesired artifacts. Instead of the normal residual blocks, ESRGAN uses residual-in-residual dense blocks that help in further improving the recovered image quality as compared to SRGAN. ESRGAN's residual-in-residual dense blocks have more skip connections compared to SRGAN's residual blocks, which is one of the major reasons for this improvement. Furthermore, the relativistic-GAN based discriminator [19] in ESRGAN predicts the relative realness of an image (real features of the HR image) which helps the generator to learn sharper edges than the standard discriminator which only predicts if the image is real or not. ESRGAN also introduces a new version of perceptual loss, which is based on the VGG network fine-tuned using the Materials in Context Database (MINC) loss [20]. This allows the perceptual loss to focus more on the material texture and thus facilitates the generator to learn realistic textures of images. Occasionally, the GAN-based methods generate undesired artefacts in the super resolved images like wrinkles on human face in case of natural images [18]. To remove this, ESRGAN and PSNR-based models can be trained individually and interpolation can be performed by combining the both the network parameters. Ref. [18] indicates that although introducing MINC loss provides only marginal improvement in the perceptual quality index, but it clearly states that MINC loss helps significantly in texture improvement. In digital rock images, we demonstrate both qualitatively and quantitatively that introducing MINC loss brings back the necessary texture present in the rock images. We call this model as **ESRGAN-MINC** in the paper.

Structure-preserving super resolution method (**SPSR**) [21], [22] is a recent super-resolution model that not only retains the merits of GAN based methods to generate perceptually pleasing features but also augments the architecture with a gradient branch and a gradient loss to alleviate the issue of geometric distortions commonly existing in the SR results of perceptual-driven methods.

It is important to note that in addition to the GAN based models mentioned in this section, we have also added one CNN based model, *viz.* **EDSR** network in our performance benchmarking for sake of completeness.

III. Materials & Method

In this section, we describe our proposed Siamese-SR model, the image restoration loss functions used, the network architecture, the figures of merit used for evaluation and the details about training and testing.

A. Proposed Model: Siamese-SR

The proposed Siamese-SR network, as shown in Fig. 2, consists of three blocks: the generator network $\tilde{\mathbf{G}}$, the relativistic discriminator $\tilde{\mathbf{D}}$ and the siamese discriminator $\tilde{\mathbf{S}}$. The generator network is a fully convolutional neural network that learns the mapping $\tilde{\mathbf{G}}$: LR $\mapsto$ HR. The relativistic discriminator learns to quantify the relative realness of the input image, whereas the siamese discriminator learns to discriminate whether the input image pair is a matching pair (HR, HR*) or a non-matching pair (SR, HR). The siamese discriminator is robust in ensuring

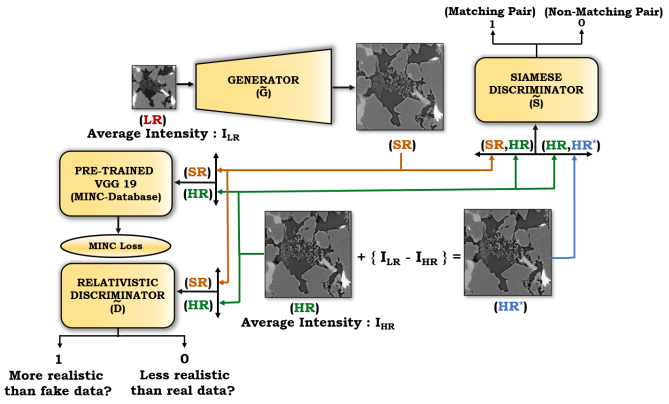

Fig. 2. Proposed Siamese-SR model architecture. A block diagram of entire SR module containing generator ($\tilde{\mathrm{G}}$), Relativistic discriminator ($\tilde{\mathrm{D}}$) and Siamese discriminator ($\tilde{\mathrm{S}}$).

that the generated SR image has semantic features similar to the corresponding HR counterpart. We also incorporate a variation of perceptual loss known as Material In Context (MINC) loss. The MINC loss helps to retain the original texture of the Digital Rock images as discussed further in Section V.

1) GAN Objective of Relativistic Discriminator: Unlike the normal discriminator which predicts whether a particular sample is real or not, a relativistic discriminator predicts the probability of relative realness of the data sample [19]. The relativistic discriminator receives feedback from both real and the generated data which in turn helps it guide the generator to generate sharper textures. Apart from this, a relativistic discriminator also stabilizes the GAN training [19]. The discriminator objective function while training is given by:

$$\max_{\tilde{D}} L_{Re}(\tilde{G},\tilde{D}) = \mathbf{E}_{HR}\left[\log\left(\tilde{\mathbf{D}}(HR,\tilde{\mathbf{G}}(LR))\right)\right] + \mathbf{E}_{LR}\left[\log\left(1-\tilde{\mathbf{D}}(\tilde{\mathbf{G}}(LR),HR)\right)\right]. \quad (1)$$

The generator objective function is given by:

$$\max_{\tilde{G}} L_{Re}(\tilde{G},\tilde{D}) = \mathbf{E}_{HR}\left[\log\left(1-\tilde{\mathbf{D}}(HR,\tilde{\mathbf{G}}(LR))\right)\right] + \mathbf{E}_{LR}\left[\log\left(\tilde{\mathbf{D}}(\tilde{\mathbf{G}}(LR),HR)\right)\right], \quad (2)$$

where $\tilde{D}(x_1, x_2)$ is mathematically represented as follows:

$$\tilde{D}(x_1,x_2) = \sigma(\mathbf{C}(x_1) - \mathbf{E}_{LR}[(\mathbf{C}(x_1)]) \ . \quad (3)$$

Here, E denotes the expectation operator over real and generated data, $\tilde{\mathrm{D}}$ denotes the relativistic discriminator network, and $\tilde{\mathrm{G}}$ denotes the generator network, $\sigma(\mathrm{x})$ is the sigmoid function, C(x) denotes a non-transformed output of the discriminator when the input is x and $\mathrm{E_{LR}}$ denotes expectation over all the generated samples which in our case is $\mathrm{G}(\tilde{LR})$. The first term in Eq.1 represents the logarithmic probability of the real data being more realistic than the generated data while the second term represents the logarithmic probability of whether the generated data is less realistic than the real data. Networks $\tilde{\mathrm{G}}$ and $\tilde{\mathrm{D}}$ are trained alternately by fixing the network parameters of one of the them and then updating the network parameters of the other.

2) GAN Objective of Siamese Discriminator: Since the siamese discriminator is based on the siamese twin network [23], [24], it takes an image pair as input as shown in Fig. 3(a) instead of taking a single image as input. If the input image pair is a matching image pair (HR, HR*), then it results in embeddings for the two images which are closer in the embedding space than the case where the input image pair is a non-matching pair (HR, SR) as shown in Fig. 3(b). The task of the siamese discriminator is to assign a high probability value to the matching input pair and a low probability value to the non-matching pair. While training a GAN with a siamese discriminator, we minimize the Jensen–Shannon (JS) divergence between matching and non-matching class distribution. When trained using the siamese discriminator, the role of the generator network is to produce an SR image which is as close as possible to the HR image so that the siamese discriminator assigns higher probability to the (HR, SR) input pair. Mathematically, the GAN objective for adversarial training using the siamese discriminator is stated as follows:

$$\min_{\tilde{G}}\max_{\tilde{D}} L_{Siamese}(\tilde{G},\tilde{S}) = \mathbf{E}_{HR}\left[\log\left(\tilde{\mathbf{S}}(HR,HR^*)\right)\right] + \mathbf{E}_{LR}\left[\log\left(1-\tilde{\mathbf{S}}(\tilde{\mathbf{G}}(LR),HR)\right)\right]. \quad (4)$$

Here, E denotes the expectation operator, $\tilde{\mathrm{S}}$ denotes the siamese discriminator network, and $\tilde{\mathrm{G}}$ denotes the generator network. The first term in Eq. 4 represents the logarithmic probability of the siamese discriminator predicting the (HR, HR*) pair as the matching sample pair. In contrast, the second term represents the logarithmic probability of the siamese discriminator predicting the $(\tilde{\mathrm{G}}(LR), HR)$ or (SR, HR) as a non-matching sample pair. Networks $\tilde{\mathrm{G}}$ and $\tilde{\mathrm{S}}$ are trained alternately by fixing the network parameters of one of them and then updating the network parameters of the other. $\tilde{\mathrm{S}}(\mathrm{x}_1, \mathrm{x}_2)$ is mathematically expanded as follows:

$$\tilde{\mathrm{S}}(x_1,x_2) = \sigma[\beta(\gamma(x_1),\gamma(x_2))] \ , \quad (5)$$

$$\beta(\mathrm{x}_1,\mathrm{x}_2) = \frac{\mathrm{x}_1.\mathrm{x}_2}{\max(\|\mathrm{x}_1\|,\|\mathrm{x}_2\|,\epsilon)} \ , \quad (6)$$

where $\sigma(\mathrm{x})$ represents the sigmoid function, $\beta(\mathrm{x}_1, \mathrm{x}_2)$ represents the cosine distance between x_1 and x_2. The value of ϵ is chosen as 10^{-8} to avoid division by zero while evaluating cosine distance. Here, $(\gamma(\mathrm{x}_1), \gamma(\mathrm{x}_2))$ represents the embeddings obtained from the convolutional block present in the siamese network for x_1 and x_2 respectively.

3) Significance of HR*: Ideally, a perfectly matching pair in our case should be (HR, HR), but the embeddings obtained from the convolutional block in the siamese network will always be the same for this image pair. In other words, irrespective of weights in the network, the output for (HR, HR) input image pair will remain the same, leading to inefficient learning. Hence, in order to avoid the inefficient learning, instead of passing (HR, HR) we pass (HR, HR*) where HR* is the slightly perturbed version of HR image, such that $\mathrm{HR}^* = \mathrm{HR} + \mathrm{I_{LR}} - \mathrm{I_{HR}}$, where $\mathrm{I_{LR}}$ and $\mathrm{I_{HR}}$ represent the average intensity of LR and HR images respectively. Note that this perturbation is not pixel specific i.e., we add a small perturbation in HR image which is the same across all pixels. Adding such type of perturbation changes the intensity

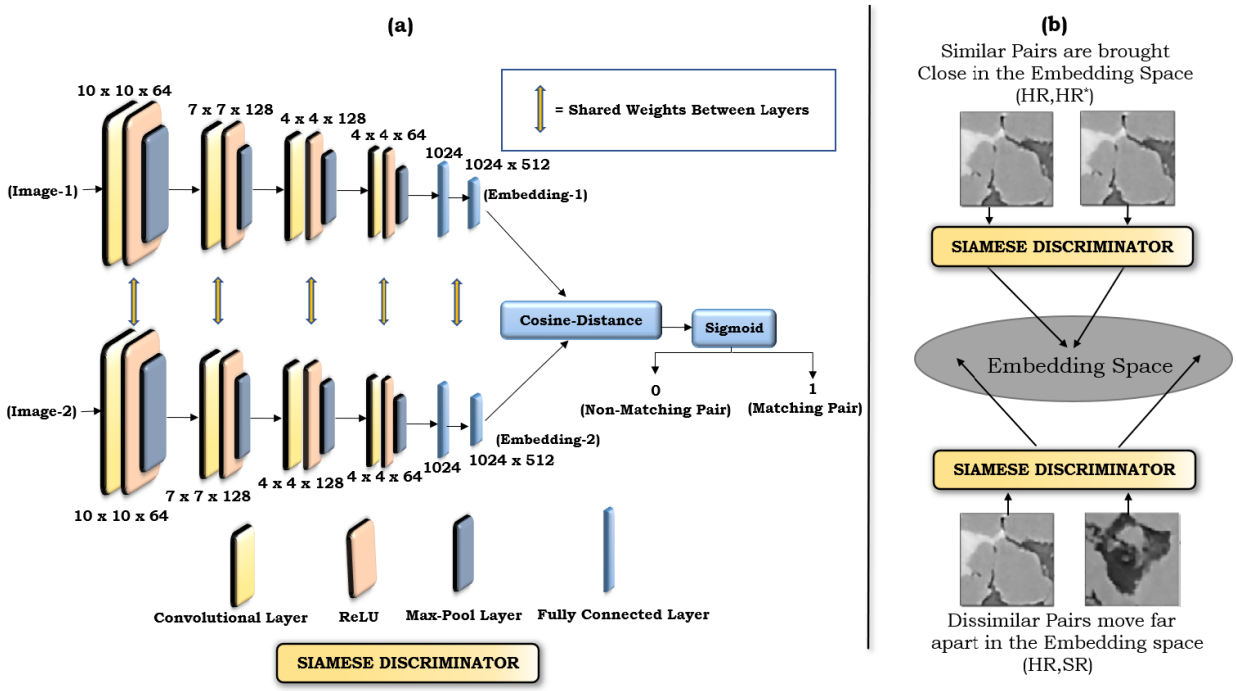

Fig. 3. (a) Basic structure of the siamese discriminator. (b) Demonstration of the working of siamese discriminator.

levels across all the pixels equally and does not introduce unnecessary high-frequency artifacts.

B. Image Restoration Losses

In this sub-section, we describe the loss functions used while training our network.

1) L_1*-Norm Based Loss Function:* The generic mean squared error loss function suffers from various limitations when used for image restoration problems. The L_2-norm based loss does not significantly match the image quality with what humans perceive [25]. Using L_1-norm based loss function instead of L_2-norm based loss function significantly reduces various image artifacts [26]. L_1-norm based cost function weighs the error differently compared to the L_2-norm based cost function; L_2-norm based loss over-weighs the error while back-propagating the error through the network layers. Mathematically, the L_1-norm based loss can be stated as follows:

$$\mathcal{L}^{\ell_1}(\mathrm{I}) = \frac{1}{\mathrm{N}} \sum_{\mathrm{i} \in \mathrm{I}} |\mathrm{x(i)} - \mathrm{y(i)}|, \tag{7}$$

where I indicates the image whose pixel indices are represented by i, x(i) represents the pixel value in the ground truth image and y(i) represents the pixel value in the predicted image.

2) Materials In Context (MINC) Loss Function: The original VGG network is trained on the ImageNet dataset, which typically contains natural images. However, since rock images are very different from natural images, we might not achieve the relevant texture in the output while using the ImageNet-trained VGG network for calculation of the perceptual loss. Instead of training the VGG network using the ImageNet database, we use the Material in Context Database (MINC) [20], which is widely used for material recognition tasks. This database consists of almost 3 million samples which belong to materials around us like stone, wood, ceramic, etc. So, MINC loss is also a perceptual loss calculated using the VGG-19 network which is trained on the MINC database. Mathematically, the MINC loss function can be expressed as follows:

$$\mathcal{L}^{\mathrm{MINC}}(\mathrm{I}, \mathrm{I}_{\mathrm{est}}) = \frac{1}{\mathrm{C} * \mathrm{W} * \mathrm{H}} ||\phi(\mathrm{I}) - \phi(\mathrm{I}_{\mathrm{est}})||_2, \tag{8}$$

where $\phi(.)$ represents the feature maps obtained from a pre-trained VGG-19 network on the MINC database. We use the 16^{th} convolutional layer of the VGG network to obtain the feature maps. Here C, W, and H represents the number of feature-maps, width, and height of the features maps.

Thus, the overall image restoration loss used in the Siamese-SR model is given by:

$$\mathcal{L}^{\mathrm{OIR}} = \alpha \mathcal{L}^{\mathrm{MINC}} + \beta \mathcal{L}^{\ell_1}, \tag{9}$$

where α and β are tunable weights. Detailed steps involved in training the Siamese-SR model are summarized in the algorithm described in the Supplementary Information file.

C. Network Architecture

The architecture of the proposed Siamese-SR model, as shown in Fig. 2, consists of the generator and the relativistic discriminator adapted from ESRGAN [18] with the addition of a siamese discriminator for improved super-resolution. The details of the three blocks are as follows:

1) Generator: The generator consists of an input convolutional layer, a stack of eight residual blocks with skip connections, a middle convolutional layer, a pixel shuffler-based upsampling block, and finally an output convolutional

layer. Instead of the normal residual blocks, the generator uses residual-in-residual dense blocks that help in further improving the recovered image quality. Residual scaling with a scale factor of 0.2 was used to improve the stability of the network [27], [28]. We observed that using more than eight residual blocks did not improve the performance and hence we have used a stack of eight residual blocks.

2) Relativistic Discriminator: The discriminator network has a chain of convolutional layers with a kernel size of 3×3, each having a LReLU (leaky ReLU) activation function. This chain is followed by two final linear layers with the same kernel size. The relativistic-GAN based discriminator [19] helps the generator to incorporate sharper edges than the standard discriminator which is only capable of producing an output of whether the image is real or not.

3) Siamese Discriminator: The siamese discriminator uses a siamese twin network [24] architecture, as shown in Fig. 3(a), consisting of series of convolutional layers which use filters of varying kernel sizes. We use 64 filters (kernel size:10×10) in the 1st convolutional layer followed by 128 filters (kernel size:7×7), 128 filters (kernel size:4×4) and 64 flters (kernel size:4×4) in the 2nd, 3rd and 4th convolutional layers respectively. Each of the four convolutional layers is followed by a ReLU activation and a MaxPool layer. The complete convolutional block is followed by a couple of fully connected layers whose output serves as the final embedding. This final embedding is then used to compute the cosine distance followed by application of a sigmoid activation function.

D. Figures of Merit

Traditionally, image based metrics such as Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM) are commonly used for quantitative evaluation of the performance of super-resolution models. However, it has been shown in the literature that PSNR is often a poor metric to judge an SR model performance as its purely based on pixel values and does not guarantee a direct correlation with a visually appealing image [29]. Particularly, PSNR favours overly smoothed images [30], which do not retain the original texture in the image. On the other hand, SSIM evaluates the quality of an image based on structures, luminance, and contrast. Hence, SSIM is a widely accepted and used metric to evaluate the perceptual quality of the images [31]. It should be noted however that most of the existing super-resolution studies [9]–[12], [17] are generally performed on synthetically generated LR images which are simply down-sampled versions of HR images. However, in our study, as shown in Fig. 4, we acquire LR volumes from a separate micro-CT scanning acquisition rather than merely down-sampling the HR volumes to generate the corresponding LR volume. Our approach of using different acquisitions for obtaining the HR and LR image pairs for training ensures that the model learns a robust mapping between an LR image and an HR image. However, since both the HR and LR images are acquired using separate micro-CT scanning acquisitions, issues like mismatch in contrast, luminance and improper image registration might occur, resulting in poor image quality assessment when using

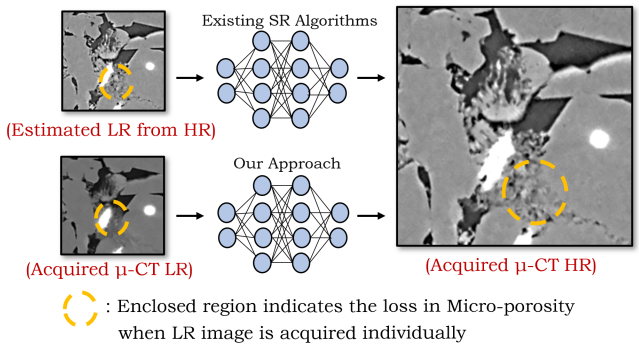

Fig. 4. Figure highlighting the difference between existing training methodologies for super-resolution using down-sampled HR images as LR images *vis-à-vis* our methodology of using different acquisitions for LR & HR images. The figure also highlights the difference in the way micro-porosity shows up as well as the difference in contrast and intensities when two separate acquisitions are used for LR & HR images.

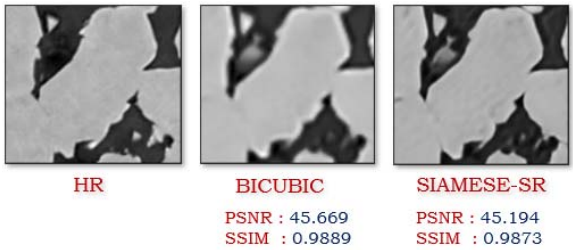

Fig. 5. Comparison of PSNR and SSIM for bicubic and super-resolved image using Siamese-SR. Bicubic image though is not perceptually good in comparison to SR image but still is given higher number for both PSNR and SSIM.

SSIM as a metric. It is also important to note that it is precisely these differences between the LR & HR images obtained from different acquisitions, because of which the additional siamese discriminator in the proposed Siamese-SR model helps in generating better super-resolution. To further illustrate with an example the misleading evaluations based on the PSNR and SSIM metrics, in Fig. 5, we show that both the PSNR and SSIM metrics for the bicubic interpolated LR image are higher compared to the SR image, even though SR has clearly a higher perceptual quality in comparison to bicubic interpolated LR image. Therefore, we propose to move away from these image based metrics and rather focus on the estimation of accurate petrophysical properties of interest as the metrics for evaluation.

Since the ultimate objective of the Digital Rock workflow is to estimate accurate petrophysical properties of interest, we propose to subject the super-resolved images to the consequent step in the workflow i.e. segmentation to calculate the porosity. Porosity is one of the most fundamental properties, the accurate estimation of which from CT images is absolutely necessary for the accuracy of the entire Digital Rock workflow [32]. Hence, the different SR models were compared based on the estimation of porosity from the super-resolved images rather than image based metrics like PSNR and SSIM. Furthermore, we also demonstrate with an example that once a more accurate porosity is obtained using super-resolution,

it also implies improvement in accuracy of the other petrophysical properties of interest that are dervied from the segmented image using Mercury Injection Capillary Pressure (MICP) simulations [33], [34] and Multiple Relaxation Time Lattice Boltzmann Method (MRTLBM) flow simulations [35], [36]. We now further describe image based porosity, MICP simulations and MRTLBM simulations below:

1) Image Based Porosity: A porous media broadly consists of two phases - the solid phase and the pore phase. The solid phase is made of various materials, while the pore phase is essentially the empty space in the media. Total porosity is defined as the ratio of total pore space volume to the total volume of the porous media. Various methodologies have been proposed in the literature to estimate porosity efficiently [37]–[40]. We estimate the porosity from segmented micro-CT rock volumes as described in Ref. [41]. We segment the super-resolved images using the spatial Fuzzy C Means (sFCM) [42] clustering algorithm in the similar fashion as presented in Ref. [41]. The sFCM segmentation algorithm requires selecting seed points (cluster centers) to initialize the segmentation process. We used the same co-ordinates for cluster centers across all the different super-resolution models as well as the HR image for consistency. For the LR image as well, we used the corresponding co-ordinates using the appropriate scaling factor between HR and LR. The way we used porosity as a metric for evaluation of the different super-resolution models is that we subject the reconstructed SR volumes as well as the HR volumes to sFCM segmentation and evaluate which of the super-resolution models leads to a porosity value closest to that obtained via the HR volume, which is our reference.

2) Mercury Injection Capillary Pressure (MICP) Simulation: The MICP simulation essentially simulates the experiment of injecting mercury into the rock sample at various applied pressures to estimate several petrophysical properties of interest. [33]. We have used the SatuDict module of the GeoDict 2020 software package [34] to run the MICP simulations and calculate the Capillary Pressure curves, from which we have derived the following petrophysical properties of interest: (a) the true porosity (φ_∞), which is a corrected version of the simple image-based porosity because it is the porosity that mercury would invade at infinite applied pressure, (b) the pore throat size distribution which can be captured quantitatively through the nature of the MICP curve represented by the pore geometrical factor G obtained by fitting the Thomeer equation [43], and (c) the displacement pressure (P_d) which is directly related to the dominant pore throat size. The MICP simulation demonstrates that the advantage of obtaining the super-resolved images is not just restricted in improving segmentation accuracy, but also in the subsequent steps based on physics simulations where petrophysical properties of interest are calculated.

3) Multiple Relaxation Time Lattice Boltzmann Method (MRTLBM) Flow Simulation: The MRTLBM method [35], which is an improvement over the Bhatnagar-Gross-Krook [44] method, is widely used to simulate single-phase flow of fluids through digital rock samples to calculate permeability, which is an important petrophysical property of interest. We have run the distributed parallel implementation [36] of the MRTLBM method on general-purpose graphics processing units, which is useful for the rapid and scalable computation of absolute permeability from high-resolution 3D micro-CT images.

TABLE I
DETAILS OF THE DATASETS UTILIZED IN THIS WORK

Rock	LR Size (Res)	HR Size (Res)	Train/Test Slices
B3	659 x 659 ($3\mu m$)	1318 x 1318 ($1.5\mu m$)	546/105
B2	534 x 534 ($2\mu m$)	1068 x 1068 ($1\mu m$)	531/103
A2	251 x 251 ($4\mu m$)	502 x 502 ($2\mu m$)	264/×
A1	299 x 299 ($4\mu m$)	598 x 598 ($2\mu m$)	264/×
SP2	340 x 340 ($5\mu m$)	680 x 680 ($2.5\mu m$)	352/×

E. Training & Testing Dataset

The dataset used in this study consists of micro-CT images of four sandstone rocks and a sphere pack. Most of the recent SR studies have used a simple down-sampling of HR images to generate the LR dataset. However, this can never entirely duplicate the original relationship between the LR and HR images. Such an approach might mislead the network to learn the reverse-down-sampling operation rather than capturing the complexities of the real differences between an LR & an HR image. All the samples have been acquired at low and high resolutions using the micro-CT scanner with a scale factor of 2x as shown in Fig. 4. Two challenges arise when using separately acquired LR and HR images for super-resolution in the Digital Rock workflow. Firstly, the HR images capture a smaller field-of-view than the corresponding LR images due to the micro CT hardware's inherent limitations [3]. Hence as a part of pre-processing, we register the LR and HR volumes and crop the LR image appropriately to match the field-of-views. The LR-HR image pairs are obtained by slicing both the volumes in the z-direction. Secondly, due to different scanning times, the LR and HR images may vary in their grayscale values at any particular location. A grayscale correction is performed by histogram matching of the LR and HR images. Although this does not mean that the corresponding grayscale values match exactly across LR and HR images, but the difference is minimized to a significant degree and the still remaining difference is handled well by the siamese discriminator. Details of the different rocks samples, the corresponding dimensions and resolution and the number of slices used for the train-test split for each rock are shown in Table I. We have restricted our testing to B2 and B3 rock samples because they have the best resolution i.e. $1 - 1.5\mu m$ for the high resolution images, which is the desired resolution range for the desired accuracy of the digital rock workflow and hence these two samples lend themselves to be the best testing dataset for clear differentiation between the different super-resolution methods for their performance in the digital rock workflow. However, to have a sufficiently large and diverse training dataset, we have used three other samples - two sandstone rocks (A1, A2) and a sphere pack (SP2) which are at a slightly

different resolution range but they have similar number of voxels for resolving the pore throats (referred to as the N value in the MICP simulations).

F. Training Details

Data augmentation operations like horizontal flip and vertical flip were performed on the training image pairs. For the training, the Adam optimizer was used and the β value was set from 1 to 0.9 for all the models. Its important to note that during the process of hyperparameter tuning and training, we observed that the discriminator network learns faster than the generator network. If the discriminator reaches its threshold, it would not propagate any non-zero gradients to the generator which inhibits the generator from learning any further. To handle this issue, we set the discriminator's learning rate to 1/10th of the generator's learning rate. After studying the results for a range of epochs, we found 100 epochs to be ideal. The hardware used for training was a workstation with an NVIDIA Quadro RTX 8000 48GB GPU. The typical training time for each of the deep learning model depends on the size of dataset, the depth of the network and the corresponding number of hyperparameters. With the current dataset and parameters, our Siamese-SR model requires about 5-6 hours to train with around 195 million trainable parameters.

IV. Results

In this section, we present the qualitative and quantitative performance benchmarking of our proposed Siamese-SR model against several different existing SR models (SRGAN, ESRGAN, ESRGAN-MINC, EDSR and SPSR). The qualitative comparison involves visual inspection of the images by the trained eye of a geologist and the quantitative comparison involves estimation of porosities from the super-resolved outputs of the different SR models as well as those of the acquired LR and HR volumes for reference. Furthermore, we present two ablation studies – one for the weights assigned to the different loss functions in our proposed Siamese-SR model and one for highlighting the impact of different loss functions that make up Siamese-SR. Another important point we discuss in this section is the evaluation of the different methods from an image-analysis point of view. As we already mentioned several reasons in Section III-D for why it is not possible to use a standard quantitative image-based metric like PSNR and SSIM, we explore a different approach here which although is qualitative but still is widely used for the evaluation of super-resolution methods - Local Attribution Maps (LAMs).

A. Qualitative Performance Benchmarking

For qualitative inspection, we show representative slices from each of the two rock volumes in the testing dataset - B3 (Fig. 6(I)) and B2 (Fig. 6(II)). The yellow arrows in this figure indicate regions of microporosity, which are the key regions that are not very well captured at low resolution leading to the lower porosity obtained from LR images, and hence these are the regions of interest that the super-resolution models must capture well. Furthermore, we zoom into certain regions represented with the green box and show the zoomed in version as an inset in the bottom left of the figures for a closer evaluation. To the trained eye of the expert geologists & petrophysicists, it was clear that the Siamese-SR model was indeed performing better than the other models but for a more easily tractable evaluation, we resort to quantiative performance benchmarking described below.

TABLE II

Comparison of Super-Resolution Methods in Terms of Estimated Porosity. Closest Porosity Values to the Actual HR Porosity Are Shown in Bold. We Also Include Percentage Error for the Estimated Porosity Considering HR Porosity as Reference

Model	B2 Porosity (p.u.) / %Error	B3 Porosity (p.u.) / %Error
LR	12.74 / -10.7	10.30 / -20.7
HR	14.27 / —	12.99 / —
SRGAN	16.04 / 12.4	15.39 / 18.5
ESRGAN	16.03 / 12.3	14.35 / 10.5
ESRGAN-MINC	15.97 / 11.9	13.82 / 6.4
EDSR	10.14 / -28.9	13.99 / 7.7
SPSR	10.78 / -24.5	14.51 / 11.7
Siamese-SR (Proposed)	**15.59 / 9.3**	**13.66 / 5.2**

B. Quantitative Performance Benchmarking

We estimated the porosity of all the rock volumes obtained using the different SR models using sFCM segmentation for quantitative comparison. We consider the estimated porosity of the HR volume as reference porosity and thus the model which estimates the closest porosity to the HR porosity is considered to perform best among all SR models. Estimated porosity values and percentage errors in the estimated porosity for the super-resolved rock volumes B2 and B3 based on the HR porosity as reference are shown in Table II. It is evident that the proposed SR method (Siamese-SR) has the least absolute error in porosity (vis-a-vis the HR porosity) as compared to the other SR models. Furthermore, as can be seen from Table II, some of the other SR methods like EDSR and SPSR are not robust because they sometimes under-predict and sometimes over-predict the porosity. Particularly, under-prediction of porosity can be considered as a serious flaw in the super-resolution because the whole idea of super-resolution is to reveal the sub-resolution porosity in the image and thereby increase the porosity estimate by resolving the under-resolved pores.

Furthermore, to demonstrate that the improvement due to super-resolution using Siamese-SR is not just restricted to the porosity, but rather carries forward to the consequent steps in the Digital Rock workflow, we have also performed Mercury Injection Capillary Pressure (MICP) simulations. Since the B3 sample showed the most significant difference in the LR and HR porosities (see Table II), we have used this

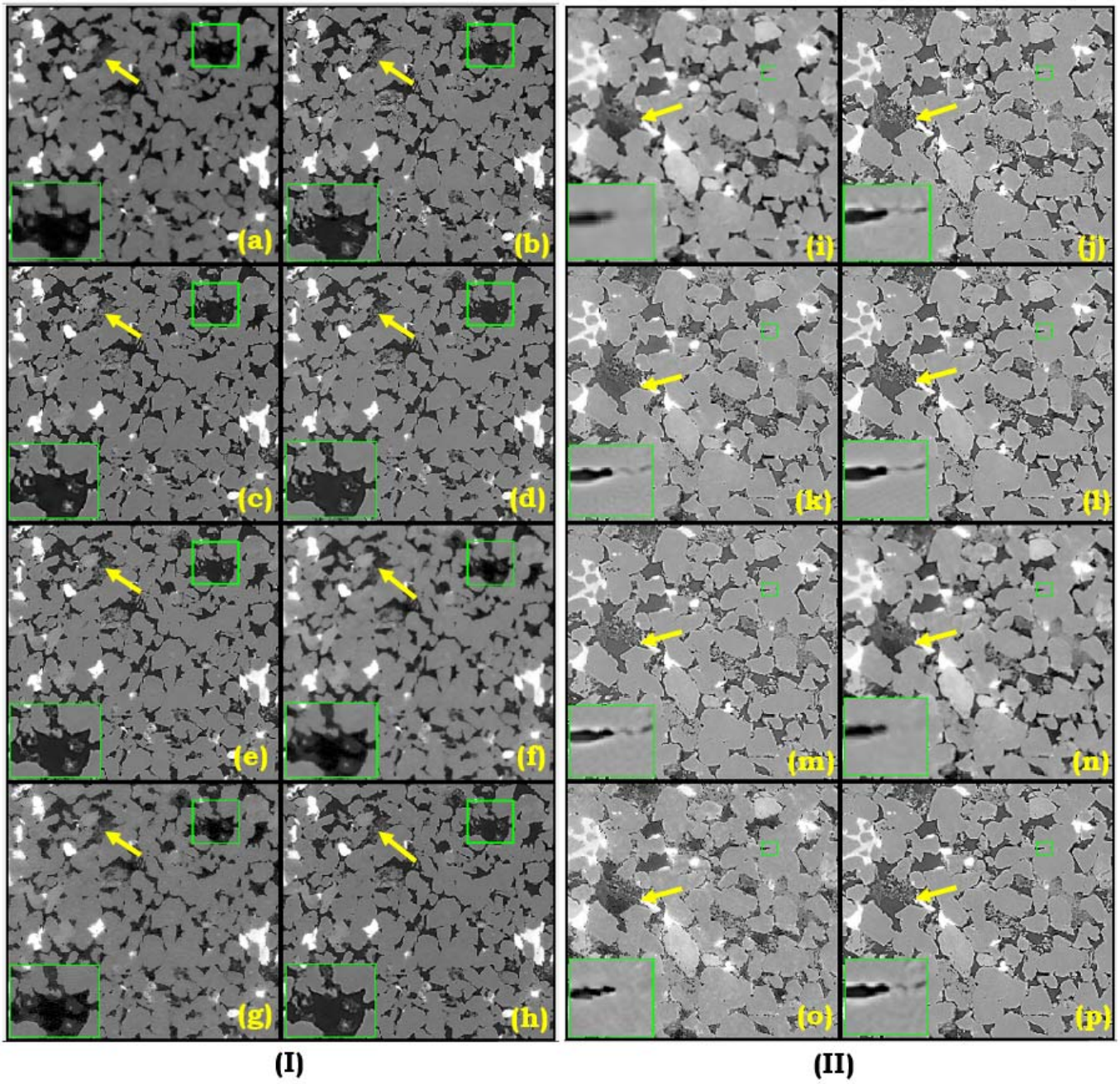

Fig. 6. Representative slice from the rock samples B3 (I) and B2 (II) for visual inspection. A region of interest (ROI) is highlighted in a green box and it's zoomed-in version is placed as an inset at the bottom left of every image. Yellow arrows indicate efficacy of Siamese-SR to capture lost micro-porosity in comparison to other models. The efficacy in capturing micro-porosity is backed by quantitative results for B3 and B2 as shown in Table:II . **[I]** (a): LR, (b): HR, (c): SRGAN, (d): ESRGAN, (e): ESRGAN-MINC, (f): EDSR, (g): SPSR, (h): Siamese-SR(Proposed). **[II]** (i): LR, (j): HR, (k): SRGAN, (l): ESRGAN, (m): ESRGAN-MINC, (n): EDSR, (o): SPSR, (p): Siamese-SR(proposed).

sample for demonstrating the effect of a better segmented image on more accurate petrophysical parameter estimation from MICP simulations and Multiple Relaxation Time Lattice Boltzmann Method (MRTLBM) flow simulations. The MICP simulations were performed using the SatuDict module of the GeoDict 2020 software package [34] with two different sets of boundary conditions. In the first case, we have used isotropic boundary conditions wherein the non-wetting phase mercury invades from Z+, Y+ and X+ directions and the wetting phase leaves from Z−, Y− and X− directions and in the second case, the Z-direction is used as the flow direction and symmetric boundary conditions are used in the other two directions. Table III summarizes the petrophysical properties of interest obtained by fitting the MICP simulation curves (shown in the supplementary information file) with the Thomeer equation [43]. It can be clearly seen that irrespective of the boundary conditions used for the MICP simulations, our proposed Siamese-SR model leads to the most accurate

TABLE III

ESTIMATED PETROPHYSICAL PROPERTIES FOR ROCK SAMPLE B3 OBTAINED FROM MICP SIMULATIONS WITH ISOTROPIC AND ANISOTROPIC BOUNDARY CONDITIONS. CLOSEST PARAMETER VALUES TO THOSE OF HR (WHICH IS USED AS REFERENCE) ARE SHOWN IN BOLD

Model	φ_∞ (Iso-) / (Aniso-tropic)	$\mathbf{P_d}$ (Iso-) / (Aniso-tropic)	$\mathbf{G}$ (Iso-) / (Aniso-tropic)
HR	0.152 / 0.152	14.0 / 14.5	0.1 / 0.1
SRGAN	0.175 / 0.175	11.0 / 11.5	0.1 / 0.1
ESRGAN	0.165 / 0.165	12.0 / 12.5	0.1 / 0.1
ESRGAN-MINC	0.16 / 0.16	13.0 / 13.5	0.1 / 0.1
EDSR	0.135 / 0.13	10.0 / 10.5	0.05 / 0.05
SPSR	0.16 / 0.16	15.0 / 15.5	0.1 / 0.1
Siamese-SR (Proposed)	**0.16 / 0.16**	**13.5 / 14.0**	**0.1 / 0.1**

TABLE IV

ESTIMATED PERMEABILITIES FOR ROCK SAMPLE B3 OBTAINED FROM MRTLBM SIMULATIONS. CLOSEST PERMEABILITY VALUE TO THE ACTUAL HR PERMEABILITY IS SHOWN IN BOLD. WE ALSO INCLUDE PERCENTAGE ERROR FOR THE ESTIMATED PERMEABILITIES CONSIDERING HR PERMEABILITY AS REFERENCE

Model	Permeability (mD)	Error (%)
HR	1.42E+02	—
SRGAN	3.26E+02	129.7
ESRGAN	2.43E+02	71.2
ESRGAN-MINC	2.03E+02	43.2
EDSR	3.10E+02	118.5
SPSR	2.00E+02	40.9
Siamese-SR (Proposed)	**1.97E+02**	**39.0**

estimation of the petrophysical properties which are closest to those obtained from the HR volumes used as reference. Table IV summarizes the permeability values obtained from MRTLBM simulations of the super-resolved outputs of the various SR models. As is clearly evident, our proposed Siamese-SR method produces permeability estimate which is closest to the HR permeability, which is used as the reference.

C. Ablation Study for the Weights of the Different Loss Functions

In this sub-section, we present a quick ablation study which demonstrates that we tried applying different weights α and β for the two loss functions as declared in Eq. 9. The images from this ablation study are shown in Fig. 7 and the quantitative results of the ablation study are summarized in Table V. As can be seen from the results, it was challenging to tune the weights in a way that we achieve the best prediction for the petrophysical properties of interest like porosity as well as permeability calculated using Multiple Relaxation Time Lattice Boltzmann Method (MRTLBM) flow simulations. Hence, we have chosen to weigh the two loss functions equally, without any loss of generality.

D. Ablation Study for the Impact of Different Loss Functions

In this sub-section, we present an ablation study to highlight the impact of the different loss functions that make up Siamese-SR. The images from the ablation study are presented in Fig. 8 and the quantitative results of the ablation study have been presented in Table VI. It can be clearly seen from the results that while the addition of the MINC loss and the Siamese discriminator individually help improve the performance, its actually their combination that makes Siamese-SR so powerful in terms of accurate prediction of petrophysical properties of interest such as porosity as well as permeability calculated using Multiple Relaxation Time Lattice Boltzmann Method (MRTLBM) flow simulations.

E. Local Attribution Maps for Evaluation of SR Methods

In this sub-section, we present the use of Local Attribution maps (LAM) [45], which are a recently developed tool used to evaluate super-resolution networks. LAMs have been developed to essentially find input features that strongly influence the network outputs. Basically, path integral gradients are employed to conduct attribution analysis as described in [45]. Although, this is not a quantitative image-based metric like PSNR or SSIM (which we have already highlighted as metrics that are not applicable due to various reasons mentioned in Section III-D), but to the best of our knowledge, LAMs are the only image-based evaluation technique for super-resolution available in the literature that ties well with the features of interest (edge-semantics) that we are interested in enhancing with super-resolution.

Firstly, to deep dive into the impact of various losses from the image enhancement perspective, we have shown Local Attribute Maps for the Siamese-SR model with both the MINC loss and the Siamese discriminator included as well as by excluding one or both of these. It can be clearly seen from the results shown in Fig. 9 that while the addition of the MINC loss and the Siamese discriminator individually help focus on different types of features, its actually their combination that makes Siamese-SR optimally focus on edge-semantics, which

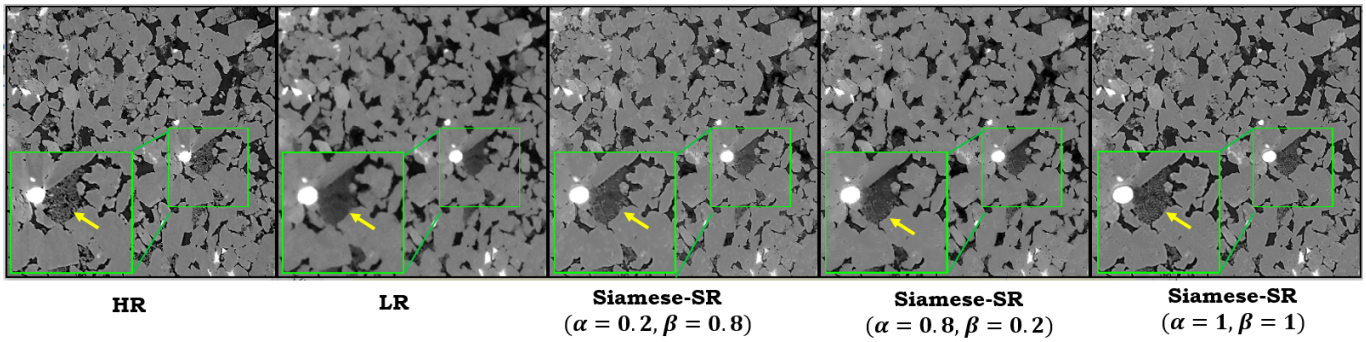

Fig. 7. Ablation study highlighting the effect of different weights for the different losses of our proposed Siamese-SR model on a representative slice from the rock sample B3. A Region of interest (ROI) is highlighted in a green box and its zoomed-in version is placed as an inset at the bottom left of every image. The quantitative results (porosity and permeability) corresponding to each sub-figure are shown in Table V.

TABLE V

ABLATION STUDY HIGHLIGHTING THE EFFECT OF DIFFERENT WEIGHTS FOR MINC LOSS AND L1 LOSS ON ESTIMATED POROSITY AND PERMEABILITY. CLOSEST POROSITY AND PERMEABILITY VALUES TO THE ACTUAL HR POROSITY AND PERMEABILITY ARE SHOWN IN BOLD

Model	α	β	Porosity (p.u.) / Error (%)	Permeability (mD) / Error (%)
HR	—	—	12.99 / —	1.42E+02 /—
Siamese-SR	0.8	0.2	14.13 / 8.8	2.53E+02 / 78.0
Siamese-SR	0.2	0.8	15.89 / 22.3	3.60E+02 / 153.5
Siamese-SR (Proposed)	**1**	**1**	**13.66 / 5.2**	**1.97E+02/ 39.0**

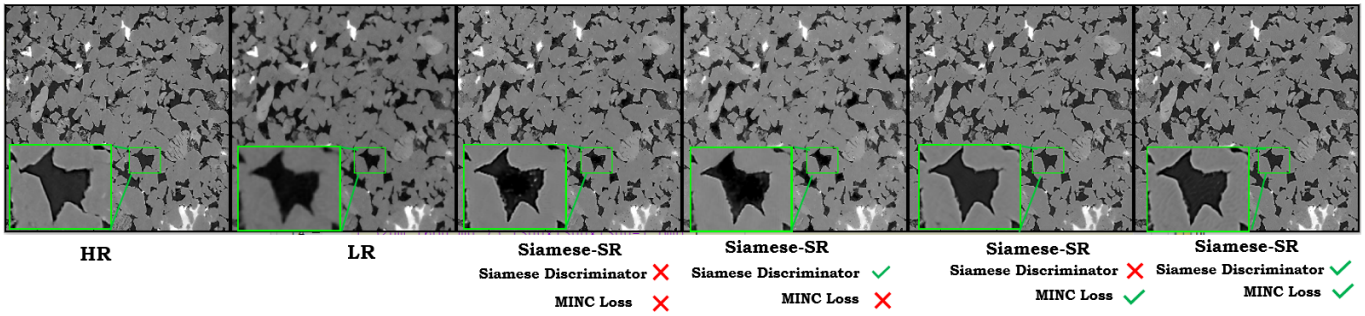

Fig. 8. Ablation study highlighting the effect of different losses on a representative slice from the rock sample B3. A Region of interest (ROI) is highlighted in a green box and its zoomed-in version is placed as an inset at the bottom left of every image. The quantitative results (porosity and permeability) corresponding to each sub-figure are shown in Table VI.

is what makes it so powerful in terms of accurate prediction of petrophysical properties of interest.

Secondly, for highlighting the difference between our proposed Siamese-SR model and other methods in terms of optimal focus on the features of interest, we have presented LAMs for the different super-resolution methods in Fig. 10. As can be seen from the comparison of the LAMs of our proposed Siamese-SR with the other methods, our proposed Siamese-SR method provides optimal focus on edge-semantics, which is what essentially translates to a more accurate porosity estimate, which is the metric that tallies well with expert petrophysical and geological opinion.

V. DISCUSSION

The main difference between the porosities estimated from LR and HR images stems from the fact that some of the micro-porosity is missed when the sample is scanned at a lower resolution and it only becomes visible at a higher resolution. This micro-porosity gets hidden in the LR images due to the coarser voxel size leading to the regions with the micro-pores getting smudged out with an intermediate intensity value which leads to misclassification during segmentation. Therefore, the task of the super-resolution models is to capture well these regions of micro-porosity from the LR image and yield a super-resolution image that has

TABLE VI

ABLATION STUDY HIGHLIGHTING THE EFFECT OF DIFFERENT LOSSES ON ESTIMATED POROSITY AND PERMEABILITY. CLOSEST POROSITY VALUES TO THE ACTUAL HR POROSITY AND PERMEABILITY ARE SHOWN IN BOLD. WE ALSO INCLUDE PERCENTAGE ERROR CONSIDERING HR POROSITY AND PERMEABILITY AS REFERENCE

Model	Siamese Discriminator	MINC Loss	Porosity (p.u.) / Error (%)	Permeability (mD) / Error (%)
HR	—	—	12.99 / —	1.42E+02 / —
Siamese-SR	✗	✗	15.19 / 16.9	2.99E+02 / 111.0
Siamese-SR	✓	✗	14.24 / 9.6	2.46E+02 / 73.1
Siamese-SR	✗	✓	13.82 / 6.4	2.03E+02 /43.2
Siamese-SR (Proposed)	✓	✓	**13.66 / 5.2**	**1.97E+02 /39.0**

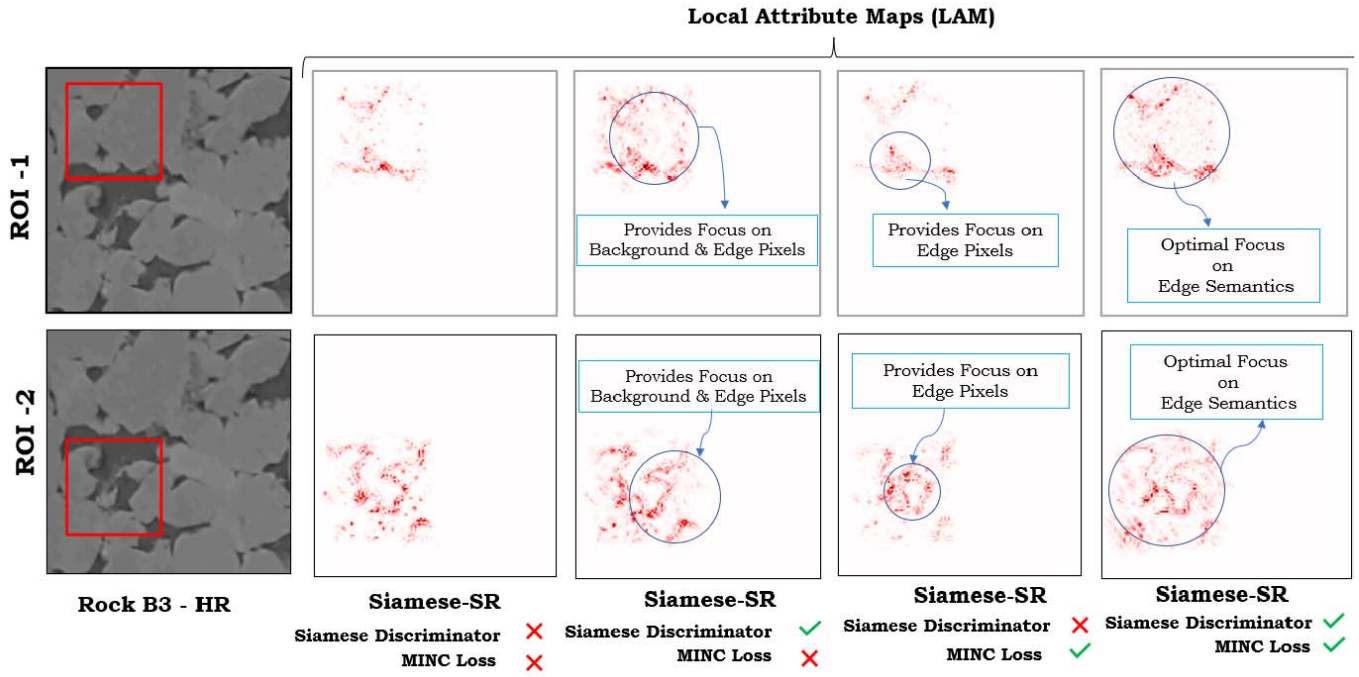

Fig. 9. Local attribution maps for two different marked-out Regions of interest (ROI-1 and ROI-2) in a super-resolved slice of rock sample B3. Qualitatively, our proposed Siamese-SR model (with both siamese siscriminator and MINC loss), achieves the optimal focus between background and edge semantics compared to other ablation settings. This is also quantitatively backed by computation of porosity and permeability values across various ablation settings as shown in Table VI.

more or less the same structure as the HR image. Since this is quite a challenging task, its understandable that the super-resolution methods are not exactly able to match the HR image porosity and do tend to slightly over-predict the porosity in the cases that we have studied but we clearly see that the proposed Siamese SR model gives the closest porosity to the HR image as can be seen from Table II). Furthermore, the superior performance of the Siamese-SR model is also evident from the accuracy of the petrophysical properties estimated via MICP simulation curves as well as permeabilities estimated from MRTLBM simulations, as can be seen from Table III and Table IV, which is a testimony to the hypothesis that a better segmented image is crucial for the better prediction of petrophysical properties from the Digital Rock workflow.

Using LR and HR images acquired separately from different micro-CT acquisitions introduces challenges like intensity variation between the LR and HR pairs as well slight differences in noise characteristics. It is to overcome these challenges that we have introduced the siamese discriminator in our Siamese-SR network, which is based on similarity learning network, also known as siamese twin network, which is robust enough to provide optimally denoising and accounting for intensity variation. The siamese discriminator checks whether the generated super-resolution (SR) image has the required semantic features such as edge information similar to the

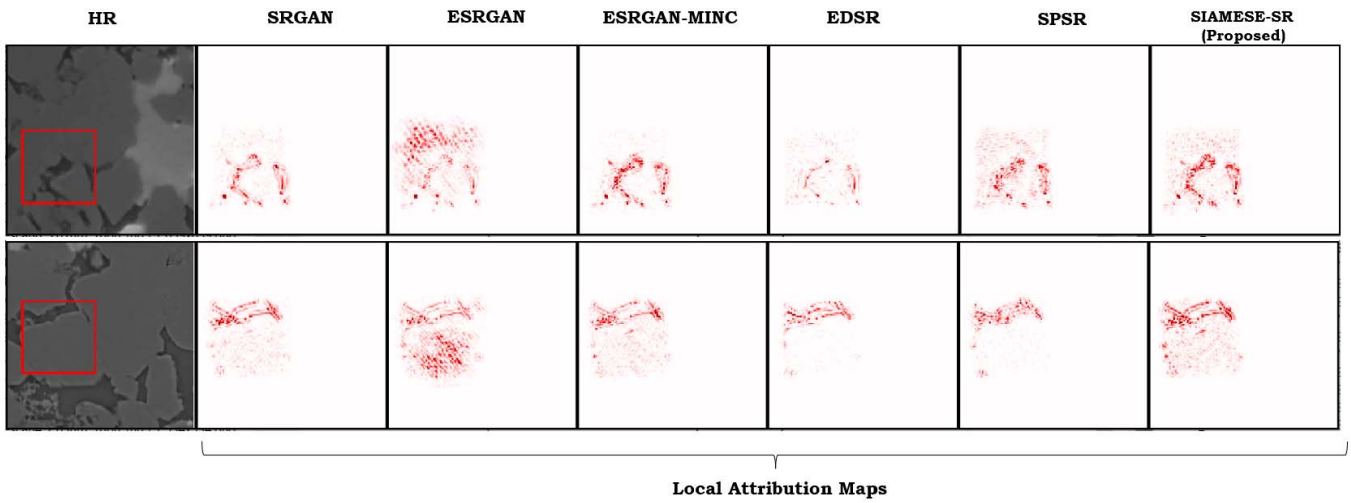

Fig. 10. Local attribution Maps of the highlighted Regions of interest (ROIs) for two sets of super-resolved images of rock sample B3 from various super-resolution methods. Qualitatively, our proposed model Siamese-SR achieves the optimal focus between background and edge semantics when compared to other models. This is quantitatively backed by petrophysical properties of interest shown in Table II, Table III and Table IV.

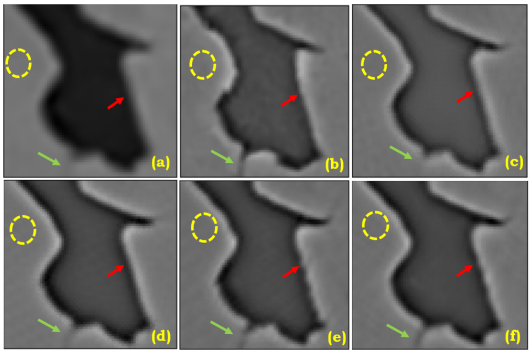

Fig. 11. Qualitative performance study showing the various super-resolution models alongside the LR and HR images as reference (a): LR, (b): HR, (c): SRGAN, (d): ESRGAN, (e): ESRGAN-MINC, (f): Siamese-SR(proposed). Yellow circle, Red arrow, Green arrow indicates texture preservation, Sharp edge profile and edge preservation respectively.

actual HR image as compared to simply determining if the SR image is real or fake image like a normal discriminator. To elucidate this further with an example, we have shown a zoomed-in part of a slice from the B2 rock sample obtained from the various super-resolution methods in Fig. 11. It can be seen that Siamese-SR turns out to be the optimal network in terms of texture preservation, sharpness of the pore-grain boundary as well as edge preservation. We observe that SRGAN and ESRGAN network output have very smooth edge profile at the grain-pore boundary (as indicated by red arrows in Fig. 11(c) and (d)) because these networks tend to produce over-smoothed images which compromises on the texture and loss of finer details (as indicated by the yellow circle and green arrows in Fig. 11 (c) and (d)). To retain these finer details and texture, the MINC loss was introduced instead of conventional perceptual loss while training the ESRGAN-MINC network. The MINC loss tends to bring back the necessary texture and finer details in the output (as indicated by green arrow and a yellow circle in Fig. 11(e)) but it tends to compromise on the sharpness at the grain-pore boundary by making the image slightly more noisy (as indicated by the red arrow in Fig. 11 (e)). The addition of the siamese discriminator provides optimal denoising without compromising finer details and also preserves the texture in the image as can be seen in Fig. 11(f).

VI. Conclusion & Scope for Future Work

We have proposed a novel super-resolution model called Siamese-SR for boosting the resolution of 3D Digital Rock images acquired using micro-CT scanning technology. The proposed Siamese-SR model improves upon the ESRGAN architecture by the addition of a siamese discriminator which provides optimal de-noising and captures important semantic features present in the rock images. The efficacy of the proposed model is backed both qualitatively using visual inspection by experts as well as quantitatively by analyzing the accuracy of the estimation of petrophysical properties of interest by subjecting the super-resolved images to the physics simulations in the consequent steps of the Digital Rock workflow. The petrophysical properites of interest that we estimate are image-based porosity, permeability from MRTLBM flow simulations and Capillary Pressure curve and other derived properties from MICP simulations.Thus, we have proposed a move away from image based metrics and rather proposed to use the quantitative metrics from the application domain based on the estimation of petrophysical properties. Furthermore, we have also used Local Attribution Maps to highlight the optimal focus of our proposed Siamese-SR model on edge-semantics which leads to our Siamese-SR model outperforming several existing SR methods like SRGAN, ESRGAN, EDSR and SPSR in terms of accurate estimation of the above-mentinoed petrophysical properties of interest. While we have applied our model to Digital Rock Physics, it is important to note that our model can also benefit the

medical image processing community, where reconstructing accurate structures and features of interest is more important than simply the perceptual quality.

A natural next step would be to extend the Siamese-SR model so that it can include the 3D context present in the 3D micro-CT volumes while training instead of training on 2D slices extracted from the volume like we have done in this present study. Hence, we are working towards exploring 3D models for Super Resolution and evaluating the same for their integration into the workflow. The 3D super-resolution models operate on 3D mini-cubes instead of 2D slices and therefore should be able to provide even more enhanced Super-Resolution in principle because they are capable of learning from the 3D context and also produce a 3D output with lesser artifacts. This will also take away the anisotropy in the way 2D super-resolution is applied by slicing along one direction and then training the model slice-by-slice. However, there are also hardware limitations based on GPU memory which can limit the size of the mini-cubes that can be used for training, let alone the higher compute times associated with the training. Thus, the application of 3D models for boosting the resolution of Digital Rock images will be separately evaluated in a future study.

ACKNOWLEDGMENTS

The authors thank Shell International Exploration and Production Inc., for permission to publish this work. They would like to thank specially Justin Freeman, Kunj Tandon, Steffen Berg, F. O. Alpak, Bochao Zhao, and Chaitanya Pradhan for discussions and Pandu Devarakota for his critical review, which significantly improved the paper.

CODE, DATA, AND MATERIALS AVAILABILITY

The paper has supplementary downloadable material provided by the authors, which includes the algorithm for training the proposed Siamese-SR model and the Mercury Injection Capillary Pressure (MICP) simulation curves. The associated code and data that support the findings of this study are available from the corresponding author upon reasonable request. Contact vishal.ahuja@shell.com for further questions about this work.

REFERENCES

[1] H. Al-Marzouqi, "Digital rock physics: Using CT scans to compute rock properties," *IEEE Signal Process. Mag.*, vol. 35, no. 2, pp. 121–131, Mar. 2018.

[2] N. Saxena, R. Hofmann, F. O. Alpak, J. Dietderich, S. Hunter, and R. J. Day-Stirrat, "Effect of image segmentation & voxel size on micro-CT computed effective transport & elastic properties," *Mar. Petroleum Geol.*, vol. 86, pp. 972–990, Sep. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0264817217302593

[3] Z. Li, Q. Teng, X. He, G. Yue, and Z. Wang, "Sparse representation-based volumetric super-resolution algorithm for 3D CT images of reservoir rocks," *J. Appl. Geophys.*, vol. 144, pp. 69–77, Sep. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0926985117303592

[4] C. Qiao *et al.*, "Evaluation and development of deep neural networks for image super-resolution in optical microscopy," *Nature Methods*, vol. 18, no. 2, pp. 194–202, Feb. 2021.

[5] Z. Chen, X. Guo, P. Y. M. Woo, and Y. Yuan, "Super-resolution enhanced medical image diagnosis with sample affinity interaction," *IEEE Trans. Med. Imag.*, vol. 40, no. 5, pp. 1377–1389, May 2021.

[6] F. Deeba *et al.*, "A plexus-convolutional neural network framework for fast remote sensing image super-resolution in wavelet domain," *IET Image Process.*, vol. 15, no. 8, pp. 1679–1687, Jun. 2021.

[7] Y. D. Wang, R. T. Armstrong, and P. Mostaghimi, "Enhancing resolution of digital rock images with super resolution convolutional neural networks," *J. Petroleum Sci. Eng.*, vol. 182, Nov. 2019, Art. no. 106261.

[8] H. Chen, X. He, Q. Teng, R. E. Sheriff, J. Feng, and S. Xiong, "Super-resolution of real-world rock microcomputed tomography images using cycle-consistent generative adversarial networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 101, no. 2, Feb. 2020, Art. no. 023305.

[9] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2015.

[10] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1637–1645.

[11] W. Shi *et al.*, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1874–1883.

[12] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1132–1140.

[13] J. Yu *et al.*, "Wide activation for efficient and accurate image super-resolution," in *Proc. 30th Brit. Mach. Vision Conf. (BMVC)*, 2018.

[14] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th Asilomar Conf. Signals, Syst. Comput.*, vol. 2, Jul. 2003, pp. 1398–1402.

[15] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[16] P. Gupta, P. Srivastava, S. Bhardwaj, and V. Bhateja, "A modified PSNR metric based on HVS for quality assessment of color images," in *Proc. Int. Conf. Commun. Ind. Appl.*, Dec. 2011, pp. 1–4.

[17] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 105–114.

[18] X. Wang *et al.*, "ESRGAN: Enhanced super-resolution generative adversarial networks," in *The Eur. Conf. Comput. Vis. Workshops (ECCVW)*, Sep. 2018, pp. 1–16.

[19] A. Jolicoeur-Martineau, "The relativistic discriminator: A key element missing from standard GAN," 2018, *arXiv:1807.00734*.

[20] S. Bell, P. Upchurch, N. Snavely, and K. Bala, "Material recognition in the wild with the materials in context database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3479–3487.

[21] C. Ma, Y. Rao, Y. Cheng, C. Chen, J. Lu, and J. Zhou, "Structure-preserving super resolution with gradient guidance," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 7769–7778.

[22] C. Ma, Y. Rao, J. Lu, and J. Zhou, "Structure-preserving image super-resolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Sep. 22, 2021, doi: 10.1109/TPAMI.2021.3114428.

[23] N. A. Kande, R. Dakhane, A. Dukkipati, and P. K. Yalavarthy, "SiameseGAN: A generative model for denoising of spectral domain optical coherence tomography images," *IEEE Trans. Med. Imag.*, vol. 40, no. 1, pp. 180–192, Jan. 2021.

[24] G. Koch, R. Zemel, and R. Salakhutdinov, "Siamese neural networks for one-shot image recognition," in *Proc. ICML Deep Learn. Workshop*, Lille, France, vol. 2, 2015, pp. 1–30.

[25] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "A comprehensive evaluation of full reference image quality assessment algorithms," in *Proc. 19th IEEE Int. Conf. Image Process.*, Sep. 2012, pp. 1477–1480.

[26] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Trans. Comput. Imag.*, vol. 3, no. 1, pp. 47–57, Mar. 2017.

[27] Y. D. Wang, R. T. Armstrong, and P. Mostaghimi, "Enhancing resolution of digital rock images with super resolution convolutional neural networks," *J. Petroleum Sci. Eng.*, vol. 182, Nov. 2019, Art. no. 106261. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0920410519306825

[28] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 4278–4284.

[29] V. Chudasama and K. Upla, "Computationally efficient progressive approach for single-image super-resolution using generative adversarial network," *J. Electron. Imag.*, vol. 30, no. 2, pp. 1–36, Jan. 2021, doi: 10.1117/1.JEI.30.2.021003.

[30] K. Zhang *et al.*, "NTIRE 2020 challenge on perceptual extreme super-resolution: Methods and results," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 2045–2057.

[31] Z. Wang, J. Chen, and S. Hoi, "Deep learning for image super-resolution: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 10, pp. 3365–3387, Mar. 2020.

[32] N. Saxena *et al.*, "Rock properties from micro-CT images: Digital rock transforms for resolution, pore volume, and field of view," *Adv. Water Resour.*, vol. 134, Dec. 2019, Art. no. 103419. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0309170819301435

[33] W. R. Purcell, "Capillary pressures–their measurement using mercury and the calculation of permeability therefrom," *J. Petroleum Technol.*, vol. 1, no. 2, pp. 39–48, Feb. 1949.

[34] L. L. Schepp *et al.*, "Digital rock physics and laboratory considerations on a high-porosity volcanic rock," *Sci. Rep.*, vol. 10, no. 1, pp. 1–16, 2020.

[35] D. d'Humieres, I. Ginzburg, M. Krafczyk, P. Lallemand, and L. S. Luo, "Multiple–relaxation–time lattice Boltzmann models in three dimensions," *Philos. Trans. Roy. Soc. London A, Math., Phys. Eng. Sci.*, vol. 360, no. 1792, pp. 437–451, 2002.

[36] F. O. Alpak, F. Gray, N. Saxena, J. Dietderich, R. Hofmann, and S. Berg, "A distributed parallel multiple-relaxation-time lattice Boltzmann method on general-purpose graphics processing units for the rapid and scalable computation of absolute permeability from high-resolution 3D micro-CT images," *Comput. Geosci.*, vol. 22, no. 3, pp. 815–832, Jun. 2018.

[37] E. M. Withjack, "Computed tomography for rock-property determination and fluid-flow visualization," *SPE Formation Eval.*, vol. 3, no. 4, pp. 696–704, Dec. 1988.

[38] S. Akin, M. Demiral, and E. Okandan, "A novel method of porosity measurement utilizing computerized tomography," *In Situ*, vol. 20, no. 4, pp. 347–365, 1996.

[39] H. Taud, R. Martinez-Angeles, J. F. Parrot, and L. Hernandez-Escobedo, "Porosity estimation method by X-ray computed tomography," *J. Petroleum Sci. Eng.*, vol. 47, nos. 3–4, pp. 209–217, Jun. 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0920410505000586

[40] L. L. Schepp *et al.*, "Digital rock physics and laboratory considerations on a high-porosity volcanic rock: Micro-XRCT data sets," Math2Market GmbH, Berlin, Germany, Tech. Rep., 2020.

[41] N. Saxena *et al.*, "Rock properties from micro-CT images: Digital rock transforms for resolution, pore volume, and field of view," *Adv. Water Resour.*, vol. 134, Dec. 2019, Art. no. 103419.

[42] K.-S. Chuang, H.-L. Tzeng, S. Chen, J. Wu, and T.-J. Chen, "Fuzzy c-means clustering with spatial information for image segmentation," *Comput. Med. Imag. Graph.*, vol. 30, pp. 9–15, Jan. 2006.

[43] J. H. M. Thomeer, "Introduction of a pore geometrical factor defined by the capillary pressure curve," *J. Petroleum Technol.*, vol. 12, no. 3, pp. 73–77, Mar. 1960, doi: 10.2118/1324-G.

[44] P. L. Bhatnagar, E. P. Gross, and M. Krook, "A model for collision processes in gases. I. Small amplitude processes in charged and neutral one-component systems," *Phys. Rev.*, vol. 94, p. 511, May 1954.

[45] J. Gu and C. Dong, "Interpreting super-resolution networks with local attribution maps," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2021, pp. 9199–9208.